université de BORDEAUX

THÈSE PRÉSENTÉE POUR OBTENIR LE GRADE DE

DOCTEUR DE L'UNIVERSITÉ DE BORDEAUX

ÉCOLE DOCTORALE SCIENCES DE LA VIE ET DE LA SANTE

SPÉCIALITÉ BIOCHIMIE

Par Lucie GRINDES

OPTIMISATION DE LA PLATEFORME CHOZN® CHO K1 DE GENERATION DE LIGNEES CELLULAIRES CLONALES POUR LA PRODUCTION DE PROTEINES RECOMBINANTES A USAGE THERAPEUTIQUE

Sous la direction de Gilles JOUCLA

Co-encadrée par Cyrielle CORBIN et Wilfrid DIERYCK

Soutenue le 7 Juillet 2022

Membres du jury :

M. JOUCLA, Gilles	Maître de conférences, CBMN	Directeur de thèse
M. DUROCHER, Yves	Professeur, CNRC	Rapporteur
M. GAILLET, Bruno	Professeur, Université de Laval	Rapporteur
Mme CLOFENT-SANCHEZ, Gisèle	Directrice de recherche, CNRS	Examinatrice, Présidente du jury
M. DIERYCK, Wilfrid	Maître de conférences, CBMN	Examinateur
Mme FRENTZEL, Julie	Docteure, Merck	Examinatrice
Mme CORBIN, Cyrielle	Docteure, Merck Biodevelopment	Invitée
M. RIBAULT, Sébastien	Docteur, Merck Biodevelopment	Invité

Remerciements

Je remercie tout d'abord ma co-encadrante de thèse, Cyrielle Corbin, de m'avoir toujours accordé le temps nécessaire pour mes innombrables questions, d'avoir mis autant d'énergie et de bienveillance pour m'encadrer. Je remercie aussi Céline Raymond, d'avoir participé à l'encadrement de la thèse. Toutes les deux, d'avoir su me guider dans les méandres de la thèse.

Je remercie mon directeur de thèse, Gilles Joucla, et mon co-encadrant, Wilfrid Dieryck, d'avoir repris le train en marche, de m'avoir toujours accordé du temps et de m'avoir encouragée dans les dernières semaines d'écriture du manuscrit.

Je remercie Sébastien Ribault d'avoir régulièrement participé au suivi de la thèse.

Je remercie les membres du jury, monsieur Yves Durocher et monsieur Bruno Gaillet, d'avoir accepté d'être rapporteurs ; madame Gisèle Clofent Sanchez, madame Julie Frentzel, et Wilfrid Dieryck, d'avoir accepté d'être examinateurs ; Cyrielle Corbin et Sébastien Ribault, d'y participer en tant qu'invités, et Gilles Joucla – et tous, de me faire l'honneur d'évaluer mon travail.

Je remercie chaleureusement mes collègues de Martillac, et plus particulièrement celles et ceux du département « DEVT LS » :

L'équipe CLD, qui m'a accueillie, formée à la CLD, ô combien de fois aidée, gentiment attribué le surnom de « Madame Plus » et qui m'a soutenue jusqu'au bout de la thèse. Je remercie donc Marine Amouriq, Mathilde Bas, Olivia Clayton, Florence Courtin, Barbara Couturaud, Cyrielle Corbin, Raphaëlle Dréan, Marilyne Failly, Camille Florimond, Clémence Justine, Emilie Navarro, Amandine Philippat, Jeanne Rouilleaux, Remy Van Geersdaele, et Murielle Vergès. Je vous en suis extrêmement reconnaissante, bien au-delà des mots.

L'équipe USP, qui m'a aussi beaucoup aidée et soutenue jusqu'au bout de la thèse. Je remercie donc Elodie Airola, Myriam Eyquard, Charlène François, Paco Hernandez, Audrey Ladonne, Charlotte Ott, Margaux Paillet, Céline Raymond, et Fanny Tessier.

L'équipe DSP, les joyeux lurons du fond de l'open-space, qui m'ont beaucoup faite rire. Je remercie donc Clément Bertrand, Emilie Cester, Arthur Destruhaut, Calypso Devigne, Guillaume Godonnier, Manon Lalanne-Claux, Alice Laly, Arthur Leclercq, Marjorie Lèvefaudes, Caroline Pennacino, Julien Raud, et Magali Toueille.

L'équipe Support, qui a fait preuve d'une aide immense, et relevant parfois du miracle, pour la préparation de milieux de culture, la saisie et la réception de commandes, et j'en passe. Je remercie donc Denis Deschamps, Stéphanie Jacquere, Hervé Magnier, et Bertrand Oeung.

Enfin, l'équipe Compliance, pour les innombrables voyages à la salle des banques, par tout temps. Je remercie donc Nora Kadi et Eugénie Thomas.

Je remercie d'autres collègues du site de Martillac, notamment : Thomas Clairembault, Christine Dinard et Isabelle Vimes du « DEVT ANA » ; Vanessa Martin et Anaïs Reignier de la « PV ANA » ; Eric Basque, Agathe Jonckeau, Cédric Mesmin et Jérémy Peyrol d'« INNOV » – qui m'ont ouvert les portes de leurs laboratoires et mis à disposition leurs équipements, qui m'ont maintes fois prêté main forte au laboratoire, et avec qui j'ai partagé d'intéressantes discussions scientifiques ou professionnelles. Je remercie aussi spécialement Vivien Le Bras et ses « réponses de Normand », qui m'a d'abord mise sur la sinueuse route des statistiques et qui m'a ensuite toujours accordé du temps pour répondre à mes questions de jeune novice.

Je remercie des collègues des Etats-Unis, particulièrement Kate Achtien, Britta Anderson, Scott Bahr, Wilson Fok, Jeffrey Galligan, Amber Petersen, James Ravelette, et David Razafsky ; de Suisse, particulièrement Julien Douet et Julie Frentzel ; d'Italie, notamment Luisa Marena ; d'Ecosse, notamment Scott McGhee ; et de Chine, notamment Weiwei Li et Vangie Wu – ces collègues que j'ai tant de fois sollicités, avec qui j'ai partagé d'intéressantes discussions scientifiques, et qui m'ont toujours apporté une aide précieuse et bienveillante.

Je remercie aussi chaleureusement les chercheurs de l'équipe SAMB : Norbert Bakalara, Charlotte Cabanne, Jean Chaudière, Agnès Hocquellet, Béatrice Langlois d'Estaintot, Majid Noubhani, Xavier Santarelli, de faire de l'école un endroit où il fait toujours bon revenir.

A défaut d'avoir côtoyé d'autres doctorants pendant ces années, j'ai pu compter sur le soutien, l'écoute et bien-sûr les anecdotes des docteurs qui m'entouraient. Je remercie donc particulièrement, et à nouveau, Cyrielle Corbin, Céline Raymond, Agathe Jonckeau et Mathilde Bas.

Je remercie Marie, Mérydith, les « Pékin Express » et les « TeleTuBBies », ces amies et amis inestimables par qui j'ai eu la chance d'être soutenue.

Je remercie Papa, Maman et Pauline, de m'avoir donné une famille tant aimante et de constamment m'apporter un indéfectible soutien. J'ai aussi une affectueuse pensée pour mes grands-parents.

Enfin, je remercie tendrement Guillaume, d'avoir partagé toutes ces années avec moi, de croire en moi avec une telle ténacité, de m'apporter autant de tendresse, et de m'ouvrir de si beaux horizons.

Résumé

Introduction

La production de protéines recombinantes thérapeutiques est une évolution majeure, résultant notamment d'avancées capitales en biologie moléculaire dans les années 1970. La production de protéines recombinantes thérapeutiques a en particulier permis de répondre aux limitations de la production de protéines thérapeutiques par des sources naturelles (au regard, notamment, des quantités produites et de la sécurité sanitaire) et de développer de nouveaux médicaments (grâce, en particulier, aux techniques d'ingénierie des protéines). Depuis l'introduction de l'insuline recombinante en 1982, 365 produits biopharmaceutiques ont reçu l'approbation des autorités sanitaires pour leur mise sur le marché. Les anticorps monoclonaux, qui dominent désormais le marché des protéines recombinantes à application thérapeutique, sont des molécules de haut poids moléculaire, à la structure complexe, dont les propriétés dépendent fortement du processus de production et qui sont potentiellement immunogènes. Depuis le milieu des années 2010, une grande variété de molécules, au format souvent dérivé d'anticorps (fragments, conjugués, ou multimères) sont en cours de développement.

Une protéine recombinante résulte de l'expression d'un ADN recombinant par une cellule ou un organisme hôte. Un système de production de protéines recombinantes est basé sur (i) un gène d'intérêt (GOI, « gene of interest »), codant pour la protéine d'intérêt (POI, « protein of interest ») ; (ii) un vecteur d'expression ou plasmide, contenant les éléments nécessaires à l'expression du GOI et éventuellement les éléments nécessaires à la sélection de clones stables ; (iii) un système hôte d'expression, assurant la production de la protéine recombinante. Le système d'expression pour la production d'un anticorps monoclonal recombinant doit être choisi avec soin. Les cellules de mammifères sont capables de produire des protéines complexes, présentant, notamment dans le cas des anticorps, des profils de glycosylation proches ou identiques à ceux rencontrés chez l'homme. Plus spécifiquement, les cellules CHO (« Chinese Hamster Ovary »), dont la vaste généalogie offre un large choix de lignées cellulaires, présentent des caractéristiques avantageuses pour la production de protéines recombinantes pour des applications thérapeutiques. Ainsi, la majorité des plateformes industrielles de développement de lignées cellulaires repose sur l'utilisation de lignées cellulaires CHO, en particulier les lignées CHO K1, CHO DHFR$^{-/-}$, CHO GS$^{-/-}$, et CHO S.

La génération d'une lignée cellulaire est la première étape du développement du procédé de production d'une protéine recombinante à usage thérapeutique. Cette étape, décisive à de nombreux points de vue, assure le développement d'une lignée cellulaire clonale, productrice de la protéine recombinante et ayant un niveau d'expression stable dans le temps. Une plateforme de génération de lignées cellulaires est composée d'une lignée hôte, d'un vecteur d'expression et d'un procédé de génération de lignée.

Les vecteurs d'expression pour l'expression stable de gènes dans les cellules de mammifères sont des vecteurs plasmidiques de quelques kilobases, qui contiennent des éléments génétiques pour : (i) l'expression et la biosynthèse de la protéine recombinante d'intérêt, (ii) l'expression d'un marqueur de sélection, qui confère un avantage sélectif aux cellules ayant intégré le plasmide, et (iii) la production du plasmide au sein du système bactérien. Les éléments permettant l'expression, soit du gène d'intérêt, soit du marqueur de sélection, sont rassemblés au sein d'une cassette d'expression, qui contient des éléments génétiques permettant le contrôle transcriptionnel et traductionnel. De nombreuses stratégies d'optimisation du vecteur d'expression ont été décrites dans la littérature : optimisation des éléments de contrôle transcriptionnel et traductionnel pour augmenter le niveau d'expression de la protéine recombinante, optimisation des éléments de contrôle transcriptionnel et traductionnel pour augmenter la stabilité du niveau d'expression de la protéine recombinante, optimisation des éléments de contrôle transcriptionnel et traductionnel pour atténuer la force du marqueur de sélection (dans une stratégie d'augmentation du niveau d'expression de la protéine recombinante), et stratégies d'intégration semi-ciblée ou ciblée du vecteur d'expression dans le génome de la lignée hôte.

Depuis les années 1980, les concepts fondamentaux du développement de lignées cellulaires sont restés inchangés. Ainsi, le développement de lignées cellulaires repose sur un schéma bien établi, qui commence par la transfection de la lignée cellulaire hôte avec un vecteur d'expression. La transfection est la première étape clef de tout processus de génération de lignées cellulaires, car elle permet l'apport intracellulaire de matériel génétique exogène nécessaire à la production de la protéine d'intérêt. Un large éventail de techniques est utilisé pour la transfection, y compris l'électroporation, la nucléofection, la transfection au phosphate de calcium, la lipofection, le transfert de gènes à médiation biolistique et polymère. Après transfection, les cellules qui ont intégré le vecteur d'expression dans leur génome sont sélectionnées, soit en l'absence d'un métabolite essentiel, soit en présence d'un antibiotique. La sélection assure la génération de lignées cellulaires stables, par opposition aux lignées cellulaires transitoires dans lesquelles l'ADN recombinant n'a pas été intégré dans le génome de la lignée cellulaire hôte et sera donc perdu lors de la division cellulaire. Les cellules survivantes sont ensuite isolées individuellement. Le clonage unicellulaire est la deuxième étape clef de tout processus de génération de lignées cellulaires, car il garantit la génération de lignées cellulaires dérivées de façon clonale à partir d'une unique cellule. Un large éventail de techniques est utilisé pour le clonage unicellulaire, notamment la dilution limite (LDC, « limiting dilution cloning »), l'encapsulation microfluidique, le tri cellulaire activé par fluorescence (FACS, « fluorescence activated cell-sorting »), le prélèvement de colonies dans des milieux semi-solides, les ensemenceurs unicellulaires et les puces microfluidiques. Les clones ainsi générés sont ensuite évalués sur des critères de croissance et de production de la protéine d'intérêt (quantité et qualité). A l'issue de ces évaluations, parfois itératives, une lignée cellulaire présentant les caractéristiques souhaitées de croissance, de productivité, et éventuellement de qualité du produit est sélectionnée. Cette lignée cellulaire sera la lignée cellulaire productrice de la protéine recombinante. La

génération de lignées cellulaires recombinantes est un procédé long, qui peut nécessiter jusqu'à six mois, de la transfection à la sélection finale des clones. Enfin, l'étude de stabilité du niveau d'expression des protéines recombinantes (également appelée étude de stabilité des clones) est une autre étude clef liée au développement des lignées cellulaires. L'étude de stabilité des clones vise à déterminer la stabilité du niveau d'expression de la protéine recombinante, et donc à définir les contraintes possibles pour les phases d'expansion et de production.

Le besoin croissant en protéines recombinantes et la compétitivité du marché ont fait de l'optimisation des plateformes de génération de lignées cellulaires un véritable défi pour l'industrie biopharmaceutique.

L'ensemble de ces notions est détaillé dans la section « Revue de la littérature » (Literature review), qui comprend six chapitres :

(i) Les protéines recombinantes à usage thérapeutiques. L'objectif de ce chapitre est de donner un aperçu des caractéristiques des protéines recombinantes thérapeutiques, en se concentrant sur les anticorps et sur l'évolution des formats d'anticorps thérapeutiques

(ii) La lignée hôte, système d'expression de la protéine recombinante. Ce chapitre commence par un préambule sur les systèmes d'expression des protéines recombinantes. L'objectif de ce chapitre est de souligner les caractéristiques des cellules CHO en tant qu'hôtes pour la production de protéines recombinantes thérapeutiques, et de présenter la vaste généalogie des cellules CHO pour comprendre les origines de la lignée hôte CHOZN® CHO K1.

(iii) Le vecteur d'expression, support de l'information génétique. Ce chapitre commence par des informations générales sur les éléments des vecteurs d'expression permettant une expression stable des gènes dans les cellules de mammifères. L'objectif de ce chapitre est de présenter les stratégies d'optimisation possibles du vecteur d'expression pour augmenter le niveau de productivité et la stabilité de la productivité.

(iv) Le processus de développement des lignées cellulaires. Ce chapitre commence par une description du déroulement général du processus de développement d'une lignée cellulaire. L'objectif de ce chapitre est de mettre l'accent sur les deux étapes clés de tout processus de développement de lignées cellulaires, c'est-à-dire la transfection et le clonage de cellules individuelles, ainsi que sur la stabilité de la productivité des cellules CHO.

(v) Aperçu de la plateforme de développement de lignées cellulaires CHOZN® CHO K1. L'objectif de ce chapitre est de fournir un aperçu détaillé de l'état initial de la plateforme de développement de la lignée cellulaire CHOZN® CHO K1, en mettant l'accent sur les éléments génétiques du vecteur d'expression et sur le processus de génération de la lignée cellulaire.

(vi) Synthèse de la revue de la littérature.

Objectifs

La thèse a pour objectif de comprendre les paramètres critiques du processus de génération de lignées cellulaires pour l'expression de protéines recombinantes dans des cellules CHO, du point de vue de chaque élément d'une plateforme de développement de lignées cellulaires : la lignée hôte, le vecteur d'expression et les étapes du processus de génération de lignées cellulaires.

Ces travaux ont directement été mis en application dans un contexte industriel. En effet, ils ont permis d'optimiser la plateforme CHOZN® CHO K1 de génération de lignées cellulaires, proposée dans le cadre des activités de sous-traitance de projets de développement et de production (CDMO, « contract development and manufacturing organization ») de Merck, afin d'augmenter la productivité des lignées générées et de réduire les délais de développement de lignées.

La plateforme CHOZN® CHO K1 de génération de lignées cellulaires a été partiellement optimisée préalablement à cette thèse, notamment en ce qui concerne l'étape de transfection (criblage des paramètres d'électroporation) et le vecteur d'expression (criblage de combinaisons d'activateurs et de promoteurs). Il est à noter que cette plateforme de développement de lignées cellulaires est caractérisée par une lignée cellulaire hôte et un vecteur d'expression qui sont libres de droits. Au vu des connaissances scientifiques issues de l'étude bibliographique, la poursuite des travaux sur le vecteur d'expression (en particulier sur l'expression de la résistance à la pression de sélection), et sur le processus de génération de lignées cellulaires ont été les pistes privilégiées.

Résultats

Les résultats sont détaillés dans la section « Résultats » (Results), qui comprend quatre chapitres :

(i) **Caractérisation de la performance de la plateforme CHOZN® CHO K1 de génération de lignées cellulaires**. Les objectifs de ce chapitre sont de définir les caractéristiques de référence de la plateforme CHOZN® CHO K1 et de définir la performance initiale de la plateforme CHOZN® CHO K1 en termes de productivité et de délais de développement de lignées cellulaires.

Le processus de développement de lignées cellulaires associé à la plateforme CHOZN® CHO K1 commence par la décongélation et la culture de la lignée hôte. La lignée hôte est ensuite transfectée par électroporation, en utilisant le vecteur d'expression pEM2V-a. Le squelette du vecteur d'expression pEM2V-a contient trois cassettes d'expression (résistance à la pression de sélection, expression de la chaîne lourde d'anticorps et de la chaîne légère d'anticorps) reposant sur des éléments génétiques classiques. La génération de lignées cellulaires monoclonales est un processus en deux étapes, divisé en (i) une étape de génération et de sélection de mini-pools (MP) et en (ii) une étape de génération et de

sélection de clones (CL). L'étape de génération de mini-pools vise à faciliter l'identification des pools fortement producteurs, tandis que l'étape de génération de clones vise à fournir une lignée cellulaire monoclonale ayant un niveau d'expression élevé de la protéine d'intérêt. Les étapes de génération de mini-pools et de clones reposent toutes deux sur un processus itératif d'identification et de sélection des cellules hautement productives : expansion et test de productivité dans des plaques 96-puits, mise à l'échelle et test de productivité dans des plaques 6-puits, et mise à l'échelle et test de productivité dans des tubes TPP®. Les délais de développement des lignées cellulaires sont estimés à 7,5 mois (de l'ADN aux banques cellulaires (pre-MCB, « pre-master cell bank »)).

Les travaux débutent donc par la caractérisation de la performance de la plateforme CHOZN® CHO K1 de génération de lignées cellulaires. Le premier objectif était de définir les caractéristiques de référence de la plateforme CHOZN® CHO K1, principalement liées à la croissance cellulaire. Le deuxième objectif était de définir la performance initiale de la plateforme CHOZN® CHO K1 en termes de productivité (mini-pools et clones) et de délais de développement des lignées cellulaires. Cette caractérisation des performances initiales de la plateforme sera par la suite considérée comme un point de comparaison pour les optimisations effectuées au cours de ce projet.

Pour cela, des mini-pools CHOZN® CHO K1 exprimant un anticorps monoclonal IgG4 ont été générés et criblés *via* un processus de génération de mini-pools. Les mini-pools les plus productifs ont été sélectionnés pour la génération de clones. Les clones CHOZN® CHO K1 ont été générés par clonage par la méthode de dilution limite, et criblés par un processus de génération de clones. La stabilité du niveau d'expression des clones a été étudiée sur 60 niveaux de doublement de la population (PDL).

Des indicateurs de référence liés à la croissance cellulaire ont été définis. La réémergence des mini-pools CHOZN® CHO K1 (le seuil de réémergence des mini-pools étant fixé à 50%) a été observée dans les 10 jours suivant la transfection. La réémergence des clones CHOZN® CHO K1 (le seuil de réémergence des clones étant fixé à 30%) a été observée à partir de 12 jours après le clonage unicellulaire. La plupart des clones ont réémergé dans les trois semaines suivant le clonage unicellulaire, avec une efficacité de clonage de 34%. Ces observations permettent d'affiner les délais requis pour la culture des mini-pools et des clones en plaques 96-puits. Le dépassement de ces seuils, c'est-à-dire une croissance cellulaire plus lente que prévu, pourrait mettre en évidence une protéine recombinante difficile à exprimer (DTE, « difficult to express »), voire une protéine recombinante cytotoxique. En outre, une corrélation possible entre la croissance en plaques 96-puits et la productivité dans des modes de culture agités (soit des plaques 24-puits profonds, soit des flasques Erlenmeyer) a été étudiée, mais n'a pas été observée.

Par ailleurs, cette caractérisation initiale a conduit à la production de clones produisant plus de 1 g/L d'un anticorps monoclonal IgG4 standard en 7,5 mois. Le niveau de productivité et les délais de nécessaires pour le développement des lignées cellulaires sont non-concurrentiels par rapport aux autres plateformes industrielles de génération de lignées cellulaires, et ont donc conduit à l'identification de

deux domaines de recherche : (i) l'optimisation du processus de génération de lignées cellulaires clonales pour réduire les délais de développement des lignées cellulaires, et (ii) l'optimisation du vecteur d'expression pour augmenter la productivité des lignées cellulaires générées.

(ii) **Optimisation du processus de développement de lignées cellulaires : mise en œuvre du procédé de clonage direct**. L'objectif de ce chapitre est d'évaluer un processus de développement de lignées cellulaires plus court, appelé clonage direct de cellules individuelles (DSCC, « direct single-cell cloning »), afin de réduire les délais nécessaires pour la génération de lignées cellulaires.

Le chapitre précédent a montré que le procédé de génération de lignées cellulaires associé à la plateforme CHOZN® CHO K1 permet de générer des clones en 7,5 mois. Les délais nécessaires pour la génération de lignées cellulaires avec la plateforme CHOZN® CHO K1 doivent être réduits. Premièrement, un processus de 7,5 mois n'est pas viable alors que toutes les activités de développement (développement de lignée cellulaire, développement de procédé de culture (USP, « up-stream Ppocess »), développement de procédé de purification (DSP, « down-stream process ») et développement analytique) et de production (préparation de la MCB, production GMP (« Good Manufacturing Practices ») et contrôle qualité) devraient être réalisées en un an. En effet, des efforts importants sont actuellement déployés pour accélérer le processus d'approbation des nouveaux médicaments, particulièrement pour les besoins médicaux encore non satisfaits. Deuxièmement, un processus de 7,5 mois n'est pas viable alors que les plateformes industrielles de développement de lignées cellulaires concurrentes permettent de développer des lignées cellulaires en 3 mois. La compétitivité et l'attractivité de la plateforme CHOZN® CHO K1sont ainsi en jeu.

Les travaux se poursuivent donc sur l'optimisation du processus de génération de lignées cellulaires associé à la plateforme CHOZN® CHO K1. L'objectif était d'évaluer un processus de développement de lignées cellulaires plus court, appelé clonage direct (DSCC), afin de réduire les délais de développement de lignées cellulaires.

Le processus de clonage direct de cellules uniques conçu consiste à : (i) transfecter la lignée cellulaire hôte, avec une quantité augmentée de cellules à électroporer mais des paramètres d'électroporation identiques à ceux utilisés dans le processus de génération de lignée long (MP-CL) ; (ii) mettre sous pression de sélection du pool de cellules transfectées ; (iii) laisser le pool transfecté atteindre le minimum de viabilité ; (iv) cloner le pool transfecté par dilution limite, sept jours après la transfection ; (v) sélectionner les clones les plus productifs, avec un processus de sélection itératif dans des plaques 96-puits (mesure de la productivité), des plaques 24-puits profonds (test de productivité batch) et des tubes TPP® (test de productivité fed-batch), identiques à ceux réalisés dans le processus de génération de lignée long (MP-CL).

Pour cela, le processus raccourci de développement de lignées cellulaires, appelé « processus court de développement de lignées cellulaires (DSCC) », a été comparé au processus traditionnel de développement de lignées cellulaires, appelé « processus long de développement de lignées cellulaires (mini-pools - clones (MP-CL)) ». Les clones CHOZN® CHO K1 ont été générés en utilisant la méthode de dilution limite, en suivant soit le processus long (MP-CL), soit le processus court (DSCC). La productivité des clones a été évaluée en plaques 96-puits, puis en plaques 24-puits profonds, puis en tubes TPP®, et finalement en bioréacteurs ambr250®. La stabilité du niveau d'expression des clones a été étudiée sur une durée correspondant à 60 doublements de la population (PDL).

La faisabilité technique du processus de clonage direct de cellules uniques a été évaluée. Premièrement, le profil de la lignée CHOZN® CHO K1 pendant l'étape de sélection a été défini. Le profil de viabilité du pool transfecté a été caractérisé par une forte diminution de la viabilité dans les huit jours suivant la transfection, atteignant un minimum entre 5 et 20 %. La réémergence du pool a été observée au plus tard trois semaines après la transfection. Deuxièmement, la réémergence des clones CHOZN® CHO K1 a été observée. Les clones générés par DSCC ont réémergé à partir de 14 jours après le clonage. Comme pour le processus MP-CL, les clones générés par le processus DSCC ont pour la plupart réémergé dans les trois semaines suivant le clonage, ce qui n'a pas d'impact sur les délais définis précédemment pour la réémergence des clones en plaques 96-puits. Des différences significatives dans l'efficacité du clonage sont observées entre les deux processus, le clonage unicellulaire direct donnant lieu à la plus faible efficacité de clonage. Ces deux points (sélection du pool transfecté et réémergence de clones) valident la faisabilité technique de mise en œuvre du processus DSCC sur la plateforme CHOZN® CHO K1, mais suggèrent également qu'il existe une marge d'optimisation considérable du processus DSCC.

Par ailleurs, l'impact du procédé DSCC sur la productivité des lignées cellulaires recombinantes a été évalué. Même si le titre maximal était assez proche entre les deux processus de génération de lignées cellulaires après les essais de productivité en fed-batch dans des tubes TPP® et des bioréacteurs ambr250®, il est notable que le titre médian est plus élevé pour les clones générés par le processus de développement de lignées cellulaires long (MP-CL). De la même manière, la productivité maximale spécifique au procédé était similaire entre les deux procédés de développement de la lignée cellulaire après les essais de productivité en fed-batch, mais on peut noter une productivité médiane spécifique au procédé plus élevée pour les clones générés avec le procédé de développement de la lignée cellulaire longue (MP-CL). Ces différences sont significatives uniquement au stade de la plaque de 24 puits.

Le procédé DSCC s'est avéré efficace pour réduire les délais de développement des lignées cellulaires. En effet, les lignées cellulaires recombinantes ont été générées en 4,5 mois grâce au processus DSCC, contre 7,5 mois pour le long processus de génération de lignées cellulaires (MP-CL), ce qui équivaut à une réduction significative de 40% du temps nécessaire pour la génération de lignées cellulaires. Cependant, le procédé DSCC a entraîné une baisse de la productivité des lignées cellulaires générées.

Le processus de clonage direct est souvent associé à des techniques avancées de tri et/ou d'ensemencement de cellules individuelles dans les autres plateformes industrielles de génération de lignées cellulaires.

Enfin, la stabilité du niveau d'expression de deux clones CHOZN® CHO K1 générés par le processus direct de clonage unicellulaire a été examinée dans le cadre d'une étude de stabilité et comparée à celle des clones générés par le processus long de développement de la lignée cellulaire (MP-CL). Les mêmes tendances ont été observées sur les clones générés avec un processus de développement de lignée cellulaire long (MP-CL), ce qui suggère que le type de processus de développement de lignée cellulaire appliqué n'influence pas la stabilité de la productivité de la lignée cellulaire recombinante.

Dans le cadre des travaux présentés ici, les efforts ont donc été orientés vers l'amélioration du vecteur d'expression associé à la plateforme.

(iii) **Optimisation du vecteur d'expression : évaluation des promoteurs pour l'expression de la résistance à la puromycine**. L'objectif de ce chapitre est d'évaluer de nouveaux promoteurs pour l'expression de la pression de sélection, afin d'augmenter la productivité des lignées de cellules clonales recombinantes.

L'identification de clones stables et producteurs dans la plateforme CHOZN® CHO K1 repose sur une sélection par antibiotique (puromycine). L'expression de la résistance à la puromycine (conférée par la puromycine N-acétyltransférase(PAC)) est sous le contrôle transcriptionnel du promoteur SV40E (« Simian virus 40 early »).

Le degré de stringence de sélection des clones est un paramètre important pour l'identification de clones fortement producteurs. Un degré élevé de sélectivité oblige les cellules à être plus productives dans le *locus* du marqueur de sélection afin de survivre au processus de sélection. Par conséquent, un marqueur de sélection affaibli peut augmenter la production de protéines recombinantes par différents mécanismes. L'augmentation de la sélectivité peut résulter d'un meilleur site d'intégration du plasmide, d'une activité transcriptionnelle ou traductionnelle plus importante, ou d'un nombre de copies de gènes plus élevé. Ces mécanismes compensatoires conduisent à une expression accrue du gène d'intérêt lié, car le marqueur de sélection et le gène d'intérêt sont généralement co-localisés. L'augmentation de la stringence de sélection des clones conduit donc à des processus de criblage de lignées cellulaires plus efficaces, en identifiant des lignées cellulaires plus productives avec un travail de laboratoire réduit. La stringence de sélection peut être renforcée en augmentant la concentration de la pression de sélection dans les milieux de culture ou en diminuant la force du marqueur de sélection. La première stratégie est rarement appliquée en raison d'une croissance cellulaire plus lente à des concentrations de pression de sélection plus élevées. En revanche, la seconde a été fréquemment décrite dans la littérature, à travers

diverses stratégies à différents stades de l'expression du marqueur de sélection : (i) au niveau transcriptionnel, en utilisant un promoteur faible ou affaibli en ciblant des sites transcriptionnels importants ; (ii) au niveau post-transcriptionnel, en utilisant des éléments déstabilisateurs de l'ARNm, tels que les éléments riches en UA (ARE, « AU-rich elements ») ; (iii) au niveau traductionnel, en utilisant les codons les moins fréquents ou des sites d'entrée internes des ribosomes (IRES, « internal ribosome entry site ») atténués ; (iv) au niveau post-traductionnel, en utilisant des éléments de déstabilisation de la protéine, tels que les séquences PEST (proline (P), glutamate (E), sérine (S), thréonine (T)) de l'ornithine décarboxylase murine (MODC, « murine ornithine decarboxylase ») ; ou (v) au niveau protéique, en utilisant des mutants atténués du marqueur de sélection.

Les travaux ont donc été concentrés sur le promoteur permettant la transcription de la puromycine N-acétyltransférase, car la transcription est traditionnellement considérée comme le facteur dominant contrôlant l'expression. De plus, cet élément génétique était facilement modifiable, au regard des techniques de laboratoire et des connaissances dont le groupe de recherche disposait. L'objectif était de diminuer la force du promoteur et donc d'augmenter la stringence de sélection des clones.

L'identification des promoteurs candidats pour l'expression de la PAC s'est appuyée sur deux stratégies. La première stratégie consistait à réduire l'activité transcriptionnelle du promoteur natif SV40E en délétant des sites transcriptionnels importants. La deuxième stratégie consistait à remplacer le promoteur natif SV40E par un promoteur endogène de la lignée CHO-K1. Trois promoteurs endogènes de la lignée CHO-K1 ont été identifiés (promoteurs H2AJ, DRC7 et RAS) et quatre promoteurs SV40E partiellement délétés ont été conçus (promoteurs SV40E_Δ72bpRII, SV40E_Δ21bpRIII-II, SV40E_ΔTATA, SV40E_ΔAT-region). L'activité transcriptionnelle réduite des promoteurs évalués par rapport au promoteur SV40E natif a été confirmée en mesurant le niveau d'expression du gène *pac* par RT-qPCR lors d'un test d'expression transitoire. Les promoteurs ayant une activité transcriptionnelle réduite ont été utilisés pour exprimer la PAC dans des vecteurs contenant la chaîne légère (LC, « light chain ») et la chaîne lourde (HC, « heavy chain ») d'un anticorps recombinant sous le contrôle de promoteurs forts classiques. Ces vecteurs ont été utilisés pour la génération de lignées cellulaires avec la plateforme CHOZN® CHO K1, selon le processus actuel de développement de lignées cellulaires. Les promoteurs ont d'abord été sélectionnés par une génération de mini-pools. Les promoteurs montrant une augmentation de la productivité du mini-pool ont ensuite été évalués par une génération de clones avec un processus de clonage direct. Une étude de stabilité du niveau d'expression de l'anticorps recombinant a été réalisée sur une durée correspondant à 60 doublements de population (PDL) avec un suivi du nombre de copies de plasmides par qPCR et une mesure du niveau d'expression de *pac*, *rituLC* et *rituHC* par RT-qPCR. Enfin, le gain de productivité obtenu sur le rituximab avec des promoteurs modifiés a été confirmé sur plusieurs protéines recombinantes (IgG1, IgG4 et protéine de fusion).

Une diminution de 38 à 77% du niveau de transcription de *pac* en expression transitoire a été observée par RT-qPCR après substitution du promoteur SV40E soit par un promoteur endogène de la lignée CHO-K1 (H2AJ, DRC7 ou RAS) soit par un promoteur SV40E modifié (SV40_Δ72bpRII, SV40_Δ21bpRIII-II, SV40_ΔTATA ou SV40_ΔAT-region). La substitution du promoteur SV40E par des promoteurs endogènes a conduit aux plus fortes diminutions (-68 à -77%) du niveau de transcription de *pac* (la différence étant statistiquement significative). Cependant, la substitution du promoteur SV40E par des promoteurs SV40E modifiés a conduit à des diminutions plus faibles (-38 à -54%) du niveau de transcription de la protéine pac (la différence étant cependant toujours statistiquement significative). L'ensemble des promoteurs candidats ont donc été retenus pour une évaluation de leur performance pour l'augmentation de la productivité des lignées recombinantes.

La substitution du promoteur natif SV40E par les promoteurs endogènes de CHO K1 ou par les promoteurs SV40E modifiés a fortement diminué la réémergence cellulaire (mini-pools et clones), ce qui indique une stringence de sélection accrue. De bonnes corrélations ont été observées entre l'activité transcriptionnelle des promoteurs évalués et la réémergence des mini-pools (avec une corrélation statistiquement significative), du pool transfecté, ou des clones (tendance de la corrélation seulement).

La productivité des mini-pools CHOZN® CHO K1 a augmenté lorsque le promoteur SV40E a été remplacé par des promoteurs endogènes de la lignée CHO-K1 (RAS) ou des promoteurs SV40E modifiés (SV40_Δ72bpRII, SV40_Δ21bpRIII-II et SV40_ΔTATA). Le titre maximal du mini-pool a été augmenté de 1,7 à 2,3 fois et le titre médian a été augmenté de 1,6 à 2,1 fois par rapport au promoteur natif SV40E, le promoteur SV40E_Δ21bpRIII-II donnant les meilleures productivités sur les deux critères. De plus, l'augmentation de la productivité des mini-pools a été confirmée sur plusieurs molécules avec les promoteurs RAS et SV40E_Δ21bpRIII-II. Par ailleurs, la productivité des clones CHOZN® CHO K1 a augmenté lorsque le promoteur SV40E a été remplacé par des promoteurs endogènes de la lignée CHO-K1 (RAS) ou des promoteurs SV40E modifiés (SV40_Δ72bpRII et SV40E_Δ21bpRIII-II - mais pas par SV40_ΔTATA). Le titre maximal du clone a été augmenté de 1,1 à 1,5 fois et le titre médian a été augmenté de 1,3 à 1,5 fois par rapport au promoteur natif SV40E, le promoteur SV40E_Δ21bpRIII-II donnant le meilleur titre maximal et le promoteur RAS donnant le meilleur titre médian. La productivité spécifique maximale des clones a été multipliée par 1,3 et le titre médian par 1,5 à 2,2 par rapport au promoteur natif SV40E, le promoteur SV40E_Δ21bpRIII-II donnant les meilleures productivités pour les deux critères.

Une étude de stabilité a été menée sur des clones générés avec le promoteur SV40E_Δ21bpRIII II, car ce promoteur a été identifié comme la meilleure condition sur la majorité des critères de productivité. Lorsque la pression de sélection a été maintenue dans le milieu de culture pendant la culture longue durée, tous les clones générés avec le promoteur SV40E_Δ21bpRIII II ont conservé au moins 70% de leur productivité initiale. Cependant, même si un plus grand nombre de clones générés avec le promoteur

SV40E_Δ21bpRIII II étaient instables lorsque la pression de sélection a été supprimée du milieu de culture pendant la culture longue durée par rapport aux clones générés avec le promoteur de type sauvage SV40E, la différence n'est pas statistiquement significative. Cette dernière partie de l'étude confirme que la stratégie de marqueur de sélection affaibli peut être appliquée sur la plateforme de développement de lignées cellulaires CHOZN® CHO K1, sans affecter les performances du procédé. De plus, cette étude de stabilité a permis d'obtenir des informations précieuses sur le comportement des clones CHOZN® CHO K1 sur une culture longue durée. Plus précisément, il apparaît que l'instabilité de la productivité des clones CHOZN® CHO K1 s'explique plutôt par la suppression de la pression de sélection.

Ainsi, la réduction de la force du promoteur pour l'expression de la résistance à la pression de sélection ouvre la porte à des processus de génération de lignées cellulaires plus efficaces, en favorisant l'identification de clones fortement producteurs (productivité supérieure à 3 g/L d'un anticorps monoclonal standard) tout en étant compatible avec des délais de génération de lignées réduits (lignées recombinantes générées en 4,5 mois).

(iv) **Développement d'une méthode de transcription inverse - réaction en chaîne par polymérase quantitative (RT-qPCR)**. L'objectif de ce chapitre est de mettre en œuvre une méthode RT-qPCR, qui est requise au chapitre 3.

L'analyse du niveau de transcription de gènes a été nécessaire à deux reprises dans les travaux présentés précédemment. Premièrement, l'analyse du niveau d'expression du gène de la puromycine N-acétyltransférase (*pac*) était nécessaire pour comparer la force des promoteurs évalués pour l'expression de la résistance à la puromycine. Deuxièmement, l'analyse des niveaux d'expression des gènes de la chaîne lourde du rituximab (*rituHC*) et de la chaîne légère du rituximab (*rituLC*) était nécessaire pour corréler le niveau de transcription au niveau de traduction pendant l'étude de stabilité du niveau d'expression du rituximab sur les clones SV40E et SV40E_Δ21bpRIII-II.

L'analyse du niveau d'expression de gènes est généralement réalisée à l'aide de puces d'ADNc ou de techniques transcription inverse - réaction en chaîne par polymérase quantitative (RT-qPCR), ou par des techniques à haut débit telles que le séquençage de l'ARN (RNA seq). Les technologies de puces d'ADNc et de RT-qPCR sont toutes deux rapides, de moyen à haut débit et peuvent être automatisées. La technologie des puces d'ADNc permet l'analyse simultanée de milliers de gènes dans deux échantillons d'ARN marqués de manière différentielle. En revanche, la technique RT-qPCR permet plutôt l'analyse parallèle de quelques gènes dans de nombreux échantillons d'ARN différents. La RT-qPCR présente une sensibilité élevée, une bonne reproductibilité et une large gamme de quantification. Elle est donc devenue la méthode de référence pour la quantification de l'ARN. Dans le contexte des travaux présentés ici, les techniques à haut débit n'avaient aucune utilité. L'intérêt d'analyse s'est concentré sur quelques gènes, à savoir *pac*, *rituHC* et *rituLC*, dans un grand nombre d'échantillons. Par conséquent, la technique

de RT-qPCR, en utilisant la technologie de détection « SYBR® Green », a été choisie pour l'analyse du niveau de transcription des gènes d'intérêt.

Le développement de la méthode de RT-qPCR a été effectué en suivant les directives d'information minimale pour la publication des expériences de PCR quantitative (MIQE, « minimum information for publication of quantitative PCR experiments ») et le langage de balisage des données de PCR en temps réel (RDML, « real-time PCR data markup language »). Les recommandations MIQE visent à fournir une méthodologie permettant de produire des données plus fiables et cohérentes à partir d'expériences de RT-qPCR. Elles couvrent notamment tous les aspects de l'acquisition des échantillons, de la conception et de la validation des essais, et de l'analyse des données.

Premièrement, la méthode d'extraction d'ARN a été mise en place, et le contrôle de la qualité des ARN extraits a été inclus dans la démarche expérimentale. Le contrôle de la qualité de l'ARN est en effet une exigence pour de nombreuses applications en aval basées sur l'ARN, y compris la RT-qPCR, et fait référence aux mesures de concentration, de pureté et d'intégrité de l'ARN. La qualité de l'ARN (en particulier sa pureté et son intégrité) influence la précision de la mesure de l'expression génique. Le contrôle de la qualité de l'ARN extrait a montré que tous les échantillons d'ARN ont un ratio A_{260}/A_{280} compris entre 2,0 et 2,2 et un ratio A_{260}/A_{230} compris entre 2,0 et 2,2, indiquant donc une solution d'ARN pure. En outre, tous les échantillons d'ARN ont un score de qualité (RQS, « RNA quality score ») supérieur à 5,0, indiquant un ARN partiellement dégradé, ou même un RQS supérieur à 8,0, indiquant un ARN intact. Bien qu'un maximum de précautions ait été pris pour une manipulation soigneuse des échantillons (échantillonnage rapide, manipulation sur glace, congélation rapide, etc.), l'intégrité de l'ARN est partiellement affectée sur plusieurs échantillons. Un tel RQS n'est pas critique pour les expériences de RT-qPCR si la taille de l'amplicon est maintenue relativement courte (inférieure à 250 bp, voire inférieure à 150 bp).

Deuxièmement, la méthode de qPCR a été mise en œuvre. Après avoir choisi la chimie de détection (colorant fluorescent) et la méthode d'analyse des données (normalisation par rapport aux gènes de référence), et conçu des amorces pour les cibles visées, les amorces pour les gènes du plasmide et les gènes de référence ont été évaluées pour leur spécificité et leur efficacité. En outre, les gènes de référence candidats ont été évalués pour la normalisation. Une normalisation précise est essentielle pour l'analyse des données d'expression génique dans les tests RT-qPCR, car elle doit, entre autres, compenser les variations de la RT-qPCR. L'utilisation de gènes de référence est la méthode la plus couramment appliquée pour la normalisation de la RT-qPCR. La communauté scientifique s'accorde à dire que le choix final des gènes de référence doit être systématiquement validé pour chaque étude de RT-qPCR. La validation des gènes de référence a été réalisée sur des gènes décrits dans la littérature, certains étant décrits comme des gènes de référence traditionnels (tels que *actb*, *b2m* et *gapdh*), d'autres étant proposés

comme de nouveaux gènes de référence. Les travaux menés ont permis de valider *rps16* et *atp5mg* comme les gènes les plus stables parmi les 16 candidats finaux à la normalisation.

Ces méthodes ont été utilisées pour l'analyse du niveau de transcription de gènes et pourraient être utilisées pour d'autres analyses sur la plateforme CHOZN® CHO K1. L'analyse du nombre de copies de gènes et la mesure du niveau de transcription pourraient être mises en œuvre en tant qu'analyses supplémentaires pour caractériser les lignées cellulaires recombinantes pendant le processus de génération des lignées cellulaires. À long terme, ces méthodes, et en particulier la RT-qPCR, pourraient également être utilisées pour la détection de marqueurs de productivité.

Conclusion

Les caractéristiques et la performance de la plateforme CHOZN® CHO K1 de génération de lignées cellulaires ont d'abord été définies au cours d'un procédé basé sur une génération de mini-pools et une génération de clones. Cette caractérisation initiale, ayant conduit à l'obtention en 7,5 mois de clones produisant plus de 1 g/L d'un anticorps monoclonal standard, a permis d'identifier deux axes de travaux de recherche : l'optimisation du procédé de génération de lignées cellulaires clonales pour la diminution des délais de développement de lignées, et l'optimisation du vecteur d'expression, pour l'augmentation de la productivité des lignées générées.

Un procédé de génération de lignées accéléré, appelé clonage direct (DSCC), a été évalué. Ce procédé a permis de réduire de près de moitié la durée du procédé de développement de lignées, mais a, en revanche, conduit à la diminution de la productivité des lignées générées.

Par ailleurs, plusieurs promoteurs atténués ont été évalués pour l'expression de la résistance à la pression de sélection, indispensable pour la génération de clones stables et producteurs. Certains promoteurs, utilisés lors d'un procédé de génération de lignées basé sur le clonage direct précédemment évalué, ont conduit à l'obtention en 4 mois de clones produisant plus de 3 g/L d'un anticorps monoclonal standard. De façon générale, l'utilisation de ces promoteurs a conduit à l'augmentation de la productivité des lignées générées. L'augmentation de la productivité des lignées a été confirmée sur plusieurs anticorps monoclonaux et une protéine de fusion Fc.

Des limites communes ont été identifiées dans les deux études et sont liées à (i) une approche expérimentale dépendant à la fois du site d'intégration du plasmide et du nombre de copies du plasmide, qui sont connus pour avoir un impact sur la productivité, (ii) une capacité de criblage limitée, aboutissant à des conclusions partielles en termes de significativité statistique, et (iii) la limitation des essais aux anticorps ou aux molécules à une ou deux chaînes. Néanmoins, et malgré ces limitations, plusieurs paramètres critiques pour l'expression de protéines recombinantes dans les cellules CHO ont été

identifiés, sur deux éléments d'une plateforme de développement de lignées cellulaires : le vecteur d'expression et le processus de développement de lignées cellulaires.

Les résultats obtenus au cours de ce travail ont une application industrielle immédiate, grâce à l'optimisation de la plateforme CHOZN® CHO K1 de génération de lignées cellulaires. Tout d'abord, le processus de clonage direct d'une seule cellule est devenu le protocole standard de développement de lignées cellulaires appliqué sur les projets CDMO avec la plateforme CHOZN® CHO K1 et inclut maintenant une combinaison de paramètres optimisés, tels que le seuil de viabilité pour le clonage d'une seule cellule et la supplémentation du milieu de clonage. Deuxièmement, les promoteurs atténués pourraient être mis en œuvre dès le prochain projet CDMO. Troisièmement, les méthodes qPCR et RT-qPCR, respectivement développées pour le suivi du nombre de copies de gènes et du niveau de transcription, ouvrent la voie à la mise en œuvre de méthodes supplémentaires pour caractériser les clones au cours du processus de développement des lignées cellulaires.

La combinaison du processus de clonage direct avec un marqueur de sélection affaibli a permis de rapprocher les performances de la plateforme CHOZN® CHO K1 de celles des plateformes industrielles concurrentes, en termes de productivité (jusqu'à 3 g/L pour un anticorps monoclonal standard) et de délais de développement des lignées cellulaires (ADN à pré-MCB en 4,5 mois). Ce travail souligne donc que, sans investissement massif en équipement ou en technologie, les performances de la plateforme CHOZN® CHO K1 peuvent être significativement améliorées, mettant ainsi en évidence le potentiel de cette plateforme.

Néanmoins, des efforts sont encore nécessaires pour continuer à améliorer les performances de la plateforme. La poursuite des travaux sur la rigueur de la sélection, la reprise de l'évaluation des promoteurs pour l'expression de protéines recombinantes, et la recherche de marqueurs associés à une haute productivité méritent particulièrement d'être explorés. Par ailleurs, et comme le montre ce travail, les développements technologiques des méthodes sont tout aussi importants. La mise en place d'un système d'intégration semi-ciblé, et d'échelles de culture miniaturisées devraient fortement contribuer à l'optimisation des performances de la plateforme. Enfin, le développement de lignées cellulaires étant sur le chemin critique de l'avancement d'une nouvelle thérapie vers les premiers essais cliniques chez l'homme, ces optimisations doivent répondre au besoin d'accélération du développement de molécules thérapeutiques recombinantes.

La recherche et la compréhension des éléments critiques pour une expression élevée et stable des protéines recombinantes dans les cellules CHO doivent donc se maintenir, afin de poursuivre l'optimisation de la plateforme CHOZN® CHO K1, qui possède des caractéristiques fortes et uniques pour être une plateforme de développement de lignées cellulaires efficace et compétitive.

Table of contents

Remerciements ..... **i**

Résumé ..... **iii**

Introduction ..... **iii**

Objectifs ..... **vi**

Résultats ..... **vi**

Conclusion ..... **xv**

Table of contents ..... **xvii**

List of figures ..... **xxv**

List of tables ..... **xxviii**

List of abbreviations ..... **xxx**

1 Introduction **1**

1.1 Historical and industrial context **1**

1.1.1 From 1980 to today: review of 40 years of recombinant protein production 1

1.1.2 Since the 2010s, monoclonal antibodies dominate the market for recombinant proteins for therapeutic use 2

1.1.3 Stages in the development of recombinant protein production processes 3

1.1.4 Progress and challenges of industrial platforms for cell line development 4

1.2 Objectives of the thesis **5**

1.3 Organization of the thesis **5**

2 Literature review **7**

2.1 Recombinant proteins for therapeutic use **7**

2.1.1 Post-translational modifications 8

2.1.1.1 Importance of post-translational modifications 8

2.1.1.2 Glycosylation, a post-translational modification for more than one third of biopharmaceuticals 9

2.1.1.2.1 Impacts of glycosylation 9

2.1.1.2.2 Glycoprotein biosynthesis and control mechanisms 9

2.1.1.2.3 Characteristics of human glycosylation 10

2.1.2 General information on antibodies 10

2.1.2.1 Structure of immunoglobulins 11

2.1.2.2 Classes d'immunoglobulines 12

2.1.2.3 Evolution of therapeutic antibody formats 15

2.1.2.3.1 From murine antibodies to human antibodies, for a decrease in immunogenicity 15

2.1.2.3.2 From antibodies to antibody derivatives, to increase the diversity of therapeutic molecules 16

2.2 The host line, expression system for the recombinant protein **18**

2.2.1 Preamble on recombinant protein expression systems 18

2.2.1.1 Prokaryotic systems, ideal for the production of non-glycosylated recombinant proteins 18

2.2.1.2 Single-cell eukaryotic systems 19
2.2.1.2.1 Yeasts, ideal for the production of small recombinant glycosylated proteins without therapeutic application 19
2.2.1.2.2 Insect cells, mainly used for the production of human and veterinary vaccines 19
2.2.1.2.3 Mammalian cells, the reference for the production of complex recombinant proteins for therapeutic applications 20
2.2.1.3 Multi-cell eukaryotic systems: transgenic plants and animals, marginal systems for the production of recombinant proteins 21
2.2.2 Characteristics of CHO lines to produce recombinant therapeutic proteins 21
2.2.3 CHO cell lines' genealogy and associated selection systems 23
2.2.3.1 From Chinese Hamster to CHO-ori cell line 26
2.2.3.2 CHO K1 cell line 26
2.2.3.3 CHO S cell line 27
2.2.3.4 CHO DXB11 (or DUK-XB11) and CHO DG44 cell lines: deletion of dihydrofolate reductase (DHFR) activity 27
2.2.3.5 CHO K1SV GS-KO and CHO GS$^{-/-}$ cell lines: deletion of glutamine synthetase (GS) activity 29
2.2.4 CHOZN® cell lines: CHOZN® CHO K1, CHOZN® GS$^{-/-}$ and CHOZN® DHFR$^{-/-}$ 29
2.3 The expression vector, support of genetic information 32
2.3.1 General information on expression vector elements for stable gene expression in mammalian cells 32
2.3.1.1 Elements for transcriptional control 32
2.3.1.2 Elements for translational control 33
2.3.2 Strategies for expression vector optimization: towards enhanced productivity and stability 33
2.3.2.1 Enhancing recombinant protein expression 34
2.3.2.2 Enhancing recombinant protein expression stability 34
2.3.2.3 Attenuating selection marker strength 34
2.3.2.3.1 Preamble on selection systems 34
2.3.2.3.1.1 Antibiotic-based systems: blasticidin, geneticin, hygromycin, puromycin, zeocin, … 34
2.3.2.3.1.2 Metabolic marker-based systems: DHFR, GS, P5CS, HPRT, … 37
2.3.2.3.1.3 Combinations of selection systems 38
2.3.2.3.2 Modulating selection stringency to increase productivity 39
2.3.2.3.2.1 Attenuating selection marker strength at transcriptional level: the example of SV40 promoter 40
2.3.2.3.2.2 Attenuating selection marker strength at post-transcriptional level 43
2.3.2.3.2.3 Attenuating selection marker strength at translational level 43
2.3.2.3.2.4 Attenuating selection marker strength at post-translational level 44
2.3.2.3.2.5 Attenuating selection marker by modifications of enzyme activity 44
2.3.2.3.2.6 Combinations of attenuation strategies 45
2.3.2.4 Targeting integration site 45
2.4 Cell line development process 46

2.4.1 Preamble on cell line development process 46
2.4.2 Transfection 47
2.4.2.1 Transfection by physical methods 47
2.4.2.1.1 Electroporation 47
2.4.2.1.2 Nucleofection 47
2.4.2.1.3 Other physical methods for transfection 48
2.4.2.2 Transfection by chemical methods 48
2.4.3 Single-cell cloning 48
2.4.3.1 Guidelines on single-cell cloning 48
2.4.3.2 Methods and technologies for single-cell cloning 49
2.4.3.2.1 Limiting dilution cloning 49
2.4.3.2.2 Cell sorter and/or seeders 49
2.4.3.3 Concept of monoclonality: monoclonal or clonally-derived cell lines? 50
2.4.4 Expression level stability study 50
2.4.4.1 Stakes and guidelines on expression level stability study 50
2.4.4.2 Insights on expression level stability of CHO-derived recombinant cell lines 51
2.4.4.3 Root causes of productivity instability 52
2.5 Overview of CHOZN® CHO K1 cell line development platform 53
2.5.1 Host cell line 53
2.5.2 Expression vector 53
2.5.2.1 Genetic elements for recombinant protein expression 55
2.5.2.2 Genetic elements for selection marker expression 57
2.5.3 Cell line development process 57
2.5.3.1 Expression vector production 60
2.5.3.2 Host cell line expansion 60
2.5.3.3 Transfection 60
2.5.3.4 Mini-pool generation 60
2.5.3.5 Clone generation 62
2.5.3.6 Recombinant protein expression level stability study 63
2.5.3.7 Media and feeds screening 64
2.5.4 Conclusion 64
2.6 Literature review synthesis 66
3 Results 67
3.1 Characterization of CHOZN® CHO K1 cell line development platform performance 67
3.1.1 Introduction 67
3.1.2 Methodology 68
3.1.3 Results 68
3.1.3.1 CHOZN® CHO K1 mini-pool generation 68
3.1.3.1.1 CHOZN® CHO K1 mini-pools reemerged nine days after transfection 69
3.1.3.1.2 CHOZN® CHO K1 mini-pools' productivity 69

3.1.3.2 CHOZN® CHO K1 clone generation 72
3.1.3.2.1 CHOZN® CHO K1 clones mainly reemerged 12 to 19 days after single-cell cloning 72
3.1.3.2.2 CHOZN® CHO K1 clones' behavior in fed-batch culture 74
3.1.3.2.3 CHOZN® CHO K1 clones' productivity 74
3.1.3.3 CHOZN® CHO K1 clones' expression level stability 78
3.1.3.3.1 Effects of long-term cultivation on CHOZN® CHO K1 clones' growth 78
3.1.3.3.2 Effects of long-term cultivation on CHOZN® CHO K1 clones' productivity 79
3.1.4 Discussion 81
3.1.5 Conclusion 84
3.1.6 Author's contributions 85
3.2 Cell line development process optimization: implementation of direct single-cell cloning 86
3.2.1 Introduction 86
3.2.2 Methodology 87
3.2.3 Design of a direct single-cell cloning (DSCC) protocol 87
3.2.4 Results 88
3.2.4.1 Shortened cell line development process decreases cloning efficiency and clone productivity 88
3.2.4.1.1 Characterization of CHOZN® CHO K1 cell behavior during selection 88
3.2.4.1.2 Impact of cell line development process on cloning efficiency and clone reemergence 89
3.2.4.1.3 Impact of cell line development process on clone productivity 92
3.2.4.2 Impact of cell line development process on clone expression level stability 95
3.2.5 Discussion 97
3.2.6 Conclusion 100
3.2.7 Author's contributions 100
3.3 Expression vector optimization: evaluation of promoters for expression of resistance to puromycin 101
3.3.1 Introduction 101
3.3.2 Hypotheses on the mechanism of action of an attenuated selection marker 102
3.3.3 Methodology 104
3.3.4 Identification of promoters for *pac* expression 104
3.3.4.1 Modified SV40 promoters: SV40E_Δ72bpRII, SV40E_Δ21bpRIII-II, SV40E_ΔTATA, and SV40E_ΔAT-region 104
3.3.4.2 CHO K1 endogenous promoters: H2AJ, DRC7, and RAS 105
3.3.5 Results 106
3.3.5.1 Candidate promoters for *pac* expression have a lower transcriptional strength than SV40E wild type promoter 106
3.3.5.2 Candidate promoters for *pac* expression decrease mini-pool reemergence but increase mini-pool productivity 108
3.3.5.2.1 Impact of promoter strengths on mini-pool reemergence 108
3.3.5.2.2 Impact of promoter strength on mini-pool productivity 110
3.3.5.3 Cellular re-emergence with RAS and SV40E_Δ21bpRIII-II promoters can be increased with simple protocol modifications 112

3.3.5.3.1 Modifications on protocol for mini-pool generation 112
3.3.5.3.2 Modifications on protocol for clone generation 115
3.3.5.4 Candidate promoters for *pac* expression increase clone productivity 116
3.3.5.4.1 Impact of promoter strength on cloning efficiency 117
3.3.5.4.2 Impact of weakened promoters on clone productivity 120
3.3.5.4.2.1 Clone productivity (titer) in 96-well plates 120
3.3.5.4.2.2 Clone productivity (titer and specific productivity) in 24-deepwell plates 121
3.3.5.4.2.3 Clone productivity (titer and specific productivity) in TPP® tubes 123
3.3.5.5 Expression level stability over 60 PDL is similar between SV40 and SV40E_Δ21bpRIII-II promoters 127
3.3.5.5.1 Effects of long-term cultivation on growth 129
3.3.5.5.2 Effects of long-term cultivation on productivity 133
3.3.5.5.2.1 Analysis of clone stability, irrespectively of the promoter used for *pac* expression 133
3.3.5.5.2.2 Analysis of clone stability, depending on promoter used for *pac* expression 135
3.3.5.6 Productivity gain obtained with RAS and SV40E_Δ21bpRIII-II promoters is confirmed on other recombinant proteins 139
3.3.6 Discussion 145
3.3.7 Conclusion 151
3.3.8 Author's contributions 151
3.4 Reverse transcription – quantitative polymerase chain reaction method development 152
3.4.1 Introduction 152
3.4.1.1 Choice of the technique for gene transcription level analysis 152
3.4.1.2 Quantitative PCR: insights on a widespread technique 152
3.4.2 Methodology: guidelines for RT-qPCR method development 154
3.4.3 Results 155
3.4.3.1 RNA extraction 155
3.4.3.2 RNA quality control 155
3.4.3.2.1 Principle 155
3.4.3.2.2 Results 157
3.4.3.3 cDNA synthesis by reverse transcription 158
3.4.3.4 Quantitative PCR 159
3.4.3.4.1 Primer design for plasmid genes 159
3.4.3.4.2 Primer design for reference genes 167
3.4.3.4.2.1 Normalization against reference genes 167
3.4.3.4.2.2 Identification of candidate reference genes and primer sequences 168
3.4.3.4.3 Evaluation of primer specificity 175
3.4.3.4.3.1 Analysis of melting curves 175
3.4.3.4.3.2 Amplicon size verification 184
3.4.3.4.4 Evaluation of PCR efficiency 187
3.4.3.4.5 Reference genes validation 189

3.4.4 Discussion ... 192
3.4.5 Conclusion ... 194
3.4.6 Author's contributions ... 194
4 Discussion ... 195
4.1 Combination of direct single-cell cloning process and attenuated selection marker for optimization of CHOZN® CHO K1 platform ... 195
4.2 Insights on stability of CHOZN® CHO K1 clones ... 202
4.2.1 Instability is independent of CHO host cell line, selection and/or amplification system, and presence or absence of selection pressure ... 202
4.2.2 Understanding underlying causes of productivity instability… ... 204
4.2.3 … to find approaches to enhance productivity stability ... 204
4.3 Towards next version of CHOZN® CHO K1 platform ... 205
4.3.1 Multiple routes to high producing recombinant cell lines ... 205
4.3.2 Reshaping cell line development, for shorter DNA to IND timelines ... 206
5 Conclusion ... 208
6 Material and methods ... 211
6.1 Analytical methods related to cell culture methods ... 211
6.1.1 Viable cell density and viability measurement ... 211
6.1.2 Cell confluence monitoring ... 211
6.1.3 Glucose and lactate concentrations measurement ... 211
6.1.4 pH measurement ... 211
6.1.5 Antibody quantification ... 213
6.2 Cell culture methods ... 215
6.2.1 Preparation of culture media ... 215
6.2.2 CHOZN® CHO K1 host cell line sub-culturing ... 216
6.2.3 Thawing and freezing of cellular banks ... 216
6.2.3.1 Thawing of cellular banks ... 216
6.2.3.2 Freezing of cellular banks ... 216
6.2.4 Transfection by electroporation ... 217
6.2.4.1 Electroporation ... 217
6.2.4.2 Controls for transfection ... 218
6.2.5 Transient gene expression ... 219
6.2.5.1 Transient recombinant protein expression ... 219
6.2.5.2 Transient puromycin N-acetyltransferase expression ... 219
6.2.6 Mini-pools generation: seeding and expansion in 96-well plates ... 219
6.2.7 Transfected bulk pool selection ... 220
6.2.8 Clone generation: seeding and expansion in 96-well plates ... 220
6.2.9 CHOZN® CHO K1 mini-pool and clone sub-culturing ... 223
6.2.9.1 Sub-culturing in 24-deepwell plates ... 223
6.2.9.2 Sub-culturing in Erlenmeyer flasks ... 223

6.2.10 Batch and fed-batch productions ... 225
6.2.10.1 Batch production in 24-deepwell plates ... 225
6.2.10.2 Fed-batch production in TPP® tubes ... 225
6.2.10.3 Fed-batch production in ambr250® bioreactors ... 227
6.3 Molecular biology methods ... 227
6.3.1 Expression vector production ... 228
6.3.2 DNA extraction ... 228
6.3.3 RNA extraction ... 228
6.3.4 DNA and RNA quality control ... 229
6.3.5 Reverse transcription ... 230
6.3.6 Quantitative PCR ... 230
6.3.7 Amplicon size analysis by electrophoresis on agarose gel ... 231
6.4 Calculation formulas ... 231
6.4.1 Doubling Time (DT) ... 231
6.4.2 Population Doubling Level (PDL) ... 232
6.4.3 Integral of Viable Cell density (IVC) ... 232
6.4.4 Specific productivity (Q_P) ... 232
6.4.5 Cloning efficiency ... 232
6.4.6 Percentage of variation ... 233
6.5 Statistical analysis ... 233
6.5.1 Assessment of normal distribution ... 234
6.5.2 One-way analysis of variance (ANOVA) ... 235
6.5.3 Comparison of means ... 235
6.5.3.1 Comparison to control group: Dunnett's test ... 235
6.5.3.2 Multiple pairwise comparisons: Tukey-Kramer test ... 236
6.5.4 Correlation matrix ... 237
References ... 239
Appendix 1: Supplementary data related to results chapter 1 ... 271
Supplementary data on clone stability ... 271
Appendix 2: Supplementary data related to results chapter 2 ... 274
Details on statistical analyses on clones' productivity ... 274
Supplementary data on clone stability ... 276
Appendix 3: Supplementary data related to results chapter 3 ... 279
Details on statistical analysis on promoters' transcriptional strength ... 279
Details on statistical analysis on mini-pools' productivity ... 285
Details on statistical analyses on clones' productivity ... 287
Clone productivity (titer) in 96-well plates ... 287
Clone productivity (titer and specific productivity) in 24-deepwell plates ... 288
Clone productivity (titer and specific productivity) in TPP® tubes ... 293
Supplementary data and details on statistical analyses on clone stability ... 296

Effects of long-term cultivation on growth 296
Effects of long-term cultivation on productivity 297
Analysis of clone stability, irrespectively of the promoter used for *pac* expression 297
Analysis of clone stability, depending on promoter used for *pac* expression 298
Details on statistical analysis on mini-pools' productivity (confirmation on several molecules) 311

List of figures

Figure 1: Stages in the development of processes for recombinant protein production 3

Figure 2: Specificities of N-glycans depending on expression system 10

Figure 3 : Schematic representation of an immunoglobulin 12

Figure 4: Therapeutic antibodies and associated nomenclature 16

Figure 5: Examples of antibody-derived molecules 17

Figure 6: Partial genealogy of CHO cell lines 25

Figure 7: Sequences targeted by zinc finger nucleases for deletions in glutamine synthetase and dihydrofolate reductase genes 31

Figure 8: Structure of SV40E promoter 42

Figure 9: Plasmid map of pEM2V-a 54

Figure 10: Plasmid map of pEM2V-c 55

Figure 11: CHOZN® CHO K1 platform cell line development process (state in 2018) 58

Figure 12: Additional CLD-linked studies: clone stability study and media and feeds screening (state in 2018) 59

Figure 13: Mini-pool confluence in 96-well plate, as a function of subculturing duration 70

Figure 14: IgG4 mAb concentration in mini-pools' supernatants and mean process specific productivity after productivity assay in 96-well plates and in 24-deepwell plates 71

Figure 15: Clone confluence in 96-wp, as a function of subculturing duration 73

Figure 16: IgG4 mAb concentration in clones' supernatants after productivity measurement in 96-well plates and productivity assays in 24-deepwell plates and TPP® tubes 75

Figure 17: Mean process specific productivities after productivity assays in 24-deepwell plates and TPP® tubes 76

Figure 18: Correlation matrix of titers and specific productivities in 96-wp, 24-dwp, and TPP® tubes 77

Figure 19: Doubling times over long-term cultivation depending on clone 79

Figure 20: Cumulated IVC, IgG4 concentrations and mean process Q_P after fed-batch productivity assay at PDL_{20}, PDL_{40}, and PDL_{60} 80

Figure 21: Viability profile of transfected bulk pool during selection 89

Figure 22: Clone confluence in 96-wp, as a function of subculturing duration and cell line development process 91

Figure 23: IgG4 mAb concentration in clones' supernatants after productivity measurement in 96-well plates and productivity assays in 24-deepwell plates, TPP® tubes and ambr250® bioreactors 93

Figure 24: Mean process specific productivities after productivity measurement in 96-well plates and productivity assays in 24-deepwell plates, TPP® tubes and ambr250® bioreactors 94

Figure 25: Cumulated IVC, IgG4 concentrations and mean process Q_P after fed-batch productivity assay at PDL_{20}, PDL_{40}, and PDL_{60} depending on clone and cell line development process 96

Figure 26: Hypotheses on the mechanism of action of an attenuated selection marker and on the impact on recombinant protein productivity 103

Figure 27: General architecture of SV40E wild type promoter and of modified SV40E promoters 105

Figure 28: pac transcription level of SV40E wild type promoter and candidate promoters 107

Figure 29: Percentage of reemerging mini-pools and correlation to pac transcription level 109

Figure 30: Rituximab concentration in mini-pools' supernatants after batch productivity assay in 96-well plate 111

Figure 31: Viability profiles during selection depending on electroporation parameters 116
Figure 32: Viability profiles during selection and cloning efficiencies 118
Figure 33: Correlation matrix between promoter transcriptional activities and selection and cloning results 119
Figure 34: Rituximab concentration in clone's supernatants after productivity measurement in 96-well plate.. 120
Figure 35: Rituximab concentration in clone's supernatants after productivity assay in 24-deepwell plates 122
Figure 36: Specific productivities after productivity assay in 24-deepwell plates 123
Figure 37: Rituximab concentration in clone's supernatants after productivity assay in TPP® tubes 125
Figure 38 : Rituximab concentration in clone's supernatants and specific productivities after productivity assay in TPP® tubes and statistical comparison of group means 126
Figure 39: Flow chart of stability study on SV40E and SV40E_ Δ21bpRIII-II clones 129
Figure 40: Doubling times over long-term cultivation depending on clone, promoter for pac expression and presence/absence of puromycin during sub-culturing. 131
Figure 41: Cumulated IVC during fed-batch productivity assay at PDL_{20}, PDL_{40}, and PDL_{60} depending on clone, promoter for pac expression and presence/absence of puromycin during sub-culturing 132
Figure 42: Rituximab concentrations after fed-batch productivity assay at PDL_{20}, PDL_{40}, and PDL_{60} depending on clone, promoter for pac expression and presence/absence of puromycin during sub-culturing 137
Figure 43: Mean process specific productivities after fed-batch productivity assay at PDL_{20}, PDL_{40}, and PDL_{60} depending on clone, promoter for pac expression and presence/absence of puromycin during sub culturing 138
Figure 44 :Rituximab, trastuzumab, pembrolizumab, nivolumab, and EPO-Fc concentration in mini-pools' supernatants after batch productivity assay in 96-well plates 144
Figure 45: Workflow for RT-qPCR method development 154
Figure 46: RNA integrity calculation and examples 157
Figure 47:Prediction DNA folding of rituximab heavy chain at 60°C 160
Figure 48:Prediction DNA folding of rituximab light chain at 60°C 161
Figure 49: Prediction 1 of DNA folding of puromycin acetyltransferase coding sequence at 60°C 163
Figure 50: Prediction 2 of DNA folding of puromycin acetyltransferase coding sequence at 60°C 164
Figure 51: Prediction of DNA folding of puromycin acetyltransferase non-coding sequence at 60°C 165
Figure 52: Amplicon size verification by agarose gel electrophoresis (plasmid genes) 184
Figure 53: Amplicon size verification by agarose gel electrophoresis (reference genes) 186
Figure 54: Quantification cycles for reference genes on validation samples 190
Figure 55: geNorm analysis results: average expression stability value (M value) and pairwise variation for determination of the optimal number of reference genes for normalization (V value) 191
Figure 56: CHOZN® CHO K1 platform cell line development process (state in 2022) 201
Figure 57: Principle of glucose and lactate concentrations measurement 212
Figure 58: Principle of pH measurement 213
Figure 59: Principle of bio-layer interferometry 214
Figure 60: Workflow for statistical analysis 233
Figure 61: Visual comparison of group means: significance of comparison circles 236
Figure 62: Construction and interpretation of a correlation matrix 238
Figure 63: Statistical comparison of group mean IgG4 mAb concentration in clone's supernatants after productivity measurement in 96-well plates, and productivity assays in 24-deepwell plates, TPP® tubes and ambr250® bioreactors 275

Figure 64: Statistical comparison of group mean specific productivity after productivity assays in 24-deepwell plates, TPP® tubes and ambr250® bioreactors 276

Figure 65: Assessment of normal distribution of RT-qPCR data 280

Figure 66: Statistical analysis on promoters' transcriptional strength 283

Figure 67: Statistical analysis on rituximab concentration in mini-pools' supernatants after batch productivity assay in 96-well plate 286

Figure 68: Rituximab concentration in clone's supernatants after productivity measurement in 96-well plate and statistical comparison of group means 288

Figure 69: Statistical comparison of group mean rituximab concentration in clone's supernatants after batch productivity assay in 24-deepwell plates 290

Figure 70: Statistical comparison of group mean specific productivities after batch productivity assay in 24-deepwell plates 292

Figure 71 : Statistical comparison of group mean rituximab concentration in clone's supernatants after fed-batch productivity assay in TPP® tubes 294

Figure 72 : Statistical comparison of group mean specific productivities after fed-batch productivity assay in TPP® tubes 295

Figure 73: Statistical comparison of group mean doubling time 297

Figure 74: Statistical comparison of group mean expression level variations between PDL_{20} and PDL_{60} 303

Figure 75:Statistical comparison of group mean specific productivity variations between PDL_{20} and PDL_{60} ... 309

Figure 76: Statistical comparison of group mean recombinant protein concentration in mini-pools' supernatants after batch productivity assay in 96-well plate 314

List of tables

Table 1: Properties of immunoglobulin classes and subclasses 13

Table 2: Codon usage in CHO cells 33

Table 3: Commonly used antibiotics for mammalian cell selection 35

Table 4: Cloning efficiency and percentages of reemerging clones as a function of subculturing duration in 96-well plates 72

Table 5: Key features of cell line development platforms 83

Table 6: Cloning efficiency and percentages of reemerging clones as a function of subculturing duration in 96-well plates 90

Table 7: Percentage of reemerging mini-pools 114

Table 8: Synthesis of best conditions for highest maximum and mean titer and highest maximum and mean specific productivity 127

Table 9: Hypotheses on expression level stability upon selection pressure removal 128

Table 10: Percentage of reemerging mini pools 141

Table 11: Synthesis of best conditions for highest maximum and mean titer 142

Table 12: Quality control of extracted RNA (concentration, purity, and integrity) 157

Table 13: Plasmid genes and associated primer sequences 166

Table 14: Reference genes candidates and associated primer sequences 170

Table 15: Reference genes candidates and associated primer sequences with characteristics (T_m, length, GC% and amplicon size) 173

Table 16: Mean C_q depending on temperature gradient and controls (no target control, no DNA control) and melt curve – pac 177

Table 17: Mean C_q depending on temperature gradient and controls (no RT control, no DNA control) and melt curve – rituHC and rituLC 178

Table 18: Mean C_q depending on temperature gradient and controls (no RT control, no cDNA control) and melt curve for reference genes 180

Table 19: Amplification efficiencies (plasmid genes) 188

Table 20: Amplification efficiencies (reference genes) 188

Table 21: Synthesis of direct single-cell cloning assays: experimental conditions and associated results 199

Table 22: Medium used for thawing of cellular banks – Details per assays 216

Table 23: Number of electroporation replicates and electroporation parameters – Details per assays 217

Table 24: Transfected bulk pool selection – Details per assays 221

Table 25: Single-cell cloning – Details per assays 222

Table 26: Mini-pool and clone sub-culturing - Details per assays 224

Table 27: Fed-batch productivity assay - Details per assays 226

Table 28: Primer sequences 231

Table 29: Cumulated IVC variations between PDL_{20}, PDL_{40}, and PDL_{60} 272

Table 30: IgG4 titer variations between PDL_{20}, PDL_{40}, and PDL_{60} 272

Table 31: Mean process Q_P variations between PDL_{20}, PDL_{40}, and PDL_{60} 273

Table 32: IgG4 mAb titer variations between PDL_{20}, PDL_{40}, and PDL_{60} (clones generated with long or short CLD process) 277

Table 33: Mean process Q_P variations between PDL_{20}, PDL_{40}, and PDL_{60} (clones generated with long or short CLD process) 278

Table 34: Rituximab titer variations between PDL_{20}, PDL_{40}, and PDL_{60} (clones with SV40E wild type promoter, w/o selection pressure during sub-culturing) 299

Table 35: Rituximab titer variations between PDL_{20}, PDL_{40}, and PDL_{60} (clones with SV40E wild type promoter, w/ selection pressure during sub-culturing) 300

Table 36: Rituximab titer variations between PDL_{20}, PDL_{40}, and PDL_{60} (clones with SV40E_Δ21bpRIII-II promoter, w/o selection pressure during sub-culturing)..... 301

Table 37: Rituximab titer variations between PDL_{20}, PDL_{40}, and PDL_{60} (clones with SV40E_Δ21bpRIII-II promoter, w/ selection pressure during sub-culturing)..... 302

Table 38: Q_P variations between PDL_{20}, PDL_{40}, and PDL_{60} (clones with SV40E wild type promoter, w/o selection pressure during sub-culturing) 305

Table 39: Q_P variations between PDL_{20}, PDL_{40}, and PDL_{60} (clones with SV40E wild type promoter, w/ selection pressure during sub-culturing) 306

Table 40: Q_P variations between PDL_{20}, PDL_{40}, and PDL_{60} (clones with SV40E_Δ21bpRIII-II promoter, w/o selection pressure during sub-culturing) 307

Table 41: Q_P variations between PDL_{20}, PDL_{40}, and PDL_{60} (clones with SV40E_Δ21bpRIII-II promoter, w/ selection pressure during sub-culturing) 308

List of abbreviations

ADC	Antibody-drug conjugate
ADCC	Antibody-dependent cellular cytotoxicity
ADCP	Antibody-dependent cellular phagocytosis
ARC	Antibody-radionuclide conjugate
ARE	AU-rich element
Asn	Asparagine
ATCC	American type culture collection
BsAb	Bispecific antibody
BGH	Bovine growth hormone
BHK	Baby hamster kidney
BLI	Bio-layer interferometry
CAM	Cell adhesion molecule
CDC	Complement-dependent cytotoxicity
CDMO	Contract development and manufacturing organization
CDR	Complementary determining region
CFE	Cell-free expression
cGMP	Current good manufacturing practices
C_H	Constant heavy
CHO	Chinese hamster ovary
C_L	Constant light
CL	Clone
CLD	Cell line development
CMR	Carcinogenic, mutagenic, and reprotoxic
CMV	Cytomegalovirus
DHFR	Dihydrofolate reductase
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
DSB	Double-strand break
DSCC	Direct single-cell cloning
DSP	Down-stream process
DT	Doubling time
ECACC	European collection of authenticated cell cultures
EF1α	Elongation factor 1α
EMA	European medicines agency
ER	Endoplasmic reticulum
ERAD	Endoplasmic reticulum-associated degradation
ExCB	Extended cell bank
Fab	Fragment antigen-binding
FACS	Fluorescence activated cell-sorting
Fc	Fragment cristallisable
FBS	Fetal bovine serum
FDA	Food and drug administration
GHT	Glycine, hypoxanthine and thymidine
GOI	Gene of interest
GS	Glutamine synthetase
HACA	Human anti-chimeric antibody
HAHA	Human anti-human antibody
HAMA	Human anti-mouse antibody
HC	Heavy chain
HDH	Homology directed repair
HEK	Human embryo kidney
HGH	Human growth hormone
HIV	human immunodeficiency virus
HSV	Herpes simplex virus
HVR	Hypervariable region
IE	Immediate early
IgA	Immunoglobulin A
IgD	Immunoglobulin D
IgE	Immunoglobulin E
IgG	Immunoglobulin G
IgM	Immunoglobulin M
IgSF	Immunoglobulin super family
IRES	Internal ribosome entry site
IS	Immune system
IVC	Integral of viable cell density
LC	Light chain
LDC	Limiting dilution cloning

mAb	Monoclonal antibody
MCB	Master cell bank
MHC	Major histocompatibility complex
MIQE	Minimum information for publication of quantitative PCR experiments
MODC	Murine ornithine decarboxylase
MP	Mini-pool
MSX	Methionine sulfoximine
MTX	Methotrexate
NA	Not applicable
NHEJ	Non homologous end joining
PAC, *pac*	Puromycin N-acetyltransferase
PCR	Polymerase chain reaction
PEST	Proline (P), glutamate (E), serine (S), threonine (T)
PDL	Population doubling level
POI	Protein of interest
Pro	Proline
PTM	Post-translational modification
Q_P	Specific productivity
qPCR	Quantitative polymerase chain reaction
rAAV	Recombinant adeno-associated virus
RCB	Research cell bank
RDML	Real-time PCR data markup language
RIN	RNA integrity number
RNA	Ribonucleic acid
RQS	RNA quality score
RT	Reverse transcription
RT-qPCR	Reverse transcription and quantitative polymerase chain reaction
scFV	Single-chain variable fragment
sdAb	Single-domain antibody
Ser	Serine
SP	Selection pressure
SV40E	Simian virus 40 early
Thr	Threonine
tPA	Tissue plasminogen activator
UPR	Unfolded protein response
USP	Up-stream process
vc	Viable cell
V_H	Variable heavy
V_L	Variable light
VLP	Virus-Like Particles
WCB	Working cell bank
w/	With
w/o	Without
ZFN	Zinc finger nuclease

1 Introduction

1.1 Historical and industrial context

1.1.1 From 1980 to today: review of 40 years of recombinant protein production

Until the 1980s, therapeutic proteins were produced by natural sources. For example, insulin, used for the treatment of diabetes, was extracted from the pancreas of cattle or pigs; human growth hormone (HGH), used for the treatment of growth disorders in children, was extracted from the pituitary gland of human cadavers. However, the supply of therapeutic proteins from natural sources was limited. For example, in the late 1970s, the production of one pound of insulin required 8,000 pounds of pancreatic glands from approximately 23,500 animals. Yet, the pharmaceutical company Eli Elly needed 56 million animals each year to meet the growing need for insulin in the United States (Fraser, 2016). In addition, sourcing therapeutic proteins from natural sources could pose significant safety risks (immune reactions, disease transmission). For example, in May 1985, all batches of HGH extracted from human pituitary glands were withdrawn from the market, after the discovery of contamination of these batches with prions, responsible for Creutzfeldt-Jakob disease (Walsh, 2000). The need to produce these therapeutic proteins by other means became imperative.

In the meantime, the 1970s saw the rise of molecular biology. Molecular biology was based on major fundamental discoveries dating from the middle of the 20th century: the identification of deoxyribonucleic acid (DNA) as the carrier of genetic information by Oswald Avery, Colin MacLeod and Maclyn McCarty in 1944; the discovery of the double helix structure of DNA by James Watson, Francis Crick, Maurice Wilkins, and Rosalind Franklin in 1953; and the discovery of restriction enzymes by Werner Arber in 1962. In 1972, Paul Berg's experiments revealed a method for covalently joining DNA fragments, thus creating hybrid DNA molecules (Jackson et al., 1972). Considered as the father of recombinant DNA, Paul Berg opened the way to production of recombinant proteins.

A recombinant protein results from the expression of an exogenous recombinant DNA by a cell or a host organism. A recombinant protein production system is based on (i) a gene of interest (GOI), coding for the protein of interest (POI); (ii) an expression vector or plasmid, containing the elements necessary for the expression of the GOI and possibly the elements necessary for the selection of stable clones; (iii) an expression host system, ensuring the production of the recombinant protein.

The production of recombinant proteins is a major progress, which has made possible to respond to the limitations of the production of therapeutic proteins by natural sources (quantities produced and safety) and which has allowed the development of new drugs, in particular through protein engineering (Walsh, 2000).

1.1.2 Since the 2010s, monoclonal antibodies dominate the market for recombinant proteins for therapeutic use

Since the introduction of recombinant insulin in 1982, 365 biopharmaceutical products have received market approval from health authorities, like the Food and Drug Administration (FDA) in the United States of America and the European Medicines Agency (EMA) in Europe. From the mid-1990s to the mid-2010s, the number of approvals per five-year period was relatively stable (54 to 60 approvals per five-year period). Since 2015, this number has doubled, with 112 approvals over the 2015-2018 period (Walsh, 2006, 2010, 2014, 2018).

Monoclonal antibodies (mAbs) now dominate the market of recombinant proteins for therapeutic applications. Over the period 2015-2018, they accounted for more than half of all approvals granted by health authorities. Major success stories include adalimumab (Humira®, AbbVie, approved in 2002), rituximab (Rituxan®, Roche, approved in 1997), infliximab (Remicade®, Johnson & Johnson, approved in 1998), trastuzumab (Herceptin®, Roche, approved in 1998) and bevacizumab (Avastin®, Roche, approved in 2004), which are the five most lucrative mAbs to date. Sales of mAbs exceeded €100 billion in 2017 (Walsh, 2018). mAbs are currently used primarily for oncology and immunology indications (Gaughan, 2016): adalimumab, rituximab, and infliximab are used for the treatment of inflammatory diseases, such as non-Hodgkin's lymphoma, rheumatoid arthritis, ankylosing spondylitis, or psoriasis, among others; trastuzumab and bevacizumab are used for the treatment of breast cancer and ovarian cancer, among others.

Other types of biopharmaceutical products are, in descending order of number of approvals, hormones, coagulation factors, enzymes, vaccines, nucleic acid products (gene therapies, DNA or RNA vaccines, interfering RNAs, ...), and engineered cell-based products, respectively (Walsh, 2018).

Nearly 2,400 biopharmaceutical products are currently in clinical trials. In the coming years, mAbs are expected to continue to dominate the market for recombinant proteins for therapeutic applications. However, since the mid-2010s, a wide variety of molecules have been in development. Since the approval of muromonab (Orthoclone OKT3®, Janssen-Cilag) in 1983, the formats of mAbs have evolved considerably, in particular with the progressive chimerization and then humanization of the variable parts, with the aim of reducing immunogenicity and therefore improving efficacy (Buss et al., 2012; J. Li & Zhu, 2010; Lu et al., 2020). The potential of mAbs, in terms of specificity and treatment efficacy, paves the way for new generations of mAb-derived molecules, some of which are already on the market: (i) antibody fragments (single-chain variable fragment (scFv), single domain antibody (SdAb), fragment antigen-binding (Fab)) (ii) antibodies conjugated or fused to another molecule (Antibody-Drug Conjugate (ADC), Antibody-Radionuclide Conjugate (ARC), Fc fusion proteins) and (iii) multimeric antibody fragments (BsAb, Bispecific Antibody, SdAb tandem, Fc tandem, dimeric

antibody fragment) (Birrer et al., 2019; Carter, 2006; Frenzel et al., 2013; Goulet & Atkins, 2020; Walsh, 2004; Wang et al., 2019).

1.1.3 Stages in the development of recombinant protein production processes

A process for producing a recombinant protein can be presented in five steps: (i) generation of a stable clonal cell line expressing GOI; (ii) production of POI, by culturing the producer line; (iii) purification of POI; (iv) formulation and aliquoting; (v) process and final product control by analytical methods (Figure 1). The first step is the cell line development (CLD) step, the second step is the USP ("UpStream Process") step, the third and fourth steps are the DSP ("DownStream Process") steps (Carrara et al., 2021; Steinmeyer & McCormick, 2008). The CLD step is decisive for the development of recombinant protein production processes. Indeed, it allows the generation of a monoclonal cell line with a high and stable expression level of the recombinant protein. This is a prerequisite to produce recombinant proteins according to current Good Manufacturing Practices (cGMP) (Frenzel et al., 2013).

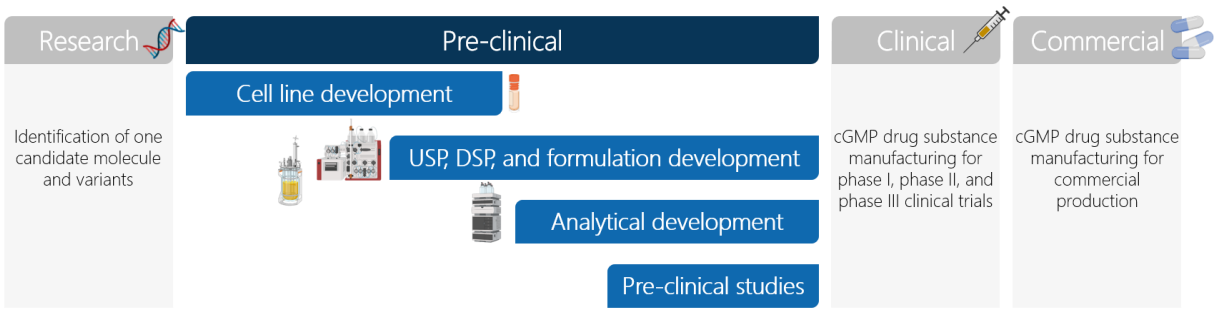


Figure 1: Stages in the development of processes for recombinant protein production

The similarities in the structure of mAbs allow a platform approach for the development of production processes. Thus, each step of the recombinant mAb production process constitutes a platform in its own right: a CLD platform, a USP platform, a DSP platform, and an analytical platform. A CLD platform includes a host cell line, an expression vector or plasmid containing the GOI, and a cell line generation process (Steinmeyer & McCormick, 2008).

The choice of host cell line is particularly important: the selected expression system must allow for high production of the POI in an active conformation (Birch & Racher, 2006; Carvalho et al., 2017). 60-70% of recombinant proteins for therapeutic use are produced in mammalian cells (F. M. Wurm, 2004). Specifically, 95% of mAbs are produced in mammalian cells (Frenzel et al., 2013). Rodent, murine, simian, or human cell lines are currently used to produce recombinant proteins (Demain & Vaishnav, 2009). Despite the variety of cell lines available, 70 to 80% of recombinant proteins for therapeutic use are produced in Chinese Hamster Ovary (CHO) cells (Jayapal, Wlashin, et al., 2007; Tripathi & Shrivastava, 2019).

Pharmaceutical companies frequently outsource development and manufacturing steps for recombinant protein production, to Contract Development and Manufacturing Organizations (CDMOs).

Development and manufacturing of therapeutic recombinant protein relies on the use of specialized equipment and facilities (Steinmeyer & McCormick, 2008). Merck End-to-End Solutions, part of the BioReliance® portfolio, Life Science, has a CDMO business, from process development to clinical and commercial production. Regarding Merck End-to-End Solutions' cell line development CDMO activity, two cell line generation platforms are currently offered: the CHOZN® CHO K1 platform and the CHOZN® GS$^{-/-}$ platform. The CHOZN® CHO K1 platform constitutes the core of the work presented in this thesis.

1.1.4 Progress and challenges of industrial platforms for cell line development

Since the 1980s, the fundamental concepts of cell line development have remained unchanged. Thus, cell line development is based on a well-established scheme, which starts with the transfection of the host cell line with an expression vector. The expression vector contains the transcriptional and translational gene elements necessary for expression and biosynthesis of the recombinant protein, as well as elements that confer a selective advantage to the cells that have integrated the plasmid. After transfection, cells that have integrated the expression vector into their genome are selected, either in the absence of an essential metabolite or in the presence of an antibiotic. The surviving cells are then individually isolated to ensure the generation of monoclonal cell lines from a single cell. The clones thus generated are then evaluated on growth and production criteria of the protein of interest (quantity and quality). At the end of these evaluations, sometimes iterative, a cell line presenting the desired characteristics of growth and productivity is selected. This cell line will be the recombinant protein-producing cell line (Lai et al., 2013; F. Li et al., 2010; F. M. Wurm, 2004).

The increasing need for recombinant proteins and the competitiveness of the market have made the optimization of cell line generation platforms a real challenge for the biopharmaceutical industry. This has resulted in a considerable increase in the productivity of cell lines generated by CLD platforms. Currently, a few companies claim CLD processes that allow the generation of recombinant cell lines producing up to 10 g/L (mAb-producing CHO-derived lines).

These productivity levels are the joint result of several decades of fundamental research and applied research and development. This work has included, but was not limited to: host cell line engineering, optimization of transcriptional and translational elements for recombinant protein biosynthesis and for expression of resistance to selection pressure, optimization of transfection and cloning methods and equipment, optimization of clone screening methods and equipment, and use of semi-targeted or targeted integration systems (Bandaranayake & Almo, 2014; Lai et al., 2013; F. M. Wurm, 2004).

Challenges remain: (i) increasing cell line's productivity, (ii) reducing cell line development timelines, to significantly decrease the time required for process development and enable faster time-to-market, and (iii) miniaturization and automation of cell line development processes. The work presented in this thesis has the ambition to participate to these research and development efforts.

1.2 Objectives of the thesis

The objective of the thesis presented here is to understand the critical parameters of the cell line generation process for the expression of recombinant proteins in CHO cells, from the point of view of each element of a cell line development platform: the host line, the expression vector, and the stages of the cell line generation process.

This work was used to optimize the CHOZN® CHO K1 cell line generation platform to produce therapeutic recombinant proteins.

1.3 Organization of the thesis

The literature review begins with a general presentation of recombinant proteins for therapeutic application, with particular emphasis on the case of monoclonal antibodies and their derivatives. The literature review then addresses the subject under the axis of each element of a cell line development platform: the host line, the expression vector, and the cell line generation process. Finally, a detailed overview of the CHOZN® CHO K1 cell line development platform initial state focuses on the genetic elements of the expression vector and on the cell line generation process.

Our work is then presented in four chapters.

In the first chapter, growth characteristics and initial performance (in terms of productivity, stability and timelines) of the CHOZN® CHO K1 cell line development platform are defined. This initial platform performance characterization is then considered as point of comparison for the optimizations made during this project and presented in the following chapters. Two main areas of work are identified following this initial characterization: (i) reduction of cell line development timelines and (ii) increase of recombinant cell lines' productivity.

In the second chapter, a shortened cell line development process, called direct single-cell cloning (DSCC), is designed and compared to the traditional cell line development process, based on a mini-pool and a clone generation. Cell line development timelines are significantly decreased but clonal cell lines productivity are also considerably reduced, strengthening the need to focus the work on productivity increase.

In the third chapter, new promoters controlling puromycin N-acetyltransferase expression are evaluated. Three CHO K1 endogenous promoters are identified and four partially deleted SV40E promoters are designed. Lower transcriptional strength, compared to SV40 early promoter, is confirmed for all evaluated promoters. Several promoters lead to an increased productivity of recombinant cell lines, which is confirmed on several recombinant proteins.

Lastly, in the fourth chapter, the methodology and the results linked to the development of the reverse transcription quantitative polymerase chain reaction (RT-qPCR) method are presented. This method was specially developed for the needs of this project, for gene copy number and gene transcription level analysis of rituximab heavy chain, rituximab light chain, and puromycin N-acetyltransferase.

2 Literature review

The literature review section consists of six chapters:

(v) Therapeutic recombinant proteins. The objective of this chapter is to provide insights on the characteristics of therapeutic recombinant proteins, with a detailed focus on antibodies and on the evolution of therapeutic antibody formats.

(vi) The host line, expression system for the recombinant protein. This chapter begins with a preamble on recombinant protein expression systems. The objective of this chapter is to highlight the characteristics of CHO cells as hosts for production of therapeutic recombinant proteins, and to present the vast genealogy of CHO cells to understand the origins of the CHOZN® CHO K1 host cell line.

(vii) The expression vector, support of genetic information. This chapter begins with general information on expression vector elements for stable gene expression in mammalian cells. The objective of this chapter is to present the possible optimization strategies of the expression vector to increase productivity level and productivity stability.

(viii) Cell line development process. This chapter begins with a description of the general workflow of a cell line development process. The objective of this chapter is to provide a focus on the two key steps of any cell line development process, i.e., transfection and single-cell cloning, as well as a focus on the productivity stability of CHO cells.

(ix) Overview of CHOZN® CHO K1 cell line development platform. The objective of this chapter is to provide a detailed overview of the initial state of the CHOZN® CHO K1 cell line development platform, with a focus on genetic elements of the expression vector and on the cell line generation process.

(x) Literature review synthesis.

2.1 Recombinant proteins for therapeutic use

Drugs currently on the market are either generated by chemical synthesis or produced by biotechnological processes. Drugs generated by chemical synthesis are generally of low molecular weight, with a well-defined structure, fully characterized, stable and with little dependence on the production process, and generally non-immunogenic. Conversely, drugs generated by biotechnological processes are generally high molecular weight proteins, complex in structure, partially characterized, sensitive to environmental conditions and highly dependent on the production process, and potentially immunogenic (Declerck, 2012). Thus, the final product is usually a more or less homogeneous mixture of the target protein and various structural variants of it. By the early 2020s, drugs produced by biotechnological processes, also known as biopharmaceuticals or biologic drugs, accounted for the

majority of drugs developed by the pharmaceutical industry (Tripathi & Shrivastava, 2019; Walsh, 2018).

To date, 365 biopharmaceuticals have been granted marketing authorization, issued by health authorities (Walsh, 2006, 2010, 2014, 2018). Recently approved biopharmaceuticals include monoclonal antibodies, hormones, clotting factors, enzymes, vaccines, nucleic acid-based products, and engineered cell-based products. Monoclonal antibodies now account for more than half of the market authorization approvals, representing an estimated $300 billion market by 2025 (Lu et al., 2020; Walsh, 2018).

2.1.1 Post-translational modifications

2.1.1.1 Importance of post-translational modifications

The complexity of the structures of biopharmaceuticals requires to use living cells or organisms to produce them. Indeed, the majority of recombinant proteins for therapeutic application exhibit post-translational modifications (PTMs) (Walsh & Jefferis, 2006). Post-translational modifications are enzymatic modifications made after translation that result in a mature protein (Kuriakose et al., 2016). The main post-translational modifications carried by proteins are: acetylation, acylation, ADP-ribosylation, amidation, γ-carboxylation, β-hydroxylation, disulfide bound formation, glycosylation, phosphorylation, proteolytic transformation, sulfation. Only certain post-translational modifications (acylation, amidation, γ-carboxylation, β-hydroxylation, disulfide bound formation, glycosylation, proteolytic transformation, and sulfation) are present on currently approved recombinant proteins and influence their stability and biological activity (Walsh & Jefferis, 2006).

Post-translational modifications are dependent on the recombinant protein production process. At the CLD stage, the choice of expression system is critical, as some expression systems are unable to produce given post-translational modifications (*e.g.*, production of aglycosylated proteins in the form of inclusion bodies in *Escherichia coli*) or produce post-translational modifications that are different from those found in humans (*e.g.*, production of glycosylation profiles that are different from those in yeast or insect cells) (Walsh & Jefferis, 2006). At the USP step, the culture conditions, namely the culture mode (batch, fed-batch or perfusion), the composition of the culture medium and nutrient solutions, and the physicochemical parameters (temperature, pH, dissolved oxygen, dissolved carbon dioxide, osmolarity), influence both the productivity and the glycosylation profile of the recombinant protein (Brülhmann et al., 2015; Cymer et al., 2018). Finally, at the DSP step, purification conditions can lead to the selection of variants of the protein with certain post-translational modifications (Walsh & Jefferis, 2006).

2.1.1.2 Glycosylation, a post-translational modification for more than one third of biopharmaceuticals

2.1.1.2.1 Impacts of glycosylation

Glycosylation is a widespread and complex PTM, which is characteristic of eukaryotic surface proteins and excreted proteins. More than one third of biopharmaceuticals are glycoproteins, *i.e.*, proteins covalently linked to linear or branched oligosaccharides, also called glycans. The presence and nature (composition, length and branching) of the attached oligosaccharide can impact the biological function, solubility, stability, protection against proteolysis, half-life, folding, addressing and immunogenicity of the protein, particularly in the case of antibodies, blood factors, interferons, and few hormones (Hossler et al., 2009; Raymond, 2015; Walsh & Jefferis, 2006).

More specifically, in the case of monoclonal antibodies, glycosylation significantly impacts the therapeutic effect of antibodies whose mechanism of action involves immune-mediated effector functions, such as antibody-dependent cellular cytotoxicity (ADCC), antibody-dependent cellular phagocytosis (ADCP) or complement-dependent cytotoxicity (CDC). As for other post-translational modifications, many parameters of the production process can impact the glycosylation profiles (Cymer et al., 2018).

2.1.1.2.2 Glycoprotein biosynthesis and control mechanisms

Oligosaccharides can bind to the oxygen atom of the hydroxyl group of a threonine (Thr) or serine (Ser) residue (O-glycosylation, consensus sequence not identified) or to the nitrogen atom of the amide group of an asparagine (Asn) residue (N-glycosylation, consensus sequence Asn-X-Ser/Thr, where X represents all amino acids except proline) (Bause, 1983; Cymer et al., 2018). Glycoprotein assembly begins in the endoplasmic reticulum (ER), where an N-glycan precursor attached to a lipid precursor is transferred to a protein under synthesis. The maturation changes of the oligosaccharide continue in the Golgi apparatus. These cleavage and elongation steps are performed by glycosidases and glycosyltransferases (Aebi, 2013; Raymond, 2015; Walsh & Jefferis, 2006).

A control mechanism, called the calnexin/calreticulin cycle, allows to verify the conformation of the glycoproteins produced in the ER and directs only folded proteins to their site of action (Aebi et al., 2009). Misfolded proteins are either directed to a new folding cycle or to a degradation pathway called endoplasmic reticulum associated degradation (ERAD) (Feige et al., 2010). Despite this control mechanism, the accumulation of misfolded proteins can lead to an unfolded protein response (UPR) (Adamson et al., 2016; Hetz et al., 2020; Omasa et al., 2010). This response allows: (i) to temporarily slow down protein synthesis in order to decrease the amount of protein in the ER (translational pathway); (ii) to activate a set of pathways for the synthesis of chaperone proteins and proteins involved in degradation (transcriptional pathway); (iii) to initiate apoptosis if the two previous pathways have failed (apoptotic pathway) (Schröder & Kaufman, 2005).

2.1.1.2.3 Characteristics of human glycosylation

Health authorities pay particular attention to glycosylation, requiring human-like glycosylation profiles for recombinant proteins, regardless of the production system used (Jefferis, 2009).

The different recombinant protein production hosts each have their own N-glycosylation machinery: glycosidases, glycosyltransferases, and variable monosaccharide synthesis capacity (Raymond, 2015). Human N-glycans are classified into three categories: oligomannosidic N-glycans, also called "highmannose", hybrid N-glycans and complex N-glycans. Human N-glycans are characterized by a penta-saccharide core, composed of three mannose residues and two N-acetylglucosamine residues (GlcNAc). The specific capabilities of the different expression systems (prokaryotic and eukaryotic) are shown on Figure 2.

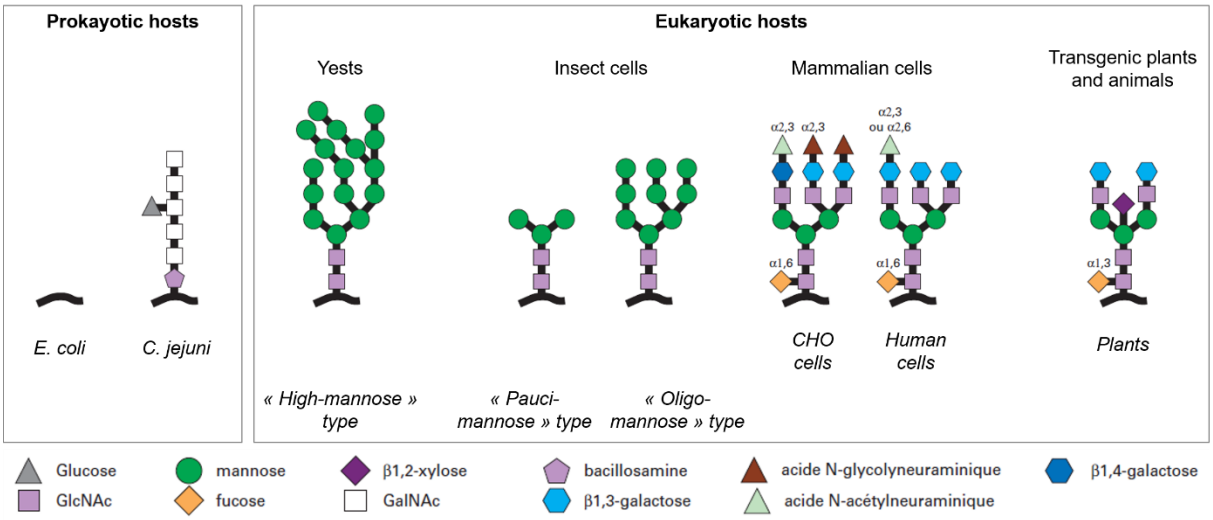


Figure 2: Specificities of N-glycans depending on expression system

The different expression systems (prokaryotic systems, single-cell eukaryotic systems, and multi-cell eukaryotic systems) have different N-glycosylation machineries, i.e. a particular set of glycosyltransferases and glycosidases, as well as variable monosaccharide synthesis capacities, which leads to different N-glycans. Adapted from Goergen et al. (2011).

2.1.2 General information on antibodies

Immunoglobulins, also commonly called antibodies, belong to the eponymous immunoglobulin superfamily (IgSF), which includes immunoglobulins, major histocompatibility complex (MHC) molecules, cell adhesion molecules, coreceptors and some cytokine receptors. These are major proteins of the adaptive immune system, secreted by plasma cells, whose function is to recognize and neutralize antigens and to recruit effector cells of the immune system (IS) (H. W. Schroeder & Cavacini, 2010) (Buss et al., 2012).

The existence of antibodies in the blood was demonstrated in 1890 by the German physician Emil von Behring and the Japanese physician Kitasato Shibasaburo. The term "antibody" was first used in 1891.

Nearly fifty years later, in 1939, the Swedish biochemist Arne Tiselius and the American immunochemist Elvin Kabat demonstrated the presence of antibodies in the γ-fraction of serum by electrophoresis (Tiselius et al., 1939). The terms γ-globulin and immunoglobulin were then used for the first time. Since then, all the work carried out by the scientific community has emphasized the complex nature of immunoglobulins (H. W. Schroeder & Cavacini, 2010).

2.1.2.1 Structure of immunoglobulins

Immunoglobulins are glycoproteins composed of two identical heavy chains (HC) and two identical light chains (LC). The heavy chains, composed of 450 amino acids and a molecular weight of about 50 kDa, are glycosylated and linked together by a variable number of disulfide bonds and non-covalent interactions. The light chains, composed of 220 amino acids and a molecular weight of about 25 kDa, are linked to the heavy chains by a disulfide bond and by non-covalent interactions. Heavy and light chains are respectively formed by four or five and two globular domains, called immunoglobulin domains (Feige et al., 2010; H. W. Schroeder & Cavacini, 2010). The immunoglobulin domains are each composed of a β-sheet of 110 amino acids, of which 60 amino acids form a loop held by an intra-catenary disulfide bond (Williams & Barclay, 1988). Thus, heavy chains are composed of an N-terminal variable immunoglobulin domain (noted V_H) and three or four C-terminal constant immunoglobulin domains (noted C_H1, C_H2, C_H3 and C_H4). Light chains are composed of an N-terminal variable light domain (noted V_L) and a C-terminal constant light domain (noted C_L). The variable domains show a very large amino acid sequence diversity between immunoglobulins. In addition, they each contain three hypervariable regions (HV), also called complementary determining regions (CDR). The antigen-binding site, called the paratope, is composed of six combined HV regions (Chiu et al., 2019; Feige et al., 2010; H. W. Schroeder & Cavacini, 2010).

The structure of immunoglobulins could be deduced by proteolysis with papain and pepsin (Porter, 1959). Enzymatic digestion with papain allows the isolation of the two antigen binding fragments (Fab) and the crystallizable fragment (Fc). The Fab contains the variable and constant domains of the light chain (V_L and C_L), as well as the variable domain and a constant domain of the heavy chain (V_H and C_H1). Fc contains the two or three constant domains (C_H2 and C_H3 or C_H2, C_H3 and C_H4) of the two heavy chains. It allows binding to Fc receptors (FcRs), which are involved in the recruitment of immune effector cells and the activation of complement. The Fab and Fc regions are separated by a hinge region (Chiu et al., 2019). The hinge region is rich in cysteines and prolines and highly flexible, in order to adapt the angle between the two Fab regions and thus facilitate antigen binding (Bloom et al., 1997).

The structure of an immunoglobulin is shown in Figure 3.

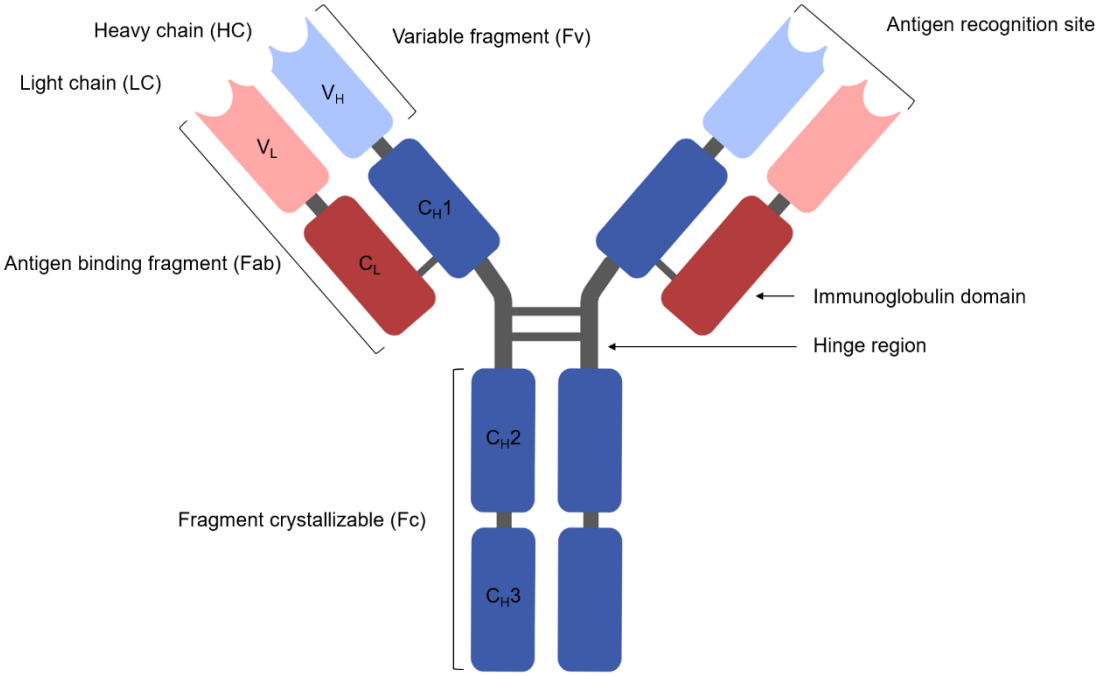


Figure 3 : Schematic representation of an immunoglobulin

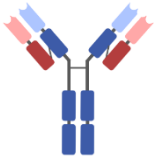
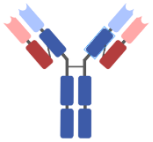
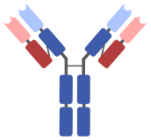
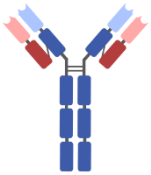
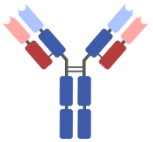
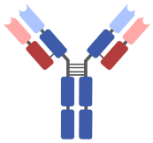
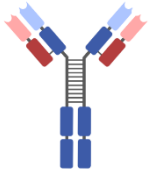
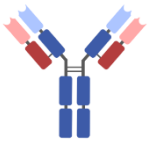
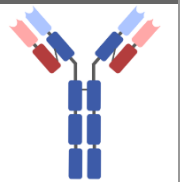
Immunoglobulins are composed of two identical heavy and two light chains, consisting of immunoglobulin domains. The two antigen-binding fragments are linked to the crystallizable fragment by a hinge region. Dark and light red: light chains (LC); dark and light blue: heavy chains (HC); V_L: light chain variable domain; C_L: light chain constant domain; V_H: heavy chain variable domain; C_Hx: heavy chain constant x domain. Adapted from Goulet & Atkins (2020). Created with BioRender.com.

2.1.2.2 Classes d'immunoglobulines

There are five classes of immunoglobulins: IgG, IgM, IgA, IgE and IgD. These classes are defined by the heavy chain isotype. IgG, IgM, IgA, IgE and IgD have γ, μ, α, ε and δ isotype heavy chains, respectively. Light chains are of κ or λ isotype, regardless of the immunoglobulin class (Feige et al., 2010). The properties of the five classes of immunoglobulins are given in Table 1.

Table 1: Properties of immunoglobulin classes and subclasses

Immunoglobulins are divided into five classes (IgA, IgD, IgE, IgG and IgM), which are further divided into two (IgA1 and IgA2) and four subclasses (IgG1, IgG2, IgG3 and IgG4). Each class and subclass is characterized by properties related to their structure, biology and effector functions. ADCC: antibody dependent cellular-mediated cytotoxicity); ADCP: antibody dependent cellular phagocytosis); CDC: complement dependent cytotoxicity. Adapted from Goulet & Atkins (2020), Keyt et al. (2020) and Steffen et al. (2020). Partially created with BioRender.com.

Class	**IgA1**	**IgA2**	**IgD**	**IgE**	**IgG1**	**IgG2**	**IgG3**	**IgG4**	**IgM**
Structure									
Light chain isotype	κ or λ	κ or λ	κ or λ	κ or λ	κ or λ	κ or λ	κ or λ	κ or λ	κ or λ
Heavy chain isotype	α1	α2	δ	ε	γ1	γ2	γ3	γ4	μ
Heavy chain domains	C_H1, C_H2, C_H3	C_H1, C_H2, C_H3	C_H1, C_H2, C_H3	C_H1, C_H2, C_H3, C_H4	C_H1, C_H2, C_H3	C_H1, C_H2, C_H3	C_H1, C_H2, C_H3	C_H1, C_H2, C_H3	C_H1, C_H2, C_H3, C_H4
Monomer structure	Monomer (serum) Dimer (external secretions)	Monomer (serum) Dimer (external secretions)	Monomer	Monomer	Monomer	Monomer	Monomer	Monomer	Monomer (membrane) Pentamer (secreted)
Valence	2 (monomer) 4 (dimer)	2	2	2	2	2	2	2	2 (monomer) 5 (pentamer)
Monomer molecular weight [kDa]	160	160	184	188	146	146	165	146	194
Numer of amino acids in the hinge region	23	10	64	0	15	12	62	12	0
Number of disulfide bonds	3	3	1	2	2	4	11	2	3

Class	IgA1	IgA2	IgD	IgE	IgG1	IgG2	IgG3	IgG4	IgM
between heavy chains									
Concentration in serum [g/L]	3	0.5	0.03	0.00005	9	3	1	0.5	1.5
Half-life in serum	6 days	6 days	3 days	2 days	8 to 21 days	20 days	7 days	21 days	5 to 10 days
Predominant antigen	Proteins	Glycolipids and glycoproteins	-	Allergens	Proteins	Glycolipids and glycoproteins	Proteins	Allergens	Glycolipids and glycoproteins
Effector function	ADCC ADCP CDC	ADCC ADCP CDC	Poorly known	ADCC	ADCC ADCP CDC	ADCC CDC	ADCC ADCP CDC	ADCC ADCP	CDC
Examples	-	-	-	-	Rituximab Trastuzumab	-	-	Nivolumab Pembrolizumab	-

In humans, IgG is the most abundant immunoglobulin. They represent 80% of the immunoglobulins in circulation in the blood. Four subclasses of IgG are defined according to their relative serum concentration: 60%, 25%, 10% and 5% respectively for IgG1, IgG2, IgG3 and IgG4 (Jefferis, 2009). These IgG subclasses differ in the number of inter-heavy chain disulfide bonds and amino acids at the hinge region. IgG1, IgG2, IgG3 and IgG4 have respectively 2, 4, 11 and 4 disulfide bonds in their hinge region.

All the monoclonal antibodies that have obtained a marketing authorization are IgG, mostly of subclass 1 and to a lesser extent of subclasses 2 and 4 (Kaplon et al., 2022; Kaplon & Reichert, 2021a; Ryman & Meibohm, 2017). Subclass 3 IgGs are not used as therapeutic antibodies. Their hinge region, containing 11 disulfide bonds and 62 amino acids long, which gives them an increased sensitivity to proteolysis, and their short half-life in serum, make any therapeutic application complex (Goulet & Atkins, 2020; Saito et al., 2019).

2.1.2.3 Evolution of therapeutic antibody formats

2.1.2.3.1 From murine antibodies to human antibodies, for a decrease in immunogenicity

In contrast to polyclonal antibodies, which are a mixture of antibodies directed against different epitopes of a given antigen, monoclonal antibodies are specific to a single epitope. The first monoclonal antibodies were created with the development of the hybridoma technique by the German immunologist George Köhler and the Argentine biochemist César Milstein (Köhler & Milstein, 1975). The hybridoma technique consists in: (i) immunizing a mouse with a given antigen; (ii) recovering secondary lymphoid organs (spleen or lymph nodes) containing B lymphocytes and plasma cells; (iii) fusing plasma cells and B lymphocytes with myeloma cells; by means of polyethylene glycol (PEG) or electrofusion; (iv) select the hybridomas thus formed in hypoxanthine, aminopterin, thymidine (HAT) medium, in which B lymphocytes and myelomas cannot grow; (v) isolate by limit dilution the hybridomas; (vi) screen the hybridomas. The hybridomas thus generated produce murine monoclonal antibodies of unique and known specificity. This technique was used for the production of muromonab (Orthoclone OKT3®, Janssen-Cilag), first monoclonal antibody to receive marketing authorization in 1986 (J. Li & Zhu, 2010).

Since then, the formats of mAbs have evolved considerably and are shown in Figure 4. Indeed, antibodies generated by the hybridoma technique led to significant immune reactions in patients, including a response that produced human anti-mouse antibodies (HAMA) (Ryman & Meibohm, 2017). Thus, in order to decrease their immunogenicity, mAbs have progressively become chimeric, then humanized, then human. Chimeric mAbs, such as infliximab (Remicade®, Janssen Biotech Inc.), the cetuximab (Erbitux®, Eli Lilly) and the rituximab (Rituxan®, Roche/Genentech), are composed of variable V_L and V_H fragments from murine sources and constant human C_L, C_H1, C_H2 and C_H3 fragments. Humanized mAbs, such as trastuzumab (Herceptine®, Roche/Genentech), the alemtuzumab

(Lemtrada®, Genzyme) and the pembrolizumab (Keytruda®, Merck) are made up of human variable and constant fragments, but have murine hypervariable regions. Today, many mAbs are fully human, such as adalimumab (Humira®, AbbVie Inc.) and the nivolumab (Opdivo®, Bristol-Myers Squibb). These successive advances have made it possible to progressively reduce the immunogenicity of mAbs, without however eliminating it entirely: some patients develop human anti-chimeric antibodies (HACA) or even human anti-human antibodies (HAHA) (Buss et al., 2012; Grilo & Mantalaris, 2019; Harding et al., 2010; Lu et al., 2020).

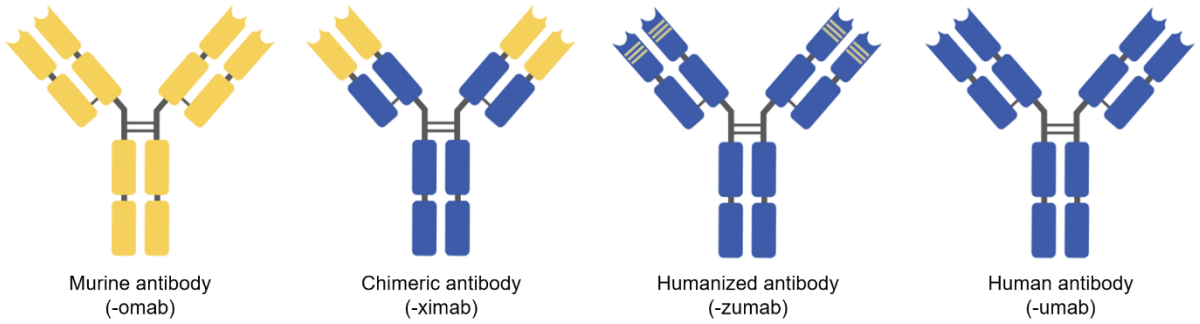


Figure 4: Therapeutic antibodies and associated nomenclature

Therapeutic antibodies can be murine (suffix -omab), chimeric (suffix -ximab), humanized (suffix -zumab) or human (suffix -umab). The International Nonproprietary Name (INN) imposes a nomenclature on the name of mAbs, indicating their origin. Adapted from Buss et al. (2012). Created with BioRender.com.

The International Nonproprietary Name (INN) imposes a nomenclature on the name of mAbs, indicating their origin: suffix -omab for murine antibodies, suffix -ximab for chimeric antibodies, suffix -zumab for humanized antibodies, and suffix -umab for human antibodies (Buss et al., 2012).

2.1.2.3.2 From antibodies to antibody derivatives, to increase the diversity of therapeutic molecules

The potential of mAbs, in terms of specificity and treatment efficacy, opens the way to new generations of mAb-derived molecules. These new molecule formats are specifically designed to have improved characteristics, such as better pharmacokinetics or increased efficacy (J. Li & Zhu, 2010). The possibilities for antibody derivatives are endless. The most common ones are shown in Figure 5 (Birrer et al., 2019; Carter, 2006; Frenzel et al., 2013; Goulet & Atkins, 2020; Wang et al., 2019):

(i) Antibody fragments: single-chain variable fragment (scFv), single domain antibody (SdAb), antigen-binding fragment (Fab), Fab where the CH1 is extended by part of the hinge region (Fab'), Fab' dimer (F(ab')2);

(ii) Antibodies conjugated or fused to another molecule: antibody-drug conjugate (ADC) molecules, antibody-radioisotope conjugate (ARC) molecules, fusion proteins with Fc fragment (Fc fusion);

(iii) Multimers of antibody fragments: bispecific antibodies (BsAb), tandem SdAb, tandem Fc, dimeric antibody fragment (Goulet & Atkins, 2020; Labrijn et al., 2019).

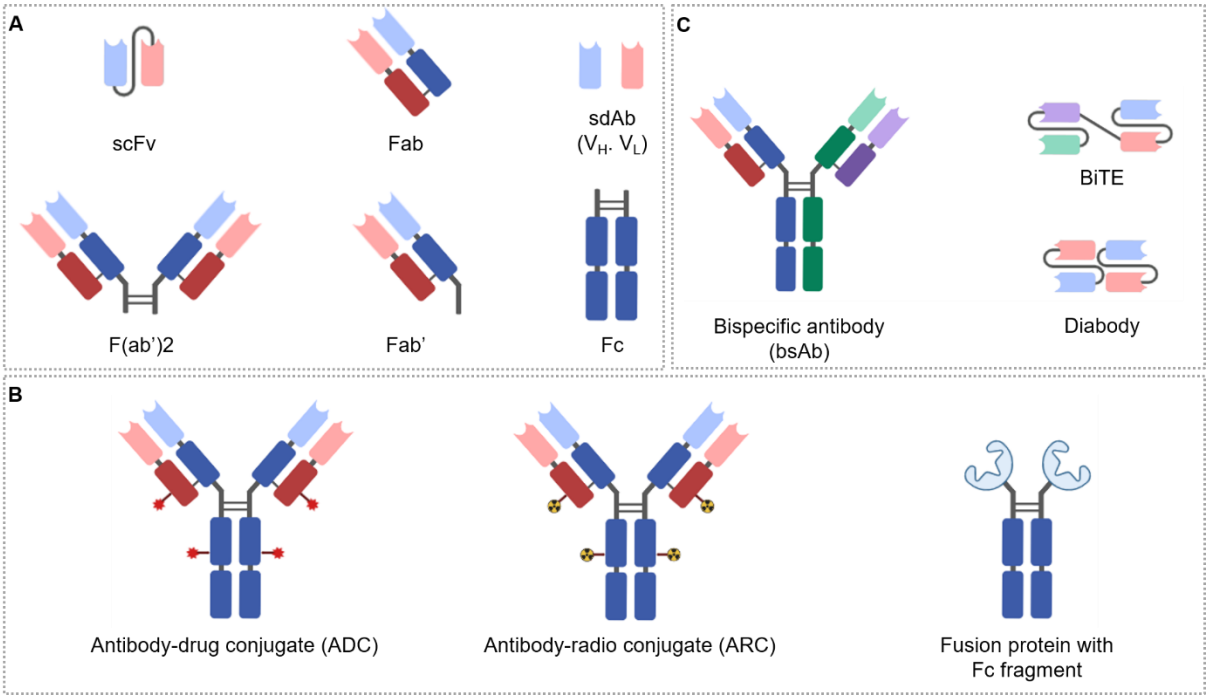


Figure 5: Examples of antibody-derived molecules

New generations of mAb-derived molecules are specifically designed to have improved characteristics, such as better pharmacokinetics or increased efficacy. ***(A)*** *Antibody fragments: single-chain variable fragment (scFv), single domain antibody (SdAb), antigen-binding fragment (Fab), Fab where the CH1 is extended by part of the hinge region (Fab'), Fab' dimer (F(ab')2).* ***(B)*** *Antibodies conjugated or fused to another molecule: antibody-drug conjugate (ADC) molecules, antibody-radioisotope conjugate (ARC) molecules, fusion proteins with Fc fragment (Fc fusion).* ***(C)*** *Multimers of antibody fragments: bispecific antibodies (BsAb), tandem SdAb, tandem Fc, dimeric antibody fragment. Adapted from Einsele et al. (2020) and Goulet & Atkins (2020). Created with BioRender.com.*

2.2 The host line, expression system for the recombinant protein

2.2.1 Preamble on recombinant protein expression systems

The choice of the expression system for a recombinant protein is decisive. It is justified by the quality and functionality of the recombinant protein produced (particularly in the case of a recombinant protein for therapeutic application in humans), as well as by the time, yield and cost of production. Several expression systems exist and are currently used industrially for the production of recombinant proteins: prokaryotic systems and eukaryotic monocellular or multicellular systems. Some of these systems are presented below.

2.2.1.1 Prokaryotic systems, ideal for the production of non-glycosylated recombinant proteins

Insulin, the first recombinant protein to obtain marketing authorization in 1982 (Humulin®, Genentech/Eli Lilly), was produced in a prokaryotic system, by the bacteria *Escherichia coli*. Bacterial systems, with *E. coli* in the lead, offer many advantages for the production of recombinant proteins. These systems allow a rapid and massive production (in the order of several grams per liter) of the protein of interest, with simple and inexpensive production processes. For these reasons, until the early 1990s, *E. coli* was the main expression host for heterologous production of recombinant proteins (D. C. Andersen & Krummen, 2002; J. R. Swartz, 2001).

However, after the successful production of recombinant insulin, difficulties were encountered in the expression of other molecules, such as human growth hormone (HGH) and tissue plasminogen activator (tPA). Indeed, recombinant proteins produced by *E. coli* are often misfolded and accumulated in the cytoplasm as insoluble and inactive aggregates, called inclusion bodies. The absence of post-translational modifications (PTMs), the difficulty of proper refolding and the presence of endotoxins limit the use of bacterial systems for the production of complex recombinant proteins (Demain & Vaishnav, 2009; Frenzel et al., 2013; Tripathi & Shrivastava, 2019).

Since then, many advances have been made to improve the expression of recombinant proteins in bacterial systems. These optimizations include: the use of inducible promoters, the use of modified bacterial strains, the co-expression of chaperone proteins, the reduction of temperature, the secretion of the protein of interest into the periplasm or extracellular medium, the reduction of the production rate of the protein of interest, the modification of the culture medium, the expression as a fusion protein, the expression of fragments of the protein of interest, and the denaturation and in vitro refolding of the protein of interest (Demain & Vaishnav, 2009; J. R. Swartz, 2001; Tripathi & Shrivastava, 2019).

In addition, since the 2010s, the so-called "cell free expression" systems (CFE systems) have also undergone significant technological progress. CFE systems are based on the use of cell extracts, mainly

from *E. coli*, to allow the in vitro synthesis of recombinant proteins (Garenne et al., 2021; Silverman et al., 2020; J. Swartz, 2006).

2.2.1.2 Single-cell eukaryotic systems

2.2.1.2.1 Yeasts, ideal for the production of small recombinant glycosylated proteins without therapeutic application

The hepatitis B vaccine, a recombinant vaccine that received marketing authorization in 1986 (Recombivax HB®, Merck), was produced in a eukaryotic system by the yeast *Saccharomyces cerevisiae*. Several yeasts are used as expression hosts for the production of recombinant proteins: *S. cerevisiae*, *Pichia pastoris*, *Hansenula polymorpha*. These systems allow, like bacterial systems, a rapid and massive production (of the order of several grams to ten grams per liter) of the protein of interest, with simple production processes and at accessible costs. On the other hand, unlike bacterial systems, some yeasts are able to synthesize proteins with certain PTMs (glycosylation, three-dimensional structure, and disulfide bonds in particular) (Huang et al., 2014; Nielsen, 2012).

Nevertheless, the secretion and correct folding of complex proteins is difficult. The glycosylation profiles carried out by some yeasts (of the "high-mannose" type) are different from those found in humans, and sometimes immunogenic, preventing any therapeutic application (Demain & Vaishnav, 2009; Durocher & Butler, 2009; Tripathi & Shrivastava, 2019).

Since then, many advances have been made to improve the expression of recombinant proteins by yeast. These optimizations include: the use of yeast strains modified for the genes involved in glycosylation, the use of inducible promoters, the modification of culture media, the optimization of culture parameters (Demain & Vaishnav, 2009; J. R. Swartz, 2001; Tripathi & Shrivastava, 2019).

2.2.1.2.2 Insect cells, mainly used for the production of human and veterinary vaccines

The cervical cancer vaccine, a recombinant vaccine that received marketing authorization in 2007 (Cervarix®, GSK), was produced in a eukaryotic system, using insect cells. Several insect lines are used as expression hosts for the production of recombinant proteins: Sf9 and Sf21 (*Spodoptera frugiperda*), S2 (*Drosophila melanogaster*), BTI-Tn-5B1-4 (High-Five™ or H5) and Tn-368 (*Trichoplusia ni*), using baculovirus as expression vector (Shen et al., 2015). These systems allow massive production (in the range of several grams to ten grams per liter) of the protein of interest. The proteins produced can be complex, the three-dimensional folding and PTMs being correctly performed by the insect cells, and are correctly addressed, thanks to an efficient cleavage of the signal peptide (Demain & Vaishnav, 2009).

However, the production of recombinant proteins in insect cells relies on a transient expression system, as the infection of insect cells by the recombinant baculovirus must be performed at each production. Moreover, the glycosylation profiles achieved by insect cells (of the "pauci-mannose" or "oligo-

mannose" type) are comparable but not identical to those found in humans (Figure 2) (Durocher & Butler, 2009; Tripathi & Shrivastava, 2019).

Many advances have been made to improve the expression of recombinant proteins by insect cells. These optimizations include: the co-expression of mammalian glycosyltransferases and the gene of interest, the use of modified lines to express mammalian glycosyltransferases, the use of modified lines free of adventitious viruses, the use of plasmids allowing the expression of multiprotein complexes (Geisler & Jarvis, 2018; Tripathi & Shrivastava, 2019; Zhai et al., 2019).

Since 2000, the baculovirus system and insect cells have been used in industrial applications. These systems are mainly used for the production of Virus-Like Particles (VLP) and recombinant vaccines, for human or veterinary applications (Durocher & Butler, 2009; Felberbaum, n.d.).

2.2.1.2.3 Mammalian cells, the reference for the production of complex recombinant proteins for therapeutic applications

In the 1980s, difficulties were encountered in producing glycosylated proteins, such as HGH, tPA, or erythropoietin (EPO), in *E. coli*. The focus then shifts to mammalian cells. tPA, produced in a eukaryotic system, by CHO (Chinese Hamster Ovary) cells, obtained marketing authorization in 1987 (Activase®, Genentech). Since the 1990s, mammalian cells have taken over from non-mammalian expression systems (Walsh, 2018). Currently, 60 to 70% of recombinant proteins for therapeutic use are produced in mammalian cells (F. M. Wurm, 2004). Several cell types are now used: murine cells (CHO cell lines (hamster ovary), NS0 and Sp2/0 (mouse myeloma), BHK (baby hamster kidney)), simian cells (Vero cell line (monkey epithelium)), and human cells (PER.C6 lines (retina), HEK293 (renal epithelium), HeLa (cervical epithelial carcinoma), HT-1080 (epithelial cells derived from connective tissue)) (Demain & Vaishnav, 2009). Despite the variety of available cell lines, 70 to 80% of recombinant proteins for therapeutic use are produced in CHO (Jayapal, Wlashin, et al., 2007; Tripathi & Shrivastava, 2019).

Mammalian cells allow the production of complex proteins, with human PTMs and glycosylation profiles close (murine cells) or identical (human cells) to those found in humans (Figure 2). The proteins produced have a biological activity comparable to that of the native protein. However, the low productivity of mammalian cells, the very high production costs and the risk of culture contamination are the main drawbacks of these expression host systems (Demain & Vaishnav, 2009; Durocher & Butler, 2009; Frenzel et al., 2013).

Many advances have been made to improve the expression of recombinant proteins by mammalian cells. These optimizations concern in particular the optimization of the producing lines and the optimization of the culture parameters in bioreactor (F. M. Wurm, 2004). The optimization of the production lines is the subject of this thesis. The literature review will therefore address this area in more detail.

2.2.1.3 Multi-cell eukaryotic systems: transgenic plants and animals, marginal systems for the production of recombinant proteins

Antithrombin III, a recombinant protein that obtained marketing authorization in 2006 (ATryn®, GTC Biotherapeutics), was produced in a multicellular eukaryotic system by transgenic goats. Several eukaryotic pluricellular systems are used for the production of recombinant proteins: transgenic plants (tobacco, corn, etc.), with trials reported as early as the 1980s, and transgenic animals (goats, sheep, etc.), with numerous trials published in the 1990s (Demain & Vaishnav, 2009).

Transgenic animals allow massive production (in the order of several tens of grams per liter) of complex glycosylated proteins, mainly in milk, sometimes in urine. Transgenic plants allow simple and inexpensive production of complex glycosylated proteins, mainly in seeds or leaves. Nevertheless, the glycosylation profiles are different from those found in humans, and are sometimes immunogenic (Figure 2). In addition, the slow development of transgenic animals and plants, public opinion against the use of genetically modified organisms, and ethical issues related to animal welfare complicate industrial applications for the production of recombinant proteins (Durocher & Butler, 2009; Pardo et al., 2009; Tripathi & Shrivastava, 2019).

To date, the use of transgenic animals or plants for the production of recombinant proteins remains marginal. However, since 2010, several molecules are currently undergoing clinical trials or being marketed (Schillberg et al., 2019; Yao et al., 2015).

2.2.2 Characteristics of CHO lines to produce recombinant therapeutic proteins

70-80% of therapeutic recombinant proteins produced in mammalian systems are produced in CHO cells (Jayapal, Wlashin, et al., 2007; Tripathi & Shrivastava, 2019). Popularity of CHO cells for therapeutic recombinant protein production finds its justification in several reasons linked to health safety, therapeutic efficiency, and ease of industrial processes.

First, CHO cells are recognized as safe hosts for production of therapeutic recombinant proteins by health authorities (Food and Drug Administration (FDA) in the United States, and European Medicines Agency (EMA) in Europe) (Bandaranayake & Almo, 2013). CHO cells have withstood the test of time for more than 30 years, therefore establishing their safety profile. Many therapeutic recombinant antibodies are currently produced in CHO cells (to name a few, rituximab, trastuzumab, pembrolizumab, and nivolumab are currently produced in CHO cells) (Chu & Robinson, 2001; Dumont et al., 2016; Lalonde & Durocher, 2017). Consideration as safe hosts mainly relies on the fact that CHO cells proved to have reduced susceptibility to viral infection, compared with other mammalian cell lines used in the production of recombinant proteins, such as baby hamster kidney (BHK) cells (Berting et al., 2010). Most human pathogenic viruses do not replicate in CHO cells: this is namely the case of human immunodeficiency virus (HIV), poliovirus, herpes simplex virus (HSV), influenza virus, and measles

virus (Jayapal, Wlaschin, et al., 2007). Conversely, the use of hamster-derived cells is perceived as species barrier protection, thus reducing the potential risk of transfer of contaminating adventitious agents to humans (Dumont et al., 2016). Many viral entry genes, especially coding for cell adhesion molecules (CAMs), plasma membrane proteins, or histone proteins, do not show any detectable expression in the CHOK1 transcriptome, therefore possibly explaining CHO cells' reduced susceptibility to viral infection (Dumont et al., 2016; X. Xu et al., 2011). Nonetheless, viral contaminations of CHO cell-derived harvests have occurred (shut down of Genzyme's Massachusetts plant being one of the latest examples (Bethencourt, 2009)), mostly due to trace levels of viruses in raw materials (Chiang et al., 2019). Though, current adventitious agent testing (AAT) methods proved to be efficient for viral detection (Berting et al., 2010).

Second, CHO cells have the capacity for efficient post-translational modifications and can produce recombinant glycoproteins with human-like glycans thanks to an advanced mammalian folding, secretion and post-translational apparatus (Butler & Spearman, 2014; Frenzel et al., 2013). Therefore, recombinant proteins are both compatible and bioactive in humans (Jayapal, Wlaschin, et al., 2007; J. Y. Kim et al., 2012). Nonetheless, CHO cells are unable to produce some types of human glycosylations, such as α-1,3/4-fucosylation and α-2,6-sialylation, and can scarcely gamma-carboxylate proteins (Lalonde & Durocher, 2017; Tripathi & Shrivastava, 2019). Conversely, CHO cells produce very low levels of non-human glycans, such as N-glycolylneuraminic acid (Neu5Gc) and galactose-α1,3-galactose (α-gal), and toward which the human immune system can produce antibodies, therefore neutralizing the therapeutic molecule (Bosques et al., 2010; Ghaderi et al., 2012).

Third, CHO cells are suitable for cell line development. CHO cells can easily be transfected, with various techniques: electroporation, nucleofection, lipofection, calcium phosphate-based transfection, or polycationic agents-based transfection (J. Y. Kim et al., 2012; F. M. Wurm, 2004). In addition, a wide variety CHO derived cell lines are available, such as CHOK1, CHOS, CHO DHFR$^{-/-}$ and CHO GS$^{-/-}$, which differ in growth and productivity characteristics (Frenzel et al., 2013; Hacker et al., 2009; Reinhart et al., 2019). Consequently, several selection systems are available, among which dihydrofolate reductase (DHFR) and glutamine synthetase (GS) selection systems, with their respective methotrexate (MTX) based and methionine sulfoximine (MSX) based gene amplification systems, are the most popular (Durocher & Butler, 2009; Lai et al., 2013; F. M. Wurm, 2004). CHO cells are also suited for large scale industrial upstream and downstream processes (Lalonde & Durocher, 2017). CHO cells have been adapted to suspension cell culture, in which they can reach high viable cell densities, and to chemically defined, animal component free, serum free culture media, therefore reducing upstream process variability (Bandaranayake & Almo, 2014). Nonetheless, mammalian cell culture media are complex, and those of CHO cells are no exception to the rule. Culture media require over fifty various components, and often require supplementation of growth factors, amino acids, reducing agents, and vitamins (Owczarek et al., 2019). Downstream process was also positively impacted by these

improvements on culture media. Viral and prion contamination risk was decreased, and reduction of protein contaminants simplified process requirements (Bandaranayake & Almo, 2014; Rodrigues et al., 2013). However, CHO based processes, as all mammalian cell-based processes, are expensive. The manufacturing of recombinant proteins produced by mammalian cells in a validated plant requires 55 to 85 million dollars per year in costs of material, manufacturing, and plant construction and validation (Demain & Vaishnav, 2009).

2.2.3 CHO cell lines' genealogy and associated selection systems

Various CHO cell lines exist and have been used for recombinant protein production over the past 40 years (F. M. Wurm, 2013). Several of these cell lines are maintained in the world's biological resource centers: the American Type Culture Collection (ATCC) and the European Collection of Authenticated Cell Cultures (ECACC), both created in the 1920s to maintain cell banks (Gayatri Dhara et al., 2018).

Although some dates, laboratory names, and researchers' names are mentioned in the literature, it is impossible today to establish the exhaustive history of CHO cell lines, from their origin (ovary of a female Hamster from China) to the currently used lines (F. M. Wurm, 2013; F. Wurm & Wurm, 2017). The CHO cell lines currently used by the biotechnology industries, in particular CHO K1, CHO S, CHO DBX11, CHO DG44, CHO K1 SV/$GS^{-/-}$, represent a large family of related cell lines. These cell lines were derived from the original CHO-ori cell line by cloning, selection, and mutagenesis mechanisms (Turilova et al., 2021). The term "quasi-species" is besides mentioned by several authors in reference to CHO cell lines (F. M. Wurm, 2013). A partial genealogy of CHO cell lines is illustrated on Figure 6.

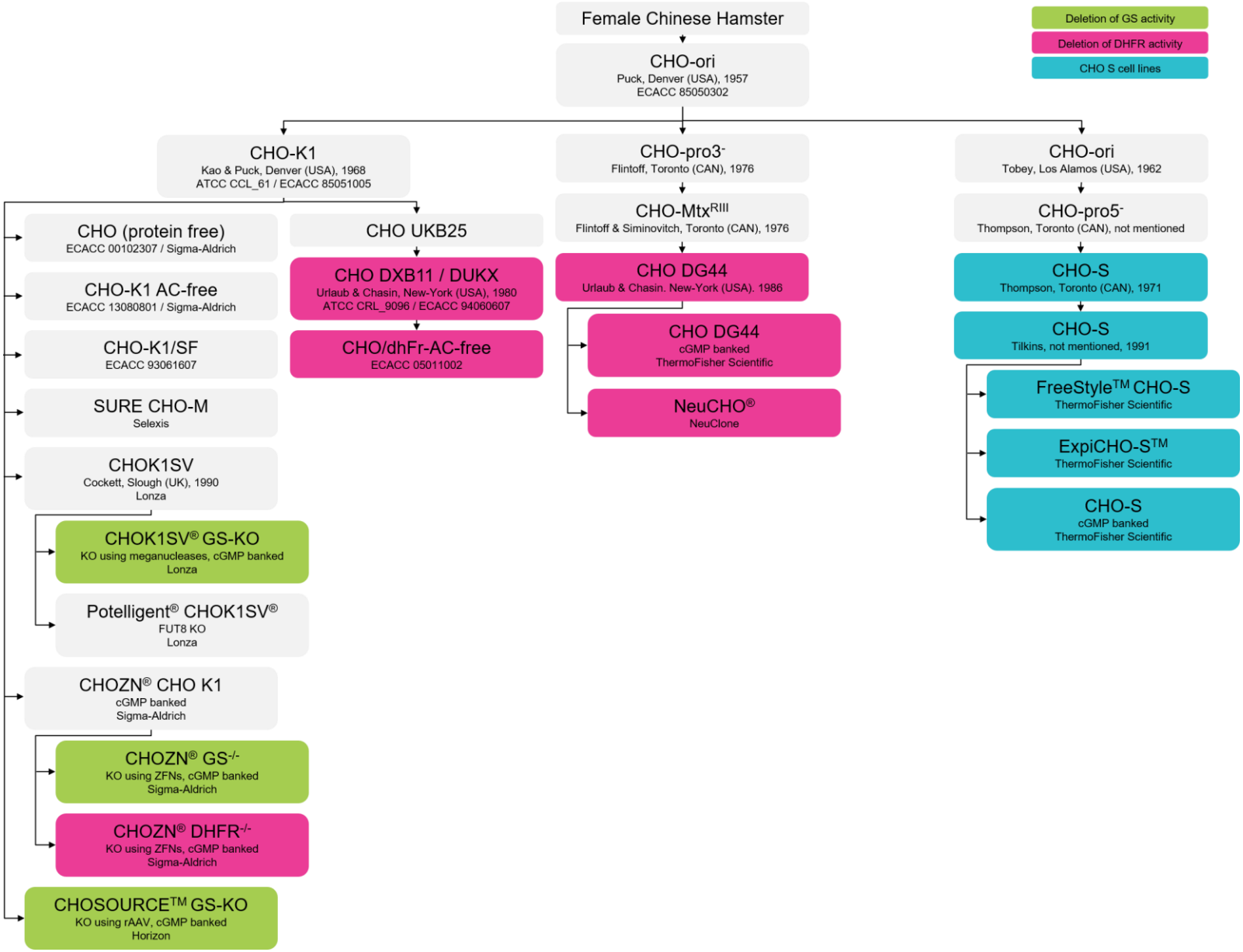

Female Chinese Hamster
CHO-ori
Puck, Denver (USA), 1957
ECACC 85050302
Deletion of GS activity
Deletion of DHFR activity
CHO S cell lines
CHO-K1
Kao & Puck, Denver (USA), 1968
ATCC CCL_61 / ECACC 85051005
CHO-pro3-
Flintoff, Toronto (CAN), 1976
CHO-ori
Tobey, Los Alamos (USA), 1962
CHO (protein free)
ECACC 00102307 / Sigma-Aldrich
CHO UKB25
CHO-Mtx RIII
Flintoff & Siminovitch, Toronto (CAN), 1976
CHO-pro5-
Thompson, Toronto (CAN), not mentioned
CHO-K1 AC-free
ECACC 13080801 / Sigma-Aldrich
CHO DXB11 / DUKX
Urlaub & Chasin, New-York (USA), 1980
ATCC CRL_9096 / ECACC 94060607
CHO DG44
Urlaub & Chasin. New-York (USA). 1986
CHO-S
Thompson, Toronto (CAN), 1971
CHO-K1/SF
ECACC 93061607
CHO/dhFr-AC-free
ECACC 05011002
CHO DG44
cGMP banked
ThermoFisher Scientific
CHO-S
Tilkins, not mentioned, 1991
SURE CHO-M
Selexis
NeuCHO®
NeuClone
FreeStyle™ CHO-S
ThermoFisher Scientific
CHOK1SV
Cockett, Slough (UK), 1990
Lonza
ExpiCHO-S™
ThermoFisher Scientific
CHOK1SV® GS-KO
KO using meganucleases, cGMP banked
Lonza
CHO-S
cGMP banked
ThermoFisher Scientific
Potelligent® CHOK1SV®
FUT8 KO
Lonza
CHOZN® CHO K1
cGMP banked
Sigma-Aldrich
CHOZN® GS-/-
KO using ZFNs, cGMP banked
Sigma-Aldrich
CHOZN® DHFR-/-
KO using ZFNs, cGMP banked
Sigma-Aldrich
CHOSOURCE™ GS-KO
KO using rAAV, cGMP banked
Horizon

Figure 6: Partial genealogy of CHO cell lines

Various CHO cell lines exist and have been used for recombinant protein production over the past 40 years. Several of these cell lines are maintained in the world's biological resource centers: the American Type Culture Collection (ATCC) and the European Collection of Authenticated Cell Cultures (ECACC), both created in the 1920s to maintain cell banks. Although some dates, laboratory names, and researchers' names are mentioned in the literature, it is impossible today to establish the exhaustive history of CHO cell lines, from their origin (ovary of a female Hamster from China) to the currently used lines. The CHO cell lines currently used by the biotechnology industries, in particular CHO K1, CHO S, CHO DBX11, CHO DG44, CHO K1 SV/GS-/-, represent a large family of related cell lines. These cell lines were derived from the original CHO ori cell line by cloning, selection, and mutagenesis mechanisms.

CHO family tree designed in 2020. Mathilde Bas (ORCID ID: 0000-0002-5492-3439) equally contributed to this work.

2.2.3.1 From Chinese Hamster to CHO-ori cell line

In 1957, Theodore Puck (University of Colorado, City of Denver, USA) received a female Chinese Hamster (*Cricetulus griseus*), sent by George Yerganian (Children's Cancer Research, City of Boston, USA) (Puck, 1957). Ovarian tissue cells, cultured and maintained for more than ten months, showed a fibroblast phenotype and a quasi-diploid karyotype. It is during this culture that the cells underwent a spontaneous phenomenon of immortalization, the origin of which is probably genetic, but is still not understood. Cloning of these cells led to the original CHO line, commonly called CHO-ori (F. M. Wurm, 2013; F. Wurm & Wurm, 2017). This line is registered with the World Biological Resource Centers under reference 85050302 (ECACC) (no known reference at ATCC).

Theodore Puck's work initially focused on the cytogenetics of mammalian cells. CHO cells had three main advantages over other available mammalian lines. First, the Chinese Hamster genome is diploid, containing only 11 pairs of large and easily identifiable chromosomes. The genome of CHO cells, derived from Chinese Hamster ovary tissue, is pseudo-diploid and also contains large and easily identifiable chromosomes, although significant structural changes are notable(F. Wurm & Wurm, 2017). Second, the cells were fast growing and robust (F. Wurm & Wurm, 2017). Third, induced or spontaneous mutations easily led to the generation of metabolic mutants (F. M. Wurm, 2013). From the late 1950s to the late 1970s, these characteristics made CHO cells very popular for cell genetic studies. It was in the 1980s, after unsuccessful attempts to produce tPA in E. coli, that CHO cells were used for the production of recombinant proteins.

CHO-ori cells are deficient in proline synthesis. This characteristic is common to all CHO-ori derived cell lines (F. M. Wurm & Hacker, 2011). In mammalian cell lines, proline synthesis normally occurs in mitochondria through two pathways. Δ-pyrroline-5-carboxylate (P5C), which is an intermediate in proline synthesis, is either formed from glutamate by P5C-synthase or from ornithine by ornithine-aminotransferase (OAT). Both pathways are dysfunctional in CHO cell lines (Sun et al., 2020).

2.2.3.2 CHO K1 cell line

CHO K1 line was derived by Fa-Ten Kao and Theodore Puck (University of Colorado, Denver, USA) in 1968 from the CHO-ori cell line (Kao & Puck, 1968; F. M. Wurm, 2013). This line is registered in world biological resource centers under the references 85051005 (ECACC) and CCL 61 (ATCC).

CHO K1 cell line is compatible with several selection systems, such as antibiotic-based systems and metabolic marker-based systems. In the case of an antibiotic-based system, recombinant CHO K1 cell lines are obtained by (i) transfecting a vector carrying the resistance gene and the gene of interest, and then (ii) selecting cells that have stably integrated the plasmid in the presence of the antibiotic in culture medium. Commonly used antibiotics are blasticidin, geneticin (G418), hygromycin, puromycin, and

zeocin (Sun et al., 2020). Cells that have stably integrated the plasmid express resistance to the antibiotic and are therefore able to grow in the presence of the antibiotic in culture medium. In the case of a metabolic marker-based system, recombinant CHO K1 cell lines are obtained by (i) transfecting a vector carrying the metabolic marker gene and the gene of interest, and then (ii) selecting cells that have stably integrated the plasmid in the absence of the metabolite in culture medium. Commonly used metabolic marker-based systems are pyrroline-5-carboxylate synthetase system and glutamine synthetase system. In pyrroline-5-carboxylate synthetase system, selection is based on CHO K1 cell line natural deficiency for proline synthesis. Cells that have stably integrated the plasmid express pyrroline-5-carboxylate synthetase, which restores proline prototrophy. Cells are therefore able to grow in the absence of proline in culture medium. In glutamine synthetase system, cells that have stably integrated the plasmid express glutamine synthetase and are therefore able to grow in the absence of L-glutamine in culture medium. However, CHO K1 endogenous glutamine synthetase activity should be inhibited by addition of methionine sulfoximine (MSX) (Cockett et al., 1990). MSX is an ATP-dependent structural analog of glutamate, which highly potently inhibits glutamine synthetase (Eisenberg et al., 2000). Depending on the concentration of MSX in cell culture medium, addition of MSX either leads to inhibition of endogenous enzyme activity, increase of selection stringency or gene amplification (Feary et al., 2017).

2.2.3.3 CHO S cell line

CHO S cell line was derived by Larry Thompson and Raymond Baker (University of Toronto, Toronto, Canada) from CHO-pro5$^-$ cells. CHO-pro5$^-$ cells were most likely derived from CHO-ori cell line handled by Robert Tobey (Los Alamos, USA); however, traceability of handling is incomplete (F. M. Wurm, 2013).

CHO S cells are characterized by their ability to grow in suspension cultures. This is remarkable as it paved the way to large scale production of therapeutic recombinant proteins in stirred-tank bioreactors (STR) (Lindskog et al., 2018).

CHO S cell line is compatible with several selection systems, mentioned and described in CHO K1 cell line.

2.2.3.4 CHO DXB11 (or DUK-XB11) and CHO DG44 cell lines: deletion of dihydrofolate reductase (DHFR) activity

The CHO DXB11 cell line, also called DUK-XB11, was derived by Gail Urlaub and Lawrence Chasin (Columbia University, New York City, USA) in 1980 from the CHO K1 line (Urlaub & Chasin, 1980). This cell line is registered with the World Biological Resource Centers under the references CRL-9096™ (ATCC) and 94060607 (ECACC). The CHO DXB11 line has been used for the production of tPA (Activase®, Genentech) (F. M. Wurm, 2013).

Gail Urlaub and Lawrence Chasin's work focused on dihydrofolate reductase (DHFR), an enzyme that catalyzes the reduction of dihydrofolate to tetrahydrofolate, a cofactor required for carbon transfer in various nucleic acid precursor biosynthesis reactions (Urlaub & Chasin, 1980). The goal of Gail Urlaub and Lawrence Chasin was to delete dihydrofolate reductase (DHFR) activity. Exposure of the CHO K1 line to chemical mutagens and gamma radiation resulted in a line with a deletion of one DHFR allele and a missense mutation (T137R) in the second DHFR allele. This mutation prevents the cells from reducing dihydrofolate to tetrahydrofolate (F. M. Wurm, 2013). However, a low but detectable level of restoration of DHFR activity in DXB11 CHO cells, due to the missense mutation of the second DHFR allele, necessitated the identification of a system completely lacking DHFR activity.

The CHO DG44 cell line was derived by Gail Urlaub and Lawrence Chasin in 1983 from the CHO-Mtx^{RIII} mutant, isolated from CHO-pro3 cells by Wayne Flintoff, Sandra Davidson and Louis Siminovitch in 1976(Flintoff et al., 1976; Urlaub et al., 1983). This line is not registered with the world biological resource centers.

Exposure of CHO-pro3- cells to ethyl methane sulfonate (EMS), a mutagenic and teratogenic organic compound inducing random genetic mutations by nucleotide substitution, generated the CHO-Mtx^{RIII} mutant. Subsequently, exposure of the CHO-Mtx^{RIII} mutant to mutagenic gamma radiation generated a line doubly deleted for the DHFR gene (bi-allelic deletion of the loci located on chromosome 2 and on the Z2 chromosomal variant) (Flintoff et al., 1976; Lindskog et al., 2018; F. M. Wurm, 2013). Since then, the CHO DG44 line has been frequently used as an expression host for the production of recombinant proteins.

DHFR activity deletion makes $DHFR^{-/-}$ cells triply auxotrophic for glycine, hypoxanthine, and thymidine, therefore requiring the addition of these compounds to culture medium. This feature makes CHO DXB11 and CHO DG44 cell lines particularly attractive for transfection and stable expression of recombinant proteins. Recombinant CHO DBX11 cell lines are obtained by (i) transfecting a vector carrying the DHFR gene and the gene of interest, and then (ii) selecting cells that have stably integrated the plasmid in the absence of the glycine, hypoxanthine, and thymidine in culture medium. Indeed, transfection of a vector carrying the gene of interest and the DHFR gene allows the selection of stable clones in glycine, hypoxanthine and thymidine-free medium (GHT^- medium), by restoration of DHFR activity (Lindskog et al., 2018; F. M. Wurm & Petropoulos, 1994).

A widespread approach for increasing productivity of recombinant $DHFR^{-/-}$ cell lines relies on the use of methotrexate (MTX). MTX is a folate derivative component that completely and irreversibly inhibit DHFR activity (F. M. Wurm & Petropoulos, 1994). Depending on the concentration of MTX in cell culture medium, stepwise addition of MTX leads to increase of selection stringency or gene amplification of the transfected DHFR gene together with the gene of interest (F. M. Wurm & Hacker,

2011; F. M. Wurm & Petropoulos, 1994). Nonetheless, use of MTX tends to be avoided as MTX is classified among carcinogenic, mutagenic, and reprotoxic (CMR) substances.

2.2.3.5 CHO K1SV GS-KO and CHO GS$^{-/-}$ cell lines: deletion of glutamine synthetase (GS) activity

Several companies have developed cell lines deleted for glutamine synthetase (GS) activity, based on the model of cell lines deleted for DHFR activity. CHO K1SV GS-KO (Lonza), CHOZN GS$^{-/-}$ (Sigma-Aldrich) and CHOSOURCE CHO-GS KO (Horizon) cell lines were derived from CHO-K1 cells, using genome editing techniques to delete both alleles of glutamine synthetase gene. More precisely, CHO K1SV GS-KO (Lonza) cell line was engineered using mega-nucleases; CHOZN GS$^{-/-}$ (Sigma-Aldrich) cell line was engineered using zinc finger nucleases (ZFN) (details in 2.2.4CHOZN® cell lines: CHOZN® CHO K1, CHOZN® GS$^{-/-}$ and CHOZN® DHFR$^{-/-}$); and CHOSOURCE CHO-GS KO (Horizon) cell line was engineered using recombinant adeno-associated virus (rAAV).

The deletion of GS activity makes GS$^{-/-}$ cells auxotrophic for glutamine, therefore requiring the addition of this compound to culture medium. This feature makes CHO GS KO cell lines particularly attractive for transfection and stable expression of recombinant proteins. Recombinant CHO GS KO cell lines are obtained by (i) transfecting a vector carrying the GS gene and the gene of interest, and then (ii) selecting cells that have stably integrated the plasmid in the absence of the glutamine in culture medium. Indeed, transfection of a vector carrying the gene of interest and the GS gene allows the selection of stable clones in glutamine-free medium (GS^{-} medium), by restoration of GS activity. Glutamine is a source of carbon and nitrogen. In the GS$^{-/-}$ system, glutamate therefore becomes the source of nitrogen, which has the advantage of producing less ammonia (known as growth inhibiting compound) than glutamine metabolism.

In the same way as for DHFR$^{-/-}$ cell lines, recombinant GS$^{-/-}$ cell lines productivity can be increased, using methionine sulfoximine (MSX), whose mode of action has been described previously (details in CHO K1 cell line).

2.2.4 CHOZN® cell lines: CHOZN® CHO K1, CHOZN® GS$^{-/-}$ and CHOZN® DHFR$^{-/-}$

In 2005, Sigma-Aldrich company acquired a CHO-K1 vial deposited at ECACC [ECACC 85051005]. The adherent CHO-K1 cells were then adapted to suspension culture in a chemically defined medium. In 2006, the CHOZN® CHO K1 cell bank was manufactured under cGMP conditions (current Good Manufacturing Practices). After complete characterization, in accordance with the requirements of the health authorities for Master Cell Banks (MCB), the CHOZN® CHO K1 line [CHOK1-1VL, SAFC] was available for commercialization in 2016 (*CHOZN® Platform Technical Bulletin*, n.d.).

SAFC company holds two other CHOZN® cell lines: the CHOZN® GS$^{-/-}$ cell line [CHOGS-1VL, SAFC] and the CHOZN® DHFR$^{-/-}$ cell line [CHODHFR-1VL, SAFC]. Both cell lines were derived from

CHO-K1 cells, which have been genetically engineered with zinc finger nucleases. Zinc finger nucleases (ZFNs) are engineered restriction enzymes created by fusing a zinc finger DNA binding domain with the catalytic DNA cleavage domain of the type IIs restriction enzyme *FokI* from *Flavobacterium okeanokoites* (De Souza, 2012; Y.-G. Kim et al., 1996; Liljedahl et al., 2003; P.-Q. Liu & Miller, 2015). As genome editing technologies, ZFNs allow a double strand break (DSB) for a given sequence (Carroll, 2011; Gaj et al., 2013; Urnov et al., 2005). The cell then employs DNA repair mechanisms: either a non-homologous end joining (NHEJ) mechanism or a homology directed repair (HDR) mechanism (Orlando et al., 2010).

The development of the CHOZN® GS$^{-/-}$ cell line began in 2009. CHO-K1 cells, adapted to suspension culture, were transfected with the zinc finger nucleases ZFNGSA9075 and ZFNGSB9372 (Figure 7-A). The ZFNGSA9075 and ZFNGSB9372 nucleases target a sequence located in exon 6 of the glutamine synthetase gene (NCBI ID: 100764163) (P.-Q. Liu & Miller, 2010). This sequence encodes amino acids critical for the catalytic function of the substrate binding domain of glutamine synthetase (Almassy et al., 1986; Liaw & Eisenberg, 1994; P.-Q. Liu & Miller, 2010). Mutations in this sequence result in the synthesis of a non-functional protein. CHO-K1 cells, transfected with ZFNGSA9075 and ZFNGSB9372 nuclease, were isolated to clone single cells. Clone 2E3, identified as monoclonal, underwent extensive characterization, aiming at verifying the double deletion on the glutamine synthetase gene by sequencing, western blotting, and culture in glutamine-deficient medium. In 2010, the CHOZN® GS$^{-/-}$ cell bank was produced under cGMP conditions from clone 2E3. After complete characterization, the CHOZN® GS$^{-/-}$ cell line was available for commercialization in 2011 (*CHOZN® Platform Technical Bulletin*, n.d.).

The development of the CHOZN® DHFR$^{-/-}$ cell line also began in the late 2000s. CHO-K1 cells, adapted to suspension culture, were transfected twice with zinc finger nucleases (Figure 7-B). The first transfection aimed at deleting one allele of the dihydrofolate reductase gene (NCBI ID: 100689028); and the second transfection aimed at deleting the second allele. CHO-K1 cells were first transfected with the zinc finger nucleases ZFN9461 and ZFN9684 (Santiago et al., 2008). The ZFN9461 and ZFN9684 nucleases target a sequence located in exon 1 of the DHFR gene. A heterozygous clone with a 7 bp deletion on one of the DHFR alleles was isolated, and retransfected- with two pairs of zinc finger nucleases ZFN9461 and ZFN9684 (used during first transfection), and ZFN9477 and ZFN9476. The ZFN9477 and ZFN9476 nucleases target a sequence located in intron 1 of the DHFR gene. A clone with a 235 bp deletion on the second allele was isolated. The CHOZN® DHFR$^{-/-}$ cell bank was produced under cGMP conditions from this clone. After complete characterization, the CHOZN® DHFR$^{-/-}$ cell line was available for commercialization (*CHOZN® ZFN-Modified DHFR-/- CHO Cell Line Technical Bulletin*, n.d.).

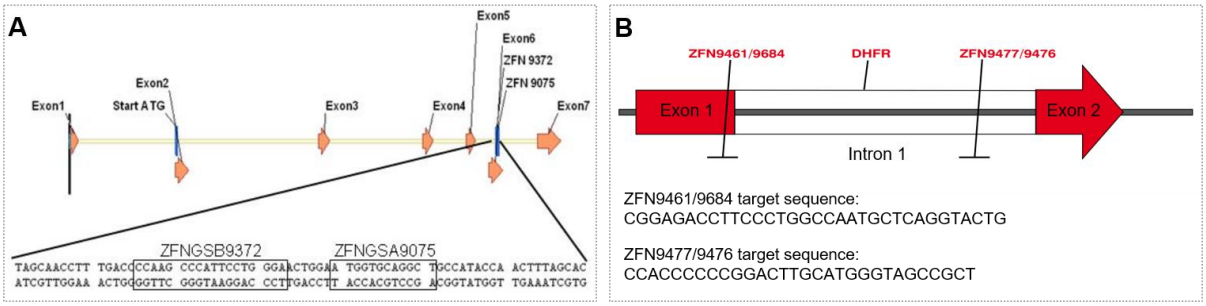


Figure 7: Sequences targeted by zinc finger nucleases for deletions in glutamine synthetase and dihydrofolate reductase genes

(A) Glutamine synthetase (GS) gene and sequences targeted by ZFNGSA9075 and ZFNGSB9372 nucleases. The glutamine synthetase gene (NCBI ID: 100764163) is located on chromosome 5 and has 8 exons. The ZFNGSA9075 and ZFNGSB9372 nucleases target the black boxed nucleotide sequences in exon 6 encoding the substrate binding domain of glutamine synthetase. (B) Dihydrofolate reductase (DHFR) gene and sequences targeted by ZFN9461/9684 and ZFN9477/9476. The dihydrofolate reductase gene (NCBI ID: 100689028) is located on chromosome 2. The ZFN9461 and ZFN9684 nucleases target a site located in exon 1. The ZFN9477 and ZFN9476 nucleases target a site located in intron 1. Adapted from (CHOZN® Platform Technical Bulletin, n.d.; CHOZN® ZFN-Modified DHFR-/- CHO Cell Line Technical Bulletin, n.d.).

The CHOZN® CHO K1, CHOZN® GS$^{-/-}$, and CHOZN® DHFR$^{-/-}$ cell lines are part of the CHOZN® platform, which is a recombinant protein expression platform in CHO cells. This platform contains media and nutrient solutions specially developed for the needs of CHOZN® lines: EX-CELL® CD CHO Fusion expansion medium [14365C/24365C, SAFC], EX-CELL® CHO Cloning medium [6366C, SAFC], EX-CELL® Advanced™ CHO Fed-batch production medium [14366C/24366C, SAFC] and EX-CELL® Advanced™ Feed 1 feeding solution [24367C/24368C, SAFC]. EX-CELL® CD CHO Fusion, EX-CELL® Advanced™ CHO Fed-batch, and EX-CELL® Advanced™ Feed 1 are chemically defined (CD), animal component free (ACF) media or solutions, which contain no hypoxanthine, thymidine, and L-glutamine. EX-CELL® CHO Cloning medium is an animal component free (ACF) medium, which contain no L-glutamine. The composition of these media and nutrient solutions is a trade secret.

2.3 The expression vector, support of genetic information

Recombinant protein production can be performed either through transient or stable expression, which are, by nature, very different. In transient gene expression, heterologous DNA is not integrated into the host cell genome and is lost through cell division in the absence of plasmid replication. Expression capabilities of the host cell line are exploited within days following transfection (Bernard, 2005). Transient gene expression is a suitable method for rapid production (as quickly as in two to four days) of small quantities of recombinant protein (tens-of-milligrams/liter range at laboratory scale, hundreds-of-milligrams/liter range after scale-up at pilot scale), namely for research purposes such as initial protein characterization, vector functionality testing and optimization, and for preclinical studies (Kaufman, 2000; Makrides, 1999, 2003; F. M. Wurm, 2004). Several cell lines are routinely used for transient gene expression, such as COS, BHK, and HEK cells (CHO cells have though long been considered less efficient) (Makrides, 1999; F. M. Wurm, 2004). On the other hand, in stable gene expression, heterologous DNA is integrated into the host cell genome (except for artificial chromosomes, such as bacterial or yeast artificial chromosomes (BAC and YAC)). Expression capabilities of the host cell line are exploited after months of labor-intensive selection of high producer clones (Bernard, 2005). Stable gene expression is a suitable method to generate high producing clones (grams/liter range after cell line development), for which properties of clonality, integration and stability of expression have been demonstrated (Cacciatore et al., 2010). Several cell lines are routinely used for stable gene expression, but as mentioned before, CHO cells are the most popular (Jayapal, F., et al., 2007).

Recombinant protein expression, either in transient or stable mode, relies on the use of an expression vector, which is, among others, the support of genetic information of the gene of interest (GOI).

2.3.1 General information on expression vector elements for stable gene expression in mammalian cells

Expression vectors for stable gene expression in mammalian cells are plasmid vector of few kilobases, which contain genetic elements for: (i) the expression of the gene of interest, (ii) the expression of a selection marker, and (iii) the production of the plasmid within bacterial system (prokaryotic origin of replication, also named replicon (ori site), and selection marker). Elements for expression, either of the gene of interest or of the selection marker, are gathered within an expression cassette, which contains genetic elements for transcriptional and translational control.

2.3.1.1 Elements for transcriptional control

Elements for transcriptional control include promoter, enhancer, introns, poly-adenylation signal and transcription terminator (Makrides, 1999, 2003). Other elements related to transcription and transcription control may be included in the vector, such as ubiquitous chromatin opening elements

(UCOEs), scaffold or matrix attachment regions (S/MARs), stabilizing anti-repressor elements (STARs), or insulators (Saunders et al., 2015).

2.3.1.2 Elements for translational control

Elements for translational control include Kozak sequence, 5' untranslated region (5' UTR), codon usage (Table 2), 3' untranslated region (3' UTR) and termination codon (Makrides, 1999, 2003). Other elements related to translation and translation control may be included in the vector, such as internal ribosome entry sites (IRES).

Table 2: Codon usage in CHO cells

Alanine: Ala/A; arginine: Arg/R; asparagine: Asn/N; aspartic acid: Asp/D; cysteine: Cys/C; glutamic acid: Glu/E; glutamine: Gln/Q; glycine: Gly/G; histidine: His/H; isoleucine: Ile/I; leucine: Leu/L; lysine: Lys/K; methionine: Met/M; phenylalanine: Phe/F; proline: Pro/P; serine: Ser/S; threonine: Thr/T; tryptophane: Trp/W; tyrosine: Tyr/Y; valine: Val/V.

Codon	Amino acid	Frequency per thousand	Codon	Amino acid	Frequency per thousand	Codon	Amino acid	Frequency per thousand	Codon	Amino acid	Frequency per thousand
UUU	Phe	19.6	UCU	Ser	16.0	UAU	Tyr	13.1	UGU	Cys	9.1
UUC	Phe	22.0	UCC	Ser	16.5	UAC	Tyr	16.4	UGC	Cys	10.3
UUA	Leu	6.4	UCA	Ser	10.3	UAA	- (stop)	0.6	UGA	- (stop)	1.2
UUG	Leu	14.1	UCG	Ser	3.4	UAG	- (stop)	0.5	UGG	Trp	13.1
CUU	Leu	13.2	CCU	Pro	16.7	CAU	His	10.2	CGU	Arg	5.6
CUC	Leu	18.4	CCC	Pro	17.0	CAC	His	12.9	CGC	Arg	9.3
CUA	Leu	7.6	CCA	Pro	15.6	CAA	Gln	10.3	CGA	Arg	7.2
CUG	Leu	38.8	CCG	Pro	4.3	CAG	Gln	33.4	CGG	Arg	10.1
AUU	Ile	17.4	ACU	Thr	14.1	AAU	Asn	17.4	AGU	Ser	11.4
AUC	Ile	24.8	ACC	Thr	20.3	AAC	Asn	21.2	AGC	Ser	16.4
AUA	Ile	6.9	ACA	Thr	15.7	AAA	Lys	24.6	AGA	Arg	10.1
AUG	Met (start)	23.0	ACG	Thr	4.5	AAG	Lys	38.4	AGG	Arg	10.2
GUU	Val	11.6	GCU	Ala	22.5	GAU	Asp	24.6	GGU	Gly	12.8
GUC	Val	15.7	GCC	Ala	25.9	GAC	Asp	28.1	GGC	Gly	21.3
GUA	Val	7.8	GCA	Ala	16.3	GAA	Glu	28.4	GGA	Gly	15.8
GUG	Val	30.1	GCG	Ala	5.0	GAG	Glu	41.1	GGG	Gly	13.4

2.3.2 Strategies for expression vector optimization: towards enhanced productivity and stability

Optimization of expression vector elements is a key area of work to increase both productivity and expression level stability of recombinant cell lines. Four axes of optimization might be identified: enhancing recombinant protein expression, enhancing recombinant protein expression stability, attenuating selection marker strength, and targeting integration site.

2.3.2.1 Enhancing recombinant protein expression

Expression vector, and more precisely genetic elements in the expression cassette of the gene of interest, may be optimized to improve recombinant protein expression. Engineered viral promoters (Moritz et al., 2015; Romanova & Noll, 2018), endogenous promoters (Tossolini et al., 2022), precise control of light chain-heavy chain ratio (Lao-Gonzalez et al., 2021; Voronina et al., 2016), optimized signal peptides (Ramezani et al., 2017) are all possible ways for vector optimization to increase productivity.

2.3.2.2 Enhancing recombinant protein expression stability

Expression vector, and more precisely genetic elements in the expression cassette of the gene of interest, may be optimized to improve recombinant protein expression stability. Ubiquitous chromatin opening elements (UCOEs) proved to significantly enhance stability of recombinant protein expression (Betts & Dickson, 2015; Nematpour et al., 2017; Neville et al., 2017; Saunders et al., 2015; F. Zhang et al., 2010), and more efficiently than other chromatin remodeling elements (such as scaffold or matrix attachment regions (S/MARs), stabilizing anti-repressor elements (STARs) and insulators (Saunders et al., 2015)). Additionally, engineered viral promoters (Moritz et al., 2015; Romanova & Noll, 2018), promoters free of CpG islands (Ho et al., 2016), synthetic promoters (A. J. Brown & James, 2016; Zúñiga et al., 2019), and endogenous promoters (Tossolini et al., 2022) are all possible ways for vector optimization to increase productivity stability.

2.3.2.3 Attenuating selection marker strength

2.3.2.3.1 Preamble on selection systems

A wide variety of selection systems are available for selection of CHO stable cell lines. These selection systems fall within two categories: antibiotic-based systems and metabolic marker-based systems (Matasci et al., 2008; F. M. Wurm, 2004).

2.3.2.3.1.1 Antibiotic-based systems: blasticidin, geneticin, hygromycin, puromycin, zeocin, …

In the case of an antibiotic-based system, recombinant CHO K1 cell lines are obtained by (i) transfecting a vector carrying the resistance gene and the gene of interest, and then (ii) selecting cells that have stably integrated the plasmid in the presence of the antibiotic in culture medium. Cells that have stably integrated the plasmid express resistance to the antibiotic and are therefore able to grow in the presence of the antibiotic in culture medium (Yeo et al., 2017).

Commonly used antibiotics are blasticidin, geneticin (G418 or G-418), hygromycin, puromycin, and zeocin (Sun et al., 2020; Yeo et al., 2017). Details about gene conferring resistance, source organism of gene conferring resistance, antibiotic mode of action and usual working concentrations of these commonly used antibiotics for mammalian cell selection are given in Table 3. Comparison studies between these antibiotics have been performed several times (Delrue et al., 2018; C. Guo et al., 2021; Lanza et al., 2013). Lanza *et al.* conducted a study to evaluate the influence of several antibiotics on

obtaining selected pools and stable cell lines in human cells (HT1080 and HEK293), and ranked antibiotics as follows: zeocin > hygromycin ~ puromycin > neomycin (Lanza et al., 2013). Similarly, Delrue *et al.* showed that geneticin and hygromycin have the highest selectivity factor respectively for BHK-21 cells and HeLa cells (Delrue et al., 2018). Lastly, Guo *et al.* showed that both level of recombinant protein expression and cell-to-cell variability in recombinant protein expression are impacted by the choice of antibiotic. Resistance to geneticin and blasticidin led to the lowest productivities, resistance to puromycin and hygromycin led to intermediate productivities and resistance to zeocin led to highest productivities in HEK293 cells (C. Guo et al., 2021). To our knowledge, no similar study was conducted with CHO cells.

Table 3: Commonly used antibiotics for mammalian cell selection

Commonly used antibiotics are blasticidin, geneticin (G418 or G-418), hygromycin, puromycin, and zeocin. bsd: blasticidin-S deaminase (NCBI GenBank ID: BAA12074.1; UniProtKB ID: P0C2P0); neo: (NCBI GenBank ID: ; UniProtKB ID:); hygB: (NCBI GenBank ID: ; UniProtKB ID:); pac: puromycin N-acetyltransferase (NCBI GenBank ID: AAA64928.1; UniProtKB ID: P13249); sh ble: bleomycin resistance protein (NCBI GenBank ID: ; UniProtKB ID: P17493). Adapted from Yeo et al., 2017.

Antibiotic	**Mode of action**	**Gene confering resistance**	**Source organism of gene conferring resistance**	**Working concentrations**
Blasticidin	Translation inhibitor (inhibits termination step of translation)	*bsd*	*Aspergillus terreus*	2 to 10 µg/mL
Geneticin (G418 or G-418)	Translation inhibitor (blocks polypeptide synthesis at 80S; inhibits chain elongation)	*neo*	Transposons Tn601 (903) or Tn5	100 to 800 µg/mL
Hygromycin	Translation inhibitor (blocks polypeptide synthesis at 80S; inhibits chain elongation)	*hygB*	*Escherichia coli*	50 to 500 µg/mL
Puromycin	Translation inhibitor (disrupts peptide transfer on ribosomes causing premature chain termination)	*pac*	*Streptomyces alboniger*	1 to 10 µg/mL
Zeocin	Complexes with DNA; Causes strand scissions	*sh ble*	*Streptoalloteichus hindustanus*	100 to 400 µg/mL

Antibiotic-based selection is a widely used approach for selection of recombinant stable cell lines, particularly in academic research, but also in industrial cell line development platforms (Priola et al., 2016). This selection method has proven many advantages. First, antibiotic-based selection is easy to

implement. Indeed, it does not require host cell line modification to generate recombinant cell lines, nor expensive raw materials or complex equipment (Priola et al., 2016; Yeo et al., 2017). Second, antibiotic-based selection is highly efficient. Aminoglycoside antibiotics (an antibiotic family to which blasticidin, geneticin, hygromycin, and puromycin namely belong) are highly cytotoxic to eukaryotic cells (Rodolosse et al., 1998). High antibiotic toxicity therefore leads to selection in few days (namely mentioned in the literature for puromycin and blasticidin). Lastly, small size of resistance genes is compatible with expression vectors (Caputo et al., 2021).

Nonetheless, antibiotic-based selection has also been criticized:

(i) Antibiotic-based selection is leading to selection of recombinant stable cell lines with poor productivities. Indeed, only low levels of the antibiotic resistance are required for the transfected cells to survive selection (Yeo et al., 2017). Non-producing clones may also survive (Priola et al., 2016). In addition, and unlike metabolic marker-based systems (namely, DHFR and GS systems), antibiotic-based selection is incompatible with gene amplification methods. Other strategies should be implemented to increase recombinant cell lines' expression level (Yeo et al., 2017).

(ii) Antibiotic-based selection may alter host cell gene expression. Expression of resistance gene and presence of antibiotic in culture medium may have some effects on the cellular transcriptome (Aviner, 2020; Budge et al., 2021; Ryu et al., 2017). It has namely been shown for hygromycin on human colorectal adenocarcinoma cells (Rodolosse et al., 1998), and more recently for puromycin on primary human erythroblasts (R. Guo et al., 2017).

(iii) The DNA elements encoding antibiotic resistance may act as transcriptional silencer on promoters carried on the same expression vector (Yeo et al., 2017). This phenomenon has namely been shown for gene encoding neomycin resistance in mammalian cell lines. However, the same study showed that the gene encoding puromycin resistance do not influence adjacent promoters (Artelt et al., 1991).

(iv) Lastly, antibiotic-based processes are not desirable in GMP production. Presence of antibiotic in culture medium can slow cellular growth and may hide low levels of adventitious agents during USP processes (Budge et al., 2021; Priola et al., 2016). In addition, the antibiotic is a contaminant that should be eliminated from the drug substance during DSP processes (Budge et al., 2021).

Selection of stable clones in the CHOZN® CHO K1 cell line development platform is based on puromycin resistance. Puromycin is a natural amino nucleoside antibiotic produced by *Streptomyces alboniger* that inhibits protein synthesis in eukaryotic and prokaryotic organisms (Vara et al., 1985). Resistance to puromycin is encoded by puromycin N-acetyltransferase (PAC), which is a member of the GCN5-related N-acetyltransferase (GNAT) family of acetyltransferases (Caputo et al., 2021). PAC

enzyme is also expressed by *S. alboniger*. First cloned in 1985, *pac* gene was then used as selection marker in transfected mammalian cell lines in 1986 (Aviner, 2020; Vara et al., 1985, 1986). Until recently, knowledge on PAC structure and PAC structure-activity relationship was surprisingly very limited. Caputo *et al.* solved the crystal structure of PAC. PAC differs from other acetyltransferases of the GNAT family by the absence of the catalytic cysteine-histidine-aspartate triad. Moreover, Caputo *et al.* identified PAC mutants with attenuated activity (Caputo et al., 2021). Details on genetic elements contained in expression cassette for puromycin resistance are given in 2.5.2.2 Genetic elements for selection marker expression.

2.3.2.3.1.2 Metabolic marker-based systems: DHFR, GS, P5CS, HPRT, …

In the case of a metabolic marker-based system, recombinant cell lines are obtained by (i) transfecting a vector carrying the metabolic marker gene and the gene of interest, and then (ii) selecting cells that have stably integrated the plasmid in the absence of the metabolite in culture medium (Lai et al., 2013).

Most used metabolic marker-based systems are dihydrofolate reductase (DHFR) and glutamine synthetase (GS) systems (Fan, Frye, et al., 2013; Gayatri Dhara et al., 2018; Lalonde & Durocher, 2017; Matasci et al., 2008). DHFR selection system works with $DHFR^{-/-}$ cell lines (namely, CHO DXB11, DUK-XB11 and CHO DG44 cell lines) (Jossé et al., 2018; Kaufman & Sharp, 1982; P. Q. Liu et al., 2010). $DHFR^{-/-}$ cell lines have been rendered auxotrophic for glycine, hypoxanthine, and thymidine (GHT) by the deletion of DHFR activity (Santiago et al., 2008; Urlaub et al., 1983; Urlaub & Chasin, 1980; F. M. Wurm & Petropoulos, 1994). GS selection system works with cell lines where GS endogenous activity has been inhibited by methionine sulfoximine (MSX) (M. E. Brown et al., 1992; Cockett et al., 1990; Feary et al., 2017), or with $GS^{-/-}$ cell lines (for example, CHO K1SV GS-KO (Lonza), CHOZN $GS^{-/-}$ (Sigma-Aldrich) and CHOSOURCE CHO-GS KO (Horizon) cell lines) (Fan et al., 2011). $GS^{-/-}$ cell lines have been rendered auxotrophic for glutamine via the deletion of GS activity (P.-Q. Liu & Miller, 2010). In both DHFR and GS systems, recombinant protein expression can be increased by exposing recombinant cell lines either to methotrexate (MTX), which inhibits DHFR enzyme's activity, or to methionine sulfoximine (MSX), which inhibits GS enzyme's activity (F. M. Wurm, 2004). However, both systems have the disadvantage to require difficult and time-consuming technique to create knock-out cell lines, as endogenous enzyme elimination relies on the use of mutagenic agents or genome-editing techniques (Teixeira et al., 2022). In addition, use of MTX or MSX suffers from the same limitations as antibiotics-based systems regarding recombinant protein expression level stability (Noh et al., 2018; Teixeira et al., 2022). Use of MTX or MSX has also been reported to increase mutation rate, resulting in possible mutations in the amino acid sequence of the recombinant protein (D. Guo et al., 2010; Roca et al., 2019). Nevertheless, DHFR and GS systems are positioned as industrial references for large-scale commercial production of recombinant therapeutic proteins (Lalonde & Durocher, 2017).

Other metabolic marker-based systems, relying on natural auxotrophy of CHO cells for specific compounds (Budge et al., 2021; Roca et al., 2019; Sun et al., 2020), on knock-out of key enzymes (Rita Costa et al., 2010), or on the use of new carbon substrates (Teixeira et al., 2022), have also been recently described.

CHO cells are auxotrophic for several amino acids, namely proline, arginine, and cysteine, due to low or absent gene expression of key enzymes in the synthesis pathways of these amino acids (Dean & Reddy, 2013; Duarte et al., 2014; Harcum & Lee, 2016; Hefzi et al., 2016). However, auxotrophy for asparagine has also been shown for CHOK1SV cells (Duarte et al., 2014) but not observed with CHO-K1 cells (Hefzi et al., 2016)– by the way strengthening the concept of "quasi-species" suggested by Florian Wurm (F. M. Wurm, 2013). Natural auxotrophy for amino acids paves the way to potential new selection systems, as highlighted by a pyrroline-5-carboxylate synthetase (P5CS) system (proline auxotrophy-based) (Budge et al., 2021; Sun et al., 2020). In addition, CHO K1 cells are dependent of polyamines (molecules derived from ornithine (such as putrescine, spermidine, and spermine) and from arginine), due to a lack or arginase activity (Holtta & Pohjanpelto, 1983). An arginase-based selection system, relying on a selection in media depleted of L-ornithine and putrescine, has recently been suggested (Roca et al., 2019).

Furthermore, new selection systems, based, like DHFR and GS systems, on the deletion of a key enzyme of cellular metabolism, were studied. This is, for example, the case of the OSCAR® selection system, relying on hypoxanthine phosphoribosyl-transferase (HPRT) based selection (Costa et al., 2012). HPRT catalyzes conversion of hypoxanthine to inosine monophosphate (IMP) and guanine to guanosine monophosphate (GMP), therefore being essential for purine synthesis (Lalonde & Durocher, 2017). HPRT has been widely used as selection marker in the hybridoma technology for monoclonal antibody generation (Sealover et al., n.d.). Nonetheless, and unlike DHFR and GS systems, use of HPRT as selection marker for recombinant stable cell lines selection has been hindered by high mutation rate of the *Hprt* locus, leading to phenotype reversion in Hprt(-/o) cell lines obtained by spontaneous or chemical mutagenesis (Sealover et al., n.d.).

Lastly, new selection systems can also rely on the use of alternative sources of glucose, such as cellobiose, which is hardly metabolized by CHO cells. In this case, selection is performed by introducing the *Neurospora crassa*-derived cellobiose utilization pathway, with cello dextrin transporter (*CDT1)* or β-glucosidase (*GH1-1*) genes as selection markers (Teixeira et al., 2022).

2.3.2.3.1.3 Combinations of selection systems

Several selection systems, either through antibiotic-antibiotic, metabolic marker-metabolic marker, or antibiotic-metabolic marker approaches, are sometimes combined. In the CHOvolution® cell line development platform (Celonic), selection of recombinant stable CHO-K1 clones is based on a dual antibiotic resistance to hygromycin and geneticin.

Combination of several selection systems also finds appropriate application for the development of bispecific antibody producing cell lines or multi-subunit protein producing cell lines, or for the expression of difficult-to-express (DTE) proteins. Fragments of bispecific antibodies are generally produced either in separate cultures or co-cultures, and secreted fragments are assembled into bispecific antibody by reduced glutathione induction (GSH), inducing disulfide bond formation (Shatz et al., 2016; Wang et al., 2019). Few reports investigate cell line development strategies to generate bispecific antibody producing cell lines. Blanco *et al.* used puromycin and hygromycin for generation of bispecific antibody producing cell lines (Blanco et al., 2020). Ding *et al.* investigated a novel approach, relying on a two-stage selection, first combining DHFR-based and puromycin-based selection, and second combining DHFR-based and hygromycin-based selection. This approach significantly increased bispecific antibody expression purity (Ding et al., 2021). In addition, Sealover *et al.* presented a work in which HPRT locus was deleted using zinc-finger nucleases in CHOZN® GS-/- cells, therefore creating a $GS^{-/-}$ and $DHFR^{-/-}$ cell line, which could be useful for producing multi-subunit recombinant proteins. However, this system still requires optimization work on selection and recovery steps, as poor cellular growth and insufficient recombinant protein productivities were obtained (Sealover et al., n.d.). Lastly, Budge et al. combined GS-based and P5CS-based systems and achieved higher concentrations of a difficult-to-expression antibody in CHOK1SV GS-KO™ cells (Lonza) (Budge et al., 2021).

2.3.2.3.2 Modulating selection stringency to increase productivity

Degree of selection stringency is an important parameter for high producing clone identification (Bandaranayake & Almo, 2014; Lai et al., 2013; F. M. Wurm, 2004). High degree of selectivity forces the cells to be more productive in the locus of the selection marker in order to survive the selection process (Lai et al., 2013; Ng et al., 2007). Therefore, a weakened selection marker can increase recombinant protein production through different mechanisms. Increasing selectivity can result in better plasmid integration site, or in higher transcriptional or translational activity, or in higher gene copy number (Lai et al., 2013). These compensatory mechanisms lead to an increased expression of the linked gene of interest (GOI), as the selection marker and the GOI are generally colocalized (Westwood et al., 2010). Increasing clone selection stringency therefore leads to more efficient cell line screening processes, by identifying higher producing cell lines with reduced lab work (F. M. Wurm, 2004). Selection stringency can be strengthened by increasing the selection agent concentration in culture media or by lowering strength of selection marker. The former strategy is rarely applied because of slower cell growth at higher selection pressure concentrations, while the latter has frequently been described in literature, through various strategies at different stages of selection marker expression (Lai et al., 2013).

2.3.2.3.2.1 Attenuating selection marker strength at transcriptional level: the example of SV40 promoter

Selection marker strength can firstly be attenuated at transcriptional level, namely by using a weak promoter to drive selection marker transcription.

Promoter driving expression of the resistance gene can be selected among weak promoters. Several studies report the use of *Herpes simplex* virus thymidine kinase (HSV-TK) promoter (Ng et al., 2012; Niwa et al., 1991), and phosphoglycerate kinase (PGK) promoter is used in expression vector of Freedom™ CHO-S™ (ThermoFisher Scientific) for expression of DHFR and of resistance to puromycin. Recently, Sacco et al. reported the use of a promoter weaker than SV40, without however specifying its name (Sacco et al., 2022).

Nonetheless, simian virus 40 (SV40) promoter and derivatives are mainly used to drive selection marker transcription (Fan et al., 2011; Fan, Kadura, et al., 2013; M. Kim et al., 2011; Sautter & Enenkel, 2005; Yang et al., 2019). SV40 promoter is also used in expression vectors of CHOK1SV cell line development platform (Lonza) for expression of GS and in expression vectors of CHOZN® cell line development platforms (CHOZN® CHO K1 and CHOZN® $GS^{-/-}$ (SAFC)) for expression of resistance to puromycin, and GS, respectively. SV40 promoter is a viral promoter known to have mid transcriptional activity, compared to cytomegalovirus (CMV) promoter (having high transcriptional activity) or to phosphoglycerate kinase (PGK) promoter (having low transcriptional activity) (Budge et al., 2020, 2021).

Structure of SV40E promoter has been extensively studied in the 1980s and is illustrated on Figure 8. Major components of SV40 wild type promoter consist of two 72-base pair repeats, three 21-base pair repeats, a 17-base pair A-T rich sequence containing the TATA box, and few palindromes (Benoist & Chambon, 1980a, 1981; Byrne et al., 1983). The two perfectly repeated 72-bp sequences are described as enhancer sequences (Byrne et al., 1983). These 72-bp repeats are believed to influence the accessibility of RNA polymerase II to the interaction site or the efficiency of RNA chain initiation (Fromm & Berg, 1983). The three GC-rich 21-bp repeats are described as major promoter components. Six copies of a C-rich octa-nucleotide having the consensus sequence 5'-YYCCGCCC-3' (Y being a pyrimidine, *i.e.*, a cytosine or a thymidine) and six copies of C-rich oligonucleotide having the consensus sequence 5'-GCCC(C)T/AA/TA/(T)CT-3' (analogous to the CAAT box consensus sequence) are overlapping the three 21-bp repeats. These two types of C-rich oligonucleotides are alternating and partially overlapping over the three 21-bp repeats (Byrne et al., 1983). These GC-rich 21-bp repeats are believed to act as RNA polymerase II recognition sequences (Fromm & Berg, 1983). Finally, the 17-bp A-T rich sequence containing the Goldberg-Hogness box (TATA box) is described as an important element for accurate initiation of transcription (Byrne et al., 1983). It has been shown that the TATA box can be deleted without drastically affecting gene expression. Deletion of the TATA box results in

the generation of multiple initiation sites, suggesting that elements could functionally substitute the TATA box. It appears that the major role of the TATA box is to fix the transcription start site within a narrow area (Benoist & Chambon, 1981). The requirements for other sequences of SV40 promoter are unclear (Hansen & Sharp, 1983).

Knowledge on sequence requirements for efficient transcription with SV40E promoter have been used several times to weaken SV40E transcriptional activity. Several examples can be quoted. A patent from Roche researchers described the deletion of the two 72-bp repeats and the partial deletion of the 21-bp repeat III, which led to a 80% decrease of promoter transcriptional activity compared to wild type SV40 promoter (Klein & Kopetzki, 2009). A few years later, a patent from Abbvie researchers described the deletion of the 72-bp repeat II and the partial deletion (at least four fifth) of the 72-bp repeat I, which led to higher maximum and mean recombinant antibody titer compared to wild type SV40 promoter (Hartman & Tso, 2013). In the same year, Eli Lilly researchers evaluated several deletions in SV40 wild type promoter, for the expression of glutamine synthetase in CHO $GS^{-/-}$ cells: deletion of the 72-bp repeat II, deletion of 72-bp repeats II and I, deletion of 72-bp repeats II and I and of 21-bp repeat III, and deletion of 72-bp repeats II and II and of 21-bp repeats III and II. For the third deletion (deletion of 72-bp repeats II and I and of 21-bp repeat III) and fourth deletion (deletion of 72-bp repeats II and II and of 21-bp repeats III and II), they observed a significantly reduced transcriptional activity. In addition, maximum and median productivity of clones generated with the third deletion was increased compared to clones generated with SV40E wild type promoter. Investigation was further conducted to understand the underlying causes of improved productivity. However, increased productivity could not be linked to higher gene copy number, letting the authors suggest that increased productivity might be due to better plasmid integration site (Fan, Kadura, et al., 2013). Recently, the fourth deletion suggested by Fan *et al.* (deletion of 72-bp repeats II and II and of 21-bp repeats III and II) was again studied by Yang *et al.*, again for the expression of glutamine synthetase in CHO $GS^{-/-}$ cells (Yang et al., 2019). Maximum and median productivity of mini-pools generated with this 189-bp deletion was increased compared to mini-pools generated with SV40E wild type promoter. However, unlike Fan *et al.*, Yang *et al.* observed a higher gene copy number of glutamine synthetase gene, as well as a higher gene copy number of recombinant antibody heavy and light chain genes, which was followed by a higher heavy chain and light chain transcription level, while glutamine synthetase gene transcription level remained stable compared to transfected cells generated with wild type SV40E promoter.

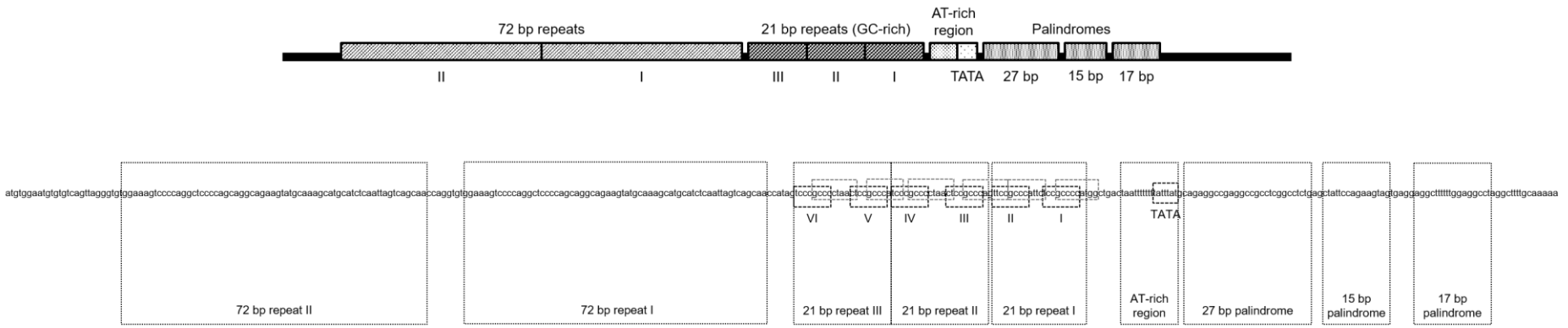


Figure 8: Structure of SV40E promoter

Major components of SV40 wild type promoter consist of two 72-base pair repeats, three 21-base pair repeats, a 17-base pair A-T rich sequence containing the TATA box, and few palindromes. The two perfectly repeated 72-bp sequences are described as enhancer sequences. The three GC-rich 21 bp repeats are described as major promoter components. The two types of C-rich oligonucleotides, which are alternating and partially overlapping over the three 21-bp repeats 17 bp, are indicated in dashed boxes. A-T rich sequence containing the TATA box is described as an important element for accurate initiation of transcription. Adapted from Byrne et al. (1983).

2.3.2.3.2.2 Attenuating selection marker strength at post-transcriptional level

Selection marker strength can secondly be attenuated at post-transcriptional level, namely by using mRNA destabilizing elements.

Once resistance gene is transcribed into RNA, stability of mRNA can be decreased using mRNA destabilizing elements, such as AU-rich elements (ARE). The nonamer 5'-UUAUUUAUU-3' was described in the 1990s as the key AU-rich sequence motif found in 3' untranslated region (3' UTR) that mediates mRNA degradation (Zubiaga et al., 1995). Ng *et al.* reported a 1.7-fold productivity increase of CHO-DG44 cells selected with DHFR-ARE selection system compared to CHO-DG44 cells selected with a non-destabilized DHFR (Ng et al., 2007). Other destabilizing signals found in mRNA, such as histone mRNA 3' terminal stem-loop, iron-responsive element (IRE), and insulin-like growth factor II long stem-loop, are less used than ARE, which can operate constitutively to destabilize mRNA (Guhaniyogi & Brewer, 2001; Ross, 1995; Zubiaga et al., 1995).

2.3.2.3.2.3 Attenuating selection marker strength at translational level

Selection marker strength can thirdly be attenuated at translational level, namely by using codon de-optimization (including start codon) and weakened IRES.

Protein synthesis can be hindered using CHO least preferred codons, therefore lowering translational efficiency. Westwood *et al.* and Chin *et al.* used this strategy to lower DHFR expression and observed that increased selection stringency was correlated with improved recombinant protein productivity (Chin et al., 2015; Westwood et al., 2010). In addition, one patent describes the use of non-ATG start codon, such as GTG and TTG codons, which resulted in high recombinant protein expression levels (Otte et al., 2009). Translation is optimal when the consensus Kozak sequence is present (Kozak, 1987, 2001), but translation might though be initiated with non-optimal sequences.

Lastly, some publications also describe the use of weakened Internal Ribosome Entry Sites (IRES), in case of multi-cistronic expression vectors. Ng *et al.* evaluated an attenuated version of the Encephalomyocarditis virus (EMCV) IRES (Ng et al., 2012), which has been reported to produce 8 to10-fold less protein than the wild type EMCV IRES (Bochkov & Palmenberg, 2006). In spite of the many reported IRES elements (Baird et al., 2006), Chin *et al.* also focused their work on the EMCV IRES and evaluated two mutations, respectively a 7-bp deletion at position 709 and a point mutation (A to C) at position 772 (Chin et al., 2015). Both reported high titer production of recombinant proteins. This strategy offers the advantage to create a close association between selection marker gene and the gene of interest, within a bi- or tri-cistronic vector, therefore preventing possible deletion of the gene of interest during gene amplification (Ng et al., 2012).

2.3.2.3.2.4 Attenuating selection marker strength at post-translational level

Selection marker strength can fourthly be attenuated at post-translational level, namely by using protein destabilizing elements.

Once mRNA is translated into protein, stability of protein can be decreased through the use of protein destabilizing elements, such as murine ornithine decarboxylase (MODC) PEST sequence. PEST sequences are enriched in proline (P), glutamate (E), serine (S) and threonine (T) and constitute a signal of rapid degradation found in proteins with short half-life (Loetscher et al., 1991). Ng *et al.* reported a 6.6-fold productivity increase of CHO-DG44 cells selected with DHFR-MODC PEST selection system compared to CHO-DG44 cells selected with a non-destabilized DHFR (Ng et al., 2007). Other degradation signals found in proteins and known to promote instability, such as N-degrons and CL peptides, are less popular than MODC PEST sequence, whose activity is independent of the ubiquitination process (Ng et al., 2007).

2.3.2.3.2.5 Attenuating selection marker by modifications of enzyme activity

Lastly, selection marker strength can be attenuated at protein level, namely by using attenuated enzyme mutants.

Attenuated mutants have been frequently described in the literature. In such strategies, wild-type selection marker is replaced by an enzyme with attenuated activity. Impaired mutants have been first engineered for neomycin phosphotransferase II (NPTII). Chen *et al.* used the mutation that involves a glutamic to aspartic acid conversion at residue 182 and which was described to reduce enzyme activity (L. Chen et al., 2004; Yenofsky et al., 1990). Sautter and Enenkel evaluated mutations in three motifs of the highly conserved C-terminal region of neomycin phosphotransferase II. They showed increased titers and specific productivities and ascribed them to a gene copy number effect and/or to integration of the expression vector into regions with higher transcriptional activity (Sautter & Enenkel, 2005). Impaired mutant strategy was again described in very recent reports. In 2019, Lin *et al.* published a work in which they evaluated attenuated glutamine synthetase mutants (Lin et al., 2019), containing R324C and R341C mutations identified in unrelated infants with congenital glutamine synthetase deficiency (Häberle et al., 2005) and novel mutations on critical residues identified from the crystal structure of mammalian glutamine synthetase (Krajewski et al., 2008). Increase of selection stringency, resulting in higher productivities, was observed. Interestingly, 100% of clones generated with GS impaired mutants were stable over long-term cultivation compared to 67% of clones generated with wild type GS enzyme and selected with MSX. In 2021, Caputo *et al.* solved the crystal structures of puromycin N-acetyltransferase. Furthermore, they designed four types of mutations to impair the acetylation activity of the enzyme: mutations in acetyl-coenzyme A and puromycin binding sites, mutations in catalytic site, and mutations aiming at altering electrostatic potential. Two of them (in catalytic site and altering

electrostatic potential) resulted in an increase of productivity of two to three folds compared to wild type enzyme in stably transfected pools of HEK293T cells (Caputo et al., 2021).

2.3.2.3.2.6 Combinations of attenuation strategies

These attenuation strategies have been here presented separately, but several reports show that they have been used in combination with each other. Several examples can be quoted: combination of a weak promoter (HSV-tk) with mRNA destabilizing element (ARE) and protein destabilizing element (MODC PEST sequence), which resulted in improvements of titer and specific productivity of 12.6 and 13.3 folds compared to non-attenuated vector (Ng et al., 2007); combination of an attenuated IRES element with protein destabilizing element (MODC PEST sequence) (Ng et al., 2012); combination of an attenuated IRES element with protein destabilizing elements and codon de-optimized selection marker (Chin et al., 2015); combination of an IRES element with attenuated enzyme mutant (Lin et al., 2019).

2.3.2.4 Targeting integration site

It has been shown that plasmid DNA insertion site into the host cell genome influence stable gene expression in CHO cells (Gaidukov et al., 2018; Hamaker & Lee, 2018; Srirangan et al., 2020). To circumvent this limitation, new approaches describe the integration of plasmid DNA into pre-characterized host cell line chromosomal *loci* (Dhiman et al., 2020; Kawabe et al., 2017, 2018; Srirangan et al., 2020).

2.4 Cell line development process

2.4.1 Preamble on cell line development process

Since the 1980s, the fundamental concepts of cell line development have remained unchanged (F. M. Wurm, 2004). Thus, cell line development is based on a well-established scheme, which starts with the transfection of the host cell line with an expression vector. Transfection is the first key step in any cell line development process, as it allows the intracellular delivery of exogenous genetic material necessary for the production of the protein of interest. A broad range of techniques are used for transfection, including (but not limited to) electroporation, nucleofection, calcium-phosphate transfection, lipofection, biolistic- and polymer-mediated gene transfer. As described before, the expression vector contains the transcriptional and translational elements necessary for expression and biosynthesis of the recombinant protein, as well as elements that confer a selective advantage to the cells that have integrated the plasmid.

After transfection, cells that have integrated the expression vector into their genome are selected, either in the absence of an essential metabolite or in the presence of an antibiotic. Selection ensures the generation of stable cell lines, as opposed to transient cell lines within which recombinant DNA has not been integrated into the genome of the host cell line and hence will be lost upon cell division (Barnes et al., 2003). Selection systems have been described before (Preamble on selection systems).

The surviving cells are then individually isolated (Lai et al., 2013; F. M. Wurm, 2004). Single-cell cloning is the second key step in any cell line development process, as it ensures the generation of clonally derived cell lines from a single cell. A broad range of techniques are used for single-cell cloning, including limiting dilution, microfluidic encapsulation, fluorescence activated cell-sorting (FACS), colony picking in semi-solid media, single-cell seeders, and microfluidics chips. The clones thus generated are then evaluated on growth and production criteria of the protein of interest (quantity and quality). At the end of these evaluations, sometimes iterative, a cell line presenting the desired characteristics of growth, productivity, and eventually product quality is selected. This cell line will be the recombinant protein-producing cell line. Generation of recombinant cell lines can take up to six months, from transfection to final clone selection (Tejwani et al., 2021).

Recombinant protein expression level stability study (also referred to as clone stability study) is an additional key study linked to cell line development. Clone stability study aims at determining recombinant protein expression level stability, and thus defining possible constraints for sub culturing and production phases.

These three steps, *i.e.*, transfection, single-cell cloning, and expression level stability study, are detailed hereinafter.

2.4.2 Transfection

Transfection refers to techniques that enable introduction of foreign nucleic acid into cells to create genetically modified cells (T. K. Kim & Eberwine, 2010). Various type of nucleic acids can be transfected into mammalian cells, including DNA, RNA, and small, non-coding RNA such as small interfering RNA (siRNA), small/short hairpin RNA (shRNA), or micro RNA (miRNA) (Fus-Kujawa et al., 2021).

Introduction of foreign nucleic acid into eukaryotic cell cytoplasm or nucleus can either be performed by transduction, which relies on the use of viral-methods, or by transfection, which relies on the use of physical methods or chemical methods (T. K. Kim & Eberwine, 2010). However, for manufacturing purposes, non-viral gene transfer remains the preferred approach to generate stable recombinant cell lines (F. M. Wurm, 2004).

2.4.2.1 Transfection by physical methods

2.4.2.1.1 Electroporation

Electroporation has become one of the most popular method for non-viral transfection, due to its ease of use, flexibility, almost no limitation on cell type and size, and yield in high transfection efficiencies in combination with low cell mortality (Shi et al., 2018). Electroporation is commonly defined as "the transient loss of semi-permeability of cell membranes subjected to electrical pulses, resulting in ion leakage, escape of metabolites, and increased uptake by cells of drugs, molecular probes, and DNA" (Tsong, 1989, 1990). The exposure of cells to an electric field leads to an increase of native biological membrane permeability, by forming transitory aqueous pores. Foreign DNA is allowed to traffic inside the cell through these pores (Escoffre et al., 2009, 2012).

Optimization of electroporation parameters is cell type specific (Slivac et al., 2017). In particular, four important characteristics of the electrical pulse may affect electroporation efficiency: pulse voltage, pulse width, pulse number, and pulse shape (exponential decay *versus* square wave). Other parameters have been reported to impact electroporation efficiency, namely: the temperature of cells before, during, and after electroporation; the concentration and the topology of DNA (linear DNA *versus* circular DNA (Escoffre et al., 2012)); the concentration and state of cells (Kumar et al., 2019).

For examples, Neon® (Thermo Fisher Scientific) and Gene Pulser Xcell™ (BioRad) devices are commercial electroporation systems; Neon® is namely used for electroporation on CHOZN® CHO K1 and CHOZN® $GS^{-/-}$ cell line development platforms.

2.4.2.1.2 Nucleofection

Electroporation was further improved, which led to a transfection technique called nucleofection. Initially developed for transfection of non-dividing and difficult-to-transfect cells, such as primary cells

(Distler et al., 2005; Quenneville et al., 2004), nucleofection is also used for transfection of mammalian cells in cell line development processes. Nucleofection relies on a precise combination of electrical parameters associated with cell-type specific reagents (Slivac et al., 2017). It enables plasmid DNA to enter directly into the nucleus, making nucleofection independent of cell division, unlike other transfection techniques (Distler et al., 2005).

For example, Nucleofector™ (Lonza) device is a commercial nucleofection system.

2.4.2.1.3 Other physical methods for transfection

Other physical methods for transfection namely include micro-injection (use of a specific needle to puncture the cell and to inject nucleic acids into the nucleus), sonoporation (transient permeabilization of cell membranes using ultrasounds), laser mediated transfection, also known as optical transfection, optoporation or phototransfection (transient permeabilization of cell membranes using laser light) (Chong et al., 2021; Fus-Kujawa et al., 2021; T. K. Kim & Eberwine, 2010; Slivac et al., 2017).

2.4.2.2 Transfection by chemical methods

Transfection by chemical methods namely include cationic polymers, cationic lipids, cationic amino acids, and calcium-phosphate transfection. The most popular cationic polymers include diethylaminoethyl-dextran (DEAD-dextran), polyethylenimine (PEI), dendritic polyamidoamine (PAMAM), poly(L-lysine), poly(N, N-dimethylaminoethyl methacrylate) (PDMAEMA) (Fus-Kujawa et al., 2021).

These methods have a common operating principle. Positively charged chemicals make nucleic acid-chemical complexes with negatively charged nucleic acids. The nucleic acid-chemical complexes are positively charged and therefore attracted by negatively charge cell membrane (T. K. Kim & Eberwine, 2010; Pham et al., 2006). The exact mechanism of entrance of nucleic acid-chemical complexes through the cell membrane are poorly known, even if phagocytosis and endocytosis are believed to be involved in this phenomenon (Fus-Kujawa et al., 2021). Similarly, the exact mechanism by which nucleic acids enter the nucleus is not well described (T. K. Kim & Eberwine, 2010).

2.4.3 Single-cell cloning

2.4.3.1 Guidelines on single-cell cloning

Regardless of the method used, single-cell cloning is an unconditional step in any cell line generation process. Single-cell cloning is indeed a requirement from guidelines from health authorities. The FDA mentions, in a supplement to the points to consider in the production and testing of new drugs and biologics produced by recombinant DNA technology, that “in most cases, a single host cell containing the expression construct is cloned to give rise to the MCB” (US Food and Drug Administration, 1992). The International Conference on Harmonization of Technical Requirements for Registration of

Pharmaceuticals for Human Use (ICH) Q5D on derivation and characterization of cell substrates used for production of biotechnological/biological products mentions that "for recombinant products, the cell substrate is the transfected cell containing the desired sequences, which has been cloned from a single cell progenitor" (European Medicines Agency, 1998). The ICH statement was expanded by the World Health Organization (WHO), which mentions that "the cloning procedure should be fully documented, with details of imaging techniques and/or appropriate statistics. For proteins derived from transfection with recombinant plasmid DNA technology, a single fully documented round of cloning is sufficient, provided that product homogeneity and consistent characteristics are demonstrated throughout the production process and within a defined cell age beyond the production process" (Evans et al., 2015).

Single-cell cloning serves two main purposes: (i) helping to identify stable, highly producing cell lines, and (ii) creating a single-cell precursor to minimize heterogeneity (Lai et al., 2013; Welch & Arden, 2019).

2.4.3.2 Methods and technologies for single-cell cloning

2.4.3.2.1 Limiting dilution cloning

The oldest method for single-cell cloning is limiting dilution cloning (LDC). Limiting dilution is still frequently used as single-cell cloning method in some cell line development processes, namely on CHOZN® CHO K1 and CHOZN® $GS^{-/-}$ cell line development platforms.

The method of single-cell cloning by limiting dilution consists in diluting the cell suspension to very low cell concentrations, generally lower than one cell per well. Clones are sub-cultured in individual wells for several weeks (generally two to four weeks) until reemergence is observed. In spite of being very simple to implement and to perform and being very gentle on cells (Tejwani et al., 2021), single-cell cloning by limiting dilution suffers from being time-consuming and low in throughput (Welch & Arden, 2019).

Cell lines resulting from one round of limiting dilution cloning are considered to follow the Poisson distribution law with regards to clonality (Zhou et al., 2018). Performing one round of limiting dilution for single-cell cloning might not be sufficient to provide high enough assurance of clonality. As a consequence, it has been recommended to perform two or more rounds of limiting dilution to increase the probability of obtaining clonally derived cell lines (Tejwani et al., 2021).

2.4.3.2.2 Cell sorter and/or seeders

Due to the limitations of limiting dilution as single-cell cloning method, advanced technologies were developed to perform single-cell cloning. These technologies can be classified into cell sorting technologies and cell seeding technologies. Automation of single-cell cloning helps to accelerate cell line development timelines, screen higher number of clones and reduce lab-work (Tejwani et al., 2021).

For examples, FACSAria™ (BD Biosciences), Cyto-Mine® (Sphere Fluidics), Beacon® (Berkeley Lights), and CellCelector™ (ALS) devices are commercial cell sorting and seeding systems. UP.SIGHT™ (Cytena), and VIPS™ (Solentim) devices are commercial cell seeding systems. Many of these equipment are currently implemented in industrial cell line development platforms: the Beacon® in Amgen platform (Diep et al., 2021), the MoFlo-XDP (Beckman Coulter) in Genentech platform (Misaghi et al., 2016), the FACSAria™ in Boehringer Ingelheim GmbH and Merck & Co, Inc. platforms, the Cyto-Mine® in Fujifilm Diosynth Biotechnology, Biogen, and Astra Zeneca platforms (Tejwani et al., 2021).

2.4.3.3 Concept of monoclonality: monoclonal or clonally-derived cell lines?

Derivation and heterogeneity of "monoclonal" cell lines have been highlighted in many reports. For examples, Pilbrough *et al.* observed that the expression level of a clone can vary significantly (standard deviation 50 to 70% of the mean) and that this variation is fully developed within only 18 days after single-cell cloning (Pilbrough et al., 2009). Chen *et al.* performed a transcriptomic analysis on subclones from a stable parental CHO clone and observed over 100 differentially expressed genes involved in various biological processes between subclones (K. Chen et al., 2017). Tharmalingam *et al.* analyzed 46 subclones from a production cell line and reported significant diversity in cell growth, protein productivity, product quality attributes and gene copy number between subclones, in spite of proven stability of the parental cell line (Tharmalingam et al., 2018). Similar observations were reported by Ko *et al.* (Ko et al., 2018).

Derivation and heterogeneity of "monoclonal" cell lines find causes in high plasticity of CHO cell lines (F. Wurm & Wurm, 2017). In addition, the high number of population doublings from single-cell cloning to production of master cell bank (MCB) may lead to genotypic and phenotypic differences between subclones (Tharmalingam et al., 2018).

Obtaining absolute clonality of CHO cell lines is impossible. Identical genetics among subclones derived from a single cell is not supported by any data (F. Wurm & Wurm, 2017). As a consequence, the terms "monoclonal" or "clonal" are inappropriate, and "monoclonal/clonal" cell lines should be best described as clonally derived cell lines (Fan et al., 2017).

2.4.4 Expression level stability study

2.4.4.1 Stakes and guidelines on expression level stability study

Generating recombinant cell lines that maintain a stable production level over time is of particular importance, from both industrial and regulatory standpoints (Barnes et al., 2003).

It is widely considered that a clone is stable if at least 70% of volumetric productivity is retained over a period that spans master cell bank (MCB) preparation, working cell bank (WCB) preparation, and GMP production, which often corresponds to at least 60 population doubling levels (PDL) (Bailey et al., 2012;

Dahodwala & Lee, 2019; Dorai et al., 2012; Lin et al., 2019). The International Conference on Harmonization of Technical Requirements for Registration of Pharmaceuticals for Human Use (ICH) Q5D on derivation and characterization of cell substrates used for production of biotechnological/biological products mentions that "at least two time points should be examined, one using cells which have received a minimal number of subcultivations, and another using cells at or beyond the limit of in vitro cell age for production use described in the marketing application. The limit of in vitro cell age for production use should be based on data derived from production cells expanded under pilot plant scale or commercial scale conditions to the proposed limit of in vitro cell age for production use or beyond" (European Medicines Agency, 1998).

2.4.4.2 Insights on expression level stability of CHO-derived recombinant cell lines

CHO cell line instability has widely been reported in previous reports and is estimated to occur in 8-63% of all recombinant cell lines (Dahodwala & Lee, 2019). Kim *et al.* observed that approximately 30% of clones derived from CHOK1SV cells isolated with GS selection system and MSX lost 30% or more of their initial productivity over long-term cultivation (M. Kim et al., 2011). Similarly, Dorai *et al.* observed that a significant percentage of clones derived from CHOK1SV cells isolated with GS selection system and MSX lost 20-30% or more of their initial productivity over 55 PDL. In addition, in a majority of cell lines evaluated in this study, the decline in volumetric productivity was less important when MSX was maintained during sub-culturing process (Dorai et al., 2012). A similar observation was made by Noh *et al.* and Weidle *et al.*, who detected a more significant decrease of specific productivity in the absence of respectively MSX or MTX than in the presence of respectively MSX or MTX (Noh et al., 2018; Weidle et al., 1988). Furthermore, Lin *et al.* observed that approximately 40% of clones derived from CHO GS KO cells isolated with GS selection system and MSX lost 30% or more of their initial productivity over 12 weeks, but this proportion was importantly decreased (15% clones were unstable) using attenuated GS mutant instead of the wild type GS enzyme (Lin et al., 2019). The withdrawal of selection pressure is a source of instability for many high producing clones (Lin et al., 2019).

As a consequence, it appears that productivity instability is independent of CHO host cell line, selection and/or amplification system, and presence/absence of selection pressure (Dahodwala & Lee, 2019). Nonetheless, in general, it seems that clones derived from CHO GS KO host cell line and selected with GS selection system and MSX are more stable than clones derived from CHO DHFR KO host cell line and selected with DHFR selection system and MTX, themselves more stable than clones derived from CHO K1 host cell line and selection either with GS selection system and MSX, or with antibiotics. Cell line development platforms with selection systems where the selection pressure is continuously maintained, even in cGMP production (*i.e.*, DHFR and GS selection systems, based on auxotrophy of

CHO DHFR$^{-/-}$ and CHO GS$^{-/-}$ for glycine, hypoxanthine, and thymidine and for glutamine respectively) are more efficient than selection systems where the selection pressure is removed.

2.4.4.3 Root causes of productivity instability

Even if several teams focused their efforts on understand the underlying causes of instability, many mechanisms still need to be understood (Ritter et al., 2016). To date, productivity instability has been explained by gene copy loss (Bailey et al., 2012; Fann et al., 2000; Jun et al., 2006; M. Kim et al., 2011), unstable chromosomal structures (Chusainow et al., 2009; Jun et al., 2006), gene silencing as a result of epigenetic regulation (Chusainow et al., 2009; M. Kim et al., 2011; Osterlehner et al., 2011; Paredes et al., 2013), or increased susceptibility to cellular stresses (Bailey et al., 2012; Chusainow et al., 2009).

Identifying the underlying causes of productivity instability may help to find approaches to enhance stability of production, which is, as well as productivity, of high importance from an industrial and regulatory point of view (Barnes et al., 2003). Indeed, several teams reported that cell line instability is clone-specific, certainly acquired during cell line development process as clones derived from the same transfection can either be stable or unstable, and possibly controllable if underlying causes are better understood (Qian et al., 2020).

2.5 Overview of CHOZN® CHO K1 cell line development platform

The work presented in this thesis aims at optimizing the CHOZN® CHO K1 cell line development platform for production of therapeutic recombinant proteins. The CHOZN® CHO K1 cell line development platform is a mammalian cell expression system for selection and scale up of clones producing a recombinant protein. This platform includes a host cell line, an expression vector, an optimized set of cGMP produced, chemically defined, growth and production media and feed, and a cell line development protocol. This chapter intends to give a detailed overview of CHOZN® CHO K1 cell line development platform initial state.

2.5.1 Host cell line

CHOZN® CHO K1 is the host cell line of the CHOZN® CHO K1 platform. As mentioned earlier, Sigma-Aldrich company acquired in 2005 a CHO-K1 vial deposited at ECACC [ECACC 85051005]. The adherent CHO-K1 cells were then adapted to suspension culture in a chemically defined medium. In 2006, the CHOZN® CHO K1 cell bank was manufactured under cGMP conditions (current Good Manufacturing Practices). After complete characterization, in accordance with the requirements of the health authorities for Master Cell Banks (MCB), the CHOZN® CHO K1 line [CHOK1-1VL, SAFC] was commercially available in 2016.

2.5.2 Expression vector

pEM2V is the plasmid of the CHOZN® CHO K1 platform. pEM2V version a (pEM2V-a) was first used and quickly replaced by pEM2V version c (pEM2V-c). Both plasmids contain three expression cassettes. Plasmid maps are shown on Figure 9 (pEM2V-a) and Figure 10 (pEM2V-c). Cassette 1 enables expression of resistance to selection pressure, i.e., puromycin N acetyltransferase, which confers resistance to puromycin to cells that have stably integrated the plasmid. Cassette 2 enables expression of recombinant antibody heavy chain. Cassette 3 enables expression of recombinant antibody light chain. Plasmids pEM2V-a and pEM2V-c also contains genetic elements for plasmid replication (ori) and production (selection marker) in bacterial host. pEM2V-c differs from pEM2V-a in the intronic and poly(A) tail sequences on the light chain cassette.



Figure 9: Plasmid map of pEM2V-a

Plasmid pEM2V-a (End-to-End Solutions Martillac, two viral promoters, plasmid version a) contains three expression cassettes. Cassette 1 (colored in green) enables expression of resistance to selection pressure, i.e., puromycin N-acetyltransferase, which confers resistance to puromycin to cells that have stably integrated the plasmid. Cassette 2 (colored in blue) enables expression of recombinant antibody heavy chain (coding sequence to be inserted in MSC1). Cassette 3 (colored in purple) enables expression of recombinant antibody light chain (coding sequence to be inserted in MSC2). Intronic and poly(A) tail sequences are identical in cassettes 2 and 3. Plasmid pEM2V-a also contains genetic elements for plasmid replication (ori, colored in yellow) and production (selection marker, colored in orange) in bacterial host. Plasmid map was generated with SnapGene® (viewer version).



Figure 10: Plasmid map of pEM2V-c

Plasmid pEM2V-c (End-to-End Solutions Martillac, two viral promoters, plasmid version c) contains three expression cassettes. Cassette 1 (colored in green) enables expression of resistance to selection pressure, i.e., puromycin N-acetyltransferase, which confers resistance to puromycin to cells that have stably integrated the plasmid. Cassette 2 (colored in blue) enables expression of recombinant antibody heavy chain (coding sequence to be inserted in MSC1). Cassette 3 (colored in purple) enables expression of recombinant antibody light chain (coding sequence to be inserted in MSC2). Enhancers, promoters, introns, and poly(A) tail sequences differ in cassettes 2 and 3. Plasmid pEM2V-c also contains genetic elements for plasmid replication (ori, colored in yellow) and production (selection marker, colored in orange) in bacterial host. Plasmid map was generated with SnapGene® (viewer version).

2.5.2.1 Genetic elements for recombinant protein expression

pEM2V-a and pEM2V-c expression vectors contain two expression cassettes for recombinant protein expression. Recombinant protein is most of the time a monoclonal antibody, whose heavy chain and

light chain coding sequences are respectively inserted in the first and second expression cassettes following selection marker.

Monoclonal antibody heavy chain expression cassette is identical in pEM2V-a and pEM2V-c. This cassette contains human cytomegalovirus (CMV) Immediate Early 1 (IE1) enhancer and promoter, Herpes Simplex Virus 5 (HSV5) intron, and Simian Vacuolating virus 40 (SV40) poly(A) tail. However, monoclonal antibody light chain expression cassette is different between pEM2V-a and pEM2V-c versions. pEM2V-a light chain cassette contains Herpes Simplex Virus 5 (HSV5) intron and SV40 poly(A) tail, while pEM2V-c light chain cassette contains Elongation Factor 1 α (EF1α) intron and Bovine Growth Hormone (BGH) poly(A) tail. Intron and poly(A) sequences of light chain cassette where modified to avoid identical genetic elements between light chain and heavy chain cassettes, which could lead to homologous recombination during cellular divisions. This is intended to enhance expression vector stability over cellular divisions, therefore improving recombinant protein expression level stability.

CHOZN® CHO K1 platform expression vector is derived from CHOZN® GS$^{-/-}$ platform expression vector (pCGS3). A screening of enhancers and promoters was performed to identify enhancer-promoter pair for heavy and light chain expression (other transcriptional and translational control elements of pCGS3 backbone being conserved). The following viral enhancers and promoters have been screened: human cytomegalovirus (CMV) promoter and enhancer and human and murine cytomegalovirus (CMV) Immediate Early 1 (IE1) enhancers and promoters. Other viral enhancers and promoters have been identified, such as human and murine cytomegalovirus (CMV) Immediate Early 2 (IE2) enhancers and promoters, but not evaluated, because of patents on related sequences.

Furthermore, inducible promoters have also been evaluated. Inducible promoters are particularly useful for expression of toxic recombinant proteins (Makrides, 1999, 2003; F. M. Wurm, 2004). Induction can be triggered by adding a nontoxic inducer to the culture medium or by modifying cell culture process parameters (temperature downshift (TDS) being the most frequent) (Bedoya-López et al., 2016; Tossolini et al., 2022). The latter was considered as preferred method for induction, especially because of the simplicity of implementation of temperature downshift during process development. Following CHO K1 endogenous, temperature inducible promoters of following genes have been screened: interleukin 33 (*Il33*) (NCBI Gene ID: 100761971), RAS like family member A (*RasI11a*) (NCBI Gene ID: 100758419), dynein regulatory complex subunit 7 (*Drc7*) (NCBI Gene ID: 100764290), and histone H2A.J (*LOC100757684*) (NCBI Gene ID: 100757684) (Bedoya-López et al., 2016). Other inducible promoters, such as RBM3 and CIRP, have been identified but not evaluated because of patents on related sequences.

2.5.2.2 Genetic elements for selection marker expression

pEM2V-a and pEM2V-c expression vectors contain an expression cassette for selection marker expression. Identification of stable and producing clones in the CHOZN® CHO K1 platform relies on puromycin-based selection. Puromycin is a natural amino nucleoside antibiotic produced by *Streptomyces alboniger* that inhibits protein synthesis in eukaryotic and prokaryotic organisms (Vara et al., 1985). Resistance to puromycin is encoded by puromycin N-acetyltransferase (PAC) (NCBI GenBank ID: AAA64928.1; UniProtKB ID: P13249), which is also expressed by *S. alboniger*. First cloned in 1985, *pac* gene was then used as selection marker in transfected mammalian cell lines in 1986 (Aviner, 2020; Vara et al., 1985, 1986).

PAC expression cassette is identical in pEM2V-a and pEM2V-c. PAC expression cassette contains Simian Vacuolating virus 40 (SV40) promoter, a Kozak sequence whose sequence is reduced to the shortest consensus motif (CCACC) for initiation (Kozak, 1987, 1999; Makrides, 1999, 2003), *pac* coding sequence, SV40 intron and poly(A) signal.

2.5.3 Cell line development process

CHOZN® CHO K1 platform cell line development process state in 2018 is shown in Figure 11. Cell line development process starts with CHOZN® CHO K1 host cell line thawing and sub-culturing. CHOZN® CHO K1 cells are then transfected by electroporation. Generation of monoclonal cell lines is a two iterative step-based process, divided in (i) a mini-pool (MP) generation and selection step and in (ii) a clone (CL) generation and selection step. The mini-pool generation step aims at easing high producing pools identification, while the clone generation step aims at delivering a clonally derived, high producing cell line. In 2018, CHOZN® CHO K1 platform cell line development process enabled DNA to pre-master cell bank (pre-MCB) in 7.5 months.

Recombinant protein expression level stability study (also referred to as clone stability study) and media and feeds screening are two additional cell line development linked studies Figure 12. Clone stability study aims at determining recombinant protein expression level stability, and thus defining possible constraints for sub culturing and production phases. Media and feeds screening aims at preselecting media and feeds platform(s) (i.e., expansion medium, production medium, and feeds) which will be used for following steps (namely, upstream process (USP) development). In 2018, 2.5 months and 2 months were respectively necessary for clone stability study and media and feeds screening.

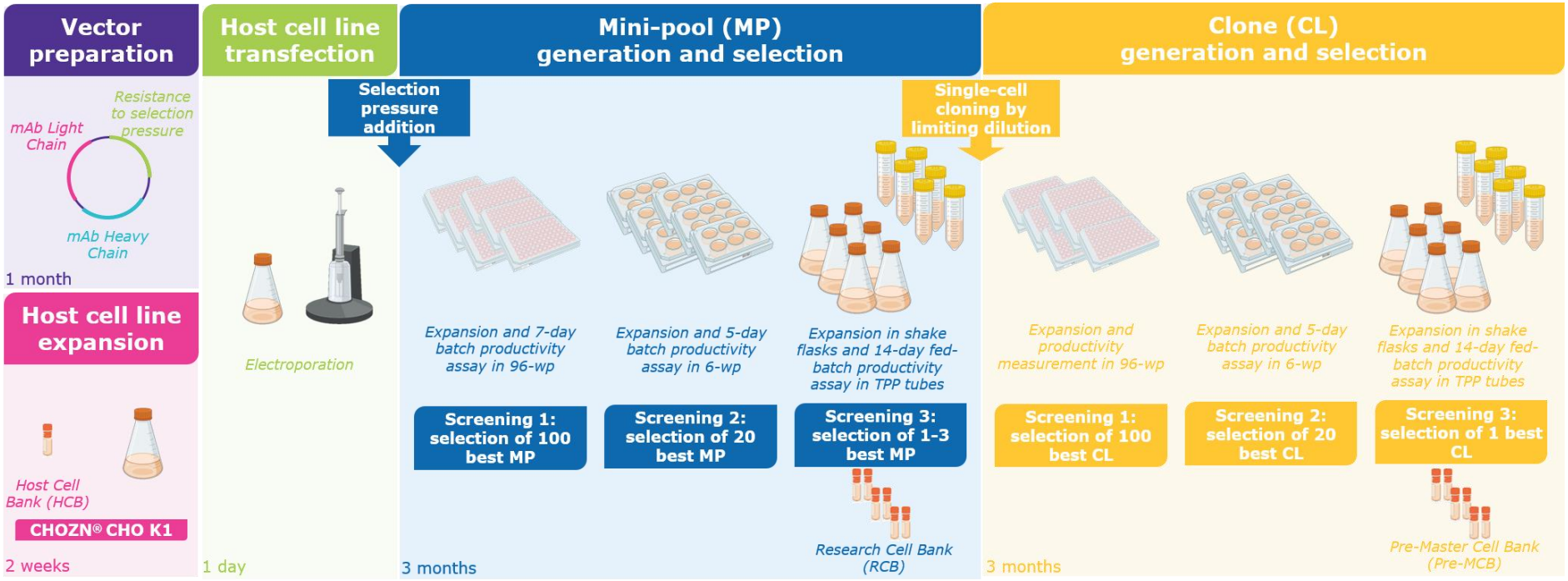


Figure 11: CHOZN® CHO K1 platform cell line development process (state in 2018)

CHOZN® CHO K1 platform cell line development process starts with host cell line thawing and subculturing. Host cell line is then transfected by electroporation. Generation of monoclonal cell lines is a two step-based process, divided in (i) a mini-pool (MP) generation and selection step (highlighted in blue) and in (ii) a clone (CL) generation and selection step (highlighted in yellow). The mini-pool generation step aims at easing high producing pools identification, while the clone generation step aims at delivering a monoclonal, high producing cell line. Both mini-pool and clone generation steps rely on iterative process for identification and selection of high producers: subculturing and productivity assay in 96-well plates, scale-up and productivity assay in 6-well plates, and scale-up and productivity assay in TPP® tubes. Biological material is stored at the end of each step, with generation of research cell banks (RCB) and pre-master cell banks (pre-MCB), respectively. In 2018, CHOZN® CHO K1 platform cell line development process enabled DNA to pre-MCB in 7.5 months.

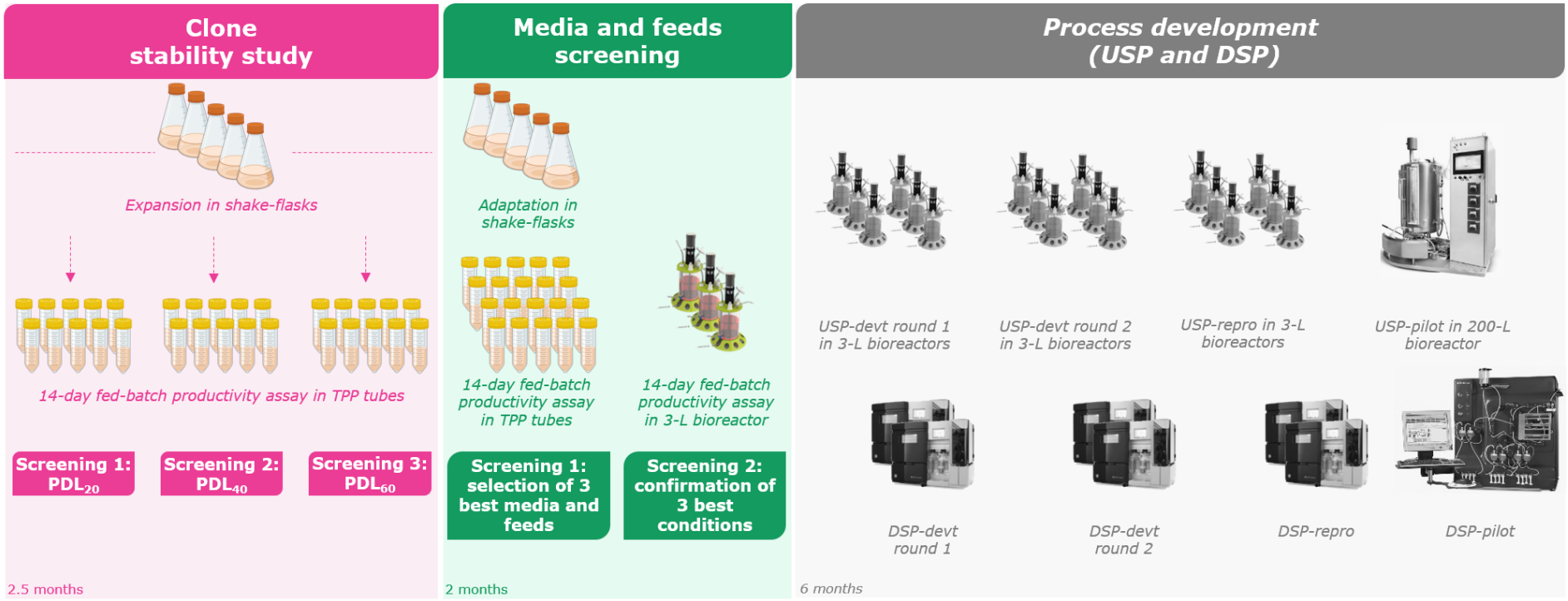


Figure 12: Additional CLD-linked studies: clone stability study and media and feeds screening (state in 2018)

Recombinant protein expression level stability study (also referred to as clone stability study) and media and feeds screening are two additional cell line development linked studies. Clone stability study aims at determining recombinant protein expression level stability, and thus defining possible constraints for sub culturing and production phases. Media and feeds screening aims at preselecting media and feeds platform(s) (i.e., expansion medium, production medium, and feeds) which will be used for following steps (namely, upstream process (USP) development). Process development (both upstream process (USP) and downstream process (DSP) development) can start once these additional studies are finishing. However, in some cases (namely fast-track projects), process development can be initiated long before these steps, sometimes even in parallel of clone generation and selection process. In 2018, 2.5 months and 2 months were respectively necessary for clone stability study and media and feeds screening.

2.5.3.1 Expression vector production

Monoclonal antibody heavy chain and light chain coding sequences are respectively inserted in multiple cloning site (MCS) 1 and 2 of pEM2v-a or pEM2v-c vector backbones. An optimization of used codons is performed for production in CHO cells. Both codon optimization and expression vector production are outsourced.

2.5.3.2 Host cell line expansion

CHOZN® CHO K1 cells are thawed and sub-cultured in Erlenmeyer flasks in EX-CELL® CD CHO Fusion supplemented with 6 mM L-glutamine **(**Figure 11**)**. EX-CELL® CD CHO Fusion medium is a chemically defined (CD), animal component free (ACF) medium developed for long-term growth of CHOZN® cells. This medium does not contain L-glutamine. It is manufactured under current good manufacturing practices (cGMP) production quality conditions. CHOZN® CHO K1 cell viability is generally maintained over 90% and doubling time between 16 to 20 hours over sub-culturing process.

2.5.3.3 Transfection

CHOZN® CHO K1 cells are then transfected *via* electroporation **(**Figure 11**)**. CHOZN® platform technical bulletin describes electroporation conditions to be used on a Gene Pulser® Xcell device [Bio-Rad]. Gene Pulser® Xcell electroporation device delivers both exponential decay and square-wave type electrical pulses, for which pulse voltage, capacitance, number, and width can be set. Electroporation conditions have been optimized by End-to-End Solutions CLD Martillac's team to use a Neon device [Thermo Fisher Scientific]. Neon device uses a pipette tip as an electroporation chamber and delivers exponential decay type electrical pulses, for which only pulse voltage, number, and width can be set. Optimized electroporation conditions result from a large parameter screening of following conditions: cell/DNA quantities ratio, pulse voltage, pulse number, and pulse width. Several additional parameters, such as DNA topology, electroporation buffer or medium, or cell preparation method, have also been previously evaluated. The best conditions highlighted by this screening resulted in the definition of electroporation protocol currently used for transfection of CHOZN® CHO K1 cells.

These electroporation protocol generally leads to a 60-80% transfection efficiency. Transfection efficiency can be determined by the expression of a fluorescent reporter, such as the green fluorescent protein (GFP) or the red fluorescent protein (RFP). Transfection can also be performed by lipid-based transfection methods, but lower transfection efficiencies are expected.

2.5.3.4 Mini-pool generation

The mini-pool selection method is based on the division of the transfected bulk pool into tinier sub-populations, called mini-pools. To provide orders of magnitude, transfected bulk pool contains approximately $5x10^6$ viable cells, while each individual mini-pool contains approximately 5,000 viable cells. The objective of the mini-pool selection method is to maintain initial cellular diversity, by creating

more variation than there would be between pools. Screening several different mini-pools therefore increases the chance of rapidly isolating stable and high-producing mini-pools.

Mini-pool generation is based on an iterative selection process **(**Figure 11**)**. 24 hours after electroporation, transfected bulk pool is seeded in 96-well plates, at a rate of 5,000 viable cells per well, in a blend of 80% EX-CELL® CD CHO Fusion medium and 20% EX-CELL® CHO Cloning medium, supplemented with 6 mM L-glutamine and 10 µg/mL puromycin. EX-CELL® CHO Cloning medium is an animal component free medium developed to support clonal survival and growth of CHOZN® cells, with results comparable to the traditional method using 10% fetal bovine serum (FBS). EX-CELL® CHO Cloning medium contains plant hydrolysates and therefore cannot be labelled as chemically defined and does not contain L-glutamine. It is manufactured under current good manufacturing practices (cGMP) production quality conditions. Each seeded well represents one mini-pool; 600 mini-pools are usually seeded. Mini-pools are cultured in 96-well plates until reemergence is observed, *i.e.*, mini-pools are 50% confluent.

A first screening is performed in 96-well plates, through a non-normalized (variable initial cell quantity, due to the impossibility to seed at a fixed viable cell density) 7-day batch productivity assay. At the end of the productivity assay, recombinant protein concentration in supernatant is measured by biolayer interferometry with protein A-coated sensors. Mini-pools are ranked according to recombinant protein concentration. 100 best producers are selected for scale-up and further evaluation.

Selected mini-pools are then sub-cultured in 6-well plates. A second screening is performed in 6-well plates, through a normalized 6-day batch productivity assay. At the end of the productivity assay, viable cell density, cellular viability and recombinant protein concentration in supernatant are measured. Mini-pools are ranked according to recombinant protein concentration and specific productivity. 20 best producers are selected for scale-up and further evaluation.

Selected mini-pools are then sub-cultured in Erlenmeyer flasks. A third and final screening is performed in TPP® tubes, through a normalized 14-day fed-batch productivity assay, in EX-CELL® Advanced™ CHO Fed-batch medium with EX-CELL® Advanced™ Feed 1 addition. EX-CELL® Advanced™ CHO Fed-batch medium is a chemically defined medium specially developed for fed-batch cultures. EX-CELL® Advanced™ Feed 1 is a highly concentrated feed developed to be conjunctively used with EX-CELL® Advanced™ CHO Fed-batch medium in fed-batch cultures. At the end of the productivity assay, viable cell density, cellular viability and recombinant protein concentration in supernatant are measured. Mini-pools are again ranked according to recombinant protein concentration and specific productivity. 1 to 3 best producers are selected for single-cell cloning. Fed-batch productivity assay yields valuable information on growth and productivity characteristics of mini-pools, which may predict the growth and productivity characteristics of clones derived from those mini-pools.

Selected mini-pools for sub-cultivation in Erlenmeyer flasks and productivity assay in TPP® tubes are cryopreserved, in 93% EX-CELL® CD CHO Fusion medium supplemented with 6 mM L-glutamine and 7% DMSO. This intermediate cellular bank is referred to as research cell bank (RCB).

CHOZN® CHO K1 platform performances over a mini-pool generation have been estimated prior the beginning of the thesis project. CHOZN® CHO K1 has been transfected with pEM2V-a vector for expression of a standard IgG4. Productivity levels reached 20 mg/L after batch productivity assay in 96-well plates, 70 mg/L after batch productivity assay in 6-well plates, and 1.3 g/L after fed-batch productivity assay in TPP® tubes.

CHOZN® platform technical bulletin also describes the bulk pool selection method, instead of the mini-pool selection method previously described. 24 hours after transfection, transfected bulk pool is seeded in T-flask with selection pressure addition. The bulk pool selection method generally leads to pool recovery from transfection and selection in approximately 10 to 21 days post transfection. Pool recovery is characterized when cellular viability is maintained above 90% and doubling time is stabilized. Though, variability generated by the bulk pool selection method is less important than that generated by the mini-pool selection method (since the transfected bulk pool is divided into tinnier sub-populations in the mini-pool generation method, making it more likely to identify high-producing mini-pools). Therefore, final titers of the top bulk pools are likely to be lower than those of top mini-pools.

2.5.3.5 Clone generation

The objective of the clone generation is to generate monoclonal, high producing cell lines.

Clone generation is based on an iterative selection process which is very similar to the one described for mini-pool generation **(**Figure 11**)**. Single-cell cloning (SCC) is performed by limiting dilution. Cells are seeded in 96-well plates at a final density of 1.5 vc/mL, *i.e.*, 0.3 viable cell per well (the CHOZN® platform technical bulletin describes a final density of 2.5 vc/mL, *i.e.*, 0.5 viable cell per well) in a blend of 80% EXCELL® CHO Cloning medium and 20% of conditioned medium, supplemented with 6 mM L-glutamine (no supplementation with puromycin, as selection pressure is withdrawn from culture medium at the end of the mini-pool step). Conditioned medium is prepared with mini-pools that will be subjected to single-cell cloning. Blending of cloning medium with conditioned medium has proved to be efficient in single-cell cloning processes, by providing the clones with beneficial nutrients and factors secreted by exponentially growing healthy cells. Furthermore, it should be noticed that puromycin has been definitively removed from the culture media since the end of the mini-pool generation process. This single-cell cloning protocol generally leads to a 20% cloning efficiency, considering that cell lines derived from this protocol follow a Poisson distribution with regards to clonality.

Identification of monoclonal cell lines is based on plate imaging. Plates are imaged one hour after cell seeding, using Cell Metric® [Solentim] plate imager. This D_0 image constitute the proof of

monoclonality. Subsequent images are then taken on D_1, D_2, and D_3 after seeding, and then twice a week to monitor clone outgrowth.

Clones are then cultured in 96-well plates until reemergence is observed, *i.e.*, clones are 30% confluent. A first screening is performed in 96-well plates, through a productivity measurement when clones are at least 30% confluent. Recombinant protein concentration in supernatant is measured by biolayer interferometry with protein A-coated sensors. Clones are ranked according to recombinant protein concentration. 100 best producers are selected for scale-up and further evaluation. Next clone generation steps (subculturing and productivity assay in 6-well plates, subculturing in Erlenmeyer flasks and productivity assay in TPP® tubes) are identical to those described for the mini-pool generation.

Selected clones for sub-cultivation in Erlenmeyer flasks and productivity assay in TPP® tubes are cryopreserved, in 93% EX-CELL® CD CHO Fusion medium supplemented with 6 mM L-glutamine and 7% DMSO. This cellular bank is referred to as pre-master cell bank (pre-MCB).

CHOZN® CHO K1 platform performances on a clone generation have not been estimated prior the beginning of the thesis project. Nonetheless, results obtained during the mini-pool generation (best mini-pool producing 1.3 g/L recombinant IgG1) give reasons to hope for clones producing at least 2 g/L (considering a 1.5-titer increase between mini-pool and clone generation).

2.5.3.6 Recombinant protein expression level stability study

Recombinant protein expression level stability study is an additional step which is performed once monoclonal cell lines are identified **(**Figure 12**)**. The objective of the expression level stability study is to determine recombinant protein expression level stability, and thus define possible constraints for sub-culturing and production phases.

1 to 10 clones can be selected for recombinant protein expression level stability study, which is generally run in biological duplicates. Furthermore, stability study can be run with and without selection pressure, with selection pressure only maintained during sub-culturing phases. Three 12-day fed-batch productivity assays are launched in TPP® tubes at pre-defined time points, corresponding to estimated population doubling levels (PDL) (respectively at PDL_{20}, PDL_{40}, and PDL_{60}). In some cases, four productivity assays can be launched, respectively at PDL_{15}, PDL_{30}, PDL_{45} and PDL_{60}. Study duration is defined to span master cell bank (MCB) preparation, working cell bank preparation (WCB), GMP production, and extended cell bank (ExCB) preparation.

A monoclonal cell line is generally considered stable if 70% or more of volumetric productivity titer is retained (or, in other words, if the variation in the production of the recombinant protein is less than 30%) over 60 to 70 generations (Dahodwala & Lee, 2019). The same threshold is applied for assessment of CHOZN® CHO K1 clones' stability. However, this conclusion might be qualified by the analysis of other available information resulting from stability study, such as specific productivity or growth profile.

2.5.3.7 Media and feeds screening

Media and feeds screening is an additional study which is performed once monoclonal cell lines are identified **(**Figure 12**)**. The objective of this screening is to preselect media and feeds platform(s) (*i.e.*, expansion medium, production medium, and feeds) which will be used for upstream process (USP) development.

Clones are first adapted in the different media. Adaptation is either performed directly (one passage) or sequentially (over several passages) during subculturing phase in Erlenmeyer flasks. Clones are likely to have different growth profiles depending on the evaluated media and may sometimes not even survive adaptation (even sequential) in some media. 20-40 media and feeds combinations are then evaluated through a 14-day fed-batch productivity assay in TPP® tubes. A confirmation study of up to the three best conditions is generally run in 3 L-bioreactors, again through a 14-day fed-batch productivity assay.

However, media and feeds screening is not mandatory. This step can be either omitted, by relying on previous results obtained with clones derived from the same host cell line, yet not producing the same recombinant protein, or combined with USP development.

2.5.4 Conclusion

As previously described, CHOZN® CHO K1 cell line development platform is characterized by a two-step cell line development process (mini-pool generation followed by clone generation) with puromycin-based selection of stable clones and random integration of the expression vector in the host cell genome. The CHOZN® CHO K1 host cell line and the pEM2V-a/c expression vector are royalty-free, making the CHOZN® CHO K1 cell line development platform a particularly attractive business proposition for small biotech companies.

CHOZN® CHO K1 platform state at the beginning of the thesis offers many working areas for optimization. Platform optimization relies on three major challenges.

Firstly, cell line productivity needs to be increased. Generation of mini-pools producing 1 g/L suggested that clones would achieve a production of 2 g/L of recombinant antibody. This has validated the beginning of the optimization of the CHOZN® CHO K1 platform but is nevertheless insufficient compared to competing platforms performances. Both expression vector and cell line development process optimization should lead to productivity increase. Secondly, cell line development process timelines need to be decreased. Two-step cell line development process is time-consuming, as seven to eight months are required from expression vector preparation to RCB. Such a process is not viable when the objective is to complete all development (cell line development, USP development, DSP development, and analytical development) and production activities (MCB preparation, GMP production, and quality control) in one year. Overall process development duration could be significantly reduced if cell line development timelines are shortened. Mainly cell line development

process optimization could lead to reduction of timelines. These two first challenges result from the need to meet the needs of generating higher producing cell lines in shortened timelines, and to ensure viable new biopharmaceutical product development, as years of development significantly contribute to the final cost of the therapeutic molecule. Thirdly, cell line development process needs to be miniaturized and ideally automatized. This third challenge may contribute to reduction of cell line development timelines.

2.6 Literature review synthesis

Recombinant proteins for therapeutic application are mainly monoclonal antibodies and antibody-derived molecules. Monoclonal antibodies are high molecular weight molecules, with a complex structure, whose properties are highly dependent on the production process and which are potentially immunogenic. Therefore, the expression system for the production of a recombinant monoclonal antibody must be chosen with care. Mammalian cells are capable of producing complex proteins, having, in particular in the case of antibodies, glycosylation profiles close or identical to those found in humans. More specifically, CHO cells, whose vast genealogy offers a large choice of cell lines, present advantageous characteristics for the production of recombinant proteins for therapeutic applications. As such, the majority of industrial cell line development platforms relies on the use CHO cell lines, in particular CHO K1, CHO DHFR$^{-/-}$, CHO GS$^{-/-}$, and CHO S cell lines.

The literature review highlights a multitude of areas of investigation to understand the critical parameters of a cell line development platform for the expression of recombinant proteins. These areas of improvement concern each of the elements of cell line development platform, *i.e.*, the host line, the expression vector, and the stages of the cell line generation process. These research lines allow to optimize a cell line development platform on three criteria: productivity of recombinant cell lines, expression level stability of recombinant cell lines, and cell line generation timelines.

Furthermore, the CHOZN® CHO K1 platform has already been partially optimized, in particular regarding the transfection step (screening of electroporation parameters) and the expression vector (screening of pairs of enhancers and promoters). It is noteworthy that this cell line development platform is characterized by a host cell line and an expression vector which are royalty-free. In the view of the scientific knowledge resulting from this bibliographic study, the continuation of the work on the expression vector (in particular on the expression of the resistance to the selection pressure), and on the process of generation of cell lines are the preferred tracks.

3 Results

The results section consists of four chapters:

(i) Characterization of CHOZN® CHO K1 cell line development platform performance. The objectives of this chapter were to define reference characteristics of the CHOZN® CHO K1 platform and to define CHOZN® CHO K1 platform initial performance in terms of productivity and cell line development timelines.

(ii) Cell line development process optimization: implementation of direct single-cell cloning. The objective of this chapter was to evaluate a shortened cell line development process, called direct single-cell cloning (DSCC), in order to reduce cell line development timelines.

(iii) Expression vector optimization: evaluation of promoters for expression of resistance to puromycin. The objective of this chapter was to evaluate novel promoters for expression of selection pressure, in order to increase recombinant clonal cell lines productivities.

(iv) Reverse transcription – quantitative polymerase chain reaction (RT-qPCR) method development. The objective of this chapter was to implement a RT-qPCR method, which was required in chapter 3.

3.1 Characterization of CHOZN® CHO K1 cell line development platform performance

3.1.1 Introduction

The CHOZN® CHO K1 cell line development platform is a mammalian cell expression system for selection and scale up of clones producing a recombinant protein. This platform includes a host cell line, an expression vector, an optimized set of cGMP produced, chemically defined, growth and production media and feed, and a cell line development protocol. Overview of the CHOZN® CHO K1 cell line development platform was previously detailed in Overview of CHOZN® CHO K1 cell line development platform.

Cell line development process associated to CHOZN® CHO K1 platform starts with host cell line thawing and subculturing. Host cell line is then transfected by electroporation, using pEM2V-a expression vector backbone (Figure 9). pEM2V-a expression vector backbone contains three expression cassettes (resistance to selection pressure, antibody heavy chain, and antibody light chain expression) relying on classical genetic elements. Generation of monoclonal cell lines is a two-step based process, divided in (i) a mini pool (MP) generation and selection step and in (ii) a clone (CL) generation and selection step (Figure 11). The mini pool generation step aims at easing high producing pools identification, while the clone generation step aims at delivering a monoclonal, high producing cell line.

Both mini pool and clone generation steps rely on iterative process for identification and selection of high producers: subculturing and productivity assay in 96-well plates, scale-up and productivity assay in 6-well plates, and scale-up and productivity assay in TPP® tubes, all steps using CHOZN® media and feed platform. Cell line development timelines are estimated to span 7.5 months (from DNA to pre-master cell banks (pre-MCB)).

We focused our work on getting acquainted with CHOZN® CHO K1 cell line development platform. Our first objective was to define reference characteristics of the CHOZN® CHO K1 platform, mostly related to cell growth (for example, reemergence profiles of mini-pools after transfection, reemergence profiles of clones after single-cell cloning). Our second objective was to define CHOZN® CHO K1 platform initial performance in terms of productivity (both mini-pools and clones) and cell line development timelines. This initial platform performance characterization will later be considered as point of comparison for the optimizations made during this project.

3.1.2 Methodology

CHOZN® CHO K1 mini-pools expressing an IgG4 monoclonal antibody were generated and screened through a mini-pool generation process. Best producing mini-pools were selected for single-cell cloning. Then, CHOZN® CHO K1 clones were generated by single-cell cloning using limiting dilution method, and screened through a clone generation. Clone expression level stability was investigated over 60 Population Doubling Levels (PDL).

3.1.3 Results

3.1.3.1 CHOZN® CHO K1 mini-pool generation

CHOZN® CHO K1 cell line development platform performance was then evaluated over a mini-pool generation. Briefly, CHOZN® CHO K1 host cell line was transfected by electroporation with plasmid pEM2V-a carrying an IgG4 monoclonal antibody heavy and light chain gene sequences. One day after transfection, mini-pools were seeded in 96-well plates and put under selection pressure. Mini-pools were cultured in 96-well plates until cellular reemergence was observed; a maximum of three weeks was allowed for reemergence. When the first of these two criteria was met, a seven-day non normalized batch productivity assay was launched in 96-well plates. At the end of the productivity assay, recombinant antibody concentration in cell culture supernatant was measured by bio-layer interferometry using protein A-coated sensors. Due to schedule constraints, the traditional subsequent scale-up (sub-culturing and batch productivity assay in 6-well plates, followed by sub-culturing in shake flasks and fed-batch productivity assay in TPP® tubes) was modified. Subsequent scale-up was thus performed in 24-deepwell plates (sub-culturing and fed-batch productivity assay).

3.1.3.1.1 CHOZN® CHO K1 mini-pools reemerged nine days after transfection

Mini-pool reemergence in 96-well plate, shown as mini-pool confluence as a function of subculturing duration is illustrated on Figure 13. Mini-pool confluence clearly increased from D_7 post mini-pool seeding (mean confluence above 20%). Mini-pool reemergence was confirmed on D_{12}, following a dilution performed on purpose on D_9. This mini-pool reemergence profile was later confirmed on several assays (data not shown). Data points of mini-pools evaluated in fed-batch culture in 24-deepwell plates are colored, according to mini-pools' final IgG4 titer (high producers being colored in red and low producers being colored in blue). There does not seem to be a correlation (either positive or negative) between the time needed for reemergence and mini-pool productivity.

3.1.3.1.2 CHOZN® CHO K1 mini-pools' productivity

Mini-pools' titers after productivity assays in 96-well plates and in 24-deepwell plates, and mean process specific productivity during productivity assay in 24-deepwell plates are shown on Figure 16. After batch productivity assay in 96-well plates, mini-pools' titers ranged from 2.6 to 20.5 mg/L, with median titer at 11.5 mg/L. After fed-batch productivity assay in 24-deepwell plates, mini-pools' titers ranged from 40.2 mg/L to 298.5 mg/L (14.6-fold increase compared to 96-well plate), with median titer at 168.8 mg/L (14.7-fold increase compared to 96-well plate). Data points of mini-pools evaluated in fed-batch culture in 24-deepwell plates are colored, according to mini-pools' final IgG4 titer (high producers being colored in red and low producers being colored in blue). There does not seem to be a correlation (either positive or negative) between titers after batch productivity assay in 96-well plates and titers after fed-batch productivity assay in 24-deepwell plates.

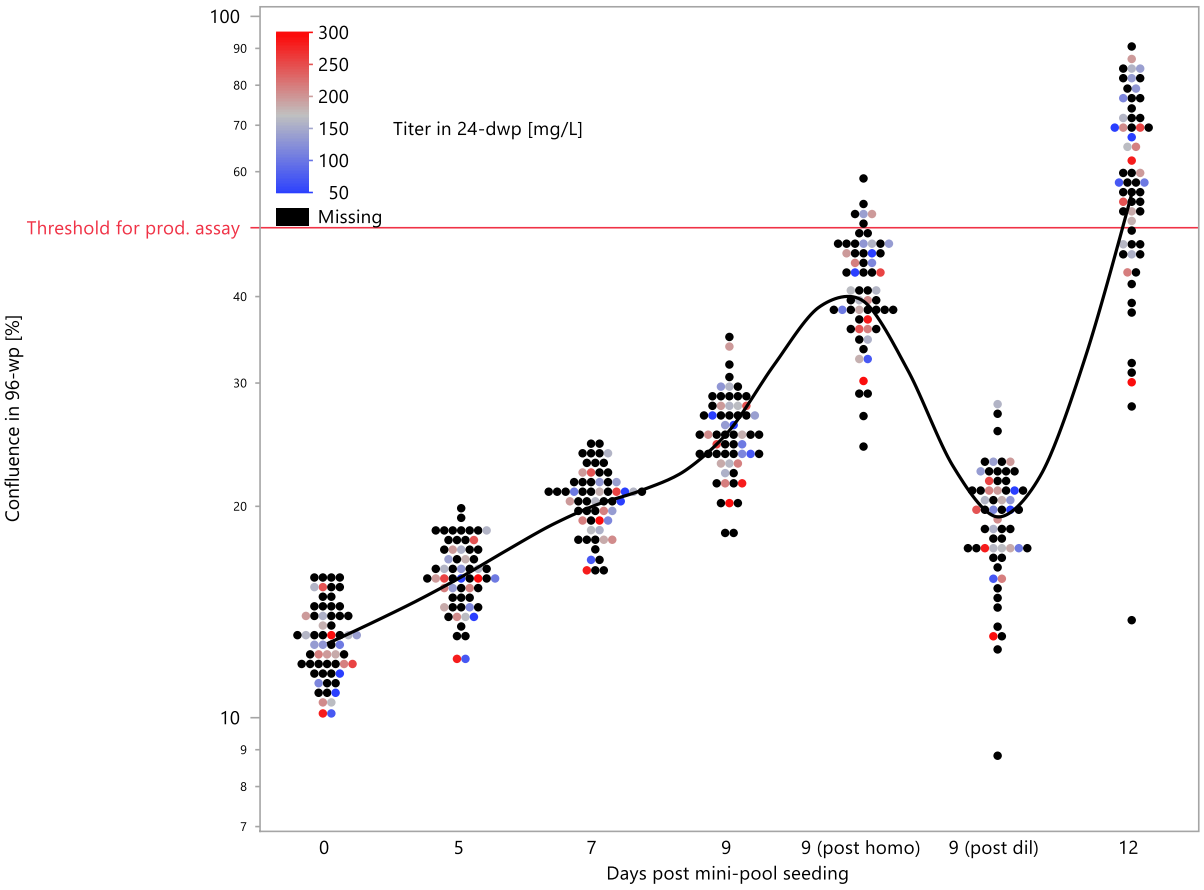
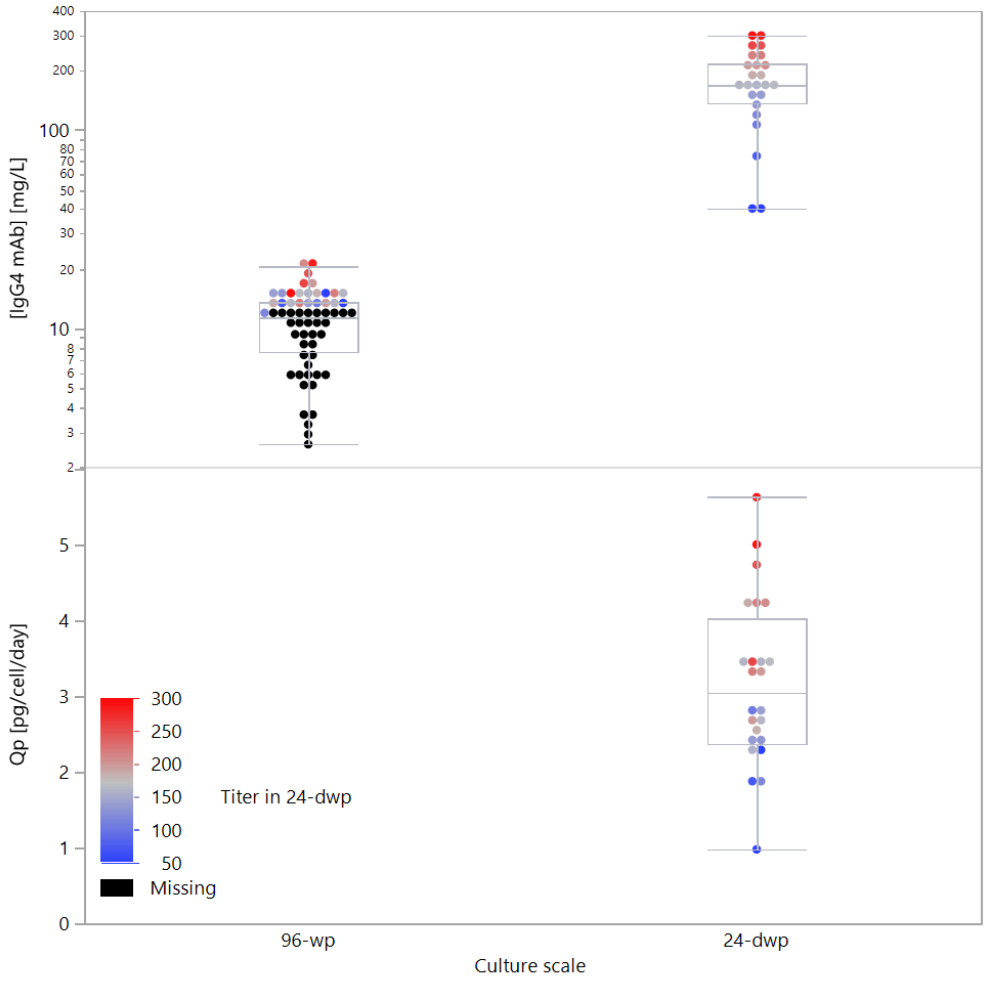


Figure 13: Mini-pool confluence in 96-well plate, as a function of subculturing duration

CHOZN® CHO K1 host cell line was transfected by electroporation with plasmid pEM2V-a carrying an IgG4 monoclonal antibody heavy and light chain gene sequences. One day after transfection, mini-pools were seeded in 96-well plates and put under selection pressure. Mini-pools were cultured in 96-well plates until cellular reemergence was observed; a maximum of three weeks was allowed for reemergence. If necessary, cell suspensions were homogenized. Cell suspensions were diluted to confirm cellular reemergence. When the first of these two criteria was met, a seven-day non normalized batch productivity assay was launched in 96-well plates. Post homo: confluence was measured after homogenization of cell suspensions; Post dil: confluence was measured after dilution of cell suspensions in a new 96-well plate. Data analysis was performed on JMP® 16 software.



Culture scale	96-well plate	24-deepwell plate
Max titer [mg/L]	20.5	298.5
Fold change on max titer compared to previous culture scale	-	14.6
Median titer [mg/L]	11.5	168.8
Fold change on median titer compared to previous culture scale	-	14.7
Max Q_P [pg/cell/day]	-	5.6
Median Q_P [pg/cell/day]	-	3.0

Figure 14: IgG4 mAb concentration in mini-pools' supernatants and mean process specific productivity after productivity assay in 96-well plates and in 24-deepwell plates

At the top: Box-and-whiskers plots of IgG4 concentration in mini-pools' supernatant after batch productivity assay in 96-well plates and fed-batch productivity assay in 24-deepwell plates. At the bottom: Box-and-whiskers plot of mean process specific productivity after batch productivity assay in 24-deepwell plates. Each dot represents one mini-pool. Data analysis was performed on JMP® 16 software.

3.1.3.2 CHOZN® CHO K1 clone generation

CHOZN® CHO K1 cell line development platform performance was then evaluated over a clone generation. Briefly, the three best mini-pools (MP 12B05 (298.5 mg/L, 5.0 pg/cell/day), MP 12D03 (288,3 mg/L, 5.6 pg/cell/day), and MP 12G04 (250.5 mg/L, 4.7 pg/cell/day)) were selected for single-cell cloning by limiting dilution method. Clones were cultured in 96-well plates until cellular reemergence was observed; a maximum of four weeks was allowed for reemergence. When the first of these two criteria was met, recombinant antibody concentration in cell culture supernatant was measured by bio-layer interferometry using protein A-coated sensors. Subsequent scale-up was first performed in 24-deepwell plates (sub-culturing and batch productivity assay) and then performed in shake flasks (sub-culturing) and TPP® tubes (fed-batch productivity assay).

3.1.3.2.1 CHOZN® CHO K1 clones mainly reemerged 12 to 19 days after single-cell cloning

Clone reemergence in 96-well plate, shown as clone confluence as a function of subculturing duration is illustrated on Figure 15. First clones (20 out of 123, *i.e.*, 16%) reemerged on D_{12} (day on which their confluence exceeded the 30% confluence threshold) after single-cell cloning at the earliest, among which was the best producer (CL 15D08, according to final titer after fed-batch productivity assay in TPP® tubes). A large number of clones then reemerged on D_{14} (32 out of 123, *i.e.*, 26%) and D_{19} (57 out of 123, *i.e.*, 46%) after single-cell cloning, among which were both high and low producers. Few clones reemerged during the fourth week of expansion (6 out of 123, *i.e.*, 5% on D_{21} and 8 out of 123, *i.e.*, 7% on D_{27}). Data points of clones evaluated in fed-batch culture in TPP® tubes are colored, according to clones' final IgG4 titer (high producers being colored in red and low producers being colored in blue). There does not seem to be a correlation between reemergence duration and final clone productivity. Cloning efficiency was calculated at 34%, according to Poisson's distribution.

Table 4: Cloning efficiency and percentages of reemerging clones as a function of subculturing duration in 96-well plates

Cell line development process	**Long process (MP-CL)**
Cloning efficiency	34%
Percentage of reemerging clones, as a function of subculturing duration in 96-wp	16% at D_{12} 26% at D_{14} 46% at D_{19} 5% at D_{21} 7% at D_{27}
Cumulated percentage of reemerging clones, as a function of subculturing duration in 96-wp	16% at D_{12} 42% at D_{14} 88% at D_{19} 93% at D_{21} 100% at D_{27}

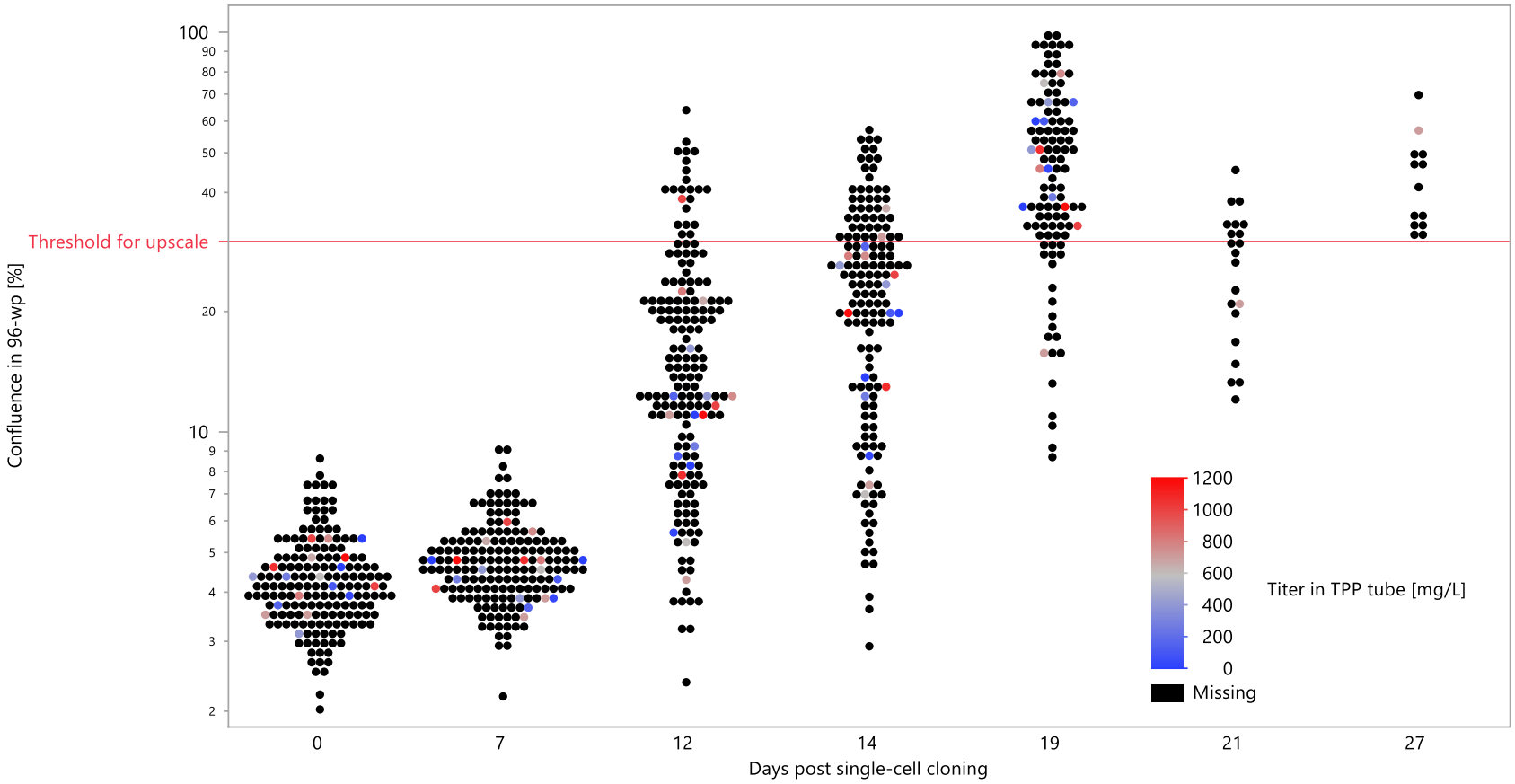


Figure 15: Clone confluence in 96-wp, as a function of subculturing duration

The three best mini-pools (MP 12B05, MP, and MP 12G04) were selected for single-cell cloning by limiting dilution method. Clones were cultured in 96-well plates until cellular reemergence was observed (30% confluence considered as confluence threshold for upscale); a maximum of four weeks was allowed for reemergence. Data analysis was performed on JMP® 16 software.

3.1.3.2.2 CHOZN® CHO K1 clones' behavior in fed-batch culture

Although CHOZN® CHO K1 clones displayed various profiles during fed-batch culture, a general trend was observed (data not shown). Viable cell density peak was mainly reached on production day 9 (PD_9), with maximum VCD ranging between 4 and $25x10^6$ viable cells/mL. Almost all cultures were maintained until PD_{14} (viability above 80%), with viability decline sometimes initiated from PD_{10}. Glucose consumption led to production of lactate, leading to a lactate concentration peak on PD_5 ranging between 1 and 1.5 g/L. Lactate were then mostly re-consumed, with few clones re-producing lactate on the last days of culture.

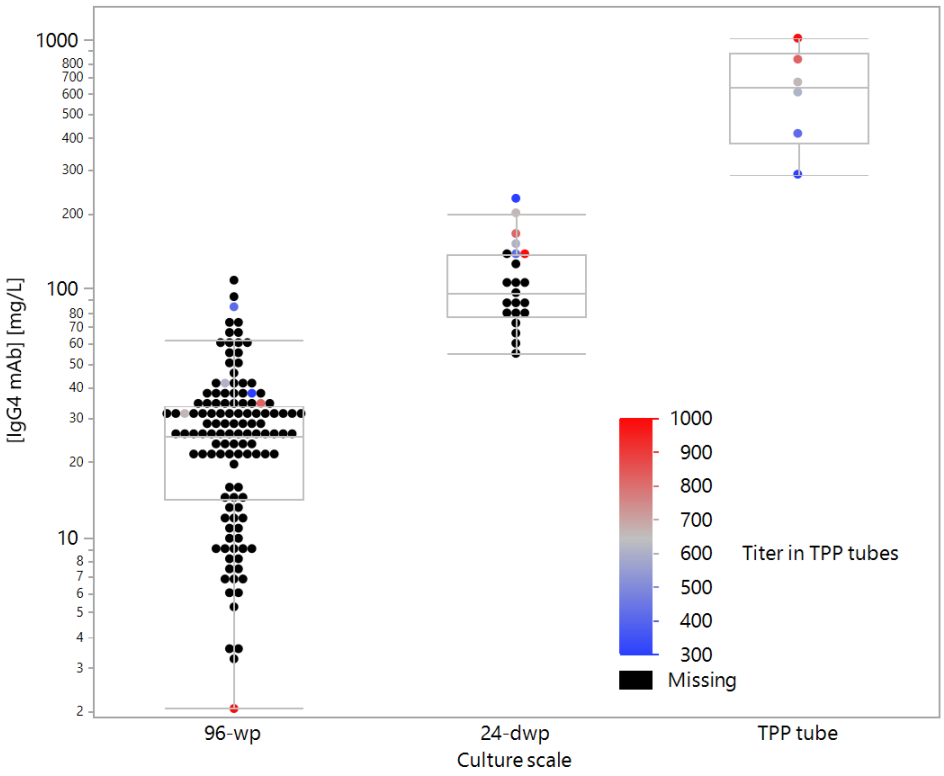
3.1.3.2.3 CHOZN® CHO K1 clones' productivity

Clones' titers after productivity measurement in 96-well plates, and productivity assays in 24-deepwell plates and TPP® tubes are shown on Figure 16. Mean process specific productivity during productivity assays in 24-deepwell plates and TPP® tubes are shown on Figure 17. After productivity measurement in 96-well plates, clones' titer ranged from 2.1 mg/L to 107.8 mg/L, with median titer at 25.5 mg/L. After batch productivity assay in 24-deepwell plates, clones' titer ranged from 54.8 mg/L to 229.7 mg/L (2.1-fold increase compared to 96-well plates), with median titer at 96.5 mg/L (3.8-fold increase compared to 96-well plates). Mean process specific productivity ranged from 2.4 pg/cell/day to 26.4 pg/cell day, with median process specific productivity at 10.5 pg/cell/day. Finally, after fed-batch productivity assay in TPP® tubes, clones' titer ranged from 287 mg/L to 1009 mg/L (4.4-fold increase compared to 24-deepwell plates), with median titer at 641 mg/L (6.6-fold increase compared to 24-deepwell plates). Mean process specific productivity ranged from 5.9 pg/cell/day to 14.4 pg/cell/day, with median process specific productivity at 9.4 pg/cell/day.

Data points of clones evaluated in fed-batch culture in TPP® tubes are colored, according to clones' final IgG4 titer or mean process specific productivity (high producers being colored in red and low producers being colored in blue). There does not seem to be a correlation (either positive or negative) between titers after productivity measurement in 96-well plates and titers after fed-batch productivity assay in TPP® tubes. Indeed, one particularly notices that best clone in TPP® tubes (CL 15D08, final titer 1008 mg/L) was the worst producer in 96-well plates (2 mg/L). On the other hand, there seems to be a better correlation between 24-deepwell plates and TPP® tubes, both on final titer and mean process specific productivity. Taking again the example of CL 15D08, this clone was the fourth best producer in 24-deepwell (ranking on final IgG4 titer) and best producer in 24-deepwell plate (ranking on mean process specific productivity).

The 24-deepwell plate culture format was first used during mini-pool generation because of schedule constraints (rather than sub-culturing and batch productivity assay in 6-well plates, followed by sub-culturing in shake flasks and fed-batch productivity assay in TPP® tubes), and applied used during clone generation. The correlation trends observed between the different scales of culture were measured with

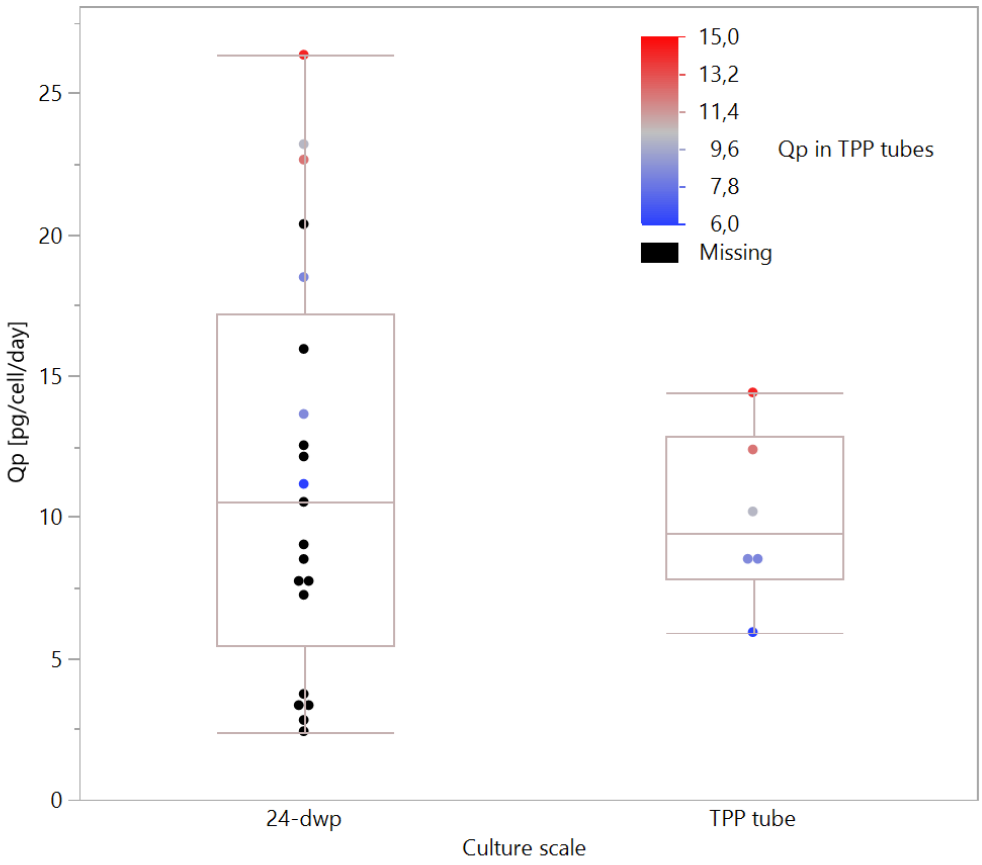
correlation matrices to confirm the potential of 24-deepwell plates for mini-pool or clone screening process optimization. Correlation matrix of titers in 96-well plate, 24-deepwell plate and TPP® tubes is shown on Figure 18-A. Correlation matrix of mean process specific productivity in 24-deepwell plate and TPP® tubes is shown on Figure 18-B**.** Absence of correlation between titer in 96-well plate and titer in shaken culture (either 24-deepwell plates or TPP® tubes) is confirmed (correlation coefficient near 0 for both correlations and correlation probability above $\alpha_{5\%}$). In contrast, a correlation trend between titers in 24-deepwell plate and in TPP® tubes is confirmed (correlation coefficient above 0.5 and correlation probability below $\alpha_{5\%}$). In addition, a strong correlation between mean process specific productivity in 24-deepwell plate and in TPP® tubes is noticed (correlation coefficient above 0.9 and correlation probability below $\alpha_{5\%}$).



Culture scale	96-well plates	24-deepwell plates	TPP® tubes
Max titer [mg/L]	107.8	229.7	1009
Fold change on max titer compared to previous culture scale	-	2.1	4.4
Median titer [mg/L]	25.5	96.5	641
Fold change on mean titer compared to previous culture scale	-	3.8	6.6

Figure 16: IgG4 mAb concentration in clones' supernatants after productivity measurement in 96-well plates and productivity assays in 24-deepwell plates and TPP® tubes

Plot of IgG4 concentration in clone's supernatant after productivity measurement in 96-well plates, batch productivity assay in 24-deepwell plates and fed-batch productivity assay in TPP® tubes. Each dot represents one clone. Fold change on maximum and median titer is calculated in relation to long cell line development process (MP-CL) at all culture scales. Data analysis was performed on JMP® 16 software.



Culture scale	24-deepwell plates	TPP® tubes
Max Q_P [pg/cell/day]	26.4	14.4
Median Q_P [pg/cell/day]	10.5	9.4

Figure 17: Mean process specific productivities after productivity assays in 24-deepwell plates and TPP® tubes

Plot of mean process specific productivities during batch productivity assay in 24-deepwell plates and fed-batch productivity assay in TPP® tubes. Each dot represents one clone. Data analysis was performed on JMP® 16 software.

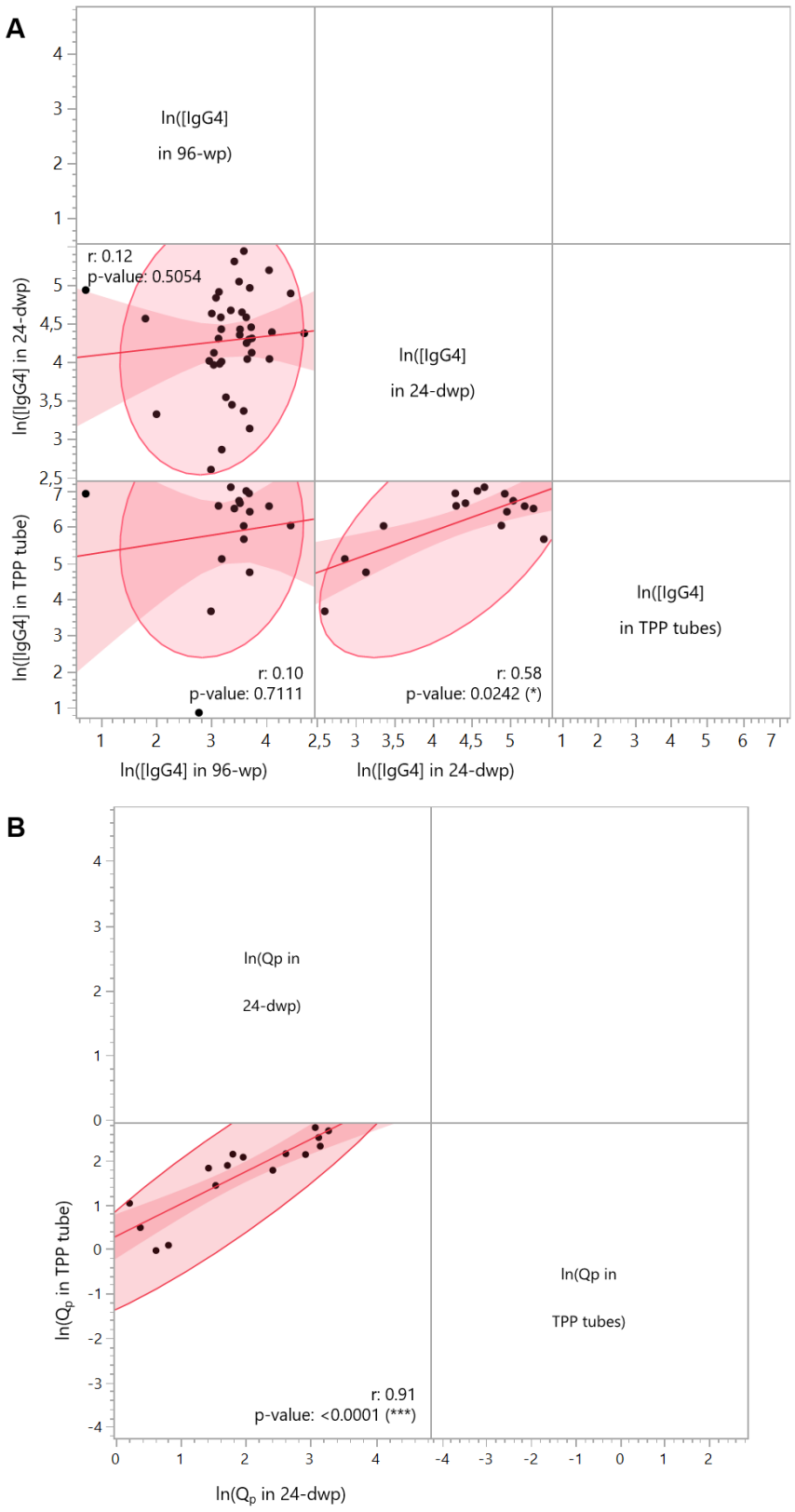


Figure 18: Correlation matrix of titers and specific productivities in 96-wp, 24-dwp, and TPP® tubes

(A) *Correlations between titers in 96-well plates, 24-deepwell plates, and TPP® tubes.* ***(B)*** *Correlations between mean process specific productivities in 24-deepwell plates, and TPP® tubes. Correlations are described by two numbers: the correlation coefficient (noted r) and the probability value (noted p-value). Data analysis was performed on JMP® 15 software.*

3.1.3.3 CHOZN® CHO K1 clones' expression level stability

CHOZN® CHO K1 cell line development platform performance was finally evaluated over a clone expression level stability study. Briefly, two top clones (CL 15D08 and CL 16E06) were sub-cultured without selection pressure in culture medium. Three 12-day fed-batch productivity assays were launched in TPP® tubes, respectively at estimated population doubling level 20, 40 and 60 (PDL_{20}, PDL_{40}, and PDL_{60}). This study duration is defined to span master cell bank (MCB) preparation, working cell bank (WCB) preparation, GMP production, and extended cell bank (ExCB) preparation. Effects of long-term cultivation (LTC) on growth and productivity characteristics was evaluated. More precisely, stability of expression level, considering both IgG4 titer and mean process specific productivity (Q_P) variations over 60 PDL, was the main criteria of interest.

3.1.3.3.1 Effects of long-term cultivation on CHOZN® CHO K1 clones' growth

Doubling time (DT) during sub-culturing over 60 PDL depending on clone is shown on Figure 19. Significant doubling time variations were observed during long-term cultivation, with three distinct phases: from thawing to D_{28}, from D_{28} to D_{54}, and from D_{54} to end of cultivation. From thawing to D_{28}, CL 15D08 doubling time was almost stabilized around 30 hours and CL 16E06 doubling time was varying between 20 and 28 hours. From D_{28} to D_{54}, CL 15D08 doubling time was more variable but decreased between 24 and 30 hours, while CL 16E06 culture showed a completely abnormal profile (sharp increase of doubling time, severe decrease of viability, bending of PDL curve). Finally, from D_{54} to end of cultivation, doubling times of both clones stabilized (between 18 and 20 hours for CL 15D08 and between 20 and 22 hours for CL 16E06).

Investigation on these abnormal doubling time profiles led to identification of culture medium batch (and more precisely specific bottles within this batch) as root cause. Indeed, CL 16E06 back-up culture, which was sub-cultured in EX-CELL® CD CHO Fusion (powder form) supplemented with 6 mM L-glutamine, showed a stable doubling time slightly lower than 20 hours. In addition, noticeable doubling time variations are perfectly concomitant with culture medium bottle change.

Cumulated integral of viable cell density (IVC) during fed-batch productivity assay at PDL_{20}, PDL_{40}, and PDL_{60} depending on clone is shown on Figure 20. Cumulated IVC variations between PDL_{20}, PDL_{40}, and PDL_{60} were calculated. Both clones showed a decrease of cumulated IVC. Decrease of cumulated IVC of CL 15D08 was though less than 30%, while decrease of cumulated IVC for CL 16E06 was higher than 30%. However, doubling time profiles do not explain observed cumulated IVC variations.

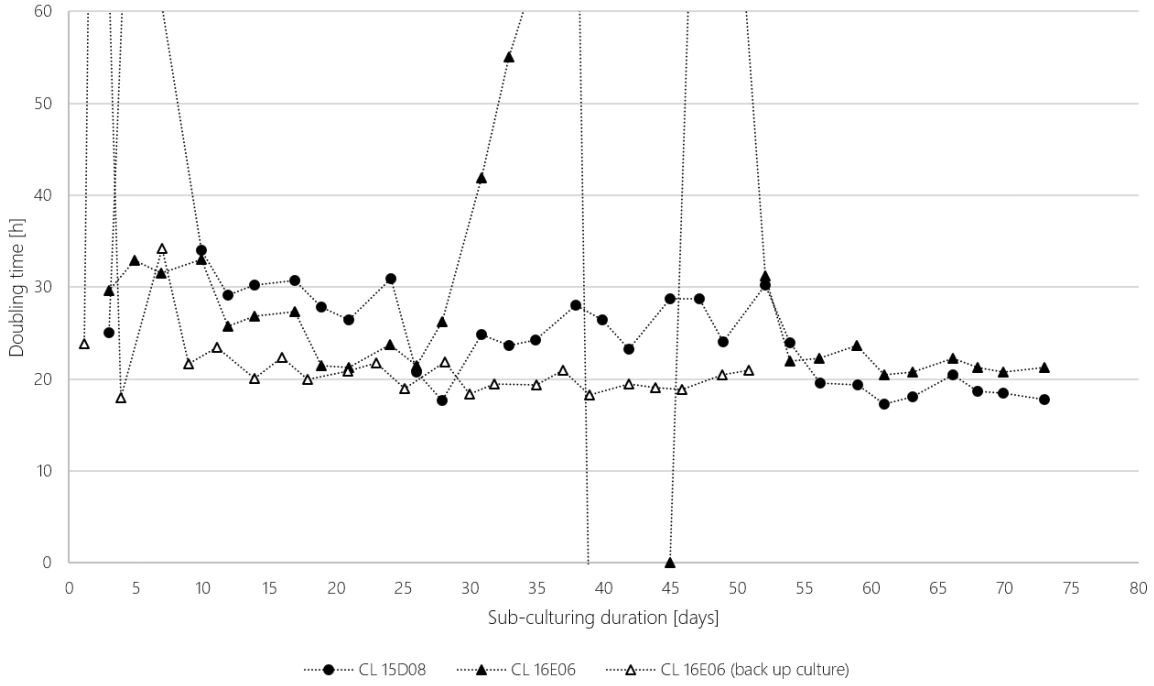


Figure 19: Doubling times over long-term cultivation depending on clone

Two top clones were sub-cultured without selection pressure in culture medium until at least PDL60. This study duration is defined to span master cell bank (MCB) preparation, working cell bank (WCB) preparation, GMP production, and extended cell bank (ExCB) preparation.

3.1.3.3.2 Effects of long-term cultivation on CHOZN® CHO K1 clones' productivity

Cumulated IVC, IgG4 concentration, and mean process specific productivity during fed-batch productivity assay at PDL_{20}, PDL_{40}, and PDL_{60} depending on clone, and cell line development process are respectively shown on Figure 20. IgG4 titer variations between PDL_{20}, PDL_{40}, and PDL_{60} are detailed in Table 30 (Appendix 1). Mean process specific productivity variations between PDL_{20}, PDL_{40}, and PDL_{60} are detailed in Table 31 (Appendix 1). Variations between PDL were calculated on the last common day of culture between the three productivity assays. For example, CL 15D08 culture was stopped at PD_{12} (during productivity assay launched at PDL_{20} and PDL_{40}) and PD_9 (during productivity assays launched at PDL_{60}). Therefore, titer and mean process specific productivity variations were calculated with titer and mean process specific productivity at PD_9 (last common day of culture between the three productivity assays) to determine whether expression level of CL 15D08 is stable or not. As a reminder, a monoclonal cell line is generally considered stable if 70% or more of volumetric productivity titer is retained (or, in other words, if the variation in the production of the recombinant protein is less than 30%) over 60 to 70 generations.

CL 15D08 expression level slightly decreases between PDL_{20} and PDL_{40}, and then sharply decreases between PDL_{40} and PDL_{60}, resulting in a final variation of more than 30% of initial titer (clone considered unstable). CL 15D08 specific productivity was though maintained stable over 60

generations. CL 16E06 expression level slightly increases between PDL_{20} and PDL_{40}, and then slightly decreases between PDL_{40} and PDL_{60}, resulting in a final variation of less than 5% (clone considered stable) of initial titer. CL 16E06 specific productivity was maintained stable over 60 generations.

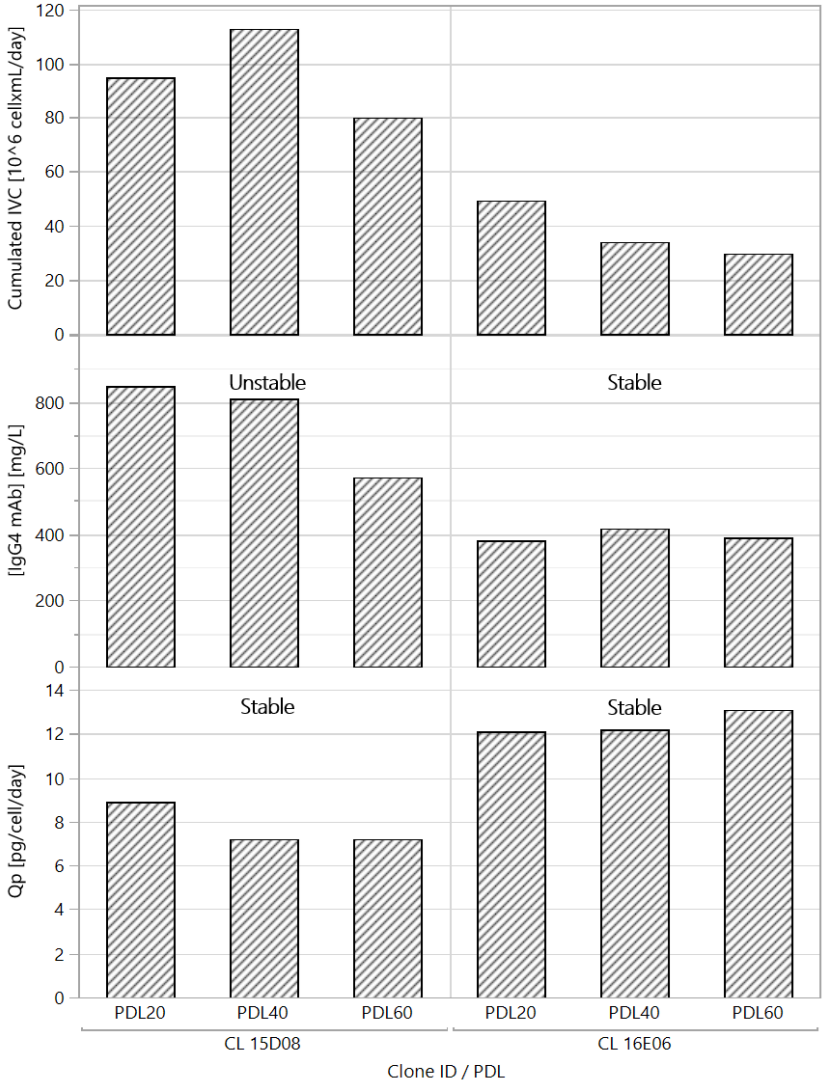


Figure 20: Cumulated IVC, IgG4 concentrations and mean process Q_P after fed-batch productivity assay at PDL_{20}, PDL_{40}, and PDL_{60}

Two top clones were sub -cultured without selection pressure in culture medium. Three 12-day fed-batch productivity assays were launched in TPP® tubes, respectively at estimated PDL_{20}, PDL_{40}, and PDL_{60}. Values displayed on the graph are the values on the last common day of culture between the three productivity assays. For example, CL 15D08 culture was stopped at PD_{12} (during productivity assay launched at $PDL2_0$ and PDL_{40}) and PD_9 (during productivity assays launched at PDL_{60}). Therefore, cumulated IVC, IgG4 concentration, and mean process specific productivity at PD_9 (last common day of culture between the three productivity assays) is plotted for the three productivity assays. IVC: Integral of viable cell density; PDL: population doubling level; Q_P: specific productivity. Data analysis was performed on JMP® 16 software.

3.1.4 Discussion

The CHOZN® CHO K1 platform is one of the cell line development platforms currently offered by Merck End-to-End Solutions for its contract development and manufacturing organization (CDMO) business. This platform, including a host cell line, an expression vector, an optimized set of media and feeds, and a cell line development protocol, is a mammalian cell expression system for selection and scale up of clones producing a recombinant protein.

This first chapter aimed at (i) describing reference characteristics of the CHOZN® CHO K1 platform, mostly related to cell growth after critical process steps (namely transfection and single-cell cloning) and (ii) defining CHOZN® CHO K1 platform initial performance in terms of productivity (both mini-pools and clones), cell line development timelines, and stability (clones).

Firstly, reference indicators related to cell growth have been defined. CHOZN® CHO K1 mini-pools' reemergence (mini-pool confluence threshold being set at 50%) was observed within ten days after transfection (first dilutions of cell suspensions having been made nine days after mini-pools' seeding). This reemergence duration is slightly shorter than reemergence duration expected for CHOZN® $GS^{-/-}$ mini-pools (10 to 21 days mentioned in the CHOZN® $GS^{-/-}$ platform technical bulletin and our internal experience confirms a CHOZN® $GS^{-/-}$ mini-pool reemergence around 12 days). This observation is consistent with host cell lines respective doubling times: CHOZN® CHO K1 host cell line doubling time (16 to 20 hours in sub-culturing process) is indeed slightly lower than that of CHOZN® $GS^{-/-}$ (18 to 22 hours in sub-culturing process).

CHOZN® CHO K1 clones' reemergence (clone confluence threshold being set at 30%) was observed from 12 days after single-cell cloning. Most clones reemerged within three weeks after single-cell cloning, with a final cloning efficiency of 34%. Here again, this cloning efficiency is higher than that obtained on cloning on CHOZN® $GS^{-/-}$ platform (usually around 20%). Similar differences have been observed in a comparative study of CHO K1 and CHO GS KO cell lines, although the selection systems used in this study were different (GS-based selection system for both cell lines, with or without MSX addition) (Noh et al., 2018).

These observations make possible to refine the required timelines for mini-pool and clone sub-culturing in 96-well plates. Mini-pool sub-culturing in 96-well plates, at first sight estimated at about three weeks, could be reduced to two weeks, with batch productivity assay still counting for one additional week. In the same way, clone sub-culturing in 96-well plates, at first sight estimated at about four weeks, could be reduced to three weeks. Exceeding of these thresholds, in other words a slower cellular growth than expected, could highlight a difficult-to-express (DTE) recombinant protein, or even a cytotoxic recombinant protein. Furthermore, possible correlation between growth in 96-well plates and productivity in shaken culture modes (either 24-deepwell plates or Erlenmeyer flasks) has been investigated. Having an indication of when high producers re-emerge (if such an observation can be

made, of course) could make it possible to concentrate lab work efforts on this period (clone sub-culturing being particularly cumbersome). Link between growth rate and specific productivity is regularly debated. Several reports support the idea that high producers are slow growers. On the contrary, other studies defend the opinion that recombinant protein expression level has no effect on recombinant cell line growth rate (Chusainow et al., 2009; Fann et al., 2000). No correlation could be established within this study. Indeed, times at which confluence was measured (D_{12}, D_{14}, D_{19}, D_{21}, and D_{27}) were certainly too far apart to establish such a correlation. In a next assay, it would be more appropriate to measure confluence daily, and to look again at a possible correlation.

Secondly, CHOZN® CHO K1 platform initial performance has been defined. CHOZN® CHO K1 mini-pools, expressing a humanized IgG4 monoclonal antibody, were generated and screened through a mini-pool generation process including batch productivity assay in 96-well plates and fed-batch productivity assay in 24-deepwell plates. CHOZN® CHO K1 mini-pool maximum titer, after fed-batch productivity assay in 24-deepwell plates, nearly reached 300 mg/L. This productivity level was much lower than that obtained on previous evaluation of the CHOZN® CHO K1 platform (best mini-pool produced 1.3 g/L of the same recombinant monoclonal antibody after fed-batch productivity assay in TPP® tubes). This low mini-pool productivity level might be explained by application of non-optimized process parameters (more especially, culture volume and orbital shaking conditions, which are key culture parameters in 24-deepwell plates (Duetz, 2007)), due to the first-time use of the 24-deepwell plate format.

CHOZN® CHO K1 clones were then generated (by single-cell cloning of the three best mini-pools previously generated) and screened through a clone generation process including productivity measurement in 96-well plates, batch productivity assay in 24-deepwell plates, and fed-batch productivity assay in TPP® tubes. CHOZN® CHO K1 clone maximum titer, after fed-batch productivity assay in TPP® tubes, reached 1 g/L. The 1 g/L productivity target was reached but not the expectation of a final titer at 2 g/L. This expectation was based on the results obtained during previous evaluation of the CHOZN® CHO K1 platform (best mini-pool producing 1.3 g/L) and productivity increase between mini-pool and clone generations (usually, a 1.5-fold titer increase between mini-pool and clone generation). Here again, this low clone productivity level might be explained by application of non-optimized process parameters for culture in 24-deepwell plates. Firstly, selection of mini-pools for single-cell cloning was based on productivity assay in 24-deepwell plates. The best mini-pools may not have been identified because of non-optimized process parameters. Secondly, cumbersome sub-culturing of clones 24-deepwell plates led to the loss of approximately 40% of cell suspensions, among which were possibly high producers.

Recombinant CHOZN® CHO K1 cell lines, producing up to 1 gL of a standard IgG4 monoclonal antibody, were thus generated within 7.5 months (all development steps from DNA to pre-master cell bank included). Both productivity level and cell line development timelines are out of the standards of

industrial cell line development platforms, naturally identifying these two criteria as the main areas of work for project continuation.

Table 5: Key features of cell line development platforms

Results of a technology scouting performed in 2021.

CDMO	CLD timelines [months]	Clone productivity [g/L]	Host cell line	Expression platform	Automation
WuXi Biologics	2.5	7-10	CHO K1	WuXia™	ClonePix
KBI BIOPHARMA	3	8	CHO-M, CHO K1	Selexis SURE Technology™	ClonePix, Beacon®
Lonza	3	5-6	CHO DG44	ApolloX™	CytoMine®
Catalent	3	2-7	CHOvolution® (CHO K1)	GPEx®, SEFEX®	FACSAria Fusion
FUJIFILM Diosynth biotechnologies	3.5	4-5	GS Xceed®, CHO K1	GS Xceed®	Beacon®
CELONIC	4	2-6	CHOZN®	GPEx®; GPEx® Boost	Beacon®
AGC Biologics	4	2-5	CHO DG44	CHEF1™	-
patheon by Thermo Fisher Scientific	4	3	CHO-STM, CHO DG44	Gibo™, Freedom™	Beacon®
MERCK	7	1-3	CHOZN®	EX-CELL®	-

An important modification has been applied on the cell line development process used in this chapter. As mentioned before, the 24-deepwell plate format, instead of the 6-well plate and TPP® tube culture formats, was chosen because of timelines constraints during mini-pool generation, and then again applied during clone generation. The 24-deepwell plates indeed combine advantages of both 6-well plates and TPP® tubes. On the one hand, like the 6-well plates, the 24-deepwell plates are a miniaturized culture scale, where an important number of conditions can be evaluated, therefore increasing screening throughput (Mora et al., 2018). On the other hand, like the TPP® tubes, the 24-deepwell plates allow shaken culture, which appears to be more relevant than static culture for reliable identification of high producers. Small but promising positive correlation between titers in 24-deepwell plates and TPP® tubes (correlation coefficient of 0.5), as well as high positive correlation between mean process specific productivities in 24-deepwell plates and TPP® tubes (correlation coefficient of 0.9) were observed. This is in line with recent reports found in the literature. It has indeed been shown that 24-deepwell plates offer a reliable platform for enabling successful ranking of cell lines candidates along cell line development activities (Chaturvedi et al., 2014; Mora et al., 2018). This culture format, used in its early days mainly for microbial culture, can now be considered as a mature alternative to culture in shaken flasks and is little by little implemented in several industrial cell line developments platforms, like those of AbbVie and Merck (Duetz, 2007; Mora et al., 2018; Rouiller et al., 2016). Further optimization work on 24-deepwell plate culture format was conducted by Merck End-to-End Solutions' cell line development team, on both CHOZN® cell lines. Optimized culture parameters have been used in the following chapters.

Lastly, expression level stability of two top CHOZN® CHO K1 clones was investigated over a stability study. This stability study was defined in the manner of a stability study conducted as part of a CDMO project. Fed-batch productivity assays were launched at estimated population doubling level 20, 40 and 60 (PDL_{20}, PDL_{40}, and PDL_{60}) – a study duration defined to span master cell bank (MCB) preparation, working cell bank (WCB) preparation, GMP production, and extended cell bank (ExCB) preparation. Selection pressure was not maintained during sub-culturing process.

Significant disturbances on cell growth (seen on viability doubling time and population doubling level profiles) have been observed for both clones. According to the investigation led root cause identification, there is no denying that cell growth has been impacted by culture medium, particularly for CL 16E06. It is striking to see that CL 15D08 and CL 16E06, although coming from the same single-cell cloning round and producing the same molecules, responded differently to this perturbation. In the case of CL 16E06, it is possible that a sub-population has been selected. This hypothesis is based on the observation of a difference of doubling times between the cell suspension that underwent growth perturbation (stabilized above 20 hours) and the back-up cell suspension (stabilized below 20 hours). This idea is also consistent with clonal heterogeneity, which has been documented several times for CHO cell lines (Scarcelli et al., 2018; Tharmalingam et al., 2018).

At least 70% of volumetric productivity was retained over 60 PDL for CL 15D08, but a titer decrease of more than 30% between PDL_{20} and PDL_{60} was observed for CL 16E06. Mean process specific productivity was though considered stable for both clones. Unfortunately, the design of the study itself does not allow for identification of root causes of CL 15D08 instability. The absence of a control condition, for which selection pressure would be maintained during sub-culturing process, and the absence of analytics (such as gene copy number monitoring) prevent from proposing hypotheses on the causes of instability.

As a consequence, conclusions that can be drawn from this clone stability seem to be limited. Although this study already opens the door to some observations on the stability of CHOZN® CHO K1 clones, a much wider stability study would be necessary to establish the real behavior CHOZN® CHO K1 clones over long-term cultivation. This wider stability study could be a least based on a largest number of clones first, ideally producing several recombinant proteins, with and without selection pressure during sub-culturing, and with additional analytics to explain a possible expression level decrease.

3.1.5 Conclusion

In this first chapter, growth characteristics and initial performance (in terms of productivity, stability and timelines) of the CHOZN® CHO K1 cell line development platform have been defined. This initial platform performance characterization will be considered as point of comparison for the optimizations made during this project and presented in the following chapters. Indeed, two main areas of work have been identified: (i) reduction of cell line development timelines and (ii) increase of recombinant cell

lines productivity. In the second chapter, an accelerated cell line development process, called direct single-cell cloning (DSCC) will be evaluated to reduce cell line development timelines. In the third chapter, several promoters driving resistance to selection pressure expression will be evaluated to increase clonal cell lines productivities.

3.1.6 Author's contributions

Literature review: Lucie Grindes[1,2] (ORCID ID: 0000-0002-6201-3355) ; **Methodology**: Lucie Grindes; **Investigation**: Lucie Grindes, Marine Amouriq[1] and Amandine Philippat[1] (significant technical help during clone generation), Myriam Eyquard[3] (significant technical help during clone stability study); **Analysis**: Lucie Grindes; **Writing – Original draft and editing**: Lucie Grindes; **Writing – Review**: Lucie Grindes, Gilles Joucla[2] (ORCID ID: 0000-0003-4192-9628), Wilfrid Dieryck[2] (ORCID ID: 0000-0002-4609-8934), Cyrielle Corbin[1] (ORCID ID: 0000-0002-6891-2511), Sébastien Ribault[4] (trade secret and intellectual property), Daniel Kasten[5] (trade secret and intellectual property); **Supervision**: Cyrielle Corbin.

[1] Cell line development team, Process development department, Merck Biodevelopment

[2] Structure and activity of biological macromolecules team, Nutraceutical, medical and food applications pole, CBMN (UMR 5842 (CNRS, University of Bordeaux, Bordeaux INP))

[3] Up-Stream process team, Process development department, Merck Biodevelopment

[4] Direction, Merck Biodevelopment

[5] Patents Life Science global team, MilliporeSigma

3.2 Cell line development process optimization: implementation of direct single-cell cloning

3.2.1 Introduction

Cell line development is a decisive step in a recombinant therapeutic molecule development, as this activity is responsible for generating the production cell line (Diep et al., 2021), and is on the critical path for advancing a novel therapeutic to first in human (FIH) clinical trials (Munro et al., 2017). As a consequence, the International Conference on Harmonization of Technical Requirements for Registration of Pharmaceuticals for Human Use (ICH) put in place guidelines in the 1990s to regulate the activities linked to cell line development (Tharmalingam et al., 2018).

Since the 1980s, the fundamental concepts of cell line development have though remained unchanged (F. M. Wurm, 2004). Thus, cell line development is based on a well-established scheme, which starts with the transfection of the host cell line with an expression vector. The expression vector contains the transcriptional and translational elements necessary for expression and biosynthesis of the recombinant protein, as well as elements that confer a selective advantage to the cells that have integrated the plasmid. After transfection, cells that have integrated the expression vector into their genome are selected, either in the absence of an essential metabolite or in the presence of an antibiotic. The surviving cells are then individually isolated to ensure the generation of clonally derived cell lines from a single cell (Lai et al., 2013; F. M. Wurm, 2004). A broad range of techniques are used for single-cell cloning, including (but not limited to) limiting dilution, microfluidic encapsulation, fluorescence activated cell-sorting (FACS), colony picking in semi-solid media, single-cell seeders, and microfluidics chips. The clones thus generated are then evaluated on growth and production criteria of the protein of interest (quantity and quality). At the end of these evaluations, sometimes iterative, a cell line presenting the desired characteristics of growth and productivity is selected. This cell line will be the recombinant protein-producing cell line.

Cell line development process associated to CHOZN® CHO K1 platform is no exception to the rule. Generation of monoclonal cell lines is a two-step based process, divided in (i) a mini pool (MP) generation and selection step and in (ii) a clone (CL) generation and selection step (Figure 11). The mini pool generation step aims at easing high producing pools identification, while the clone generation step aims at delivering a clonally derived, high producing cell line. Both mini pool and clone generation steps rely on iterative process for identification and selection of high producers: subculturing and productivity assay in 96-well plates, scale-up and productivity assay in 6-well plates, and scale-up and productivity assay in TPP® tubes. Cell line development timelines are estimated to span 7.5 months (from DNA to pre-master cell banks (pre-MCB)). Similar two-step based processes, with mini-pool and clone

generation, have been reported by process development teams from Amgen and Bristol-Myers Squibb (Diep et al., 2021; Fan et al., 2017).

Cell line development timelines associated to CHOZN® CHO K1 platform need to be decreased. Firstly, a 7.5-month process is not viable when all development (cell line development, upstream development, downstream development, and analytical development) and production activities (MCB preparation, GMP production, and quality control) should be completed within one year. Indeed, significant efforts are currently made to accelerate new medicine approval process, especially for unmet medical needs (Munro et al., 2017). In the same impulse, breakthrough designation and Priority Medicines (PRIME) designation have been recently introduced, respectively by FDA in 2012 and by EMA in 2016. Nivolumab and pembrolizumab are two examples of monoclonal antibodies that benefited from accelerated approval by health authorities (Fan et al., 2017). Secondly, a 7.5-month process is not viable when most competitors' industrial cell line development platforms enable cell line development within 3 months. The competitiveness and attractiveness of the CHOZN® CHO K1 platform are at stake.

We focused our work on the cell line development process applied on the CHOZN® CHO K1 platform. Our objective was to evaluate a shortened cell line development process, called direct single-cell cloning (DSCC), in order to reduce cell line development timelines.

3.2.2 Methodology

A shortened cell line development process, called "short cell line development process (direct single-cell cloning (DSCC))", has been designed, to be compared to the traditional cell line development process, called "long cell line development process (mini-pools – clones (MP-CL))". CHOZN® CHO K1 clones were generated by single-cell cloning using limiting dilution method, either following the long process (MP-CL) or the short process (DSCC). Clones' productivity was evaluated in 96-well plates, 24-deepwell plates, TPP® tubes, and additionally in ambr250® bioreactors. Clone expression level stability was investigated over 60 Population Doubling Levels (PDL).

3.2.3 Design of a direct single-cell cloning (DSCC) protocol

The designed direct single-cell cloning process consisted in:

(i) Transfecting the host cell line, with increased quantity of cells to be electroporated but identical electroporation parameters to that used in long CLD process (MP-CL).

(ii) Putting the transfected bulk pool under selection pressure, with addition of 10 µg/mL puromycin in culture medium 24 hours after transfection.

(iii) Letting the transfecting bulk pool reach the lowest viability. A viability decrease was expected, resulting from transfection and, above all, from selection pressure addition.

(iv) Single-cell cloning the transfected bulk pool by limiting dilution, seven days after transfection. A seven-day period was considered as the minimum duration of selection

between transfection and single-cell cloning and planning constraints were not allowing for a longer selection duration. Puromycin was planned to be withdrawn from culture media at this stage, following the model of clone generation in the long CLD process (MP-CL), where single-cell cloning and clone generation are performed without selection pressure.

(v) Selection of best producing clones, with iterative selection process in 96-well plates (productivity measurement), 24-deepwell plates (batch productivity assay), and TPP® tubes (fed-batch productivity assay), identical to those performed in long CLD process (MP-CL).

3.2.4 Results

3.2.4.1 Shortened cell line development process decreases cloning efficiency and clone productivity

The designed short cell line development process (DSCC) was compared to the traditional long cell line development process (MP-CL) over a clone generation.

Briefly, for the long cell line development process, CHOZN® CHO K1 host cell line was transfected by electroporation and mini-pools were generated. Three best mini-pools were selected for single-cell cloning. On the other hand, for the short cell line development process, CHOZN® CHO K1 host cell line was transfected by electroporation and transfected bulk pool was let to reach the lowest viability. The two cell line development processes (long process (MP-CL) *versus* short process (DSCC)) were run in parallel from single-cell cloning step. Clones were cultured in 96-well plates until cellular reemergence was observed; a maximum of four weeks was allowed for reemergence. When the first of these two criteria was met, recombinant antibody concentration in cell culture supernatant was measured by bio-layer interferometry using protein A-coated sensors. Subsequent scale-up was first performed in 24-deepwell plates (sub-culturing and batch productivity assay) and then performed in shake flasks (sub-culturing) and TPP® tubes (fed-batch productivity assay). Lastly, clones' behavior and productivity was confirmed in ambr250® bioreactors.

CHOZN® CHO K1 transfected bulk pool behavior was first characterized during selection step. Cloning efficiency and clone productivity were considered as the main criteria of interest for evaluation of the short cell line development (DSCC) strategy.

3.2.4.1.1 Characterization of CHOZN® CHO K1 cell behavior during selection

Viability profile of transfected bulk pool during selection (*i.e.*, between transfection and single-cell cloning) is shown on Figure 21 (red curve). A sharp viability decrease was observed between viability prior transfection and viability 24 hours after electroporation (62%). Viability was maintained around 60% until day 4 (D_4). Then, a sharp viability decrease was again observed, with transfected bulk pool reaching a minimum viability (16.2%) on D_7. As the whole volume of cell suspension was used for single-cell cloning, recovery profile of the transfected bulk pool is not known.

Several clone generations following the direct single-cell cloning protocol were later performed. Viability profiles of corresponding transfected bulk pools are shown on Figure 21 (black and grey curves). Sharp viability decrease was observed for all conditions within the 24 hours following transfection. However, viability range was wide, ranging from 37% to 85.2% one day after electroporation. The viability plateau between D_1 and D_4 observed on DSCC assays 1 and 2 was not observed on DSCC assays 3 and 4. Instead, viability drop continues until D_8. Lowest viability was predominantly reached on D_8 and ranged in most cases between 5 and 20%. Transfected bulk pool recovery was observed in the following days, at the latest three weeks after transfection.

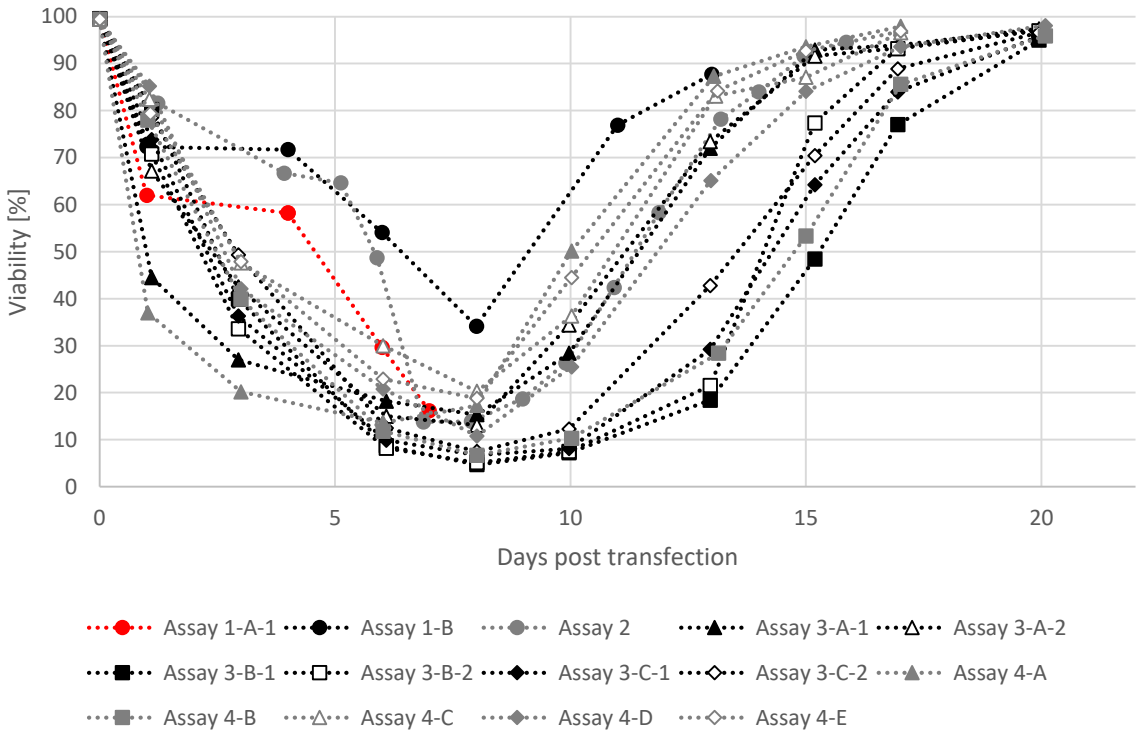


Figure 21: Viability profile of transfected bulk pool during selection

CHOZN® CHO K1 host cell line was transfected by electroporation. 24 hours after transfection, transfected bulk cell suspension was centrifuged, and cell pellet was resuspended at 1x10^6 viable cells/mL in culture medium supplemented with puromycin. Transfected bulk pool was let to reach the lowest viability and recover. Transfected bulk pool was centrifuged when viability lost 20 points (or more) compared to previous time point and/or if viability was below 60% and passaged at a seeding density of 1x10^6 vc/mL as long as viability was below 60%. Viability curve of transfected bulk pool used for clone generation with direct single-cell cloning process is highlighted in red. Viability curves of transfected bulk pools later performed are highlighted in black and grey. Details on experimental conditions of these assays are given in Table 21.

3.2.4.1.2 Impact of cell line development process on cloning efficiency and clone reemergence

Cloning efficiencies for both CLD processes are detailed in Table 6. Lower cloning efficiency was obtained with short CLD process (DSCC) (16%), compared to long CLD process (MP-CL) (34%). This difference is directly linked to cell suspension viability on the day of single-cell cloning. Indeed,

viability of cell suspension cloned for short CLD process was lower (16.2%, transfected bulk pool having experienced expected viability decrease after transfection and selection pressure addition and not having yet recovered from selection) than that of cell suspension cloned for long CLD process (98.6%, blend of three healthy, exponentially growing mini-pools).

Clone confluence in 96-well plate, as a function of subculturing duration and cell line development process is shown on Figure 22. Details on percentage on reemerging clones and cumulated percentage of reemerging clones are given in Table 6. First re-emerging clones were observed from 12 days after cloning, for both CLD process, although in greater proportion for long CLD process (16%) compared to short CLD process (2%). Clone reemergence is faster for clones generated with long CLD process (42% as cumulated percentage) compared to clones generated with short CLD process (34% as cumulated percentage) until D_{14}. From D_{19} though, cumulated percentages of reemerging clones are equivalent between both CLD processes (88% at D_{19}, 93% at D_{21} and 100% at D_{27}). Finally, for one CLD process as for the other, no correlation is observed between reemergence duration and productivity level in TPP® tubes.

Table 6: Cloning efficiency and percentages of reemerging clones as a function of subculturing duration in 96-well plates

Cell line development process	**Long process (MP-CL)**	**Short process (DSCC)**
Cell suspension viability on the day of single-cell cloning	98.6%	16.2%
Cloning efficiency	34%	16%
Percentage of reemerging clones, as a function of subculturing duration in 96-wp	16% at D_{12} 26% at D_{14} 46% at D_{19} 5% at D_{21} 7% at D_{27}	2% at D_{12} 26% at D_{14} 59% at D_{19} 4% at D_{21} 9% at D_{27}
Cumulated percentage of reemerging clones, as a function of subculturing duration in 96-wp	16% at D_{12} 42% at D_{14} 88% at D_{19} 93% at D_{21} 100% at D_{27}	2% at D_{12} 28% at D_{14} 87% at D_{19} 91% at D_{21} 100% at D_{27}

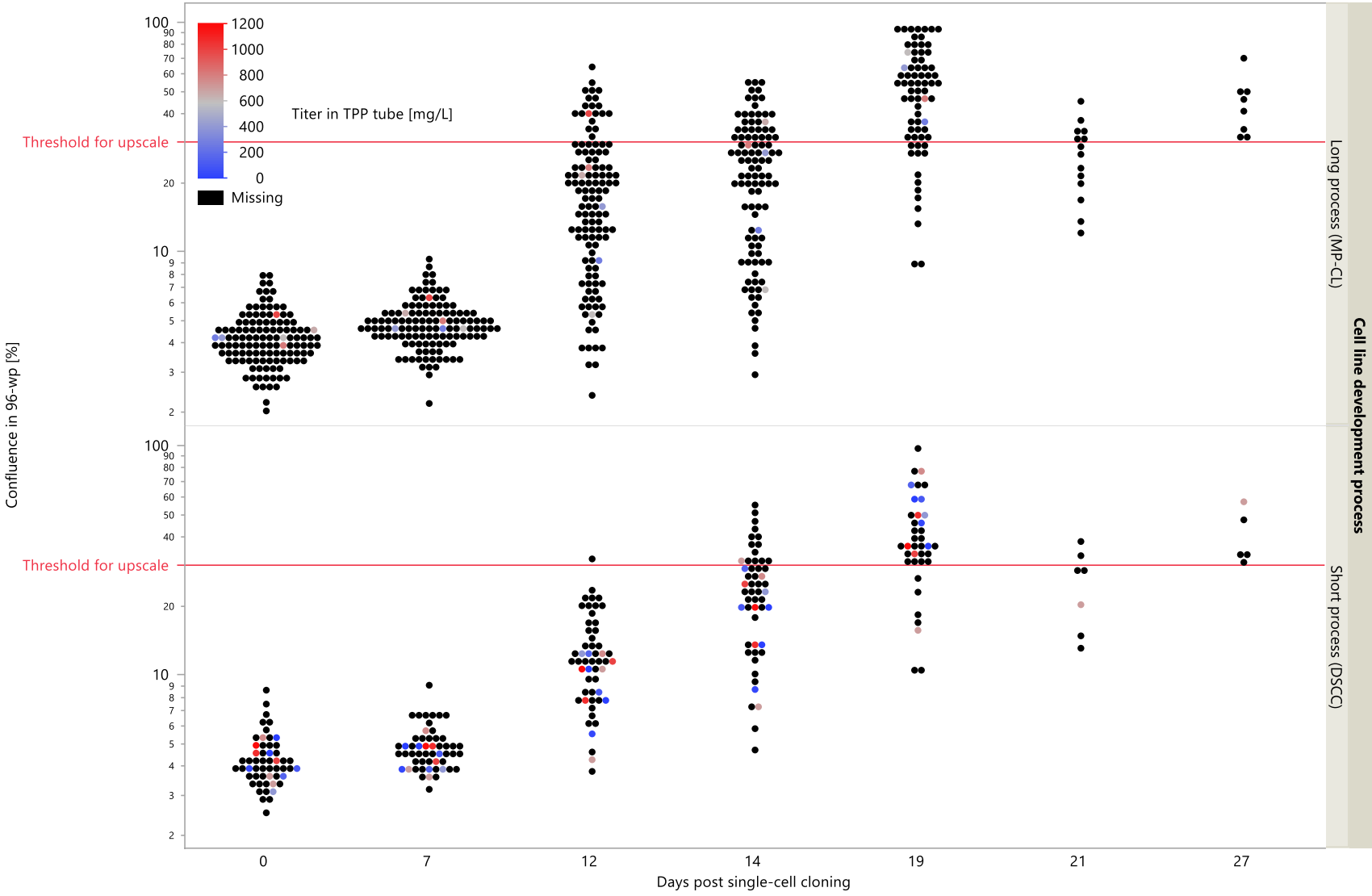


Figure 22: Clone confluence in 96-wp, as a function of subculturing duration and cell line development process

3.2.4.1.3 Impact of cell line development process on clone productivity

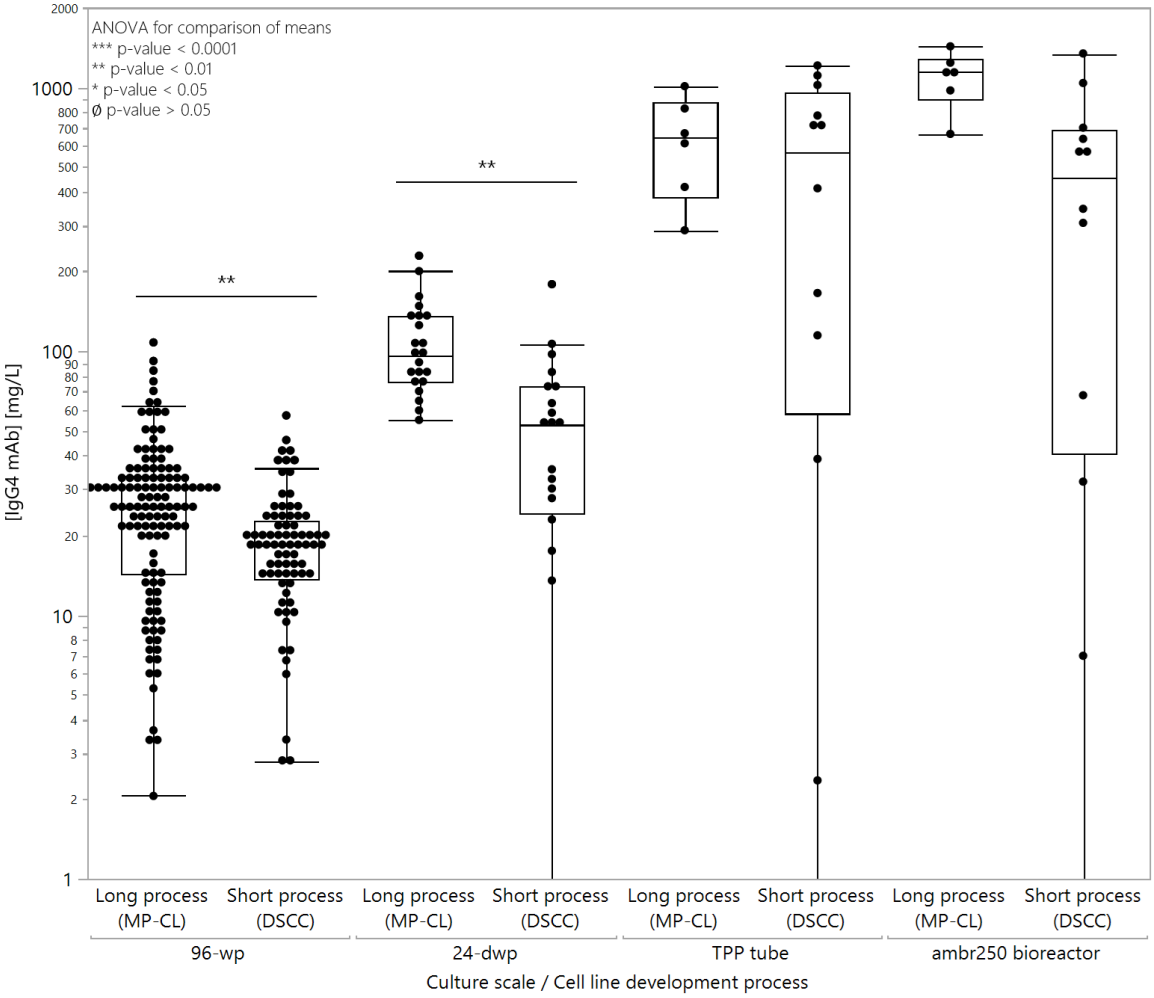
The impact of cell line development process on clone productivity was then evaluated. Comparison of clone productivity between both processes was performed at each stage, *i.e.*, in 96-well plates, 24-deepwell plates, TPP® tubes, and additionally in ambr250® bioreactors.

Clones' titers after productivity measurement in 96-well plates, and productivity assays in 24-deepwell plates, TPP® tubes, and ambr250® bioreactors are shown on Figure 23. Maximum IgG4 titer is higher for clones generated with long CLD process (MP-CL) at 96-well plate, 24-deep-well and ambr250® bioreactor stages, compared to clones generated with short CLD process (DSCC). Short CLD process leads to a 0.8-fold change on IgG4 maximum titer in 24-deepwell plates and a 0.9-fold change on IgG4 maximum titer in ambr250® bioreactors. In addition, median IgG4 titer is higher for clones generated with long CLD process (MP-CL) at all culture stages, compared to clones generated with short CLD process (DSCC). Short CLD process leads to a 0.9 (in TPP® tubes) to 0.4-fold change (in ambr250® bioreactors) on IgG4 median titer.

Significance of mean titer difference between clones generated *via* long or short CLD process in 96-well plates, 24-deepwell plates, TPP® tubes, and ambr250® bioreactors was evaluated by a statistical analysis (details of which can be found in Appendix 2). Mean IgG4 titer is significantly higher for clones generated by long CLD process (MP-CL) in 96-well plates and 24-deepwell plates.

Mean process specific productivity during productivity assays in 24-deepwell plates, TPP® tubes, and ambr250® bioreactors are shown on Figure 24. Similar to observations made on IgG4 titer, mean process Q_P is higher for clones generated with long CLD process (MP-CL) at 24-deep-well and ambr250® bioreactor stages, compared to clones generated with short CLD process (DSCC). Short CLD process leads to a 0.8-fold change on maximum process Q_P in 24-deepwell plates and a 0.7-fold change on maximum process Q_P in ambr250® bioreactors. In addition, median mean process Q_P is higher for clones generated with long CLD process (MP-CL) at all culture stages, compared to clones generated with short CLD process (DSCC). Short CLD process leads to a 0.4-fold change on mean process Q_P in ambr250® bioreactors and a 0.2-fold change on mean process Q_P in 24-deepwell plates.

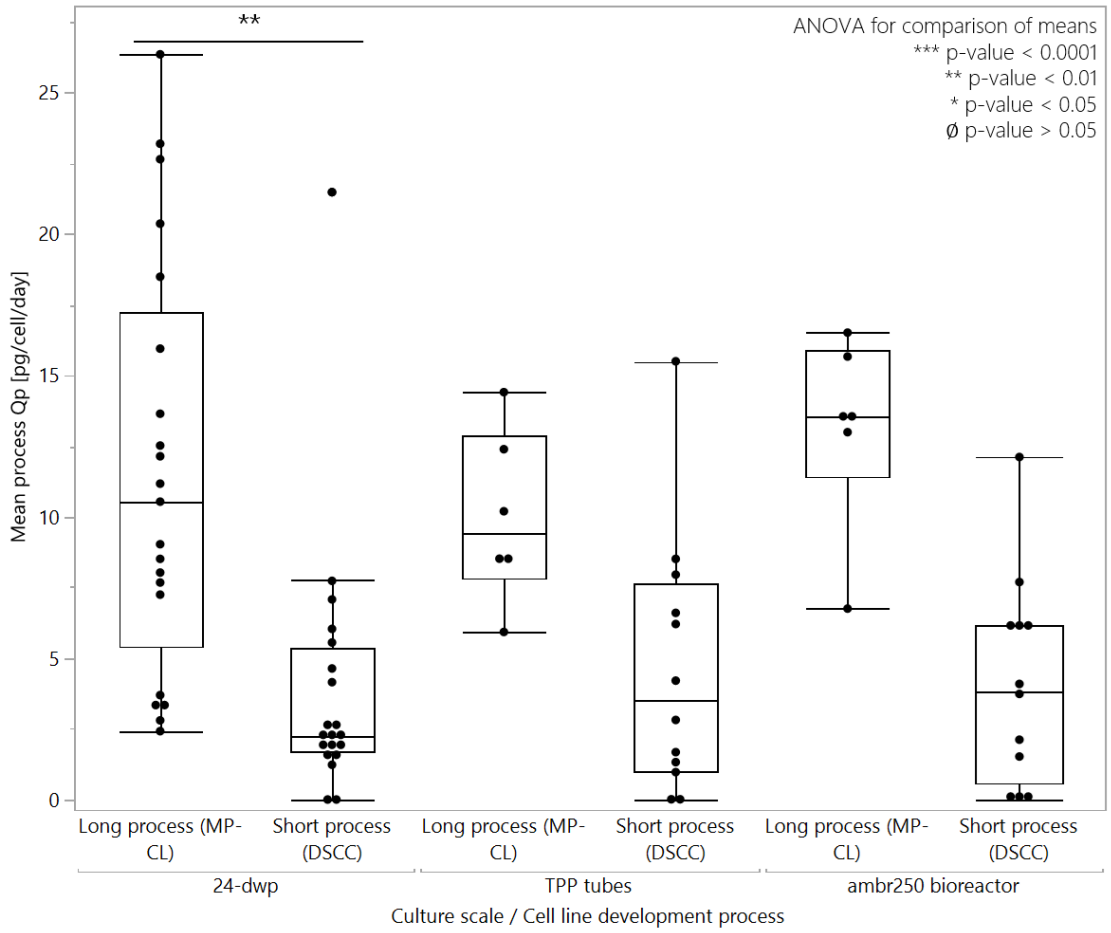
Significance of mean process Q_P difference between clones generated *via* long or short CLD process in 24-deepwell plates, TPP® tubes, and ambr250® bioreactors was evaluated by a statistical analysis (details of which can be found in Appendix 2). Mean specific productivity is significantly higher for clones generated by long CLD process (MP-CL) in 24-deepwell plates.



Culture scale	96-wp		24-dwp		TPP® tubes		ambr250® bioreactors	
CLD process	Long process (MP-CL)	Short process (DSCC)	Long process (MP-CL)	Short process (DSCC)	Long process (MP-CL)	Short process (DSCC)	Long process (MP-CL)	Short process (DSCC)
Max titer [mg/L]	107.8	57.0	229.7	179.2	1009	1209	1428	1342
Fold change on max titer	-	0.5	-	0.8	-	1.2	-	0.9
Median titer [mg/L]	22.5	18.4	96.5	52.9	641	564	1139	451
Fold change on mean titer	-	0.8	-	0.5	-	0.9	-	0.4

Figure 23: IgG4 mAb concentration in clones' supernatants after productivity measurement in 96-well plates and productivity assays in 24-deepwell plates, TPP® tubes and ambr250® bioreactors

Box-and-whiskers plots of IgG4 concentration in clone's supernatant after productivity measurement in 96-well plates, batch productivity assay in 24-deepwell plates and fed-batch productivity assays in TPP® tubes and ambr250® bioreactors. Each dot represents one clone. Boxplot horizontal lines respectively indicate, top to bottom, the maximum, the upper quartile (Q3), the median, the lower quartile (Q1) and the minimum IgG4 concentration. Fold change on maximum and median titer is calculated in relation to long cell line development process (MP-CL) at all culture scales. Data analysis was performed on JMP® 16 software.



Culture scale	24-dwp		TPP® tubes		ambr250® bioreactors	
CLD process	Long process (MP-CL)	Short process (DSCC)	Long process (MP-CL)	Short process (DSCC)	Long process (MP-CL)	Short process (DSCC)
Max Q_P [mg/L]	26.4	21.5	14.4	15.5	16.5	12.1
Fold change on max Q_P	-	0.8	-	1.1	-	0.7
Median Q_P [mg/L]	10.5	2.2	9.4	3.5	13.6	3.8
Fold change on mean Q_P	-	0.2	-	0.4	-	0.3

Figure 24: Mean process specific productivities after productivity measurement in 96-well plates and productivity assays in 24-deepwell plates, TPP® tubes and ambr250® bioreactors

Box-and-whiskers plots of mean process specific productivities during batch productivity assay in 24-deepwell plates and fed-batch productivity assays in TPP® tubes and ambr250® bioreactors. Each dot represents one clone. Boxplot horizontal lines respectively indicate, top to bottom, the maximum, the upper quartile (Q3), the median, the lower quartile (Q1) and the minimum mean process specific productivity. Fold change on maximum and median mean process specific productivity is calculated in relation to long cell line development process (MP-CL) at all culture scales. Data analysis was performed on JMP® 16 software.

3.2.4.2 Impact of cell line development process on clone expression level stability

Lower clone productivity was observed for clones generated with short cell line development process (DSCC), compared to clones generated with long cell line development process (MP-CL). Impact of cell line development process on expression level stability was further investigated.

Briefly, two top clones, respectively generated by long (MP-CL) and short (DSCC) cell line development processes, were sub-cultured without selection pressure in culture medium. Three 12-day fed-batch productivity assays were launched in TPP® tubes, respectively at estimated population doubling level 20, 40 and 60 (PDL_{20}, PDL_{40}, and PDL_{60}). This study duration is defined to span master cell bank (MCB) preparation, working cell bank (WCB) preparation, GMP production, and extended cell bank (ExCB) preparation. Effects of long-term cultivation (LTC) on growth and productivity characteristics was evaluated. More precisely, stability of expression level, considering both IgG4 titer and mean process specific productivity (Q_P) variations over 60 PDL, was the main criteria of interest.

This stability study was conducted over four clones: CL 15D08 and CL 16E06 have been generated with long CLD process and CL 38A06 and CL 38C02 have been generated with short CLD process. Results of CL 15D08 and CL 16E06 have been previously presented (CHOZN® CHO K1 clones' expression level stability); results of CL 38A06 and CL 38C02 will be detailed hereinafter.

Cumulated IVC, IgG4 concentration, and mean process specific productivity during fed-batch productivity assay at PDL_{20}, PDL_{40}, and PDL_{60} depending on clone, and cell line development process are respectively shown on Figure 25. IgG4 titer variations between PDL_{20}, PDL_{40}, and PDL_{60} are detailed in Table 32 (in Appendix 2). Mean process specific productivity variations between PDL_{20}, PDL_{40}, and PDL_{60} are detailed in Table 33 (in Appendix 2). Variations between PDL were calculated on the last common day of culture between the three productivity assays. For example, CL 15D08 culture was stopped at PD_{12} (during productivity assay launched at PDL_{20} and PDL_{40}) and PD_9 (during productivity assays launched at PDL_{60}). Therefore, titer and mean process specific productivity variations were calculated with titer and mean process specific productivity at PD_9 (last common day of culture between the three productivity assays) to determine whether expression level of CL 15D08 is stable or not.

CL 38A06 behaves similarly to CL 16E06. IgG4 titer variation and mean process specific productivity variation between PDL_{20} and PDL_{60} are less than 30% (and even less than 10% for IgG4 titer variation and less than 20% for mean process specific productivity variation). Slight increase in mean process specific productivity is explained by slight decrease of cumulated IVC while IgG4 titer is stable.

CL 38C02 behaves similarly to CL 15D08. IgG4 titer variation between PDL_{20} and PDL_{60} is greater than 30%. Decrease of productivity takes place between PDL_{40} and PDL_{60} (IgG4 titer variation between PDL_{20} and PDL_{60} being less than 5% for both clones). However, decrease in mean process specific productivity results in a stable mean process specific productivity over PDL_{60} for both clones.

The behavior of CHOZN® CHO K1 clones generated with short cell line development process (DSCC) does not seem to differ from clones generated with long cell line development process (MP-CL), in terms of titer and mean process specific productivity stability. No statistical analysis was done on these results (insufficient number of samples per condition).

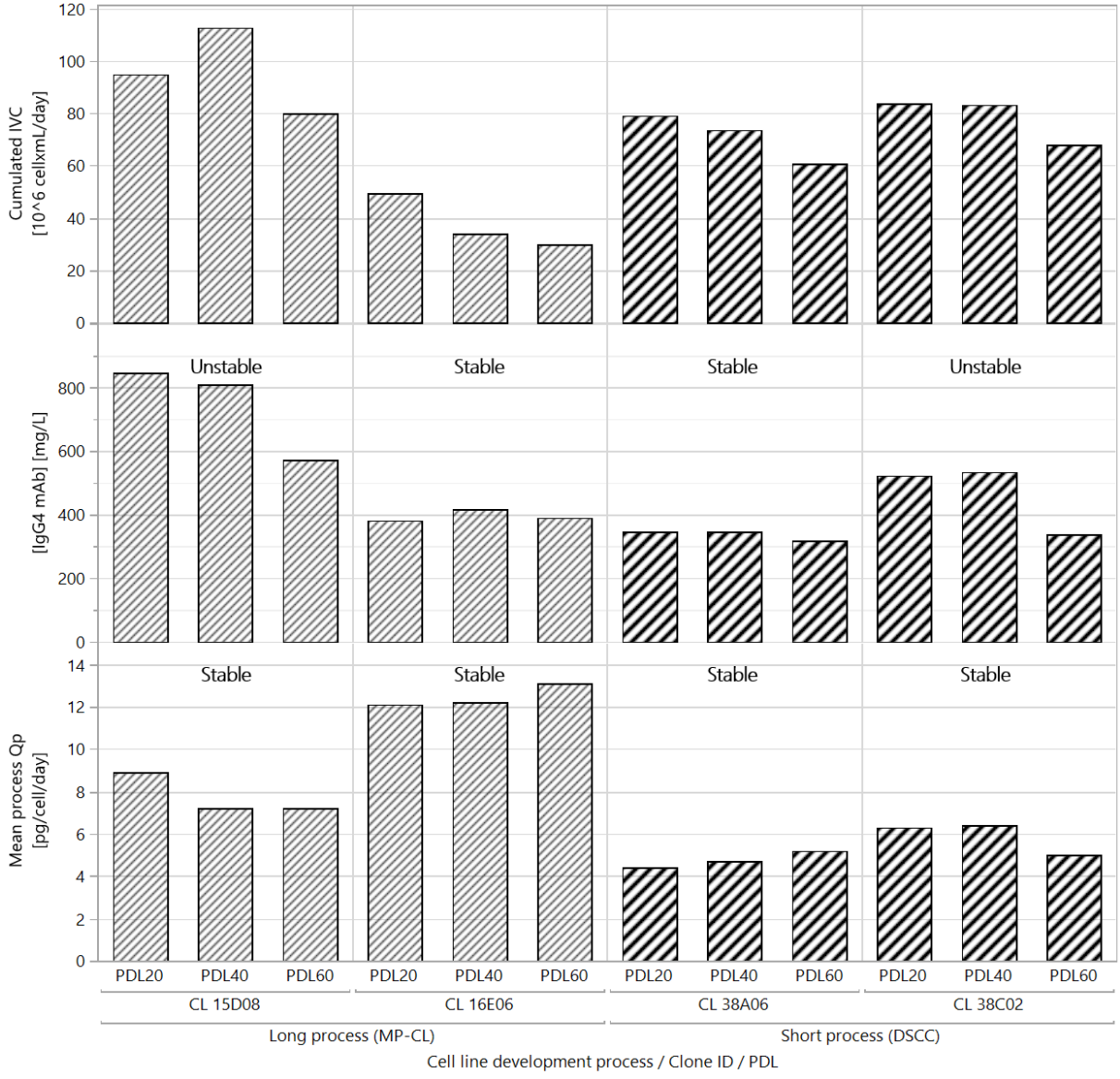


Figure 25: Cumulated IVC, IgG4 concentrations and mean process Q_P after fed-batch productivity assay at PDL_{20}, PDL_{40}, and PDL_{60} depending on clone and cell line development process

Two top clones, respectively from long CLD process and short CLD process conditions, were sub -cultured without selection pressure in culture medium. Three 12-day fed-batch productivity assays were launched in TPP® tubes, respectively at estimated PDL_{20}, PDL_{40}, and PDL_{60}. Values displayed on the graph are the values on the last common day of culture between the three productivity assays. For example, CL 15D08 culture was stopped at PD_{12} (during productivity assay launched at $PDL2_0$ and PDL_{40}) and PD_9 (during productivity assays launched at PDL_{60}). Therefore, cumulated IVC, IgG4 concentration, and mean process specific productivity at PD_9 (last common day of culture between the three productivity assays) is plotted for the three productivity assays. IVC: Integral of viable cell density; PDL: population doubling level; Q_P: specific productivity. Data analysis was performed on JMP® 16 software.

3.2.5 Discussion

Cell line development process associated to CHOZN® CHO K1 platform is a two-step based process, divided in (i) a mini pool (MP) generation and selection step and in (ii) a clone (CL) generation and selection step. The mini pool generation step aims at easing high producing pools identification, while the clone generation step aims at delivering a clonally derived, high producing cell line. Both mini pool and clone generation steps rely on an iterative process for identification and selection of high producers. Applying this cell line development process enables to generate recombinant cell lines within 7.5 months (from DNA to pre-master cell banks (pre-MCB)).

Cell line development is a decisive step in a recombinant therapeutic molecule development, as this activity is responsible for generating the production cell line (Diep et al., 2021), and is on the critical path for advancing a novel therapeutic to first-in-human (FIH) clinical trials (Munro et al., 2017). However, cell line development timelines should be drastically reduced, for two main reasons. Firstly, efforts are currently made to accelerate new medicine approval process, especially for unmet medical needs (Munro et al., 2017). Secondly, CHOZN® CHO K1 platform timelines exceed the standards of most competitors' industrial platforms, which enable cell line development within 3 months.

A shortened cell line development process, called direct single-cell cloning (DSCC) was designed, in order to reduce cell line development timelines. The designed direct single-cell cloning process consisted in: (i) transfecting the host cell line, (ii) putting the transfected bulk pool under selection pressure 24 hours after transfection, (iii) letting the transfecting bulk pool reach the lowest viability, (iv) single-cell cloning the transfected bulk pool by limiting dilution, seven days after transfection, and (v) identifying and selecting best producing clones, following usual iterative selection process.

This second chapter aimed at evaluating the direct single-cell cloning process on several criteria: technical feasibility, productivity and stability of generated cell lines, and cell line development timelines.

Firstly, technical feasibility of direct single-cell cloning process has been evaluated. Profile of CHOZN® CHO K1 cell line during selection has been defined. Selection phase, *i.e.*, the period between transfection and single-cell cloning, should allow the elimination of cells that have not integrated the plasmid (non-producers). Transfected bulk pool viability profile was characterized by a sharp viability decrease within eight days following transfection, reaching a minimum between 5 and 20%, followed by a bulk pool recovery, at the latest three weeks after transfection. This transfected bulk pool viability profile is consistent with previously defined reemergence of CHOZN® CHO K1 mini-pools, nine days after transfection. In addition, similar transfected bulk pool profiles have been reported for other CHO cell lines. CHO $GS^{-/-}$ bulk pool viability decreased until seven days after transfection, reaching a minimum viability between 50 and 60%, followed by a bulk pool recovery 17 days after transfection (Fan et al., 2011). If viability profiles are similar, one can though notice a significant difference in

viability minimum. This suggests a stringent selection applied on CHOZN® CHO K1 transfected bulk pool with puromycin. Another example of selection of a CHO (lineage not being mentioned) bulk pool shows almost identical viability than ours, with bulk pool viability decrease until eight to ten days after transfection, reaching a minimum viability between 10 and 20%, followed by bulk pool recovery at the latest three weeks after transfection (Diep et al., 2021).

Reemergence of CHOZN® CHO K1 clones generated with direct single-cell cloning was observed from 12 days after single-cell cloning (on reemerging clone), but more exactly from 14 days after single-cell cloning. A slightly later reemergence was observed for clones generated with direct single-cell cloning process. However, as for long cell line development process, clones generated with direct single-cell cloning process mostly reemerged within three weeks after single-cell cloning, therefore not impacting defined timelines for clone reemergence in 96-well plates. Significant differences in cloning efficiencies are observed between the two processes, direct single-cell cloning resulting in the lowest cloning efficiency. Both later reemergence and lower cloning efficiency obtained on direct single-cell cloning process may be explained by the very low viability of the transfected bulk pool on the day of cloning (16%) compared to the viability of the mini-pool suspension (98.6%, blend of three healthy, exponentially growing mini-pools). Moreover, it cannot be excluded that transfected bulk pool used for single-cell cloning might not have reached the lowest viability prior single-cell cloning. As the whole volume of cell suspension was used for single-cell cloning, recovery profile of the transfected bulk pool is not known. Indeed, single-cell cloning was performed seven days after transfection, though minimum viability appears to be reached eight days after transfection, as observed on all direct single-cell cloning runs performed afterwards. This incomplete viability decrease prior cloning might have contributed to the lower cloning efficiency.

These two points validate technical feasibility of implementation of direct single-cell cloning process on the CHOZN® CHO K1 platform, but also suggest that there is considerable scope for optimizing the direct single-cell cloning process.

Secondly, impact of direct single-cell cloning on recombinant cell lines' productivity has been evaluated through a clone selection process, including productivity measurement in 96-well plates, batch productivity assay in 24-deepwell plates, and fed-batch productivity assay in TPP® tubes. Even if maximum titer was fairly close between both cell line development processes after fed-batch productivity assays in TPP® tubes and ambr250® bioreactors, one can notice a higher median titer for clones generated with long cell line development process (MP-CL). In the same way, maximum process specific productivity was similar between both cell line development process after the fed-batch productivity assays, but one can notice a higher median process specific productivity for clones generated with long cell line development process (MP-CL). These differences are significant (p-value below $\alpha_{5\%}$) only at the 24-deepwell plate stage. Significance of difference could not be determined at

further scales (namely, TPP® tubes and ambr250® bioreactors) because the assumption of normality of data distribution was not respected.

Recombinant cell lines have been generated within 4.5 months with direct single-cell cloning process, compared to 7.5 months for long cell line development process (MP-CL). This equal to a significant reduction of 40% of time needed for cell line development.

Lastly, expression level stability of two top CHOZN® CHO K1 clones generated with direct single-cell process was investigated over a stability study, and compared to those of clones generated with long cell line development process (MP-CL). This stability study was defined in the manner of a stability study conducted as part of a CDMO project. Fed-batch productivity assays were launched at estimated population doubling level 20, 40 and 60 (PDL_{20}, PDL_{40}, and PDL_{60}) – a study duration defined to span master cell bank (MCB) preparation, working cell bank (WCB) preparation, GMP production, and extended cell bank (ExCB) preparation. Selection pressure was not maintained during sub-culturing process.

Disturbances on cell growth (seen on viability doubling time and population doubling level profiles) observed on clones generated with long cell line development process (MP-CL) in previous chapter were also observed on clones generated with direct single-cell process. CL 38A06 and CL 38C02 cell growth was similar to those observed for CL 15D08. At least 70% of volumetric productivity was retained over 60 PDL for CL 38A06, but a titer decrease of more than 30% between PDL_{20} and PDL_{60} was observed for CL 38C02. Mean process specific productivity was though considered stable for both clones. Same trends have been observed on clones generated with long cell line development process (MP-CL), which suggests that the type of cell line development process applied does not influence recombinant cell line productivity stability. However, the limits of this study are as previously stated (Discussion): disturbance of cell growth might have impacted productivity, and stability study design does not allow to have a complete picture of CHOZN® CHO K1 clones (regardless of cell line development process applied) behavior over long term cultivation and to understand underlying root causes of productivity instability.

Direct single-cell cloning process proved to be efficient for cell line development timelines reduction but led to a decrease of productivity of generated cell lines. Direct single-cell cloning process is often associated with advanced techniques for single-cell sorting and/or seeding in industrial cell line development platforms (Tejwani et al., 2021). For example, implementation of the Beacon (Berkeley Lights) cell sorter and seeder enabled Amgen's cell line development team to reduce by 30% cell line development timelines while generating clones having comparable growth, productivity and product quality than clones generated from standard cell line development process (including an eight-week pool generation and screening step and fluorescence activated cell sorting (FACS) cloning) (Diep et al., 2021). Similarly, use of the MoFlo-XDP (Beckman Coulter) cell sorter and seeder enabled Genentech's

cell line development team to save four weeks in cell line development timelines while generating clones with similar titer and product quality than clones generated from standard cell line development process (including one pool generation and screening step and single-cell cloning by limiting dilution) (Misaghi et al., 2016). Numerous other examples, such as Boehringer Ingelheim GmbH and Merck & Co, Inc. (both with FACSAria™ (Becton Dickinson) cell sorter and seeder), Fujifilm Diosynth Biotechnology (with Cyto-Mine® (Sphere Fluidics) cell sorter and seeder), Biogen, and Astra Zeneca – to name a few) could be detailed (Tejwani et al., 2021). However, in the context of the work presented here, efforts have been orientated towards expression vector improvement.

3.2.6 Conclusion

In this second chapter, a shortened cell line development process, called direct single-cell cloning (DSCC), has been designed and compared to the traditional cell line development process, based on a mini-pool and a clone generation. Cell line development timelines were significantly decreased but clonal cell lines productivity was also considerably reduced. Further direct single-cell process optimization was conducted by Merck End-to-End Solutions' cell line development team, namely on viability threshold for single-cell cloning and supplementation of cloning medium.

Efforts for CHOZN® CHO K1 platform optimization were then focused on the expression vector. In the third chapter, several promoters driving resistance to selection pressure expression will be evaluated to increase clonal cell lines productivities.

3.2.7 Author's contributions

Literature review: Lucie Grindes[1,2] (ORCID ID: 0000-0002-6201-3355) ; **Methodology**: Lucie Grindes; **Investigation**: Lucie Grindes, Marine Amouriq[1] and Amandine Philippat[1] (significant technical help during clone generation), Myriam Eyquard[3] (significant technical help during cultures in ambr®250 bioreactors and clone stability study); **Analysis**: Lucie Grindes; **Writing – Original draft and editing**: Lucie Grindes; **Writing – Review**: Lucie Grindes, Gilles Joucla[2] (ORCID ID: 0000-0003-4192-9628), Wilfrid Dieryck[2] (ORCID ID: 0000-0002-4609-8934), Cyrielle Corbin[1] (ORCID ID: 0000-0002-6891-2511), Sébastien Ribault[4] (trade secret and intellectual property), Daniel Kasten[5] (trade secret and intellectual property); **Supervision**: Cyrielle Corbin.

[1] Cell line development team, Process development department, Merck Biodevelopment

[2] Structure and activity of biological macromolecules team, Nutraceutical, medical and food applications pole, CBMN (UMR 5842 (CNRS, University of Bordeaux, Bordeaux INP))

[3] Up-Stream process team, Process development department, Merck Biodevelopment

[4] Direction, Merck Biodevelopment

[5] Patents Life Science global team, MilliporeSigma

3.3 Expression vector optimization: evaluation of promoters for expression of resistance to puromycin

3.3.1 Introduction

Identification of stable and producing clones in the CHOZN® CHO K1 platform relies on puromycin-based selection. Puromycin-based selection of stably transfected cells is widespread due to the efficiency of selection (high toxicity leading to selection in few days), the small size of *pac* gene (600 bp) and the low cost of puromycin (Aviner, 2020; Caputo et al., 2021). Puromycin is a natural amino nucleoside antibiotic produced by *Streptomyces alboniger* that inhibits protein synthesis in eukaryotic and prokaryotic organisms (Vara et al., 1985). *S. alboniger* also express a puromycin N-acetyltransferase (PAC) that confers resistance to puromycin. First cloned in 1985, *pac* gene was then used as selection marker in transfected mammalian cell lines in 1986 (Aviner, 2020; Vara et al., 1985, 1986).

Degree of selection stringency is an important parameter for high producing clone identification (Bandaranayake & Almo, 2014; Lai et al., 2013; F. M. Wurm, 2004). High degree of selectivity forces the cells to be more productive in the locus of the selection marker in order to survive the selection process (Lai et al., 2013; Ng et al., 2007). Therefore, a weakened selection marker can increase recombinant protein production through different mechanisms. Increasing selectivity can result in better plasmid integration site, or in higher transcriptional or translational activity, or in higher gene copy number (Lai et al., 2013). These compensatory mechanisms lead to an increased expression of the linked gene of interest (GOI), as the selection marker and the GOI are generally co-localized (Westwood et al., 2010). Increasing clone selection stringency therefore leads to more efficient cell line screening processes, by identifying higher producing cell lines with reduced lab work (F. M. Wurm, 2004). Selection stringency can be strengthened by increasing the selection pressure concentration in culture media or by lowering strength of selection marker. The former strategy is rarely applied because of slower cell growth at higher selection pressure concentrations, while the latter has frequently been described in literature, through various strategies at different stages of selection marker expression (Lai et al., 2013).

Promoter driving expression of the resistance gene can either be selected among weak promoters (use of a weak *Herpes simplex* virus thymidine kinase (HSV-TK) promoter (Niwa et al., 1991)) or be weakened by targeting important transcriptional sites (engineered deletions in SV40 promoter (Fan, Kadura, et al., 2013; Hartman & Tso, 2013; Klein & Kopetzki, 2009; Sacco et al., 2022; Yang et al., 2019)). Once resistance gene is transcribed into RNA, stability of mRNA can be decreased through the use of mRNA destabilizing elements, such as AU-rich elements (ARE) (Ng et al., 2007). Protein synthesis can then be hindered using least preferred codons, therefore lowering translational efficiency (Chin et al., 2015; Westwood et al., 2010). Some patents also describe the use of non-ATG start codon

(like GTG or TTG codons) or weakened Internal Ribosome Entry Sites (IRES) to make protein synthesis more difficult, when the resistance gene translation initiation is under the control of an IRES (Chin et al., 2015; Otte et al., 2009). Once mRNA is translated into protein, stability of protein can be decreased through the use of protein destabilizing elements, such as murine ornithine decarboxylase (MODC) PEST sequences (Chin et al., 2015; Ng et al., 2007, 2012). Lastly, attenuated mutants were also frequently described in the literature. In such strategies, wild-type selection marker is replaced by an enzyme with attenuated activity. Impaired mutants were engineered for neomycin phosphotransferase II (L. Chen et al., 2004; Sautter & Enenkel, 2005), glutamine synthetase (Lin et al., 2019) and puromycin N-acetyltransferase enzymes (Caputo et al., 2021).

We focused our work on the promoter driving puromycin N-acetyltransferase transcription, as transcription is traditionally considered as the dominant factor controlling expression (Barnes et al., 2003). In addition, this genetic element was easily modifiable, with regard to the laboratory techniques and the knowledge we had at our disposal. Our objective was to decrease promoter strength and therefore increase clone selection stringency. The results presented hereinafter will be subjected to publication.

3.3.2 Hypotheses on the mechanism of action of an attenuated selection marker

Hypotheses were formulated regarding the mechanism of action of an attenuated selection marker, and how an attenuated selection marker might impact the recombinant protein productivity (Figure 26):

(i) Attenuated selection marker would lead to selection of clones for which plasmid integrated in a naturally transcriptionally highly active zone.

Transcription level is higher in genomic sites referred to as “transcriptional hot-spots”. Therefore, transcription level of selection marker in such genomic *loci* results in a sufficient level of selection marker to survive selection process. Once selection pressure is applied, only clones that have integrated plasmid in genomic *loci* where transcriptional activity is high are capable of surviving selection process. High transcriptional activity at the genomic *loci* of insertion also benefits to the gene of interest, which is co-localized on the expression vector. Thus, it results in higher expression of the recombinant protein (Ng et al., 2007).

(ii) Attenuated selection marker would lead to creation of an artificial transcriptionally highly active zone.

Once selection pressure is applied, only clones that have a better transcriptional mechanism are capable of surviving selection process. Better transcriptional mechanism also benefits to the gene of interest, which is co-localized on the expression vector. Thus, it results in higher expression of the recombinant protein (Ng et al., 2007).

(iii) Attenuated selection marker would lead to selection of clones for which plasmid integrated in multiple copies.

Due to higher plasmid copy number, expression of selection marker is sufficient for cells to survive selection process. Once selection pressure is applied, only clones that have integrated a high number of plasmid copies are capable of surviving selection process. The gene of interest, which is co-localized on the expression vector, is also present in multiple copies. Thus, it results in higher expression of the recombinant protein (Ng et al., 2007).

(iv) Attenuated selection marker would lead to multiplication of plasmid copies.

Once selection pressure is applied, gene amplification occurs. Several molecular mechanisms underlying gene amplification have been suggested, such as unequal sister chromatid exchange, telomeric fusions and bridge-breakage-fusion cycles, centromere recombination, and chromosome breakage-acentric element models, that would result in multiplication of plasmid copies (Omasa, 2002; F. M. Wurm & Petropoulos, 1994; Yoshikawa, Nakanishi, Itami, et al., 2000). Gene amplification also benefits to the gene of interest, which is co-localized on the expression vector. Thus, it results in higher expression of the recombinant protein.

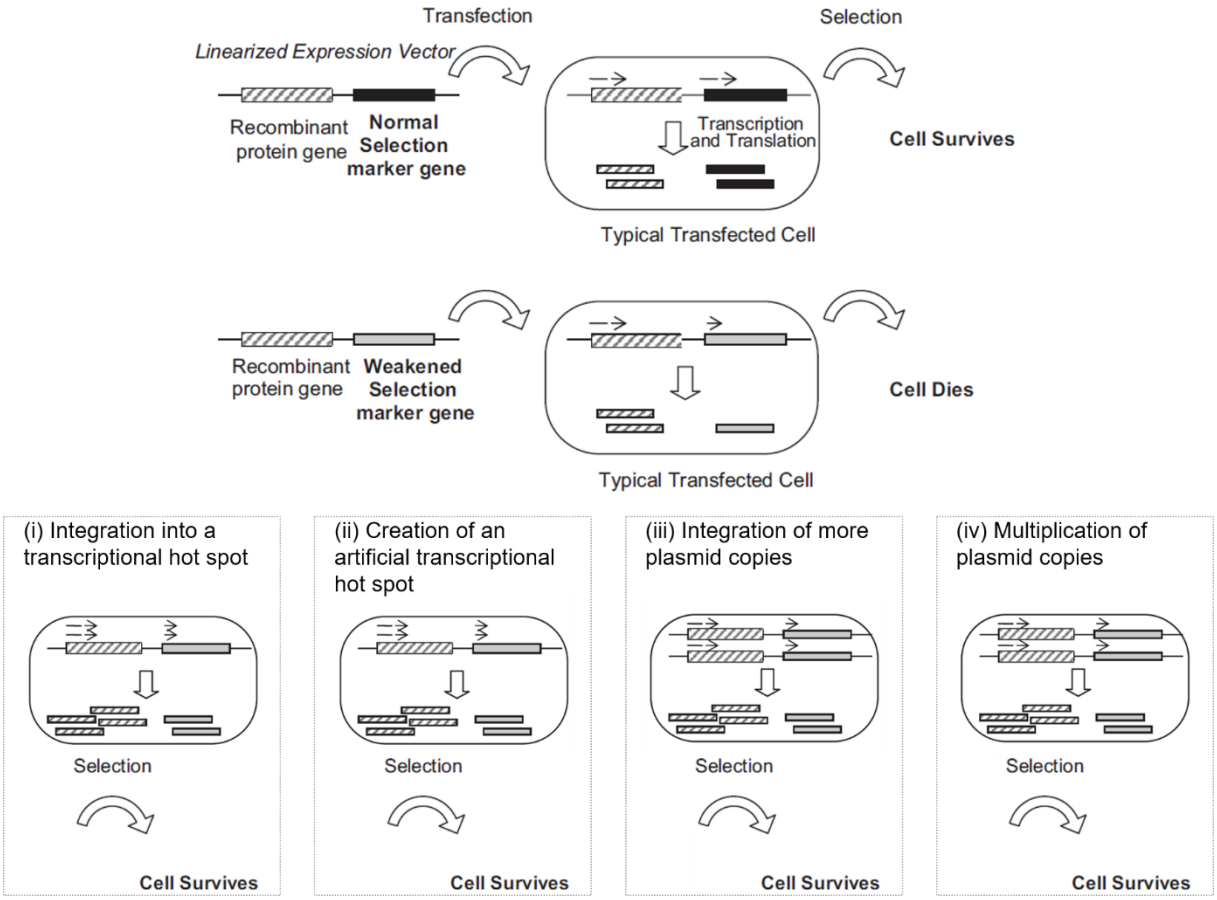


Figure 26: Hypotheses on the mechanism of action of an attenuated selection marker and on the impact on recombinant protein productivity

Hypotheses were formulated regarding the mechanism of action of an attenuated selection marker, and how an attenuated selection marker might impact the recombinant protein productivity. Hypothesis (i):

Attenuated selection marker would lead to selection of clones for which plasmid integrated in a naturally transcriptionally highly active zone. Hypothesis (ii): Attenuated selection marker would lead to creation of an artificial transcriptionally highly active zone. Hypothesis (iii): Attenuated selection marker would lead to selection of clones for which plasmid integrated in multiple copies. Hypothesis (iv): Attenuated selection marker would lead to multiplication of plasmid copies. Adapted from Ng et al., 2007.

3.3.3 Methodology

Until now, expression of puromycin N-acetyltransferase was driven by Simian virus early (SV40E) promoter (both in pEM2V-a and pEM2V-c expression vector versions).

Three CHO K1 endogenous promoters have been identified and four deleted SV40E promoters have been designed. Reduced transcriptional activity of evaluated promoters compared to SV40E wild type promoter was confirmed by measuring *pac* expression level by RT-qPCR during a transient expression assay. Promoters with weakened transcriptional activity have been used to drive PAC expression in vector containing recombinant antibody light chain (LC) and heavy chain (HC) under the control of a classical strong promoter. These vectors were used for cell line generation with CHOZN® CHO K1 host cell line expressing an IgG1 monoclonal antibody (rituximab) under the current cell line development process. Promoters were first screened through a mini-pool generation. Promoters showing an increase of mini-pool productivity were then evaluated through a clone generation with a direct single cell cloning cell line development process. Stability study of expression level of recombinant antibody was performed over 60 Population Doubling Levels (PDL) with plasmid copy number tracking by qPCR and expression level of PAC, LC, and HC measurement by RT-qPCR. Lastly, productivity gain obtained on rituximab with modified promoters was confirmed on several recombinant proteins.

3.3.4 Identification of promoters for *pac* expression

Identification of candidate promoters for *pac* expression relied on two strategies. The first strategy was to reduce the transcriptional activity of SV40E wild type promoter by deleting important transcriptional sites. The second strategy was to substitute SV40E wild type promoter by a CHO K1 endogenous promoter.

3.3.4.1 Modified SV40 promoters: SV40E_Δ72bpRII, SV40E_Δ21bpRIII-II, SV40E_ΔTATA, and SV40E_ΔAT-region

The first strategy was to partially delete SV40E wild type promoter by targeting important transcriptional sites. General architecture of SV40E wild type promoter and of proposed modified SV40E promoters is shown on Figure 27. Briefly, SV40E wild type promoter consists of two 72-base pair repeats, three 21-base pair repeats, a 17-base pair A-T rich sequence containing the TATA box, and few palindromes (Benoist & Chambon, 1980b, 1981; Byrne et al., 1983).

The two 72-bp repeats are described as enhancer sequences (Byrne et al., 1983; Fan, Kadura, et al., 2013; Hartman & Tso, 2013; Klein & Kopetzki, 2009). Repeat II was deleted in candidate promoter

SV40E_Δ72bpRII. The three 21-bp repeats are described as major promoter components (Byrne et al., 1983; Fan, Kadura, et al., 2013). Repeats III and II were deleted in candidate promoter SV40E_Δ21bpRIII-II. Finally, the 17-bp A-T rich sequence containing the TATA box is described as an important element for accurate initiation of transcription (Byrne et al., 1983). The TATA box was deleted in candidate promoter SV40E_ΔTATA and the whole A-T rich sequence was deleted in SV40E_ΔAT-region.

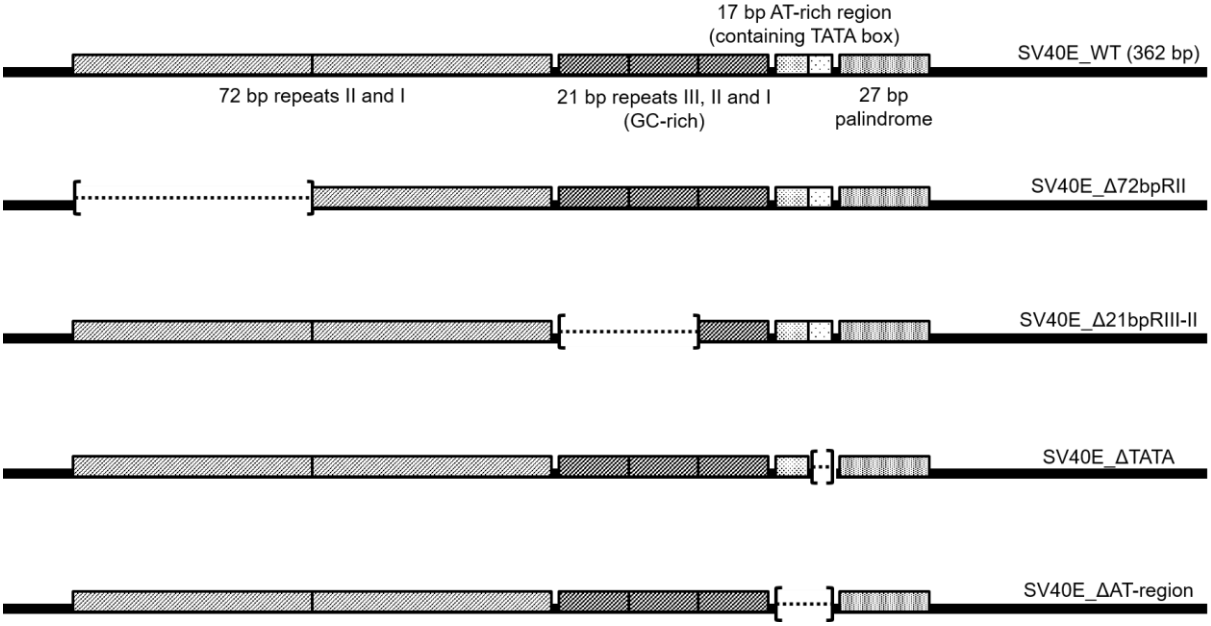


Figure 27: General architecture of SV40E wild type promoter and of modified SV40E promoters

Deletions are displayed as dashed lines in brackets. Adapted from Byrne et al., 1983.

3.3.4.2 CHO K1 endogenous promoters: H2AJ, DRC7, and RAS

The second strategy was to substitute SV40E wild type promoter by a CHO K1 endogenous promoter. Indeed, it is widely considered that mammalian endogenous promoters have lower transcriptional activity than viral promoters (Tossolini et al., 2022). Therefore, mammalian promoters could be used for *pac* expression. This approach, to our knowledge never described before, is innovative and thus distinguish itself from strategies relying on engineered-promoters.

Several CHO K1 endogenous promoters have been previously evaluated, before the beginning of the work presented here, for their inducibility potential in response to a temperature downshift (TDS) for recombinant protein expression: interleukin 33 (*Il33*) (NCBI Gene ID: 100761971), RAS like family 11 member A (*RasI11a*) (NCBI Gene ID: 100758419), dynein regulatory complex subunit 7 (*Drc7*) (NCBI Gene ID: 100764290), and histone H2A.J (*LOC100757684*) (NCBI Gene ID: 100757684). These promoters have been identified by researchers from National Autonomous University of Mexico, investigating the effects of a temperature downshift on the transcriptome (Bedoya-López et al., 2016). Nonetheless, from our unpublished results, they proved to be less efficient than other viral promoters

evaluated within the same study. However, for the purpose of the work presented here, these promoters have been subject to renewed interest.

3.3.5 Results

3.3.5.1 Candidate promoters for *pac* expression have a lower transcriptional strength than SV40E wild type promoter

It was first necessary to confirm the reduced transcriptional activity of candidate promoters, by comparison to current promoter (SV40E). To avoid bias linked to plasmid integration in the genome, which would be due to putatively decreased promoter strength, transcriptional activity was measured over a transient gene expression assay.

Briefly, the expression plasmids carrying candidate promoters for *pac* expression have been constructed, and CHOZN® CHO K1 host cell line was transfected by electroporation with theses plasmids (two independent transfection series, for which three replicates of transfection per plasmid were performed, *i.e.*, six biological replicates per condition). Transfected bulk pools were cultured in T-flasks for four days. At the end of the transient gene expression assay, *pac* transcription level was measured by quantitative PCR (RT-qPCR) (three technical replicates per sample). RT-qPCR method has been developed for this purpose (Reverse transcription – quantitative polymerase chain reaction method development). *pac* transcription level was normalized both to rituximab heavy chain (rituHC) and rituximab light chain (rituLC) transcripts expression levels. To have a single normalization result, it is proposed to normalize *pac* transcription level by the geometric mean of rituHC and rituLC relative quantities. This normalization method is based on the model of normalization by geometric mean of reference genes relative quantities.

Normalized *pac* transcript expression levels, are shown on Figure 28. All candidate promoters for *pac* transcription have on average a lower transcriptional strength than wild type SV40E promoter. CHO K1 endogenous promoters H2AJ and DRC7 exhibit the lowest *pac* transcription levels. RAS shows a higher *pac* transcription level than H2AJ and DRC7, but lower level than all SV40E modified promoters. Furthermore, it is noticeable that SV40E modified promoters differ in transcriptional strength. SV40E_Δ72bpRII enables the highest *pac* transcription level, while SV40E_Δ21bpRIII-II, SV40E_ΔTATA and SV40E_AT-region seem to give similar *pac* transcription levels.

Significance of mean transcription level difference between candidate and SV40E wild type promoters was evaluated by a statistical analysis (details of which can be found in Appendix 3). All candidate promoters show a significantly reduced mean transcriptional activity compared to SV40E wild type promoter (all p-values below $\alpha_{5\%}$). Deleted SV40E promoters (SV40E_Δ72bpRII, SV40E_Δ21bpRIII-II, SV40E_ΔTATA, and SV40E_ΔAT-region) display similar promoter mean transcriptional activities. The same observation is made for CHO K1 endogenous promoters (H2AJ, DRC7 and RAS). Interestingly, SV40E_ΔAT-region and RAS mean transcriptional activities are

similar. All other comparisons (SV40E_Δ72bpRII /SV40E_Δ21bpRIII-II /SV40E_ΔTATA *vs* H2AJ /DRC7 /RAS, and SV40E_ΔAT-region *vs* H2AJ /DRC7) show that promoter have statistically different mean transcriptional activities.

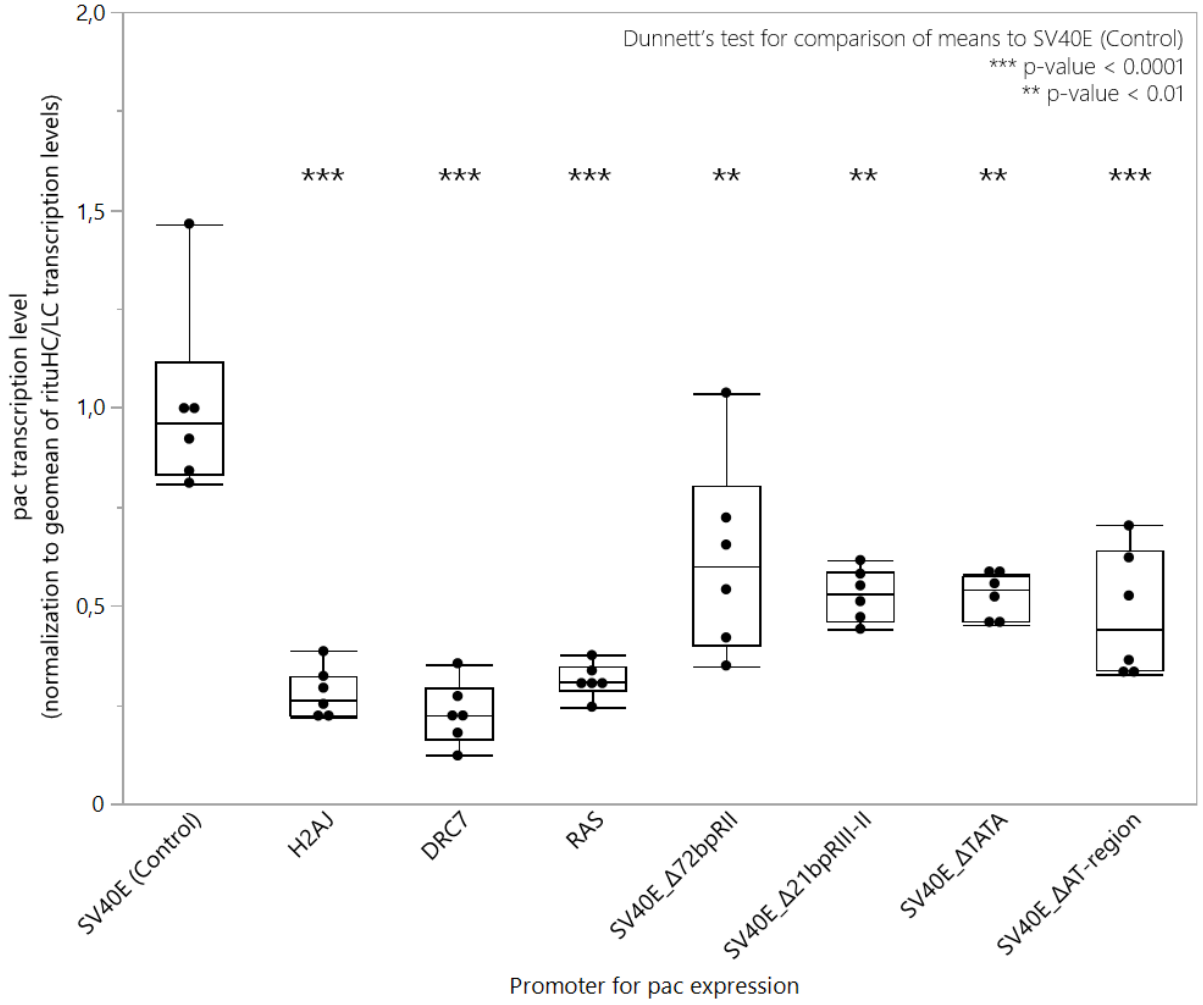


Figure 28: pac transcription level of SV40E wild type promoter and candidate promoters

Box-and-whiskers plot of pac transcription level. Transcriptional activities of SV40E wild type promoter and candidate promoters were measured over a transient transfection assay. pac transcript expression level was measured by quantitative PCR (RT-qPCR) and normalized to geometric mean of rituHC and rituLC relative quantities. Two independent transfection series, for which three replicates of transfection per plasmid were performed, i.e., six biological replicates per condition. Each dot represents biological replicate. Boxplot horizontal lines respectively indicate, top to bottom, the maximum, the upper quartile (Q_3), the median, the lower quartile (Q_1) and the minimum pac transcription level. Significance of transcription level difference between candidate and wild type SV40E promoters was evaluated by a statistical analysis, including assessment of normal distribution, one-way analysis of variance (ANOVA) and Dunnett's test (α 5%). Data analysis was performed JMP® 15 software.

3.3.5.2 Candidate promoters for *pac* expression decrease mini-pool reemergence but increase mini-pool productivity

Reduction of transcriptional strength, compared to SV40E wild type promoter, was previously demonstrated for all candidate promoters for *pac* expression. A first screening through a mini-pool generation was performed to evaluate candidate promoters.

The mini-pool strategy allows to create small pools of stably transfected cells, by dividing the original transfected bulk pool into smaller populations (*i.e.*, mini-pools). This strategy has the advantages to (i) generate higher variation between mini-pools than there would be between pools, therefore resulting in a more accurate evaluation of conditions between each other's; (ii) generate higher quantity of samples per condition than there would be with a clone generation, due to better reemergence of mini-pools compared to clones with equal number of seeded wells, and (iii) generate results in a relatively short time (approximately five to six weeks, from host cell bank thawing to 96-well plate productivity assay).

Briefly, CHOZN® CHO K1 host cell line was transfected by electroporation with plasmids carrying candidate promoters for *pac* expression. One day after transfection, mini-pools were seeded in 96-well plates and put under selection pressure. Mini-pools were cultured in 96-well plates until cellular reemergence was observed; a maximum of three weeks was allowed for reemergence. When the first of these two criteria was met, a seven-day non normalized batch productivity assay was launched in 96-well plates (viable cell density not being controlled at seeding). At the end of the productivity assay, recombinant antibody concentration in cell culture supernatant was measured by bio-layer interferometry using protein A coated sensors. Percentage of reemerging mini-pools and mini-pool productivity were the main criteria of interest for evaluation of candidate promoters for *pac* expression.

3.3.5.2.1 Impact of promoter strengths on mini-pool reemergence

Percentage of reemerging mini-pools in 96-well plates are shown on Figure 29-A. No mini-pool reemergence was observed for CHOZN® CHO K1 cells transfected with plasmids containing H2AJ and DRC7 promoters. Mini-pool reemergence was severely decreased for cells transfected with plasmids containing RAS promoter or partial deletions in SV40 promoter (SV40E_Δ21bpRIII-II and SV40_ΔAT-region). However, mini-pool reemergence of cells transfected with SV40E_ΔAT region plasmid was not steadily observed when similar assay was repeated (data not shown). Lastly, other partial deletions in SV40 promoter (SV40_Δ72bpRII and SV40_ΔTATA) had limited impact on mini-pool recovery, compared to control condition.

Correlation matrix of percentage of reemerging mini-pools 96-well plates and *pac* transcription level is shown on Figure 29-B. Percentage of reemerging mini-pools positively correlates (correlation coefficient above 0.8) with *pac* transcription level. Lower transcriptional activity than SV40E wild type promoter resulted in decreased mini-pool reemergence compared to control condition for all candidate

promoters. For example, lowest transcriptional activities were obtained with H2AJ and DRC7, for which mini-pool reemergence was totally inhibited. Thus, it seems that H2AJ and DRC7 do not enable sufficient expression of *pac* with respect to puromycin concentration (10 µg/mL) used for stable gene expression. Lowest percentages of reemerging mini-pools were obtained with RAS and SV40E_AT-region, for which promoter transcriptional activity was lowest than other SV40-partially deleted promoters. Highest *pac* transcription levels were obtained with SV40E_Δ72bpRII, SV40E_Δ21bpRIII-II, and SV40E_ΔTATA, for which mini-pool reemergence broadly ranged from 23 to 88%.

A

Promoter for pac expression	Percentage of reemerging mini-pools
SV40E (Control)	98%
H2AJ	0%
DRC7	0%
RAS	3%
SV40E_Δ72bpRII	78%
SV40E_Δ21bpRIII-II	23%
SV40E_ΔTATA	88%
SV40E_ΔAT-region	2%

B

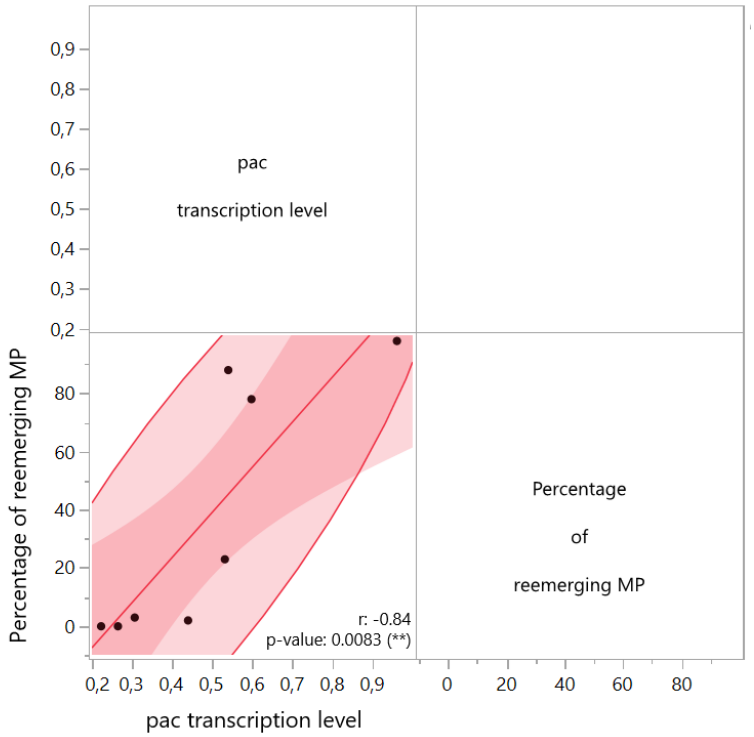


Figure 29: Percentage of reemerging mini-pools and correlation to pac transcription level

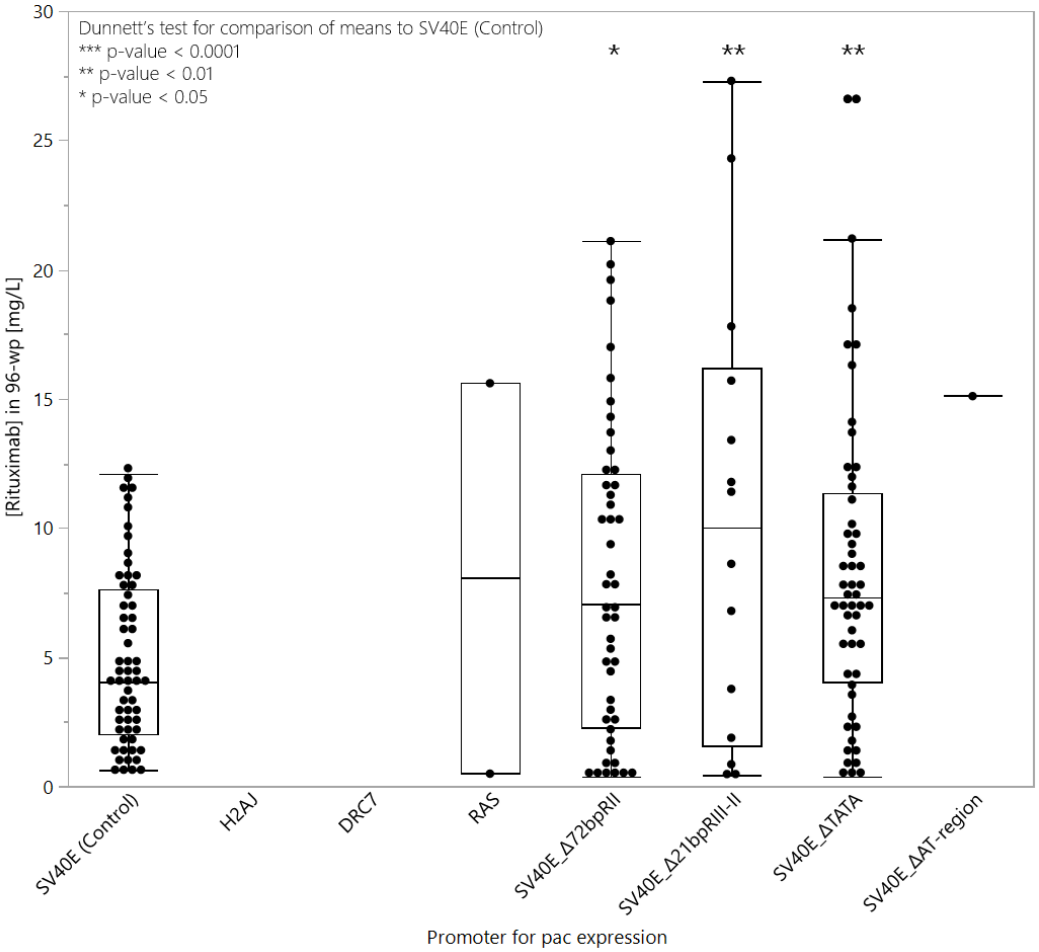
CHOZN® CHO K1 host cell line was transfected by electroporation with plasmids carrying candidate promoters for pac expression. One day after transfection, mini-pools were seeded in 96-well plates and put under selection pressure. Mini-pools were cultured in 96-well plates until cellular reemergence was

observed; a maximum of three weeks has been allowed for reemergence. ***(A)*** *Percentage of reemerging mini-pools is expressed as the ratio of number of mini-pools which have a confluence delta higher than 20% between production day 0 (PD_0) and production day 7 (PD_7) and of total number of seeded mini-pools per condition (one 96-well plate per condition, i.e., 60 mini-pools per condition).* ***(B)*** *Correlation matrix of percentage of reemerging mini-pools and pac transcription level. Data analysis was performed on JMP® 15 software.*

3.3.5.2.2 Impact of promoter strength on mini-pool productivity

Mini-pools' titers (rituximab concentration in supernatant) after batch productivity assay in 96-well plate are shown on Figure 30. Maximum titer is increased in all conditions, compared to SV40E wild type condition. Highest maximum titer is reached with SV40E_Δ21bpRIII-II promoter (27.3 mg/L, *i.e.*, a 2.3-fold increase on maximum titer compared to SV40E wild type promoter). Mean titer is also increased in all conditions, compared to SV40E wild type condition. Highest mean titer is again reached with SV40E_Δ21bpRIII-II promoter (10.3 mg/L, *i.e.*, a 2.1-fold increase on mean titer compared to SV40E wild type promoter).

Significance of mean titer difference between candidate and SV40E wild type promoters was evaluated by a statistical analysis (details of which can be found in Appendix 3). All candidate promoters show a significantly increased mean titer compared to SV40E wild type promoter.



Promoter for pac expression	SV40E (Control)	RAS	SV40E _Δ72bpRII	SV40E _Δ21bpRIII-II	SV40E _ΔTATA	SV40E _ΔAT-region
Max titer [mg/L]	12.1	15.6	21.1	**27.3**	26.6	15.1
Fold increase on max titer	-	1.3	1.7	**2.3**	2.2	1.2
Mean titer [mg/L]	4.9	8.0	7.9	**10.3**	8.3	15.1
Fold increase on mean titer	-	-	1.6	**2.1**	1.7	-

Figure 30: Rituximab concentration in mini-pools' supernatants after batch productivity assay in 96-well plate

Box-and-whiskers plots of rituximab concentration in mini-pools' supernatant after batch productivity assay in 96-well plates. Each dot represents one mini-pool. Data are shown only for growing mini-pools (i.e., ΔPD_7-PD_0 confluence ≥ 20% on 96-well plate productivity assay). Boxplot horizontal lines respectively indicate, top to bottom, the maximum, the upper quartile (Q3), the median, the lower quartile (Q1) and the minimum rituximab concentration. Fold increase on maximum titer is calculated in relation to SV40E (control) maximum titer for all conditions. Fold increase on mean titer is calculated in relation to SV40E (control) maximum titer for all conditions having more than 10 samples. Significance of mean titer difference between candidate and wild type SV40 promoters was evaluated by a statistical analysis, including assessment of normal distribution, one-way analysis of variance (ANOVA) and Dunnett's test (α 5%). This analysis was limited to conditions having more than 10 samples (exclusion of RAS and SV40E_ΔAT-region conditions).

3.3.5.3 Cellular re-emergence with RAS and SV40E_Δ21bpRIII-II promoters can be increased with simple protocol modifications

Increase of mini-pool productivity, compared to SV40E wild type promoter, was previously observed with RAS, SV40_Δ72bpRII, SV40_Δ21bpRIII-II, and SV40_ΔTATA promoters. Nonetheless, it was also demonstrated that mini-pool reemergence was more or less severely decreased with these promoters, due to lower transcriptional activity of the evaluated promoters compared to SV40E wild type promoter. More precisely, RAS and SV40E_Δ21bpRIII-II promoters decreased the most mini-pool reemergence (DRC7 and H2AJ promoters being excluded, as no reemergence was observed), with respectively 3% and 23% of reemergence. Though showing a promising potential for recombinant protein productivity increase, these promoters can only be implemented in next plasmid version if they do not increase cell line development timelines. Therefore, improvement of cellular reemergence should be proven to further evaluation of these promoters for *pac* expression. Two strategies have been investigated:

(i) Modifications on protocol for mini-pool generation. In this case, cell line development process would still rely on two steps (mini-pool generation followed by clone generation). Though being the least preferred option, as cell line development timelines would not be reduced, this strategy still allows productivity increase.

(ii) Modifications on protocol for clone generation. In this case, cell line development process would rely only on one step (clone generation, also referred to as direct single-cell cloning). This was the preferred option, as improvements would target both cell line development timelines (reduced process) and productivity levels (productivity increase).

3.3.5.3.1 Modifications on protocol for mini-pool generation

Three types of modifications on protocol for mini-pool generation have been investigated:

(i) Modification of mini-pool seeding density. We hypothesized that increasing viable cell seeding density would lead to better cellular recovery from selection process. Mini-pools are usually seeded at 5,000 vc/well; proposed seeding densities were 10,000 vc/well, 20,000 vc/well, 40,000 vc/well, and 80,000 vc/well, which was the estimated maximum density for cell confluence monitoring with Cell Metric (Solentim).

(ii) Modification of mini-pool seeding medium. Two approaches, although contradictory, have been evaluated:

  (i) Supplementation of mini-pool seeding medium. We hypothesized that supplementing mini-pool seeding medium would provide additional nutrients, that would improve cellular reemergence. Several transfection and/or cloning supplements have previously been identified, such as Transfectory supplement (Irvine Scientific, 91148-500ML) as transfection supplement for transient expression, ClonaCell™-CHO ACF supplement

(STEMCELL™ Technologies, 03820) as cloning supplement, and InstiGRO™ CHO^{PLUS} (SAL Scientific, RS-1205) as cloning supplement. Although none of these supplements are described for the intended use, they are supposed to increase cell recovery, which is the desired purpose of this study. As the composition of supplements solutions is a trade secret, deeper understanding of mechanisms underlying improved cellular reemergence would not be possible. Mini-seeding medium was either supplemented by 1X ClonaCell™-CHO ACF supplement, or 2.5 to 10% Transfectory supplement (weekly sequential increase). In addition, replacement of EX-CELL® CD CHO Fusion by conditioned medium (as for cloning medium) was evaluated.

(ii) Depletion of mini-pool seeding medium. Is has been incidentally observed that mini-pool seeded in a blend of mini-pool seeding medium and PBS better reemerged than mini-pool seeded only in mini-pool seeding medium. Surprisingly, same kind of observation was reported in a Cell Series 2020 presentation ("Speeding up CLD for the expression of monoclonal antibody and bispecific antibody formats", Moritz B., Ichnos Sciences): dilution of cloning medium with PBS 1:1 proved to significantly increase cloning efficiency. The underlying mechanisms of action remain unclear and mainly raise from incidental observation.

(iii) Modification of electroporation parameters. We hypothesized that softening electroporation parameters would lead to better cellular recovery from transfection. Host cell line is usually transfected at 1,600 V, 10 ms, 3 pulses; proposed transfections parameters were 1,400 V, 20 ms, 2 pulses. This parameter combination results from a screening of electroporation parameters, during the study that led to establishment of CHOZN® CHO K1 cell line development process (work done previous the work presented here). The 1,400 V, 20 ms, 2 pulses condition was the second-best electroporation condition.

Briefly, CHOZN® CHO K1 host cell line was transfected by electroporation with plasmids either carrying RAS or SV40E_Δ21bpRIII-II promoters. One day after transfection, mini-pools were seeded in 96-well plates and put under selection pressure. Mini-pools were cultured in 96-well plates until cellular reemergence was observed; a maximum of three weeks was allowed for reemergence. Percentage of reemerging mini-pools was the main criteria of interest for evaluation protocol modifications.

Percentage of reemerging mini-pools in 96-well plates are shown on Table 7. All evaluated modifications improved mini-pool reemergence, compared to control condition, for both promoters. It should be noticed that percentage of reemerging mini-pools with the control process is higher on this assay than on the initial screening of promoters (as a reminder, 3% and 23% of reemergence respectively for RAS and SV40E_Δ21bpRIII-II promoters) – which corresponds to the variability frequently

observed. Though, the overall trend is the same, with lower reemergence again obtained with RAS promoter, compared to SV40E_Δ21bpRIII-II promoter.

Highest reemergence percentages were obtained with modification of mini-pool seeding density, for both promoters. For RAS promoter, modification of mini-pool seeding medium, either by supplementation or by depletion (except for replacement of EX-CELL® CD CHO Fusion by conditioned medium) improved mini-pool reemergence to a lower extent compared to modification of electroporation parameters. For SV40E_Δ21bpRIII-II promoter, too slight differences are observed to draw a similar conclusion.

Several levers of action are therefore possible to improve mini-pool reemergence, in case direct single-cell cloning process proves to be inapplicable with candidate promoters for *pac* expression.

Table 7: Percentage of reemerging mini-pools

CHOZN® CHO K1 host cell line was transfected by electroporation with plasmids either carrying RAS or SV40E_Δ21bpRIII-II promoters driving pac expression. One day after transfection, mini-pools were seeded in 96-well plates and put under selection pressure. Mini-pools were cultured in 96-well plates until cellular reemergence was observed; a maximum of three weeks has been allowed for reemergence. Percentage of reemerging mini-pools is expressed as the ratio of number of mini-pools which have a confluence delta higher than 20% between production day 0 (PD_0) and production day 7 (PD_7) and of total number of seeded mini-pools per condition (one 96-well plate per condition, i.e., 60 mini-pools per condition).

Process modification	**Percentage of reemerging mini-pools [%]**	
	RAS	**SV40_Δ21bpRIII-II**
Control - Electroporation parameters: 1,600 V, 10 ms, 3 pulses - Mini-pool seeding density: 5,000 vc/well - Mini-pool seeding medium: 80% EX-CELL® CHO Cloning + 20% EX-CELL® CD CHO Fusion + 6 mM L-Gln + 10 µg/mL puromycin	23%	95%
10,000 vc/well	87%	100%
20,000 vc/well	93%	100%
40,000 vc/well	100%	100%
80,000 vc/well	100%	100%
60% EX-CELL® CHO Cloning + 15% EX-CELL® CD CHO Fusion + 25% PBS + 6 mM L-Gln + 10 µg/mL puromycin	47%	97%
40% EX-CELL® CHO Cloning + 10% EX-CELL® CD CHO Fusion + 50% PBS + 6 mM L-Gln + 10 µg/mL puromycin	48%	97%

Process modification	Percentage of reemerging mini-pools [%]	
	RAS	SV40_Δ21bpRIII-II
80% EX-CELL® CHO Cloning + 20% EX-CELL® CD CHO Fusion + 6 mM L-Gln + 10 µg/mL puromycin + 1:40 ClonaCell™-CHO ACF	57%	97%
80% EX-CELL® CHO Cloning + 20% EX-CELL® CD CHO Fusion + 6 mM L-Gln + 10 µg/mL puromycin + 2.5 to 10% Transfectory supplement	52%	98%
80% EX-CELL® CHO Cloning + 20% conditioned medium + 6 mM L-Gln + 10 µg/mL puromycin	68%	98%
1,400 V, 20 ms, 2 pulses	68%	98%

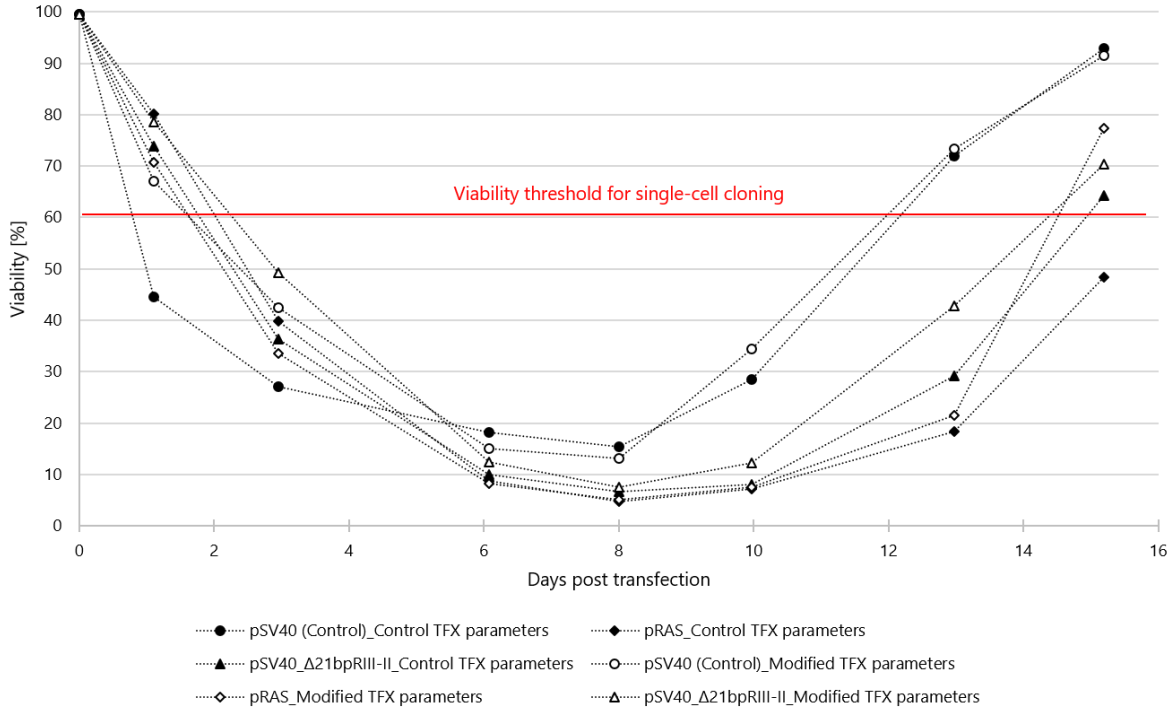
3.3.5.3.2 Modifications on protocol for clone generation

Proposed modification of electroporation parameters, earlier evaluated for mini-pool generation, was taken up for clone generation.

Briefly, CHOZN® CHO K1 host cell line was transfected by electroporation with plasmids either carrying SV40E, RAS or SV40E_Δ21bpRIII-II promoters. One day after transfection, selection pressure was applied to culture medium (puromycin addition), then bulk transfection pool was let to recover. Viability profile during selection and recovery duration were the main criteria of interest for evaluation of electroporation parameters.

Viability profiles during selection (*i.e.*, between transfection and single-cell cloning) are shown on Figure 31. Bulk pools showed on the whole similar profiles. A sharp viability decrease was observed within the first days following transfection and addition of selection pressure, reaching a minimum viability on day 8 (D_8). For both RAS and SV40E_Δ21bpRIII-II, lowest viability was reached for pool generated with control transfection parameters, but very small difference is noticed with modified transfection parameters. Afterwards, viability increases until full pool recovery. Same recovery profiles are observed between SV40E wild type promoter conditions. However, slightly earlier recovery is observed for RAS and SV40E_Δ21bpRIII-II promoters' conditions.

Difficulty of direct single-cell cloning method mainly relies in selection period, after transfection and prior limiting dilution cloning. This assay proved that direct single-cell cloning is technically feasible with RAS and SV40E_Δ21bpRIII-II promoters. As SV40E_Δ72bpRII and SV40E_ΔTATA promoters showed higher transcriptional activities than RAS and SV40E_Δ21bpRIII-II promoters, it was supposed that selection would be more easily achievable.



Promoter for pac expression	SV40E (Control)		RAS		SV40E_Δ72bpRII	
Electroporation parameters	1,600 V, 10 ms, 3 pulses (Control)	1,400 V, 20 ms, 2 pulses	1,600 V, 10 ms, 3 pulses (Control)	1,400 V, 20 ms, 2 pulses	1,600 V, 10 ms, 3 pulses (Control)	1,400 V, 20 ms, 2 pulses
Min viability [%]	15.3 (on D_8)	13.1 (on D_8)	4.7 (on D_8)	5.1 (on D_8)	6.7 (on D_8)	7.5 (on D_8)
Day on which 60% viability threshold is exceeded	Between D_{11}-D_{13}	Between D_{11}-D_{13}	Between D_{12}-D_{14}	Not reached before 15 days post transfection	Between D_{14}-D_{16}	Between D_{14}-D_{16}

Figure 31: Viability profiles during selection depending on electroporation parameters

CHOZN® CHO K1 host cell line was transfected by electroporation with plasmids either carrying SV40E, RAS or SV40E_Δ21bpRIII-II promoters. Two electroporation parameters were used: 1,600 V, 10 ms, 3 pulses (standard parameters) or 1,400 V, 20 ms, 2 pulses (modified parameters). One day after transfection, selection pressure was applied to culture medium (puromycin addition), then bulk transfection pool was let to recover.

3.3.5.4 Candidate promoters for *pac* expression increase clone productivity

Increase of mini-pool productivity, compared to SV40E wild type promoter, was previously demonstrated with RAS, SV40E_Δ72bpRII, SV40E_Δ21bpRIII-II, and SV40E_ΔTATA promoters. Evaluation of candidate promoters for *pac* expression was continued through a clone generation with direct single-cell cloning protocol.

The direct single-cell cloning strategy has the advantage to (i) generate clones using the bulk transfected pool, in which cellular diversity is high, and (ii) be the straightest method to generate clone-linked results (approximately 4 months, from host cell line thawing to fed-batch productivity assay).

Briefly, CHOZN® CHO K1 host cell line was transfected by electroporation with plasmids carrying candidate promoters for *pac* expression. Electroporation parameters were modified (1,400 V, 20 ms, 2 pulses instead of 1,600 V, 10 ms, 3 pulses), based on previous observations (Modifications on protocol for mini-pool generation). Selection pressure was applied to culture medium one day after transfection. Bulk transfection pool was let to recover and single-cell cloning was performed when viability reached the 60% threshold. Single-cell cloning medium was modified (blend of 80% EX-CELL® CHO Cloning medium and 20% of conditioned medium, supplemented with 6 mM L-glutamine, 10 µg/mL puromycin, and 1X ClonaCell™-CHO ACF supplement), based on previous observations (Modifications on protocol for mini-pool generation). Clones were cultured in 96-well plates until cellular reemergence was observed; a maximum of four weeks was allowed for reemergence. When the first of these two criteria was met, recombinant antibody concentration in cell culture supernatant was measured by bio-layer interferometry using protein A coated sensors. Subsequent scale-up was first performed in 24-deepwell plates (sub-culturing and batch productivity assay) and then performed in shake flasks (sub-culturing) and TPP® tubes (fed-batch productivity assay). Clone generation was performed twice for SV40E wild type, RAS, and SV40E_Δ21bpRIII-II promoters but once only for SV40E_Δ72bpRII and SV40E_ΔTATA promoters. Recombinant antibody concentration in supernatant and clone's specific productivity were the main criteria of interest for evaluation of candidate promoters for *pac* expression.

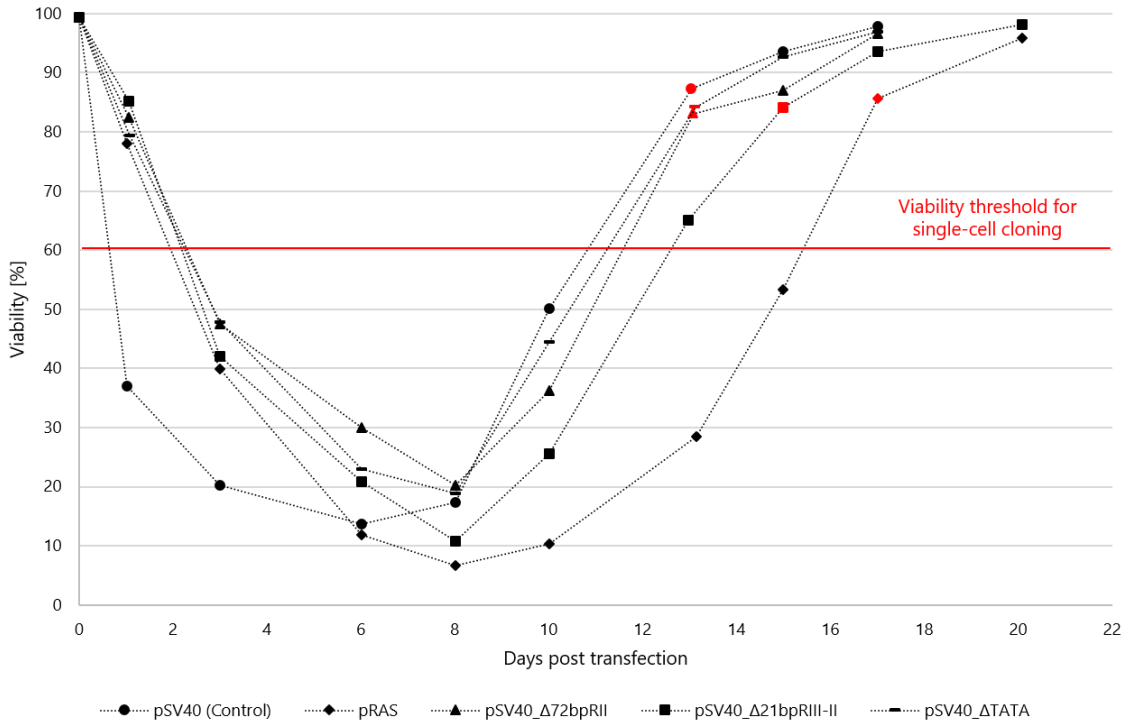
Additionally, to better understand mechanisms underlying productivity increase with candidate promoters for *pac* expression, plasmid copy number determination was planned on generated clones. qPCR method was developed for this purpose (Reverse transcription – quantitative polymerase chain reaction method development). Unfortunately, determination plasmid copy number on generated clones could not be achieved before the writing of this manuscript.

3.3.5.4.1 Impact of promoter strength on cloning efficiency

Viability profiles during selection (*i.e.*, between transfection and single-cell cloning) are shown on Figure 32. Transfected bulk pools generated with plasmids containing the five different promoters showed, on the whole, similar profiles. A sharp viability decrease was observed within the first days following transfection and addition of selection pressure, reaching a minimum viability on day 8 (D_8). Lowest minimum viability was measured for bulk pool generated with RAS promoter (6.7%); highest minimum viability was measured for bulk pool generated with SV40E_Δ72bpRII (20.3%). It is noticed that minimum bulk pool viability well correlates with promoter transcriptional strength, except for SV40E wild type promoter, whose viability profile already distinguished it from other conditions in the early days of selection. Afterwards, viability increases, and full pool recovery is reached on D_{20} at the latest. Earliest recovery is observed for pool generated with SV40E wild type promoter, which was then followed, in order of recovery, by SV40E_Δ72bpRII and SV40E_ΔTATA, SV40E_Δ21bpRIII-II, and

finally RAS. Correlation matrix of recovery duration and *pac* transcription level is shown on Figure 33. Recovery duration negatively correlates (correlation coefficient above 0.8 (in absolute value)) with *pac* transcription level. This observation corroborates previous results (positive correlation between percentage of reemerging mini-pools and *pac* transcription level).

Single-cell cloning is performed once the 60% viability threshold is exceeded. Corresponding time point is highlighted in red on the figure. All pools were single-cell cloned around 80-90% viability. However, it might be noticed that bulk pool generated with SV40E_Δ21bpRIII-II promoter was not single-cell cloned on D_{13} (i.e., after the viability criterion has been exceeded), but rather on D_{15} (when pool viability was around 80-90% viability). This decision was taken to ensure as much as possible equivalent pool viabilities on the day of limiting dilution cloning, as it has been shown that pool viability reflects cellular diversity within the pool (low viability would correspond to high cellular diversity, while loss of diversity would have occurred once recovery is achieved) (Diep et al., 2021).



Promoter for pac expression	SV40E (Control)	RAS	SV40E_Δ72bpRII	SV40E_Δ21bpRIII-II	SV40E_ΔTATA
Min viability [%]	13.8 (on D_6)	6.7 (on D_8)	20.3 (on D_8)	10.8 (on D_8)	18.8 (on D_8)
Day on which 60% viability threshold is exceeded	Between D_{10}-D_{12}	Between D_{14}-D_{16}	Between D_{12}-D_{14}	Between D_{10}-D_{12}	Between D_{10}-D_{12}
Cloning efficiency [%]	36	29	30	31	34

Figure 32: Viability profiles during selection and cloning efficiencies

CHOZN® CHO K1 host cell line was transfected by electroporation with plasmids carrying candidate promoters for pac expression. Electroporation parameters were modified (1,400 V, 20 ms, 2 pulses instead of 1,600 V, 10 ms, 3 pulses). Selection pressure was applied to culture medium one day after

transfection. Bulk transfection pool was let to recover and single-cell cloning was performed when viability reached the 60% threshold. Corresponding time point is highlighted in red. Cloning efficiency was calculated as the ratio of number of reemerging clones by theoretical number of clones. Theoretical number of clones is determined using the Poisson distribution law.

Cloning efficiencies ranged from 29% (RAS promoter) to 36% (SV40E wild type promoter). Correlation matrix of cloning efficiencies and *pac* transcription level is shown on Figure 33**.** Cloning efficiency positively correlates (correlation coefficient above 0.8) with *pac* transcription level. Again, this observation corroborates with previous results (positive correlation between percentage of reemerging mini-pools and *pac* transcription level and negative correlation between bulk pool recovery duration and *pac* transcription level).

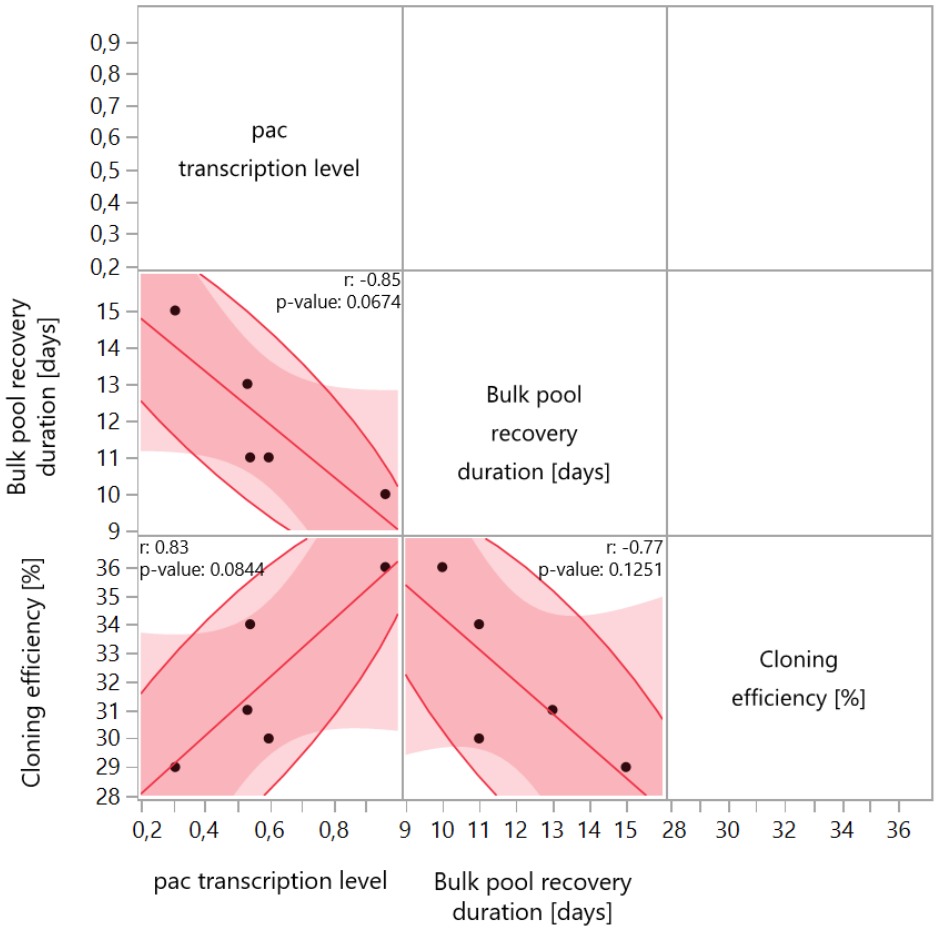


Figure 33: Correlation matrix between promoter transcriptional activities and selection and cloning results

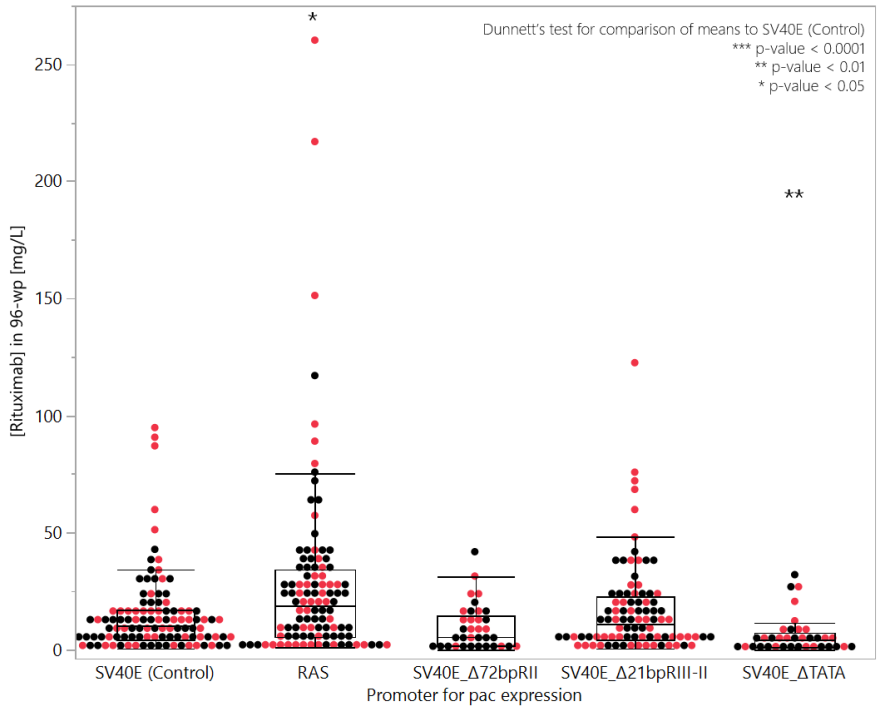
Correlations between bulk pool recovery duration and pac transcription level, cloning efficiency and pac transcription level, and cloning efficiency and bulk pool recovery duration. Correlations are described by two numbers: the correlation coefficient (noted r) and the probability value (noted p-value). Data analysis was performed on JMP® 15 software.

3.3.5.4.2 Impact of weakened promoters on clone productivity

3.3.5.4.2.1 Clone productivity (titer) in 96-well plates

Clones' titers (rituximab concentration in supernatant) after productivity measurement in 96-well plate are shown on Figure 34-A. Maximum and median titers are both only increased in RAS condition, with highest maximum titer reaching 260.2 mg/L (*i.e.*, a 2.0-fold increase on maximum titer compared to SV40E wild type promoter) and highest median titer reaching 19.0 mg/L (*i.e.*, a 1.8-fold increase on maximum titer compared to SV40E wild type promoter). Clone's titers of SV40E_Δ21bpRIII-II condition are equivalent to control condition. However, clone's titers of SV40E_Δ72bpRII and SV40E_ΔTATA conditions are lower than control condition.

Significance of mean titer difference between candidate and SV40E wild type promoters was evaluated by a statistical analysis (details of which can found in Appendix 3). Only RAS promoter shows a significantly increased mean titer compared to SV40E wild type promoter.



Promoter for pac expression	SV40E (Control)	RAS	SV40E_Δ72bpRII	SV40E_Δ21bpRIII-II	SV40_ΔTATA
Max titer [mg/L]	130.0	260.2	41.8	122.5	32.0
Fold increase on max titer	-	2.0	0.3	0.9	0.2
Median titer [mg/L]	10.8	19.0	5.4	11.1	4.1
Fold increase on median titer	-	1.8	0.5	1.0	0.4

Figure 34: Rituximab concentration in clone's supernatants after productivity measurement in 96-well plate

Box-and-whiskers plot of rituximab concentration in clone's supernatants after productivity measurement in 96-well plate. Each dot represents one clone. Boxplot horizontal lines respectively

indicate, top to bottom, the maximum, the upper quartile (Q_3), the median, the lower quartile (Q_1) and the minimum rituximab concentration. Dots outside box-and-whiskers plot are considered as outliers. A point is an outlier if it exceeds the upper outlier limit ($Q_3+1.5\times IQR$) or is less than the lower outlier limit ($Q_1-1.5\times IQR$), where $IQR = Q_3-Q_1$. Fold increase on maximum (or median) titer is calculated in relation to SV40E (control) maximum (or median) titer for all conditions. Clones selected for scale-up in 24-deepwell plates are highlighted in red. Significance of mean titer difference between candidate and wild type SV40 promoters was evaluated by a statistical analysis, including assessment of normal distribution, one-way analysis of variance (ANOVA) and Dunnett's test (α 5%). Data analysis was performed on JMP® 15 software.

3.3.5.4.2.2 Clone productivity (titer and specific productivity) in 24-deepwell plates

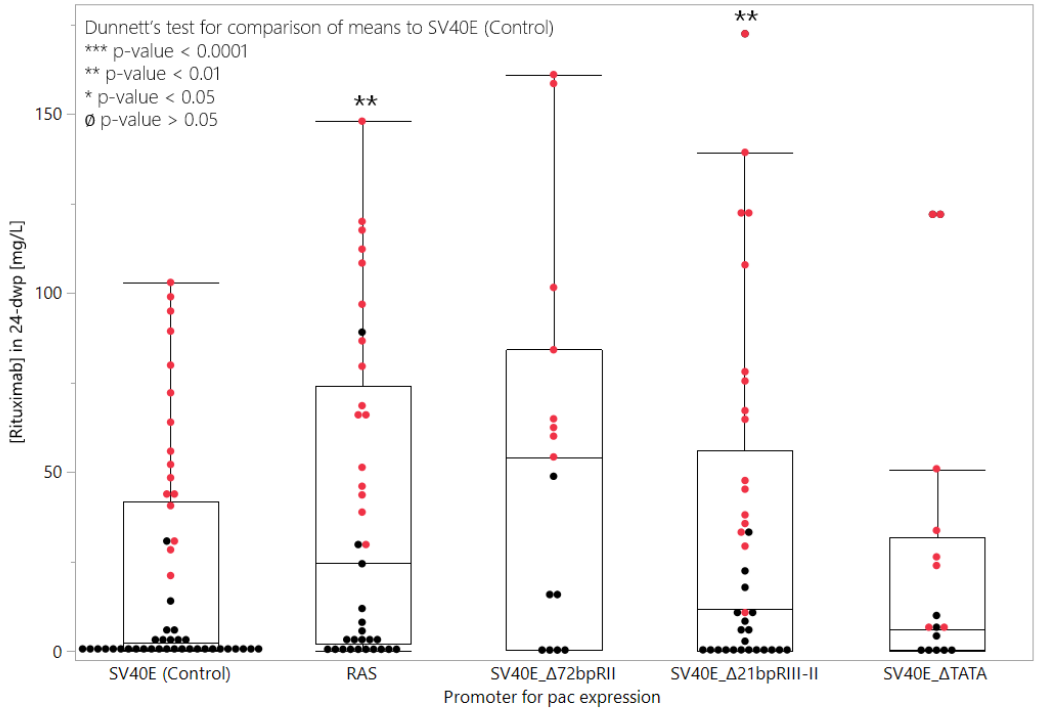
Clones were then selected for sub-culturing and productivity assay in 24-deepwell plates. Selection for scale-up was usually performed by ranking clones according to recombinant protein concentration and 100 top clones were selected for further evaluation. Nonetheless, this selection proved to be inefficient for identification of high producers, as shown by poor correlation factor between titers in 96-well plates and titers in TPP® tubes. Therefore, clones were randomly selected, as follows: after ranking according to recombinant protein concentration, six best producers, twelve medium producers and six worst producers were selected for expansion. That way, 24 clones per condition were randomly selected for sub-culturing and productivity assay in 24-deepwell plates (these clones are highlighted in red on Figure 35 and Figure 36).

Clones' titers after productivity assay in 24-deepwell plate are shown on Figure 35. Maximum titer is increased in all conditions, compared to SV40E wild type condition. Highest maximum titer is reached with SV40E_Δ21bpRIII-II promoter (172.2 mg/L, *i.e.*, a 1.7-fold increase on maximum titer compared to SV40E wild type promoter). Median titer is also increased in all conditions, compared to SV40E wild type condition. Highest mean titer is reached with SV40E_Δ72bpRII promoter (54.1 mg/L, *i.e.*, a 22.5-fold increase on mean titer compared to SV40E wild type promoter).

In addition, 24-deepwell plate culture format allows for determination of specific productivity. Specific productivity is defined as the ratio of rituximab concentration normalized by the quantity of accumulated viable cells. Therefore, specific productivity expresses the quantity of recombinant protein that one cell can produce in one day. Clones' specific productivities after productivity assay in 24-deepwell plate are shown on Figure 36. Maximum specific productivity is increased in SV40E_Δ72bpRII and SV40E_Δ21bpRIII-II conditions, compared to SV40E wild type condition. Highest specific productivity is reached with SV40E_Δ21bpRIII-II promoter (39.5 pg/cell/day, *i.e.*, a 2.1-fold increase on maximum specific productivity compared to SV40E wild type promoter). Mean specific productivity is increased in all conditions compared to SV40E wild type condition. Highest mean specific productivity is reached with SV40E_Δ72bpRII promoter (5.1 pg/cell/day, *i.e.*, a 25.5-fold increase on maximum specific productivity compared to SV40E wild type promoter).

Significance of mean titer and specific productivity differences in 24-deepwell plates between candidate and SV40E wild type promoters was evaluated by a statistical analysis (details of which can be found

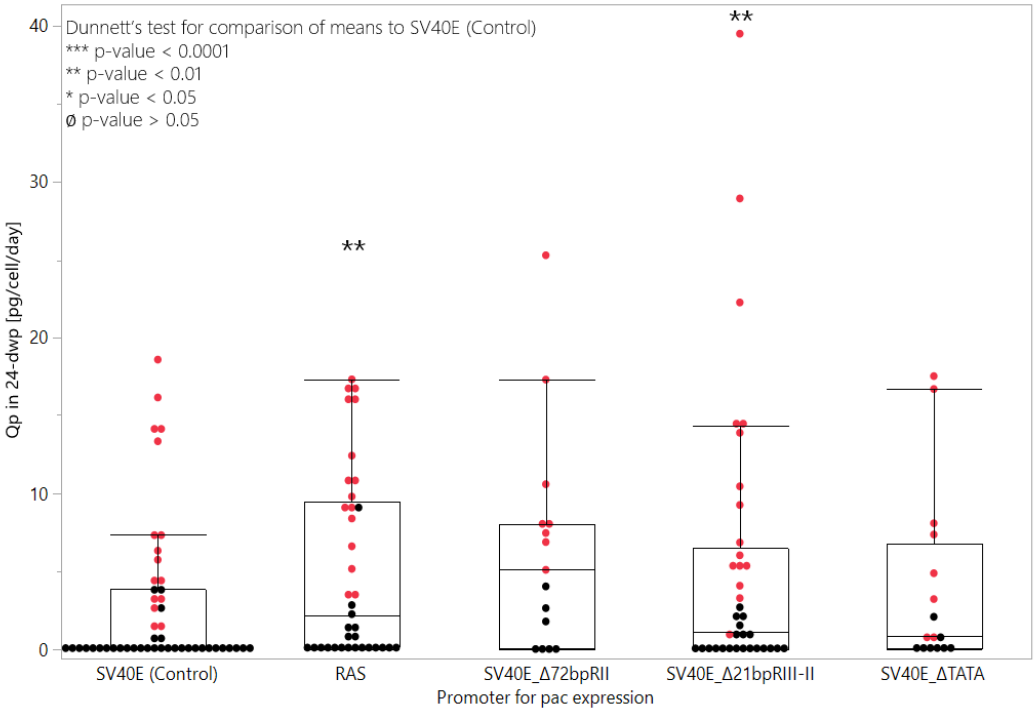
in Appendix 3). RAS and SV40E_Δ21bpRIII-II promoters show a significantly increased mean titer and mean process specific productivity compared to SV40E wild type promoter.



Promoter for pac expression	SV40E (Control)	RAS	SV40E_Δ72bpRII	SV40E_Δ21bpRIII-II	SV40E_ΔTATA
Max titer [mg/L]	102.8	147.8	160.8	**172.2**	122.7
Fold increase on max titer	-	1.4	1.6	**1.7**	1.2
Median titer [mg/L]	2.4	24.3	**54.1**	11.4	5.8
Fold increase on median titer	-	10.1	**22.5**	4.8	2.4

Figure 35: Rituximab concentration in clone's supernatants after productivity assay in 24-deepwell plates

Box-and-whiskers plot of rituximab concentration in clone's supernatants after productivity assay in TPP® tubes. Each dot represents one clone. Boxplot horizontal lines respectively indicate, top to bottom, the maximum, the upper quartile (Q_3), the median, the lower quartile (Q_1) and the minimum rituximab concentration. Dots outside box-and-whiskers plot are considered as outliers. A point is an outlier if it exceeds the upper outlier limit ($Q_3+1.5\times IQR$) or is less than the lower outlier limit ($Q_1-1.5\times IQR$), where $IQR = Q_3-Q_1$. Fold increase on maximum (or median) titer is calculated in relation to SV40E (control) maximum (or median) titer for all conditions. Clones selected for expansion in Erlenmeyer flasks and productivity assay in TPP® tubes are highlighted in red. Significance of mean titer difference between candidate and wild type SV40 promoters was evaluated by a statistical analysis, including assessment of normal distribution, one-way analysis of variance (ANOVA) and Dunnett's test (α 5%). Data analysis was performed on JMP® 15 software.



Promoter for pac expression	SV40E (Control)	RAS	SV40E_Δ72bpRII	SV40E_Δ21bpRIII-II	SV40E_ΔTATA
Max Q_p [pg/cell/day]	18.6	17.2	25.3	**39.5**	17.5
Fold increase on max Q_p	-	0.9	1.4	**2.1**	0.9
Median Q_p [pg/cell/day]	0.2	2.3	**5.1**	1.2	0.9
Fold increase on median Q_p	-	11.5	**25.5**	6	4.5

Figure 36: Specific productivities after productivity assay in 24-deepwell plates

Box-and-whiskers plot of clone's specific productivities (Q_P) after productivity assay in TPP® tubes. Each dot represents one clone. Boxplot horizontal lines respectively indicate, top to bottom, the maximum, the upper quartile (Q_3), the median, the lower quartile (Q_1) and the minimum Q_P. Dots outside box-and-whiskers plot are considered as outliers. A point is an outlier if it exceeds the upper outlier limit (Q_3+1.5×IQR) or is less than the lower outlier limit (Q_1-1.5×IQR), where IQR = Q_3-Q_1. Fold increase on maximum (or median) titer is calculated in relation to SV40E (control) maximum (or median) Q_P for all conditions. Clones selected for expansion in Erlenmeyer flasks and productivity assay in TPP® tubes are highlighted in red. Significance of mean Q_P difference between candidate and wild type SV40 promoters was evaluated by a statistical analysis, including one-way analysis of variance (ANOVA) and Dunnett's test (α 5%). Data analysis was performed on JMP® 15 software.

3.3.5.4.2.3 Clone productivity (titer and specific productivity) in TPP® tubes

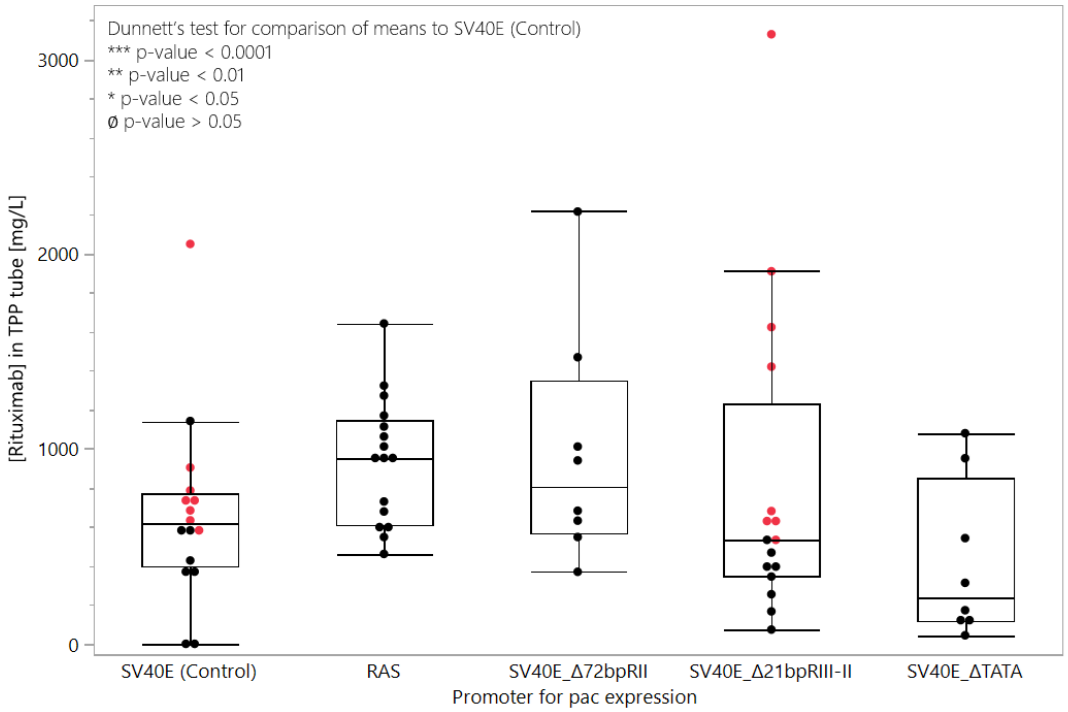
Clones were then selected for sub-culturing in Erlenmeyer flasks and productivity assay in TPP® tubes. Implementation of 24-deepwell plate culture format instead of the 6-well plate culture format proved to improve correlation factor between titers in mid-process evaluation and titers in TPP® tubes. Therefore, a relatively reliable selection can be made from 24-deepwell plate based on titers and specific productivities. Clones were selected as follows: after ranking according to recombinant protein

concentration, top 8 clones per condition were selected for next steps. If two clones displayed similar titers, specific productivity was considered as additional criteria of choice. That way, 8 clones per condition were selected for next step (these clones are highlighted in red on Figure 37 (titers) and on Figure 38 (specific productivities)).

Clones' titers after productivity assay in TPP® tubes are shown on Figure 37. Maximum titer is increased in SV40E_Δ72bpRII and SV40E_Δ21bpRIII-II conditions, compared to SV40E wild type condition. Highest maximum titer is reached with SV40E_Δ21bpRIII-II promoter (3127 mg/L, *i.e.*, a 1.5-fold increase on maximum titer compared to SV40E wild type promoter). Mean titer is increased in RAS and SV40E_Δ72bpRII conditions, compared to SV40E wild type condition. Highest mean titer is reached with RAS promoter (956 mg/L, *i.e.*, a 1.5-fold increase on mean titer compared to SV40E wild type promoter).

Clones' specific productivities after productivity assay in TPP® tubes are shown on Figure 38. Maximum specific productivity is increased in SV40E_Δ21bpRIII-II condition, compared to SV40E wild type condition (17.7 pg/cell/day, *i.e.*, a 1.3-fold increase on maximum specific productivity compared to SV40E wild type promoter). Mean specific productivity is increased in RAS, SV40E_Δ72bpRII, and SV40E_Δ21bpRIII-II conditions compared to SV40E wild type condition. Highest mean specific productivity is reached with SV40E_Δ21bpRIII-II promoter (6.5 pg/cell/day, *i.e.*, a 2.2-fold increase on maximum specific productivity compared to SV40E wild type promoter).

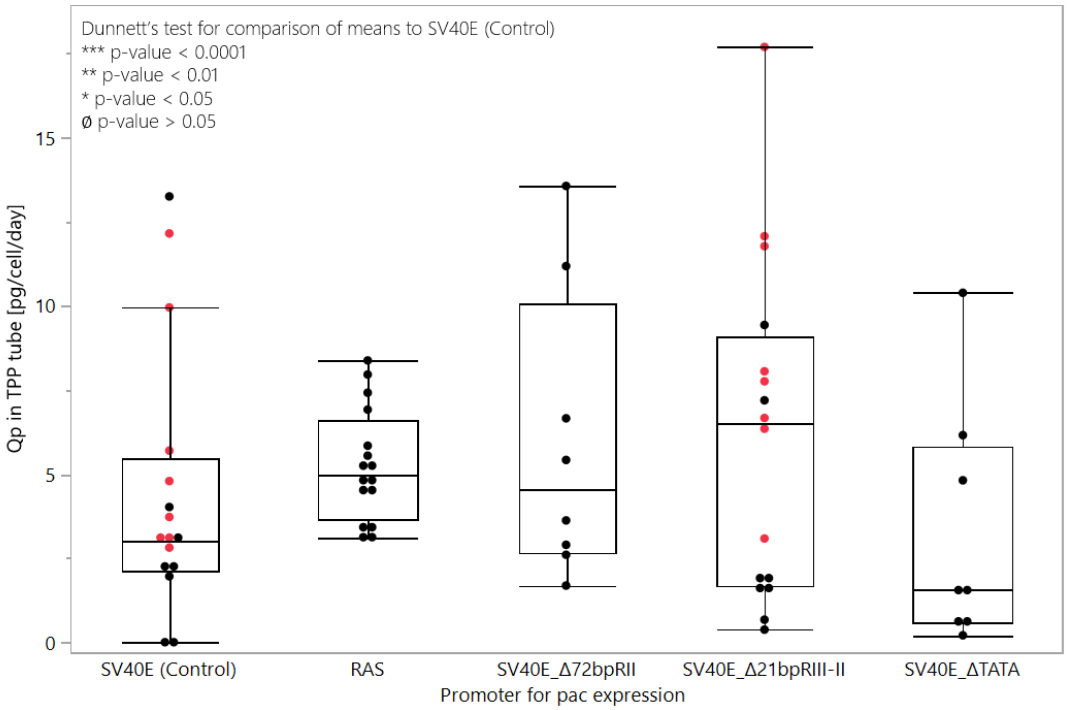
Significance of mean titer and specific productivity differences in TPP® tubes between candidate and SV40E wild type promoters was evaluated by a statistical analysis (details of which can be found in Appendix 3). No condition leads to a significant increase of mean titer nor of mean process specific productivity.



Promoter for pac expression	SV40E (Control)	RAS	SV40E_Δ72bpRII	SV40E_Δ21bpRIII-II	SV40E_ΔTATA
Max titer [mg/L]	2052	1644	2218	**3127**	1082
Fold increase on max titer	-	0.8	1.1	**1.5**	0.5
Median titer [mg/L]	618	**956**	808	535	236
Fold increase on median titer	-	**1.5**	1.3	0.8	0.4

Figure 37: Rituximab concentration in clone's supernatants after productivity assay in TPP® tubes

Box-and-whiskers plot of rituximab concentration in clone's supernatants after productivity assay in TPP® tubes. Each dot represents one clone. Boxplot horizontal lines respectively indicate, top to bottom, the maximum, the upper quartile (Q_3), the median, the lower quartile (Q_1) and the minimum rituximab concentration. Dots outside box-and-whiskers plot are considered as outliers. A point is an outlier if it exceeds the upper outlier limit ($Q_3+1.5\times IQR$) or is less than the lower outlier limit ($Q_1-1.5\times IQR$), where $IQR = Q_3-Q_1$. Fold increase on maximum (or median) titer is calculated in relation to SV40E (control) maximum (or median) titer for all conditions. Clones selected for expression level stability study are highlighted in red. Significance of mean titer difference between candidate and wild type SV40 promoters was evaluated by a statistical analysis, including assessment of normal distribution, one-way analysis of variance (ANOVA) and Dunnett's test (α 5%). Data analysis was performed on JMP® 15 software.



Promoter for pac expression	SV40E (Control)	RAS	SV40E_Δ72bpRII	SV40E_Δ21bpRIII-II	SV40E_ΔTATA
Max Q_p [pg/cell/day]	13.3	8.4	13.6	**17.7**	10.4
Fold increase on max Q_p	-	0.6	1.0	**1.3**	0.8
Median Q_p [pg/cell/day]	3.0	5.0	4.5	**6.5**	1.6
Fold increase on median Q_p	-	1.7	1.5	**2.2**	0.5

Figure 38 : Rituximab concentration in clone's supernatants and specific productivities after productivity assay in TPP® tubes and statistical comparison of group means

Box-and-whiskers plot of clone's specific productivities (Q_P) after productivity assay in TPP® tubes. Each dot represents one clone. Boxplot horizontal lines respectively indicate, top to bottom, the maximum, the upper quartile (Q_3), the median, the lower quartile (Q_1) and the minimum Q_P. Dots outside box-and-whiskers plot are considered as outliers. A point is an outlier if it exceeds the upper outlier limit (Q_3+1.5×IQR) or is less than the lower outlier limit (Q_1-1.5×IQR), where IQR = Q_3-Q_1. Fold increase on maximum (or median) titer is calculated in relation to SV40E (control) maximum (or median) Q_P for all conditions. Clones selected for expression level stability study are highlighted in red. Significance of mean Q_P difference between all candidate promoters was evaluated by a statistical analysis, including assessment of normal distribution, one-way analysis of variance (ANOVA) (α 5%). Data analysis was performed on JMP® 15 software.

Synthesis of best conditions, considering maximum and mean titer, as well as maximum and mean specific productivity is given in Table 8. SV40E_Δ21bpRIII-II promoter appears to be the best condition for clone productivity increase.

Table 8: Synthesis of best conditions for highest maximum and mean titer and highest maximum and mean specific productivity

Criteria	Best condition
Max titer	SV40E_Δ21bpRIII-II
Mean titer	RAS without statistically significant difference at TPP® tube stage (ANOVA) with statistically significant difference at 24-deepwell plate stage (Dunnett's test)
Max Q_p	SV40E_Δ21bpRIII-II
Mean Q_p	SV40E_Δ21bpRIII-II without statistically significant difference at TPP® tube stage (ANOVA) with statistically significant difference at 24-deepwell plate stage (Dunnett's test)

3.3.5.5 Expression level stability over 60 PDL is similar between SV40 and SV40E_Δ21bpRIII-II promoters

Mini-pool productivity increase and, above all, clone productivity increase were previously observed with several candidate promoters for *pac* expression. More precisely, SV40E_Δ21bpRIII-II promoter was identified as the best promoter for increasing clone maximum titer, as well as maximum and mean specific productivity. Besides, it has been shown that about 50% of CHOZN® CHO K1 clones maintain a stable expression level once selection pressure is removed from sub-culturing process (CHOZN® CHO K1 clones' expression level stability). Since the proposed improvement of the expression vector concerns the expression of resistance to selection pressure, it cannot be excluded that recombinant protein expression level is impacted by removal of selection pressure.

Several hypotheses were initially formulated regarding the mechanism of action of an attenuated selection marker (Hypotheses on the mechanism of action of an attenuated selection marker). For each hypothesis, assumptions were formulated regarding the impact on plasmid copy number and/or transcription level, and subsequently on expression level stability, upon selection pressure withdrawal. These new hypotheses are summarized in Table 9. According to these hypotheses, expression level stability may be impacted by weaning from selection pressure. Therefore, assessment of expression level stability of clones generated with SV40E wild type promoter and SV40E_Δ21bpRIII-II promoter, with and without selection pressure in culture medium, was required to continue evaluating SV40E_Δ21bpRIII-II promoter performance.

Furthermore, it is known that productivity instability finds causes on genomic, transcriptomic, and proteomic levels (Dahodwala & Lee, 2019). Some root causes can easily be identified, such as loss of plasmid copy number or variation of heavy and light chains transcription levels. Monitoring of plasmid copy number and heavy and light chains transcription levels was therefore included in this study.

Table 9: Hypotheses on expression level stability upon selection pressure removal

Hypotheses on the mechanism of action of an attenuated selection marker	**(i) Selection of clones for which plasmid integrated in a naturally transcriptionally highly active zone**	**(ii) Creation of an artificial transcriptionally highly active zone**	**(iii) Selection of clones for which plasmid integrated in multiple copies**	**(iv) Multiplication of plasmid copies**
Impact of selection pressure withdrawal on plasmid copy number and/or transcription level	Not impacted	Possibly impacted	Possibly impacted	Possibly impacted
Impact on expression level stability	Not impacted	Possibly impacted	Possibly impacted	Possibly impacted

Briefly, eight top clones, respectively from SV40E wild type and SV40E_Δ21bpRIII-II conditions, were sub-cultured each with and without selection pressure in culture medium. Three 12-day fed-batch productivity assays were launched in TPP® tubes, respectively at estimated population doubling level 20, 40 and 60 (PDL_{20}, PDL_{40}, and PDL_{60}) (Figure 39). This study duration is defined to span master cell bank (MCB) preparation, working cell bank (WCB) preparation, GMP production, and extended cell bank (ExCB) preparation. Effects of long-term cultivation (LTC) on growth and productivity characteristics was evaluated with and without selection pressure during sub-culturing. More precisely, stability of expression level, considering both rituximab titer and mean process specific productivity (Q_P) variations over 60 PDL, was the main criteria of interest for validation of SV40E_Δ21bpRIII-II promoter for *pac* expression.

In addition, plasmid copy number was monitored during sub-culturing process at PDL_{10}, PDL_{20}, PDL_{30}, PDL_{40}, PDL_{50}, PDL_{60} and PDL_{70} by qPCR. Furthermore, rituximab light chain and heavy chain transcription level was monitored during the three productivity assays, at production day 5 (PD_5), PD_7, PD_{10} and PD_{12} by RT-qPCR (Figure 39). qPCR and RT-qPCR methods have been developed for this purpose. Unfortunately, complete monitoring of plasmid copy number and analysis of rituximab light chain and heavy chain transcription levels could not be achieved before the writing of this manuscript.

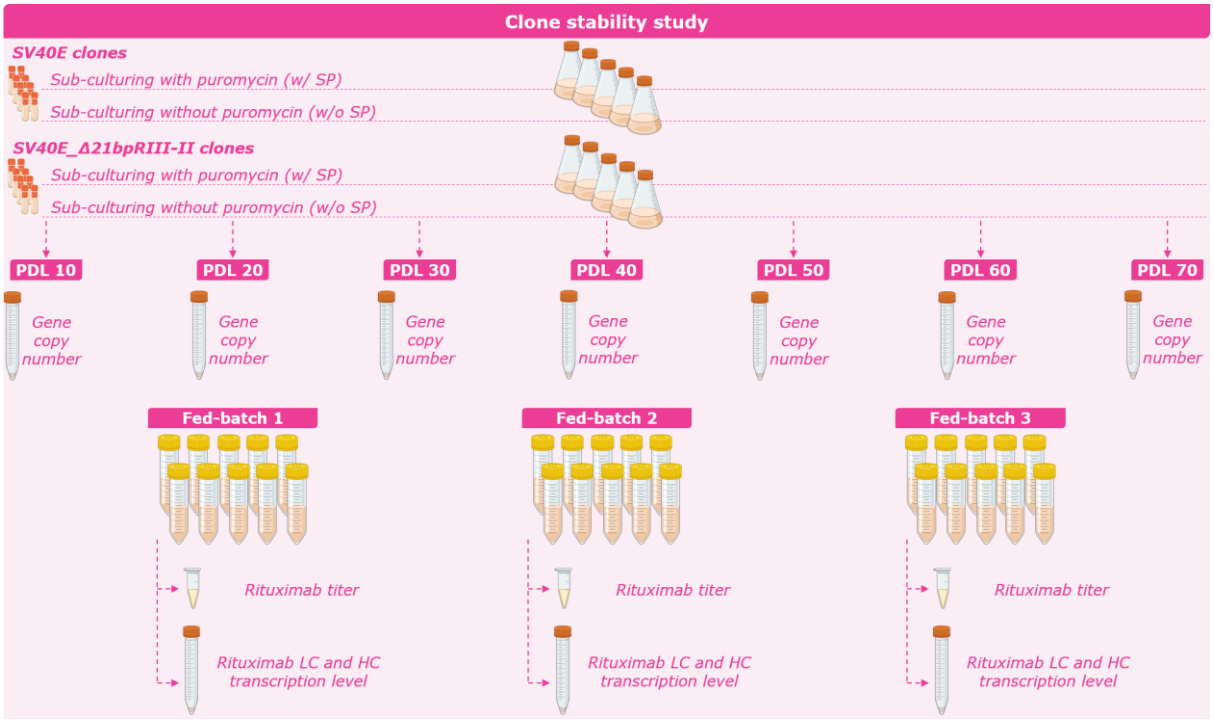


Figure 39: Flow chart of stability study on SV40E and SV40E_ Δ21bpRIII-II clones

Eight top clones, respectively from SV40E wild type and SV40E_Δ21bpRIII-II conditions, were sub-cultured each with and without selection pressure in culture medium. Three 12-day fed-batch productivity assays were launched in TPP® tubes, respectively at estimated population doubling level 20, 40 and 60 (PDL_{20}, PDL_{40}, and PDL_{60}). Plasmid copy number was monitored during sub-culturing process at PDL_{10}, PDL_{20}, PDL_{30}, PDL_{40}, PDL_{50}, PDL_{60} and PDL_{70} by qPCR. Rituximab light chain and heavy chain transcription level was monitored during the three productivity assays, at production day 5 (PD_5), PD_7, PD_{10} and PD_{12} by RT-qPCR. w/ SP: with selection pressure: w/o SP: without selection pressure; PDL: population doubling level; LC: light chain; HC: heavy chain. Partially created with BioRender.com.

3.3.5.5.1 Effects of long-term cultivation on growth

Doubling time (DT) during sub-culturing over 60 PDL, depending on clone, promoter for *pac* expression, and presence/absence of puromycin in culture medium is shown on Figure 40. Doubling time variations are observed over the first 25 days of sub-culturing. Afterwards, doubling time stabilization around 18 to 20 hours is noticed. This overall trend towards a progressive decrease in doubling time over long-term cultivation is observed for all conditions, regardless of clone, promoter for *pac* expression, and presence/absence of puromycin in culture medium.

Significance of mean doubling time difference was evaluated by a statistical analysis (details of which can be found in Appendix 3). Mean titer variability in condition without selection pressure during sub-culturing is higher and significantly different from condition with selection pressure during sub-culturing. Besides, mean specific productivity variability in condition without selection pressure during sub-culturing is higher and significantly different from condition with selection pressure during sub-culturing. Nonetheless, doubling time of clones generated with SV40E_Δ21bpRIII-II promoter

ranges from 18 to 22 hours, which is perfectly suitable for further process development steps and industrial scale-up.

Cumulated integral of viable cell density (IVC) during fed-batch productivity assay at PDL_{20}, PDL_{40}, and PDL_{60} depending on clone, promoter for *pac* expression, and presence/absence of puromycin during sub-culturing is shown on Figure 41. Cumulated IVC variations between PDL_{20}, PDL_{40}, and PDL_{60} were calculated. All clones (except CL 15A10 w/o selection pressure) showed a stable cumulated IVC over 60 PDL (data not shown). Though, slight and steady increase of cumulated IVC is noticed for several clones: CL 02B09, CL 02E06, and CL 02H09 (generated with SV40E wild type promoter, and cultivated with selection pressure during sub-culturing), CL 02D11, CL 02E06, CL 02H09, CL 03E05, and CL 05E12 (generated with SV40E wild type promoter, and cultivated without selection pressure during sub-culturing), and CL 11F08, CL 13D11, CL 13E01, CL 15A10, CL 22G03, and CL 24H08 (generated with SV40E_Δ21bpRIII-II promoter, and cultivated without selection pressure during sub-culturing). It clearly appears that increase of cumulated IVC mainly concerns condition without selection pressure during sub-culturing. This suggests possible underlying modifications, at genomic, transcriptomic, and/or proteomic levels, in response to withdrawal of selection pressure from culture medium.

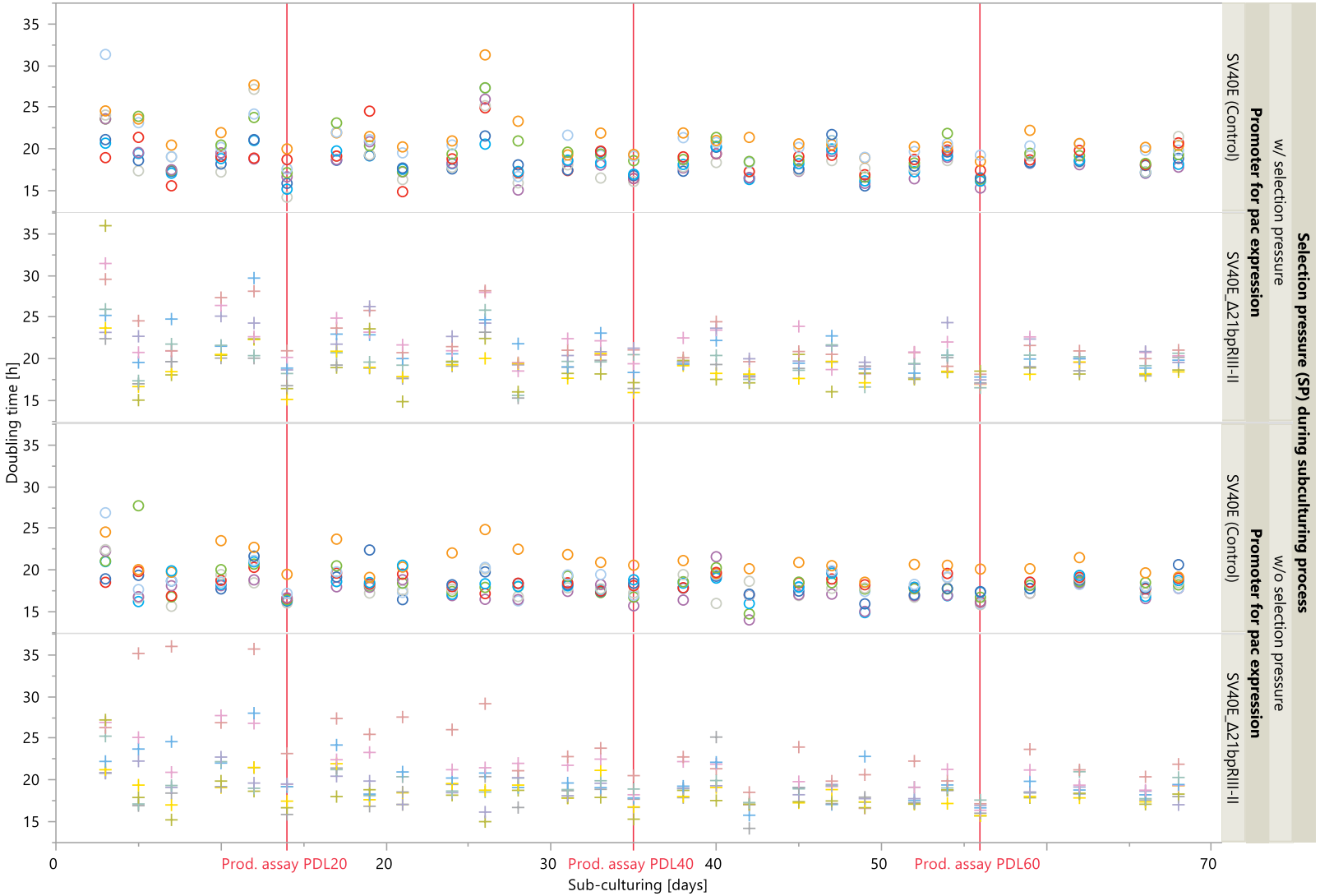


Figure 40: Doubling times over long-term cultivation depending on clone, promoter for pac expression and presence/absence of puromycin during sub-culturing.

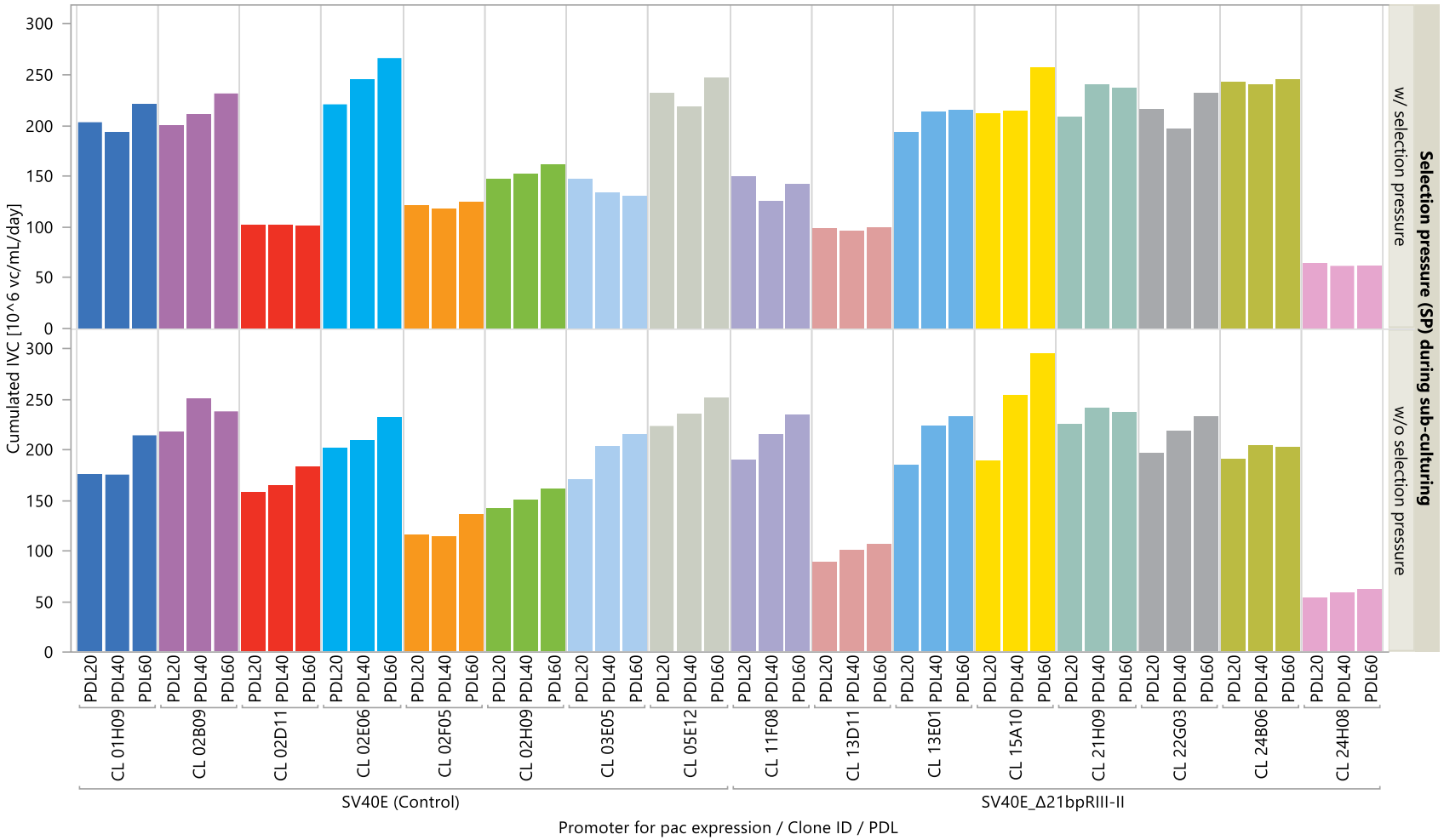


Figure 41: Cumulated IVC during fed-batch productivity assay at PDL_{20}, PDL_{40}, and PDL_{60} depending on clone, promoter for pac expression and presence/absence of puromycin during sub-culturing

Cumulated IVC displayed on the graph are the cumulated IVC on the last common day of culture between the three productivity assays. Data analysis was performed on JMP® 15 software.

3.3.5.5.2 Effects of long-term cultivation on productivity

Rituximab concentration and mean process specific productivity during fed-batch productivity assay at PDL_{20}, PDL_{40}, and PDL_{60} depending on clone, promoter for *pac* expression, and presence/absence of puromycin during sub-culturing are respectively shown on Figure 42 and Figure 43. Rituximab titer variations between PDL_{20}, PDL_{40}, and PDL_{60} are detailed in Table 34, Table 35, Table 36, and Table 37 (in Appendix 3). Mean process specific productivity variations between PDL_{20}, PDL_{40}, and PDL_{60} are detailed in Table 38, Table 39, Table 40, and Table 41 (in Appendix 3). Variations between PDL were calculated on the last common day of culture between the three productivity assays. For example, CL 01H09 culture was stopped at PD_{12} (during productivity assay launched at PDL_{20}) and PD_{10} (during productivity assays launched at PDL_{40} and PDL_{60}). Therefore, titer and mean process specific productivity variations were calculated with titer and mean process specific productivity at PD_{10} (last common day of culture between the three productivity assays) to determine whether expression level of CL 01H09 is stable or not.

3.3.5.5.2.1 Analysis of clone stability, irrespectively of the promoter used for *pac* expression

Irrespectively of the promoter used for *pac* expression, four clone profiles can be distinguished, regarding recombinant protein expression level stability (Table 34, Table 35, Table 36, and Table 37):

(i) Clones for which recombinant protein expression is stable over 60 PDL, whether selection pressure is maintained or not during sub-culturing process. This is namely the case of CL 01H09, CL 02B09, and CL 02F05 (generated with SV40E wild type promoter) (3 out of 16 clones, *i.e.*, 18.8%), and CL 13D11, CL 15A10, and CL 24H08 (generated with SV40E_Δ21bpRIII-II promoter) (6 out of 16 clones, *i.e.*, 37.5%).

(ii) Clones for which recombinant protein expression is stable over 60 PDL if selection pressure is maintained during sub-culturing process, but unstable if selection pressure is withdrawn. This is namely the case of CL 02H09, CL 03E05, CL 05F12 (generated with SV40E wild type promoter) (3 out of 16 clones, *i.e.*, 18.8%), and CL 11F08, CL 13E01, CL 21H09, CL 22G03, and CL 24B06 (generated with SV40E_Δ21bpRIII-II promoter) (8 out of 16 clones, *i.e.*, 50%). It should though be noticed that CL 02H09 and CL 22G03 are very close to the stability criterion.

(iii) Clones for which recombinant protein expression is unstable over 60 PDL, whether selection pressure is maintained or not during sub-culturing process. This is namely the case of CL 02E06 (generated with SV40E wild type promoter) (1 out of 16 clones, *i.e.*, 6.3%), although very close to the stability criterion).

(iv) Clones for which recombinant protein expression is stable over 60 PDL if selection pressure is withdrawn during sub-culturing process, but unstable if selection pressure is maintained.

This is namely the case of CL 02D11 (generated with SV40E wild type promoter) (1 out of 16 clones, *i.e.*, 6.3%).

Similar observations can be made from mean process specific productivity stability. Again, irrespectively of the promoter used for *pac* expression, three clone profiles can be distinguished (Table 38, Table 39, Table 40, and Table 41):

(i) Clones for which mean process Q_P is stable over 60 PDL, whether selection pressure is maintained or not during sub-culturing process. This is namely the case of CL 01H09, CL 02B09, CL 02E06, and CL 02F05 (generated with SV40E wild type promoter), and CL 13D11 and CL 22G03 (generated with SV40E_Δ21bpRIII-II promoter) (6 out of 16 clones, *i.e.*, 37.5%).

(ii) Clones for which mean process Q_P is stable over 60 PDL if selection pressure is maintained during sub-culturing process, but unstable if selection pressure is withdrawn. This is namely the case of CL 02H09, CL 03E05, and CL 05E12 (generated with SV40E wild type promoter), and CL 11F08, CL 13E01, CL 15A10, CL 21H09, CL 24B06, and CL 24H08 (generated with SV40E_Δ21bpRIII-II promoter) (9 out of 16 clones, *i.e.*, 56.3%).

(iii) Clones for which mean process Q_P is stable over 60 PDL if selection pressure is withdrawn during sub-culturing process, but unstable is selection pressure is maintained. This is namely the case of CL 02D11 (generated with SV40E wild type promoter) (1 out of 16 clones, *i.e.*, 6.3%).

It is noticed that most evaluated clones fall in the same categories, whether rituximab titer or mean process specific productivity are considered as stability parameter.

(i) Clone is stable over 60 PDL, whether selection pressure is maintained or not during sub-culturing process (CL 01H09, CL 02B09, CL 02F05, CL 13D11, and CL 22G03). In this case, it is likely that plasmid copy number is stable and that transcriptional activity around plasmid integration site is maintained during long-term cultivation.

(ii) Clone is stable over 60 PDL if selection pressure is maintained during sub-culturing process, but unstable if selection pressure is withdrawn (CL 02H09, CL 03E05, CL 05E12, CL 11F08, CL 13E01, CL 21H09, and CL 24B06). In this case, two hypotheses might explain instability upon selection pressure withdrawal. Plasmid copies can be lost and/or (both mechanisms being independent) transcriptional activity around plasmid integration site can be reduced during sub-culturing when selection pressure is withdrawn. These

assumptions mainly rely on the fact that expression of resistance to puromycin is no longer necessary when puromycin is withdrawn from culture medium.

(iii) Clone stable over 60 PDL if selection pressure is withdrawn during sub-culturing process, but unstable if selection pressure is maintained (CL 02D11).

Conversely, conclusion on CL 15A10 and CL 24H08 stability should be based on analysis of growth (cumulated IVC) and productivity results (rituximab titer and mean process specific productivity). In the case of CL 15A10, mean process specific productivity instability is explained by increase of cumulated IVC while rituximab titer is stable. Increase of cumulated IVC is itself explicated by a slight decrease of clone doubling time during sub-culturing process. However, in the case of CL 24H08, rituximab titer and cumulated IVC are stable, resulting in mean process specific productivity instability when selection pressure is withdrawn. Previously mentioned hypotheses, regarding plasmid copy loss and/or decrease of transcriptional activity, might explain instability upon selection pressure withdrawal.

Significance of mean titer variability and mean specific productivity variability differences between condition with puromycin and condition without puromycin during sub-culturing (irrespective of the promoter used for *pac* expression) was evaluated by a statistical analysis (details of which can be found in Appendix 3). No statistically significant difference was observed between comparisons of interest, namely SV40E wild type promoter *versus* SV40E_Δ21bpRIII-II promoter with selection pressure during sub-culturing, and SV40E wild type promoter *versus* SV40E_Δ21bpRIII-II promoter without selection pressure during sub-culturing.

3.3.5.5.2.2 Analysis of clone stability, depending on promoter used for *pac* expression

Rituximab titer is stable over 60 PDL for four clones (50%) generated with SV40E wild type promoter (CL 01H09, CL 02B09, CL 02D11, and CL 02F05) and three clones (37.5%) generated with SV40E_Δ21bpRIII-II promoter (CL 13D11, CL 15A10, and CL 24H08), when selection pressure is withdrawn from sub-culturing process. Rituximab titer is stable over 60 PDL for six clones (75%) generated with SV40E wild type promoter (CL 01H09, CL 02B09, CL 02F05, CL 02H09, CL 03E05, and CL 05E12) and eight clones (100%) generated with SV40E_Δ21bpRIII-II promoter (CL 11F08, CL 13D11, CL 13E01, CL 15A10, CL 21H09, CL 22G03, CL 24B06, and CL 24H08), when selection pressure is maintained during sub-culturing process (Figure 42).

Mean process specific productivity is stable over 60 PDL for five clones (62.5%) generated with SV40E wild type promoter (CL 01H09, CL 02B09, CL 02D11, CL 02E06, and CL 02F05) and two clones (25%) generated with SV40E_Δ21bpRIII-II promoter (CL 13D11 and CL 22G03), when selection pressure is withdrawn from sub-culturing process. Mean process specific productivity is stable over 60 PDL for seven clones (87.5%) generated with SV40E wild type promoter (CL 01H09, CL 02B09, CL 02E06, CL 02F05, CL 02H09, CL 03E05, and CL 05E12) and eight clones (100%) generated with SV40E_Δ21bpRIII-II promoter (CL 11F08, CL 13D11, CL 13E01, CL 15A10, CL 21H09, CL 22G03,

CL 24B06, and CL 24H08), when selection pressure is maintained during sub-culturing process (Figure 43).

Significance of mean titer variability and mean specific productivity variability differences between SV40E wild type and SV40E_Δ21bpRIII-II promoters was evaluated by a statistical analysis (details of which can be found in Appendix 3). No statistically significant difference was observed between comparisons of interest, namely SV40E wild type promoter *versus* SV40E_Δ21bpRIII-II promoter with selection pressure during sub-culturing, and SV40E wild type promoter *versus* SV40E_Δ21bpRIII-II promoter without selection pressure during sub-culturing.

From these observations, it appears that the use of SV40E_Δ21bpRIII-II promoter instead of SV40E wild type promoter for *pac* expression does not significantly impact titer and mean process specific productivity stability over 60 PDL. As previously demonstrated, CHOZN® CHO K1 clone productivity instability is rather explained by removal of selection pressure. Therefore, this study confirms that the weakened selection marker strategy can be applied on CHOZN® CHO K1 cell line development platform, without affecting process performance.

From these observations, it appears that the use of SV40E_Δ21bpRIII-II promoter instead of SV40E wild type promoter for *pac* expression does not significantly impact titer and mean process specific productivity stability over 60 PDL. As previously demonstrated, CHOZN® CHO K1 clone productivity instability is rather explained by removal of selection pressure. Therefore, this study confirms that the weakened selection marker strategy can be applied on CHOZN® CHO K1 cell line development platform, without affecting process performance.

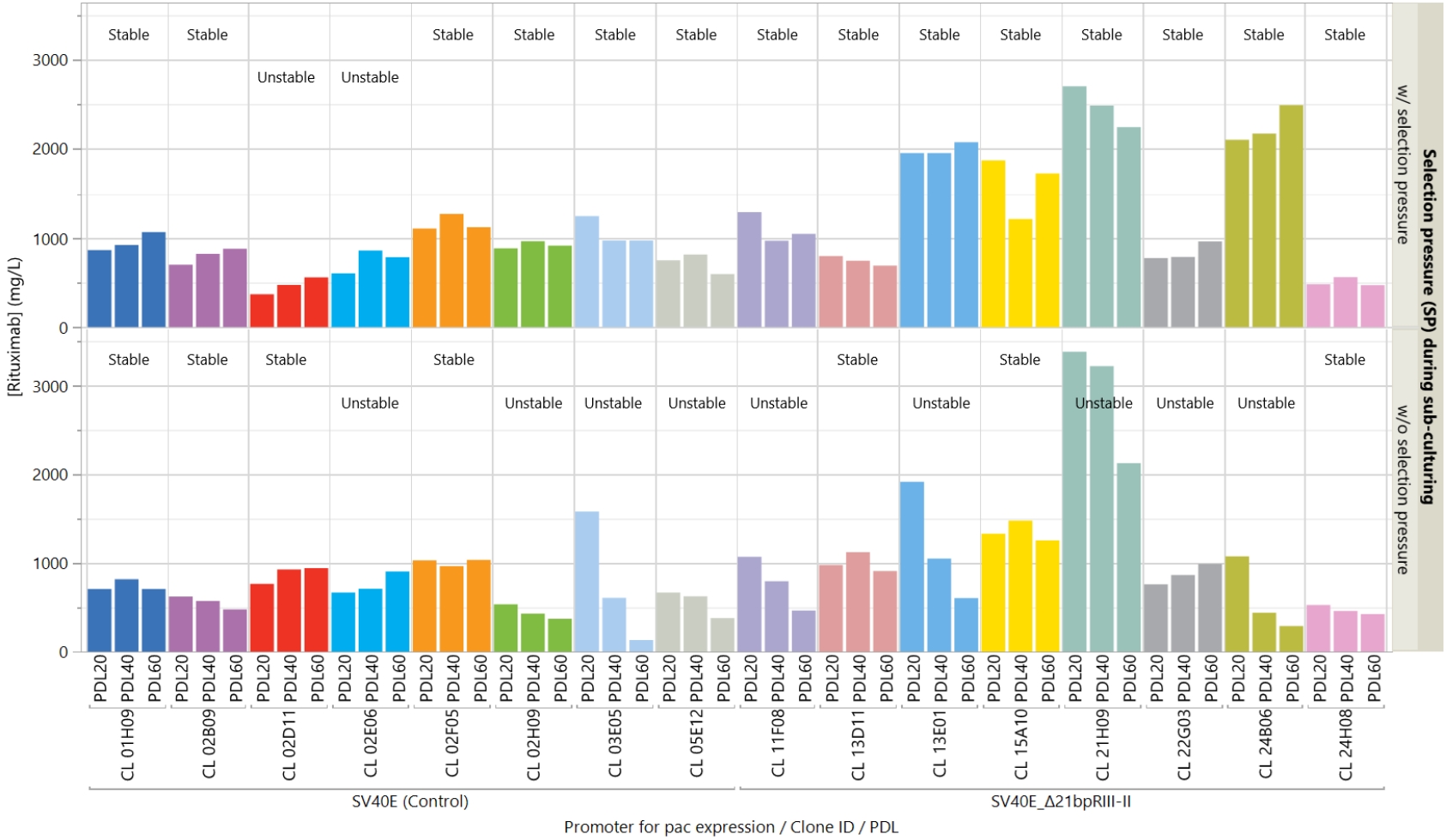


Figure 42: Rituximab concentrations after fed-batch productivity assay at PDL_{20}, PDL_{40}, and PDL_{60} depending on clone, promoter for pac expression and presence/absence of puromycin during sub-culturing

Rituximab concentrations displayed on the graph are the concentrations on the last common day of culture between the three productivity assays. Conclusions on clone stability or instability were drawn from these values. Data analysis was performed on JMP® 15 software.

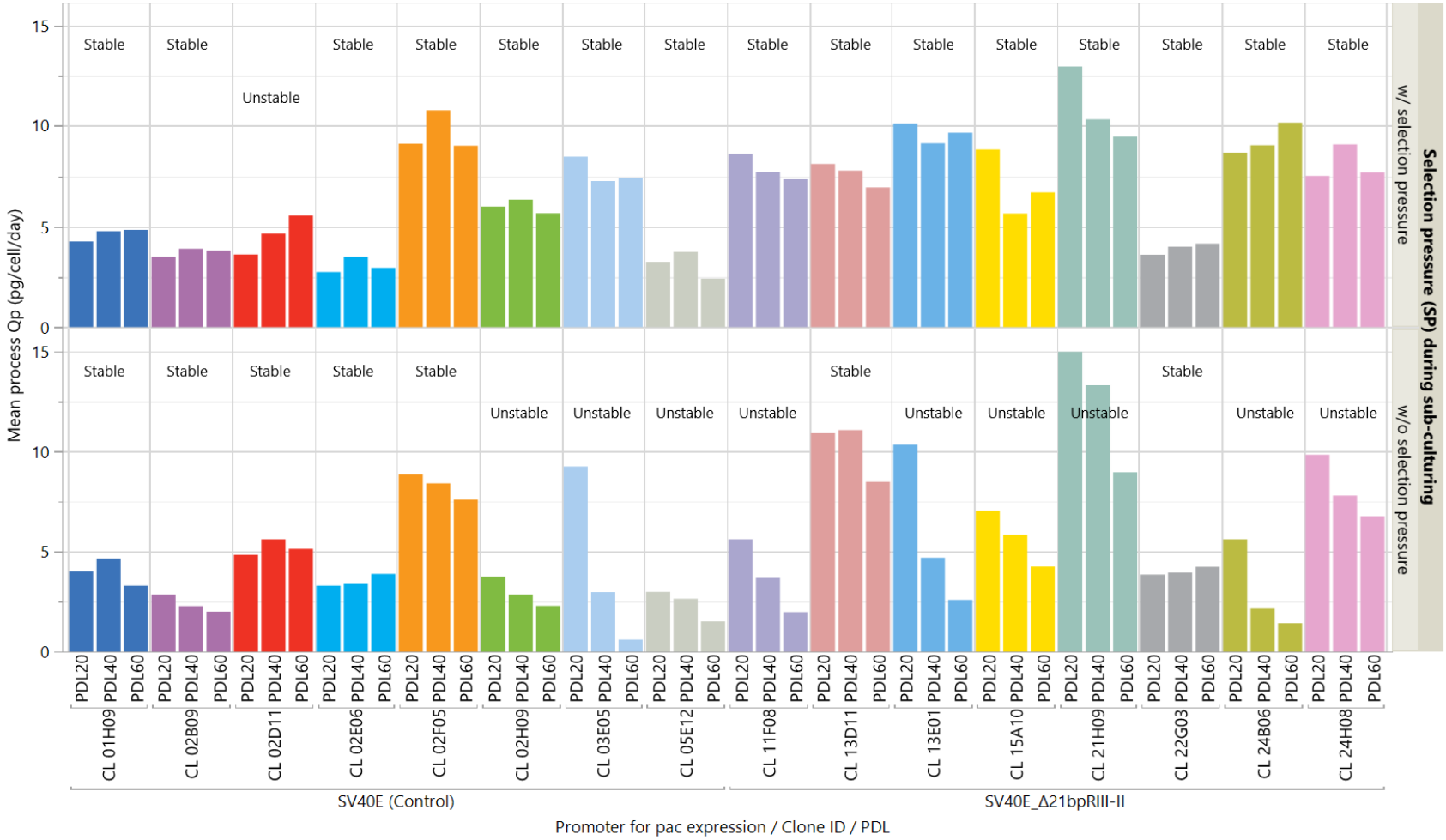


Figure 43: Mean process specific productivities after fed-batch productivity assay at PDL_{20}, PDL_{40}, and PDL_{60} depending on clone, promoter for pac expression and presence/absence of puromycin during sub culturing

Mean process specific productivities displayed on the graph are the mean process specific productivities on the last common day of culture between the three productivity assays. Conclusions on clone stability or instability were drawn from these values. Data analysis was performed on JMP® 15 software.

3.3.5.6 Productivity gain obtained with RAS and SV40E_Δ21bpRIII-II promoters is confirmed on other recombinant proteins

Mini-pool productivity increase and, above all, clone productivity increase were previously demonstrated with several candidate promoters for *pac* expression. Nonetheless, all assays achieved so far have been performed with rituximab, an IgG_1 chimeric antibody. Therefore, it was necessary to validate the gain of productivity obtained with RAS and SV40E_Δ21bpRIII-II promoters on other recombinant proteins.

Two IgG_1 (rituximab and trastuzumab), two IgG_4 (pembrolizumab and nivolumab), and a F_C-fusion protein (EPO-F_C) were selected for confirmation study. Selected recombinant proteins for confirmation study had to meet several criteria: (i) protein should have a maximum of two chains, so that the coding sequence can be inserted in the two multiple cloning sites (MCS) of pEM2v-c vector backbone; (ii) protein should display affinity for protein A, so that titration method by bio-layer interferometry using protein A biosensors can be used; (iii) protein should be used for therapeutic applications. Therefore, several types of molecules were excluded from the scope of this study. To name a few:

(i) Other IgG sub-classes, namely IgG_2 and IgG_3. Even if some IgG_2 are used as therapeutic antibodies (tremelimumab, tezepelumab, meplazumab, and tanezumab can be cited as very recent examples (Kaplon et al., 2020, 2022; Kaplon & Reichert, 2021b)), they are much less frequent than IgG_1 and IgG_4. On the other hand, IgG_3 are not (or very rarely) used as therapeutic antibodies (Kaplon et al., 2020, 2022; Kaplon & Reichert, 2021b; Walsh, 2014, 2018), because of their large hinge region which renders them highly sensitive to proteolysis and therefore not useable as therapeutic molecules (Saito et al., 2019).

(ii) Other immunoglobulin classes, namely IgA, IgD, IgE, and IgM. These immunoglobulin classes are not (or very rarely) used as therapeutic antibodies (Kaplon et al., 2020, 2022; Kaplon & Reichert, 2021b; Walsh, 2014, 2018), because of shorter serum half-life (one week for IgA, two days for IgE, and ten days for IgM, instead of three weeks for IgG (Goulet & Atkins, 2020)) or difficult expression (for example due, in the case of IgM, to large protein size and complexity (Keyt et al., 2020)).

(iii) Bispecific antibodies. As they target two different epitopes, bispecific antibodies are constituted of four different chains. Yet, pEM2v-c backbone is not suited for expression of molecules with more than two chains.

(iv) Derivatives from variable fragments (F_V) or antigen-binding fragments (F_{ab}), such as single chain variable fragment (scFv), antigen-binding fragments (F_{ab}), single-domain antibody

(sdAb), F(ab')2 or Fab'. These derivatives do not arbor the F_C fragment which is responsible for protein A affinity. Therefore, titration with protein A-method is not possible.

(v) Conjugated molecules, namely antibody-drug conjugate (ADC) and antibody-radioisotope conjugate (ARC). Conjugation of antibody to drug (ADC) or radioisotope (ARC) is performed after downstream processing (DSP) (Baah et al., 2021; Tsuchikama, 2018; Yamada & Ito, 2019). Therefore, cell line development process for ADC or ARC is unchanged compared to a cell line development process for an IgG_1 or IgG_4.

The mini-pool strategy, for the same reasons as detailed before (Candidate promoters for *pac* expression decrease mini-pool reemergence but increase mini-pool productivity), was applied for confirmation study. Briefly, CHOZN® CHO K1 host cell line was transfected by electroporation with plasmids carrying candidate promoters for *pac* expression and previously mentioned recombinant protein coding sequences. Electroporation parameters were modified (1,400 V, 20 ms, 2 pulses instead of 1,600 V, 10 ms, 3 pulses), based on previous observations (Modifications on protocol for mini-pool generation). One day after transfection, mini-pools were seeded in 96-well plates and put under selection pressure. Mini-pool seeding density was modified (10,000 vc/well instead of 5,000 vc/well), based again on previous observations (Modifications on protocol for mini-pool generation). Mini-pools were cultured in 96-well plates until cellular reemergence was observed; a maximum of three weeks was allowed for reemergence. When the first of these two criteria was met, a seven-day non normalized batch productivity assay was launched in 96-well plates (viable cell density not being controlled at seeding). At the end of the productivity assay, recombinant antibody concentration in cell culture supernatant was measured by bio-layer interferometry using protein A coated sensors. Percentage of reemerging mini-pools and mini-pool productivity were the main criteria of interest for evaluation of RAS and SV40E_Δ21bpRIII-II promoters for *pac* expression.

Percentage of reemerging mini-pools in 96-well plates are given in Table 10. RAS and SV40E_Δ21bpRIII-II promoters display identical mini-pool reemergence percentages for trastuzumab, and EPO-F_C expression. Minor reemergence differences should be noticed for rituximab (slightly lower for RAS and SV40E_Δ21bpRIII-II promoters), pembrolizumab (slightly lower for RAS promoter, lower for SV40E_Δ21bpRIII-II promoter) and nivolumab (lowest reemergence for SV40E wild type promoter) expression. Nonetheless, despite identical or similar reemergence percentage between SV40E wild type promoter and RAS and SV40E_Δ21bpRIII-II promoters, reemergence duration is generally shorter with SV40E wild type promoter, compared to RAS and SV40E_Δ21bpRIII-II promoters.

Besides, it should be noted that mini-pool protocol applied here considers a combination of modified parameters previously evaluated (Modifications on protocol for mini-pool generation): softened electroporation parameters and increased viable cell seeding density. Comparison can be established with former results. Reemergence of rituximab producing mini-pools generated with both RAS and

SV40E_Δ21bpRIII-II promoters is the highest from all assays performed so far (RAS: 3% during first assay, 23% during second assay with control conditions, 87% during second assay with modified seeding density, and 68% during second assay with modified electroporation parameters; SV40E_Δ21bpRIII-II: 23% during first assay, 95% during second assay with control conditions, 100% during second assay with modified seeding density, and 98% during second assay with modified electroporation parameters). Combination of modifications might be an additional lever for enhanced cellular recovery.

Table 10: Percentage of reemerging mini pools

CHOZN® CHO K1 host cell line was transfected by electroporation (electroporation parameters being modified) with plasmids carrying candidate promoters for pac expression. One day after transfection, mini-pools were seeded in 96-well plates (seeding density being modified) and put under selection pressure. Mini-pools were cultured in 96-well plates until cellular reemergence was observed; a maximum of three weeks has been allowed for reemergence. Percentage of reemerging mini-pools is expressed as the ratio of number of mini-pools which have a confluence delta higher than 20% between production day 0 (PD_0) and production day 7 (PD_7) and of total number of seeded mini-pools per condition (one 96-well plate per condition, i.e., 60 mini-pools per condition).

Recombinant protein	**Percentage of reemerging mini-pools**			**Duration of reemergence [days]**		
	SV40E (Control)	**RAS**	**SV40_Δ21b pRIII-II**	**SV40E (Control)**	**RAS**	**SV40_Δ21b pRIII-II**
Rituximab	100%	98%	99%	9	11	12
Trastuzumab	100%	100%	100%	10	14	10
Pembrolizumab	100%	98%	72%	10	14	17
Nivolumab	95%	100%	100%	17	10	14
EPO-F_C	100%	100%	100%	14	14	10

Mini-pools' titers (recombinant protein concentration in supernatant) after batch productivity assay in 96-well plate are shown on Figure 44. RAS promoter increases maximum titer for rituximab (which confirms candidate promoter screening results), trastuzumab, nivolumab, and EPO-F_C production, compared to SV40E wild type promoter. Though, median titer is increased for all evaluated recombinant proteins. SV40E_Δ21bpRIII-II promoter increases maximum titer for rituximab (which, again, confirms candidate promoter screening results, even though it should be noticed that the maximum titer is lower than previously obtained), trastuzumab, and EPO-F_C production, compared to SV40E wild type promoter. Median titer is increased in the same conditions.

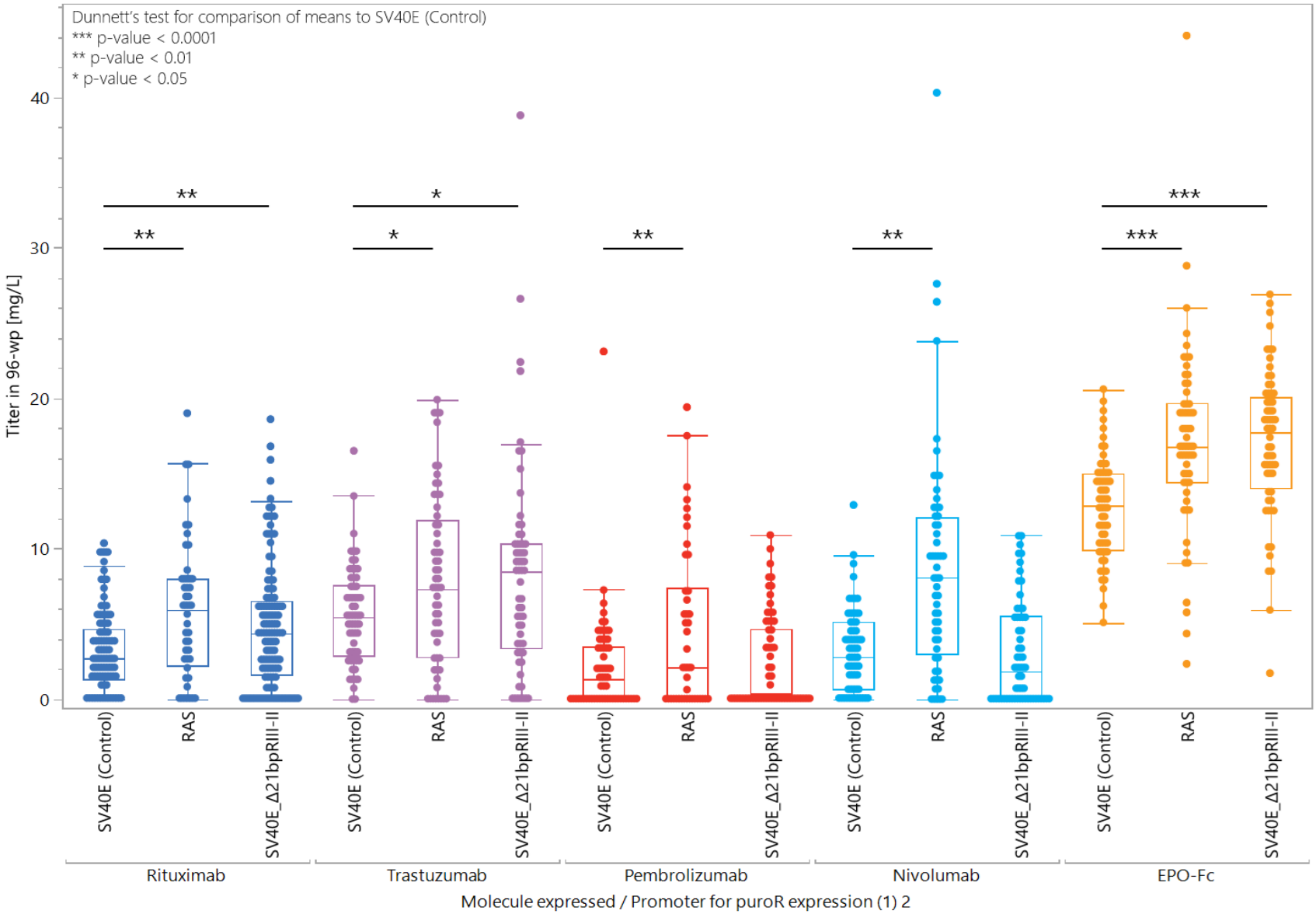
Significance of mean titer difference between candidate and SV40E wild type promoters was evaluated by a statistical analysis (details of which can be found in Appendix 3). RAS promoter shows a

significantly higher mean titer than SV40E wild type promoter for rituximab, trastuzumab, and pembrolizumab expression. RAS promoter also shows a significantly higher mean titer than SV40E wild type promoter for nivolumab and EPO-F_C expression, though it should be kept in mind that non-parametric tests are less robust than parametric ones. SV40E_Δ21bpRIII-II promoter shows a significantly higher mean titer than SV40E wild type promoter for rituximab and trastuzumab. SV40E_Δ21bpRIII-II promoter also shows a significantly higher mean titer than SV40E wild type promoter for EPO-F_C expression (same consideration about non-parametric assay to be considered).

Both RAS and SV40E_Δ21bpRIII-II promoters increase mini-pool productivity on several standard monoclonal antibodies and one fusion protein. Synthesis of best conditions, considering maximum and mean titer, is given in Table 11. RAS promoter appears to be the best condition for both highest maximum and mean titer, for the majority of evaluated recombinant proteins.

Table 11: Synthesis of best conditions for highest maximum and mean titer

Recombinant protein	**Best condition for**	
	Highest maximum titer	**Highest mean titer**
Rituximab	RAS	RAS with statistically significant mean difference
Trastuzumab	SV40E_Δ21bpRIII-II	SV40E_Δ21bpRIII-II with statistically significant mean difference
Pembrolizumab	SV40E (Control)	RAS with statistically significant mean difference
Nivolumab	RAS	RAS with statistically significant mean difference (non-parametric test)
EPO-F_C	RAS	RAS with statistically significant mean difference (non-parametric test)


Dunnett's test for comparison of means to SV40E (Control)
*** p-value < 0.0001
** p-value < 0.01
* p-value < 0.05
Titer in 96-wp [mg/L]
0
10
20
30
40
**
**
*
*
**
**
***
***
SV40E (Control)
RAS
SV40E_Δ21bpRIII-II
Rituximab
Trastuzumab
Pembrolizumab
Nivolumab
EPO-Fc
Molecule expressed / Promoter for puroR expression (1) 2

Promoter for pac expression	SV40E (Control)	RAS	SV40E_Δ21bpRIII-II	SV40E (Control)	RAS	SV40E_Δ21bpRIII-II	SV40E (Control)	RAS	SV40E_Δ21bpRIII-II	SV40E (Control)	RAS	SV40E_Δ21bpRIII-II	SV40E (Control)	RAS	SV40E_Δ21bpRIII-II
Molecule	Rituximab			Trastuzumab			Pembrolizumab			Nivolumab			EPO-Fc		
Max titer [mg/L]	10.1	**19.0**	18.6	16.8	**20.4**	39.1	**23.1**	19.4	10.9	12.9	**40.3**	10.9	17.2	**41.6**	23.4
Fold increase on max titer	-	**1.9**	1.8	-	**1.2**	2.3	-	0.8	0.5	-	**3.1**	0.8	-	**2.4**	1.4
Median titer [mg/L]	2.7	**5.9**	4.3	5.4	7.4	**8.6**	1.3	**3.9**	0.3	2.8	**9.4**	1.8	9.9	13.6	**14.4**
Fold increase on median titer	-	**2.2**	1.6	-	1.4	**1.6**	-	**3.0**	0.2	-	**3.3**	0.6	-	1.4	**1.4**

Figure 44 :Rituximab, trastuzumab, pembrolizumab, nivolumab, and EPO-Fc concentration in mini-pools' supernatants after batch productivity assay in 96-well plates

Box-and-whiskers plots of rituximab (dark blue), trastuzumab (purple), pembrolizumab (red), nivolumab (light blue), and EPO-F_C fusion protein (orange) concentration in mini-pools' supernatants after batch productivity assay in 96-well plates. Each dot represents one mini-pool. Data are shown only for growing mini-pools (i.e., ΔPD_7-PD_0 confluence ≥ 20% on 96-well plate productivity assay). Boxplot horizontal lines respectively indicate, top to bottom, the maximum, the upper quartile (Q3), the median, the lower quartile (Q1) and the minimum concentration. Dots outside box-and-whiskers plot are considered as outliers. A point is an outlier if it exceeds the upper outlier limit (Q_3+1.5×IQR) or is less than the lower outlier limit (Q_1-1.5×IQR), where IQR (interquartile range) = Q_3-Q_1. Fold increase on maximum or median titer is calculated in relation to SV40E (control) maximum or median titer for all conditions. Data analysis was performed on JMP® 15 software.

3.3.6 Discussion

Identification of stable and producing clones in the CHOZN® CHO K1 cell line development platform relies on puromycin-based selection. Puromycin-based selection of stably transfected cells is widespread technique due to the high efficiency of selection, the small size of *pac* gene and the low cost of puromycin (Aviner, 2020; Caputo et al., 2021; Yeo et al., 2017).

However, expression of puromycin N-acetyltransferase (PAC), which confers resistance to puromycin (Vara et al., 1985, 1986), is currently under the control of SV40E promoter. As transcriptional strength of SV40E promoter is relatively strong, significant levels of PAC could be expressed, even if the expression vector was not integrated in high copy number nor in highly active transcriptional sites (Fan, Kadura, et al., 2013). In addition, only low levels of PAC are necessary for cells to survive selection process (Yeo et al., 2017). This could still allow low producing cells to survive in stable transfected pools (Fan et al., 2011; Fan, Kadura, et al., 2013; Yeo et al., 2017).

As a consequence, optimizing antibiotic resistance protein is a relevant strategy to improve the productivity of recombinant cell lines (Ho et al., 2012; Yeo et al., 2017). Degree of selection stringency is an important parameter for high producing clone identification (Bandaranayake & Almo, 2014; Fussenegger et al., 1999; F. M. Wurm, 2004). High degree of selectivity forces the cells to be more productive in the locus of the selection marker in order to survive the selection process, resulting in better plasmid integration site, or in higher transcriptional or translational activity, or in higher gene copy number (Lai et al., 2013; Ng et al., 2007). Numerous strategies may be implemented to attenuate selection marker in order to increase the productivity of recombinant cell lines, among which the use of weak promoters to drive the expression of resistance to selection pressure (Chin et al., 2015; Fan, Kadura, et al., 2013; Hartman & Tso, 2013; Klein & Kopetzki, 2009; Sacco et al., 2022).

This third chapter aimed at evaluating novel promoters for expression of resistance to selection pressure on several criteria: technical feasibility, productivity and stability of generated cell lines, and cell line development timelines. Identification of candidate promoters for *pac* expression relied on two strategies. The first strategy was to reduce SV40 transcriptional activity by deleting important transcriptional sites. We relied on previously published work (either focusing on characterization of SV40E promoter structure (Benoist & Chambon, 1980b, 1981; Byrne et al., 1983; Wildeman, 1988) or on the use of SV40E promoter and derivatives for mammalian cell line development (Fan, Kadura, et al., 2013; Hartman & Tso, 2013; Klein & Kopetzki, 2009; Yang et al., 2019). The second strategy was to substitute SV40E wild type promoter by a CHO K1 endogenous promoter. To our knowledge, this approach has never been described before.

A 38 to 77% decrease in *pac* transcription level in transient expression was observed by RT-qPCR measurement after substituting SV40E promoter either by a CHO K1 endogenous promoter (H2AJ, DRC7 or RAS) or a modified SV40E promoter (SV40_Δ72bpRII, SV40_Δ21bpRIII-II, SV40_ΔTATA

or SV40_ΔAT-region). Substitution of SV40E promoter by endogenous promoters led to the highest decreases (-68 to -77%) in *pac* transcription level (difference being statistically significant). However, substitution of SV40E promoter by modified SV40E promoters led to lower decreases (-38 to -54%) in *pac* transcription level (difference being though still statistically significant). Deletion of 72 bp repeat II led to a 38% decrease in *pac* transcription level, while, for the same deletion, no significant change was observed in GS mRNA level by Fan *et al.* (Fan, Kadura, et al., 2013). Our results are closer to those of Klein & Kopetzki, who however deleted both 72 bp repeats and obtained a deleted SV40E promoter having lost 80% of wild type SV40E promoter activity (Klein & Kopetzki, 2009). Deletion of 21 bp repeats III and II led to a 45% decrease in *pac* transcription level, which is very similar to the decrease observed by Fan *et al.* (Fan, Kadura, et al., 2013) and consistent with reports from the 1980s describing the 21 bp repeat region as a major promoter component for early transcription (Benoist & Chambon, 1980b, 1981; Byrne et al., 1983; Wildeman, 1988). Lastly, deletions in the AT-region, either targeting only the TATA box or the whole AT-region, led to a decrease of 44% and 54% respectively and no comparison can be established with other studies (the TATA box region has not been targeted in other studies using modified SV40E for expression of selection pressure).

Substitution of SV40E promoter by either CHO K1 endogenous promoters or modified SV40E promoters severely decreased cellular reemergence, which is an indication of increased selection stringency. Good correlations were observed between transcriptional activity of evaluated promoters and either mini-pool (with statistically significant correlation), transfected bulk pool, or clone reemergence (trend of correlation only (not statistically significant)). One can though notice unexpected results, with regards to promoter strength and cellular reemergence. SV40E_Δ21bpRIII-II and SV40E_ΔTATA promoters led to almost identical decrease of *pac* transcription level (-45% and -44%, respectively). However, a greater number of mini-pools reemerged and duration of transfected bulk pool reemergence was shorter with SV40E_ΔTATA compared to SV40E_Δ21bpRIII-II. Previously published reports mentioned that the removal of the TATA box leads to generation of multiple cap sites. More precisely, it was hypothesized that SV40E sequence elements can functionally substitute the TATA box when this one is lacking, but that the TATA box is more readily recognized by RNA polymerase and/or transcription factors than these sequence elements in the wild-type promoter (Benoist & Chambon, 1981). This would explain results in stable expression: generation of multiple cap sites enables sufficient transcription level of *pac* with regards to puromycin concentration for selection, resulting in decreased mini-pool reemergence compared to SV40E promoter, but higher reemergence compared to other SV40E modified promoters. However, results in transient expression are not explained. A different regulation of the SV40E_ΔTATA promoter in transient and stable expression cannot be excluded.

Several assays have been conducted to improve both mini-pool and transfected bulk pool reemergence, focusing on electroporation parameters (both for mini-pools and transfected bulk pools) and on seeding

density and seeding medium (for mini-pools), in particular for showing the technical feasibility of combining a direct single-cell cloning (DSCC) process with a selection marker attenuation strategy. However, in order to increase selection stringency only, enhancing reemergence might have been counterproductive, maybe by allowing the reemergence of low-producing cells.

Productivity of CHOZN® CHO K1 mini-pools was increased when SV40E promoter was substituted by either CHO K1 endogenous promoters (RAS) or modified SV40E promoters (SV40_Δ72bpRII, SV40_Δ21bpRIII-II, and SV40_ΔTATA). Mini-pool maximum titer improved from a 1.7 to a 2.3-fold increase and median titer improved from a 1.6 to a 2.1-fold increase compared to SV40E wild type promoter, with SV40E_Δ21bpRIII-II promoter giving the best productivities on both criteria. Yang *et al.* also conducted their study on mini-pools and reported a 3.8-fold increase on mini-pool median titer and a 2-fold increase on mini-pool maximum titer with the modified SV40E promoter (a 189-bp deletion covering 72 bp repeats II and I and 21 bp repeats III and II) compared to SV40E wild type promoter (Yang et al., 2019). The increase factors are of the same order of magnitude and statistically significant in both studies.

Productivity of CHOZN® CHO K1 clones was increased when SV40E promoter was substituted by either CHO K1 endogenous promoters (RAS) or modified SV40E promoters (SV40_Δ72bpRII, and SV40E_Δ21bpRIII-II – but not by SV40_ΔTATA). Clone maximum titer improved from a 1.1 to a 1.5-fold increase and median titer improved from a 1.3 to a 1.5-fold increase compared to SV40E wild type promoter, with SV40E_Δ21bpRIII-II promoter leading to the best maximum titer and RAS promoter leading to the best median titer. Clone maximum specific productivity improved up to a 1.3-fold increase and median titer improved from a 1.5 to a 2.2-fold increase compared to SV40E wild type promoter, with SV40E_Δ21bpRIII-II promoter giving the best productivities on both criteria. Fan *et al.* also conducted their study on clones and reported a 1.8-fold increase on clone maximum titer and a 3-fold increase on clone median titer with the modified SV40E promoter (deletion of 72 bp repeats and deletion of 21 bp repeats III and II) compared to SV40E wild type promoter (Fan, Kadura, et al., 2013).

Increase of productivity observed by Fan *et al.* could not be linked to high plasmid copy number and was hypothesized to result from integration of the expression vector in a transcriptional hot spot (Fan, Kadura, et al., 2013). On the other hand, increase of productivity observed by Yang *et al.* was attributed to a significant higher copy number of selection marker, heavy chain and light chain genes in the modified SV40E promoter compared to SV40E wild type promoter (Yang et al., 2019). Similar analyses were planned in our study, namely the analysis of plasmid copy number on generated clones (but unfortunately could not be done before the writing of this manuscript), to try to understand mechanism(s) underlying productivity increase.

Most studies, if not all, published so far have been performed on only one molecule. In the industrial context of the thesis, it is important to show the robustness of the strategy on several molecules, ideally

of different types. Therefore, a confirmation study was conducted on several industrially relevant molecules, including IgG1 (rituximab and trastuzumab), IgG4 (pembrolizumab and nivolumab), and a fusion protein (EPO-Fc). This panel of molecules complete that of previous studies, which have been conducted with adalimumab, a human IgG1 antibody (Fan, Kadura, et al., 2013). However, the results obtained so far, during mini-pool and clone generation for evaluation of candidate promoters with rituximab, and during mini-pool generation for confirmation on several molecules, do not clearly indicate one ideal promoter for *pac* expression for the next version of the expression vector associated to the CHOZN® CHO K1 platform. SV40E_Δ21bpRIII-II promoter was identified as the best promoter for increasing clone maximum titer, as well as maximum and mean specific productivity. If top clones are highly interesting with regards to productivity, that of the other clones decreases very quickly, resulting in a greatly heterogeneous population of clones. Extrapolation of results to a 50-plate clone generation (usual amount of 96-well plates for a standard clone generation) indicates that all clones screened during the final fed-batch productivity assay would at least produce 1 g/L of recombinant antibody, with five clones producing at least 3 g/L and 15 clones producing between 1 and 2 g/L). On the other hand, RAS promoter was identified as the best promoter for increasing clone median titer. The population of clones generated with RAS promoter is the most homogeneous, whether for titer or for mean process specific productivity. Extrapolation of results to a 50-plate clone generation indicates that all clones screened during the final fed-batch productivity assay would be producing between 1 and 2 g/L of recombinant antibody. Results of the confirmation study on several molecules tend to indicate that RAS promoter is the best condition for both highest maximum and mean titer, for the majority of evaluated recombinant proteins. Nonetheless, SV40E_Δ21bpRIII-II promoter led to an increase on maximum and median titer for both IgG1 and the fusion protein. The fact that SV40E_Δ21bpRIII-II promoter had no effect on IgG4 production is more likely attributed to chance than to a link between this promoter and the γ4 isotype of heavy chain.

Furthermore, only few studies report the analysis of the stability of expression level over long-term cultivation. In particular, this point might have not been taken into account by Fan *et al.* and Yang *et al.* as their cell line platform was relying on a CHO $GS^{-/-}$ cell line. In the case of the CHOZN® CHO K1 platform however, the withdrawal of selection pressure might impact plasmid copy number and/or transcription level, and subsequently expression level stability. This stability study was conducted on clones generated with SV40E_Δ21bpRIII-II promoter, as this promoter was identified as the best conditions on the majority of productivity criteria. When selection pressure was maintained in subculturing medium during long-term cultivation, all clones generated with SV40E_Δ21bpRIII-II promoter retained at least 70% of their initial productivity. However, even if more clones generated with SV40E_Δ21bpRIII-II promoter were unstable when selection pressure was removed from sub-culturing during long-term cultivation compared to clones generated with SV40E wild type promoter, the difference is not statistically significant. This last part of the study confirms that the weakened selection

marker strategy can be applied on CHOZN® CHO K1 cell line development platform, without affecting process performance. In addition, this stability study gave precious insights on the CHOZN® CHO K1 clone behavior over long-term cultivation. Specifically, it appears that CHOZN® CHO K1 clone productivity instability is rather explained by removal of selection pressure.

Several limitations can though be identified in this study. First of all, our approach is both dependent of the integration site of the expression vector in the host cell genome and of gene copy number. Indeed, current configuration of expression vector and cell line development process enable random integration of the plasmid in the genome of the host cell line. Yet, it has been shown that insertion site influence stable gene expression in CHO cells (Gaidukov et al., 2018; Hamaker & Lee, 2018; Srirangan et al., 2020). More unbiased approaches could be applied to circumvent possible position and copy number effects. A first methodology would rely on the use of an expression vector and a cell line development process enabling targeted integration of the plasmid in a pre-defined *locus*. A second methodology, even more adequate than the first one, would rely on the use of artificial chromosomes, such as bacterial artificial chromosomes (BACs). This approach was recently used by Reinhart *et al.* to evaluate host-cell specific differences between three frequently used CHO cell lines (Reinhart et al., 2019). Both approaches have the advantage of being independent of the insertion site, but are though gene copy number-dependent. Normalization to gene copy number can easily be performed by RT-qPCR.

This first bias could be mitigated by an increased screening capacity. Screening a larger number of clones would be more reliable and statistically significant. Our screening capacity was defined by laboratory and human resources. Increase of productivity was observed at all cultures scales (96-well plates, 24-deepwell plates, and TPP® tubes, as well as during mini-pool and clone generation), though statistically significant increase on mean titer or specific productivity could only be shown in 96-well plates and 24-deepwell plates (either during mini-pool or clone generation). It is highly probable that statistically significant difference could be shown if number of screened clones in TPP® tubes was increased.

Furthermore, our selection marker attenuation strategy was only focused on the promoter driving *pac* expression. Yet, several articles report the combined use of two or more selection marker attenuation strategies. For example, mRNA and protein destabilizing elements have been combined (Ng et al., 2007), and an attenuated IRES element was combined with both protein destabilizing elements and codon de-optimized selection marker (Chin et al., 2015). To further our work, similar strategies could be implemented, as weakened/weak promoters are compatible with all approaches, except IRES (if used in the same expression cassette). Indeed, in this case, light and heavy chain coding sequences are positioned within the same cistron than selection marker, and using a weakened or weak promoter would result in a decrease of antibody expression. Finding the right balance of weakened or destabilizing

elements for attenuation of selection marker is highly important, as over-weakening selection marker would lead to cells failing to survive selection process (Yeo et al., 2017).

Lastly, attenuation of selection marker was sometimes presented as a strategy for identification of "hot-spots" in the host cell genome (Fan, Kadura, et al., 2013). Indeed, one hypothesis which was formulated regarding the mechanism of action of an attenuated selection marker was that an attenuated selection marker would lead to selection of clones for which plasmid integrated in a naturally transcriptionally highly active zone (Ng et al., 2007). In this study, CL 13E01, CL 21H09, and CL 24B06, which are the three top clones of the SV40E_Δ21bpRIII-II condition, proved to be unstable during long-term cultivation when selection pressure is withdrawn from culture medium during subculturing. Therefore, it is very unlikely that high productivity of these clones results from integration of plasmid in a natural transcriptional hot-spot (otherwise, expression level would be stable over time even upon removal of selection pressure). High productivity would instead result of creation of an artificial transcriptionally highly active zone, or integration of plasmid in multiple copies, or multiplication of plasmid copies. As a consequence, if the strategy of attenuation of selection marker is to be used for the identification of "hot-spots", it is probably necessary to increase again selection stringency, for example by combining attenuation approaches, and to significantly increase screening capacity. The probability of isolating an "hyper-producer" is exceedingly low (Seth et al., 2007), thus requiring screening of thousands clones (Hacker et al., 2009).

Because transcription is usually considered to be the dominant factor controlling expression (Barnes et al., 2003), we focused our work on the promoter driving selection pressure. This genetic element was easily modifiable, with regard to the laboratory techniques and the knowledge we had at our disposal. However, several other strategies have been explored by other teams to attenuate selection marker strength: use of mRNA destabilizing elements (Ng et al., 2007), use of non-ATG start codon (Blokland et al., 2007; Otte et al., 2009), use of least preferred codons (Chin et al., 2015; Westwood et al., 2010), use of weakened IRES (Chin et al., 2015; Ho et al., 2012; Kwaks et al., 2003, 2005; Ng et al., 2007; Otte et al., 2009; Yeo et al., 2017), use of 2A peptide (Z. Liu et al., 2017), use of protein destabilizing elements (Chin et al., 2015; Ng et al., 2007), and use of attenuated enzyme mutants (Caputo et al., 2021; L. Chen et al., 2004; Lin et al., 2019; Sautter & Enenkel, 2005). To our knowledge, there are no studies comparing the efficiency of these attenuation strategies; however, this might be relevant and could help to choose the most efficient approach. The methodology applied here for measurement of promoters' transcription level, *i.e.*, measurement of transcriptional activity over a transient gene expression assay, could be applied for this purpose. RT-qPCR method would of course not be appropriate (as some attenuation strategies are targeting protein expression or stability), but instead, PAC protein level could be measured by enzyme-linked immunosorbent assay (ELISA). This methodology could work for all attenuation strategies except attenuated enzyme mutants. Moreover, as reemergence duration of transfected bulk pool during selection seemed to be negatively correlated with *pac* expression level, one

can suggest to rank attenuation strategies according to reemergence duration of transfected bulk pool (as an alternative methodology, especially for in the case of use of attenuated enzyme mutants). This methodology was sometimes used in previously published reports as preliminary evidence of strength of selection stringency (Fan et al., 2011; Fan, Kadura, et al., 2013). Nevertheless, confirmation of productivity increase in stable expression, either with mini-pools or clones, is necessary.

3.3.7 Conclusion

In this third chapter, new promoters driving puromycin N-acetyltransferase expression have been evaluated. Three CHO K1 endogenous promoters have been identified and four partially deleted SV40E promoters have been designed. Lower transcriptional strength, compared to SV40 early promoter, was confirmed for all evaluated promoters. Several promoters led to an increased productivity of recombinant cell lines, which was confirmed on several recombinant proteins. Lowering the strength of promoter for expression of selective pressure resistance opens the door to more efficient cell line development processes, by fostering identification of high-producing clones while being compatible with reduced process timelines.

In the fourth chapter, the development steps that led to the implementation of the RT-qPCR method, which was necessary for transcriptional strength measurement of evaluated promoters, will be detailed.

3.3.8 Author's contributions

Literature review: Lucie Grindes[1,2] (ORCID ID: 0000-0002-6201-3355); **Methodology**: Lucie Grindes; **Investigation**: Lucie Grindes, Marine Amouriq[1] (significant technical help during first clone generation), Florence Courtin[1] and Olivia Clayton[1] (significant technical help during second clone generation), Camille Florimond[1] (significant technical help during expression level stability study, confirmation study), Christine Dinard[3] (RNA integrity analysis); **Analysis**: Lucie Grindes, Camille Florimond (expression level stability study and confirmation study); **Writing – Original draft and editing**: Lucie Grindes; **Writing – Review**: Lucie Grindes, Gilles Joucla[2] (ORCID ID: 0000-0003-4192-9628), Wilfrid Dieryck[2] (ORCID ID: 0000-0002-4609-8934), Cyrielle Corbin (ORCID ID: 0000-0002-6891-2511), Sébastien Ribault[5] (trade secret and intellectual property), Daniel Kasten[6] (trade secret and intellectual property); **Supervision**: Cyrielle Corbin (lead), Céline Raymond (supporting).

[1] Cell line development team, Process development department, Merck Biodevelopment

[2] Structure and activity of biological macromolecules team, Nutraceutical, medical and food applications pole, CBMN (UMR 5842 (CNRS, University of Bordeaux, Bordeaux INP))

[3] Analytical development department, Merck Biodevelopment

[4] Up-Stream process team, Process development department, Merck Biodevelopment

[5] Direction, Merck Biodevelopment

[6] Patents Life Science global team, MilliporeSigma

3.4 Reverse transcription – quantitative polymerase chain reaction method development

3.4.1 Introduction

Gene transcription level analysis was required twice in the work previously presented. First, analysis of puromycin acetyltransferase (*pac*) gene expression level was necessary to compare the strength of evaluated promoters for expression of resistance to puromycin. Second, analysis of rituximab heavy chain (*rituHC*) and rituximab light chain (*rituLC*) genes expression levels was required to correlate transcription level to translation level during the stability study of rituximab expression level on SV40E and SV40EΔ21bpRIII-II clones.

3.4.1.1 Choice of the technique for gene transcription level analysis

Gene expression level analysis is commonly performed using cDNA microarrays or reverse transcription quantitative polymerase chain reaction (RT-qPCR) techniques (Kozera & Rapacz, 2013; Radonić et al., 2004), or by high-throughput techniques such as RNA-sequencing (RNA-seq). Both cDNA microarrays and RT-qPCR technologies are rapid, medium-to-high throughput and can be automated. cDNA microarray technology allows the simultaneous analysis of thousands of genes in two differentially labelled RNA samples. On the other hand, RT-qPCR technique rather allows the parallel analysis of few genes in many different RNA samples (Derveaux et al., 2010; Vandesompele et al., 2002). RT-qPCR shows high sensitivity, good reproducibility, and a wide quantification range (Fleige & Pfaffl, 2006). It has thus become the gold standard method for quantification of RNA (Nolan et al., 2006; Streit et al., 2009).

In the context of the study presented here, high-throughput techniques, such as RNA-sequencing, had no use. Interest was focused on few genes, namely *pac*, *rituHC*, and *rituLC*, in a large number of samples. Therefore, RT-qPCR technique was chosen for gene transcription level analysis.

3.4.1.2 Quantitative PCR: insights on a widespread technique

Polymerase chain reaction (PCR) technique was described by Saiki *et al.* in 1985 and Mullis *et al.* in 1987 (Mullis & Faloona, 1987; Saiki et al., 1985). This technique enables *in vitro* amplification of DNA sequences by repeating denaturation, annealing, and extension reactions, and by using specific nucleotide primers and a DNA polymerase. The discovery of *Thermus aquaticus*, thermophile eubacteria, and the use of its thermostable DNA polymerase are at the origin of the technique development. A few years later, analysis of PCR kinetics was first performed by Higuchi *et al.* in 1992, leading to the quantitative PCR (qPCR) technique (Higuchi et al., 1992, 1993). qPCR technique was further improved with the development of hydrolysis probes, thanks to the demonstration of 5’ nuclease activity of *Taq* DNA polymerase by Holland *et al.* in 1991 (Holland et al., 1991). Nuclease activity of

Taq DNA polymerase enables the cleavage of a target specific probe during qPCR reaction, allowing the monitoring of a fluorescent signal, which is generated only in case of specific hybridization between probe and target (Livak et al., 1995; Raymaekers et al., 2009).

RT-qPCR is a succession of four steps: (i) the RNA extraction; (ii) the reverse transcription of RNA into complementary DNA (cDNA) by a reverse transcriptase; (iii) the amplification of the cDNA by a DNA polymerase during a PCR reaction; (iv) the detection and quantification of amplification products in real-time (Nolan et al., 2006).

In vitro amplification of cDNA is based on the repetition of three steps: (i) denaturation of double stranded DNA (dsDNA) to obtain single-stranded DNA (ssDNA), generally at 95°C; (ii) annealing of nucleotide primers complementary to a specific ssDNA sequence, generally between 40-65°C; (iii) extension by a thermostable DNA polymerase, generally between 55-72°C. Quantity of target sequence is duplicated in each cycle assuming 100% PCR reaction efficiency (Kozera & Rapacz, 2013).

Two different detection chemistries can be used for detection and quantification of amplification products: "TaqMan"-based detection and "SYBR® Green"-based detection (Huggett & Bustin, 2011; Thellin et al., 2009). "TaqMan"-based detection uses a hydrolysis probe specific to target gene to detect target as it accumulates during PCR cycles. The hydrolysis probe contains a fluorescent reporter dye on the 5' end and a quencher dye on the 3' end. When the probe is intact, the reporter dye emission is quenched by the quencher dye by fluorescence resonance energy transfer (FRET). After annealing of the nucleotide primers and hydrolysis probe to the target sequence, the probe is cleaved by the 5' nuclease activity of Taq DNA polymerase during extension. Once separated from the quencher, the reporter dye emits its fluorescence. Increase in fluorescence intensity is proportional to the amount of cDNA target. "SYBR® Green"-based detection uses a dsDNA binding dye to detect target as it accumulates during PCR cycles. When added to the sample, the dye bounds to all dsDNA. cDNA target sequence is amplified during repetition of PCR cycles. The dye then binds to each new copy of dsDNA, resulting in an increase in fluorescence intensity proportional to the amount of cDNA target (Huggett & Bustin, 2011). In the present work, "SYBR® Green" dye was chosen as detection chemistry.

Analysis of RT-qPCR experiments is commonly performed either by absolute quantification or by relative quantification. Absolute quantification method determines the input target copy number by relating the PCR (quantification cycle (C_q)) signal to a standard curve. This analysis method is required when the determination of the exact target copy number is necessary. However, reporting the relative change in gene expression is most of the time both sufficient and more relevant. Therefore, RT-qPCR data are commonly analyzed with relative quantification method. Relative quantification method determines the change in expression of the target gene relative to a control group (Raymaekers et al., 2009). For example, but not limited to, the $2^{-\Delta\Delta Ct}$ calculation method allows analysis of relative changes in gene expression in RT-qPCR assays (Livak & Schmittgen, 2001).

3.4.2 Methodology: guidelines for RT-qPCR method development

RT-qPCR method development was conducted following the minimum information for publication of quantitative PCR experiments (MIQE) guidelines and the real-time PCR data markup language (RDML) (Bustin, 2010; Bustin et al., 2009, 2010; Lefever et al., 2009).

In late 2000s, the scientific community witnessed technical deficiencies in many publications. These deficiencies were namely related to sample preparation and storage, nucleic acid quality, choice of reverse transcription primers, choice of primers and probes for qPCR, and data analysis (Bustin, 2010; Bustin et al., 2009; Bustin & Nolan, 2017). Therefore, the MIQE guidelines aim to provide a methodology to produce more reliable and consistent data from RT-qPCR experiments (Taylor et al., 2019). They namely cover every aspect of sample acquisition, assay design and validation, and data analysis (Bustin et al., 2010).

RT-qPCR method development was conducted following the steps described in Figure 45.

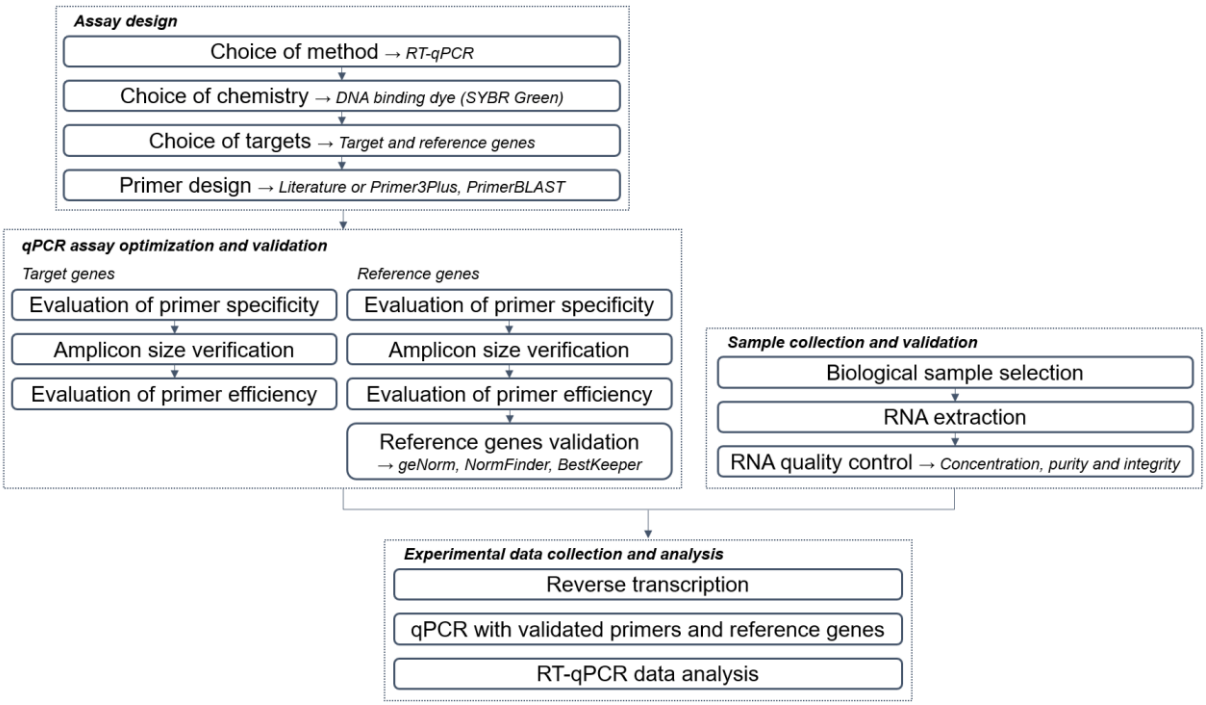


Figure 45: Workflow for RT-qPCR method development

Workflow for RT-qPCR method development is divided into four main parts: assay design (choice of method, chemistry and targets, and primer design), qPCR assay optimization and validation (evaluation of primer specificity, amplicon size verification and evaluation of primer efficiency, both for target genes and reference genes, as well as reference genes validation), sample collection and validation (RNA extraction and RNA quality control) and lastly experimental data collection and analysis. Adapted from Nolan et al., 2006.

3.4.3 Results

3.4.3.1 RNA extraction

Biological samples are lysed and homogenized in a denaturing guanidine-isothiocyanate-containing buffer, which denatures proteins and then inactivates RNases. The lysate is then passed through a spin column that binds and eliminates genomic DNA. Ethanol is added to the flow-through to provide appropriate binding conditions for RNA. Sample is then applied to a spin column, where total RNA binds to the membrane and contaminants are eliminated during repeated washing steps. RNA molecules longer than 200 nucleotides are finally eluted in water.

3.4.3.2 RNA quality control

3.4.3.2.1 Principle

RNA quality control is a requirement for many further RNA-based downstream applications, such as RT-qPCR, micro-arrays, ribonuclease-protection assays, *in situ* hybridization, Northern blot analysis, RNA mapping, *in vitro* translation, cDNA library construction, or RNA sequencing (Bustin et al., 2009; Fleige & Pfaffl, 2006; A. Schroeder et al., 2006). RNA quality control refers to RNA concentration, purity, and integrity measurements. RNA quality (especially purity and integrity) influences the accuracy of gene expression measurement. It has been shown that the variability of RT-qPCR assays is increased with decreasing RNA quality and that concerns about RNA quality rise when length of the amplified target length is increased (Fleige & Pfaffl, 2006).

RNA is very sensitive to degradation, mainly due to inadequate sample handling and storage conditions (Pérez-Novo et al., 2005). Furthermore, RNA extraction method should allow purification of intact total RNA free of protein, free of genomic DNA, free of enzymatic inhibitors for reverse transcription and PCR reactions, free of any substances which complex essential reaction co-factors, and free of nucleases (Bustin & Nolan, 2004; Fleige & Pfaffl, 2006).

RNA concentration and purity are commonly measured using spectrophotometric methods. Nucleic acids (both DNA and RNA) absorb in ultra-violet at $\lambda_{260\ nm}$. Nucleic acid concentration is then calculated with Beer-Lambert law equation, which relates absorbance to concentration. Additional measurements of nucleic acid solution at $\lambda_{280\ nm}$ (specific wavelength for proteins) and $\lambda_{230\ nm}$ (specific wavelength for background and possible contaminants) enables to calculate A_{260}/A_{280} and A_{260}/A_{230} ratios, which indicates RNA purity. More precisely, A_{260}/A_{280} indicates potential protein contamination. Pure RNA solution has a A_{260}/A_{280} ranging from 1.8 to 2.2. A_{260}/A_{230} ratio indicates potential contamination by other components (salts, isothiocyanate, guanidine, carbohydrates, urea, EDTA, …). Pure RNA solution has a A_{260}/A_{230} ranging from 2.0 to 2.2 (Fleige & Pfaffl, 2006).

RNA integrity is commonly assessed either by denaturating agarose gel-electrophoresis or by microfluidic capillary electrophoresis performed on lab-on-chip technologies, such as 2100 Bioanalyzer

(Agilent Technologies), Experion™ Automated Electrophoresis System (Bio-Rad Laboratories) or LabChip GXII Touch HT (Perkin Elmer). In both techniques, RNA integrity measurement relies on the inspection of 18S and 28S ribosomal RNA bands (Fleige & Pfaffl, 2006). While denaturating agarose gel-electrophoresis is a tedious and low-throughput technique that requires important amounts of sample, capillary electrophoresis is the most suitable method for high-throughput and objective analysis of RNA integrity (Fleige et al., 2006; A. Schroeder et al., 2006).

Capillary electrophoresis enables electrophoretic separation of very small amount of RNA sample (down to 200 pg total RNA) on a micro-fluidic chip according to their molecular weight. Detection of RNA fragments is based on laser-induced fluorescence detection. RNA band sizes are estimated thanks to the use of an RNA ladder as a mass and size standard. Depending on the equipment, RNA integrity is referred to as RNA Integrity Number (RIN) (2100 Bioanalyzer (Agilent Technologies)) or RNA Quality Score (RQS) (LabChip GXII Touch HT (Perkin Elmer), used in this study). Calculation of RQS considers the 18S and 28S peak areas, the 18S and 28S peak heights and the fast and total region areas (Nolan et al., 2006; A. Schroeder et al., 2006) (Figure 46-A). RQS and RIN values are given with a 1-10 scale rating, with 1 being the most degraded profile and 10 being the most intact (Figure 46-B). It is generally accepted that RQS or RIN values above 8.0 indicate intact, high quality RNA samples, between 5.0 and 8.0 moderately degraded RNA samples, and below 5.0 degraded samples (Fleige & Pfaffl, 2006; Padhi et al., 2018). Even if experimental protocols should be designed to obtain high-quality RNA samples, moderately degraded RNA samples can be analyzed if PCR products are kept short (below 250 bp) and if target gene expression is normalized with one or more reference genes (Fleige & Pfaffl, 2006; Nolan et al., 2006).

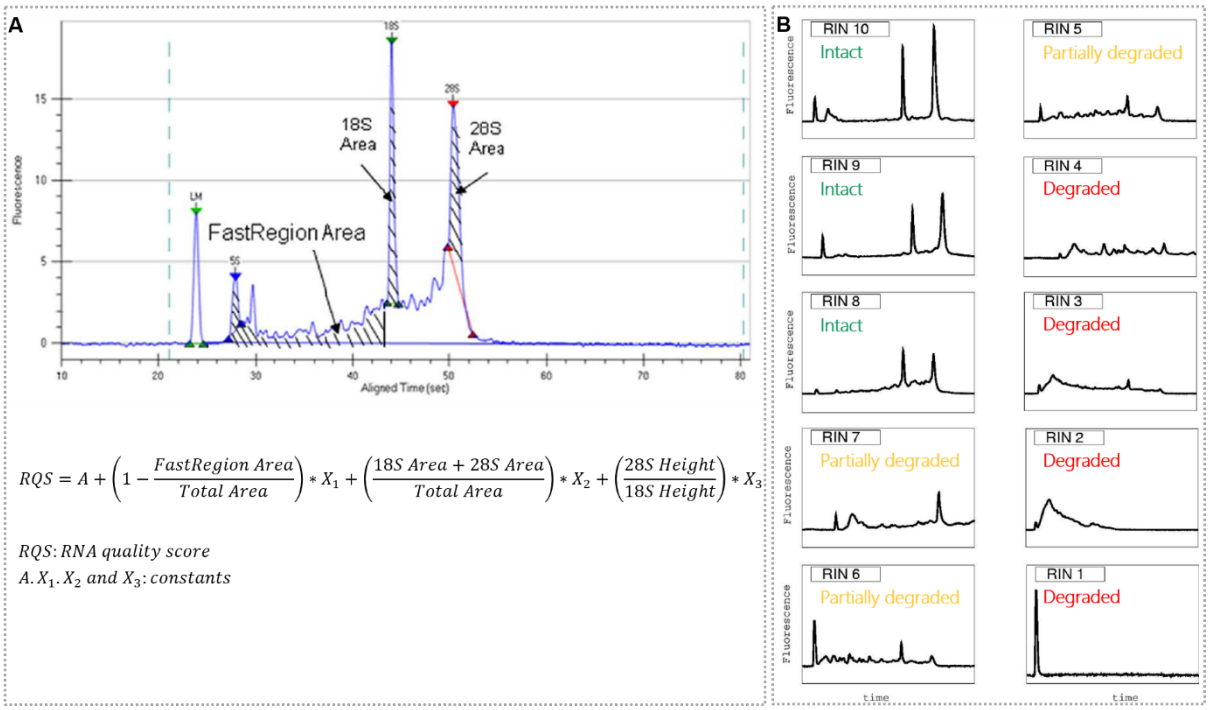


Figure 46: RNA integrity calculation and examples

(A) *Equation for RNA integrity calculation. Calculation of RIN or RQS considers the 18S and 28S peak widths, the 18S and 28S peak heights and the fast and total region widths.* ***(B)*** *RIN scores with associated chromatograms. RIN or RQS are given with a 1-10 scale rating, with 1 being the most degraded profile and 10 being the most intact. RIN or RQS above 8.0 indicate intact, high quality RNA samples, between 5.0 and 8.0 moderately degraded RNA samples, and below 5.0 degraded samples. The gradual degradation of ribosomal RNA is reflected by a continuous shift towards shorter fragment sizes. Adapted from Caliper LiveScience's application note 402 and Schroeder et al., 2006.*

3.4.3.2.2 Results

RNA concentration, purity and integrity results of RNA samples used for reference gene evaluation are given in Table 12: Quality control of extracted RNA (concentration, purity, and integrity)Table 12. A_{260}/A_{280} and A_{260}/A_{230} ratios are within defined range for all RNA samples. Analysis of RNA quality shows RQS ranging from 3.2 to 9.5. All RNA samples have a RQS above 5.0, indicating partially degraded or even a RQS above 8.0, indicating intact RNA. Though two RNA samples (fed-batch PD_{12}, both biological replicates) have a RQS around 3.25, meaning that RNA is degraded. This is explained by the very low viability of cell cultures (respectively 8.2 and 8.5%, end of decline phase).

Table 12: Quality control of extracted RNA (concentration, purity, and integrity)

TF: T-flask; PD: production day; ST: spin tube; SF: shake flask.

A_{260}/A_{280} and A_{260}/A_{230} ratios indicate RNA purity. More precisely, A_{260}/A_{280} indicates potential protein contamination. Pure RNA solution has a A_{260}/A_{280} ranging from 2.0 to 2.2. A_{260}/A_{230} ratio indicates potential contamination by other components (salts, isothiocyanate, guanidine, carbohydrates, urea, EDTA, …). Pure RNA solution has a A_{260}/A_{230} ranging from 2.0 to 2.2.

RNA quality score (RQS) indicates RNA integrity. Intact RNA has a RQS ranging from 8.0 to 10; partially degraded RNA has a RQS ranging from 5.0 to 8.0; degraded RNA has a RQS below 5.0.

Sample	[RNA] [ng/µL]	A_{260}/A_{280} [-]	A_{260}/A_{230} [-]	RQS [-]
Transfection Negative control TF_{01}	639.7	2.1	2.1	6.9
Transfection Negative control TF_{02}	773.0	2.1	2.2	6.4
Transfection Positive control TF_{01}	807.8	2.1	2.3	6.2
Transfection Positive control TF_{02}	711.4	2.1	2.3	6
Fed-batch PD_{05} ST_{01}	994.9	2.1	1.9	8.4
Fed-batch PD_{05} ST_{02}	942.8	2.1	2.3	8.3
Fed-batch PD_{07} ST_{01}	2487.6	2.1	2.3	9.1
Fed-batch PD_{07} ST_{02}	1206.9	2.1	2.3	8.8
Fed-batch PD_{09} ST_{01}	1044.3	2.1	2.2	8.3
Fed-batch PD_{09} ST_{02}	1120.7	2.1	2.2	8.8
Fed-batch PD_{12} ST_{01}	325.9	2.1	1.9	3.3
Fed-batch PD_{12} ST_{02}	495.1	2.0	2.1	3.2
Batch exponential phase SF_{01}	679.0	2.1	2.1	8.8
Batch exponential phase SF_{02}	571.9	2.1	2.1	9.0
Batch stationary phase	550.0	2.1	2.1	9.5

3.4.3.3 cDNA synthesis by reverse transcription

RNA sample is briefly incubated with a genomic DNA (gDNA) eliminating buffer to remove contaminating gDNA. RNA sample is then reverse transcribed into cDNA using a master mix composed of reverse transcriptase enzyme (contains an optimized blend of Omniscript Reverse Transcriptase, Sensiscript Reverse Transcriptase and RNase inhibitor), reverse transcription buffer (containing dNTPs) and reverse transcription primer mix (containing an optimized blend of oligo-dT and random primers dissolved in water). Reverse transcription reaction is then inactivated.

The reverse transcription reaction is though poorly understood and therefore stands as the most uncertain step in gene expression analysis. Yield of reverse transcription reaction can namely be impacted by reaction inhibitors present in RNA sample, RNA secondary and tertiary structures, variation in priming efficiencies, and properties of the reverse transcriptase (Bustin et al., 2015; Nolan et al., 2006; Stahlberg et al., 2004).

3.4.3.4 Quantitative PCR

3.4.3.4.1 Primer design for plasmid genes

Several rules are jointly considered for oligonucleotide design for qPCR: primer length between 18-24 bp, primer melting temperature (T_m) between 58-60°C, primer GC content between 30-70%, no more than two C or G in last five positions at 3' end of primer, no more than four constitutive guanines, no primer dimer, length of amplicon between 75-400 bp (Raymaekers et al., 2009). Nonetheless, length of amplicon is more frequently reduced between 75-250 bp, or even between 75-150 bp, mainly because of the difficulty of obtaining intact RNA samples (Fleige & Pfaffl, 2006). Moreover, primers are often designed either to span introns or to flank introns, to avoid PCR amplification from contaminant genomic DNA.

Primers were designed for puromycin acetyltransferase (*pac*), rituximab heavy chain (*rituHC*) and rituximab light chain (*rituLC*) genes and transcripts detection.

Nucleic acid structure (namely stem and loop secondary structures of DNA and RNA) might impact reverse transcription and qPCR reactions efficiencies (Bustin et al., 2009). Hence, design of oligonucleotides should take the folding of RNA and cDNA templates into consideration, to avoid highly structured regions where hybridization of primers to target would prove inefficient (Bustin et al., 2009; Nolan et al., 2006). MIQE guidelines suggest the use of mfold for RNA or DNA folding prediction (Bustin et al., 2009). mfold is a computational molecular biology web server. It was first developed in the late 1980s for RNA folding prediction, and then updated in the late 1990s for both RNA and DNA folding prediction (Zuker, 2003). Once zones implied in secondary structures are identified, primer design can be conducted, with the help of tools such as Primer3Plus or PrimerBLAST (Untergasser et al., 2012).

DNA sequences of pac, rituHC and rituLC were analyzed on mfold, with following parameters: linear sequence, folding temperature 60°C, 50 mM Na^+, 1.5 mM Mg^{2+}.

Predicted DNA structures of *rituHC* and *rituLC* are respectively shown on Figure 47 and Figure 48. *RituHC* and *rituLC* display a limited number of zones involved in stem or loop secondary structures. Therefore, these zones were systematically excluded for primer design, corresponding to: <72-109 b>, <449-501 b>, <544-569 b>, <583-601 b>, <646-657 b>, <1083-1093 b>, <1300-1318 b>, <1320-1344 b> for *rituHC*, and to: <24-59 b>, <133-187 b>, <257-266 b>, <301-314 b>, <346-365 b>, <401-430 b>, <443-451 b>, <672-700 b> for *rituLC*. Primer design was performed on Primer3Plus, with default parameters except for amplicon size, which was limited between 75-150 bp. Four pairs of primers were designed for both *rituHC* and *rituLC*. Primer's characteristics (T_m, length, GC% and amplicon size) are listed in Table 13.

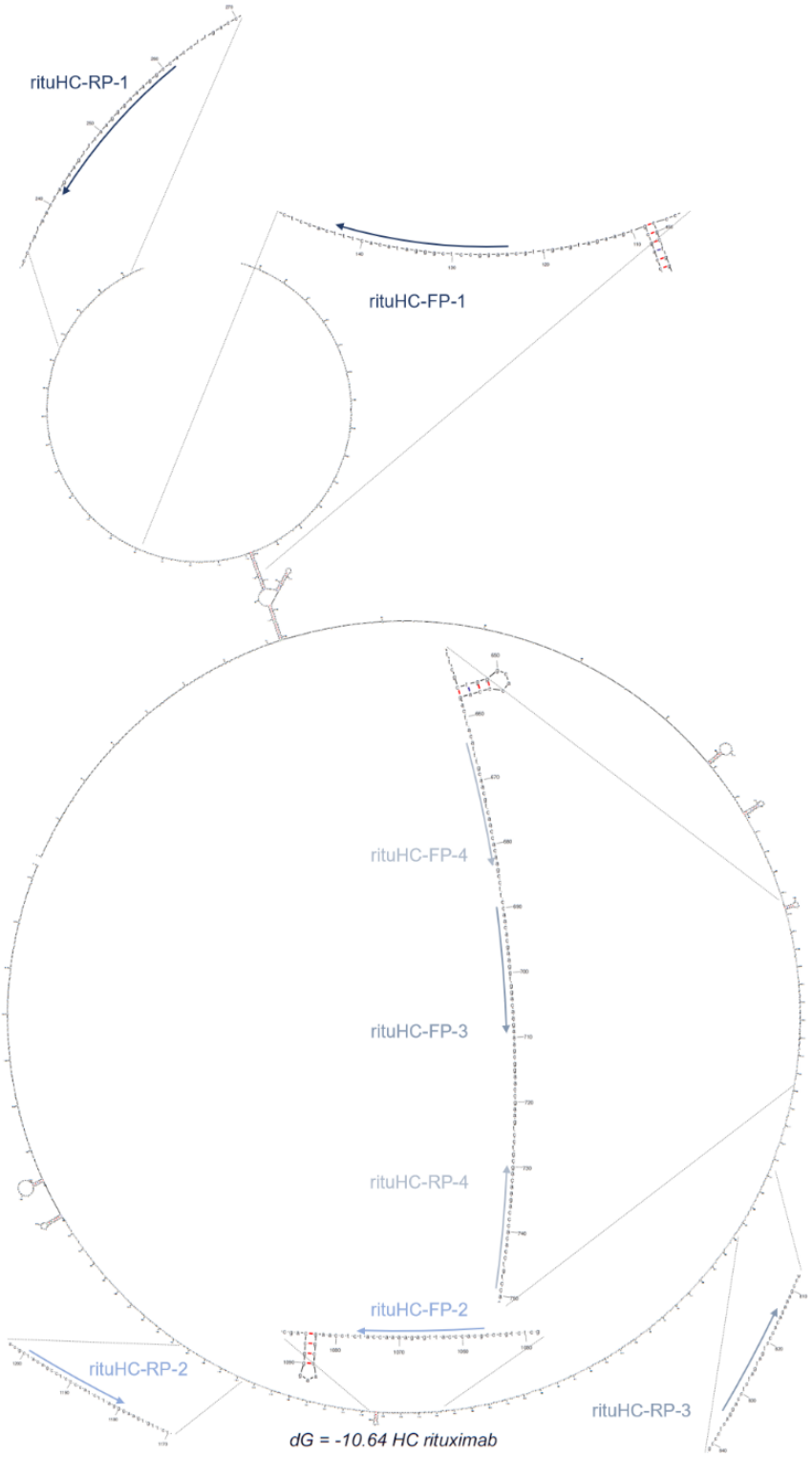


Figure 47:Prediction DNA folding of rituximab heavy chain at 60°C

DNA structure of rituximab heavy chain at 60°C was analyzed on mfold, with these parameters: linear sequence, folding temperature 60°C, 50 mM Na^+, 1.5 mM Mg^{2+}. Following zones were excluded for primer design: <72-109 b>, <449-501 b>, <544-569 b>, <583-601 b>, <646-657 b>, <1083-1093 b>, <1300-1318 b>, <1320-1344 b>. The four designed forward and reverse primer pairs (rituHC-FP/RP-1, rituHC-FP/RP-2, rituHC-FP/RP-3, and rituHC-FP/RP-4) are represented.

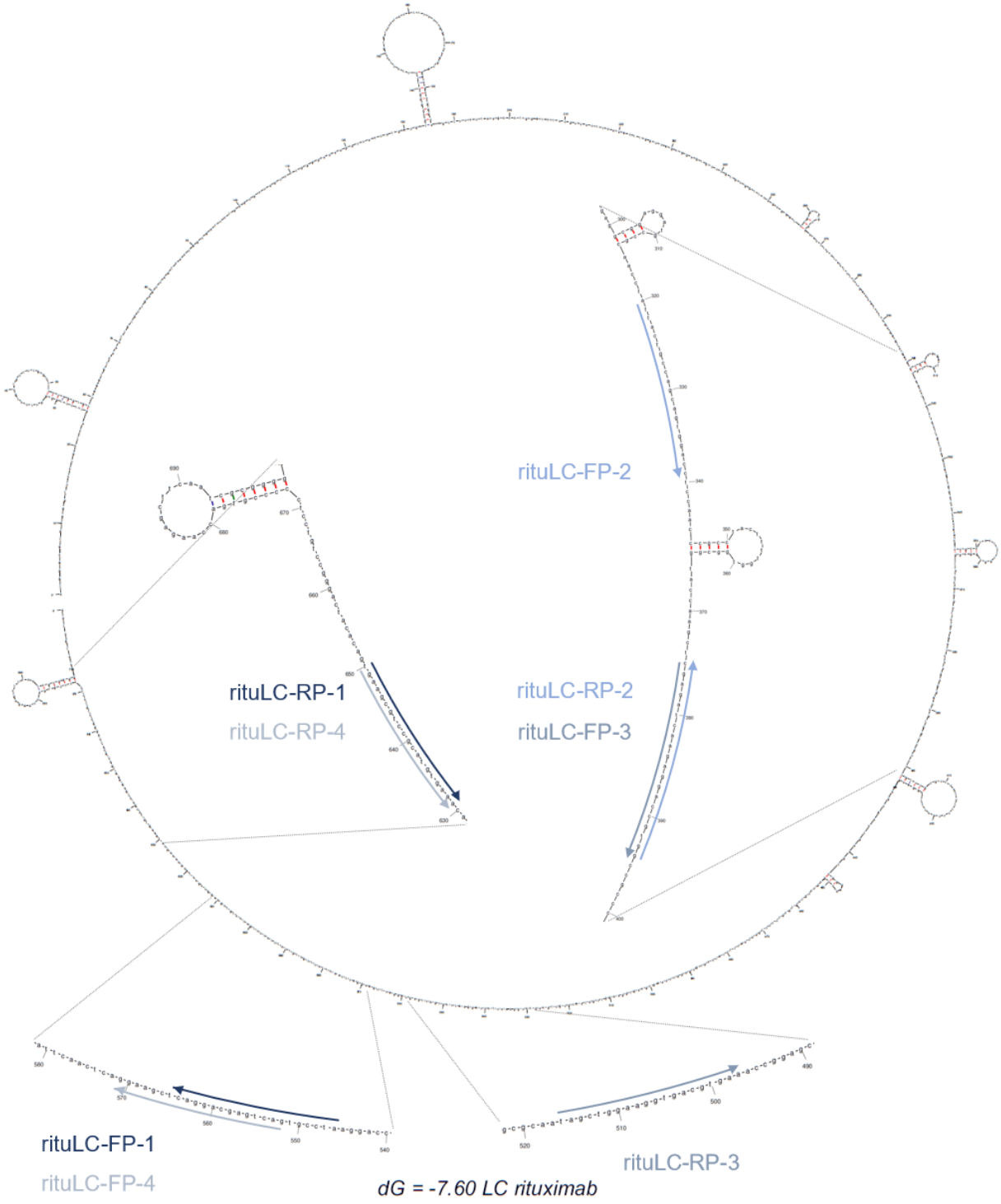


Figure 48:Prediction DNA folding of rituximab light chain at 60°C

DNA structure of rituximab light chain at 60°C was analyzed on mfold, with these parameters: linear sequence, folding temperature 60°C, 50 mM Na^+, 1.5 mM Mg^{2+}. Following zones were excluded for primer design: <24-59 b>, <133-187 b>, <257-266 b>, <301-314 b>, <346-365 b>, <401-430 b>, <443-451 b>, <672-700 b>. The four designed forward and reverse primer pairs (rituLC-FP/RP-1, rituLC-FP/RP-2, rituLC-FP/RP-3, and rituLC-FP/RP-4) are represented.

Two DNA structures, with equivalent free energy (ΔG), were predicted for *pac*. They are respectively shown on Figure 49 and Figure 50. *pac* displays a highly structured DNA with many stem-loop

structures, explained by a very high content of guanines and cytosines (GC% of 72.8). The two predicted DNA structures are identical until the stem-loop finishing on base 374. Structure 1 is then composed of two stem structures followed by three stem-loop structures, while structure 2 is then composed of four stem-loop structures. Primer design was performed trying to avoid zones involved in stem or loop secondary structures, corresponding to: <30-40 b>, <69-108 b>, <189-222 b>, <232-245 b>, <255-267 b>, <273-289 b>, <296-314 b>, <342-352 b>, <356-374 b>, <384-427 b>, <446-475 b>, <522-541 b>, <548-571 b> for *pac* structure 1, and to: <30-40 b>, <69-108 b>, <189-222 b>, <232-245 b>, <255-267 b>, <273-289 b>, <296-314 b>, <342-352 b>, <356-374 b>, <384-395 b>, <415-440 b>, <446-475 b>, <522-541 b>, <548-571 b> for *pac* structure 2. Primer design was performed on Primer3Plus, with default parameters except for amplicon size, which was limited between 75-250 bp. Five pairs of primers were designed for *pac*; their positions on DNA folding predictions 1 and 2 are shown on Figure 49 and Figure 50. Primer's characteristics (T_m, length, GC% and amplicon size) are listed in Table 13.

As primer design was complex in *pac* coding sequence due to highly structured DNA at 60°C, another primer design was performed in *pac* non-coding sequence. One DNA structure was predicted for *pac* non-coding sequence at 60°C as shown on Figure 51. *pac* non-coding sequence displays only two stem-loop structures of less than 15 nucleotides each. Therefore, these zones were systematically excluded for primer design, corresponding to: <205-219 b> and <363-375 b>. Primer design was performed on Primer3Plus, with default parameters except for amplicon size, which was limited between 75-150 bp. Three pairs of primers were designed; their positions on DNA folding prediction are shown on Figure 51. Primer's characteristics (T_m, length, GC% and amplicon size) are listed in Table 13.

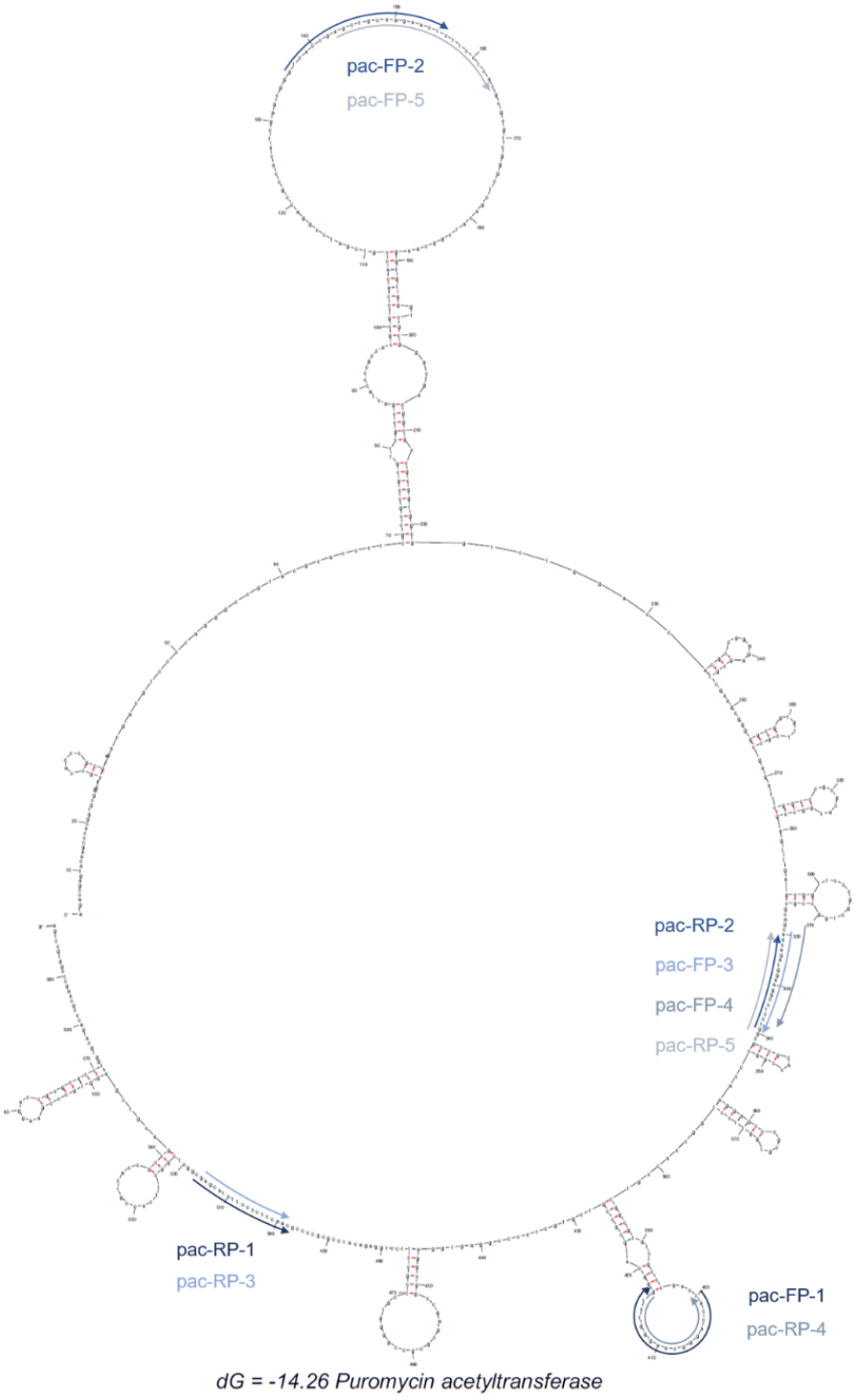


Figure 49: Prediction 1 of DNA folding of puromycin acetyltransferase coding sequence at 60°C

DNA structure of puromycin acetyltransferase coding sequence at 60°C was analyzed on mfold, with these parameters: linear sequence, folding temperature 60°C, 50 mM Na^+, 1.5 mM Mg^{2+}. Following zones were excluded for primer design: <30-40 b>, <69-108 b>, <189-222 b>, <232-245 b>, <255-267 b>, <273-289 b>, <296-314 b>, <342-352 b>, <356-374 b>, <384-427 b>, <446-475 b>, <522-541 b>, <548-571 b>. The five designed forward and reverse primer pairs (pac-FP/RP-1, pac-FP/RP-2, pac-FP/RP-3, pac-FP/RP-4, and pac-FP/RP-5) are represented.

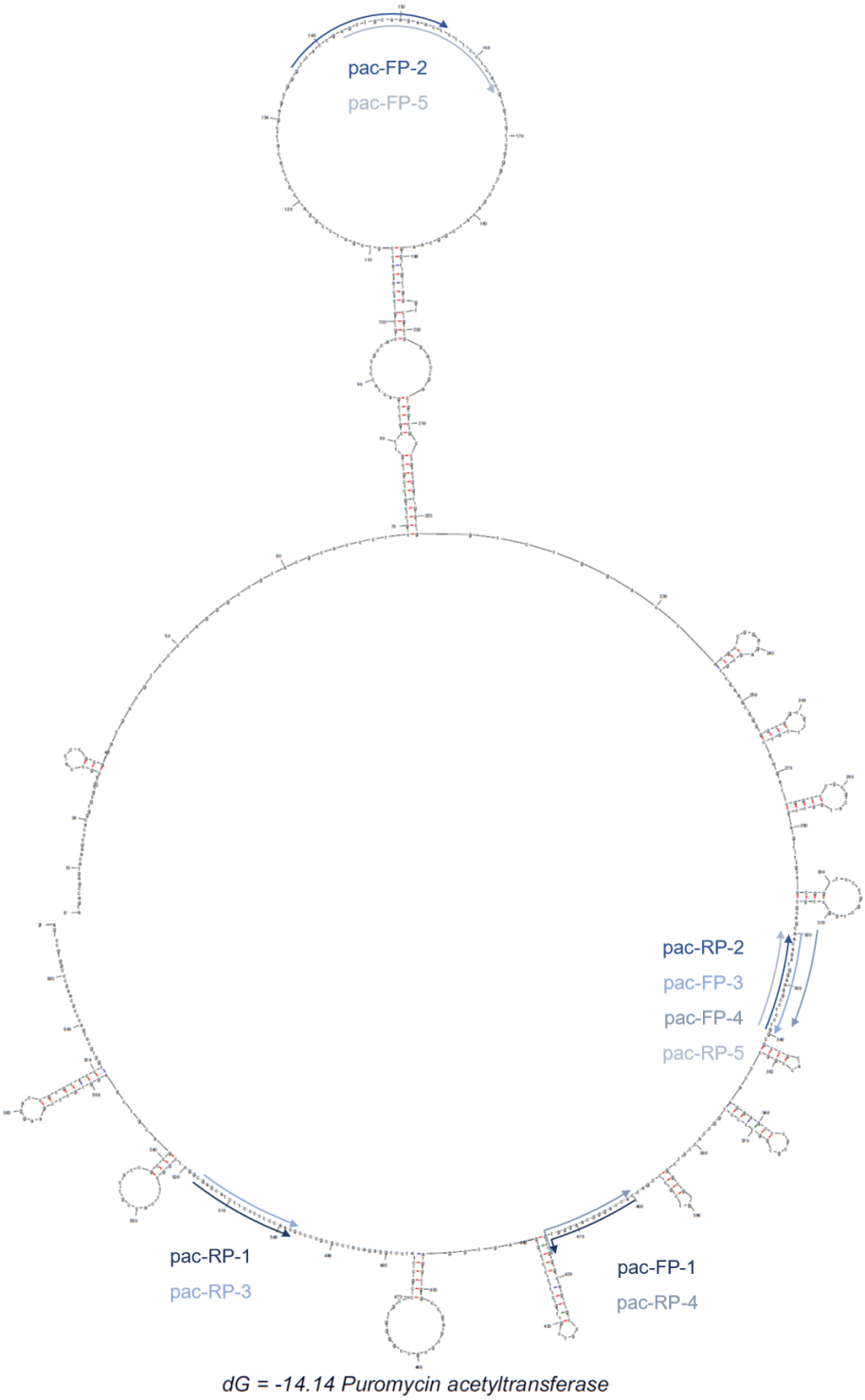


Figure 50: Prediction 2 of DNA folding of puromycin acetyltransferase coding sequence at 60°C

DNA structure of puromycin acetyltransferase coding sequence at 60°C was analyzed on mfold, with these parameters: linear sequence, folding temperature 60°C, 50 mM Na^+, 1.5 mM Mg^{2+}. Following zones were excluded for primer design: <30-40 b>, <69-108 b>, <189-222 b>, <232-245 b>, <255-267 b>, <273-289 b>, <296-314 b>, <342-352 b>, <356-374 b>, <384-395 b>, <415-440 b>, <446-475 b>, <522-541 b>, <548-571 b>. The five designed forward and reverse primer pairs (pac-FP/RP-1, pac-FP/RP-2, pac-FP/RP-3, pac-FP/RP-4, and pac-FP/RP-5) are represented.

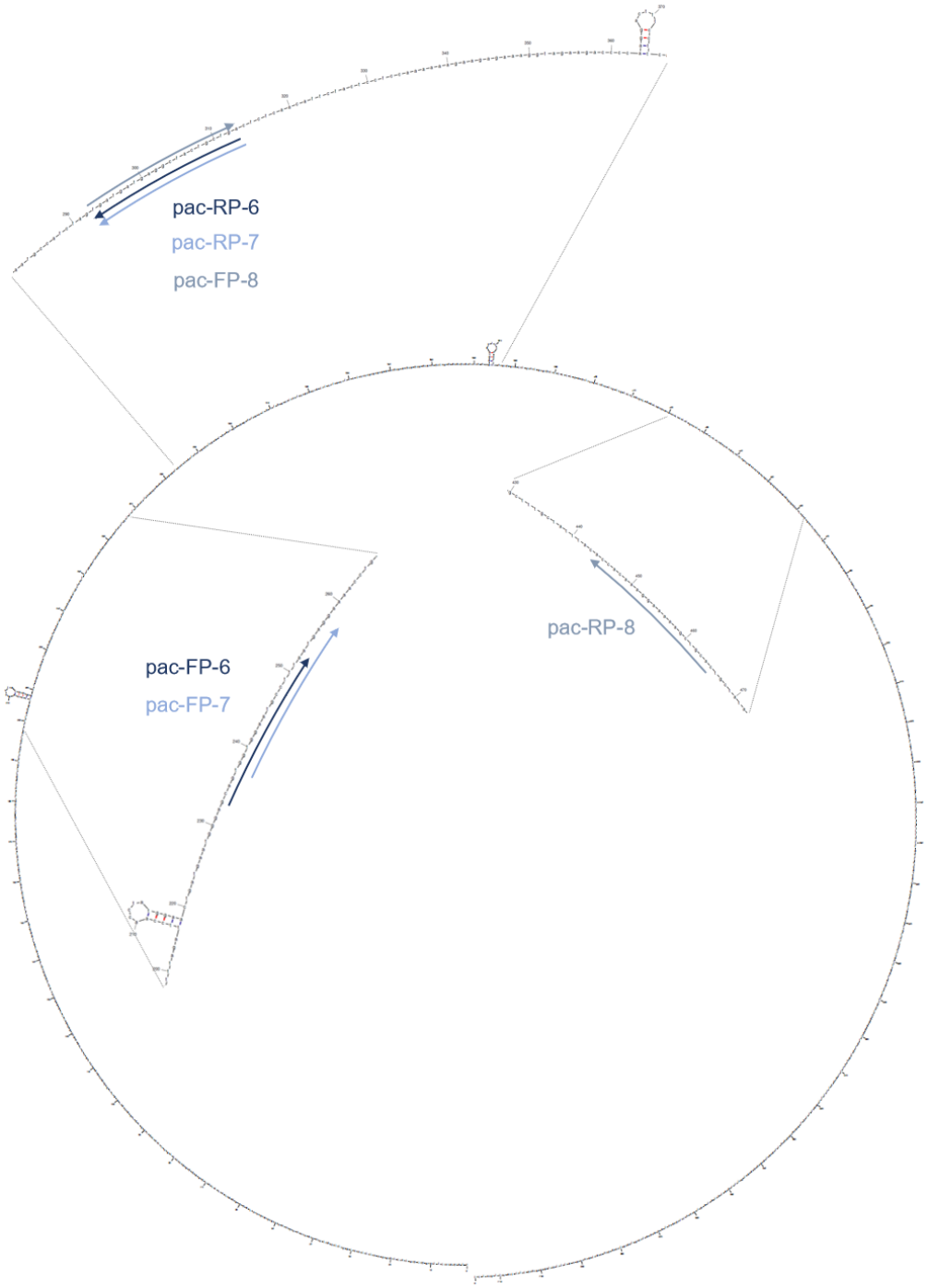


Figure 51: Prediction of DNA folding of puromycin acetyltransferase non-coding sequence at 60°C

DNA structure of puromycin acetyltransferase non-coding sequence at 60°C was analyzed on mfold with these parameters: linear sequence, folding temperature 60°C, 50 mM Na^+, 1.5 mM Mg^{2+}. Following zones were excluded for primer design: <205 219 b>, <363 375 b>. The three designed forward and reverse primer pairs (pac FP/RP 6, pac FP/RP 7, and pac FP/RP 8) are represented.

Table 13: Plasmid genes and associated primer sequences

FP: forward primer; RP: reverse primer; T_m: melting temperature; GC%: percentage in guanines and cytosines; bp: base pair.

pac: puromycin acetyltransferase; rituHC: rituximab heavy chain; rituLC: rituximab light chain.

Several rules are jointly considered for oligonucleotide design for qPCR: primer length between 18-24 bp, primer T_m between 58-60°C, GC content between 30-70%, no more than two C or G in last five positions at 3' end of primer, no more than four constitutive guanines, no primer dimer, length of amplicon between 75-400 bp (Raymaekers et al., 2009). *Nonetheless, length of amplicon is more frequently reduced between 75-250 bp, or even between 75-150 bp, mainly because of the difficulty of obtaining intact RNA samples* (Fleige & Pfaffl, 2006).

Primer3Plus was used for primer design, with default parameters, except amplicon size between 75-250 bp.

Gene	FP name	FP sequence (5'-3')	FP T_m [°C]	FP length [bp]	FP GC%	RP name	RP sequence (5'-3')	RP T_m [°C]	RP length [bp]	RP GC%	Amplicon size [bp]
pac	pac-FP-1	caccagggcaagggtctg	63	18	66.7	pac-RP-1	gctcgtagaaggggaggttg	61	20	60	118
pac	pac-FP-2	gtcaccgagctgcaagaact	61	20	55	pac-RP-2	caggaggccttccatctgt	60	19	57.9	204
pac	pac-FP-3	aacagatggaaggcctcctg	59	20	55	pac-RP-3	ctcgtagaaggggaggttgc	60	20	60	197
pac	pac-FP-4	caacagatggaaggcctcc	59	19	57.9	pac-RP-4	cagaccttgccctggtg	61	18	66.7	101
pac	pac-FP-5	agctgcaagaactcttcctcac	60	22	50	pac-RP-5	caggaggccttccatctgtt	62	20	55	196
pac	pac-FP-6	gcagtggtggaatgccttta	61	20	50	pac-RP-6	tcagcagtagcctcatcatca	60	21	47.6	81
pac	pac-FP-7	tggtggaatgcctttaatgag	60	21	42.9	pac-RP-7	tcagcagtagcctcatcatca	60	21	47.6	77
pac	pac-FP-8	accacaaaggaaaaagctgcac	60	22	45.5	pac-RP-8	agcagacactctatgcctgt	58	20	50	135
rituHC	rituHC-FP-1	aaggcctcgggatacacttt	60.0	20	50	rituHC-RP-1	ggcctttcccttgaacttct	59.7	20	50	138
rituHC	rituHC-FP-2	gcacccattgagaaaaccat	59.8	20	45	rituHC-RP-2	atgtcggaggggtagaatcc	60.2	55	3	143
rituHC	rituHC-FP-3	caacacgaaggtggacaaga	59.7	20	50	rituHC-RP-3	cctggagatcatcagggtgt	59.9	20	55	145
rituHC	rituHC-FP-4	atttgcaacgtcaaccacaa	60.0	20	40	rituHC-RP-4	ggacaggtgtgggtcttgtc	60.4	20	60	86
rituLC	rituLC-FP-1	atccgtgactgagcaggact	59.9	20	55	rituLC-RP-1	acttcgcaggcgtacacttt	59.9	20	50	105
rituLC	rituLC-FP-2	attactgccagcagtggacc	60.0	20	55	rituLC-RP-2	ccacggtcctcttgatctcg	59.9	20	60	75
rituLC	rituLC-FP-3	cgagatcaagaggaccgtgg	59.9	20	60	rituLC-RP-3	atcgaccttccactgcactt	59.3	20	50	142
rituLC	rituLC-FP-4	gactgagcaggactcgaagg	59.8	20	60	rituLC-RP-4	acttcgcaggcgtacacttt	60.3	20	50	99

3.4.3.4.2 Primer design for reference genes

3.4.3.4.2.1 Normalization against reference genes

Precise normalization is essential for gene expression data analysis in RT-qPCR assays. Normalization should, among other things, compensate RT-qPCR variations, which are mainly due to: difference in amount of starting material between samples, difference in RNA extraction efficiency, difference in RNA purity and integrity, difference in reverse transcription reaction efficiency, template cDNA sample loading variation, difference in PCR reaction efficiency (Bustin et al., 2005; Chervoneva et al., 2010; Huggett et al., 2005; Kubista et al., 2006; Radonić et al., 2004). The use of reference genes (frequently misnamed "housekeeping genes", according to MIQE guidelines (Bustin et al., 2009)) is the most routinely applied method for RT-qPCR normalization (Kozera & Rapacz, 2013; Kubista et al., 2006; Raymaekers et al., 2009).

A gene should meet several criteria before being considered as a reference gene. Its expression level should have minimal innate variability and should be unaffected by experimental conditions; it should not be co-regulated with the target gene. Furthermore, its expression level should be similar to the target gene and should vary in accordance with the experimental error associated with the technique (Chervoneva et al., 2010; Kozera & Rapacz, 2013). Metabolism-related genes involved in essential processes for cellular maintenance, or structure-related genes were first used as reference genes. To name a few, glyceraldehyde-3-phosphate dehydrogenase gene (*gapdh*), β-actin gene (*actb*), β-2-microglobulin gene (*b2m*), or ribosomal RNA's 18S and 28S genes were frequently mentioned in the literature (Radonić et al., 2004; Thellin et al., 1999). However, many studies revealed that traditional reference genes had significant variability in expression level and that normalization against only one reference gene gave inconsistent results. Consequently, many other reference genes candidates have been widely studied, such as (but not limited to) hypoxanthine-guanine phosphoribosyl-transferase (*hprt*), peptidyl prolyl isomerase A (*ppia*), glucose 6-phosphate dehydrogenase (*g6pdh*), TATA-box binding protein (*tbp*), α-tubulin (*tub*), porphobilinogen deaminase (*pbgd*), ribosomal protein L13 (*l13*) (Radonić et al., 2004).

Therefore, there is a consensus among the scientific community that the final choice of reference genes should be systematically validated for each RT-qPCR study (Bustin et al., 2009; Chervoneva et al., 2010; Kozera & Rapacz, 2013; Radonić et al., 2004). Moreover, normalization against three or more validated reference genes was considered as the most reliable and universally applicable method (Derveaux et al., 2010).

Authors generally advise to evaluate reference gene candidates on at least 10 different samples (Vandesompele et al., 2002). Several calculation methodologies have been described to identify the best reference genes: geNorm (Vandesompele et al., 2002), NormFinder (C. L. Andersen et al., 2004), and BestKeeper (Pfaffl et al., 2004) are some well-known examples. Details on these calculation

methodologies are given in section “Reference genes validation”. Evaluation of reference genes candidates was conducted following these guidelines.

3.4.3.4.2.2 Identification of candidate reference genes and primer sequences

Reference candidate genes and associated primers have been identified in the literature. Reference genes candidates and associated primer sequences are listed in Table 14 and primers’ characteristics are listed in Table 15. Compliance with previously described rules on primer design have been verified, as well as the samples (cell line and cell culture conditions) used for reference genes validation. Similarities with our *modus operandi* (e.g., for RNA extraction, reverse transcription reaction, or qPCR reaction) were considered as a plus.

Bahr et al. evaluated 12 reference genes candidates (*actr5*, *ampd2*, *ap3d1*, *cog1*, *eftud2*, *eif3i*, *hirip3*, *nap1l1*, *pabpn1*, *pms2*, *yaf2*, and *vezt*) based on microarray expression results and compared their variability to three traditional reference genes (*gapdh*, *actb*, and *b2m*) (Bahr et al., 2009). Samples represented batch and fed-batch cultures in 125-mL shake flasks, TPP bioreactors tubes and static six-well plates, from two parental and four IgG-producing CHO cell lines from CHO K1 lineage. Authors mention that Primer3 software has been used for primer design. Primers are 18-25 bp with T_m of 64-65°C, and amplicons are 156-249 bp, which agree with recommendations on primer design (Raymaekers et al., 2009). Additional similarity with our protocol is about RNA extraction, which was performed with RNeasy® Mini Kit (Qiagen). However, data analysis does not refer to advanced calculation methodologies (such as geNorm, NormFinder or BestKeeper) for identification of best reference genes.

More recently, Brown et al. evaluated 8 reference genes candidates (*pkar1a*, *fkbp1a*, *mmadhc*, *gnb1*, *tmed2*, *actb*, *pgam1* and *gapdh*) based on transcriptomic datasets (A. J. Brown et al., 2018). Samples represented batch cultures in static 24-well plates, shaking 96-well plates, and 125mL shake flasks from CHO cell lines (CHO lineage not mentioned). Experimental technique-specific (lipofection, nucleofection, hypothermia) samples were also included. Authors mention that PrimerBLAST software has been used for primer design. Primers are 19-23 bp with T_m of 60°C and span exon-exon boundaries. Amplicons are 95-187 bp, which agree with recommendations on primer design (Raymaekers et al., 2009). Additional similarity with our protocol is about RNA extraction, which was performed with RNeasy® Mini Kit (Qiagen), and reverse transcription reaction, which was performed with QuantiTect® Reverse Transcription Kit (Qiagen). Data analysis refers to both geNorm, NormFinder and BestKeeper.

Finally, Ma et al. re-evaluated 11 reference gene candidates previously assessed either by Bahr et al. or Brown et al, and evaluated 9 novel reference gene candidates (*eif3k*, *akr1a1*, *rps16*, *atp5mg*, *aprt*, *gpx1*, *pgk1*, *atp5f1* and *gusb*) (Ma et al., 2020). Samples represented long-term cultivation and fed-batch cultures in 125mL shake flasks, from one parental and five Fc-fusion protein-producing CHO cell lines from CHO K1 lineage. Authors mention that Primer3 software has been used for primer design. Primers are 18-25 bp with T_m of 60°C, and amplicons are 95-224 bp, which agree with recommendations on

primer design (Raymaekers et al., 2009). No additional similarity with our protocol is noticed. Data analysis refers to both geNorm, NormFinder, ΔCt (Silver et al., 2006) and BestKeeper.

Table 14: Reference genes candidates and associated primer sequences

Reference candidate genes and associated primers have been identified in the literature, from studies conducted in 2009, 2018 and 2020 (Bahr et al., 2009; A. J. Brown et al., 2018; Ma et al., 2020). *Compliance with previously described rules on primer design have been verified, as well as the samples (cell line and cell culture conditions) used for reference genes validation. Similarities with our modus operandi (e.g., for RNA extraction, reverse transcription reaction, or qPCR reaction) were considered as a plus.*

Gene description	RefSeq ID	Forward primer name	Forward primer sequence (5'-3')	Reverse primer name	Reverse primer sequence (5'-3')	Reference
Beta actin	NM_009735	actb-FP-1	gctcttttccagccttcctt	actb-RP-1	gagccagagcagtgatctcc	(Bahr et al., 2009)
Beta actin	NM_001244575.1	actb-FP-2	tgacccagatcatgtttgagacc	actb-RP-2	caggatggcatgagggagag	(A. J. Brown et al., 2018)
Beta actin	NM_001244575.1	actb-FP-3	ctgtgctatgttgccctgga	actb-RP-3	gccacaggattccatacccag	(Ma et al., 2020)
Actin related protein 5	NM_175419	actr5-FP-1	ccagaatggagaaggagctg	actr5-RP-1	ggcacaatgttccttgaggt	(Bahr et al., 2009; Ma et al., 2020)
Actin related protein 5	Not provided	actr5-FP-2	tgctggatgcctggtatgg	actr5-RP-2	acccaggctccactgtcttct	Internal
Aldo-keto reductase family 1 member A1	XM_003497372.4	akr1a1-FP-1	ggcttggaggtgactgcttat	akr1a1-RP-1	gaacctgccatctgagcaaga	(Ma et al., 2020)
Adenine phosphoribosyl-transferase	XM_003495090.2	aprt-FP-1	ctccttccgagcttccatcc	aprt-RP-1	ctagggaggggccaaacaag	(Ma et al., 2020)
ATP synthase peripheral stalk-membrane subunit b	XM_027393891.1	atp5f1-FP-1	gagcacatgatggactgggt	atp5f1-RP-1	tgcctgtttaactctctctggc	(Ma et al., 2020)
ATP synthase, H+ transporting, mitochondrial Fo complex subunit G	XM_007651137.2	atp5mg-FP-1	atggccaagttcgtccgtaa	atp5mg-RP-1	gggttgggggaaacagttca	(Ma et al., 2020)
Beta-2-microglobulin	NM_009735	b2m-FP-1	acggagtttacacccactgc	b2m-RP-1	cagacctccatgatgcttga	(Bahr et al., 2009)
Beta-2-microglobulin	NM_001246674.2	b2m-FP-2	tggctcacacggagtttaca	b2m-RP-2	catgtctcgttcccaggtga	(Ma et al., 2020)

Gene description	RefSeq ID	Forward primer name	Forward primer sequence (5'-3')	Reverse primer name	Reverse primer sequence (5'-3')	Reference
Component of oligomeric Golgi complex 1	NM_013581	cog1-FP-1	actagcctccagccagatca	cog1-RP-1	gcaggtgagtcgtcttcctc	(Bahr et al., 2009)
Eukaryotic translation initiation factor 3 subunit I	NM_018799	eif3i-FP-1	ccacaacttccaccaggatt	eif3i-RP-1	atgcggacgtaaccatcttc	(Bahr et al., 2009; Ma et al., 2020)
Eukaryotic translation initiation factor 3 subunit K	XM_007644386.2	eif3k-FP-1	agcccaaaaacatcgtggaga	eif3k-RP-1	tcattgaggaaggggcagaag	(Ma et al., 2020)
FK506 binding protein 1A	XM_003499952.2	fkbp1a-FP-1	ctctcgggacagaaacaagc	fkbp1a-RP-1	gacctacactcatctgggctac	(A. J. Brown et al., 2018; Ma et al., 2020; Yang et al., 2019)
Glyceraldehyde-3-phosphate dehydrogenase	NM_008084	gapdh-FP-1	gccaagagggtcatcatctc	gapdh-RP-1	ccttccacaatgccaaagtt	(Bahr et al., 2009)
Glyceraldehyde-3-phosphate dehydrogenase	NM_001244854.2	gapdh-FP-2	gaaagctgtggcgtgatgg	gapdh-RP-2	tacttggcaggtttctccag	(A. J. Brown et al., 2018)
Glyceraldehyde-3-phosphate dehydrogenase	Not mentionned	gapdh-FP-3	aactttggcattgtggaagg	gapdh-RP-3	acacgttgggggtaggaaca	Nguyen et al., 2019
Glyceraldehyde-3-phosphate dehydrogenase	NM_001244854.2	gapdh-FP-4	tgtaaagctcatttcctggtatgac	gapdh-RP-4	tgtgggggttattggacagg	Ma et al., 2020
G protein subunit beta 1	NM_001246701.1	gnb1-FP-1	ccatatgtttctttcccaatggc	gnb1-RP-1	aagtcgtcgtacccagcaag	Brown et al., 2018 ; Xiong et al., 2019 ; Ma et al., 2020
Glutathione peroxidase 1	NM_001256788.1	gpx1-FP-1	cggacatcaggagaatgcca	gpx1-RP-1	gtaaagagcgggtgagcctt	Ma et al., 2020
Glucuronidase beta	XM_027413803.1	gusb-FP-1	ttggtgccaactcctttcgt	gusb-RP-1	ttgtccctgcgtaccagttc	Ma et al., 2020
HIRA interacting protein 3	NM_172746	hirip3-FP-1	cgttatattcgggcttgtgg	hirip3-RP-1	ggtcgacctgaactgctgat	Bahr et al., 2009 ; Ma et al., 2020
Hypoxanthine phosphoribosyl-transferase 1	XM_027432079.1	hprt-FP-1	tacctcaccgctttctcgtg	hprt-RP-1	tcactaatcacgacgctggg	Ma et al., 2020

Gene description	RefSeq ID	Forward primer name	Forward primer sequence (5'-3')	Reverse primer name	Reverse primer sequence (5'-3')	Reference
Methylmalonic aciduria (cobalamin deficiency) with homocystinuria	XM_003513988.2	mmadhc-FP-1	tgtcacctcaatgggactgc	mmadhc-RP-1	caggtgcatcactactctgaaac	Brown et al., 2018 ; Ma et al., 2020
Poly(A) binding protein nuclear 1	NM_019402	pabpn1-FP-1	gtggccatcctaaagggttt	pabpn1-RP-1	cgggagctgttgtaattggt	Bahr et al., 2009 ; Ma et al., 2020
Phosphoglycerate mutase 1	XM_007617182.1	pgam1-FP-1	aggcgcaggtaaagatctgg	pgam1-RP-1	tgcgatccttactgatgttgc	Brown et al., 2018
Phosphoglycerate kinase 1	NM_001246725.1	pgk1-FP-1	caaggggaaccaagtccctc	pgk1-RP-1	gagctctagactggcaccac	Ma et al., 2020
Protein kinase, cAMP-dependant, regulatory, type I, alpha	XM_003500312.2	pkar1a-FP-1	gaatcctcatgggaagcactctg	pkar1a-RP-1	ccttgcaccacgatcttctg	Brown et al., 2018
Ribosomal protein S16	XM_003503853.3	rps16-FP-1	tgaagggtggtggtcatgtg	rps16-RP-1	tcagctacaagcagggttcg	Ma et al., 2020
Transmembrane p24 trafficking protein 2	XM_007648402.1	tmed2-FP-1	gcccacatggatgggacatac	tmed2-RP-1	tgatgagcttctgtctccatgtc	Brown et al., 2018
Vezatin, adherents' junctions transmembrane protein	NM_172538	vezt-FP-1	gtgtgaaagtggggctgaat	vezt-RP-1	gttcctgcatggtggtgaat	Bahr et al., 2009
YY1-associated factor 2	NM_024189	yaf2-FP-1	catccatggcaaagctgtaa	yaf2-RP-1	tttggtaaaaggccaatgaa	Bahr et al., 2009

Table 15: Reference genes candidates and associated primer sequences with characteristics (T_m, length, GC% and amplicon size)

FP: forward primer; RP: reverse primer; T_m: melting temperature; GC%: percentage in guanines and cytosines.

Several rules are jointly considered for oligonucleotide design for qPCR: primer length between 18-24 bp, primer T_m between 58-60°C, GC content between 30-70%, no more than two C or G in last five positions at 3' end of primer, no more than four constitutive guanines, no primer dimer, length of amplicon between 75-400 bp (Raymaekers et al., 2009). *Nonetheless, length of amplicon is more frequently reduced between 75-250 bp, or even between 75-150 bp, mainly because of the difficulty of obtaining intact RNA samples* (Fleige & Pfaffl, 2006).

Primer T_m, length, and GC% were calculated with Oligo Analysis Tool (Eurofins Genomics). Amplicon sizes were calculated with Primer3Plus.

Gene	FP name	FP sequence (5'-3')	FP T_m [°C]	FP length [bp]	FP GC%	RP name	RP sequence (5'-3')	RP T_m [°C]	RP length [bp]	RP GC%	Amplicon size [bp]
actb	actb-FP-1	gctcttttccagccttcctt	57.3	19	50	actb-RP-1	gagccagagcagtgatctcc	61.4	20	60	284
actb	actb-FP-2	tgacccagatcatgtttgagacc	60.6	23	47.8	actb-RP-2	caggatggcatgagggagag	61.4	20	60	173
actb	actb-FP-3	ctgtgctatgttgccctgga	59.3	20	55	actb-RP-3	gccacaggattccatacccag	61.8	21	57.1	174
actr5	actr5-FP-1	ccagaatggagaaggagctg	59.3	20	55	actr5-RP-1	ggcacaatgttccttgaggt	57.3	20	50	197
actr5	actr5-FP-2	tgctggatgcctggtatgg	58.8	19	57.9	actr5-RP-2	acccaggctccactgtcttct	61.8	21	57.1	67
akr1a1	akr1a1-FP-1	ggcttggaggtgactgcttat	59.8	21	52.4	akr1a1-RP-1	gaacctgccatctgagcaaga	59.8	21	52.4	154
aprt	aprt-FP-1	ctccttccgagcttccatcc	61.4	20	60	aprt-RP-1	ctagggaggggccaaacaag	61.4	20	60	116
atp5f1	atp5f1-FP-1	gagcacatgatggactgggt	59.3	20	55	atp5f1-RP-1	tgcctgtttaactctctctggc	60.3	22	50	180
atp5mg	atp5mg-FP-1	atggccaagttcgtccgtaa	57.3	20	50	atp5mg-RP-1	gggttgggggaaacagttca	59.3	20	55	127
b2m	b2m-FP-1	acggagtttacacccactgc	59.3	20	55	b2m-RP-1	cagacctccatgatgcttga	57.3	20	50	119
b2m	b2m-FP-2	tggctcacacggagtttaca	57.3	20	50	b2m-RP-2	catgtctcgttcccaggtga	59.3	20	55	104
cog1	cog1-FP-1	actagcctccagccagatca	59.3	20	55	cog1-RP-1	gcaggtgagtcgtcttcctc	61.4	20	60	192
eif3i	eif3i-FP-1	ccacaacttccaccaggatt	57.3	20	50	eif3i-RP-1	atgcggacgtaaccatcttc	57.3	20	50	166
eif3k	eif3k-FP-1	agcccaaaaacatcgtggaga	57.9	21	47.6	eif3k-RP-1	tcattgaggaaggggcagaag	59.8	21	52.4	134

Gene	FP name	FP sequence (5'-3')	FP T_m [°C]	FP length [bp]	FP GC%	RP name	RP sequence (5'-3')	RP T_m [°C]	RP length [bp]	RP GC%	Amplicon size [bp]
fkbp1a	fkbp1a-FP-1	ctctcgggacagaaacaagc	59.3	20	55	fkbp1a-RP-1	gacctacactcatctgggctac	62.1	22	54.5	95
gapdh	gapdh-FP-1	gccaagagggtcatcatctc	59.3	20	55	gapdh-RP-1	ccttccacaatgccaaagtt	55.3	20	45	173
gapdh	gapdh-FP-2	gaaagctgtggcgtgatgg	58.8	19	57.9	gapdh-RP-2	tacttggcaggtttctccag	59.3	20	50	187
gapdh	gapdh-FP-3	aactttggcattgtggaagg	55.3	20	45	gapdh-RP-3	acacgttgggggtaggaaca	55.3	20	55	223
gapdh	gapdh-FP-4	tgtaaagctcatttcctggtatgac	59.7	25	40	gapdh-RP-4	tgtgggggttattggacagg	59.3	20	55	178
gnb1	gnb1-FP-1	ccatatgtttctttcccaatggc	58.9	23	43.5	gnb1-RP-1	aagtcgtcgtacccagcaag	59.3	20	55	184
gpx1	gpx1-FP-1	cggacatcaggagaatgcca	59.3	20	55	gpx1-RP-1	gtaaagagcgggtgagcctt	59.3	20	55	138
gusb	gusb-FP-1	ttggtgccaactcctttcgt	57.3	20	57.3	gusb-RP-1	ttgtccctgcgtaccagttc	59.3	20	55	187
hirip3	hirip3-FP-1	cgttatattcgggcttgtgg	57.3	20	50	hirip3-RP-1	ggtcgacctgaactgctgat	59.3	20	55	229
hprt	hprt-FP-1	tacctcaccgctttctcgtg	59.3	20	55	hprt-RP-1	tcactaatcacgacgctggg	59.3	20	55	124
mmadhc	mmadhc-FP-1	tgtcacctcaatgggactgc	59.3	20	55	mmadhc-RP-1	caggtgcatcactactctgaaac	60.6	23	47.8	145
pabpn1	pabpn1-FP-1	gtggccatcctaaagggttt	57.3	20	50	pabpn1-RP-1	cgggagctgttgtaattggt	57.3	20	50	205
pgam1	pgam1-FP-1	aggcgcaggtaaagatctgg	59.3	20	55	pgam1-RP-1	tgcgatccttactgatgttgc	57.9	21	47.6	96
pgk1	pgk1-FP-1	caaggggaaccaagtccctc	61.4	20	60	pgk1-RP-1	gagctctagactggcaccac	61.4	20	60	161
pkar1a	pkar1a-FP-1	gaatcctcatgggaagcactctg	62.4	23	52.2	pkar1a-RP-1	ccttgcaccacgatcttctg	59.3	20	55	160
rps16	rps16-FP-1	tgaagggtggtggtcatgtg	59.3	20	55	rps16-RP-1	tcagctacaagcagggttcg	59.3	20	55	154
tmed2	tmed2-FP-1	gcccacatggatgggacatac	61.8	21	57.1	tmed2-RP-1	tgatgagcttctgtctccatgtc	60.6	23	47.8	131
vezt	vezt-FP-1	gtgtgaaagtggggctgaat	57.3	20	50	vezt-RP-1	gttcctgcatggtggtgaat	57.3	20	50	200
yaf2	yaf2-FP-1	catccatggcaaagctgtaa	55.3	20	45	yaf2-RP-1	tttggtaaaaggccaatgaa	51.1	20	60	162

3.4.3.4.3 Evaluation of primer specificity

Primer specificity was evaluated by analysis melting curves and by DNA electrophoresis of the PCR products.

3.4.3.4.3.1 Analysis of melting curves

A thermal gradient, followed by melt curve analysis, was performed for each pair of primer to (i) confirm annealing temperature; (ii) confirm primer specificity with unique product amplification; (iii) verify the absence of primer-dimer formation; (iv) verify the absence of genomic DNA contamination (Raymaekers et al., 2009; Taylor et al., 2019).

Quant Studio 7Pro real-time thermal cycler allows to perform a six-level temperature gradient on the 96-well plate bloc, with maximum difference of 5°C between adjacent zones.

Pooled DNA (from several DNA extractions) from CHOZN® CHO K1 clone expressing pac, rituHC and rituHC was used for evaluation of primer specificity for plasmid gene primers. Two negative controls were included: pooled DNA from CHOZN® CHO K1 host cell line was used as no target control (*pac*$^-$, *rituHC*$^-$, *rituLC*$^-$), and no template control (NTC). Temperature levels were defined with minimum and maximum primers' T_m. For example, among *pac* forward and reverse primers, lowest primer T_m was 58°C and highest T_m was 63°C, resulting in a 58-to-63°C temperature gradient. Results of thermal gradient and melt curve are listed in Table 16 (for *pac*-specific primer pairs) and in Table 17 (for *rituHC*- and *rituLC*-specific primer pairs). Primer pairs were selected for evaluation of primer efficiency if no amplification was detected on negative controls and if a single peak was observed on the melt curve of DNA samples.

Primer pairs pac-FP/RP-5 and rituLC-FP/RP-1 were the only pairs respectively designed for *pac* and *rituLC* detection to fulfill criteria for selection for evaluation of primer specificity and chosen for verification of amplicon size and evaluation of primer efficiency. All primer pairs designed for *rituHC* detection met the criteria for selection during evaluation of primer efficiency. However, to reduce the number of experimental assays to be done, only rituHCFP-/RP-1 pair was chosen for verification of amplicon size and evaluation of primer efficiency.

For other primer pairs, four main cases were encountered:

(i) Single peak on melt curve of DNA samples; single peak on melt curve of no target control resulting of late amplification (generally with C_q above 30); no amplification on NTC. In this case, primer dimerization is excluded, as a single peak is observed on the DNA samples' melt curves and as there is no amplification on the NTC. Late amplification noticed on the no target control suggests non-specific amplification on the host cell genome. This problem could be circumvented by reducing number of PCR cycles, as target-specific and non-

specific amplification C_q are very distinct (10 cycles or more). This is namely the case of primer pairs pac-FP/RP-1, pac-FP/RP-2, pac-FP/RP-3, pac-FP/RP-7, and rituLC-FP/RP-4.

(ii) Single peak on melt curve of DNA samples; single peak on melt curve of no target control and NTC resulting of late amplification (generally with C_q above 30); peaks' T_m of DNA samples and negative controls are different. In this case, primer-dimer is suggested by amplification on the NTC. However, this is inconsistent with DNA samples' melt curves profiles, which show a single peak. Late amplification noticed on the no target control suggests non-specific amplification on the host cell genome. Well-to-well DNA contamination from no target control to no template control could possibly explain amplification on NTC control if peaks' T_m are the same. This is namely the case of primer pair rituLC-FP/RP-3.

(iii) Single peak on melt curve of DNA samples; single peak on melt curve of no target control and NTC resulting of late amplification (generally with C_q above 30); peaks' T_m of DNA samples and negative controls are equal. In this case, primer-dimer is suggested by amplification on the NTC. However, this is inconsistent with DNA samples' melt curves profiles, which show a single peak. Late amplification noticed on the no target control suggests non-specific amplification on the host cell genome. Well-to-well DNA contamination from DNA samples control to negative controls could possibly explain amplification on negative controls. This is namely the case of primer pairs pac-FP/RP-6 and pac-FP/RP-8. A retest would be required for proper evaluation of these primer pairs.

(iv) Two peaks on melt curve of DNA samples, single peak on melt curve of no target control and NTC resulting of late amplification (generally with C_q above 30); peaks' T_m of DNA samples and negative controls are equal. In this case, primer-dimer is highly probable. This is namely the case of primer pairs pac-FP/RP-4 and rituLC-FP/RP-2.

Table 16: Mean C_q depending on temperature gradient and controls (no target control, no DNA control) and melt curve – pac

A thermal gradient with melt curve analysis was performed for each pair of primer to (i) confirm annealing temperature; (ii) confirm primer specificity with unique product amplification; (iii) verify the absence of primer-dimer formation; (iv) verify the absence of genomic DNA contamination.

Pooled DNA from CHOZN® CHO K1 clone expressing pac was used for evaluation of primer specificity. Pooled DNA from CHOZN® CHO K1 host cell line (pac^-) was used as no target control. Temperature gradient was applied from 58°C to 63°C, according to minimum and maximum pac-primers' T_m.

Gene	Primer pair	Pooled DNA from CHOZN® CHO K1 clone expressing *pac*						No target control	No DNA control	Melt curves	Primer pair selection for evaluation of efficiency
		Mean C_q 58°C [-]	Mean C_q 59°C [-]	Mean C_q 60°C [-]	Mean C_q 61°C [-]	Mean C_q 62°C [-]	Mean C_q 63°C [-]				
pac	pac-FP/RP-1	22.9	22.9	22.8	22.8	22.7	22.9	34.2	No amplification	DNA samples: single peak (T_m 89.3°C) No target control: single peak (T_m 75.6°C)	No
pac	pac-FP/RP-2	22.5	22.2	22.1	22.1	22.2	22.4	34.7	No amplification	DNA samples: single peak (T_m 90.5°C) No target control: several peaks	No
pac	pac-FP/RP-3	24.2	23.5	23.2	23.2	23.3	23.5	33.7	No amplification	DNA samples: single peak (T_m 91.2°C) No target control: several peaks	No
pac	pac-FP/RP-4	23.3	23.5	23.6	23.7	23.8	24.2	27.3	29.7	DNA samples: two peaks (T_m 78.8°C and T_m 87.9°C) below 61°C; single peak (T_m 87.9°C) above 62°C No target control: single peak (T_m 78.9°C) No DNA control: single peak (T_m 76.2°C)	No
pac	pac-FP/RP-5	**22.4**	**22.1**	**21.9**	**21.9**	**21.9**	**22.5**	**No amplification**	**No amplification**	**DNA samples: single peak (T_m 90.7°C)**	**Yes**
pac	pac-FP/RP-6	21.5	21.4	21.3	21.3	21.5	22.0	34.8	No amplification	DNA samples: single peak (T_m 76.5°C) No target control: single peak (T_m 76.5°C)	No
pac	pac-FP/RP-7	21.9	21.8	21.9	22.1	23.3	26.6	34.1	No amplification	DNA samples: single peak (T_m 76.1°C) No target control: single peak (T_m 74.8°C)	No
pac	pac-FP/RP-8	20.6	20.6	20.6	20.6	20.7	20.9	32.4	33.8	DNA samples: single peak (T_m 74.2°C) No target and no DNA controls: single peak (T_m 74.2°C)	No

Table 17: Mean C_q depending on temperature gradient and controls (no RT control, no DNA control) and melt curve – rituHC and rituLC

A thermal gradient with melt curve analysis was performed for each pair of primer to (i) confirm annealing temperature; (ii) confirm primer specificity with unique product amplification; (iii) verify the absence of primer-dimer formation; (iv) verify the absence of genomic DNA contamination.

Pooled DNA from CHOZN® CHO K1 clone expressing rituHC and rituHC was used for evaluation of primer specificity. Pooled DNA from CHOZN® CHO K1 host cell line (rituHC- and rituLC-) was used as no target control. Temperature gradient was applied from 57°C to 62°C, according to minimum and maximum rituHC- and rituLC-primers' T_m.

Gene	Primer pair	Pooled DNA from CHOZN® CHO K1 clone expressing *rituHC* and *rituLC*						No target control	No DNA control	Melt curves	Primer pair selection for evaluation of efficiency
		Mean C_q 57°C [-]	Mean C_q 58°C [-]	Mean C_q 59°C [-]	Mean C_q 60°C [-]	Mean C_q 61°C [-]	Mean C_q 62°C [-]				
rituHC	rituHC-FP/RP-1	**22.2**	**22.2**	**22.3**	**22.3**	**22.4**	**22.8**	**No amplification**	**No amplification**	**DNA samples: single peak (T_m 83.7°C)**	**Yes**
rituHC	rituHC-FP/RP-2	21.9	21.9	22.0	21.9	22.1	22.2	No amplification	No amplification	DNA samples: single peak (T_m 81.5°C)	Yes
rituHC	rituHC-FP/RP-3	23.0	22.9	22.9	22.9	23.4	24.1	No amplification	No amplification	DNA samples: single peak (T_m 84.4°C)	Yes
rituHC	rituHC-FP/RP-4	22.5	22.3	22.3	22.4	22.5	22.9	No amplification	No amplification	DNA samples: single peak (T_m 80.6°C)	Yes
rituLC	rituLC-FP/RP-1	**22.4**	**22.4**	**22.5**	**22.5**	**22.6**	**22.9**	**No amplification**	**No amplification**	**DNA samples: single peak (T_m 81.9°C)**	**Yes**
rituLC	rituLC-FP/RP-2	24.0	23.9	23.8	23.9	24.1	24.1	26.5	27.0	DNA samples: two peaks (T_m 77.1°C and 81.9°C) No target and no DNA controls: one peak (T_m 77.1°C)	No
rituLC	rituLC-FP/RP-3	21.9	22.0	22.0	21.9	22.1	22.2	29.2	29.6	DNA samples: single peak (T_m 84.9°C) No target and no DNA controls: single peak (T_m 75.9°C)	No
rituLC	rituLC-FP/RP-4	22.8	22.9	22.8	23.0	23.0	23.1	38.0	No amplification	DNA samples: single peak (T_m 81.4°C) No target control: single peak (T_m 73.5°C) below 60°C	No

Pooled cDNA from CHOZN® CHO K1 host cell line was used for evaluation of primer specificity for reference genes. Two negative controls were included: no reverse transcription control (no RT control) and no template control (NTC). A 55, 57-60, 62°C temperature gradient was applied to all reference gene primer pairs. Results of thermal gradient and melt curve are listed in Table 18. Primer pairs were selected for evaluation of primer efficiency if no amplification was detected on negative controls and if a single peak was observed on the melt curve of DNA samples. This is the case of primer pairs actbFP/RP1, akr1a1FP/RP1, atp5mgFP/RP1, b2mFP/RP1, cog1FP/RP1, eif3iFP/RP1, eif3kFP/RP1, gapdhFP/RP1, gnb1FP/RP1, gpx1FP/RP1, gusbFP/RP1, hirip3FP/RP1, pabpn1FP/RP1, pgkFP/RP1, rps16FP/RP1, tmed2FP/RP1, veztFP/RP1, and yaf2FP/RP1. Furthermore, combined annealing-extension at 60°C was confirmed, as lowest C_q were most of the time obtained at this temperature.

For other primer pairs, two main cases were encountered:

(i) Single peak on melt curve of DNA samples; no peak or single peak on melt curve of no RT control, resulting of late amplification (generally with C_q above 30); no peak or single peak on NTC, also resulting of late amplification (generally with C_q above 30); peaks' T_m of DNA samples and negative controls are equal. In this case, primer-dimer is suggested by amplification on the NTC. However, this is inconsistent with DNA samples' melt curves profiles, which show a single peak. Well-to-well DNA contamination from DNA samples control to negative controls could possibly explain amplification on negative controls. A retest would be required for proper evaluation of these primer pairs. This is namely the case of primer pairs actb-FP/RP-2, fkbp1a-FP/RP-1, gapdh-FP/RP-2, mmadhc-FP/RP-1, pgam-FP/RP-1 and pkar1a-FP/RP-1.

(ii) Single peak on melt curve of DNA samples; single peak on melt curve of no RT control, resulting of late amplification (generally with C_q above 30); no peak on NTC; peaks' T_m of DNA samples and no RT control are different. In this case, amplification of on the no RT control suggests genomic DNA contamination. However, this is inconsistent with results of primer pairs for which no amplification was detected, showing that potential remaining genomic DNA was removed during RT reaction. Further investigation would be required to know the identity of this amplicon. This is namely the case of primer pairs actb-FP/RP-3, actr5-FP/RP-1/2, aprt-FP/RP-1, atp5f1-FP/RP-1, b2m-FP/RP-2, gapdh-FP/RP-4, and hprt-FP/RP-1.

Table 18: Mean C_q depending on temperature gradient and controls (no RT control, no cDNA control) and melt curve for reference genes

C_q: quantification cycle; RT: reverse transcription.

A thermal gradient with melt curve analysis was performed for each pair of primer to (i) confirm annealing temperature; (ii) confirm primer specificity with unique product amplification; (iii) verify the absence of primer-dimer formation; (iv) verify the absence of genomic DNA contamination.

Pooled cDNA from CHOZN® CHO K1 host cell line was used for evaluation of primer specificity for reference genes.

Results of selected reference genes for normalization (atp5mg and rps16) are in bold.

Gene	Primer pair	Pooled cDNA from CHOZN® CHO K1 host cell line						No target control	No DNA control	Melt curves	Primer pair selection for evaluation of efficiency
		Mean C_q 55°C [-]	Mean C_q 57°C [-]	Mean C_q 58°C [-]	Mean C_q 59°C [-]	Mean C_q 60°C [-]	Mean C_q 62°C [-]				
actb	actb-FP/RP-1	15.8	16.9	17.8	18.6	19.8	22.8	No amplification	No amplification	cDNA samples: single peak (T_m 83.4°C)	Yes
actb	actb-FP/RP-2	13.5	13.1	13.4	13.2	13.1	13.4	32.7	No amplification	cDNA samples: single peak (T_m 83.4°C) No RT control: single peak (T_m 83.4°C)	No
actb	actb-FP/RP-3	14.0	13.8	13.7	13.6	13.6	13.6	35.1	35.0	cDNA samples: single peak (T_m 84.2°C) No RT and no cDNA controls: single peak (T_m 78.1°C)	No
actr5	actr5-FP/RP-1	22.6	22.5	22.3	22.3	22.4	22.9	34.4	33.9	cDNA samples: single peak (T_m 83.9°C) No RT and no cDNA controls: single peak (T_m 74.8°C)	No
actr5	actr5-FP/RP-2	22.1	22.0	22.1	22.2	22.2	22.2	35.7	35.0	cDNA samples: single peak (T_m 81.1°C) No RT and no cDNA controls: single peak (T_m 77.4°C)	No

Gene	Primer pair	Pooled cDNA from CHOZN® CHO K1 host cell line						No target control	No DNA control	Melt curves	Primer pair selection for evaluation of efficiency
		Mean C_q 55°C [-]	Mean C_q 57°C [-]	Mean C_q 58°C [-]	Mean C_q 59°C [-]	Mean C_q 60°C [-]	Mean C_q 62°C [-]				
akr1a1	akr1a1-FP/RP-1	16.9	16.6	16.6	16.5	16.4	16.8	No amplification	No amplification	cDNA samples: single peak (T_m 84.0°C)	Yes
aprt	aprt-FP/RP-1	19.0	19.0	19.1	19.1	19.0	19.2	31.9	31.4	cDNA samples: single peak (T_m 83.8°C) No RT and no cDNA controls: single peak (T_m 76.1°C)	No
atp5f1	atp5f1-FP/RP-1	16.4	16.5	16.4	16.2	16.0	16.3	35.3	35.6	cDNA samples: single peak (T_m 80.1°C) No RT and no cDNA controls: single peak (T_m 76.8°C)	No
atp5mg	atp5mg-FP/RP-1	**16.4**	**16.2**	**16.1**	**16.1**	**16.3**	**16.6**	**No amplification**	**No amplification**	**cDNA samples: single peak (T_m 84.2°C)**	**Yes**
b2m	b2m-FP/RP-1	17.7	17.8	17.7	17.7	17.8	18.1	No amplification	No amplification	cDNA samples: single peak (T_m 80.8°C)	Yes
b2m	b2m-FP/RP-2	16.9	16.7	16.8	16.7	16.6	16.8	34.9	35.3	cDNA samples: single peak (T_m 81.8°C) No RT control: single peak (T_m 75.6°C)	No
cog1	cog1-FP/RP-1	23.3	22.8	22.6	22.7	22.4	22.5	No amplification	No amplification	cDNA samples: single peak (T_m 85.3°C)	Yes
eif3i	eif3i-FP/RP-1	16.5	16.0	16.2	16.0	16.2	16.3	No amplification	No amplification	cDNA samples: single peak (T_m 82.2°C)	Yes
eif3k	eif3k-FP/RP-1	24.8	24.7	24.3	24.2	24.4	24.5	No amplification	No amplification	cDNA samples: single peak (T_m 81.1°C)	Yes
fkbp1a	fkbp1a-FP/RP-1	17.7	17.7	17.6	17.6	17.6	18.0	36.2	36.0	cDNA samples: single peak (T_m 80.0°C) No RT and cDNA controls: single peak (T_m 80.0°C)	No

Gene	Primer pair	Pooled cDNA from CHOZN® CHO K1 host cell line						No target control	No DNA control	Melt curves	Primer pair selection for evaluation of efficiency
		Mean C_q 55°C [-]	Mean C_q 57°C [-]	Mean C_q 58°C [-]	Mean C_q 59°C [-]	Mean C_q 60°C [-]	Mean C_q 62°C [-]				
gapdh	gapdh-FP/RP-1[(1)]	-	-	-	-	-	-	-	No amplification	cDNA samples: single peak (T_m 81.4°C)	Yes
gapdh	gapdh-FP/RP-2	12.9	12.6	12.5	12.6	12.4	12.5	37.1	No amplification	cDNA samples: single peak (T_m 85.3°C) No RT control: single peak (T_m 85.3°C)	No
gapdh	gapdh-FP/RP-3	14.4	13.7	13.7	13.6	13.6	13.8	No amplification	No amplification	cDNA samples: single peak (T_m 86.5°C)	Yes
gapdh	gapdh-FP/RP-4	11.9	11.5	11.3	11.4	11.5	11.5	35.7	35.5	cDNA samples: single peak (T_m 84.7°C) No RT and no cDNA control: single peak (T_m 74.7°C)	No
gnb1	gnb1-FP/RP-1	16.1	16.1	16.0	16.1	16.3	17.5	No amplification	No amplification	cDNA samples: single peak (T_m 82.6°C)	Yes
gpx1	gpx1-FP/RP-1	17.5	17.4	17.3	17.4	17.3	17.3	No amplification	No amplification	cDNA samples: single peak (T_m 81.1°C)	Yes
gusb	gusb-FP/RP-1	19.8	19.1	18.9	18.7	18.6	18.8	No amplification	No amplification	cDNA samples: single peak (T_m 83.8°C)	Yes
hirip3	hirip3-FP/RP-1	21.5	21.0	21.1	21.2	21.2	21.5	No amplification	No amplification	cDNA samples: single peak (T_m 84.7°C)	Yes
hprt	hprt-FP/RP-1	20.0	19.6	19.5	19.4	19.4	19.5	36.3	37.4	cDNA samples: single peak (T_m 87.5°C) No RT and no cDNA controls: single peak (T_m 75.7°C)	No
mmadhc	mmadhc-FP/RP-1	17.9	18.2	18.4	18.3	18.4	18.5	34.8	34.8	cDNA samples: single peak (T_m 80.0°C) No RT and no cDNA controls: single peak (T_m 80.0°C)	No

Gene	Primer pair	Pooled cDNA from CHOZN® CHO K1 host cell line						No target control	No DNA control	Melt curves	Primer pair selection for evaluation of efficiency
		Mean C_q 55°C [-]	**Mean C_q 57°C [-]**	**Mean C_q 58°C [-]**	**Mean C_q 59°C [-]**	**Mean C_q 60°C [-]**	**Mean C_q 62°C [-]**				
pabpn1	pabpn1-FP/RP-1	20.2	20.0	19.9	19.8	20.0	20.1	No amplification	No amplification	cDNA samples: single peak (T_m 82.8°C)	Yes
pgam1	pgam1-FP/RP-1	18.5	18.2	18.1	18.2	17.8	18.0	No amplification	33.4	cDNA samples: single peak (T_m 80.1°C) No RT and no cDNA controls: single peak (T_m 79.8°C)	No
pgk1	pgk1-FP/RP-1	16.6	16.5	16.3	16.2	16.2	16.2	No amplification	No amplification	cDNA samples: single peak (T_m 82.1°C)	Yes
pkar1a	pkar1a-FP/RP-1	18.4	17.9	17.9	17.8	17.7	18.0	35.4	36.0	cDNA samples: single peak (T_m 80.9°C) No RT and no cDNA controls: single peak (T_m 80.2°C)	No
rps16	rps16-FP/RP-1	**14.1**	**16.7**	**13.5**	**13.5**	**13.5**	**13.6**	**No amplification**	**No amplification**	**cDNA samples: single peak (T_m 80.8°C)**	**Yes**
tmed2	tmed2-FP/RP-1	18.8	18.3	18.2	18.1	18.2	18.2	No amplification	No amplification	cDNA samples: single peak (T_m 77.9°C)	Yes
vezt	vezt-FP/RP-1	21.0	21.6	21.6	21.5	21.5	21.8	No amplification	No amplification	cDNA samples: single peak (T_m 85.2°C)	Yes
yaf2	yaf2-FP/RP-1	23.9	22.8	22.7	23.3	24.8	34.4	No amplification	No amplification	cDNA samples: single peak (T_m 78.4°C)	Yes

3.4.3.4.3.2 Amplicon size verification

Amplicon size should be verified by agarose gel electrophoresis to confirm melt curve results (Raymaekers et al., 2009; Taylor et al., 2019).

Amplicon size was verified by agarose gel electrophoresis. Results are shown on Figure 52 (for plasmid genes) and on Figure 53 (for reference genes). All PCR products were of expected size. Melt curve results are confirmed for each pair of primers. Amplification products were not sequence analyzed.

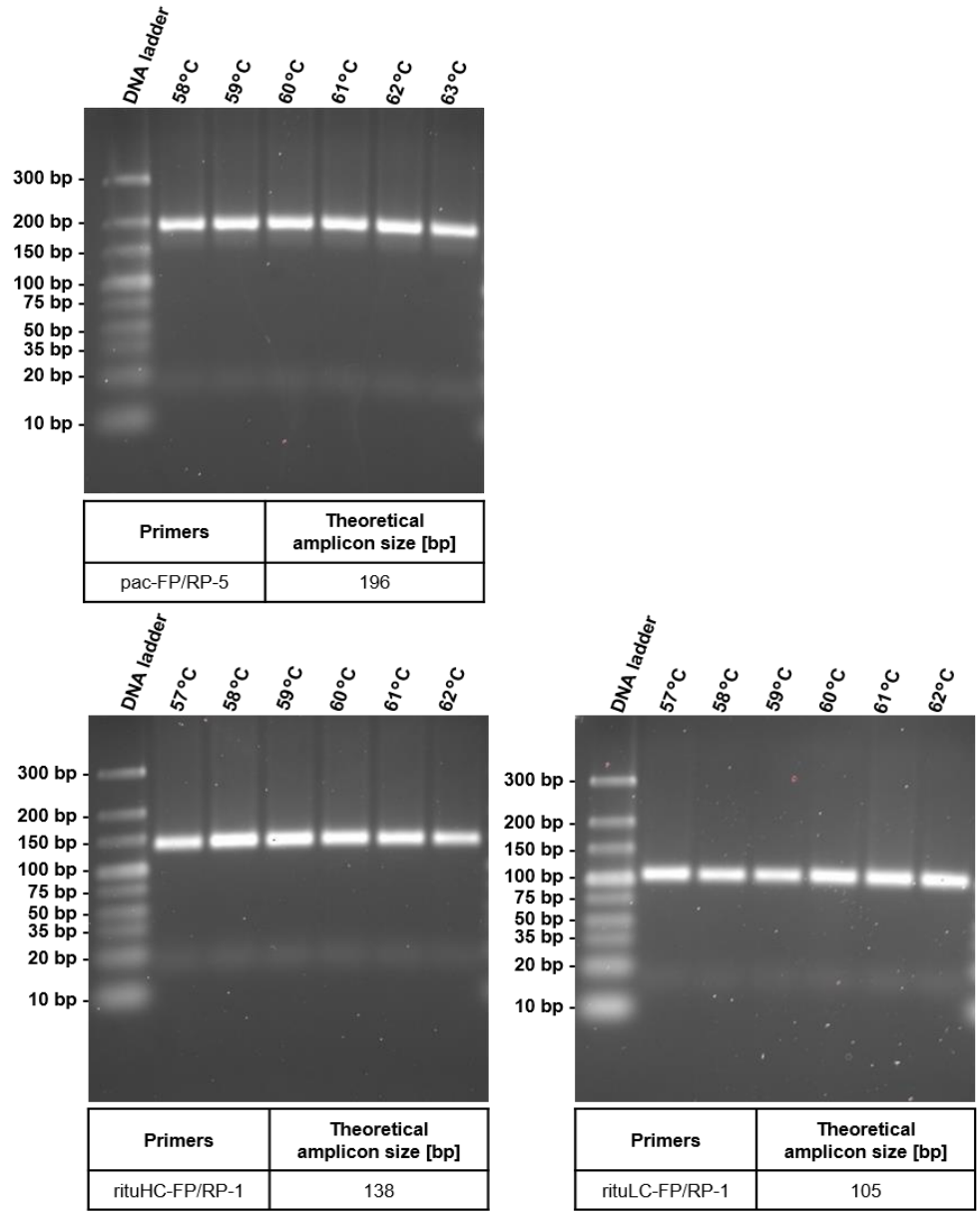
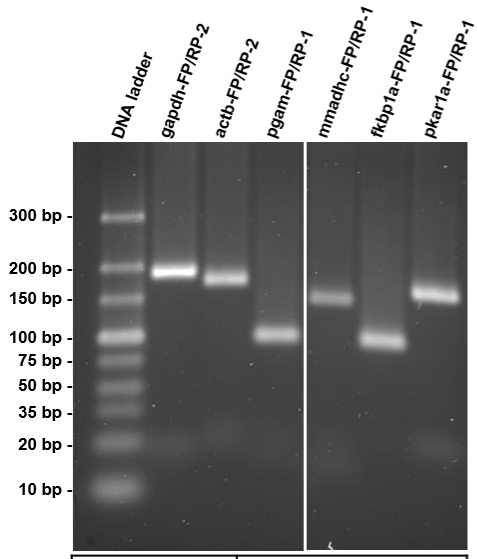
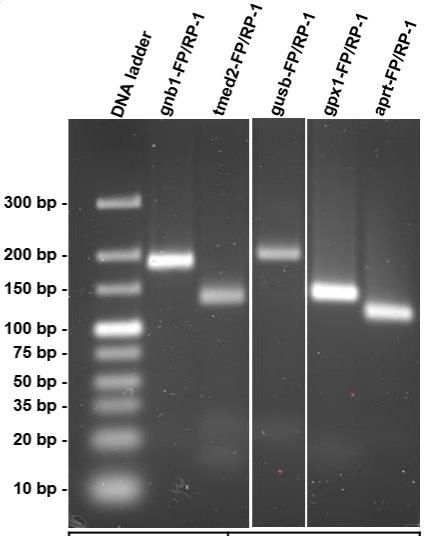


Figure 52: Amplicon size verification by agarose gel electrophoresis (plasmid genes)

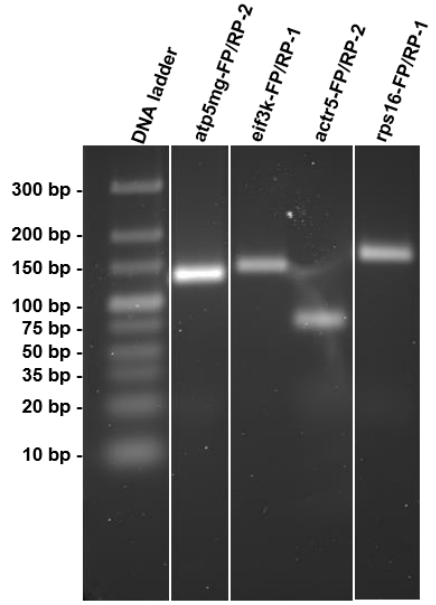
Size of PCR amplified products was verified on agarose gel electrophoresis (3% agarose). DNA ladder: Ultra Low Range DNA ladder (ThermoFischer Scientific), designed for sizing double-stranded DNA in the range of 10 bp to 300 bp. DNA was stained with SYBR™ Safe DNA Gel Stain in 1X TAE (ThermoFischer Scientific). Images were taken with a ChemiDoc™ Touch Gel Imaging System (BioRad).



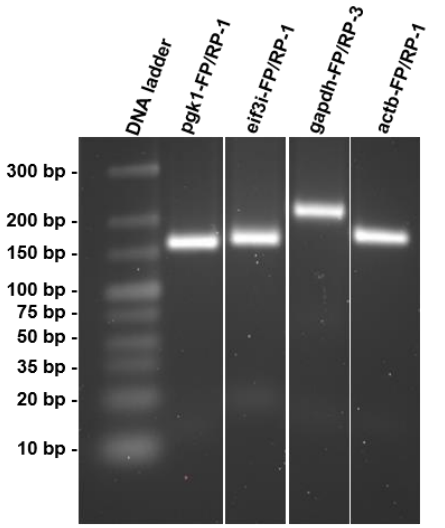
Primers	Theoretical amplicon size [bp]
gapdh-FP/RP-2	187
actb-FP/RP-1	173
pgam-FP/RP-1	96
mmadhc-FP/RP-1	145
fkbp1a-FP/RP-1	95
pkar1a1-FP/RP1	160



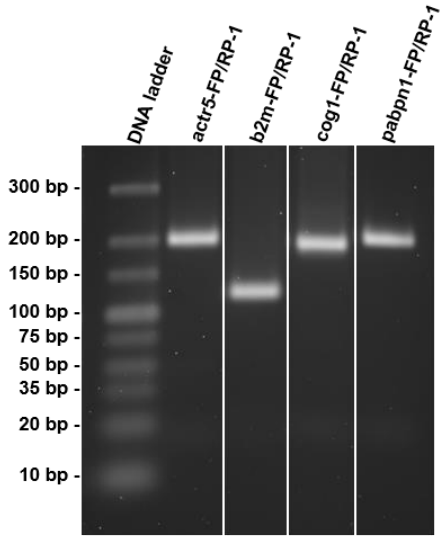
Primers	Theoretical amplicon size [bp]
gnb1-FP/RP-1	184
tmed2-FP/RP-1	131
gusb-FP/RP-1	187
gpx1-FP/RP-1	138
aprt-FP/RP-1	116



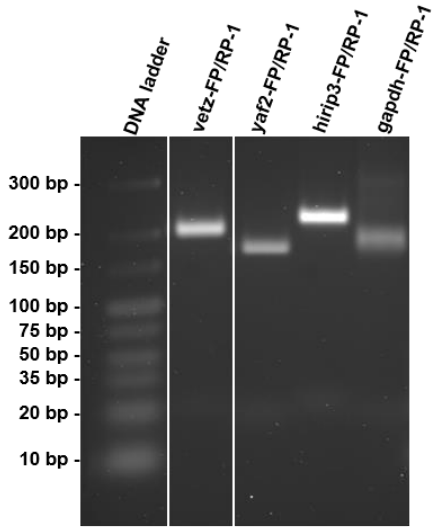
Primers	Theoretical amplicon size [bp]
atp5mg-FP/RP-1	127
eif3k-FP/RP-1	134
actr5-FP/RP-2	67
rps16-FP/RP-1	154
atp5mg-FP/RP-1	127



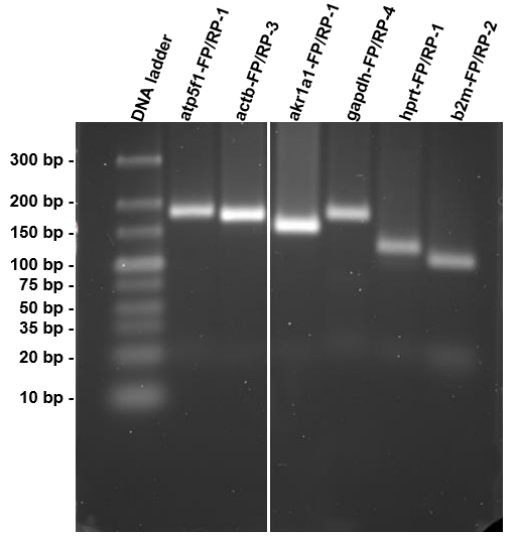
Primers	Theoretical amplicon size [bp]
pgk1-FP/RP-1	161
eif3i-FP/RP-1	166
gapdh-FP/RP-3	223
actb-FP/RP-1	284



Primers	Theoretical amplicon size [bp]
actr5-FP/RP-1	197
b2m-FP/RP-1	119
cog1-FP/RP-1	192
pabpn1-FP/RP-1	205



Primers	Theoretical amplicon size [bp]
yaf2-FP/RP-1	200
vetz-FP/RP-1	162
hirip3-FP/RP-1	229
gapdh-FP/RP-1	173



Primers	Theoretical amplicon size [bp]
atp5f1-FP/RP-1	180
actb-FP/RP-3	174
akr1a1-FP/RP-1	154
gapdh-FP/RP-4	178
hprt-FP/RP-1	124
b2m-FP/RP-2	104

Figure 53: Amplicon size verification by agarose gel electrophoresis (reference genes)

Size of PCR amplified products was verified on agarose gel electrophoresis (3% agarose). DNA ladder: Ultra Low Range DNA ladder (ThermoFischer Scientific), designed for sizing double-stranded DNA in the range of 10 bp to 300 bp. DNA was stained with SYBR™ Safe DNA Gel Stain in 1X TAE

(ThermoFischer Scientific). Images were taken with a ChemiDoc™ Touch Gel Imaging System (BioRad).

3.4.3.4.4 Evaluation of PCR efficiency

PCR characteristics (namely slope (mentioned as *m*), efficiency (mentioned as *E*), and coefficient of correlation (mentioned as R^2) of linear regression) should be determined for each pair of primers (Bustin et al., 2009). When relative quantification is performed, the choice of the calculation method for normalization depends on PCR efficiencies. The $2^{-\Delta\Delta Ct}$ calculation method allows analysis of relative changes in gene expression when PCR efficiencies are the same (Livak & Schmittgen, 2001). Otherwise, other mathematical methods, such as the Pfaffl model, have been described to correct for differences in amplification efficiencies (Pfaffl, 2001).

PCR characteristics are defined from a standard curve generally performed with 10-fold serial dilutions of a mixture of representative cDNA template (Derveaux et al., 2010; Raymaekers et al., 2009). C_q values (plotted on y-axis) are plotted against a calibrator with arbitrary units, which can be either the logarithm of cDNA template's concentrations, number of cDNA target copies or dilution factors (plotted on x-axis) (Raymaekers et al., 2009).

(i) The slope results from the linear regression on the standard curve. Its value should ideally be equal to -3.3219 or comprised between -3.6 and -3.1 (Raymaekers et al., 2009). Some authors reduce this interval between -3.5 and -3.2 (Nolan et al., 2006).

(ii) PCR efficiency is calculated as follows $E = 10^{\left(\frac{-1}{m}\right)} - 1$. PCR efficiency should ideally be equal to 1 (100%) or comprised between 90-110% (Raymaekers et al., 2009; Taylor et al., 2019). 100% efficient PCR reaction means that the target amplicon quantity is doubled at each cycle (Nolan et al., 2006). Several parameters can broadly influence PCR efficiency, such as PCR inhibitors, PCR enhancers, cDNA degradation, cDNA concentration, amplicon size, secondary structure, primer quality (Raymaekers et al., 2009).

(iii) Coefficient of correlation is the measure of the closeness of experimental values to the linear regression and indicates reproducibility of replicates. Coefficient of correlation should ideally be equal to 1 or above 0.99 (Nolan et al., 2006; Raymaekers et al., 2009).

For each primer pair, PCR efficiency was evaluated with three independent 10-fold serial dilutions of pooled representative cDNA templates. Slope, PCR efficiency and coefficient of correlation of linear regression for each pair of primers are described in Table 19 (plasmid genes) and Table 20 (reference genes). Primer pairs for *pac*, *rituHC* and *rituLC* detection were all validated. Primers pairs for *actb*, *akr1a1*, *atp5mg*, *b2m*, *cog1*, *eif3i*, *gapdh*, *gnb1*, *gpx1*, *gusb*, *hirip3*, *pabpn1*, *pgk1*, *rps16*, *tmed2*, and *vezt* fulfilled the criteria previously mentioned both for slope, efficiency, and coefficient of correlation. They were therefore selected for reference gene validation.

Table 19: Amplification efficiencies (plasmid genes)

Three independent 10-fold serial dilutions of pooled DNA templates were performed in technical triplicates to determine amplification efficiencies. Slope, amplification efficiencies and coefficient of correlation (R^2) were determined from standard curves, using the equations $C_q = m * log\, Conc. + y_{intercept}$ *and* $E = 10^{\left(\frac{-1}{m}\right)} - 1$ *linear regression analysis, respectively.*

Gene	Primer pair	Slope (*m*) [-]	Efficiency (*E*) [%]	R^2 [-]	Primer pair validation
pac	pac-FP/RP-5	**-3.313**	**100.4**	**0.998**	**Yes**
rituHC	rituHC-FP/RP-1	**-3.287**	**101.5**	**0.996**	**Yes**
rituLC	rituLC-FP/RP-1	**-3.371**	**98.0**	**0.997**	**Yes**

Table 20: Amplification efficiencies (reference genes)

Three independent 10-fold serial dilutions of pooled cDNA templates were performed in technical triplicates to determine amplification efficiencies. Slope, amplification efficiencies and coefficient of correlation (R^2) were determined from standard curves, using the equations $C_q = m * log\, Conc. + y_{intercept}$ *and* $E = 10^{\left(\frac{-1}{m}\right)} - 1$ *linear regression analysis, respectively.*

Results of selected reference genes for normalization (atp5mg and rps16) are in bold.

Gene	Primer pair	Slope (*m*) [-]	Efficiency (*E*) [%]	R^2 [-]	Primer pair selection for reference genes validation
actb	actb-FP/RP-1	-3.301	101	0.999	Yes
akr1a1	akr1a1-FP/RP-1	-3.459	95	0.999	Yes
atp5mg	atp5mg-FP/RP-1	**-3.484**	**94**	**1.000**	**Yes**
b2m	b2m-FP/RP-1	-3.430	96	0.999	Yes
cog1	cog1-FP/RP-1	-3.297	101	0.991	Yes
eif3i	eif3i-FP/RP-1	-3.428	96	0.999	Yes
eif3k	eif3k-FP/RP-1	-4.095	76	0.968	No
gapdh	gapdh-FP/PR-1	-3.242	103	0.996	Yes
gnb1	gnb1-FP/RP-1	-3.598	90	0.996	Yes
gpx1	gpx1-FP/RP-1	-3.416	96	0.999	Yes
gusb	gusb-FP/RP-1	-3.429	96	0.992	Yes
hirip3	hirip3-FP/RP-1	-3.421	96	0.991	Yes
pabpn1	pabpn1-FP/RP-1	-3.272	102	0.999	Yes
pgk1	pgk1-FP/RP-1	-3.405	97	0.999	Yes
rps16	rps16-FP/RP-1	**-3.417**	**96**	**0.999**	**Yes**
tmed2	tmed2-FP/RP-1	-3.484	94	0.999	Yes
vezt	vezt-FP/RP-1	-3.419	96	0.997	Yes
yaf2	yaf2-FP/RP-1	-3.159	107	0.936	No

3.4.3.4.5 Reference genes validation

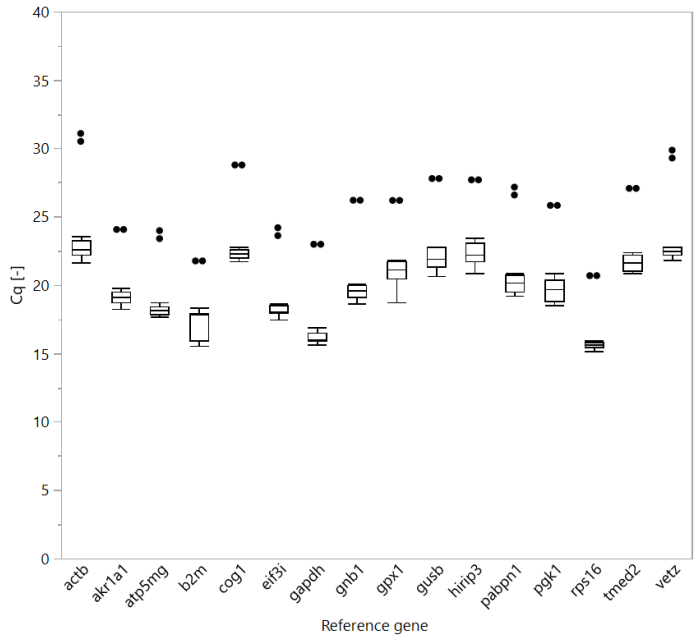
Reference genes validation was achieved on a set of samples reflecting experimental conditions in which RT-qPCR assays will further be performed. Therefore, samples were collected from batch culture, fed-batch culture, and transient expression culture. C_q values of all reference gene candidates across batch culture, fed-batch culture, and transient expression culture are plotted in Figure 54-A. The mean C_q of each reference gene varies from 16.2 (*rps16*) to 23.6 (*actb*) **(**Figure 54-B**)**. Variable transcript level is observed, from high gene expression level (mean C_q below 20) for *akr1a1*, *atp5mg*, *b2m*, *eif3i*, *gapdh*, and *rps16* to medium gene expression level (mean C_q between 20 and 30) for *actb*, *cog1*, *gnb1*, *gpx1*, *gusb*, *hirip3*, *pabpn1*, *pgk1*, *tmed2* and *vezt*. It might be noticed that some experimental points (samples from fed-batch culture at PD_{12}) are considered as outliers. Very low viability of cell cultures resulted in low quality RNA, therefore explaining higher C_q. Standard deviation (SD) and coefficient of variation (CV) were calculated to give a preliminary assessment of transcript level stability **(**Figure 54-B**)**. All reference gene candidates showed a CV below 15%.

geNorm algorithm was further used (i) to rank reference genes candidates according to their expression level stability, and (ii) to determine the appropriate number of reference genes to be used for normalization in RT-qPCR assays. geNorm is a Visual Basic Application (VBA) used in Microsoft Excel. geNorm calculates the gene-stability measure (mentioned as *M* value), based on non-normalized expression levels for several genes in a set of samples. The gene-stability measure relies on the principle that the expression ratio of two reference genes is equal in all samples, irrespective of the experimental treatment. Variation of the expression ratio of two reference genes indicates that one or both genes are not constantly expressed. The *M* value is calculated by averaging pair-wise variation of each gene with all other evaluated genes. Genes with the lowest *M* value are the most stably expressed. Ranking of the reference genes candidates according to the gene expression stability measure is based on standard deviation of logarithmically transformed expression ratios from the most variable to the most stable two genes by stepwise elimination of the least stable one (Vandesompele et al., 2002). geNorm also calculates the systemic variation (mentioned as *V* value) as the pairwise variation for repeated RT-qPCR experiments on the same gene, reflecting variation due to equipment and enzymatic reaction variations. The *V* value indicates the minimum number of reference genes required for accurate normalization in RT-qPCR experiments, based on a 0.15 cut-off value below which the inclusion of an additional reference gene is not required (Vandesompele et al., 2002).

Gene stability measures of reference gene candidates are depicted in Figure 54-A. *rps16* and *atp5mg* are identified as the two most stably expressed genes, *akr1a1* being the third one. Pairwise variations for determination of the optimal number of reference genes for normalization are depicted in Figure 54-B. The $V_{2/3}$ is below the cut-off value of 0.15, meaning that two reference genes are required for normalization of RT-qPCR experiments. Inclusion of a third reference gene has no significant effect.

Based on average expression stability values, on pairwise variation ($V_{n/n+1}$) analysis between the normalization factors, and on literature's guidelines, *rps16* and *atp5mg* were chosen as reference genes for normalization of RT-qPCR results.

A



B

Gene	Primer pair	Mean C_q	SD	CV (%)
actb	actb-FP/FP-1	23.6	2.9	12.2
akr1a1	akr1a1-FP/RP-1	19.7	1.8	9.2
atp5mg	atp5mg-FP/RP-1	**18.8**	**1.9**	**10.3**
b2m	b2m-FP/RP-1	17.7	1.9	10.7
cog1	cog1-FP/RP-1	23.1	2.3	10.0
eif3i	eif3i-FP/RP-1	18.9	2.0	10.8
gapdh	gapdh-FP/RP-1	17.0	2.4	14.3
gnb1	gnb1-FP/RP-1	20.3	2.4	12.0
gpx1	gpx1-FP/RP-1	21.4	2.1	9.8
gusb	gusb-FP/RP-1	22.7	2.2	9.5
hirip3	hirip3-FP/RP-1	22.8	2.1	9.1
pabpn1	pabpn1-FP/RP-1	20.9	2.4	11.4
pgk1	pgk1-FP/RP-1	20.3	2.3	11.5
rps16	rps16-FP/RP-1	**16.2**	**1.8**	**11.2**
tmed2	tmed2-FP/RP-1	22.3	2.0	9.1
vezt	vetz-FP/RP-1	23.3	2.5	10.8

Figure 54: Quantification cycles for reference genes on validation samples

C_q: quantification cycle; SD: standard deviation; CV: coefficient of variation.

Box-and-whiskers plots of quantification cycles (C_q) values from RT-qPCR reactions using 15 different cell culture conditions (batch culture, fed-batch culture, and transient expression). All experiments were

performed in technical triplicates. Boxplot horizontal lines respectively indicate, top to bottom, the maximum, the upper quartile (Q3), the median, the lower quartile (Q1) and the minimum C_q. Outliers are indicated by dots. A point is an outlier if it exceeds the upper outlier limit (Q_3+1.5×IQR) or is less than the lower outlier limit (Q_1-1.5×IQR), where IQR = Q_3-Q_1 Data analysis was performed on JMP® 15 software.

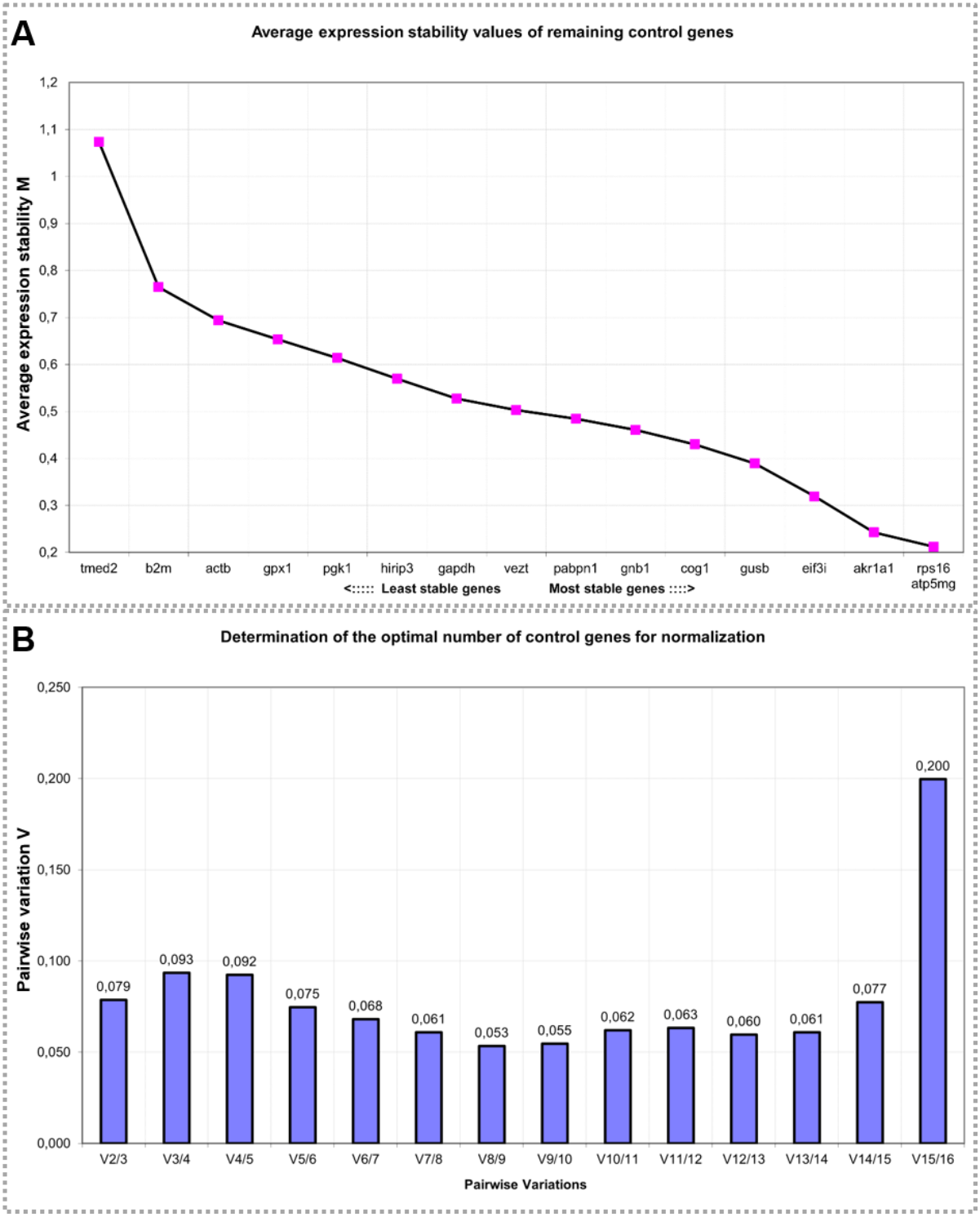


Figure 55: geNorm analysis results: average expression stability value (M value) and pairwise variation for determination of the optimal number of reference genes for normalization (V value)

(A) *Average expression stability values (M values) of remaining reference genes during stepwise exclusion of the least stable reference gene in different samples (batch culture, fed-batch culture, and transient expression culture).* ***(B)*** *Pairwise variation ($V_{n/n+1}$) analysis between the normalization factors to determine the number of reference genes required for accurate normalization. Cut-off value is 0.15. Data analysis was performed with geNorm applet.*

3.4.4 Discussion

The need to analyze transcription levels (first for evaluation of promoters' transcriptional strength, and second for correlation of transcription levels with translation levels during expression level stability study) led to the development of a RT-qPCR method. An RT-qPCR method is a succession of several steps (sampling from cell culture, RNA extraction, cDNA synthesis by reverse transcription, and qPCR), each of which are critical for the overall success of the analysis. Therefore, method development was conducted following the minimum information for publication of quantitative PCR experiments (MIQE) guidelines and the real-time PCR data markup language (RDML), to ensure proper implementation and validation of the method (Bustin, 2010; Bustin et al., 2009, 2010; Lefever et al., 2009).

First, RNA extraction method was set up, and RNA quality control included in the experimental design. RNA quality control is indeed a requirement for many further RNA-based downstream applications, including RT-qPCR, and refers to RNA concentration, purity, and integrity measurements (Bustin et al., 2009; Fleige & Pfaffl, 2006; A. Schroeder et al., 2006). RNA quality (especially purity and integrity) influences the accuracy of gene expression measurement. Quality control of extracted RNA showed that all RNA samples have a A_{260}/A_{280} ranging from 2.0 to 2.2 and a A_{260}/A_{230} ranging from 2.0 to 2.2, therefore indicating pure RNA solution. In addition, all RNA samples have a RQS above 5.0, indicating partially degraded RNA, or even a RQS above 8.0, indicating intact RNA. Although a maximum of precautions has been taken for careful handling of samples (rapid sampling, handling on ice, quick freezing, etc.), RNA integrity is partially impacted on several samples. Such a RQS is not critical for RT-qPCR experiments if amplicon size is kept relatively short (below 250 bp, or even below 150 bp) (Fleige & Pfaffl, 2006). Nonetheless, RNA extraction procedure could be further improved to ensure systematic extraction of high-quality RNA. For example, RNase inhibitors or storage reagents (such as RNA*later*™ (ThermoFisher Scientific)) could be used for this purpose (Fleige & Pfaffl, 2006). RNase inhibitors are recombinant enzymes used to inhibit RNase activity, which can be co-purified with isolated RNA and compromise further RNA-based downstream applications. Stabilization and storage reagents rapidly permeate tissues to stabilize and protect cellular RNA.

Second, qPCR method has been implemented. After choosing detection chemistry (fluorescent dye) and data analysis method (normalization against reference genes), and designing primer for intended targets, primers for plasmid genes and reference genes were evaluated for specificity and efficacy. In addition, candidate reference genes were evaluated for normalization. Precise normalization is essential for gene expression data analysis in RT-qPCR assays, as it should, among other things, compensate RT-qPCR variations (Bustin et al., 2005; Chervoneva et al., 2010; Huggett et al., 2005; Kubista et al., 2006;

Radonić et al., 2004). The use of reference genes is the most routinely applied method for RT-qPCR normalization (Kozera & Rapacz, 2013; Kubista et al., 2006; Raymaekers et al., 2009). There is a consensus among the scientific community that the final choice of reference genes should be systematically validated for each RT-qPCR study (Bustin et al., 2009; Chervoneva et al., 2010; Kozera & Rapacz, 2013; Radonić et al., 2004). Validation of reference genes was conducted on genes described in the literature, some of them being described as traditional reference genes (such as *actb*, *b2m*, and *gapdh*), others being proposed as novel reference genes (Bahr et al., 2009; A. J. Brown et al., 2018; Ma et al., 2020). Our study resulted in validation of *rps16* and *atp5mg* as the most stable genes among 16 final candidates for normalization.

rps16 (NCBI Gene ID: 100773230) encodes the 40S ribosomal protein S16-like protein (UniProt ID: A0A061HTB4), which is a ribosomal structural protein component of the 40S ribosome sub-unit and a member of the S9P family of ribosomal proteins (Ma et al., 2020). *rps16* gene is located on chromosome 9 and protein is localized in cytoplasm. *rps16* was first evaluated as reference gene in CHO cells in Ma's work in 2020. Nonetheless, *rps16* is reported as reference gene in other organisms (namely, insects such as the ant *Atta sexdens*) (do Livramento et al., 2018), and related genes (namely, *rps13*, *rps18*, and *rps23* – all encoding structural constituent proteins of the 40S ribosome) have been used as reference genes in other cells types (induced pluripotent stem cells (iPSC), bladder or colon cancer cells) (C. L. Andersen et al., 2004; Panina et al., 2018). On the other hand, *atp5mg* (NCBI Gene ID: 100757815) encodes ATP synthase membrane sub-unit g (UniProt ID: A0A3L7HSE4), which produces ATP from ADP in the presence of a proton gradient across the membrane, itself generated by electron transport complexes of the respiratory chain. *atp5mg* gene is located on chromosome 4 and protein is localized in mitochondrion. As for *rps16*, *atp5mg* was first evaluated as reference gene in CHO cells in Ma's work in 2020, and has not, to our knowledge, been reported for use in RT-qPCR analyses before. *rps16* and *atp5mg* are therefore involved in various molecular functions and biological processes, which makes them even more suitable for normalization as they are not likely to be co-regulated (Ma et al., 2020).

Our results partially support those of the literature. Indeed, *rps16* and *atp5mg* were also identified in Ma's work as the most stable genes in fed-batch cultures (according to geNorm and BestKeeper rankings, as well as overall ranking) (Ma et al., 2020). Similarities in dataset collection (samples taken from exponential and stationary phases in fed-batch culture of recombinant cell lines from CHO-K1 lineage) could indeed lead to expect similar, though not identical, results. Indeed, some discrepancies might be noticed. For example, *b2m* and *gpx1*, which are respectively ranked third and fourth most stable genes (according to geNorm ranking) in Ma's work, were evaluated among the four least stable genes in our study. Our dataset was mainly composed of samples taken in fed-batch culture, but not only: samples taken from sub-culturing flasks, as well as samples taken from transient expression cultures, were included for evaluation of candidate reference genes. It has been attempted to perform candidate

reference genes evaluation on individual type of samples (*i.e.*, samples from fed-batch culture, samples from sub-culturing, and samples from transient expression individually), but insufficient number of samples per condition would have led to flawed conclusion (Vandesompele et al., 2002).

In addition, it might be noticed that commonly used reference genes, such as *gapdh* or *actb*, are among the least stable genes across evaluated genes, highlighting the importance of proper reference gene evaluation for normalization in RT-qPCR assays, as mentioned in many publications (Bustin et al., 2009; Chervoneva et al., 2010; Derveaux et al., 2010; Kozera & Rapacz, 2013; Radonić et al., 2004).

3.4.5 Conclusion

In this fourth and last chapter, several methods (namely, nucleic acid extraction, nucleic acid quality control, reverse transcription, and quantitative PCR) have been validated for implementation on the CHOZN® CHO K1 platform. These methods have been used for gene transcription level analysis (results presented in the third chapter) and could be used for further analyses on this platform. Gene copy number analysis and transcription level measurement of recombinant protein could be implemented as additional analytics to characterize recombinant cell lines during cell line development process. In the long term, these methods, and particularly RT-qPCR, could be used for the detection of markers indicating a high level of productivity.

3.4.6 Author's contributions

Literature review: Lucie Grindes[1,2] (ORCID ID: 0000-0002-6201-3355); **Methodology**: Lucie Grindes; **Investigation**: Lucie Grindes, Christine Dinard[3] (significant technical help for RNA integrity analysis); **Analysis**: Lucie Grindes; **Writing – Original draft and editing**: Lucie Grindes; **Writing – Review**: Lucie Grindes, Gilles Joucla[2] (ORCID ID: 0000-0003-4192-9628), Wilfrid Dieryck[2] (ORCID ID: 0000-0002-4609-8934), Cyrielle Corbin[1] (ORCID ID: 0000-0002-6891-2511), Sébastien Ribault[4] (trade secret and intellectual property), Daniel Kasten[5] (trade secret and intellectual property); **Supervision**: Cyrielle Corbin.

[1] Cell line development team, Process development department, Merck Biodevelopment

[2] Structure and activity of biological macromolecules team, Nutraceutical, medical and food applications pole, CBMN (UMR 5842 (CNRS, University of Bordeaux, Bordeaux INP))

[3] Analytical development department, Merck Biodevelopment

[4] Direction, Merck Biodevelopment

[5] Patents Life Science global team, MilliporeSigma

4 Discussion

4.1 Combination of direct single-cell cloning process and attenuated selection marker for optimization of CHOZN® CHO K1 platform

In the first chapter, growth characteristics and initial performance (in terms of productivity, stability and timelines) of the CHOZN® CHO K1 cell line development platform have been defined and led to identification of two main areas of work: (i) reduction of cell line development timelines and (ii) increase of recombinant cell lines productivity. In the second chapter, an accelerated cell line development process, called direct single-cell cloning (DSCC) was evaluated to reduce cell line development timelines. In the third chapter, several promoters driving resistance to selection pressure expression were evaluated to increase clonal cell lines productivities. Both parameters have been separately discussed in respective chapters.

Reducing cell line development timelines from 7.5 to 4.5 months *via* direct single-cell cloning protocol proved to be technically feasible with CHOZN® CHO K1 cell line development platform. Nonetheless, significant decrease of clones' productivity was observed, compared to the traditional cell line development process based on mini-pool and clone generation.

In spite of this productivity decrease, direct single-cell cloning became the standard cell line development protocol for several studies conducted during this thesis. Indeed, the direct single-cell cloning strategy has the advantage to (i) generate clones using the bulk transfected pool, in which cellular diversity is high, and (ii) be the straightest method to generate clone-linked results (approximately 4 months, from host cell line thawing to fed-batch productivity assay, compared to approximately 7 months for traditional cell line development process). The use of a non-optimized direct single-cloning process was not troublesome in the context of the thesis, as long as the comparison assays (for example, comparison of promoters for puromycin resistance expression (3.3 Expression vector optimization: evaluation of promoters for expression of resistance to puromycin)) were performed following the same experimental conditions.

As a consequence, knowledge about the behavior of the CHOZN® CHO K1 cell line during direct single-cell cloning, and more precisely during selection, has been acquired and confirmed over assays. In all direct single-cell assays performed so far, a sharp viability decrease was observed within the first days following transfection and addition of selection pressure, generally reaching a minimum viability between 5-15% on day 7 or 8 (D_7 or D_8). Afterwards, viability increased until full pool recovery. The 60% viability threshold, later defined as viability threshold for single-cell cloning, was reached at the earliest between D_8-D_{11} and at the latest between D_{14}-D_{16}. Longer recovery duration was namely due to

lower strength of selection marker (3.3 Expression vector optimization: evaluation of promoters for expression of resistance to puromycin).

In addition, data has been accumulated during these assays, without these tests being process optimization assays per se. Experimental conditions and associated results regarding cell behavior during selection, cloning efficiency and clone productivity of all direct-single cell cloning runs are detailed in Table 21. Several parameters have been modified over assays, namely:

(i) Viability threshold for single-cell cloning. First direct single-cell cloning assay was performed when viability was the lowest (or, more precisely, supposed to be the lowest, as the whole cell suspension was used for single-cell cloning, therefore preventing the observation of pool recovery). Resulting cloning efficiencies were the lowest (10 to 16%) of all dataset. Viability threshold for single-cell cloning was later increased to 60%. Resulting cloning efficiencies were consecutively increased (21% cloning efficiency on assay 2, compared to 16% cloning efficiency on assay 1-A-1, *i.e.*, a 1.3-fold increase on cloning efficiency between assays). Similar results have been obtained by Misaghi *et al*, who showed that CHO K1 clone recovery rate increases when selection duration increases, but suspected that increased clone recovery rate might be due to an increase of sister clones derived from the same parental cell (Misaghi et al., 2016). Finding the right viability target for single-cell cloning is a balance between cloning efficiency and clone productivity. On the one hand, increasing viability target would result in increased cloning efficiency. On the other hand, it has been shown that pool viability reflects cellular diversity within the pool (low viability would correspond to high cellular diversity, while loss of diversity would have occurred once recovery is achieved) (Diep et al., 2021). Some groups investigated a fast-track cloning approach, reported in a Cell Series 2020 presentation ("Improved host cell lines for accelerated manufacture of biologics", Köhler R., Novartis, NIBR Biologics Center – Cell line screening and development). This approach consists in performing cloning as soon as possible after transfection during viability decrease and before minimum viability and led to the generation of high producing clones (up to 8.5 g/L after fed-batch productivity assay in ambr250® bioreactors – though this approach was also combined with host cell line engineering (CHO-C8DEL cell line), expression vector optimization and advanced technology equipment for single-cell cloning (FACS). Therefore, increasing viability target would result in lower clone productivity, probably because the best producers are overtaken by faster growing clones. Two other levers of action can be explored: single-cell cloning method and cloning medium composition.

(ii) Single-cell cloning method. All cloning assays were performed with limiting dilution method, except assay 1-A-2, which has been performed using a cell sorter-seeder (Cyto-Mine® (Sphere Fluidics)). Lower cloning efficiency was obtained with this cloning method

compared to limiting dilution cloning method (10% cloning efficiency with cell sorter-seeder on assay 1-A-2, compared to 16% cloning efficiency with limiting dilution method on assay 1-A-2 – all other parameters being equal). Even though cell sorting might be stressful for cells, 10 to 25-fold higher clone seeding density with cell seeder equipment (compared to limiting dilution cloning seeding density) should result in higher cloning efficiency. Our results with cell sorter-seeder should though be mitigated. Indeed, they were obtained during an equipment demonstration, therefore cloning parameters were not optimized for the CHOZN® CHO K1 cell line development platform.

However, selection of high producing clones is facilitated with the use of a cell sorter. Higher producing clones were identified on assay 1-A-2 (single-cell cloning using a cell sorter-seeder) compared to assay 1-A-1 (single-cell cloning using limiting dilution method) (1.7-fold increase on maximum titer, 3.6-fold increase on mean titer, and 1.9-fold increase on maximum specific productivity between assays 1-A-2 and 1-A-1).

(iii) Cloning medium composition. First single-cell assays have been performed using a blend of 80% EX-CELL® CHO Cloning medium and 20% of conditioned medium, supplemented with 6 mM L-glutamine, and 10 µg/mL puromycin. Several cloning supplements have been evaluated on CHOZN® GS$^{-/-}$ platform and ClonaCell™-CHO ACF supplement (Stemcell™ Technologies) proved to increase cloning efficiency and clone productivity. This supplement was therefore used in cloning medium for assays 3 and 4. The results obtained on these assays corroborate those obtained on CHOZN® GS$^{-/-}$ platform. Cloning efficiency is increased in these conditions (46% cloning efficiency on assay 3-A-2, compared to a 21% cloning efficiency on assay 2, *i.e.*, a 2.2-fold increase – all other parameters being equal, except electroporation parameters, which are believed to impact mainly viability profile during selection). In addition, the use of a supplement may promote the re-emergence of high-producing clones that would not have re-emerged without the supplement (maximum titer of 1144 mg/L on assay 3-A-2, compared to a maximum titer of 528 mg/L on assay 2, *i.e.*, a 2.2-fold increase – all other parameters being equal, except electroporation parameters).

(iv) Clone seeding density. First single-cell cloning assays have been performed seeding clones at 0.3 viable cell/well. Clone seeding density was increased to 0.5 viable cell/well for assays 3 and 4. This results in increased number of screened clones (21 monoclonal wells for one 96-well plate seeded at 0.3 viable cell/well *versus* 29 monoclonal wells for one 96-well plate seeded at 0.5 viable cell/well, according to Poisson's distribution law), increasing the chances of identifying rare, high producing clones. Nonetheless, clone seeding density can only be increased to a certain extent when using the limiting dilution cloning method. Though, beyond a certain clone seeding density, probability of monoclonality is indeed exceeded by probability of polyclonality, according to Poisson's distribution law. Single-

cell seeders (capable of seeding clones at 1 viable cell/well) have proved to be an effective alternative to limiting dilution cloning method (Yim & Shaw, 2018).

(v) Maintenance of selection pressure over clone generation. Selection pressure was planned to be removed and effectively removed from culture medium at the time of single-cell cloning for first single-cell cloning assay. Selection pressure was though re-added in culture medium from subculturing in 24-deepweell plates. Indeed, time spent under selection pressure appeared to be insufficient (only seven days, between transfection and single-cell cloning). This was later confirmed on a DSCC of CHOZN® CHO K1, where a significant clone productivity decrease was observed between 6-well plates (culture stage equivalent to 24-deepwell plates) and TPP® tubes, when the selection pressure was removed (data not shown). As a consequence, selection pressure was maintained in culture medium from selection to subculturing in Erlenmeyer flasks for all following DSCC assays.

(vi) Promoter driving expression of resistance to selection pressure. Third and fourth single-cell cloning assays have been performed using SV40E wild type promoter, modified SV40E promoters and one CHO K1 endogenous promoter. Substitution of SV40E wild type promoter by either RAS promoter, SV40_Δ72bpRII promoter, or SV40E_Δ21bpRIII-II promoter resulted in an increased productivity – all other parameters being equal.

Interestingly, this synthesis of all results obtained so far highlights a trend towards better platform performance, with enhanced cloning efficiency and above all, increased clone productivity as the assays were conducted. The combination of the direct single-cell cloning process with a weakened selection marker made it possible to bring the performance of the CHOZN® CHO K1 platform closer to current industrial standards, with regards to productivity and cell line development timelines. CHOZN® CHO K1 platform cell line development process updated to state in 2022 is shown in Figure 56.

Limitations previously identified to these studies are recurring and are related to (i) an experimental approach being both dependent of plasmid integration site and plasmid copy number, which are known to have an impact on productivity, (ii) a limited screening capacity, resulting in partial conclusions in terms of statistical significance, and (iii) limitation of assays and optimizations to antibodies or two-chain molecules. It seems necessary to dwell on these points for further evaluation assays.

Table 21: Synthesis of direct single-cell cloning assays: experimental conditions and associated results

Expression vector backbone: Details on plasmids pEM2V-a and pEM2V-c are given in 2.5.2 Expression vector. pEM2E is identical to pEM2V-a, with the exception of the enhancer-promoter couple for heavy chain expression (murine CMV-IE1 enhancer and RAS promoter on pEM2E, instead of human CMV-IE1 enhancer and promoter on pEM2V-a) and promoter for light chain expression (RAS promoter on pEM2E, instead of murine CMV-IE1 promoter on pEM2V-a).

Promoter for puro^R expression: Resistance to puromycin is encoded by is encoded by puromycin N-acetyltransferase (PAC). Expression of pac gene is under the control of Pr_{SV40} in pEM2V-a and pEM2V-c expression vector backbones. Pr_{RAS}, $Pr_{SV40_\Delta 72bpRII}$, $Pr_{SV40_\Delta 21bpRIII\text{-}II}$, and $Pr_{SV40_\Delta TATA}$ have lower transcriptional strength compared to Pr_{SV40}. Details on these promoters are given in 3.3.4 Identification of promoters for pac expression.

Molecule: IgG_4 is a humanized mAb. Rituximab is a chimeric IgG_1.

Electroporation parameters: Traditional electroporation parameters are 1,600 V – 10 ms – 3 pulses. Modified electroporation parameters are 1,400 V – 20 ms – 2 pulses. Details on these electroporation parameters are given in 3.3.5.3.1 Modifications on protocol for mini-pool generation.

Single-cell cloning medium: Basal cloning medium composition is 80% EX-CELL® CHO Cloning medium and 20% of conditioned medium, supplemented with 6 mM L-glutamine, and 10 µg/mL puromycin. Supplemented cloning medium composition is 80% EX-CELL® CHO Cloning medium and 20% of conditioned medium, supplemented with 6 mM L-glutamine, 10 µg/mL puromycin, and 1X ClonaCell™-CHO ACF supplement (Stemcell™ Technologies, 03820).

Maintenance of selection pressure over clone generation: SF: culture medium was supplemented with 10 µg/mL puromycin during selection, subculturing in 24-deepwell plates and Erlenmeyer flasks, but not during subculturing in 96-well plates. SF: culture medium was supplemented with 10 µg/mL puromycin all along clone generation, i.e., during selection, clone subculturing in 96-well plates, 24-deepwell plates, and Erlenmeyer flasks.*

Assay	Expression vector backbone	Promoter for puro^R expression	Recombinant protein	Electroporation parameters	Lowest viability [%]	Pool recovery duration [days]	Viability on the day of single-cell cloning [%]	Single-cell cloning method	Clone seeding density [cells/well]	Cloning medium supplementation	Cloning efficiency [%]	Selection pressure maintained until	Max titer (in TPP® tubes) [mg/L]	Mean titer (in TPP® tubes) [mg/L]	Max Q_P (in TPP® tubes) [pg/cell/day]	Mean Q_P (in TPP® tubes) [pg/cell/day]
1-A-1	pEM2V-a	Pr_{SV40}	IgG4	1,600 V – 10 ms – 3 pulses	16.2 (D_7)		16.2	Limiting dilution	0.3	No	16	SF*	720	208	8	1,4
1-A-2	pEM2V-a	Pr_{SV40}	IgG4	1,600 V – 10 ms – 3 pulses	16.2 (D_7)		16.2	Cell sorter-seeder	1.0	No	10	SF*	1209	746	15,5	5,2
1-B	pEM2E	Pr_{SV40}	IgG4	1,600 V – 10 ms – 3 pulses	34.2 (D_8)	D_8-D_{11}										
2	pEM2V-c	Pr_{SV40}	Rituximab	1,600 V – 10 ms – 3 pulses	13.8 (D_7)	D_{11}-D_{13}	87.5	Limiting dilution	0.3	No	21	SF	528	135,5	5,1	1,8
3-A-1	pEM2V-c	Pr_{SV40}	Rituximab	1,600 V – 10 ms – 3 pulses	15.3 (D_8)											
3-A-2	pEM2V-c	Pr_{SV40}	Rituximab	1,400 V – 20 ms – 2 pulses	13.1 (D_8)	D_{11}-D_{13}	73.4	Limiting dilution	0.5	Yes	46	SF	1144	554	13,3	4,4
3-B-1	pEM2V-c	Pr_{RAS}	Rituximab	1,600 V – 10 ms – 3 pulses	4.7 (D_8)											
3-B-2	pEM2V-c	Pr_{RAS}	Rituximab	1,400 V – 20 ms – 2 pulses	5.1 (D_8)	D_{12}-D_{14}	77.4	Limiting dilution	0.5	Yes	48	SF	1640	1035	8	5,9
3-C-1	pEM2V-c	$Pr_{SV40_\Delta21bpRIII\text{-}II}$	Rituximab	1,600 V – 10 ms – 3 pulses	6.7 (D_8)											
3-C-2	pEM2V-c	$Pr_{SV40_\Delta21bpRIII\text{-}II}$	Rituximab	1,400 V – 20 ms – 2 pulses	7.5 (D_8)	D_{14}-D_{16}	70.5	Limiting dilution	0.5	Yes	34	SF	1552	704	11,9	7,1
4-A	pEM2V-c	Pr_{SV40}	Rituximab	1,400 V – 20 ms – 2 pulses	13.8 (D_6)	D_{10}-D_{12}	87.3	Limiting dilution	0.5	Yes	36	SF	2052	759	12,2	4,3
4-B	pEM2V-c	Pr_{RAS}	Rituximab	1,400 V – 20 ms – 2 pulses	6.7 (D_8)	D_{14}-D_{16}	85.6	Limiting dilution	0.5	Yes	29	SF	1172	864	5,6	4,6
4-C	pEM2V-c	$Pr_{SV40_\Delta72bpRII}$	Rituximab	1,400 V – 20 ms – 2 pulses	20.3 (D_8)	D_{12}-D_{14}	83.1	Limiting dilution	0.5	Yes	30	SF	2218	984	13,6	6
4-D	pEM2V-c	$Pr_{SV40_\Delta21bpRIII\text{-}II}$	Rituximab	1,400 V – 20 ms – 2 pulses	10.8 (D_8)	D_{10}-D_{12}	84.1	Limiting dilution	0.5	Yes	31	SF	3127	934	17,7	5,1
4-E	pEM2V-c	$Pr_{SV40_\Delta TATA}$	Rituximab	1,400 V – 20 ms – 2 pulses	18.8 (D_8)	D_{10}-D_{12}	84.2	Limiting dilution	0.5	Yes	34	SF	1082	418	10,4	3,3

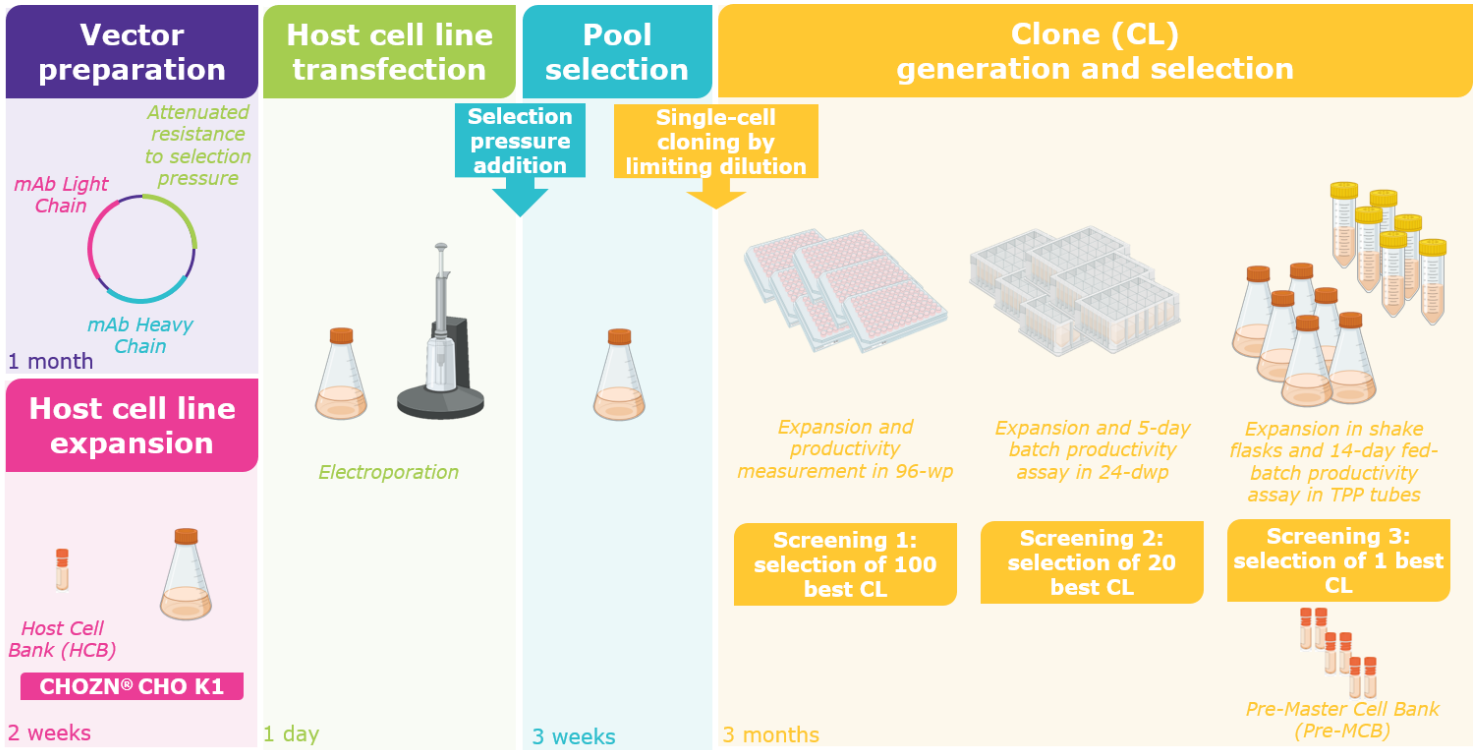


Figure 56: CHOZN® CHO K1 platform cell line development process (state in 2022)

CHOZN® CHO K1 platform cell line development process starts with host cell line thawing and subculturing. Host cell line is then transfected by electroporation. Generation of monoclonal cell lines is a two step-based process, divided in (i) a pool selection step (highlighted in blue) and in (ii) a clone (CL) generation and selection step (highlighted in yellow). The pool selection step aims at selecting stably transfected cells, while the clone generation step aims at delivering a clonally-derived, high producing cell line. Clone generation steps rely on iterative process for identification and selection of high producers: subculturing and productivity assay in 96-well plates, scale-up and productivity assay in 24-deepwell plates, and scale-up and productivity assay in TPP® tubes. Biological material is stored at the end of clone generation step, with generation of pre-master cell banks (pre-MCB). In 2022, CHOZN® CHO K1 platform cell line development process enabled DNA to pre-MCB in 4.5 months.

4.2 Insights on stability of CHOZN® CHO K1 clones

4.2.1 Instability is independent of CHO host cell line, selection and/or amplification system, and presence or absence of selection pressure

Two stability studies have been conducted on CHOZN® CHO K1 clones. The first one aimed at characterizing the impact of long-term cultivation on growth and productivity characteristics of clones either generated by long (MP-CL) or short (DSCC) cell line development process. The second one aimed at determining the impact of the attenuation of selection marker (by the use of weakened/weak promoters) on growth and productivity characteristics over long-term cultivation. These two studies tell us a lot about stability of CHOZN® CHO K1 clones.

It is widely accepted that a clone is stable if at least 70% of volumetric productivity is retained over a period that spans master cell bank (MCB) preparation, working cell bank (WCB) preparation, GMP production, and extended cell bank (ExCB) preparation, *i.e.*, at least 60 population doubling levels (PDL) (Bailey et al., 2012; Dahodwala & Lee, 2019; Dorai et al., 2012; European Medicines Agency, 1998; Lin et al., 2019).

In a similar way on these two studies, conducted with clones generated through different clone generations and producing two different molecules, half of the clones were considered unstable, *i.e.*, less than 70% of volumetric productivity was retained over 60 PDL, when selection pressure was withdrawn from sub-culturing medium. More precisely, the second stability study enabled the observation of predominantly three types of clones:

(i) Clones that were stable over 60 PDL, whether selection pressure was maintained or not during sub-culturing process (approximately 45%). This clone profile is highly desirable within an industrial context, as process robustness is, among others, impacted by cell line stability (Dahodwala & Lee, 2019; Qian et al., 2020).

(ii) Clones that were stable over 60 PDL, if selection pressure was maintained during sub-culturing process, but unstable if selection pressure is withdrawn (approximately 50%). Maintaining selection pressure during sub-culturing process in cGMP production is possible but undesired (F. Wurm & Wurm, 2017). Indeed, selection pressure is considered as a

contaminant in the manufacturing process (Mignon et al., 2015), that must be removed from the drug substance (Budge et al., 2021).

(iii) Clones that were unstable, whether selection pressure was maintained or not during sub-culturing process (approximately 5%). This clone profile is undesirable and rejected at the end of the cell line stability study from further development steps (Barnes et al., 2003).

CHO cell line instability has widely been reported in previous reports and is estimated to occur in 8-63% of all recombinant cell lines (Dahodwala & Lee, 2019). Our observations are consistent with previous reports. Kim *et al.* observed that approximately 30% of clones derived from CHOK1SV cells isolated with GS selection system and MSX lost 30% or more of their initial productivity over long-term cultivation (M. Kim et al., 2011). Similarly, Dorai *et al.* observed that a significant percentage of clones derived from CHOK1SV cells isolated with GS selection system and MSX lost 20-30% or more of their initial productivity over 55 PDL. In addition, in a majority of cell lines evaluated in this study, the decline in volumetric productivity was less important when MSX was maintained during sub-culturing process (Dorai et al., 2012). A similar observation was made by Noh *et al.* and Weidle *et al.*, who detected a more significant decrease of specific productivity in the absence of respectively MSX or MTX than in the presence of respectively MSX or MTX (Noh et al., 2018; Weidle et al., 1988). Furthermore, Lin *et al.* observed that approximately 40% of clones derived from CHO GS KO cells isolated with GS selection system and MSX lost more 30% or more of their initial productivity over 12 weeks, but this proportion was importantly decreased (15% clones were unstable) using attenuated GS mutant instead of the wild type GS enzyme (Lin et al., 2019). This last observation is consistent with our experience with CHO GS KO cells isolated with GS selection system (CHOZN® $GS^{-/-}$ platform), for which approximately 20% of clones are unstable over 60 PDL (unpublished results). The withdrawal of selection pressure is a source of instability for many high producing clones (Lin et al., 2019). These findings show that productivity instability is independent of CHO host cell line, selection and/or amplification system, and presence/absence of selection pressure (Dahodwala & Lee, 2019). Nonetheless, in general, it seems that clones derived from CHO GS KO host cell line and selected with GS selection system and MSX are more stable than clones derived from CHO DHFR KO host cell line and selected with DHFR selection system and MTX, themselves more stable than clones derived from CHO K1 host cell line and selection either with GS selection system and MSX, or with antibiotics. Cell line development platforms with selection systems where the selection pressure is continuously maintained, even in cGMP production (*i.e.*, DHFR and GS selection systems, based on auxotrophy of CHO $DHFR^{-/-}$ and CHO $GS^{-/-}$ for glycine, hypoxanthine, and thymidine and for glutamine respectively) are more efficient than selection systems where the selection pressure is removed.

4.2.2 Understanding underlying causes of productivity instability…

Even if several teams focused their efforts on understand the underlying causes of instability, many mechanisms still need to be understood (Ritter et al., 2016). To date, productivity instability has been explained by gene copy loss (Bailey et al., 2012; Fann et al., 2000; Jun et al., 2006; M. Kim et al., 2011), unstable chromosomal structures (Chusainow et al., 2009; Jun et al., 2006), gene silencing as a result of epigenetic regulation (Chusainow et al., 2009; M. Kim et al., 2011; Osterlehner et al., 2011; Paredes et al., 2013), or increased susceptibility to cellular stresses (Bailey et al., 2012; Chusainow et al., 2009). Here, unfortunately, the design itself of the first stability study did not allow for identification of root causes of clone instability. The absence of a control condition, for which selection pressure would be maintained during sub-culturing process, and the absence of analytics prevented from proposing hypotheses on the causes of instability. However, the design of the second stability study was much more complete, including a control condition and additional analytics (namely, gene copy number analysis and heavy chain and light chain transcription level measurement). Regrettably, these analyses could not be performed before the end of this project, therefore still letting the door open to numerous questions.

It will be interesting to continue research efforts to understand the underlying causes of productivity instability in CHOZN® CHO K1 clones. In first instance, simple analyses can be performed, such as gene copy number measurement by quantitative PCR (qPCR) or digital droplet PCR (ddPCR) (Whale et al., 2012). Analysis of gene copy number has been widely used to explain productivity decrease over time by gene copy loss (Bailey et al., 2012; Betts & Dickson, 2015; Fann et al., 2000; M. Kim et al., 2011; Nematpour et al., 2017; Qian et al., 2020). Fluorescence *in situ* hybridization (FISH) has also been commonly used to reveal plasmid integration site and possible chromosomal rearrangements (Pallavicini et al., 1990; Weidle et al., 1988; Yoshikawa, Nakanishi, Ogura, et al., 2000). In second instance, if productivity instability is not explained by gene copy loss, further analyses could be performed to attribute instability to molecular mechanisms linked to protein expression (transcriptional, translational, and post-translational levels). Measurement of heavy chain and light chain transcription level by RT-qPCR could indicate gene silencing, directing, for example to bisulfite sequencing to further detect DNA methylation (M. Kim et al., 2011; Kurdyukov & Bullock, 2016; Spencer et al., 2015). Recently, omics technologies (and especially transcriptomics) have also been used to characterize cell line instability (H. Li et al., 2015; Qian et al., 2020).

4.2.3 … to find approaches to enhance productivity stability

Identifying the underlying causes of productivity instability may help to find approaches to enhance stability of production, which is, as well as productivity, of high importance from an industrial and regulatory point of view (Barnes et al., 2003). Indeed, several teams reported that cell line instability is clone-specific, certainly acquired during cell line development process as clones derived from the same

transfection can either be stable or unstable, and possibly controllable if underlying causes are better understood (Qian et al., 2020). Many approaches have been described by research groups and sometimes implemented in industrial cell line development platforms. For example, expression systems compatible with targeted integration of the expression within the host cell genome are thought to be relevant to avoid position effect (Barnes et al., 2003). Moreover, ubiquitous chromatin opening elements (UCOEs) proved to significantly enhance stability of recombinant protein expression (Betts & Dickson, 2015; Nematpour et al., 2017; Neville et al., 2017; Saunders et al., 2015; F. Zhang et al., 2010), and more efficiently than other chromatin remodeling elements (such as scaffold or matrix attachment regions (S/MARs), stabilizing anti-repressor elements (STARs) and insulators (Saunders et al., 2015)). Additionally, engineered viral promoters (Moritz et al., 2015; Romanova & Noll, 2018), promoters free of CpG islands (Ho et al., 2016), synthetic promoters (A. J. Brown & James, 2016; Zúñiga et al., 2019), and endogenous promoters (Tossolini et al., 2022) are all possible ways for vector optimization to decrease productivity instability.

Lastly, new selection systems, and especially those relying on natural auxotrophy of CHO K1 cells (Budge et al., 2021; Roca et al., 2019; Sun et al., 2020), could be of particular interest (Dahodwala & Lee, 2019). Selection systems based on natural auxotrophy of host cell line indeed offer the advantages (i) to keep selection pressure even during cGMP production, which is more effective than selection systems were selection pressure has to be removed (Lin et al., 2019), and (ii) to avoid deploying complex genome editing technologies that require costly development and significant validation efforts (Kelley et al., 2018).

4.3 Towards next version of CHOZN® CHO K1 platform

4.3.1 Multiple routes to high producing recombinant cell lines

Literature review revealed a multitude of optimization areas for CHOZN® CHO K1 platform. Based on the literature review and on the optimization work done on the CHOZN® CHO K1 platform previously, we decided to focus our work on selection stringency.

There are, of course, multiple routes of optimization to high producing recombinant cell lines. Seth *et al.* interestingly explained that acquiring superior attributes from several functional classes, such as protein secretion, energy metabolism, redox balance or growth/death control, would lead to higher productivities. Combinations of attributes could be additive, augmentative, or non-additive with regard to productivity. If different combinations of attributes could confer the same level of enhanced productivity, it is though highly probable that high producing cell lines bear enhanced attributes of multiple significant pathways (Seth et al., 2007).

Research and understanding of the critical elements for high and stable expression of recombinant proteins in CHO cells must therefore continue. Results obtained with weakened/weak promoters driving selection marker expression confirmed that increasing selection stringency is an interesting approach for improving recombinant cell line productivity. Several suggestions have been made to further this work, namely increasing again selection stringency, with combination of attenuated selection markers, and evaluating these elements with a novel selection system, relying on natural auxotrophy of CHO K1 cells.

4.3.2 Reshaping cell line development, for shorter DNA to IND timelines

Cell line development is part of the global context of recombinant therapeutic molecule development, which encompasses all activities of cell line development, process development (upstream and downstream processes), analytical development, and GMP production for clinical trials. Standard industry timelines from DNA to investigational new drug/investigational medicinal product (IND/IMP, *i.e.*, request to obtain authorization from heath authorities to administer an investigational drug or biological product to humans) application/dossier filling are 18 months, and, in the fastest cases, 10 to 12 months (Z. Zhang et al., 2021). However, significant efforts are currently made to accelerate these timelines, especially for unmet medical needs (Fan et al., 2017), or, as shown in the 2020s, in the context epidemics or pandemics (Kelley, 2020; Z. Zhang et al., 2021). Health authorities have namely developed accelerated approval processes, such as breakthrough designation (introduced by FDA in 2012) and Priority Medicines (PRIME) designation (introduced by EMA in 2016).

Cell line development is a decisive step in a recombinant therapeutic molecule development, as this activity is responsible for generating the production cell line (Diep et al., 2021), and is on the critical path for advancing a novel therapeutic to first-in-human (FIH) clinical trials (Munro et al., 2017). Nonetheless, generation of a recombinant cell line is a particularly long and tedious process, spanning 3 to 4 months in most industrial cell line development platforms. Initial cell line generation timelines with CHOZN® CHO K1 platform were spanning 7.5 months, and were reduced to 4.5 months (*i.e.*, a reduction of 40%) with the implementation of direct single-cell cloning (DSCC) process. The implementation of shortened processes was an absolute necessity and efforts should pursue in this direction.

As well as health authorities, biopharmaceutical industries have also proposed creative approaches to further shorten new drug development timelines. Use of transient pools (Rodriguez-Conde et al., 2022; Stuible et al., 2018; G. Xu et al., 2022), stable pools (Munro et al., 2017; Rajendra et al., 2017; Rajendran et al., 2021; Schmieder et al., 2022), mini-pools (Fan et al., 2017; G. Xu et al., 2022), and pools of clones (Bolisetty et al., 2020; Hu et al., 2017) are all strategies that are now increasingly being described in the literature for early production of therapeutic recombinant protein for pre-clinical and/or clinical studies. These approaches are often relying on semi-targeted or targeted integration strategies, as well as

comparative studies of pool-derived *versus* clone-derived material (Fan et al., 2017; G. Xu et al., 2022). In contrast, even if the importance of single-cell cloning has been highly debated within the industry (Frye et al., 2016; Scarcelli et al., 2018; F. Wurm & Wurm, 2017), clonality is though still a requirement for health authorities (Welch & Arden, 2019), making single-cell cloning an essential step in any cell line development process.

The next optimization works on the CHOZN® CHO K1 platform should follow this trend, which might lead to an important reshaping of cell line generation process. The different activities linked to cell line development should therefore be thought out and arranged in such a way as to (i) reduce cell line development timelines, (ii) initiate the USP development steps at the earliest, and (iii) allow the generation of biological material for early studies, later required for the IND.

5 Conclusion

The objective of the thesis presented here was to understand the critical parameters of the cell line generation process for the expression of recombinant proteins in CHO cells, from the point of view of each element of a cell line development platform: the host line, the expression vector, and the stages of the cell line generation process. This work was part of the broader context of the optimization of the CHOZN® CHO K1 cell line generation platform, which is one of the two cell line generation platforms currently offered by Merck End-to-End Solutions for its CDMO activity.

The characteristics and performance of the CHOZN® CHO K1 cell line generation platform were first defined in a process based on a mini-pool generation and a clone generation. This initial characterization led to the production of clones producing more than 1 g/L of a standard monoclonal antibody in 7.5 months. Both productivity level and cell line development timelines were out of the standards of industrial cell line development platforms, and therefore steered to the identification of two areas of research: optimization of the clonal cell line generation process to reduce cell line development times, and optimization of the expression vector to increase the productivity of generated cell lines.

Cell line development process associated to CHOZN® CHO K1 platform was a two-step based iterative process, divided in (i) a mini pool generation and selection step, aiming at easing high producing pools identification and in (ii) a clone generation and selection step, aiming at delivering a clonally derived, high producing cell line. An accelerated cell line development process, called direct single-cell cloning, was evaluated. Mini-pool generation and selection step was replaced by a 3-week bulk pool selection step, followed by the clone generation and selection step (remained unchanged). This first direct single-cell cloning assay is considered as the proof of concept: it indeed showed the technical feasibility of this process on the CHOZN® CHO K1 platform. This process reduced the duration of the line development process by almost half, but, as one might expect, led to a decrease in the productivity of the generated cell lines.

Previous optimization work on CHOZN® CHO K1 platform focused on genetic elements for recombinant protein expression. However, few investigations had been carried out on the elements allowing the expression of resistance to selection pressure, while selection stringency has been described as a major lever for increasing recombinant cell lines' productivity. As transcription is traditionally considered as the dominant factor controlling expression, this research work was focused on the promoter driving puromycin N-acetyltransferase transcription. Several weakened promoters (partial deletions in SV40E wild type promoter) and weak promoters (substitution of SV40E wild type promoter by CHO K1 endogenous promoters) were evaluated for the expression of resistance to selection pressure. Some promoters, used with direct single-cell cloning process previously evaluated, led to the identification in 4 months of clones producing more than 3 g/L of a standard monoclonal antibody. In

general, the use of these promoters led to an increase of productivity of generated cell lines, without significant impact on expression level stability over 60 PDL. This productivity increase was confirmed on several monoclonal antibodies (both IgG1 and IgG4) and a Fc fusion protein.

Common limitations have been identified on both studies and are related to (i) an experimental approach being both dependent of plasmid integration site and plasmid copy number, which are known to have an impact on productivity, (ii) a limited screening capacity, resulting in partial conclusions in terms of statistical significance, and (iii) limitation of assays to antibodies or two-chain molecules. Nonetheless, and in spite of these limitations, several critical parameters for the expression of recombinant proteins in CHO cells have been identified, on two elements of a cell line development platform: the expression vector and the cell line development process.

The results obtained during this work have an immediate industrial application, through the optimization of the CHOZN® CHO K1 cell line generation platform. First of all, the direct single-cell cloning process has become the standard cell line development protocol applied on CDMO projects with CHOZN® CHO K1 platform and now includes combination of optimized parameters, such as viability threshold for single-cell cloning and cloning medium supplementation. Secondly, attenuated promoters could be implemented as soon as the next CDMO project. Thirdly, the qPCR and RT-qPCR methods, respectively developed for gene copy number and transcription level monitoring, pave the way for the implementation of additional methods to characterize clones during cell line development process. Lastly, these optimizations could also be incorporated in the CHOZN® $GS^{-/-}$ platform, which is, to date, the second cell line development platform offered for Merck End-to-End Solutions' CDMO activity. Contrary to CHOZN® CHO K1 platform, a direct single-cell cloning assay run with CHOZN® $GS^{-/-}$ platform failed, as proof of concept, to show technical feasibility of this process for cell line generation with CHOZN® $GS^{-/-}$ platform. A lot of work will certainly be necessary to achieve the implementation of the direct single-cell cloning process on the CHOZN® $GS^{-/-}$ platform. However, weakened/weak promoters are likely to be easily and quickly implemented in the CHOZN® $GS^{-/-}$ platform. Indeed, similar work, aiming at increasing selection stringency, although sometimes targeting other attenuation strategies, proved to be efficient in CHO $GS^{-/-}$ cells selected with GS selection system.

The combination of the direct single-cell cloning process with a weakened selection marker made it possible to bring the performance of the CHOZN® CHO K1 platform closer to current industrial standards, with regards to productivity (up to 3 g/L for a standard monoclonal antibody) and cell line development timelines (DNA to pre-MCB in 4.5months). This work therefore emphasizes that, without massive investment on equipment or technology, performance of the CHOZN® CHO K1 platform can be significantly improved, thus highlighting the potential of this platform.

Nonetheless, efforts are still necessary to continue improving the performance of the platform. The continuation of the work on selection stringency, the resumption of evaluation of promoters (namely endogenous promoters) for recombinant protein expression, and research of markers associated to high productivity particularly deserve to be explored. Moreover, and as shown in this work, the technological developments of methods are just as important. Implementation of a semi-targeted integration system, and of miniaturized culture scales are expected to highly contribute to platform performance optimization. Lastly, as cell line development is on the critical path for advancing a novel therapeutic to first-in-human clinical trials, these optimizations must meet the need for acceleration for recombinant therapeutic molecule development.

Research and understanding of the critical elements for high and stable expression of recombinant proteins in CHO cells must therefore continue, to pursue optimization of the CHOZN® CHO K1 platform, that has strong and unique characteristics to be an efficient and competitive cell line development platform.

6 Material and methods

6.1 Analytical methods related to cell culture methods

6.1.1 Viable cell density and viability measurement

Viable cell concentration and cell viability were measured using a Vi-CELL™ XR Cell Viability Analyzer (Beckman Coulter) automated cell counter. Vi-CELL™ XR counting method relies on Trypan blue dye exclusion method combined with digital image analysis. Vi-CELL™ XR automatically performs cell suspension sampling (at least 500 µL of sample being previously deposited in a sampling cup), Trypan blue staining, image analysis and cleaning steps.

Following parameters were used for viable cell concentration and cell viability measurements: minimum diameter: 7 µm; maximum diameter: 30 µm; 30 images; 2 aspirate cycles; 2 Trypan blue mixing cycles; cell brightness: 85%; cell sharpness: 60%; viable cell spot brightness: 85%; viable cell spot area: 7%; minimum circularity: 0%; decluster degree: medium.

6.1.2 Cell confluence monitoring

Cell confluence was monitored using a Cell Metric® (Solentim) plate imager, often coupled with an automated platform including an Orbitor™ (Thermo Fisher Scientific) automated arm and a Cytomat ™ (Thermo Fisher Scientific) automated incubator.

A scan profile was defined according to the culture plate reference, the focus (automatically defined according to the culture plate reference), the reading mode, the plate map, and the cell detection parameters. Specific scan profiles were defined and used respectively for CHOZN® CHO K1 mini-pools and CHOZN® CHO K1 clones.

6.1.3 Glucose and lactate concentrations measurement

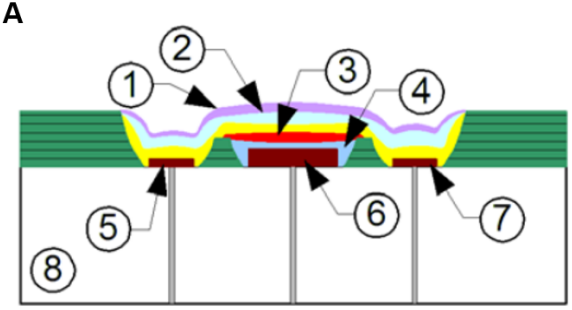
Glucose and lactate concentrations were measured using an ABL90 FLEX (Radiometer) gas analyzer. ABL90 FLEX gas analyzer is a blood gas analyzer generally used by hospital departments for analysis of blood gas, metabolites, electrolytes, and oximetry. Glucose and lactate concentrations measurement relies on specific enzymatic reaction coupled to the amperometric measuring principle, where the magnitude of an electrical current flowing through an electrode is related to the concentration of a substance being oxidized or reduced at an electrode of the chain. The amperometric measuring principle is illustrated on Figure 57.

6.1.4 pH measurement

pH was measured using the ABL90 FLEX gas analyzer. pH measurement relies on the potentiometric measuring principle, where the potential of an electrode chain recorded at the voltmeter is related to the

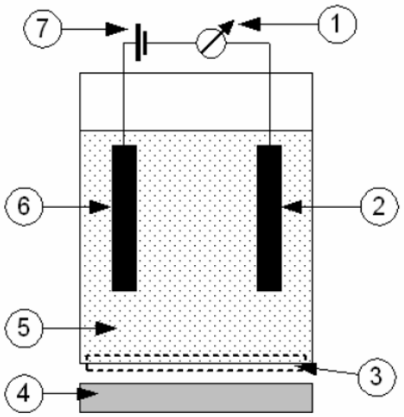
concentration of a substance via the Nernst equation. The potentiometric measuring principle is illustrated on Figure 58.

A



Item	Part	Description
①	Biocompatible layer	Biocompatible layer
②	Outer membrane	Permeable to glucose/lactate
③	Enzyme layer	Contains glucose/lactate oxidase
④	Inner membrane	Cellulose acetate
⑤	Reference	Ag/AgCl electrode
⑥	Anode	Platinium electrode
⑦	Cathode	Platinium electrode
⑧	Electrode base	The structural platform on which the sensor is formed

B



Item	Part	Description
①	Amperemeter	Measures the current flowing through the circuit
②	Cathode	Negative electrode where a reduction reaction occurs and electrons are consumed
③	Membrane	Consists of four layers
④	Sample	Contacts the membrane
⑤	Electrolyte	Provides electrical contact betwen the anode and the cathode
⑥	Anode	Positive electrode where an oxydation reaction occurs and electrons are released
⑦	Applied voltage	Applies the necessary potential for the reduction or oxidation reaction under study

C

Glucose oxidase

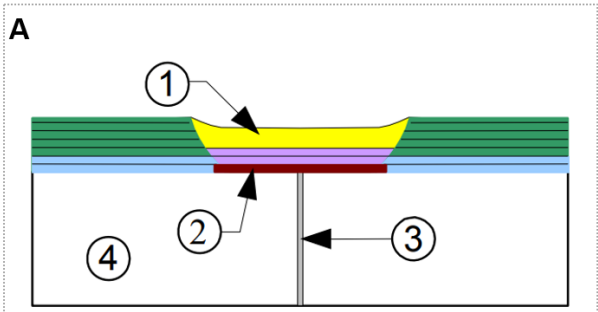
$$\text{Glucose} + H_2O + O_2 \rightarrow \text{Gluconic acid} + H_2O_2$$

Lactate oxidase

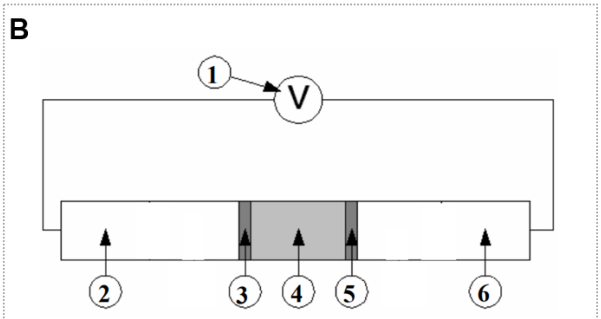
$$\text{Lactate} + H_2O + O_2 \rightarrow \text{Pyruvate} + H_2O_2$$

Figure 57: Principle of glucose and lactate concentrations measurement

(A) *Glucose or lactate sensor. Glucose and lactate sensors are three-electrode sensors consisting of an internal silver/silver chloride reference electrode, a platinum auxiliary electrode, and a platinum anode. The sensors are covered by a multi-layer membrane bound to the sensor board.* ***(B)*** *Electrode chain for glucose and lactate concentrations measurement. Glucose and lactate concentrations measurement relies on the amperometric measuring principle, where the magnitude of an electrical current flowing through an electrode is related to the concentration of a substance being oxidized or reduced at an electrode of the chain.* ***(C)*** *Glucose and lactate oxidation reactions. Adapted from ABL90 FLEX reference manual.*



Item	Part	Description
①	Membrane	Ion-selective membrane that is in direct contact with the sample or calibration solution and that is sensitive to H^+ ions
②	Solid-state contact	The point of electrical and ionic contact with the membrane
③	Electrical contact	The point of electrical contact between the sensor and the analyzer
④	Electrode base	The structural platform on which the electrode is formed



Item	Part	Description
①	Voltmeter	Measures the voltage potential in the circuit
②	Reference electrode	Provides electrical connection to the voltmeter
③	Liquid junction	Point of contact between the reference sensor and the sample
④	Sample	The unknown liquid being measured
⑤	Membrane	An ion-sensitive membrane, which is sensitive to H^+ ions
⑥	Solid-state contact	Provides electrical connection to the voltmeter

Figure 58: Principle of pH measurement

(A) *pH sensor. pH sensor is of solid-state design with a H+ sensitive PVC membrane.* ***(B)*** *Electrode chain for pH measurement. pH measurement relies on the potentiometric measuring principle, where the potential of an electrode chain recorded at the voltmeter is related to the concentration of a substance via the Nernst equation. Adapted from ABL90 FLEX reference manual.*

6.1.5 Antibody quantification

Antibody quantification was performed by bio-layer interferometry (BLI), which is illustrated on Figure 59. Bio-layer interferometry is an optical analytical technique relying on the measure of interference patterns between waves of light down-directed in optic biosensors. Surface of biosensors is composed of two interfaces: an internal reference layer and a biocompatible layer on which ligand molecules (protein A, in the case of antibody quantification) are immobilized. Light is directed down the biosensor and is reflected from both layers. Interference pattern of reflected beams is detected by a charge coupled device (CCD) captor. When the target molecule (IgG_1, IgG_2, or IgG_4, in the case of antibody quantification) binds to ligand molecules immobilized to the biocompatible layer, the distance between the two layers increases, resulting in a shift in the interference pattern of the reflected light. Therefore, changes in the spectral pattern of the reflected light indicate the quantity of target molecule in the sample.

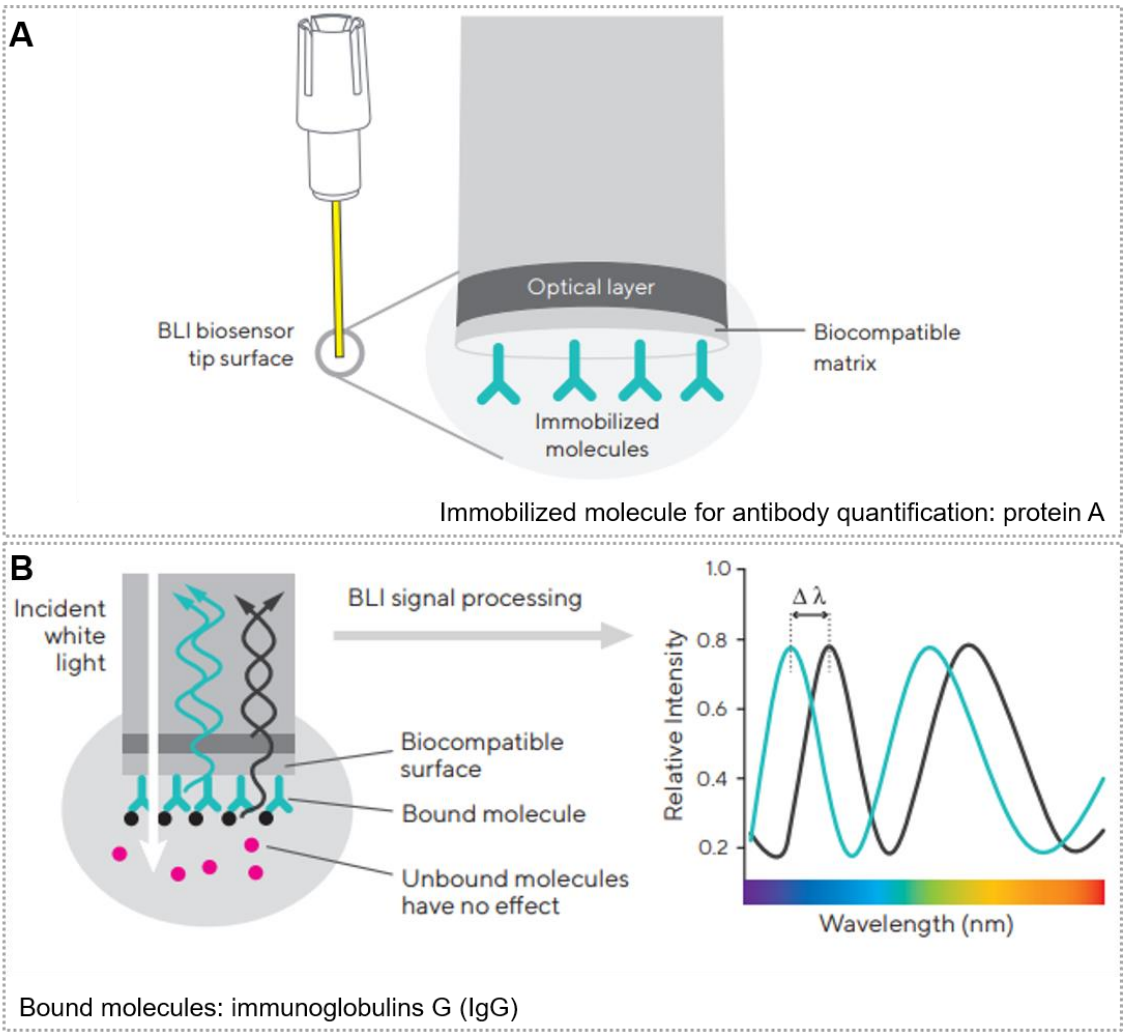


Figure 59: Principle of bio-layer interferometry

(A) *Details on surface of bio-layer interferometry sensors. Surface of biosensors is composed of two interfaces: an internal reference layer and a biocompatible layer on which ligand molecules (protein A, in the case of antibody quantification) are immobilized.* ***(B)*** *Principle of bio-layer interferometry. Light is directed down the biosensor and is reflected from both layers. Interference pattern of reflected beams is detected by a charge coupled device (CCD) captor. Changes in the spectral pattern of the reflected light indicates the quantity of target molecule in the sample. Adapted from Application Note Biomolecular Binding Kinetics Assays on the Octet® Platform (Sartorius).*

Antibody quantification was performed using Octet RED96 (Sartorius) and protein A (ProA) coated biosensors (Sartorius, 18-5012). ProA biosensors were equilibrated in 0.02% Tween 20, 150 mM NaCl, 1 mg/mL BSA, PBS 1X (equilibration buffer) for 30 minutes prior to use. Binding rates of each sample were measured for 2 minutes. Biosensors were regenerated 5 seconds, 400 rpm in 10 mM glycine-HCl (regeneration buffer) and equilibrated 5 seconds, 400 rpm in equilibration buffer with three cycles, nine times. Titration plates were filled with 200 µL of sample, either undiluted or diluted in equilibration buffer. Dilution factor was estimated according to culture stage and re-tests were performed if required.

Raw data were acquired using Octet Data Acquisition software 9.0 (Sartorius) and analyzed using Octet Data Analysis software 9.0 (Sartorius). The concentrations of monoclonal antibody in supernatants were

calculated from a standard curve generated in-lab. IgG4 standard (internally produced), rituximab standard (Sigma-Aldrich, MSQC17-0.5 mg), trastuzumab standard (Sigma-Aldrich, MSQC22-0.5 mg), pembrolizumab standard (Sigma-Aldrich, MSQC24-0.5 mg), nivolumab standard (Sigma-Aldrich, MSQC26-0.5 mg) and EPO-Fc standard (internally produced) were respectively used for IgG4, rituximab, trastuzumab, pembrolizumab, nivolumab, and EPO-Fc fusion quantification.

6.2 Cell culture methods

6.2.1 Preparation of culture media

EX-CELL® CD CHO Fusion medium is a chemically defined (CD), animal component free (ACF) medium developed for long-term growth of CHOZN® cells. EX-CELL® CD CHO Fusion medium was either used in liquid form (Sigma-Aldrich, 14365C) or in powder form (Sigma-Aldrich, 24365C).

EX-CELL® CHO Cloning medium is an animal component free medium developed to support clonal survival and growth of CHOZN® cells. EX-CELL® CHO Cloning medium was used in liquid form (Sigma-Aldrich, C6366).

EX-CELL® Advanced CHO Fed-batch medium is a chemically defined medium specially developed for fed-batch cultures. EX-CELL® Advanced CHO Fed-batch medium was either used in liquid form (Sigma-Aldrich, 14366C) or in powder form (Sigma-Aldrich, 24366C). The concentrated feed media added in fed-batch cultures will be detailed in specific paragraphs (6.2.10.2 Fed-batch production in TPP® tubes).

Conditioned medium (sometimes referred to as cell secretome) is a medium prepared with healthy exponentially growing cells and which contains enzymes, growth factors, cytokines and hormones, and other soluble mediators that are fundamental in the process of cell growth. CHOZN® CHO K1 host cell line suspension volume necessary for a seeding at $1.0x10^6$ viable cells/mL was centrifuged for 5 minutes, 200 *g*, at room temperature. Supernatant was removed and cell pellet was resuspended in EX-CELL® CD CHO Fusion supplemented with 6 mM L-glutamine. 24 hours after seeding, cell suspension was centrifuged for 5 minutes, 2,440 *g*, at room temperature. Supernatant was collected and sterile filtered on a Stericup-GV Sterile Vacuum Filtration System 0.2 µm PVDF membrane (Millipore, SCGVU01RE). The obtained conditioned medium was either freshly used or stored at 4°C for 5 days maximum.

All culture media were supplemented with 6 mM L-glutamine (Gibco, 25030081). Depending on conditions, culture media were also supplemented with 10 µg/mL puromycin (Gibco, A1113803).

6.2.2 CHOZN® CHO K1 host cell line sub-culturing

CHOZN® CHO K1 cells (Sigma-Aldrich, CHOK1-1VL) were cultured in EX-CELL® CD CHO Fusion medium supplemented with 6 mM L-glutamine in 125 mL Erlenmeyer flasks (Corning, 431143), at 37°C, 80% relative humidity, 5% CO_2, and 150 rpm (10 mm shaking diameter) or 130 rpm (25 mm shaking diameter). Cells were passaged at a seeding density of $4x10^5$ viable cells/mL every two days, or $3x10^5$ viable cells/mL every three days, or $2x10^5$ viable cells/mL every four days.

6.2.3 Thawing and freezing of cellular banks

6.2.3.1 Thawing of cellular banks

Cellular bank vial was thawed using the ThawSTAR® (Stemcell) automated cell thawing system. Thawed cellular suspension was diluted in 10 mL of thawing medium and centrifuged for 5 minutes, 200 *g*, at room temperature. Cell pellet was resuspended in 10 mL of thawing medium. Details on thawing medium composition are given in Table 22.

Viable cell concentration and viability were then measured. Cell suspension was diluted to usual seeding densities in 125 mL Erlenmeyer flasks (or in any appropriate shaken container, namely TPP® tube (TPP, 87050) or 24-deepwell plate (Enzyscreen, CR1424cl), depending on cell suspension volume). Erlenmeyer flasks were incubated at 37°C, 80% relative humidity, 5% CO_2, 150 rpm (10 mm shaking diameter) or 130 rpm (25 mm shaking diameter).

6.2.3.2 Freezing of cellular banks

CHOZN® CHO K1 host cell line and derived cell lines were cryopreserved. Appropriate cell suspension volume ($10x10^6$ cells per vial) was pelleted by centrifugation for 10 minutes, 200 *g*, at room temperature. Cell culture supernatant was removed. Cell pellets were carefully dried and re-suspended in freezing medium (1 mL per vial). Freezing medium was a blend of 93% (v/v) EX-CELL® CD CHO Fusion supplemented with 6 mM L-glutamine and 7% (v/v) DMSO (Sigma-Aldrich, D2438) sterile filtered using Millex-VV Syringe Filter Unit, 0.1 µm, PVDF (Millipore, SLVV033RS). Vials were stored in Mr. Frosty™ Freezing Container (ThermoFisher Scientific, 5100-0001) filled with isopropanol (SAFC, 1.00995) at < -70°C, at least 18 hours and at most 7 days. Vials were then stored at -196°C for long-term storage.

Table 22: Medium used for thawing of cellular banks – Details per assays

Cell type	Thawing medium	References to results
CHOZN® CHO K1 host cell line	EX-CELL® CD CHO Fusion + 6 mM L-glutamine	3.1 Characterization of CHOZN® CHO K1 cell line development platform performance 3.2 Cell line development process optimization: implementation of direct single-cell cloning

Cell type	Thawing medium	References to results
		3.3 Expression vector optimization: evaluation of promoters for expression of resistance to puromycin
CHOZN® CHO K1 mini-pools	EX-CELL® CD CHO Fusion + 6 mM L-glutamine	
CHOZN® CHO K1 clones	EX-CELL® CD CHO Fusion + 6 mM L-glutamine	3.1 Characterization of CHOZN® CHO K1 cell line development platform performance 3.2 Cell line development process optimization: implementation of direct single-cell cloning 3.3 Expression vector optimization: evaluation of promoters for expression of resistance to puromycin
	EX-CELL® CD CHO Fusion + 6 mM L-glutamine + 10 µg/mL puromycin	3.3 Expression vector optimization: evaluation of promoters for expression of resistance to puromycin

6.2.4 Transfection by electroporation

6.2.4.1 Electroporation

Transfection is performed by electroporation. For one electroporation replicate, $5.5x10^6$ viable cells were centrifuged for 10 minutes, 400 *g*, at room temperature. Supernatant was removed, and cell pellet was resuspended in 110 µL of EX-CELL® CD CHO Fusion supplemented with 6 mM L-glutamine. Cell suspension was mixed with 33 µg DNA (expression vector). Cells were electroporated using a Neon device (Thermo Fisher Scientific) and 100 µL Neon kit (Thermo Fisher Scientific). Electroporated cells were seeded in pre-warmed EX-CELL® CD CHO Fusion supplemented with 6 mM L-glutamine in T-flasks (Corning, 431463) and incubated at 37°C, 80% relative humidity, and 5% CO_2. Details on electroporation mix, electroporation parameters, number of electroporation replicates, and seeding of electroporation replicates are given in Table 23.

Table 23: Number of electroporation replicates and electroporation parameters – Details per assays

Electroporation mix (cells-DNA-medium)	Electroporation parameters	Number of electroporation replicates per T-flask	Seeding of electroporation replicates	References to results
$22x10^6$ viable cells; 132 µg DNA; 440 µL of medium	1,600 V, 10 ms, 3 pulses	3	15 mL (TF 75 cm²) 1 TF / condition	3.1 Characterization of CHOZN® CHO K1 cell line development platform performance
$22x10^6$ viable cells; 132 µg DNA; 440 µL of medium	1,600 V, 10 ms, 3 pulses	5	22.5 mL (TF 75 cm²)	3.2 Cell line development process optimization: implementation of direct single-cell cloning

Electroporation mix (cells-DNA-medium)	Electroporation parameters	Number of electroporation replicates per T-flask	Seeding of electroporation replicates	References to results
			1 TF / condition	
$22x10^6$ viable cells; 132 µg DNA; 440 µL of medium	1,600 V, 10 ms, 3 pulses	1	5 mL (TF 25 cm²) 3 TF / condition	3.3 Expression vector optimization: evaluation of promoters for expression of resistance to puromycin
$5.5x10^6$ viable cells; 33 µg DNA; 110 µL of medium	1,600 V, 10 ms, 3 pulses	1	5 mL (TF 25 cm²) 1 TF / condition	3.3 Expression vector optimization: evaluation of promoters for expression of resistance to puromycin
$22x10^6$ viable cells; 132 µg DNA; 440 µL of medium	1,600 V, 10 ms, 3 pulses	3	15 mL (TF 75 cm²) 3 TF / condition	3.3 Expression vector optimization: evaluation of promoters for expression of resistance to puromycin
$5.5x10^6$ viable cells; 33 µg DNA; 110 µL of medium	1,400 V, 20 ms, 2 pulses	1	5 mL (TF 25 cm²) 1 TF / condition	3.3 Expression vector optimization: evaluation of promoters for expression of resistance to puromycin
$22x10^6$ viable cells; 132 µg DNA; 440 µL of medium	1,600 V, 10 ms, 3 pulses	5	15 mL (TF 75 cm²) 2 TF / condition	3.3 Expression vector optimization: evaluation of promoters for expression of resistance to puromycin
$22x10^6$ viable cells; 132 µg DNA; 440 µL of medium	1,400 V, 20 ms, 2 pulses	5	15 mL (TF 75 cm²) 2 TF / condition	3.3 Expression vector optimization: evaluation of promoters for expression of resistance to puromycin
$11x10^6$ viable cells; 66 µg DNA; 220 µL of medium	1,400 V, 20 ms, 2 pulses	1	5 mL (TF 25 cm²) 1 TF / condition	3.3 Expression vector optimization: evaluation of promoters for expression of resistance to puromycin

6.2.4.2 Controls for transfection

Controls for transfection are performed for each series of electroporation to determine transfection efficiency. Negative control is performed by replacing DNA solution volume by water. Positive control is performed with a green fluorescent protein (GFP) expression vector. Negative and positive control electroporations are performed following the same protocol as described before.

24 hours after transfection, $1x10^6$ viable cells from both negative and positive cellular suspensions were centrifuged for 5 minutes, 200 *g*, at room temperature. Supernatant was removed, and cell pellet was resuspended in 500 µL of DPBS (Gibco, 14190). To determine transfection efficiency, fluorescence intensity was measured on a Guava® easyCyte™ (Luminex) flow cytometer. Negative control cell suspension was used to validate acquisition parameters and positive control cell suspension was used to determine transfection efficiency. Raw data were acquired and analyzed using Guava® InCyte software (Luminex).

6.2.5 Transient gene expression

6.2.5.1 Transient recombinant protein expression

24 hours after transfection, 3 mL of transfected bulk pool cell suspension were transferred in a 6-well plate incubated at 37°C, 80% relative humidity, and 5% CO_2 for transient recombinant protein expression assay. Four days after seeding, cell suspension viable cell density and viability were measured. Cell suspensions were subsequently centrifuged 5 minutes, 200 *g*, at room temperature. Supernatants were stored at < -70°C for later antibody titer measurement.

6.2.5.2 Transient puromycin N-acetyltransferase expression

Transfected bulk pool cell suspensions were incubated at 37°C, 80% relative humidity, and 5% CO_2 for transient puromycin N-acetyltransferase expression assay. Two days after seeding, 2 mL of fresh EX-CELL® CD CHO Fusion supplemented with 6 mM L-glutamine were added in each T-flask. Four days after seeding, cell suspension viable cell density and viability were measured. $5x10^6$ cells were pelleted by centrifuging appropriate cell suspension volume for 5 minutes at 400 *g*. Cell culture supernatant was removed. Cell pellets were carefully dried and stored at < -70°C for later *pac* transcription level measurement by RT-qPCR.

6.2.6 Mini-pools generation: seeding and expansion in 96-well plates

24 hours after transfection, transfected bulk cell suspensions were centrifuged for 5 minutes, 200 *g*, at room temperature. Supernatants were removed, and cell pellets were resuspended at $2.5x10^4$ viable cells/mL in 20% (v/v) EX-CELL® CD CHO Fusion, 80% EX-CELL® CHO Cloning (Sigma-Aldrich, C6366) supplemented with 6 mM L-glutamine and 10 µg/mL puromycin (Gibco, A1113803). 96-well plates (Corning, 3595) were seeded with 200 µL of cell suspension (*i.e.*, 5,000 viable cells/well). External wells of the 96-well-plates (*i.e.*, columns 1 and 12 and rows A and H) were filled with sterile DPBS (Gibco, 14190). Plates were incubated at 37°C, 80% relative humidity, and 5% CO_2.

Mini-pool recovery was monitored by confluence measurement using Cell Metric plate imager (Solentim). Plates were imaged one hour after mini-pool seeding and then at least twice a week. Medium renewal was performed by removing 100 µL of supernatant and adding 120 µL of EX-CELL® CD CHO Fusion supplemented with 6 mM L-glutamine and 10 µg/mL puromycin twice a week.

Once confluence approximately reached 50%, mini-pools were diluted in EX-CELL® CD CHO Fusion supplemented with 6 mM L-glutamine and 10 µg/mL puromycin in a new 96-well plate. Dilution factor was determined by targeting at least 20% of confluence.

When confluence approximately reached 50% after the first dilution, productivity assay was launched. Mini-pools were diluted 1:5 in 200 µL of EX-CELL® CD CHO Fusion supplemented with 6 mM L-glutamine and 10 µg/mL puromycin in a new 96-well plate. Plates were incubated at 37°C, 80% relative

humidity, and 5% CO_2 for seven days. Confluence and recombinant antibody concentration were measured at the end of the productivity assay. Best producing mini-pools were selected for scale-up in 24-deepwell plates.

6.2.7 Transfected bulk pool selection

24 hours after transfection, transfected bulk cell suspension was centrifuged for 5 minutes, 200 *g*, at room temperature. Supernatant was removed, and cell pellet was resuspended at $1x10^6$ viable cells/mL in EX-CELL® CD CHO Fusion supplemented with 6 mM L-glutamine and 10 µg/mL puromycin. Details on culture containers and incubation parameters are given in Table 24.

Viable cell density and viability of transfected bulk pool was measured every two or three days. Transfected bulk pool was centrifuged for 5 minutes, 200 *g*, at room temperature when viability lost 20 points (or more) compared to previous time point and/or if viability was below 60%. Supernatant was removed and cell pellet was resuspended in fresh EX-CELL® CD CHO Fusion supplemented with 6 mM L-glutamine and 10 µg/mL puromycin.

Transfected bulk pool was passaged at a seeding density of $1x10^6$ viable cells/mL in EX-CELL® CD CHO Fusion supplemented with 6 mM L-glutamine and 10 µg/mL puromycin in Erlenmeyer flask (or in any appropriate shaken container, namely TPP® tube or 24-deepwell plate, depending on cell suspension volume) as long as viability was below 60%. Transfected bulk pool was then passaged at a seeding density of $4x10^5$ viable cells/mL every two days, or $3x10^5$ viable cells/mL every three days until full pool recovery.

Defined target for single-cell cloning per assays are given in Table 25.

6.2.8 Clone generation: seeding and expansion in 96-well plates

Single-cell cloning was performed with limiting dilution technique. Details on clone seeding density and cloning medium composition per assay are given in Table 25. Individual clones were seeded in 96-well plates. Plates were incubated at 37°C, 80% relative humidity, and 5% CO_2.

Clone recovery was monitored by confluence measurement using Cell Metric plate imager. Plates were imaged on the day of limiting dilution cloning, then ideally but not necessarily on the three first days following limiting dilution cloning, and then at least twice a week. Monoclonality was assessed by verification of plate images taken following limiting dilution cloning. 30 µL of culture medium (details on medium composition per assay given in Table 25) was added once a week. Once clone confluence reached a target confluence (details on confluence target per assay given in Table 25), supernatant was sampled for antibody titration and single colonies were scaled-up in 24-deepwell plates.

Table 24: Transfected bulk pool selection – Details per assays

<table>
<tr><th>Culture containers</th><th>Incubation parameters</th><th>Target for single-cell cloning</th><th>References to results</th></tr>
<tr><td>Seeding in static container (TF 75 cm² or TF 25 cm², depending on electroporation conditions)

First passage (24 hours after transfection) in static container (TF 75 cm² or TF 25 cm², depending on cell suspension volume)

Following passages in shaken container (Erlenmeyer flask, TPP® tube or 24-deepwell plate, depending on cell suspension volume)</td><td rowspan="2">T-flasks: 37°C, 80% relative humidity, 5% CO_2

Erlenmeyer flask: 37°C, 80% relative humidity, 5% CO_2, 150 rpm (10 mm shaking diameter) or 130 rpm (25 mm shaking diameter)

TPP® tube or 24-deepwell plate: 37°C, 80% relative humidity, 5% CO_2, 320 rpm (25 mm shaking diameter)</td><td>7 days post transfection</td><td>3.1 Characterization of CHOZN® CHO K1 cell line development platform performance

3.2 Cell line development process optimization: implementation of direct single-cell cloning</td></tr>
<tr><td>Seeding in static container (TF 25 cm² or TF 75 cm², depending on electroporation conditions)

First passage (24 hours after transfection) and following passages in shaken container (Erlenmeyer flask, TPP® tube or 24-deepwell plate, depending on cell suspension volume)</td><td>When viability recovers to 60%</td><td>3.3 Expression vector optimization: evaluation of promoters for expression of resistance to puromycin</td></tr>
</table>

Table 25: Single-cell cloning – Details per assays

Clone seeding density	**Cloning medium composition**	**Culture medium composition (for medium addition during culturing in 96-well plates)**	**Confluence target for up-scale in 24-deepwell plates**	**Selection for up-scale in 24-deepwell plates**	**References to results**
0.3 viable cell/well	20% (v/v) conditioned medium 80% EX-CELL® CHO Cloning Supplementation with 6 mM L-glutamine	EX-CELL® CD CHO Fusion supplemented with 6 mM L-glutamine	30%	Best producers and worst producers (ranking on antibody titer)	3.1 Characterization of CHOZN® CHO K1 cell line development platform performance 3.2 Cell line development process optimization: implementation of direct single-cell cloning
0.3 viable cell/well	20% (v/v) conditioned medium 80% EX-CELL® CHO Cloning Supplementation with 6 mM L-glutamine and 10 µg/mL puromycin	EX-CELL® CD CHO Fusion supplemented with 6 mM L-glutamine and 10 µg/mL puromycin	80%	Best producers (ranking on antibody titer)	3.2 Cell line development process optimization: implementation of direct single-cell cloning
0.5 viable cell/well	20% (v/v) conditioned medium 80% EX-CELL® CHO Cloning Supplementation with 6 mM L-glutamine, 10 µg/mL puromycin, and 1X ClonaCell™-CHO ACF (ClonaCell, 03820)	EX-CELL® CD CHO Fusion supplemented with 6 mM L-glutamine and 10 µg/mL puromycin	80%	Random selection of 6 best producers, 12 medium producers, 6 worst producers (ranking on antibody titer)	3.3 Expression vector optimization: evaluation of promoters for expression of resistance to puromycin

6.2.9 CHOZN® CHO K1 mini-pool and clone sub-culturing

6.2.9.1 Sub-culturing in 24-deepwell plates

Selected mini-pools or clones for up-scale in 24-deepwell plates were seeded in 3 mL of culture medium (details on medium composition per assay given in Table 26). 24-deepwell plates were incubated at 37°C, 80% relative humidity, 5% CO_2, 320 rpm (25 mm shaking diameter). Cell suspensions were let to recover from shaking culture at least one week.

One week after seeding in 24-deepwell plates, viable cell concentration and cell viability were measured. Cells were passaged either every two days, every three days, or every four days in 3 mL of culture medium (applying viable cell seeding densities mentioned before). Cell suspensions were centrifuged for 5 minutes, 200 *g*, at room temperature if viability was below 80%. Supernatant was removed and cell pellet was resuspended in fresh culture medium.

Productivity assay was launched in 24-deepwell when viability was maintained over 95% and when doubling times were stabilized.

6.2.9.2 Sub-culturing in Erlenmeyer flasks

Selected mini-pools or clones for up-scale in 125-mL Erlenmeyer flasks were seeded in 15 mL of culture medium (details on medium composition per assay given in Table 26). Erlenmeyer flasks were incubated at 37°C, 80% relative humidity, 5% CO_2, 150 rpm (10 mm shaking diameter) or 130 rpm (25 mm shaking diameter).

Two or three days after seeding in Erlenmeyer flasks, viable cell concentration and cell viability were measured. Cells were passaged either every two days, every three days, or every four days in 3 mL of culture medium (applying viable cell seeding densities mentioned before). Cell suspensions were centrifuged for 5 minutes, 200 *g*, at room temperature if viability was below 80%. Supernatant was removed and cell pellet was resuspended in fresh culture medium.

Productivity assay was launched in TPP® tubes when viability was maintained over 95% and when doubling times were stabilized.

Table 26: Mini-pool and clone sub-culturing - Details per assays

Cell type	Culture scale	Incubation parameters	Culture medium	References to results
CHOZN® CHO K1 mini-pools	24-deepwell plate	37°C, 80% relative humidity, 5% CO_2, 320 rpm (25 mm shaking diameter)	EX-CELL® CD CHO Fusion supplemented with 6 mM L-glutamine and 10 µg/mL puromycin	3.1 Characterization of CHOZN® CHO K1 cell line development platform performance
	Erlenmeyer flask	37°C, 80% relative humidity, 5% CO_2, 150 rpm (10 mm shaking diameter) or 130 rpm (25 mm shaking diameter)	EX-CELL® CD CHO Fusion supplemented with 6 mM L-glutamine and 10 µg/mL puromycin	3.1 Characterization of CHOZN® CHO K1 cell line development platform performance
CHOZN® CHO K1 clones	24-deepwell plate	37°C, 80% relative humidity, 5% CO_2, 320 rpm (25 mm shaking diameter)	EX-CELL® CD CHO Fusion supplemented with 6 mM L-glutamine and 10 µg/mL puromycin	3.2 Cell line development process optimization: implementation of direct single-cell cloning 3.3 Expression vector optimization: evaluation of promoters for expression of resistance to puromycin
	Erlenmeyer flask	37°C, 80% relative humidity, 5% CO_2, 150 rpm (10 mm shaking diameter) or 130 rpm (25 mm shaking diameter)	EX-CELL® CD CHO Fusion supplemented with 6 mM L-glutamine and 10 µg/mL puromycin	3.3 Expression vector optimization: evaluation of promoters for expression of resistance to puromycin
	Erlenmeyer flask	37°C, 80% relative humidity, 5% CO_2, 150 rpm (10 mm shaking diameter) or 130 rpm (25 mm shaking diameter)	EX-CELL® CD CHO Fusion supplemented with 6 mM L-glutamine	3.3 Expression vector optimization: evaluation of promoters for expression of resistance to puromycin

6.2.10 Batch and fed-batch productions

6.2.10.1 Batch production in 24-deepwell plates

Batch productivity assay in 24-deepwell plates was launched when doubling times were stabilized, PDL curve was linear and viability was above 95%. Cell suspensions were seeded at $6x10^5$ viable cells/mL in 3 mL of EX-CELL® CD CHO Fusion supplemented with 6 mM L-glutamine and 10 µg/mL puromycin. 24-deepwell plates were incubated at 37°C, 80% relative humidity, 5% CO_2, 320 rpm (25 mm shaking diameter) for four to five days.

Cell concentration, cell viability and antibody concentration were measured at the end of the productivity assay. Supernatants were harvested by centrifugation for 5 minutes, 200 *g*, at room temperature and were stored at < -70°C for later antibody titer measurement.

6.2.10.2 Fed-batch production in TPP® tubes

Fed-batch productivity assay in TPP® tubes was launched when doubling times were stabilized, PDL curve was linear and viability was above 95%. Cell suspensions were centrifuged for 5 minutes, 200 *g*, at room temperature. Supernatants were removed, and cell pellets were resuspended in 26 mL of EX-CELL® Advanced CHO Fed-batch supplemented with 6 mM L-glutamine. TPP® tubes were incubated at 37°C, 80% relative humidity, 5% CO_2, 320 rpm (25 mm shaking diameter). Details on viable cell seeding density, feeding strategy, glucose monitoring strategy and production duration are given in Table 27.

Viable cell concentration, cell viability, glucose and lactate concentrations, pH, and antibody concentration were measured on PD_3, PD_5, PD_7, PD_9, PD_{11} and PD_{14}. Supernatants were harvested by centrifugation for 5 minutes, 200 *g*, at room temperature and were stored at < -70°C for later antibody titer measurement.

Table 27: Fed-batch productivity assay - Details per assays

Seeding VCD [x10^6 vc/mL]	**Feeding strategy**	**Glucose monitoring strategy**	**Production duration**	**References to results**
0.3	5% (v/v) EX-CELL® Advanced CHO Feed 1 on PD_3, PD_5, PD_7, PD_9, and PD_{11}	If $[Glc]_{res}$ < 4 g/L, adjustment of [Glc] to 6 g/L with a 300 g/kg D-glucose solution	14 days	3.1 Characterization of CHOZN® CHO K1 cell line development platform performance
0.3	2% (v/v) HyClone™ Cell Boost™ 7a supplement from PD_3 to PD_{13} 0.2% (v/v) HyClone™ Cell Boost™ 7b Supplement from PD_3 to PD_{13}	If $[Glc]_{res}$ < 4 g/L, bolus* targeting approximately 6 g/L with a 300 g/kg D-glucose solution (bolus calculation based on known Glc consumption of CHOZN® CHO K1 cells in fed-batch culture)	14 days	3.3 Expression vector optimization: evaluation of promoters for expression of resistance to puromycin
1.2	2% (v/v) HyClone™ Cell Boost™ 7a supplement from PD_1 to PD_{11} 0.2% (v/v) HyClone™ Cell Boost™ 7b Supplement from PD_1 to PD_{11}	If $[Glc]_{res}$ < 4 g/L, bolus* targeting approximately 6 g/L with a 300 g/kg D-glucose solution (bolus calculation based on known Glc consumption of CHOZN® CHO K1 cells in fed-batch culture)	12 days	3.3 Expression vector optimization: evaluation of promoters for expression of resistance to puromycin
1.2	2% (v/v) HyClone™ Cell Boost™ 7a supplement from PD_1 to PD_{11} 0.2% (v/v) HyClone™ Cell Boost™ 7b Supplement from PD_1 to PD_{11}	If $[Glc]_{res}$ < 4 g/L, bolus* targeting approximately 6 g/L with a 300 g/kg D-glucose solution (bolus calculation based on fixed values of $[Glc]_{res}$)	12 days	3.3 Expression vector optimization: evaluation of promoters for expression of resistance to puromycin

*Bolus: The glucose bolus corresponds to the addition of a pre-defined volume of a 300 g/kg D-glucose solution. Bolus value was either calculated with data of previous fed-batch cultures of CHOZN® CHO K1 cells or calculated with fixed values of residual glucose concentration ($[Glc]_{res}$) (bolus value calculated for $[Glc]_{res}$ at 0 g/L and applied for $[Glc]_{res}$ ranging between 0 and 2 g/L, and bolus value calculated for $[Glc]_{res}$ at 2 g/L and applied for $[Glc]_{res}$ ranging between 2 and 4 g/L).

6.2.10.3 Fed-batch production in ambr250® bioreactors

Fed-batch production assay was launched in 250-mL bioreactors, using ambr250® modular mammalian bioreactors (Sartorius, 0012A23) and an ambr250® modular benchtop workstation (Sartorius).

100 mL of EX-CELL® Advanced CHO Fed-batch medium supplemented with 6 mM L-glutamine were introduced in the bioreactor and defined amounts of EX-CELL® Advanced™ CHO Feed 1 feeding solution, 300 g/kg D-glucose (Merck Millipore, 137048) solution, FoamAway™ irradiated AoF (ThermoFisher Scientific, A1036902) antifoam solution, and 105.99 g/kg sodium carbonate (Na_2CO_3) (Merck Millipore, 137014) solution were introduced in adapted tanks 24 hours before inoculation. Temperature was set to 37°C, agitation was set to 287 rpm, airflow in head space was set to 2.4 mL/min, and air in sparging lines was set to 0.1 mL/min overnight to equilibrate pH probe. pH was set at 7.00 with a deadband of 0.50 unit of pH before inoculation.

Cell suspensions were seeded at $0.3x10^6$ viable cells/mL. Volume was adjusted to 220 mL with EX-CELL® Advanced CHO Fed-batch medium supplemented with 6 mM L-glutamine. 0.16 g/kg of FoamAway™ irradiated AoF antifoam solution was added to the inoculum.

Temperature set point was 37°C. pH set point was 7.00 with a dead-band of 0.50 unit of pH. pH was regulated with carbon dioxide (CO_2) in head-space (maximum flow rate of 2 mL/minute, PID 10/0/0) in case of basification and with 105.99 g/kg sodium carbonate (Na_2CO_3) (maximum flow rate of 2.5 mL/hour, PID 10/0/0) in case of acidification. Partial pressure of dioxygen (pO_2) set point was 30% and was regulated with O_2 in open pipe (maximum flow rate of 28.6 mL/minute, PID 0.5/500/0). Air was added in open pipe (0.1 mL/minute continuously, PID 0.1/500/0) and in surface (2.4 mL/minute continuously). Stirring speed set point was 200 rpm. Antifoam was manually added when needed (maximum addition of 2.5 mL over process duration).

Cells were fed with 5.5% (v/daily volume) of EX-CELL® Advanced CHO Feed 1 on PD_3, PD_5, and PD_7, and 5.6% (v/daily volume) of EX-CELL® Advanced CHO Feed 1 on PD_9, PD_{11}, and PD_{13}. 300 g/kg D-glucose was added up to 6 g/L, taking into consideration glucose in feed.

Viable cell concentration, cell viability, glucose and lactate concentrations, pH, pO_2, pCO_2, osmolality, and antibody concentration were measured daily. pO_2 and pCO_2 were measured using the ABL90 FLEX gas analyzer. Osmolality was measured using an OsmoPRO® Multi-Sample Micro-Osmometer (Advanced Instruments) osmometer. Supernatants were harvested by centrifugation for 5 minutes, 200 *g*, at room temperature and were stored at < -70°C for later antibody titer measurement.

Production was stopped at PD_{16} or whenever viability fell below 60%. Program was set on ambr250® Definition software (Sartorius). Program was launched and process was monitored on ambr250® Run Time software (Sartorius). Recording of process parameters was viewed on ambr250® Results Viewer software (Sartorius).**Molecular biology methods**

6.3.1 Expression vector production

pEM2V-a (Figure 9) or pEM2V-c (Figure 10) were used as vector backbone for plasmid construction which has been outsourced (Solvias). All plasmids were supplied after DNA sequence checking.

6.3.2 DNA extraction

Genomic DNA was extracted using the DNeasy® Blood & Tissue Kit (Qiagen, 69506), following the manufacturer's protocol. Protocol details (especially for centrifugation speeds, as they are not clearly stated in manufacturer's protocol) are given hereinafter.

$5x10^6$ cells were pelleted by centrifuging appropriate cell suspension volume for 5 minutes at 400 *g*. Cell culture supernatant was removed. Cell pellets were carefully dried and stored at -70°C.

Cell pellets were allowed to thaw until the pellets could be dislodged by gently flicking the tubes. Cells were disrupted by adding 200 µL of buffer AL. Complete loosening of the cell pellet was ensured by 15 seconds vortexing. Lysates were incubated at 56°C for 10 minutes. 200 µL of ethanol 96-100% (Fisher Scientific, 15552393) were added to the sample. Sample and ethanol were mix thoroughly by vortexing 15 seconds at full speed. Mixture was pipetted into a DNeasy Mini spin column placed in a 2 mL collection tube. DNeasy Mini spin column was centrifuged for 1 minute at ≥ 6000 *g* (*i.e.*, 7,500 *g*). Flow-through and collection tube were discarded. DNeasy Mini spin column was placed in a new 2 mL collection tube. 500 µL of buffer AW1 were added to the DNeasy Mini spin column. DNeasy Mini spin column was centrifuged for 1 minute at ≥ 6000 *g* (*i.e.*, 7,500 *g*). Flow-through and collection tube were discarded. DNeasy Mini spin column was placed in a new 2 mL collection tube. 500 µL of buffer AW2 were added to the DNeasy Mini spin column. DNeasy Mini spin column was centrifuged for 5 minutes at 16,000 *g*. Flow-through and collection tube were discarded. If carryover of ethanol occurred, collection tube was emptied and reused in another centrifugation for 1 minute at 16,000 *g*. DNeasy Mini spin column was transferred to a new 1.5 mL or 2 mL micro-tube. 100 µL of buffer AE were added to the DNeasy Mini spin column and incubated for 1 minute at room temperature (15-25°C). DNeasy Mini spin column was centrifuged for 1 minute at ≥ 6000 *g* (*i.e.*, 7,500 *g*). For increased DNA yield, last step was repeated: 100 µL of buffer AE were added to the DNeasy Mini spin column and incubated for 1 minute at room temperature (15-25°C). DNeasy Mini spin column was centrifuged for 1 minute at ≥ 6000 *g* (*i.e.*, 7,500 *g*). Aliquot was sampled for concentration, and purity measurement. DNA was either immediately used or stored at -20°C.

6.3.3 RNA extraction

RNA extraction principle has been described previously, in 3.4.3.1 RNA extraction. Total RNA was extracted using the RNeasy® Plus Mini Kit (Qiagen, 74136), following the manufacturer's protocol. Protocol details (especially for centrifugation speeds, as they are not clearly stated in manufacturer's protocol) are given hereinafter.

$10x10^6$ cells were pelleted by centrifuging appropriate cell suspension volume for 5 minutes at 400 *g*. Cell culture supernatant was removed. Cell pellets were carefully dried on ice and immediately stored at -70°C.

Cell pellets were allowed to thaw until the pellets could be dislodged by gently flicking the tubes. Cells were disrupted by adding 600 µL of buffer RLT Plus supplemented with 1% (v/v) β-mercapto-ethanol (SAFC, M6250-100ML). Complete loosening of the cell pellet was ensured by 30 seconds vortexing. Up to 700 µL of lysate was pipetted into a QIAshredder spin column (Qiagen, 79656) placed in a 2 mL collection tube. QIAshredder spin column was centrifuged for 2 minutes at maximum speed (*i.e.*, 16,000 *g*). The homogenized lysate was transferred to a gDNA Eliminator spin column placed in a 2 mL collection tube. gDNA Eliminator spin column was centrifuged for 30 seconds at ≥ 8,000 *g* (*i.e.*, 9,200 *g*). Column was discarded and the flow-through saved. 600 µL of ethanol 70% (Fisher Scientific, 15542393) were added to the flow-through. Up to 700 µL of the sample was transferred into a RNeasy spin column placed in a 2 mL collection tube and then was centrifuged for 15 seconds at ≥ 8,000 g (*i.e.*, 9,200 *g*). Flow-through was discarded. 700 µl of buffer RW1 were added to the RNeasy spin column. RNeasy spin column was centrifuged for 15 seconds at ≥ 8,000 g (*i.e.*, 9,200 *g*). Flow-through was discarded. 500 µl of buffer RPE supplemented with 4 volumes of ethanol 100% (Fisher Scientific, 15552393) were added to the RNeasy spin column. RNeasy spin column was centrifuged for 15 seconds at ≥ 8,000 g (*i.e.*, 9,200 *g*). Flow-through was discarded. 500 µl of buffer RPE supplemented with 4 volumes of ethanol 100% (Fisher Scientific, 15552393) were added to the RNeasy spin column. RNeasy spin column was centrifuged for 2 minutes at ≥ 8,000 g (*i.e.*, 9,200 *g*). Flow-through was discarded and RNeasy spin column placed in a new 2 mL collection tube. RNeasy spin column was centrifuged at full speed (*i.e.*, 16,000 *g*) for 1 minute. Flow-through was discarded and RNeasy spin column placed in a new 1.5 mL collection tube. 50 µL of RNase-free water was added to the spin column membrane. RNeasy spin column was centrifuged for 1 minute at ≥ 8,000 g (*i.e.*, 9,200 *g*) to elute total RNA. Previous step was repeated using eluted RNA. Aliquot was sampled for concentration, purity, and integrity measurement. RNA was either immediately used or stored at -70°C.

6.3.4 DNA and RNA quality control

DNA and RNA concentration and purity (A_{260}/A_{230} and A_{260}/A_{280} ratios) were analyzed by spectrophotometry on a NanoDrop™ One (ThermoFischer Scientific).

RNA integrity was analyzed by microfluidic capillary electrophoresis on a LabChip GXII Touch HT (Perkin Elmer), using a DNA 5K/RNA/CZE HT chip (Perkin Elmer, 760435) and associated RNA reagent kit (Perkin Elmer, P/N CLS960010), following manufacturer's instructions. RNA Quality Score (RQS) was calculated using LabChip GX Reviewer software.

6.3.5 Reverse transcription

Reverse transcription (RT) principle has been described previously, in 3.4.3.3 cDNA synthesis by reverse transcription. RNA was reverse transcribed into cDNA using QuantiTect® Reverse Transcription Kit (Qiagen, 205313), following the manufacturer's protocol. Protocol details are given hereinafter.

Template RNA was thawed on ice. gDNA Wipeout Buffer, Quantiscript® Reverse Transcriptase, Quantiscript® RT buffer, RT Primer Mix and RNase-free water were thawed at room temperature (15-25°C). Each solution was mixed by flicking the tubes, centrifuged briefly to collect residual liquid from the sides of the tubes, and then kept on ice. Genomic DNA elimination reaction was prepared in a 0.2mL reaction tube (Applied Biosystems, N8010540) on ice with: 2 µL of gDNA Wipeout Buffer, 1 µg template RNA, and RNase free water (qs 14 µL) (volumes per reaction; master mix was prepared if setting up more than one reaction with a volume 10% greater than required for the total number of reactions). Reaction tube was incubated for 2 minutes at 42°C using a 2720 Thermal Cycler (Applied Biosystems) and was then immediately placed on ice. Reverse transcription master mix was prepared on ice with 1 µL Quantiscript® Reverse Transcriptase, 4 µL 5X Quantiscript® RT buffer and 1 µL RT Primer Mix (volumes per reaction; master mix was prepared if setting up more than one reaction with a volume 10% greater than required for the total number of reactions). Reverse-transcription master mix was added to each tube containing template RNA. Reaction tube was incubated for 15 minutes at 42°C, then for 3 minutes at 95°C using a 2720 Thermal Cycler. Reaction tube was placed on ice at the end of the reaction. cDNA was either immediately used or stored at -20°C.

6.3.6 Quantitative PCR

Quantitative PCR principle has been described previously, in 3.4.1.2 Quantitative PCR: insights on a widespread technique. qPCR was performed using QuantiNova™ SYBR® Green PCR Kit (Qiagen, 208056), following the manufacturer's protocol, on a QuantStudio™ 7 Pro (Applied Biosystems) real-time cycler. Protocol details are given hereinafter.

2X QuantiNova™ SYBR® Green PCR master mix, QuantiNova™ ROX reference dye, QuantiNova™ Yellow template dilution buffer, template cDNA, forward and reverse primers (sequences in Table 28), and RNase-free water were allowed to thaw at room temperature (15-25°C). PCR reaction mix was prepared with: 1X QuantiNova™ SYBR® Green PCR master mix, 1X QuantiNova™ ROX reference dye, 0.7 µM forward primer, 0.7 µM reverse primer, and RNase free water (qs 20 µL) (volumes per reaction; master mix was prepared if setting up more than one reaction with a volume 10% greater than required for the total number of reactions). PCR reaction mix was dispensed in 96-well PCR plates (Applied Biosystems, 4346906). ≤ 100 ng template cDNA was added to the 96-well PCR plates. PCR cycle was run in maximal/fast mode as follows: hot-start activation for 2 minutes at 95°C, followed by 40 cycles of denaturation for 5 seconds at 95°C and combined annealing-extension for 10 seconds at 60°C. Melt curve analysis was performed from 60°C to 95°C. Reaction mixtures containing no cDNA

template and reaction mixtures containing products from reverse transcription reactions performed in the absence of reverse transcriptase were used as negative controls. All samples were run in technical triplicate. Mean quantification cycle (C_q) values were used for data analysis.

Table 28: Primer sequences

Following table only lists validated primer sequences for pac, rituHC, and rituLC (plasmid genes) and atp5mg and rps16 (reference genes) detection. Primer sequences evaluated for specificity and efficacy for plasmid genes and reference genes detection are listed in the Table 13 (plasmid genes) and Table 14 (reference genes).

Gene	Forward primer sequence (5'-3')	Reverse primer sequence (5'-3')
pac	agctgcaagaactcttcctcac	caggaggccttccatctgtt
rituHC	aaggcctcgggatacacttt	ggcctttcccttgaacttct
rituLC	atccgtgactgagcaggact	acttcgcaggcgtacacttt
atp5mg	atggccaagttcgtccgtaa	gggttgggggaaacagttca
rps16	tgaagggtggtggtcatgtg	tcagctacaagcagggttcg

6.3.7 Amplicon size analysis by electrophoresis on agarose gel

DNA samples were diluted in 1X TrackIt™ Cyan/Yellow Loading Buffer (ThermoFisher Scientific,10482035). Amplicon size (in base pair) was determined using Ultra-Low Range DNA Ladder (ThermoFisher Scientific, 10597012) molecular weight marker. DNA samples were migrated on 3% TAE Mini ReadyAgarose™ precast gel (Biorad, 1613017), in 1X Tris acetate EDTA (TAE) buffer (VWR chemicals, 444125D), at 100 V during approximately one hour, using a Wide Mini-Sub Cell GT Cell (Biorad) migration chamber and a PowerPac™ Basic Power Supply (Biorad) generator. Gels were stained using SYBR™ Safe DNA DNA Gel Stain (ThermoFisher Scientific, S33111) staining solution, for 15 to 30 minutes, in the dark and under agitation. Gels were imaged by transillumination, using the Chemidoc Touch™ imaging system (Biorad).

6.4 Calculation formulas

6.4.1 Doubling Time (DT)

Doubling time (DT), expressed in hours ([h]), was calculated as follows:

$$DT\ [hours] = \frac{\ln(2) \times (T_t - T_{t-1}) \times 24}{\ln\left(\frac{VCD_t}{VCD_{t-1}}\right)}$$

With: T_t: considered time point [day]; T_{t-1}: previous time point [day]; VCD_t: viable cell density at considered time point [viable cells/mL]; VCD_{t-1}: viable cell density at previous time point [viable cells/mL].

6.4.2 Population Doubling Level (PDL)

Population Doubling Level (PDL), expressed without unit ([-]), was calculated as follows:

$$PDL\ [-] = \log 2 \times \left(\frac{Qtc_t}{Qvc_{t-1}}\right) + PDL_{t-1}$$

With: Qtc_t: total cell quantity at considered time point [cells]; Qvc_{t-1}: initial viable cell quantity at previous time point [cells]; PDL_{t-1}: population doubling level at previous time point [-].

6.4.3 Integral of Viable Cell density (IVC)

Integral of Viable Cell density (IVC), expressed in viable cells times day per milliliter ([viable cell x day/mL]), was calculated as follows:

$$IVC_{Of\ the\ day}\ [vc \times day/mL] = \frac{\left(VCD_{t-1_{Post\ feed}} - VCD_t\right) \times (T_t - T_{t-1})}{2}$$

$$IVC_{Cumulated}\ [vc \times day/mL] = \frac{\left(VCD_{t-1_{Post\ feed}} - VCD_t\right) \times (T_t - T_{t-1})}{2} + IVC_{Cumulated_{t-1}}$$

With: $VCD_{t\text{-}1\ post\ feed}$: viable cell density post feed at previous time point [viable cells/mL]; VCD_t: viable cell density at considered time point [viable cells/mL]; T_t: considered time point [day]; T_{t-1}: previous time point [day]; IVC_{t-1}: cumulated integral of viable cell density at previous time point ([viable cell x day/mL].

6.4.4 Specific productivity (Q_P)

Specific productivity (Q_P), expressed in picograms per cell per day ([pg/cell/day]), was calculated as follows:

$$Q_{P_{Of\ the\ day}}\ [pg/cell/day] = \frac{mAb\ titer_t - mAb\ titer_{t-1}}{IVC_{Of\ the\ day}}$$

$$Q_{P_{Process}}\ [pg/cell/day] = \frac{mAb\ titer_{Final}}{IVC_{Cumulated}}$$

With: $mAb\ titer_t$: antibody titer at considered time point [mg/L]; $mAb\ titer_{t-1}$: antibody titer at previous time point [mg/L]; $mAb\ titer_{Final}$: final antibody titer [mg/L]; $IVC_{Of\ the\ day}$: integral of viable cell density between considered time point and previous time point [viable cell x day/mL]; $IVC_{Cumulated}$: cumulated integral of viable cell density over process duration [viable cell x day/mL].

6.4.5 Cloning efficiency

Cloning efficiency was calculated as follows:

$$Cloning\ efficiency\ [\%] = \frac{Number\ of\ reemerging\ clones}{Theoretical\ number\ of\ clones}$$

Theoretical number of clones is determined using the Poisson distribution law ($\mathcal{P}(\lambda)$) as follows:

$$Theoretical\ number\ of\ clones = Total\ number\ of\ wells \times P(1)$$
$$= Total\ number\ of\ wells \times VCD_{Cloning} \times exp^{-VCD_{Cloning}}$$

With: $VCD_{Cloning}$: viable cell density for clone seeding [viable cells/well].

6.4.6 Percentage of variation

Percentage of variation was calculated as follows:

$$\Delta X\ [\%] = \frac{X_{Final\ value} - X_{Initial\ value}}{X_{Initial\ value}}$$

6.5 Statistical analysis

Statistical analysis was conducted following steps described in Figure 60. Statistical analysis was performed using JMP® software (versions 15 and 16).

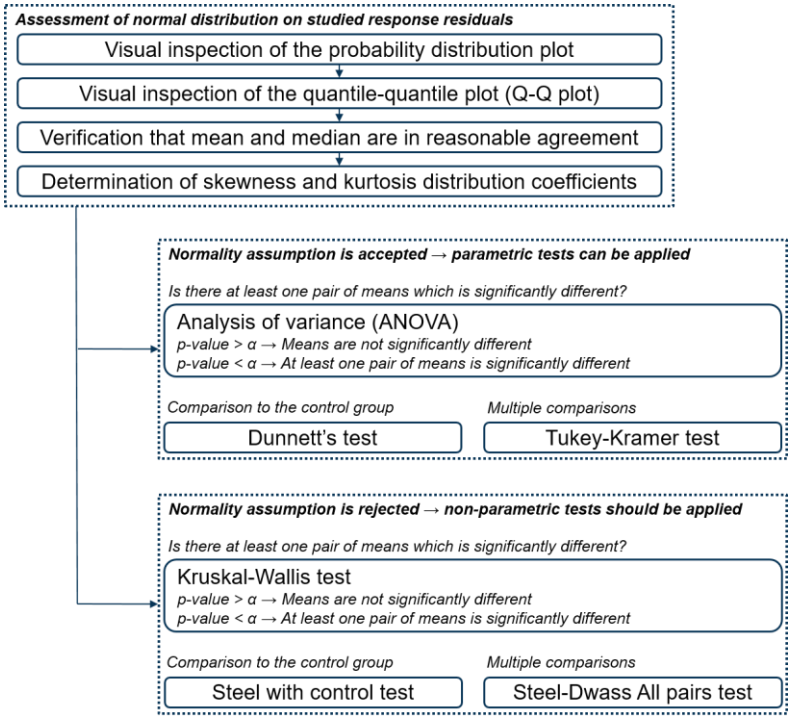


Figure 60: Workflow for statistical analysis

Assessment of normality of data distribution, which is ideally performed on studied response residuals, is a prerequisite to determine which statistical test will be suited for data analysis. If normality assumption is accepted, parametric tests can be applied. ANOVA indicates if at least one pair of means is significantly different. In this case, analysis is further investigated, either with a Dunnett's test (for

comparisons to the control group) or with a Tukey-Kramer test (for multiple comparisons). Otherwise, if normality assumption is rejected, non-parametric tests should be applied.

6.5.1 Assessment of normal distribution

Assessment of normality of data distribution is a prerequisite to determine which statistical test will be suited for data analysis. Assessment of normal distribution is ideally performed on studied response residuals (remaining variations once potential group-related effects have been removed). If data are normally distributed and if the variability within groups is similar between groups, then parametric tests can be applied. Otherwise, non-parametric tests should be used for analysis.

A normal distribution (noted $\mathcal{N}(\mu, \sigma)$, with μ the mean and σ the standard deviation) is characterized by a symmetrical bell probability density function curve, centered on the mean (μ). Normal distribution of data can be assessed by:

(i) Visual inspection of the probability distribution plot. The probability distribution plot is a graphical tool to plot experimental data according to their frequency. If experimental data points distribution follows a normal distribution, frequency histogram fits to a normal distribution curve. In practice, this is checked by visual inspection of the probability distribution plot. Probability distribution plot is performed with a distribution analysis and normal distribution curve can be displayed with "Continuous Fit → Fit Normal" option on JMP® software.

(ii) Visual inspection of the quantile-quantile plot (Q-Q plot). The quantile-quantile plot is a graphical tool to evaluate the adequacy of the fit of a given distribution to a theoretical model. If experimental data points distribution follows well the chosen theoretical distribution, experimental quantiles should be equal to the quantiles associated to the theoretical model. In practice, this is often checked by visual inspection of the quantile-quantile plot: plotted points should fit in "the thickness of the pencil". Quantile-quantile plot can be performed in a distribution analysis with "Normal Quantile Plot" option on JMP® software.

(iii) Verification that calculated mean and median are in reasonable agreement (mean of normally distributed data coincide with the median).

(iv) Determination of skewness and kurtosis distribution coefficients. A normal distribution is characterized by a zero-skewness coefficient and a zero-kurtosis coefficient. A negative skewness coefficient indicates right skewness, while a positive skewness coefficient indicates left skewness. A negative kurtosis coefficient indicates that the distribution is platykurtic, *i.e.*, flatter than a normal density, while a positive kurtosis coefficient indicates that the distribution is leptokurtic, *i.e.*, less flattened than a normal density. In practice, they should be comprised between ± 2 to consider that experimental data are normally

distributed. Both skewness and kurtosis coefficients are available in summary statistics of distribution analysis on JMP® software.

6.5.2 One-way analysis of variance (ANOVA)

One-way analysis of variance (ANOVA) is a statistical method for comparing means of three or more groups, if normality assumption is accepted. ANOVA results indicate whether at least one pair of means is significantly different. Nonetheless, these results do not indicate which pair of means are different. ANOVA tests the null hypothesis (noted H_0) that three or more population means are equal ($H_0: \mu_1 = \mu_2 = \cdots = \mu_n$, with μ_i being the mean of the i-th level of the factor) *versus* the alternative hypothesis (noted H_α) that at least one mean is different. The test calculates a probability value (noted p-value) for all tested groups and compares this p-value to the error risk value (α). If the p-value is lower than the α-threshold, the H_0 hypothesis is rejected, *i.e.*, at least one mean is different.

One-way ANOVA can be performed by "Fit Y by X → Means/ANOVA" on JMP® software.

In the case where the normality assumption is not acceptable, the Kruskal-Wallis test (non-parametric test) is used as an alternative to ANOVA.

6.5.3 Comparison of means

ANOVA results indicate if at least one pair of means is significantly different, but do not highlight which pair of means are different. An additional test is therefore required to know which pair of means are different. If comparisons a set of means against the mean of the control group are the comparisons of interest, a Dunnett's test should be applied. Otherwise, if multiple pairwise comparisons are the comparisons of interest, a Tukey-Kramer test should be applied.

6.5.3.1 Comparison to control group: Dunnett's test

The Dunnett's test can be applied for comparison of a set of means against the mean of a control group. Dunnett's test tests the null hypothesis that the control group and the tested group means are equal ($H_0: \mu_{Control\ group} = \mu_{Evaluated\ group}$, with μ being the mean) *versus* the alternative hypothesis that the control group and the tested group means are different. If the p-value is lower than the α-threshold, the H_0 hypothesis is rejected, *i.e.*, the control group and the tested group means are different. It should nevertheless be mentioned that Dunnett's test does not compare the groups other than the control group among themselves.

Dunnett's test can be performed by "Fit Y by X → Compare Means → With control, Dunnett's" on JMP® software. Results are also displayed with a comparison circles plot, which is a visual representation of group mean comparisons (Figure 61). Circles for means that are significantly different either do not intersect, or intersect slightly, so that the outside angle of intersection is less than 90

degrees. Circles for means that are not significantly different intersect by an angle of more than 90 degrees and are nested.

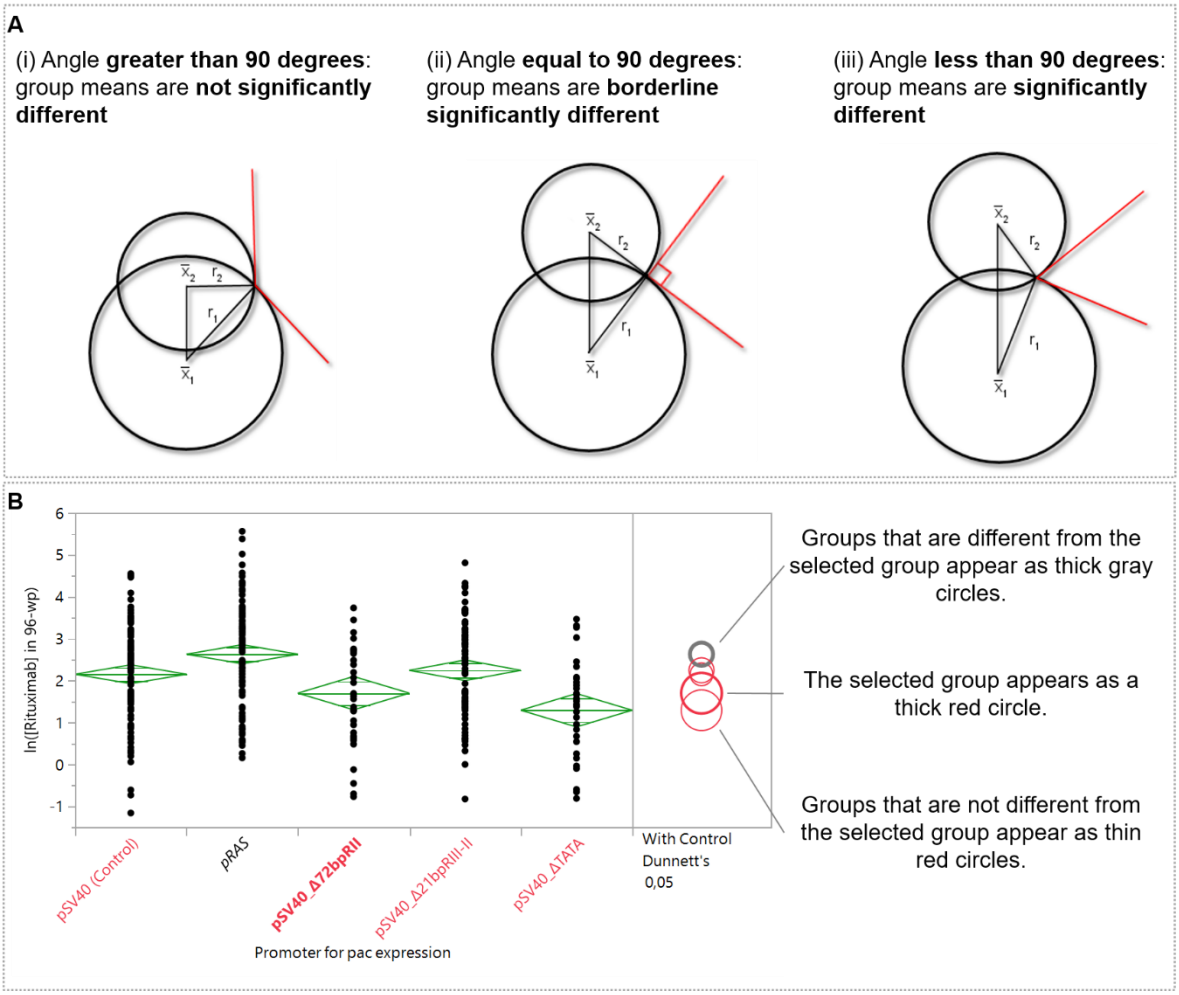


Figure 61: Visual comparison of group means: significance of comparison circles

(A) *Significance of comparison circles. Circles for means that are significantly different either do not intersect, or intersect slightly, so that the outside angle of intersection is less than 90 degrees. Circles for means that are not significantly different intersect by an angle of more than 90 degrees and are nested.* ***(B)*** *Significance of circles' color on JMP® software.*

In the case where the normality assumption is not acceptable, the Steel with control test (non-parametric test) is used as an alternative to Dunnett's test.

6.5.3.2 Multiple pairwise comparisons: Tukey-Kramer test

The Tukey-Kramer test can be applied for multiple pairwise comparisons, *i.e.*, to determine whether there is a difference between the mean of all possible pairs, by comparing all group means to all other group means.

It should be noticed that comparisons to the control group are included while performing a Tukey-Kramer test. Nonetheless, the Dunnett's test is more relevant and powerful than Tukey-Kramer test for

comparison of a set of means against the mean of a control group, because of reduced number of tests compared to all pairwise comparisons (Lee & Lee, 2018). In the case where comparisons to control and pairwise comparisons were both of interest, a two-test based analysis strategy was applied. First, comparisons to the control group were performed with a Dunnett's test. Second, data from control group were excluded and multiple pairwise comparisons were performed with a Tukey-Kramer test.

Tukey-Kramer test can be performed by "Fit Y by X → Compare Means → All pairs, Tukey HSD" on JMP® software. Results are also displayed with a comparison circles plot (Figure 61).

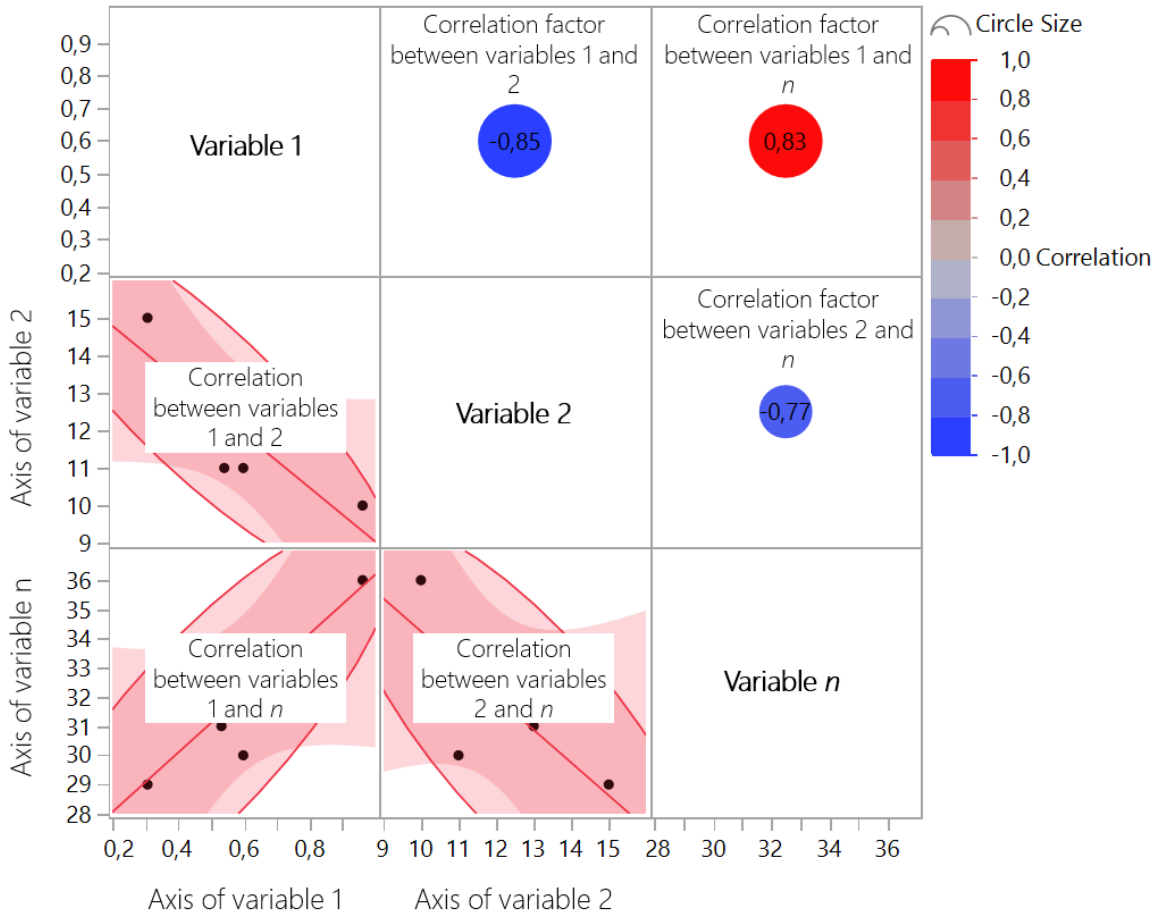
In the case where the normality assumption is not acceptable, the Steel-Dwass All pairs test (non-parametric test) is used as an alternative to Tukey-Kramer test.

6.5.4 Correlation matrix

Multivariate analysis was used to determine correlation between variables. Correlation is a statistical measure of a linear relationship between variables. It expresses the extent to which two variables are linearly related.

Correlations are described by two numbers: the correlation coefficient (noted *r*) and the probability value (noted p-value). The correlation coefficient measures the strength of the relationship between two variables and varies from -1 to +1. The closer *r* is to 1 (in absolute value), the stronger the linear relationship is. Positive *r* values indicate a positive correlation (both variables tend to increase together) while negative *r* values indicate a negative correlation (the values of one variable tend to decrease while the values of the other variable tend to increase). The p-value measures the probability of correlation and is the result of a test of the null hypothesis that the correlation between variables is zero (no linear relationship) *versus* the alternative hypothesis that the variables correlate (linear relationship). If the p-value is lower than the α-threshold, the H_0 hypothesis is rejected, *i.e.*, there is a linear relationship between variables.

Multivariate analysis can be performed by "Multivariate methods → Multivariate" on JMP® software. Results are displayed in a correlation matrix (probability values of correlations) and a scatterplot matrix (correlation coefficients).



Correlation Probability

	Variable 1	Variable 2	Variable *n*	
Variable 1		<,0001	0,0674	0,0844
Variable 2	0,0674		<,0001	0,1251
Variable *n*	0,0844	0,1251		<,0001

Figure 62: Construction and interpretation of a correlation matrix

Correlation matrix between n variables. Correlation is a statistical measure of a linear relationship between variables and expresses the extent to which two variables are linearly related. Correlations are described by two numbers: the correlation coefficient (noted r) and the probability value (noted p-value).

References

Adamson, B., Norman, T. M., Jost, M., Cho, M. Y., Nuñez, J. K., Chen, Y., Villalta, J. E., Gilbert, L. A., Horlbeck, M. A., Hein, M. Y., Pak, R. A., Gray, A. N., Gross, C. A., Dixit, A., Parnas, O., Regev, A., & Weissman, J. S. (2016). A Multiplexed Single-Cell CRISPR Screening Platform Enables Systematic Dissection of the Unfolded Protein Response. *Cell*, *167*(7), 1867-1882.e21. https://doi.org/10.1016/j.cell.2016.11.048

Aebi, M. (2013). Biochimica et Biophysica Acta N-linked protein glycosylation in the ER ☆. *BBA - Molecular Cell Research*, *1833*(11), 2430–2437. https://doi.org/10.1016/j.bbamcr.2013.04.001

Aebi, M., Bernasconi, R., Clerc, S., & Molinari, M. (2009). N-glycan structures : recognition and processing in the ER. *Trends in Biochemical Sciences*, *35*(2), 74–82. https://doi.org/10.1016/j.tibs.2009.10.001

Almassy, R. J., Janson, C. A., Hamlin, R., Xuong, N. H., & Eisenberg, D. (1986). Novel subunit-subunit interactions in the structure of glutamine synthetase. *Nature*, *323*(6086), 304–309. https://doi.org/10.1038/323304a0

Andersen, C. L., Jensen, J. L., & Ørntoft, T. F. (2004). *Normalization of Real-Time Quantitative Reverse Transcription-PCR Data : A Model-Based Variance Estimation Approach to Identify Genes Suited for Normalization , Applied to Bladder and Colon Cancer Data Sets*. 5245–5250.

Andersen, D. C., & Krummen, L. (2002). Recombinant protein expression for therapeutic applications. *Current Opinion in Biotechnology*, *13*, 117–123. https://doi.org/10.1002/0471650129.dob0608

Artelt, P., Grannemann, R., Stocking, C., Friel, J., Bartsch, J., & Hauser, H. (1991). The prokaryotic neomycin-resistance-encoding gene acts as a transcriptional silencer in eukaryotic cells. *Gene*, *99*(2), 249–254. https://doi.org/10.1016/0378-1119(91)90134-W

Aviner, R. (2020). The science of puromycin: From studies of ribosome function to applications in biotechnology. *Computational and Structural Biotechnology Journal*, *18*, 1074–1083. https://doi.org/10.1016/j.csbj.2020.04.014

Baah, S., Laws, M., & Rahman, K. M. (2021). Antibody–Drug Conjugates : A Tutorial Review. *Molecules*, *26*(2943).

Bahr, S. M., Borgschulte, T., Kayser, K. J., & Lin, N. (2009). Using microarray technology to select housekeeping genes in Chinese hamster ovary cells. *Biotechnology and Bioengineering*, *104*(5), 1041–1046. https://doi.org/10.1002/bit.22452

Bailey, L. A., Hatton, D., Field, R., & Dickson, A. J. (2012). Determination of Chinese hamster ovary

cell line stability and recombinant antibody expression during long-term culture. *Biotechnology and Bioengineering*, *109*(8), 2093–2103. https://doi.org/10.1002/bit.24485

Baird, S. D., Turcotte, M., Korneluk, R. G., & Holcik, M. (2006). Searching for IRES. In *RNA* (Vol. 12, Issue 10, pp. 1755–1785). https://doi.org/10.1261/rna.157806

Bandaranayake, A. D., & Almo, S. C. (2013). Recent advances in mammalian protein production. *FEBS Letters*, *588*(2), 253–260. https://doi.org/10.1016/j.febslet.2013.11.035

Bandaranayake, A. D., & Almo, S. C. (2014). Recent advances in mammalian protein production. *FEBS Letters*, *588*(2), 253–260. https://doi.org/10.1016/j.febslet.2013.11.035

Barnes, L. M., Bentley, C. M., & Dickson, A. J. (2003). Stability of protein production from recombinant mammalian cells. In *Biotechnology and Bioengineering* (Vol. 81, Issue 6, pp. 631–639). https://doi.org/10.1002/bit.10517

Bause, E. (1983). *Structural requirements of N-glycosylation of proteins*. *209*, 331–336.

Bedoya-López, A., Estrada, K., Sanchez-Flores, A., Ramírez, O. T., Altamirano, C., Segovia, L., Miranda-Ríos, J., Trujillo-Roldán, M. A., & Valdez-Cruz, N. A. (2016). Effect of temperature downshift on the transcriptomic responses of Chinese hamster ovary cells using recombinant human tissue plasminogen activator production culture. *PLoS ONE*, *11*(3), 1–26. https://doi.org/10.1371/journal.pone.0151529

Benoist, C., & Chambon, P. (1980a). Deletions covering the putative promoter region of early mRNAs of simian virus 40 do not abolish T-antigen expression. *Proceedings of the National Academy of Sciences of the United States of America*, *77*(7 II), 3865–3869. https://doi.org/10.1073/pnas.77.7.3865

Benoist, C., & Chambon, P. (1980b). Deletions covering the putative promoter region of early mRNAs of simian virus 40 do not abolish T-antigen expression. *Proceedings of the National Academy of Sciences of the United States of America*, *77*(7 II), 3865–3869. https://doi.org/10.1073/pnas.77.7.3865

Benoist, C., & Chambon, P. (1981). In vivo sequence requirements of the SV40 early promoter region. *Nature*, *290*(5804), 304–310. https://doi.org/10.1038/290304a0

Bernard, A. (2005). Production of proteins by transient expression. In S. Ozturk & W. S. Hu (Eds.), *Cell Culture Technology for Pharmaceutical and Cell-Based Therapies* (CRC press).

Berting, A., Farcet, M. R., & Kreil, T. R. (2010). Virus susceptibility of Chinese hamster ovary (CHO) cells and detection of viral contaminations by adventitious agent testing. *Biotechnology and Bioengineering*, *106*(4), 598–607. https://doi.org/10.1002/bit.22723

Bethencourt, V. (2009). Virus stalls Genzyme plant. *Nature Biotechnology*, *27*(8), 681–681. https://doi.org/10.1038/nbt0809-681a

Betts, Z., & Dickson, A. J. (2015). Assessment of UCOE on Recombinant EPO Production and Expression Stability in Amplified Chinese Hamster Ovary Cells. *Molecular Biotechnology*, *57*(9), 846–858. https://doi.org/10.1007/s12033-015-9877-y

Birch, J. R., & Racher, A. J. (2006). Antibody production. *Advanced Drug Delivery Reviews*, *58*(5–6), 671–685. https://doi.org/10.1016/j.addr.2005.12.006

Birrer, M. J., Moore, K. N., Betella, I., & Bates, R. C. (2019). Antibody-Drug Conjugate-Based Therapeutics: State of the Science. *Journal of the National Cancer Institute*, *111*(6), 538–549. https://doi.org/10.1093/jnci/djz035

Blanco, N., Williams, A. J., Tang, D., Zhan, D., Misaghi, S., Kelley, R. F., & Simmons, L. C. (2020). *Tailoring translational strength using Kozak sequence variants improves bispecific antibody assembly and reduces product - related impurities in CHO cells*. *March*, 1946–1960. https://doi.org/10.1002/bit.27347

Blokland, H. J. M. Van, Kwaks, T. H. J., Sewalt, R. G. A. B., Verhees, J. A., Klaren, V. N. A., Siersma, T. K., Korse, J. W. M., Teunissen, N. C., Botschuijver, S., Mer, C. Van, Man, S. Y., & Otte, A. P. (2007). *A novel , high stringency selection system allows screening of few clones for high protein expression*. *128*, 237–245. https://doi.org/10.1016/j.jbiotec.2006.09.023

Bloom, J., Madanat, M., Marriott, D., Wong, T., Chan, S., & Bloom, J. W. (1997). Intrachain disulfide bond in the core hinge region of human IgG4. *Prorein Science*, *6*, 407415.

Bochkov, Y. A., & Palmenberg, A. C. (2006). Translational efficiency of EMCV IRES in bicistronic vectors is dependent upon IRES sequence and gene location. *BioTechniques*, *41*(3), 283–292. https://doi.org/10.2144/000112243

Bolisetty, P., Tremml, G., Xu, S., & Khetan, A. (2020). Enabling speed to clinic for monoclonal antibody programs using a pool of clones for IND-enabling toxicity studies. *MAbs*, *12*(1). https://doi.org/10.1080/19420862.2020.1763727

Bosques, C. J., Collins, B. E., Meador Ill, J. W., Sarvaiya, H., Murphy, J. L., DelloRusso, G., Bulik, D. A., Hsu, I.-H., Washburn, N., Sipsey, S. F., Myette, J. R., Raman, R., Shriver, Z., Sasisekharan, R., & Venkataraman, G. (2010). Chinese hamster ovary cels can produce galactose-α-1,3-galactose antigens on proteins. *Nature Biotechnology*, *28*(11), 1153–1156. https://doi.org/10.1038/nbt1110-1153.Chinese

Brown, A. J., Gibson, S., Hatton, D., & James, D. C. (2018). Transcriptome-Based Identification of the

Optimal Reference CHO Genes for Normalisation of qPCR Data. *Biotechnology Journal*, *13*(1). https://doi.org/10.1002/biot.201700259

Brown, A. J., & James, D. C. (2016). Precision control of recombinant gene transcription for CHO cell synthetic biology. *Biotechnology Advances*, *34*(5), 492–503. https://doi.org/10.1016/j.biotechadv.2015.12.012

Brown, M. E., Renner, G., Field, R. P., & Hassell, T. (1992). Process development for the production of recombinant antibodies using the glutamine synthetase (GS) system. *Cytotechnology*, *9*(1–3), 231–236. https://doi.org/10.1007/BF02521750

Brülhmann, D., Jordan, M., Hemberger, J., Sauer, M., Stettler, M., & Broly, H. (2015). Tailoring Recombinant Protein Quality by Rational Media Design. *Biotechnology Progress*, *31*(3). https://doi.org/10.1002/btpr.2089

Budge, J. D., Knight, T. J., Povey, J., Roobol, J., Brown, I. R., Singh, G., Dean, A., Turner, S., Jaques, C. M., Young, R. J., Racher, A. J., & Smales, C. M. (2020). Data for engineering lipid metabolism of Chinese hamster ovary (CHO) cells for enhanced recombinant protein production. *Data in Brief*, *29*(October 2019), 203–216. https://doi.org/10.1016/j.dib.2020.105217

Budge, J. D., Roobol, J., Singh, G., Mozzanino, T., Knight, T. J., Povey, J., Dean, A., Turner, S. J., Jaques, C. M., Young, R. J., Racher, A. J., & Smales, C. M. (2021). A proline metabolism selection system and its application to the engineering of lipid biosynthesis in Chinese hamster ovary cells. *Metabolic Engineering Communications*, *13*(April). https://doi.org/10.1016/j.mec.2021.e00179

Buss, N. A. P. S., Henderson, S. J., McFarlane, M., Shenton, J. M., & De Haan, L. (2012). Monoclonal antibody therapeutics: History and future. *Current Opinion in Pharmacology*, *12*(5), 615–622. https://doi.org/10.1016/j.coph.2012.08.001

Bustin, S. A. (2010). Why the need for qPCR publication guidelines? The case for MIQE. *Methods*, 217–226.

Bustin, S. A., Beaulieu, J. F., Huggett, J., Jaggi, R., Kibenge, F. S. B., Olsvik, P. A., Penning, L. C., & Toegel, S. (2010). MIQE précis: Practical implementation of minimum standard guidelines for fluorescence-based quantitative real-time PCR experiments. *BMC Molecular Biology*, *11*. https://doi.org/10.1186/1471-2199-11-74

Bustin, S. A., Benes, V., Garson, J. A., Hellemans, J., Huggett, J., Kubista, M., Mueller, R., Nolan, T., Pfaffl, M. W., Shipley, G. L., Vandesompele, J., & Wittwer, C. T. (2009). The MIQE guidelines: Minimum information for publication of quantitative real-time PCR experiments. *Clinical Chemistry*, *55*(4), 611–622. https://doi.org/10.1373/clinchem.2008.112797

Bustin, S. A., Benes, V., Nolan, T., & Pfaffl, M. W. (2005). Quantitative real-time RT-PCR - A perspective. *Journal of Molecular Endocrinology*, *34*, 597–601.

Bustin, S. A., Dhillon, H. S., Kirvell, S., Greenwood, C., Parker, M., Shipley, G. L., & Nolan, T. (2015). Variability of the Reverse Transcription Step : Practical Implications. *Clinical Chemistry*, *61*(1), 202–212. https://doi.org/10.1373/clinchem.2014.230615

Bustin, S. A., & Nolan, T. (2004). Pitfalls of Quantitative Real-Time Reverse-Transcription Polymerase Chain Reaction. *Journal of Biomolecular Techniques*, *15*(3), 155–166.

Bustin, S. A., & Nolan, T. (2017). Talking the talk , but not walking the walk : RT-qPCR as a paradigm for the lack of reproducibility in molecular research. *European Society for Clinical Investigation Journal Foundation*, 756–774. https://doi.org/10.1111/eci.12801

Butler, M., & Spearman, M. (2014). The choice of mammalian cell host and possibilities for glycosylation engineering. *Current Opinion in Biotechnology*, *30*, 107–112. https://doi.org/10.1016/j.copbio.2014.06.010

Byrne, B. J., Davis, M. S., Yamaguchi, J., Bergsma, D. J., & Subramanian, K. N. (1983). Definition of the simian virus 40 early promoter region and demonstration of a host range bias in the enhancement effect of the simian virus 40 72-base-pair repeat. *Proceedings of the National Academy of Sciences of the United States of America*, *80*(3 I), 721–725. https://doi.org/10.1073/pnas.80.3.721

Cacciatore, J. J., Chasin, L. A., & Leonard, E. F. (2010). Gene amplification and vector engineering to achieve rapid and high-level therapeutic protein production using the Dhfr-based CHO cell selection system. *Biotechnology Advances*, *28*(6), 673–681. https://doi.org/10.1016/j.biotechadv.2010.04.003

Caputo, A. T., Eder, O. M., Bereznakova, H., Pothuis, H., Ardevol, A., Newman, J., Nuttall, S., Peat, T. S., & Adams, T. E. (2021). Structure-guided selection of puromycin N-acetyltransferase mutants with enhanced selection stringency for deriving mammalian cell lines expressing recombinant proteins. *Scientific Reports*, *11*(1), 1–11. https://doi.org/10.1038/s41598-021-84551-9

Carrara, S. C., Ulitzka, M., Grzeschik, J., Kornmann, H., Hock, B., & Kolmar, H. (2021). From cell line development to the formulated drug product: The art of manufacturing therapeutic monoclonal antibodies. *International Journal of Pharmaceutics*, *594*(December 2020). https://doi.org/10.1016/j.ijpharm.2020.120164

Carroll, D. (2011). Genome Engineering With Zinc-Finger Nucleases. *Genetics*, *188*(August), 773–782. https://doi.org/10.1534/genetics.111.131433

Carter, P. J. (2006). Potent antibody therapeutics by design. *Nature Reviews Immunology*, *6*(5), 343–357. https://doi.org/10.1038/nri1837

Carvalho, L. S., Silva, O. B. da, Almeida, G. C. da, Oliveira, J. D. de, Parachin, N. S., & Carmo, T. S. (2017). Production Processes for Monoclonal Antibodies. *Fermentation Processes*, *February*. https://doi.org/10.5772/64263

Chaturvedi, K., Sun, S. Y., O'Brien, T., Liu, Y. J., & Brooks, J. W. (2014). Comparison of the behavior of CHO cells during cultivation in 24-square deep well microplates and conventional shake flask systems. *Biotechnology Reports*, *1*, 22–26. https://doi.org/https://doi.org/10.1016/j.btre.2014.04.001

Chen, K., Li, D., Li, H., Li, B., Li, J., Huang, L., Xu, X., Jiang, L., Jiang, C., Gu, H., & Fang, J. (2017). Genetic analysis of heterogeneous sub-clones in recombinant Chinese hamster ovary cells. *Applied Microbiology and Biotechnology*, *101*(14), 5785–5797.

Chen, L., Xie, Z., Teng, Y., Wang, M., Shi, M., Qian, L., Hu, M., Feng, J., Yang, X., Shen, B., & Guo, N. (2004). Highly efficient selection of the stable clones expressing antibody-IL-2 fusion protein by a dicistronic expression vector containing a mutant neo gene. *Journal of Immunological Methods*, *295*(1–2), 49–56. https://doi.org/10.1016/j.jim.2004.08.013

Chervoneva, I., Li, Y., Schulz, S., Croker, S., Wilson, C., Waldman, S. A., & Hyslop, T. (2010). *Selection of optimal reference genes for normalization in quantitative RT-PCR.*

Chiang, A. W. T., Li, S., Kellman, B. P., Chattopadhyay, G., Zhang, Y., Kuo, C. C., Gutierrez, J. M., Ghazi, F., Schmeisser, H., Ménard, P., Bjørn, S. P., Voldborg, B. G., Rosenberg, A. S., Puig, M., & Lewis, N. E. (2019). Combating viral contaminants in CHO cells by engineering innate immunity. *Scientific Reports*, *9*(1), 1–15. https://doi.org/10.1038/s41598-019-45126-x

Chin, C. L., Chin, H. K., Chin, C. S., Lai, E. T., & Ng, S. K. (2015). Engineering selection stringency on expression vector for the production of recombinant human alpha1-antitrypsin using Chinese Hamster ovary cells. *BMC Biotechnology*, *15*(1), 1–15. https://doi.org/10.1186/s12896-015-0145-9

Chiu, M. L., Goulet, D. R., Teplyakov, A., & Gilliland, G. L. (2019). Antibody Structure and Function: The Basis for Engineering Therapeutics. *Antibodies*, *8*(4), 55. https://doi.org/10.3390/antib8040055

Chong, Z. X., Yeap, S. K., & Ho, W. Y. (2021). Transfection types, methods and strategies: A technical review. In *PeerJ* (Vol. 9). https://doi.org/10.7717/peerj.11165

CHOZN® Platform Technical Bulletin. (n.d.). SAFC.

https://www.sigmaaldrich.com/deepweb/assets/sigmaaldrich/product/documents/245/459/chozn-platform-technical-bulletin.pdf

CHOZN® ZFN-Modified DHFR-/- CHO Cell Line Technical Bulletin. (n.d.). SAFC. https://www.sigmaaldrich.com/deepweb/assets/sigmaaldrich/product/documents/345/510/chozn_zfn_tech_bull.pdf

Chu, L., & Robinson, D. K. (2001). Industrial choices for protein production by large-scale cell culture. *Current Opinion in Biotechnology*, *12*(2), 180–187. https://doi.org/10.1016/S0958-1669(00)00197-X

Chusainow, J., Yang, Y. S., Yeo, J. H. M., Ton, P. C., Asvadi, P., Wong, N. S. C., & Yap, M. G. S. (2009). A study of monoclonal antibody-producing CHO cell lines: What makes a stable high producer? *Biotechnology and Bioengineering*, *102*(4), 1182–1196. https://doi.org/10.1002/bit.22158

Cockett, M. I., Bebbington, C. R., & Yarranton, G. T. (1990). High level expression of tissue inhibitor of metalloproteinases in CHO cells using GS gene amplification. *Nature*, *8*, 662–667.

Costa, A. R., Rodrigues, M. E., Henriques, M., Melton, D., Cunnah, P., Oliveira, R., & Azeredo, J. (2012). Evaluation of the OSCAR™ system for the production of monoclonal antibodies by CHO-K1 cells. *International Journal of Pharmaceutics*, *430*(1–2), 42–46. https://doi.org/10.1016/j.ijpharm.2012.03.028

Cymer, F., Beck, H., Rohde, A., & Reusch, D. (2018). Biologicals Therapeutic monoclonal antibody N-glycosylation – Structure , function and therapeutic potential. *Biologicals*, *52*(September 2017), 1–11. https://doi.org/10.1016/j.biologicals.2017.11.001

Dahodwala, H., & Lee, K. H. (2019). The fickle CHO: a review of the causes, implications, and potential alleviation of the CHO cell line instability problem. *Current Opinion in Biotechnology*, *60*(August 2018), 128–137. https://doi.org/10.1016/j.copbio.2019.01.011

De Souza, N. (2012). Primer: Genome editing with engineered nucleases. *Nature Methods*, *9*(1), 27. https://doi.org/10.1038/nmeth.1848

Dean, J., & Reddy, P. (2013). Metabolic analysis of antibody producing CHO cells in fed-batch production. *Biotechnology and Bioengineering*, *110*(6), 1735–1747. https://doi.org/10.1002/bit.24826

Declerck, P. J. (2012). Biologicals and biosimilars : a review of the science and its implications. *Genetics and Biosimilars Initiative Journal*, *1*(1), 13–16.

Delrue, I., Pan, Q., Baczmanska, A. K., Callens, B. W., & Verdoodt, L. L. M. (2018). Determination of

the Selection Capacity of Antibiotics for Gene Selection. *Biotechnology Journal*, *13*(8). https://doi.org/10.1002/biot.201700747

Demain, A. L., & Vaishnav, P. (2009). Production of recombinant proteins by microbes and higher organisms. *Biotechnology Advances*, *27*(3), 297–306. https://doi.org/10.1016/j.biotechadv.2009.01.008

Derveaux, S., Vandesompele, J., & Hellemans, J. (2010). How to do successful gene expression analysis using real-time PCR. *Methods*, *50*(4), 227–230. https://doi.org/10.1016/j.ymeth.2009.11.001

Dhiman, H., Campbell, M., Melcher, M., Smith, K. D., & Borth, N. (2020). Predicting favorable landing pads for targeted integrations in Chinese hamster ovary cell lines by learning stability characteristics from random transgene integrations. *Computational and Structural Biotechnology Journal*, *18*, 3632–3648. https://doi.org/10.1016/j.csbj.2020.11.008

Diep, J., Le, H., Le, K., Zasadzinska, E., Tat, J., Yam, P., Zastrow, R., Gomez, N., & Stevens, J. (2021). Microfluidic chip based single cell cloning to accelerate biologic production timelines. *Biotechnology Progress*.

Ding, M., Shen, L., Xiao, L., Liu, X., & Hu, J. (2021). A cell line development strategy to improve a bispecific antibody expression purity in CHO cells. *Biochemical Engineering Journal*, *166*(April 2020), 107857. https://doi.org/10.1016/j.bej.2020.107857

Distler, J. H. W., Jüngel, A., Kurowska-Stolarska, M., Michel, B. A., Gay, R. E., Gay, S., & Distler, O. (2005). Nucleofection: A new, highly efficient transfection method for primary human keratinocytes. *Experimental Dermatology*, *14*(4), 315–320. https://doi.org/10.1111/j.0906-6705.2005.00276.x

do Livramento, K. G., Freitas, N. C., Máximo, W. P. F., Zanetti, R., & Paiva, L. V. (2018). Gene expression profile analysis is directly affected by the selected reference gene: The case of leaf-cutting atta sexdens. *Insects*, *9*(1). https://doi.org/10.3390/INSECTS9010018

Dorai, H., Corisdeo, S., Ellis, D., Kinney, C., Chomo, M., Hawley-Nelson, P., Moore, G., Betenbaugh, M. J., & Ganguly, S. (2012). Early prediction of instability of chinese hamster ovary cell lines expressing recombinant antibodies and antibody-fusion proteins. *Biotechnology and Bioengineering*, *109*(4), 1016–1030. https://doi.org/10.1002/bit.24367

Duarte, T. M., Carinhas, N., Barreiro, L. C., Carrondo, M. J. T., Alves, P. M., & Teixeira, A. P. (2014). Metabolic responses of CHO cells to limitation of key amino acids. *Biotechnology and Bioengineering*, *111*(10), 2095–2106. https://doi.org/10.1002/bit.25266

Duetz, W. A. (2007). Microtiter plates as mini-bioreactors: miniaturization of fermentation methods.

Trends in Microbiology, *15*(10), 469–475. https://doi.org/10.1016/j.tim.2007.09.004

Dumont, J., Euwart, D., Mei, B., Estes, S., & Kshirsagar, R. (2016). Human cell lines for biopharmaceutical manufacturing: history, status, and future perspectives. *Critical Reviews in Biotechnology*, *36*(6), 1110–1122. https://doi.org/10.3109/07388551.2015.1084266

Durocher, Y., & Butler, M. (2009). Expression systems for therapeutic glycoprotein production. *Current Opinion in Biotechnology*, *20*, 700–707. https://doi.org/10.1016/j.copbio.2009.10.008

Eisenberg, D., Gill, H. S., Pfluegl, G. M. U., & Rotstein, S. H. (2000). Structure-function relationships of glutamine synthetases. *Biochimica et Biophysica Acta*, *1477*, 122–145.

Escoffre, J.-M., Nikolova, B., Mallet, L., Henri, J., Favard, C., Golzio, M., Teissie, J., Tsoneva, I., & Rols, M.-P. (2012). New Insights in the Gene Electrotransfer Process: Evidence for the Involvement of the Plasmid DNA Topology. *Current Gene Therapy*, *12*(5), 417–422. https://doi.org/10.2174/156652312802762554

Escoffre, J.-M., Portet, T., Wasungu, L., Teissié, J., Dean, D., & Rols, M. P. (2009). What is (Still not) known of the mechanism by which electroporation mediates gene transfer and expression in cells and tissues. In *Molecular Biotechnology* (Vol. 41, Issue 3, pp. 286–295). https://doi.org/10.1007/s12033-008-9121-0

European Medicines Agency. (1998). *ICH Topic Q 5 D Quality of biotechnological products: derivation and characterisation of cell substrates used for production of biotechnological/biological products*. https://doi.org/10.32388/yokp53

Evans, K., Albanetti, T., Venkat, R., Schoner, R., Savery, J., Miro-Quesada, G., Rajan, B., & Groves, C. (2015). Assurance of monoclonality in one round of cloning through cell sorting for single cell deposition coupled with high resolution cell imaging. *Biotechnology Progress*, *31*(5), 1172–1178. https://doi.org/10.1002/btpr.2145

Fan, L., Frye, C. C., & Racher, A. J. (2013). The use of glutamine synthetase as a selection marker: recent advances in Chinese hamster ovary cell line generation processes. *Pharmaceutical Bioprocessing*, *1*(5), 487–502. https://doi.org/10.4155/pbp.13.56

Fan, L., Kadura, I., Krebs, L. E., Hatfield, C. C., Shaw, M., & Frye, C. C. (2011). Improving the efficiency of CHO cell line generation using glutamine synthetase gene knockout cells. *Biotechnology and Bioengineering*, *109*(4), 1007–1015.

Fan, L., Kadura, I., Krebs, L. E., Larson, J. L., Bowden, D. M., & Frye, C. C. (2013). Development of a highly-efficient CHO cell line generation system with engineered SV40E promoter. *Journal of Biotechnology*, *168*(4), 652–658. https://doi.org/10.1016/j.jbiotec.2013.08.021

Fan, L., Rizzi, G., Bierilo, K., Tian, J., Chon Yee, J., Russell, R., & Das, T. K. (2017). Comparative study of therapeutic antibody candidates derived from mini-pool and clonal cell lines. *Biotechnology Progress*, *33*(6), 1456–1462.

Fann, C. H., Guirgis, F., Chen, G., Lao, M. S., & Piret, J. M. (2000). Limitations to the amplification and stability of human tissue-type plasminogen activator expression by Chinese hamster ovary cells. *Biotechnology and Bioengineering*, *69*(2), 204–212. https://doi.org/10.1002/(sici)1097-0290(20000720)69:2<204::aid-bit9>3.0.co;2-z

Feary, M., Racher, A. J., Young, R. J., & Smales, C. M. (2017). Methionine sulfoximine supplementation enhances productivity in GS–CHOK1SV cell lines through glutathione biosynthesis. *Biotechnology Progress*, *33*(1), 17–25. https://doi.org/10.1002/btpr.2372

Feige, M. J., Hendershot, L. M., & Buchner, J. (2010). How antibodies fold. *Trends in Biochemical Sciences*, *35*(4), 189–198. https://doi.org/10.1016/j.tibs.2009.11.005.How

Felberbaum, R. S. (n.d.). *The baculovirus expression vector system : A commercial manufacturing platform for viral vaccines and gene therapy vectors*. https://doi.org/10.1002/biot.201400438

Fleige, S., & Pfaffl, M. W. (2006). *RNA integrity and the effect on the real-time qRT-PCR performance*. *27*, 126–139. https://doi.org/10.1016/j.mam.2005.12.003

Fleige, S., Walf, *Æ*. V., Huch, *Æ*. S., & Pfaffl, M. W. (2006). *Comparison of relative mRNA quantification models and the impact of RNA integrity in quantitative real-time*. 1601–1613. https://doi.org/10.1007/s10529-006-9127-2

Flintoff, W. F., Davidson, S., & Siminovitch, L. (1976). Isolation and partial characterization of three methotrexate resistant phenotypes from CHO cells. *Somatic Cell Genetics*, *2*(3), 245–261.

Frenzel, A., Hust, M., & Schirrmann, T. (2013). Expression of recombinant antibodies. *Frontiers in Immunology*, *4*(JUL), 1–20. https://doi.org/10.3389/fimmu.2013.00217

Fromm, M., & Berg, P. (1983). Simian virus 40 early- and late-region promoter functions are enhanced by the 72-base-pair repeat inserted at distant locations and inverted orientations. *Molecular and Cellular Biology*, *3*(6), 991–999. https://doi.org/10.1128/mcb.3.6.991-999.1983

Frye, C., Deshpande, R., Estes, S., Francissen, K., Joly, J., Lubiniecki, A., Munro, T., Russell, R., Wang, T., & Anderson, K. (2016). Industry view on the relative importance of "clonality" of biopharmaceutical-producing cell lines. *Biologicals*, *44*(2), 117–122. https://doi.org/10.1016/j.biologicals.2016.01.001

Fus-Kujawa, A., Prus, P., Bajdak-Rusinek, K., Teper, P., Gawron, K., Kowalczuk, A., & Sieron, A. L. (2021). An Overview of Methods and Tools for Transfection of Eukaryotic Cells in vitro. In

Frontiers in Bioengineering and Biotechnology (Vol. 9). https://doi.org/10.3389/fbioe.2021.701031

Fussenegger, M., Bailey, J. E., Hauser, H., & Mueller, P. P. (1999). Genetic optimization of recombinant glycoprotein production by mammalian cells. *Trends in Biotechnology*, *17*(1), 35–42. https://doi.org/10.1016/S0167-7799(98)01248-7

Gaidukov, L., Wroblewska, L., Teague, B., Nelson, T., Zhang, X., Liu, Y., Jagtap, K., Mamo, S., Allen Tseng, W., Lowe, A., Das, J., Bandara, K., Baijuraj, S., Summers, N. M., Lu, T. K., Zhang, L., & Weiss, R. (2018). A multi-landing pad DNA integration platform for mammalian cell engineering. *Nucleic Acids Research*, *46*(8), 4072–4086. https://doi.org/10.1093/nar/gky216

Gaj, T., Gersbach, C. A., & Barbas, C. F. (2013). ZFN, TALEN, and CRISPR/Cas-based methods for genome engineering. *Trends in Biotechnology*, *31*(7), 397–405. https://doi.org/10.1016/j.tibtech.2013.04.004

Garenne, D., Haines, M. C., Romantseva, E. F., Freemont, P., Strychalski, E. A., & Noireaux, V. (2021). Cell-free gene expression. *Nature Reviews Methods Primers*, *1*(1). https://doi.org/10.1038/s43586-021-00046-x

Gaughan, C. L. (2016). The present state of the art in expression, production and characterization of monoclonal antibodies. *Molecular Diversity*, *20*(1), 255–270. https://doi.org/10.1007/s11030-015-9625-z

Gayatri Dhara, V., Mukesh Naik, H., Majewska, N. I., & Betenbaugh, M. J. (2018). Recombinant Antibody Production in CHO and NS0 Cells : Differences and Similarities. *BioDrugs*, *32*(6), 571–584. https://doi.org/10.1007/s40259-018-0319-9

Geisler, C., & Jarvis, D. L. (2018). Adventitious viruses in insect cell lines used for recombinant protein expression. *Protein Expression and Purification*, *144*(November 2017), 25–32. https://doi.org/10.1016/j.pep.2017.11.002

Ghaderi, D., Zhang, M., Hurtado-Ziola, N., & Varki, A. (2012). Production platforms for biotherapeutic glycoproteins. Occurrence, impact, and challenges of non-human sialylation. *Biotechnology and Genetic Engineering Reviews*, *28*, 147–176. https://doi.org/10.5661/bger-28-147

Goulet, D. R., & Atkins, W. M. (2020). Considerations for the Design of Antibody-Based Therapeutics. *Journal of Pharmaceutical Sciences*, *109*(1), 74–103. https://doi.org/10.1016/j.xphs.2019.05.031

Grilo, A. L., & Mantalaris, A. (2019). The Increasingly Human and Profitable Monoclonal Antibody Market. *Trends in Biotechnology*, *37*(1), 9–16. https://doi.org/10.1016/j.tibtech.2018.05.014

Guhaniyogi, J., & Brewer, G. (2001). Regulation of mRNA stability in mammalian cells. *Gene*, *265*(1–

2), 11–23. https://doi.org/10.1016/S0378-1119(01)00350-X

Guo, C., Fordjour, F. K., Tsai, S. J., Morrell, J. C., & Gould, S. J. (2021). Choice of selectable marker affects recombinant protein expression in cells and exosomes. *Journal of Biological Chemistry*, *297*(1), 1–13. https://doi.org/10.1016/j.jbc.2021.100838

Guo, D., Gao, A., Michels, D. A., Feeney, L., Eng, M., Chan, B., Laird, M. W., Zhang, B., Yu, X. C., Joly, J., Snedecor, B., & Shen, A. (2010). Mechanisms of unintended amino acid sequence changes in recombinant monoclonal antibodies expressed in Chinese Hamster Ovary (CHO) cells. *Biotechnology and Bioengineering*, *107*(1), 163–171. https://doi.org/10.1002/bit.22780

Guo, R., Lee, Y., Byrnes, C., & Miller, J. (2017). Puromycin Selection Confounds the RNA-Seq Profiles of Primary Human Erythroblasts. *Transcriptomics: Open Access*, *05*(01), 1–3. https://doi.org/10.4172/2329-8936.1000140

Häberle, J., Görg, B., Rutsch, F., Schmidt, E., Toutain, A., Benoist, J.-F., Gelot, A., Suc, A.-L., Höhne, W., Schliess, F., Häussinger, D., & Koch, H. G. (2005). Congenital Glutamine Deficiency with Glutamine Synthetase Mutations. *New England Journal of Medicine*, *353*(18), 1926–1933. https://doi.org/10.1056/nejmoa050456

Hacker, D. L., De Jesus, M., & Wurm, F. M. (2009). 25 years of recombinant proteins from reactor-grown cells - Where do we go from here? *Biotechnology Advances*, *27*(6), 1023–1027. https://doi.org/10.1016/j.biotechadv.2009.05.008

Hamaker, N. K., & Lee, K. H. (2018). Site-specific integration ushers in a new era of precise CHO cell line engineering. *Current Opinion in Chemical Engineering*, *22*, 152–160. https://doi.org/10.1016/j.coche.2018.09.011

Hansen, U., & Sharp, P. A. (1983). Sequences controlling in vitro transcription of SV40 promoters. *The EMBO Journal*, *2*(12), 2293–2303. https://doi.org/10.1002/j.1460-2075.1983.tb01737.x

Harcum, S. W., & Lee, K. H. (2016). CHO Cells Can Make More Protein. *Cell Systems*, *3*(5), 412–413. https://doi.org/10.1016/j.cels.2016.11.007

Harding, F. A., Stickler, M. M., Razo, J., & Dubridge, R. B. (2010). *The immunogenicity of humanized and fully human antibodies Residual immunogenicity resides in the CDR regions*. *2*(3), 256–265.

Hartman, T., & Tso, Y. J. (2013). *Vectors and host cells comprising a modified SV40 promoter for protein expression* (Patent No. WO 2013/081612 Al).

Hefzi, H., Ang, K. S., Hanscho, M., Bordbar, A., Ruckerbauer, D., Lakshmanan, M., Orellana, C. A., Baycin-Hizal, D., Huang, Y., Ley, D., Martinez, V. S., Kyriakopoulos, S., Jiménez, N. E., Zielinski, D. C., Quek, L. E., Wulff, T., Arnsdorf, J., Li, S., Lee, J. S., … Lewis, N. E. (2016). A

Consensus Genome-scale Reconstruction of Chinese Hamster Ovary Cell Metabolism. *Cell Systems*, *3*(5), 434-443.e8. https://doi.org/10.1016/j.cels.2016.10.020

Hetz, C., Zhang, K., & Kaufman, R. J. (2020). Mechanisms, regulation and functions of the unfolded protein response. In *Nature Reviews Molecular Cell Biology* (Vol. 21, Issue 8, pp. 421–438). https://doi.org/10.1038/s41580-020-0250-z

Higuchi, R., Dollinger, G., Walsh, S. P., & Griffith, R. (1992). Simultaneous amplification and detection of specific DNA sequences. *Biotechnology*, *10*.

Higuchi, R., Fockler, C., Dollinger, G., & Watson, R. (1993). Kinetic PCR Analysis: Real-time Monitoring of DNA Amplification Reactions. *Biotechnology*, *11*, 1026–1030. http://www.nature.com/naturebiotechnology

Ho, S. C. L., Bardor, M., Feng, H., Mariati, Tong, Y. W., Song, Z., Yap, M. G. S., & Yang, Y. (2012). IRES-mediated Tricistronic vectors for enhancing generation of high monoclonal antibody expressing CHO cell lines. *Journal of Biotechnology*, *157*(1), 130–139. https://doi.org/10.1016/j.jbiotec.2011.09.023

Ho, S. C. L., Koh, E. Y. C., Soo, B. P. C., Chao, S. H., & Yang, Y. (2016). Evaluating the use of a CpG free promoter for long-term recombinant protein expression stability in Chinese hamster ovary cells. *BMC Biotechnology*, *16*(1), 1–13. https://doi.org/10.1186/s12896-016-0300-y

Holland, P. M., Abramson, R. D., Watson, R., & Gelfand, D. H. (1991). Detection of specific PCR product by utilizing the 5'-3' exonuclease activity of Thermus aquaticus DNA polymerase. *Proceedings of the National Academy of Sciences*, *88*, 7276–7280.

Holtta, E., & Pohjanpelto, P. (1983). Polyamine starvation causes accumulation of cadaverine and its derivatives in a polyamine-dependent strain of Chinese-hamster ovary cells. *Biochemical Journal*, *210*(3), 945–948. https://doi.org/10.1042/bj2100945

Hossler, P., Khattak, S. F., & Li, Z. J. (2009). Optimal and consistent protein glycosylation in mammalian cell culture. *Glycobiology*, *19*(9), 936–949. https://doi.org/10.1093/glycob/cwp079

Hu, Z., Hsu, W., Pynn, A., Ng, D., Quicho, D., Adem, Y., Kwong, Z., Mauger, B., Joly, J., Snedecor, B., Laird, M. W., Andersen, D., & Shen, A. (2017). A strategy to accelerate protein production from a pool of clones in Chinese hamster ovary cells for toxicology studies. *Biotechnology Progress*, *33*(6), 1449–1455.

Huang, M., Bao, J., & Nielsen, J. (2014). Biopharmaceutical protein production by Saccharomyces cerevisiae : current state and future prospects. *Pharmaceutical Bioprocessing*, *2*, 167–182.

Huggett, J., & Bustin, S. A. (2011). *Standardisation and reporting for nucleic acid quantification*.

https://doi.org/10.1007/s00769-011-0769-y

Huggett, J., Dheda, K., Bustin, S. A., & Zumla, A. (2005). Real-time RT-PCR normalisation ; strategies and considerations. *Genes and Immunity*, *6*, 279–284. https://doi.org/10.1038/sj.gene.6364190

Jackson, D. A., Symonst, R. H., & Berg, P. (1972). Biochemical Method for Inserting New Genetic Information into DNA of Simian Virus 40: Circular SV40 DNA Molecules Containing Lambda Phage Genes and the Galactose Operon of Escherichia coli. *Proceedings of the National Academy of Sciences*, *69*(10), 2904–2909.

Jayapal, K. P., F., W. K., & Wei-Shou, H. (2007). *Recombinant protein therapeutics from CHO cells - 20 years and counting*. *105*(2), 332–333. https://doi.org/10.1378/chest.105.2.332

Jayapal, K. P., Wlaschin, K. F., Hu, W. S., & Yap, M. G. S. (2007). Recombinant protein therapeutics from CHO Cells - 20 years and counting. *Chemical Engineering Progress*, *103*(10), 40–47.

Jayapal, K. P., Wlashin, K. F., & Hu, W.-S. (2007). *Recombinant protein therapeutics from CHO cells - 20 years and counting*. *105*(2), 332–333. https://doi.org/10.1378/chest.105.2.332

Jefferis, R. (2009). Glycosylation as a strategy to improve antibody-based therapeutics. *Nature Reviews Drug Discovery*, *8*(3), 226–234. https://doi.org/10.1038/nrd2804

Jossé, L., Zhang, L., & Smales, C. M. (2018). Application of microRNA Targeted 3′UTRs to Repress DHFR Selection Marker Expression for Development of Recombinant Antibody Expressing CHO Cell Pools. *Biotechnology Journal*, *13*(10), 1–8. https://doi.org/10.1002/biot.201800129

Jun, S. C., Kim, M. S., Hong, H. J., & Lee, G. M. (2006). Limitations to the development of humanized antibody producing chinese hamster ovary cells using glutamine synthetase-mediated gene amplification. *Biotechnology Progress*, *22*(3), 770–780. https://doi.org/10.1021/bp060004t

Kao, F. T., & Puck, T. T. (1968). Genetics of somatic mammalian cells, VII. Induction and isolation of nutritional mutants in Chinese hamster cells. *Proceedings of the National Academy of Sciences of the United States of America*, *60*(4), 1275–1281. https://doi.org/10.1073/pnas.60.4.1275

Kaplon, H., Chenoweth, A., Crescioli, S., & Reichert, J. M. (2022). Antibodies to watch in 2022. *MAbs*, *14*(1).

Kaplon, H., Muralidharan, M., Schneider, Z., & Reichert, J. M. (2020). Antibodies to watch in 2020. *MAb*, *12*(1).

Kaplon, H., & Reichert, J. M. (2021a). Antibodies to watch in 2021. *MAbs*, *13*(1). https://doi.org/10.1080/19420862.2020.1860476

Kaplon, H., & Reichert, J. M. (2021b). Antibodies to watch in 2021. *MAbs*, *13*(1).

https://doi.org/10.1080/19420862.2020.1860476

Kaufman, R. J. (2000). Overview of vector design for mammalian gene expression. *Applied Biochemistry and Biotechnology - Part B Molecular Biotechnology*, *16*(2), 151–160. https://doi.org/10.1385/MB:16:2:151

Kaufman, R. J., & Sharp, P. A. (1982). Amplification and expression of sequences cotransfected with a modular dihydrofolate reductase complementary DNA gene. *Journal of Molecular Biology*, *159*(4), 601–621. https://doi.org/10.1016/0022-2836(82)90103-6

Kawabe, Y., Inao, T., Komatsu, S., Huang, G., Ito, A., Omasa, T., & Kamihira, M. (2017). Improved recombinant antibody production by CHO cells using a production enhancer DNA element with repeated transgene integration at a predetermined chromosomal site. *Journal of Bioscience and Bioengineering*, *123*(3), 390–397. https://doi.org/10.1016/j.jbiosc.2016.10.011

Kawabe, Y., Komatsu, S., Komatsu, S., Murakami, M., Ito, A., Sakuma, T., Nakamura, T., Yamamoto, T., & Kamihira, M. (2018). Targeted knock-in of an scFv-Fc antibody gene into the hprt locus of Chinese hamster ovary cells using CRISPR/Cas9 and CRIS-PITCh systems. *Journal of Bioscience and Bioengineering*, *125*(5), 599–605. https://doi.org/10.1016/j.jbiosc.2017.12.003

Kelley, B. (2020). Developing therapeutic monoclonal antibodies at pandemic pace. *Nature Biotechnology*, *38*(5), 540–545. https://doi.org/10.1038/s41587-020-0512-5

Kelley, B., Kiss, R., & Laird, M. (2018). A different perspective: How much innovation is really needed for monoclonal antibody production using mammalian cell technology? In *Advances in Biochemical Engineering/Biotechnology* (Vol. 165, pp. 443–462). https://doi.org/10.1007/10_2018_59

Keyt, B. A., Baliga, R., Sinclair, A. M., Carroll, S. F., & Peterson, M. S. (2020). *Structure , Function , and Therapeutic Use of IgM Antibodies*. *Figure 1*.

Kim, J. Y., Kim, Y. G., & Lee, G. M. (2012). CHO cells in biotechnology for production of recombinant proteins: Current state and further potential. *Applied Microbiology and Biotechnology*, *93*(3), 917–930. https://doi.org/10.1007/s00253-011-3758-5

Kim, M., O'Callaghan, P. M., Droms, K. A., & James, D. C. (2011). A mechanistic understanding of production instability in CHO cell lines expressing recombinant monoclonal antibodies. *Biotechnology and Bioengineering*, *108*(10), 2434–2446. https://doi.org/10.1002/bit.23189

Kim, T. K., & Eberwine, J. H. (2010). Mammalian cell transfection: the present and the future. *Analytical and Bioanalytical Chemistry*, *397*, 3173–3178.

Kim, Y.-G., Cha, J., & Chandrasegaran, S. (1996). Hybrid restriction enzymes : Zinc finger fusions to

Fok I cleavage domain. *Proceedings of the National Academy of Sciences of the United States of America*, *93*(February), 1156–1160.

Klein, C., & Kopetzki, E. (2009). *Promoter* (Patent No. WO 2009/003622 Al).

Ko, P., Misaghi, S., Hu, Z., Zhan, D., Tsukuda, J., Yim, M., Sanford, M., Shaw, D., Shiratori, M., Snedecor, B., Laird, M., & Shen, A. (2018). Probing the importance of clonality: Single cell subcloning of clonally derived CHO cell lines yields widely diverse clones differing in growth, productivity, and product quality. *Biotechnology Progress*, *34*(3), 624–634. https://doi.org/10.1002/btpr.2594

Köhler, G., & Milstein, C. (1975). Continuous cultures of fused cells secreting antibody of predefined specificity. *Nature*, *256*(Mopc 21), 495–497.

Kozak, M. (1987). At least six nucleotides preceding the AUG initiator codon enhance translation in mammalian cells. *Journal of Molecular Biology*, *196*, 947–950.

Kozak, M. (1999). Initiation of translation in prokaryotes and eukaryotes. *Gene*, *234*, 187–208.

Kozak, M. (2001). *Constraints on reinitiation of translation in mammals*. *29*(24), 5226–5232.

Kozera, B., & Rapacz, M. (2013). Reference genes in real-time PCR. *Journal of Applied Genetics*, *54*, 391–406. https://doi.org/10.1007/s13353-013-0173-x

Krajewski, W. W., Collins, R., Holmberg-Schiavone, L., Jones, T. A., Karlberg, T., & Mowbray, S. L. (2008). Crystal structures of mammalian glutamine synthetases illustrate substrate-induced conformational changes and provide opportunities for drug and herbicide design. *Journal of Molecular Biology*, *375*(1), 217–228.

Kubista, M., Andrade, J. M., Bengtsson, M., Forootan, A., Jonak, J., Lind, K., Sindelka, R., Sjoback, R., Sjogreen, B., Strombom, L., Stahlberg, A., & Zoric, N. (2006). The real-time polymerase chain reaction. *Molecular Aspects of Medicine*, *27*, 95–125. https://doi.org/10.1016/j.mam.2005.12.007

Kumar, P., Nagarajan, A., & Uchil, P. D. (2019). Electroporation. In M. R. Green & J. Sambrook (Eds.), *Molecular cloning collection*.

Kurdyukov, S., & Bullock, M. (2016). DNA methylation analysis: Choosing the right method. *Biology*, *5*(1), 3. https://doi.org/10.3390/biology5010003

Kuriakose, A., Chirmule, N., & Nair, P. (2016). Immunogenicity of Biotherapeutics : Causes and Association with Posttranslational Modifications. *Journal of Immunology Research*, *2016*.

Kwaks, T. H. J., Barnett, P., Hemrika, W., Siersma, T., Sewalt, R. G. A. B., Satijn, D. P. E., Brons, J. F., Van Blokland, R., Kwakman, P., Kruckeberg, A. L., Kelder, A., & Otte, A. P. (2003).

Identification of anti-repressor elements that confer high and stable protein production in mammalian cells. *Nature Biotechnology*, *21*(5), 553–558. https://doi.org/10.1038/nbt814

Kwaks, T. H. J., Sewalt, R. G. A. B., Van Blokland, R., Siersma, T. J., Kasiem, M., Kelder, A., & Otte, A. P. (2005). Targeting of a histone acetyltransferase domain to a promoter enhances protein expression levels in mammalian cells. *Journal of Biotechnology*, *115*(1), 35–46. https://doi.org/10.1016/j.jbiotec.2004.07.012

Labrijn, A. F., Janmaat, M. L., Reichert, J. M., & Parren, P. W. H. I. (2019). Bispecific antibodies: a mechanistic review of the pipeline. *Nature Reviews Drug Discovery*, *18*(8), 585–608. https://doi.org/10.1038/s41573-019-0028-1

Lai, T., Yang, Y., & Ng, S. K. (2013). Advances in mammalian cell line development technologies for recombinant protein production. *Pharmaceuticals*, *6*(5), 579–603. https://doi.org/10.3390/ph6050579

Lalonde, M., & Durocher, Y. (2017). Therapeutic glycoprotein production in mammalian cells. *Journal of Biotechnology*, *251*(December 2016), 128–140. https://doi.org/10.1016/j.jbiotec.2017.04.028

Lanza, A. M., Kim, D. S., & Alper, H. S. (2013). Evaluating the influence of selection markers on obtaining selected pools and stable cell lines in human cells. *Biotechnology Journal*, *8*(7), 811–821. https://doi.org/10.1002/biot.201200364

Lao-Gonzalez, T., Bueno-Soler, A., Duran-Hernandez, A., Sosa-Aguiar, K., Hinojosa-Puerta, L. E., Hernandez-Garcia, T., de la Luz-Hernandez, K. R., Palacios-Oliva, J., & Boggiano-Ayo, T. (2021). Screening and selection strategy for the establishment of biosimilar to trastuzumab-expressing CHO-K1 cell lines. *AMB Express*, *11*(1). https://doi.org/10.1186/s13568-020-01157-6

Lee, S., & Lee, D. K. (2018). What is the proper way to apply the multiple comparison test? *Korean Journal of Anesthesiology*, *70*.

Lefever, S., Hellemans, J., Pattyn, F., Przybylski, D. R., Taylor, C., Geurts, R., Untergasser, A., & Vandesompele, J. (2009). RDML: Structured language and reporting guidelines for real-time quantitative PCR data. *Nucleic Acids Research*, *37*(7), 2065–2069. https://doi.org/10.1093/nar/gkp056

Li, F., Vijayasankaran, N., Shen, A., Kiss, R., & Amanullah, A. (2010). Cell culture processes for monoclonal antibody production. *MAbs*, *2*(5), 466–479. https://doi.org/10.4161/mabs.2.5.12720

Li, H., Chen, K., Wang, Z., Li, D., Lin, J., Yu, C., Yu, F., Wang, X., Huang, L., Jiang, C., Gu, H., & Fang, J. (2015). Genetic analysis of the clonal stability of Chinese hamster ovary cells for recombinant protein production. *Molecular BioSystems*, *12*(1), 102–109.

https://doi.org/10.1039/c5mb00627a

Li, J., & Zhu, Z. (2010). Research and development of next generation of antibody-based therapeutics. *Acta Pharmacologica Sinica*, *31*(9), 1198–1207. https://doi.org/10.1038/aps.2010.120

Liaw, S. H., & Eisenberg, D. (1994). Structural model for the reaction mechanism of glutamine synthetase, based on five crystal structures of enzyme-substrate complexes. *Biochemistry*, *33*(3), 675–681.

Liljedahl, M., Aspland, S. E., & Segal, D. J. (2003). *Methods and compositions for using zinc finger endonucleases to enhance homologous recombination* (Patent No. US 2003/0232410 A1).

Lin, P. C., Chan, K. F., Kiess, I. A., Tan, J., Shahreel, W., Wong, S. Y., & Song, Z. (2019). Attenuated glutamine synthetase as a selection marker in CHO cells to efficiently isolate highly productive stable cells for the production of antibodies and other biologics. *MAbs*, *11*(5), 965–976. https://doi.org/10.1080/19420862.2019.1612690

Lindskog, E. K., Fischer, S., Wenger, T., & Schulz, P. (2018). Host Cells. In G. Jagschies, E. Lindskog, K. Lacki, & P. M. Galliher (Eds.), *Biopharmaceutical processing: development, design, and implementation of manufacturing processes* (Elsevier, pp. 111–130).

Liu, P.-Q., & Miller, J. C. (2010). *Methods and compositions for inactivating glutamine synthetase gene expression* (Patent No. US 2010/0129869 A1).

Liu, P.-Q., & Miller, J. C. (2015). *Methods and compositions for inactivating glutamine synthetase gene expression* (Patent No. AU 2015201300 A1).

Liu, P. Q., Chan, E. M., Cost, G. J., Zhang, L., Wang, J., Miller, J. C., Guschin, D. Y., Reik, A., Holmes, M. C., Mott, J. E., Collingwood, T. N., & Gregory, P. D. (2010). Generation of a triple-gene knockout mammalian cell line using engineered zinc-finger nucleases. *Biotechnology and Bioengineering*, *106*(1), 97–105. https://doi.org/10.1002/bit.22654

Liu, Z., Chen, O., Wall, B. J., Zheng, M., Zhou, Y., Wang, L., Ruth Vaseghi, H., Qian, L., & Liu, J. (2017). Systematic comparison of 2A peptides for cloning multi-genes in a polycistronic vector. *Scientific Reports*, *7*(1), 1–9.

Livak, K. J., Flood, S. J. A., Marmaro, J., Giusti, W., & Deetz, K. (1995). Oligonucleotides with fluorescent dyes at opposite ends provide a quenched probe system useful for detecting PCR product and nucleic acid hybridization. *Genome Research*, *4*, 357–362.

Livak, K. J., & Schmittgen, T. D. (2001). Analysis of Relative Gene Expression Data Using Real- Time Quantitative PCR and the 2-Delta Delta Ct Method. *Methods*, *25*, 402–408. https://doi.org/10.1006/meth.2001.1262

Loetscher, P., Pratt, G., & Rechsteiner, M. (1991). The C terminus of mouse ornithine decarboxylase confers rapid degradation on dihydrofolate reductase: Support for the pest hypothesis. *Journal of Biological Chemistry*, *266*(17), 11213–11220. https://doi.org/10.1016/s0021-9258(18)99150-7

Lu, R. M., Hwang, Y. C., Liu, I. J., Lee, C. C., Tsai, H. Z., Li, H. J., & Wu, H. C. (2020). Development of therapeutic antibodies for the treatment of diseases. *Journal of Biomedical Science*, *27*(1), 1–30. https://doi.org/10.1186/s12929-019-0592-z

Ma, X., Zhang, L., Zhang, L., Wang, C., Guo, X., Yang, Y., & Wang, L. (2020). *Validation and identification of reference genes in Chinese hamster ovary cells for Fc-fusion protein production*. 690–702. https://doi.org/10.1177/1535370220914058

Makrides, S. C. (1999). Components of vectors for gene transfer and expression in mammalian cells. *Protein Expression and Purification*, *17*(2), 183–202. https://doi.org/10.1006/prep.1999.1137

Makrides, S. C. (2003). Vectors for gene expression in mammalian cells. In *Gene Transfer and Expression in Mammalian Cells*.

Matasci, M., Hacker, D. L., Baldi, L., & Wurm, F. M. (2008). Recombinant therapeutic protein production in cultivated mammalian cells: current status and future prospects. *Drug Discovery Today: Technologies*, *5*(2–3), 37–42. https://doi.org/10.1016/j.ddtec.2008.12.003

Mignon, C., Sodoyer, R., & Werle, B. (2015). Antibiotic-free selection in biotherapeutics: Now and forever. *Pathogens*, *4*(2), 157–181. https://doi.org/10.3390/pathogens4020157

Misaghi, S., Shaw, D., Louie, S., Nava, A., Simmons, L., Snedecor, B., Poon, C., Paw, J. S., Gilmour-Appling, L., & Cupp, J. E. (2016). Slashing the timelines: Opting to generate high-titer clonal lines faster via viability-based single cell sorting. *Biotechnology Progress*, *32*(1), 198–207. https://doi.org/10.1002/btpr.2204

Mora, A., Zhang, S. S., Carson, G., Nabiswa, B., Hossler, P., & Yoon, S. (2018). Sustaining an efficient and effective CHO cell line development platform by incorporation of 24-deep well plate screening and multivariate analysis. *Biotechnology Progress*, *34*(1), 175–186. https://doi.org/10.1002/btpr.2584

Moritz, B., Becker, P. B., & Göpfert, U. (2015). CMV promoter mutants with a reduced propensity to productivity loss in CHO cells. *Scientific Reports*, *5*, 1–8. https://doi.org/10.1038/srep16952

Mullis, K. B., & Faloona, F. A. (1987). Specific synthesis of DNA in vitro via a polymerase-catalyzed chain reaction. *Methods in Enzymology*, *155*.

Munro, T. P., Le, K., Le, H., Zhang, L., Stevens, J., Soice, N., Benchaar, S. A., Hong, R. W., & Goudar, C. T. (2017). Accelerating patient access to novel biologics using stable pool-derived product for

non-clinical studies and single clone-derived product for clinical studies. *Biotechnology Progress*, *33*(6), 1476–1482. https://doi.org/10.1002/btpr.2572

Nematpour, F., Mahboudi, F., Vaziri, B., Khalaj, V., Ahmadi, S., & Ahmadi, M. (2017). *Evaluating the expression profile and stability of different UCOE containing vector combinations in mAb-producing CHO cells*. 1–9. https://doi.org/10.1186/s12896-017-0330-0

Neville, J. J., Orlando, J., Mann, K., McCloskey, B., & Antoniou, M. N. (2017). Ubiquitous Chromatin-opening Elements (UCOEs): Applications in biomanufacturing and gene therapy. *Biotechnology Advances*, *35*(5), 557–564. https://doi.org/10.1016/j.biotechadv.2017.05.004

Ng, S. K., Tan, T. R. M., Wang, Y., Ng, D., Goh, L. T., Bardor, M., Wong, V. V. T., & Lam, K. P. (2012). Production of Functional Soluble Dectin-1 Glycoprotein Using an IRES-Linked Destabilized-Dihydrofolate Reductase Expression Vector. *PLoS ONE*, *7*(12). https://doi.org/10.1371/journal.pone.0052785

Ng, S. K., Wang, D. I. C., & Yap, M. G. S. (2007). Application of destabilizing sequences on selection marker for improved recombinant protein productivity in CHO-DG44. *Metabolic Engineering*, *9*(3), 304–316. https://doi.org/10.1016/j.ymben.2007.01.001

Nielsen, J. (2012). *Production of biopharmaceutical proteins by yeast.pdf*. 207–211.

Niwa, H., Yamamura, K., & Miyazaki, J. (1991). Efficient selection for high-expression transfectants with a novel eukaryotic vector. *Gene*, *108*(2), 193–199. https://doi.org/10.1016/0378-1119(91)90434-D

Noh, S. M., Shin, S., & Lee, G. M. (2018). Comprehensive characterization of glutamine synthetase-mediated selection for the establishment of recombinant CHO cells producing monoclonal antibodies. *Scientific Reports*, *8*(1), 1–11. https://doi.org/10.1038/s41598-018-23720-9

Nolan, T., Hands, R. E., & Bustin, S. A. (2006). Quantification of mRNA using real-time RT-PCR. *Nature Protocols*, *1*(3), 1559–1582. https://doi.org/10.1038/nprot.2006.236

Omasa, T. (2002). Gene Amplification and Its Application in Cell and Tissue Engineering. *Journal of Bioscience and Bioengineering*, *94*(6), 600–605. https://doi.org/10.1016/s1389-1723(02)80201-8

Omasa, T., Onitsuka, M., & Kim, W.-D. (2010). Cell Engineering and Cultivation of Chinese Hamster Ovary (CHO) Cells. *Current Pharmaceutical Biotechnology*, *11*(3), 233–240. https://doi.org/10.2174/138920110791111960

Orlando, S. J., Santiago, Y., Dekelver, R. C., Freyvert, Y., Boydston, E. A., Moehle, E. A., Choi, V. M., Gopalan, S. M., Lou, J. F., Li, J., Miller, J. C., Holmes, M. C., Gregory, P. D., Urnov, F. D., Cost, G. J., Biosciences, S., & Bvld, C. (2010). *Zinc-finger nuclease-driven targeted integration into*

mammalian genomes using donors with limited chromosomal homology. *38*(15), 1–15. https://doi.org/10.1093/nar/gkq512

Osterlehner, A., Simmeth, S., & Göpfert, U. (2011). Promoter methylation and transgene copy numbers predict unstable protein production in recombinant chinese hamster ovary cell lines. *Biotechnology and Bioengineering*, *108*(11), 2670–2681. https://doi.org/10.1002/bit.23216

Otte, P. A., Van Blokland, H. J. M., Kwaks, J. T. H., & Sewalt, R. G. A. B. (2009). *Selection of host cells expressing protein at high levels* (Patent No. US 2009/0098601 A1). http://dx.doi.org/10.1016/j.ccr.2011.01.031

Owczarek, B., Gerszberg, A., & Hnatuszko-Konka, K. (2019). A brief reminder of systems of production and chromatography based recovery of recombinant protein biopharmaceuticals. *BioMed Research International*.

Padhi, B. K., Singh, M., Rosales, M., Pelletier, G., & Cakmak, S. (2018). Biomolecular Detection and Quanti fi cation A PCR-based quantitative assay for the evaluation of mRNA integrity in rat samples. *Biomolecular Detection and Quantification*, *15*(December 2017), 18–23. https://doi.org/10.1016/j.bdq.2018.02.001

Pallavicini, M. G., DeTeresa, P. S., Rosette, C., Gray, J. W., & Wurm, F. M. (1990). Effects of methotrexate on transfected DNA stability in mammalian cells. *Molecular and Cellular Biology*, *10*(1), 401–404. https://doi.org/10.1128/mcb.10.1.401-404.1990

Panina, Y., Germond, A., Masui, S., & Watanabe, T. M. (2018). Validation of Common Housekeeping Genes as Reference for qPCR Gene Expression Analysis During iPS Reprogramming Process. *Scientific Reports*, *8*(8716).

Pardo, R., Engelhard, M., Hagen, K., Jørgensen, R. B., Rehbinder, E., Schnieke, A., Szmulewicz, M., & Thiele, F. (2009). The role of means and goals in technology acceptance A differentiated landscape of public perceptions of pharming. *European Molecular Biology Organization Reports*, *10*(10), 1069–1075.

Paredes, V., Park, J. S., Jeong, Y., Yoon, J., & Baek, K. (2013). Unstable expression of recombinant antibody during long-term culture of CHO cells is accompanied by histone H3 hypoacetylation. *Biotechnology Letters*, *35*(7), 987–993. https://doi.org/10.1007/s10529-013-1168-8

Pérez-Novo, C. A., Claeys, C., Speleman, F., Cauwenberge, P. Van, Bachert, C., & Vandesompele, J. (2005). Impact of RNA quality on reference gene expression stability. *BioTechniques*, *39*(1), 1–3.

Pfaffl, M. W. (2001). A new mathematical model for relative quantification in real-time RT-PCR. *Nuclei*, *29*(9), 2069–2082. https://doi.org/10.1111/j.1365-2966.2012.21196.x

Pfaffl, M. W., Tichopad, A., Prgomet, C., & Neuvians, T. P. (2004). Determination of stable housekeeping genes, differentially regulated target genes and sample integrity: BestKeeper - Excel-based tool using pair-wise correlations. *Biotechnology Letters*, *26*, 509–515.

Pham, P. L., Kamen, A., & Durocher, Y. (2006). Large-scale transfection of mammalian cells for the fast production of recombinant protein. *Molecular Biotechnology*, *34*(2), 225–237. https://doi.org/10.1385/MB:34:2:225

Pilbrough, W., Munro, T. P., & Gray, P. (2009). Intraclonal protein expression heterogeneity in recombinant CHO cells. *PLoS ONE*, *4*(12). https://doi.org/10.1371/journal.pone.0008432

Porter, R. (1959). The hydrolysis of rabbit y-globulin and antibodies with crystalline papain. *The Biochemical Journal*, *73*(1948), 119–126. https://doi.org/10.1042/bj0730119

Poulain, A. (2018). *Caractérisation du système inductible au cumate pour la production de protéines thérapeutiques en cellules CHO*.

Priola, J. J., Calzadilla, N., Baumann, M., Borth, N., Tate, C. G., & Betenbaugh, M. J. (2016). High-throughput screening and selection of mammalian cells for enhanced protein production. *Biotechnology Journal*, *11*(7), 853–865. https://doi.org/10.1002/biot.201500579

Puck, T. T. (1957). The genetics of somatic mammalian cells. In *Advances in biological and medical physics* (Vol. 5, Issue 49). ACADEMIC PRESS INC. https://doi.org/10.1016/b978-1-4832-3111-2.50006-7

Qian, Y., Sowa, S. W., Aron, K. L., Xu, P., Langsdorf, E., Warrack, B., Aranibar, N., Tremml, G., Xu, J., McVey, D., Reily, M., Khetan, A., Borys, M. C., & Li, Z. J. (2020). New insights into genetic instability of an industrial CHO cell line by orthogonal omics. *Biochemical Engineering Journal*, *164*(June), 107799. https://doi.org/10.1016/j.bej.2020.107799

Quenneville, S. P., Chapdelaine, P., Rousseau, J., Beaulieu, J., Caron, N. J., Skuk, D., Mills, P., Olivares, E. C., Calos, M. P., & Tremblay, J. P. (2004). Nucleofection of muscle-derived stem cells and myoblasts with φC31 integrase: Stable expression of a full-length-dystrophin fusion gene by human myoblasts. *Molecular Therapy*, *10*(4), 679–687. https://doi.org/10.1016/j.ymthe.2004.05.034

Radonić, A., Thulke, S., Mackay, I. M., Landt, O., Siegert, W., & Nitsche, A. (2004). Guideline to reference gene selection for quantitative real-time PCR. *Biochemical and Biophysical Research Communications*, *313*(4), 856–862. https://doi.org/10.1016/j.bbrc.2003.11.177

Rajendra, Y., Balasubramanian, S., McCracken, N., Norris, D., Lian, Z., Schmitt, M., Frye, C. C., & Barnard, G. C. (2017). Evaluation of piggyBac-mediated CHO pools to enable material generation

to support GLP toxicology studies. *Biotechnology Progress*, *33*(6), 1436–1448.

Rajendran, S., Balasubramanian, S., Webster, L., Lee, M., Vavilala, D., Kulikov, N., Choi, J., Tang, C., Hunter, M., Wang, R., Kaur, H., Karunakaran, S., Sitaraman, V., Minshull, J., & Boldog, F. (2021). Accelerating and de-risking CMC development with transposon-derived manufacturing cell lines. *Biotechnology and Bioengineering*, *118*(6), 2301–2311. https://doi.org/10.1002/bit.27742

Ramezani, A., Mahmoudi Maymand, E., Yazdanpanah-Samani, M., Hosseini, A., Toghraie, F. S., & Ghaderi, A. (2017). Improving Pertuzumab production by gene optimization and proper signal peptide selection. *Protein Expression and Purification*, *135*, 24–32. https://doi.org/10.1016/j.pep.2017.04.013

Raymaekers, M., Smets, R., Maes, B., & Cartuyvels, R. (2009). Checklist for optimization and validation of real-time PCR assays. *Journal of Clinical Laboratory Analysis*, *23*, 145–151.

Raymond, C. (2015). *Production d'IgG sialylées en CHO et impact sur leurs fonctions effectrices*. Montreal.

Reinhart, D., Damjanovic, L., Kaisermayer, C., Sommeregger, W., Gili, A., Gasselhuber, B., Castan, A., Mayrhofer, P., Grünwald-Gruber, C., & Kunert, R. (2019). Bioprocessing of Recombinant CHO-K1, CHO-DG44, and CHO-S: CHO Expression Hosts Favor Either mAb Production or Biomass Synthesis. *Biotechnology Journal*, *14*(3), 1–11. https://doi.org/10.1002/biot.201700686

Rita Costa, A., Elisa Rodrigues, M., Henriques, M., Azeredo, J., & Oliveira, R. (2010). Guidelines to cell engineering for monoclonal antibody production. *European Journal of Pharmaceutics and Biopharmaceutics*, *74*(2), 127–138. https://doi.org/10.1016/j.ejpb.2009.10.002

Ritter, A., Voedisch, B., Wienberg, J., Wilms, B., Geisse, S., Jostock, T., & Laux, H. (2016). Deletion of a telomeric region on chromosome 8 correlates with higher productivity and stability of CHO cell lines. *Biotechnology and Bioengineering*, *113*(5), 1084–1093. https://doi.org/10.1002/bit.25876

Roca, B. C., Lao, N., Barron, N., Doolan, P., & Clynes, M. (2019). An arginase-based system for selection of transfected CHO cells without the use of toxic chemicals. *Journal of Biological Chemistry*, *294*(49), 18756–18768. https://doi.org/10.1074/jbc.RA119.011162

Rodolosse, A., Barbat, A., Chantret, I., Lacasa, M., Brot-Laroche, E., Zweibaum, A., & Rousset, M. (1998). Selecting agent hygromycin B alters expression of glucose-regulated genes in transfected Caco-2 cells. *American Journal of Physiology - Gastrointestinal and Liver Physiology*, *274*(5 37-5), 931–938. https://doi.org/10.1152/ajpgi.1998.274.5.g931

Rodrigues, M. E., Rita Costa, A., Henriques, M., Cunnah, P., Melton, D. W., Azeredo, J., & Oliveira,

R. (2013). Advances and drawbacks of the adaptation to serum-free culture of CHO-K1 cells for monoclonal antibody production. *Applied Biochemistry and Biotechnology*, *169*, 1279–1291.

Rodriguez-Conde, S., Inman, S., Lindo, V., Amery, L., Tang, A., Okorji-Obike, U., Du, W., Bosch, B.-J., Wichgers Schreur, P. J., Kortekaas, J., Sola, I., Enjuanes, L., Kerry, L., Mahal, K., Hulley, M., & Daramola, O. (2022). Suitability of transiently expressed antibodies for clinical studies: product quality consistency at different production scales. *MAbs*, *14*(1). https://doi.org/10.1080/19420862.2022.2052228

Romanova, N., & Noll, T. (2018). Engineered and Natural Promoters and Chromatin-Modifying Elements for Recombinant Protein Expression in CHO Cells. *Biotechnology Journal*, *13*(3), 1–13. https://doi.org/10.1002/biot.201700232

Ross, J. (1995). mRNA stability in mammalian cells. *Microbiological Reviews*, *59*(3), 423–450. https://doi.org/10.1128/mmbr.59.3.423-450.1995

Rouiller, Y., Bielser, J. M., Brühlmann, D., Jordan, M., Broly, H., & Stettler, M. (2016). Screening and assessment of performance and molecule quality attributes of industrial cell lines across different fed-batch systems. *Biotechnology Progress*, *32*(1), 160–170.

Ryman, J. T., & Meibohm, B. (2017). Pharmacokinetics of Monoclonal Antibodies. *CPT Pharmacometrics Syst. Pharmacol*, *6*, 576–588. https://doi.org/10.1002/psp4.12224

Ryu, A. H., Eckalbar, W. L., Kreimer, A., Yosef, N., & Ahituv, N. (2017). Use antibiotics in cell culture with caution: Genome-wide identification of antibiotic-induced changes in gene expression and regulation. *Scientific Reports*, *7*(1), 1–9. https://doi.org/10.1038/s41598-017-07757-w

Sacco, S. A., Tuckowski, A. M., Trenary, I., Kraft, L., Betenbaugh, M. J., Young, J. D., & Smith, K. D. (2022). Attenuation of glutamine synthetase selection marker improves product titer and reduces glutamine overflow in Chinese hamster ovary cells. *Biotechnology and Bioengineering*, *February*. https://doi.org/10.1002/bit.28084

Saiki, R. K., Scharf, S., Faloona, F., Mullis, K. B., Horn, G. T., Erlich, H. A., & Arnheim, N. (1985). Enzymatic Amplification of Beta-Globin Genomic Sequences and Restriction Site Analysis for Diagnosis of Sickle Cell Anemia. *Science*, *230*.

Saito, S., Namisaki, H., Hiraishi, K., Takahashi, N., & Iida, S. (2019). A stable engineered human IgG3 antibody with decreased aggregation during antibody expression and low pH stress. *Protein Science*, *28*(5), 900–909. https://doi.org/10.1002/pro.3598

Santiago, Y., Chan, E., Liu, P.-Q., Orlando, S., Zhang, L., Urnov, F. D., Holmes, M. C., Guschin, D., Waite, A., Miller, J. C., Rebar, E. J., Gregory, P. D., Klug, A., & Collingwood, T. N. (2008).

Targeted genome knockout in mammalian cells by using engineered zinc-finger nucleases. *Proceedings of the National Academy of Sciences*, *105*(15), 5809–5814.

Saunders, F., Sweeney, B., Antoniou, M. N., Stephens, P., & Cain, K. (2015). Chromatin function modifying elements in an industrial antibody production platform - Comparison of UCOE, MAR, STAR and cHS4 elements. *PLoS ONE*, *10*(4), 1–20. https://doi.org/10.1371/journal.pone.0120096

Sautter, K., & Enenkel, B. (2005). Selection of high-producing CHO cells using NPT selection marker with reduced enzyme activity. *Biotechnology and Bioengineering*, *89*(5), 530–538. https://doi.org/10.1002/bit.20374

Scarcelli, J. J., Hone, M., Beal, K., Ortega, A., Figueroa, B., Starkey, J. A., & Anderson, K. (2018). Analytical subcloning of a clonal cell line demonstrates cellular heterogeneity that does not impact process consistency or robustness. *Biotechnology Progress*, *34*(3), 602–612. https://doi.org/10.1002/btpr.2646

Schillberg, S., Raven, N., Spiegel, H., Rasche, S., & Buntru, M. (2019). Critical analysis of the commercial potential of plants for the production of recombinant proteins. *Frontiers in Plant Science*, *10*. https://doi.org/10.3389/fpls.2019.00720

Schmieder, V., Fieder, J., Drerup, R., Gutierrez, E. A., Guelch, C., Stolzenberger, J., Stumbaum, M., Mueller, V. S., Higel, F., Bergbauer, M., Bornhoefft, K., Wittner, M., Gronemeyer, P., Braig, C., Huber, M., Reisenauer-Schaupp, A., Mueller, M. M., Schuette, M., Puengel, S., … Fischer, S. (2022). Towards maximum acceleration of monoclonal antibody development: Leveraging transposase-mediated cell line generation to enable GMP manufacturing within 3 months using a stable pool. *Journal of Biotechnology*, *349*, 53–64. https://doi.org/10.1016/j.jbiotec.2022.03.010

Schröder, M., & Kaufman, R. J. (2005). ER stress and the unfolded protein response. *Mutation Research*, *569*, 29–63. https://doi.org/10.1016/j.mrfmmm.2004.06.056

Schroeder, A., Mueller, O., Stocker, S., Salowsky, R., Leiber, M., Gassmann, M., Lightfoot, S., Menzel, W., Granzow, M., & Ragg, T. (2006). *The RIN : an RNA integrity number for assigning integrity values to RNA measurements*. *14*, 1–14. https://doi.org/10.1186/1471-2199-7-3

Schroeder, H. W., & Cavacini, L. (2010). Structure and function of immunoglobulins. *Journal of Allergy and Clinical Immunology*, *125*(2 SUPPL. 2), S41–S52. https://doi.org/10.1016/j.jaci.2009.09.046

Sealover, N., Kobayashi, K., George, H., Kayser, K., & Lin, N. (n.d.). *A Novel CHO Host Cell Line Using HPRT as an Amplifiable Selectable Marker for Therapeutic Protein Production*. 4.

Seth, G., Charaniya, S., Wlaschin, K. F., & Hu, W. S. (2007). In pursuit of a super producer-alternative paths to high producing recombinant mammalian cells. *Current Opinion in Biotechnology*, *18*(6),

557–564. https://doi.org/10.1016/j.copbio.2007.10.012

Shatz, W., Ng, D., Dutina, G., Wong, A. W., Dunshee, D. R., Sonoda, J., Shen, A., & Scheer, J. M. (2016). An efficient route to bispecific antibody production using single-reactor mammalian co-culture. *MAbs*, *8*(8), 1487–1497. https://doi.org/10.1080/19420862.2016.1234569

Shen, X., Pitol, A. K., Bachmann, V., Hacker, D. L., Baldi, L., & Wurm, F. M. (2015). A simple plasmid-based transient gene expression method using High Five cells. *Journal of Biotechnology*, *216*, 67–75. https://doi.org/10.1016/j.jbiotec.2015.10.007

Shi, J., Ma, Y., Zhu, J., Chen, Y., Sun, Y., Yao, Y., Yang, Z., & Xie, J. (2018). A review on electroporation-based intracellular delivery. In *Molecules* (Vol. 23, Issue 11). https://doi.org/10.3390/molecules23113044

Silver, N., Best, S., Jiang, J., & Thein, S. L. (2006). Selection of housekeeping genes for gene expression studies in human reticulocytes using real-time PCR. *BMC Molecular Biology*, *9*, 1–9. https://doi.org/10.1186/1471-2199-7-33

Silverman, A. D., Karim, A. S., & Jewett, M. C. (2020). Cell-free gene expression: an expanded repertoire of applications. *Nature Reviews Genetics*, *21*(3), 151–170. https://doi.org/10.1038/s41576-019-0186-3

Slivac, I., Guay, D., Mangion, M., Champeil, J., & Gaillet, B. (2017). Non-viral nucleic acid delivery methods. *Expert Opinion on Biological Therapy*, *17*(1), 105–118.

Spencer, S., Gugliotta, A., Koenitzer, J., Hauser, H., & Wirth, D. (2015). Stability of single copy transgene expression in CHOK1 cells is affected by histone modifications but not by DNA methylation. *Journal of Biotechnology*, *195*(1), 15–29. https://doi.org/10.1016/j.jbiotec.2014.12.009

Srirangan, K., Loignon, M., & Durocher, Y. (2020). The use of site-specific recombination and cassette exchange technologies for monoclonal antibody production in Chinese Hamster ovary cells: retrospective analysis and future directions. *Critical Reviews in Biotechnology*, *40*(6), 833–851. https://doi.org/10.1080/07388551.2020.1768043

Stahlberg, A., Hakansson, J., Xian, X., Semb, H., & Kubista, M. (2004). Properties of the Reverse Transcription Reaction in mRNA Quantification. *Clinical Chemistry*, *515*, 509–515. https://doi.org/10.1373/clinchem.2003.026161

Steinmeyer, D. E., & McCormick, E. L. (2008). The art of antibody process development. *Drug Discovery Today*, *13*(13–14), 613–618. https://doi.org/10.1016/j.drudis.2008.04.005

Streit, S., Michalski, C. W., Erkan, M., Kleeff, J., & Friess, H. (2009). Northern blot analysis for

detection and quantification of RNA in pancreatic cancer cells and tissues. *Nature Protocols*, *4*(1), 37–43. https://doi.org/10.1038/nprot.2008.216

Stuible, M., van Lier, F., Croughan, M. S., & Durocher, Y. (2018). Beyond preclinical research: production of CHO-derived biotherapeutics for toxicology and early-phase trials by transient gene expression or stable pools. *Current Opinion in Chemical Engineering*, *22*, 145–151. https://doi.org/10.1016/J.COCHE.2018.09.010

Sun, T., Kwok, W. C., Chua, K. J., Lo, T. M., Potter, J., Yew, W. S., Chesnut, J. D., Hwang, I. Y., & Chang, M. W. (2020). Development of a Proline-Based Selection System for Reliable Genetic Engineering in Chinese Hamster Ovary Cells. *ACS Synthetic Biology*, *9*(7), 1864–1872. https://doi.org/10.1021/acssynbio.0c00221

Swartz, J. (2006). Developing cell-free biology for industrial applications. *Journal of Industrial Microbiology and Biotechnology*, *33*(7), 476–485. https://doi.org/10.1007/s10295-006-0127-y

Swartz, J. R. (2001). Advances in Escherichia coli production of therapeutic proteins. *Current Opinion in Biotechnology*, *12*(2), 195–201. https://doi.org/10.1016/S0958-1669(00)00199-3

Taylor, S. C., Nadeau, K., Abbasi, M., Lachance, C., Nguyen, M., & Fenrich, J. (2019). The Ultimate qPCR Experiment: Producing Publication Quality, Reproducible Data the First Time. *Trends in Biotechnology*, *37*(7).

Teixeira, A. P., Stücheli, P., Ausländer, S., Ausländer, D., Schönenberger, P., Hürlemann, S., & Fussenegger, M. (2022). CelloSelect – A synthetic cellobiose metabolic pathway for selection of stable transgenic CHO cell lines. *Metabolic Engineering*, *70*(January), 23–30. https://doi.org/10.1016/j.ymben.2022.01.001

Tejwani, V., Chaudhari, M., Rai, T., & Sharfstein, S. T. (2021). High-throughput and automation advances for accelerating single-cell cloning, monoclonality and early phase clone screening steps in mammalian cell line development for biologics production. *Biotechnology Progress*, *37*(6), 1–13. https://doi.org/10.1002/btpr.3208

Tharmalingam, T., Barkhordarian, H., Tejeda, N., Daris, K., Yaghmour, S., Yam, P., Lu, F., Goudar, C., Munro, T., & Stevens, J. (2018). Characterization of phenotypic and genotypic diversity in subclones derived from a clonal cell line. *Biotechnology Progress*, *34*(3), 613–623.

Thellin, O., ElMoualij, B., Heinen, E., & Zorzi, W. (2009). A decade of improvements in quantification of gene expression and internal standard selection. *Biotechnology Advances*, *27*, 323–333.

Thellin, O., Zorzi, W., Lakaye, B., Borman, B. De, Coumans, B., Hennen, G., Grisar, T., Igout, A., & Heinen, E. (1999). Housekeeping genes as internal standards : use and limits. *Journal of*

Biotechnology, *75*, 291–295.

Tiselius, B. Y. A., D, P., & Kabat, E. A. (1939). An electrophoretic study of immune sera and purified antibody preparations. *Journal of Experimental Medecine*, *69*(7), 119–131.

Tossolini, I., Gugliotta, A., Lopez Diaz, F., Kratje, R., & Prieto, C. (2022). Screening of CHO-K1 endogenous promoters for expressing recombinant proteins in mammalian cell cultures. *Plasmid*, *120*. https://doi.org/10.1016/j.plasmid.2022.102620

Tripathi, N. K., & Shrivastava, A. (2019). Recent Developments in Bioprocessing of Recombinant Proteins: Expression Hosts and Process Development. *Frontiers in Bioengineering and Biotechnology*, *7*(December). https://doi.org/10.3389/fbioe.2019.00420

Tsong, T. Y. (1989). Electroporation of cell membranes. *Electroporation and Electrofusion in Cell Biology*, 149–163.

Tsong, T. Y. (1990). On electroporation of cell membranes and some related phenomena. In *Journal of Electroanalytical Chemistry* (Vol. 299, Issue 3, pp. 271–295). https://doi.org/10.1016/0022-0728(90)87529-S

Tsuchikama, K. (2018). Antibody-drug conjugates : recent advances in conjugation and linker chemistries. *Protein & Cell*, *9*(1), 33–46. https://doi.org/10.1007/s13238-016-0323-0

Turilova, V. I., Goryachaya, T. S., & Yakovleva, T. K. (2021). Chinese hamster ovary cell line DXB - 11 : chromosomal instability and karyotype heterogeneity. *Molecular Cytogenetics*, 1–13. https://doi.org/10.1186/s13039-021-00528-3

Untergasser, A., Cutcutache, I., Koressaar, T., Ye, J., Faircloth, B. C., Remm, M., & Rozen, S. G. (2012). Primer3-new capabilities and interfaces. *Nucleic Acids Research*, *40*(15), 1–12. https://doi.org/10.1093/nar/gks596

Urlaub, G., & Chasin, L. A. (1980). *Isolation of Chinese hamster cell mutants deficient in dihydrofolate reductase activity*. *77*(7), 4216–4220.

Urlaub, G., Käs, E., Carothers, A. M., & Chasin, L. A. (1983). Deletion of the diploid dihydrofolate reductase locus from cultured mammalian cells. *Cell*, *33*(2), 405–412. https://doi.org/10.1016/0092-8674(83)90422-1

Urnov, F. D., Miller, J. C., Lee, Y. L., Beausejour, C. M., Rock, J. M., Augustus, S., Jamieson, A. C., Porteus, M. H., Gregory, P. D., & Holmes, M. C. (2005). Highly efficient endogenous human gene correction using designed zinc-finger nucleases. *Nature*, *435*(7042), 646–651. https://doi.org/10.1038/nature03556

US Food and Drug Administration. (1992). Supplement to the Points to consider in the production and

testing of new drugs and biologics produced by recombinant DNA technology: nucleic acid characterization and genetic stability. In *Pharmacy Today*. https://doi.org/10.1016/s1042-0991(15)32122-8

Vandesompele, J., Preter, K. De, Poppe, B., Roy, N. Van, & Paepe, A. De. (2002). *Accurate normalization of real-time quantitative RT -PCR data by geometric averaging of multiple internal control genes*. 1–12.

Vara, J., Malpartida, F., Hopwood, D. A., & Jiménez, A. (1985). Cloning and expression of a puromycin N-acetyl transferase gene from Streptomyces alboniger in Streptomyces lividans and Escherichia coli. *Gene*, *33*(2), 197–206. https://doi.org/10.1016/0378-1119(85)90094-0

Vara, J., Portela, A., Ortin, J., & Jimenez, A. (1986). Expression in mammalian cells of a gene from Streptomyces alboniger conferring puromycin resistance. *Nucleic Acids Research*, *32*(DATABASE ISS.), 8235–8251. https://doi.org/10.1093/nar/gkh100

Voronina, E. V., Seregin, Y. A., Litvinova, N. A., Shvets, V. I., & Shukurov, R. R. (2016). Design of a stable cell line producing a recombinant monoclonal anti-TNFα antibody based on a CHO cell line. *SpringerPlus*, *5*(1). https://doi.org/10.1186/s40064-016-3213-2

Walsh, G. (2000). Biopharmaceutical benchmarks. *Nature Biotechnology*, *32*(10), 992–1000. https://doi.org/10.1038/nbt.3040

Walsh, G. (2004). *Second-generation biopharmaceuticals*. *58*, 185–196. https://doi.org/10.1016/j.ejpb.2004.03.012

Walsh, G. (2006). Biopharmaceutical benchmarks 2006. *Nature Biotechnology*, *24*(7), 769–776. https://doi.org/10.1038/nbt0706-769

Walsh, G. (2010). Biopharmaceutical benchmarks 2010. *Nature Biotechnology*, *32*(10), 992–1000. https://doi.org/10.1038/nbt.3040

Walsh, G. (2014). Biopharmaceutical benchmarks 2014. *Nat Biotech*, *32*(10), 992–1000. https://doi.org/10.1038/nbt.3040

Walsh, G. (2018). Biopharmaceutical benchmarks 2018. *Nature Biotechnology*, *36*(12), 1136–1145. https://doi.org/10.1038/nbt.4305

Walsh, G., & Jefferis, R. (2006). Post-translational modifications in the context of therapeutic proteins. *Nature Biotechnology*, *24*(10), 1241–1252. https://doi.org/10.1038/nbt1252

Wang, Q., Chen, Y., Park, J., Liu, X., Hu, Y., Wang, T., McFarland, K., & Betenbaugh, M. J. (2019). Design and Production of Bispecific Antibodies. *Antibodies*, *8*(3), 43. https://doi.org/10.3390/antib8030043

Weidle, U. H., Buckel, P., & Wienberg, J. (1988). Amplified expression constructs for human tissue-type plasminogen activator in Chinese hamster ovary cells: instability in the absence of selective pressure. *Gene*, *66*(2), 193–203. https://doi.org/10.1016/0378-1119(88)90356-3

Welch, J. T., & Arden, N. S. (2019). Considering "clonality": A regulatory perspective on the importance of the clonal derivation of mammalian cell banks in biopharmaceutical development. *Biologicals*, *62*(October), 16–21. https://doi.org/10.1016/j.biologicals.2019.09.006

Westwood, A. D., Rowe, D. A., & Clarke, H. R. G. (2010). Improved recombinant protein yield using a codon deoptimized DHFR selectable marker in a CHEF1 expression plasmid. *Biotechnology Progress*, *26*(6), 1558–1566. https://doi.org/10.1002/btpr.491

Whale, A. S., Huggett, J. F., Cowen, S., Speirs, V., Shaw, J., Ellison, S., Foy, C. A., Scott, D. J., & Le, L. (2012). *Comparison of microfluidic digital PCR and conventional quantitative PCR for measuring copy number variation*. *40*(11). https://doi.org/10.1093/nar/gks203

Wildeman, A. G. (1988). Regulation of SV40 early gene expression. *Biochemistry and Cell Biology*, *66*(6), 567–577. https://doi.org/10.1139/o88-067

Wurm, F. M. (2004). Production of recombinant protein therapeutics in cultivated mammalian cells. *Nature Biotechnology*, *22*(11), 1393–1398. https://doi.org/10.1038/nbt1026

Wurm, F. M. (2013). CHO Quasispecies—Implications for Manufacturing Processes. *Processes*, *1*, 296–311. https://doi.org/10.3390/pr1030296

Wurm, F. M., & Hacker, D. (2011). First CHO genome. *Nature Biotechnology*, *29*(8), 718–720. https://doi.org/10.1038/nbt.1943

Wurm, F. M., & Petropoulos, C. J. (1994). Plasmid integration, amplification fna cytogenetics in CHO cells: questions and comments. *Biologicals*, *22*, 95–102. https://doi.org/10.1006/biol.1994.1015

Wurm, F., & Wurm, M. (2017). Cloning of CHO Cells, Productivity and Genetic Stability—A Discussion. *Processes*, *5*(2), 20. https://doi.org/10.3390/pr5020020

Xu, G., Yu, C., Wang, W., Fu, C., Liu, H., Zhu, Y., Li, Y., Liu, C., Fu, Z., Wu, G., Li, M., Guo, S., Yu, X., Du, J., Yang, Y., Duan, M., Cui, Y., Feng, H., & Wang, L. (2022). Quality comparability assessment of a SARS-CoV-2-neutralizing antibody across transient, mini-pool-derived and single-clone CHO cells. In *mAbs* (Vol. 14, Issue 1). https://doi.org/10.1080/19420862.2021.2005507

Xu, X., Nagarajan, H., Lewis, N. E., Pan, S., Cai, Z., Liu, X., Chen, W., Xie, M., Wang, W., Hammond, S., Andersen, M. R., Neff, N., Passarelli, B., Koh, W., Fan, H. C., Wang, J., Gui, Y., Lee, K. H., Betenbaugh, M. J., … Wang, J. (2011). The genomic sequence of the Chinese hamster ovary

(CHO)-K1 cell line. *Nature Biotechnology*, *29*(8), 735–741. https://doi.org/10.1038/nbt.1932

Yamada, K., & Ito, Y. (2019). Recent Chemical Approaches for Site-Specific Conjugation of Native Antibodies: Technologies toward Next-Generation Antibody–Drug Conjugates. *ChemBioChem*, *20*, 2729–2737.

Yang, B., Zhou, J., Zhao, H., Wang, A., Lei, Y., Xie, Q., & Xiong, S. (2019). Study of the mechanism for increased protein expression via transcription potency reduction of the selection marker. *Bioprocess and Biosystems Engineering*, *42*(5), 799–806. https://doi.org/10.1007/s00449-019-02083-z

Yao, J., Weng, Y., Dickey, A., & Wang, K. Y. (2015). Plants as factories for human pharmaceuticals: Applications and challenges. *International Journal of Molecular Sciences*, *16*(12), 28549–28565. https://doi.org/10.3390/ijms161226122

Yenofsky, R. L., Fine, M., & Pellow, J. W. (1990). A mutant neomycin phosphotransferase II gene reduces the resistance of transformants to antibiotic selection pressure. *Proceedings of the National Academy of Sciences of the United States of America*, *87*(9), 3435–3439. https://doi.org/10.1073/pnas.87.9.3435

Yeo, J. H. M., Ho, S. C. L., Mariati, M., Koh, E., Tay, S. J., Woen, S., Zhang, P., & Yang, Y. (2017). Optimized Selection Marker and CHO Host Cell Combinations for Generating High Monoclonal Antibody Producing Cell Lines. *Biotechnology Journal*, *12*(12), 1–11. https://doi.org/10.1002/biot.201700175

Yim, M., & Shaw, D. (2018). Achieving greater efficiency and higher confidence in single-cell cloning by combining cell printing and plate imaging technologies. *Biotechnology Progress*, *34*(6), 1454–1459. https://doi.org/10.1002/btpr.2698

Yoshikawa, T., Nakanishi, F., Itami, S., Kameoka, D., Omasa, T., Katakura, Y., Kishimoto, M., & Suga, K. I. (2000). Evaluation of stable and highly productive gene amplified CHO cell line based on the location of amplified genes. *Cytotechnology*, *33*(1–3), 37–46. https://doi.org/10.1023/a:1008111328771

Yoshikawa, T., Nakanishi, F., Ogura, Y., Oi, D., Omasa, T., Katakura, Y., Kishimoto, M., & Suga, K. I. (2000). Amplified gene location in chromosomal DNA affected recombinant protein production and stability of amplified genes. *Biotechnology Progress*, *16*(5), 710–715. https://doi.org/10.1021/bp000114e

Zhai, Y., Zhang, D., Yu, L., Sun, F., & Sun, F. (2019). SmartBac, a new baculovirus system for large protein complex production. *Journal of Structural Biology: X*, *1*. https://doi.org/10.1016/j.yjsbx.2019.100003

Zhang, F., Frost, A. R., Blundell, M. P., Bales, O., Antoniou, M. N., & Thrasher, A. J. (2010). A Ubiquitous Chromatin Opening Element (UCOE) confers resistance to DNA methylation-mediated silencing of lentiviral vectors. *Molecular Therapy*, *18*(9), 1640–1649. https://doi.org/10.1038/mt.2010.132

Zhang, Z., Chen, J., Wang, J., Gao, Q., Ma, Z., Xu, S., Zhang, L., Cai, J., & Zhou, W. (2021). Reshaping cell line development and CMC strategy for fast responses to pandemic outbreak. *Biotechnology Progress*, *37*(5), 1–17. https://doi.org/10.1002/btpr.3186

Zhou, Y., Shaw, D., Lam, C., Tsukuda, J., Yim, M., Tang, D., Louie, S., Laird, M. W., Snedecor, B., & Misaghi, S. (2018). Beating the odds: The poisson distribution of all input cells during limiting dilution grossly underestimates whether a cell line is clonally-derived or not. *Biotechnology Progress*, *34*(3), 559–569. https://doi.org/10.1002/btpr.2560

Zubiaga, A. M., Belasco, J. G., & Greenberg, M. E. (1995). The nonamer UUAUUUAUU is the key AU-rich sequence motif that mediates mRNA degradation. *Molecular and Cellular Biology*, *15*(4), 2219–2230. https://doi.org/10.1128/mcb.15.4.2219

Zuker, M. (2003). Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Research*, *31*(13), 3406–3415. https://doi.org/10.1093/nar/gkg595

Zúñiga, R. A., Gutiérrez-González, M., Collazo, N., Sotelo, P. H., Ribeiro, C. H., Altamirano, C., Lorenzo, C., Aguillón, J. C., & Molina, M. C. (2019). Development of a new promoter to avoid the silencing of genes in the production of recombinant antibodies in chinese hamster ovary cells. *Journal of Biological Engineering*, *13*(1), 1–12. https://doi.org/10.1186/s13036-019-0187-y

Appendix 1: Supplementary data related to results chapter 1

Supplementary data on clone stability

Supplementary data related to 3.1.3.2.3 CHOZN® CHO K1 clones' productivity is detailed hereinafter.

Table 29: Cumulated IVC variations between PDL_{20}, PDL_{40}, and PDL_{60}

Variations between PDL were calculated on the last common day of culture between the three productivity assays.

Clone ID	PDL	Last production day (PD)	Cumulated IVC on last PD [cells/mL/day]	Last common production day (PD)	Cumulated IVC on last common PD [cells.mL/day]	ΔCumulated IVC between PDL_{20}-PDL_{40}	ΔCumulated IVC between PDL_{40}-PDL_{60}	ΔCumulated IVC between PDL_{20}-PDL_{60}	Conclusion on expression level stability
CL 15D08	PDL_{20}	PD_{12}	148.1	PD_9	94.8	+19.2%		-15.7%	**Stable**
	PDL_{40}	PD_{12}	178.4	PD_9	113		-29.3%		
	PDL_{60}	PD_9	79.9	PD_9	79.9				
CL 16E06	PDL_{20}	PD_9	49.4	PD_7	49.4	-31.2%		-39.7%	Unstable
	PDL_{40}	PD_{12}	71.4	PD_7	34		-12.3%		
	PDL_{60}	PD_7	29.7	PD_7	29.8				

Table 30: IgG4 titer variations between PDL_{20}, PDL_{40}, and PDL_{60}

Variations between PDL were calculated on the last common day of culture between the three productivity assays.

Clone ID	PDL	Last production day (PD)	[IgG4] on last PD [mg/L]	Last common production day (PD)	[IgG4] on last common PD [mg/L]	Δ[IgG4] between PDL_{20}-PDL_{40}	Δ[IgG4] between PDL_{40}-PDL_{60}	Δ[IgG4] between PDL_{20}-PDL_{60}	Conclusion on expression level stability
CL 15D08	PDL_{20}	PD_{12}	1247.3	PD_9	846.6	-4.3%		-32.4%	Unstable
	PDL_{40}	PD_{12}	1150	PD_9	809.9		-29.4%		
	PDL_{60}	PD_9	572.0	PD_9	572.0				
CL 16E06	PDL_{20}	PD_9	715.6	PD_7	381.5	9.4%		2.1%	**Stable**
	PDL_{40}	PD_{12}	1230.9	PD_7	417.3		-6.7%		
	PDL_{60}	PD_7	389.4	PD_7	389.4				

Table 31: Mean process Q_P variations between PDL_{20}, PDL_{40}, and PDL_{60}

Variations between PDL were calculated on the last common day of culture between the three productivity assays.

Clone ID	**PDL**	**Last production day (PD)**	**Mean process Q_p on last PD [pg/cell/day]**	**Last common production day (PD)**	**Mean process Q_p on last common PD [pg/cell/day]**	**ΔQ_p between PDL_{20}-PDL_{40}**	**ΔQ_p between PDL_{40}-PDL_{60}**	**ΔQ_p between PDL_{20}-PDL_{60}**	**Conclusion on mean process Q_p stability**
CL 15D08	PDL_{20}	PD_{12}	8.4	PD_9	8.9	-19.1%		-19.1	**Stable**
	PDL_{40}	PD_{12}	6.4	PD_9	7.2		0%		
	PDL_{60}	PD_9	7.2	PD_9	7.2				
CL 16E06	PDL_{20}	PD_9	14.5	PD_7	12.1	0.8%		8.3%	**Stable**
	PDL_{40}	PD_{12}	17.3	PD_7	12.2		7.3%		
	PDL_{60}	PD_7	13.1	PD_7	13.1				

Appendix 2: Supplementary data related to results chapter 2

Details on statistical analyses on clones' productivity

Statistical analysis related to 3.2.4.1.3 Impact of cell line development process on clone productivity is detailed hereinafter.

Significance of mean titer difference between long cell line development process (MP-CL) and short cell line development process (DSCC) in 96-well plates, 24-deepwell plates, TPP® tubes and ambr250® bioreactors was evaluated by a statistical analysis, including assessment of normal distribution of data, and ANOVA. Assessment of normal distribution was performed on studied ln-transformed response residuals and normality was demonstrated only for 96-well plate and 24-deepwell plate datasets (data not shown). Comparison of mean titer between long cell line development process (MP-CL) and short cell line development process (DSCC) is shown on Figure 63. Mean titer in 96-well plates and 24-deepwell plates is significantly lower with short cell line development process (DSCC) compared to long cell line development process (MP-CL) (p-value below $\alpha_{5\%}$). Normality was not demonstrated for TPP® tube and ambr250® bioreactor datasets; therefore, non-parametric assay was performed (Wilcoxon exact test). Mean titer in ambr250® bioreactors is significantly lower with short cell line development process (DSCC) compared to long cell line development process (MP-CL) (p-value below $\alpha_{5\%}$).

Similar analysis was performed on mean process specific productivity. Significance of mean process specific productivity difference between long cell line development process (MP-CL) and short cell line development process (DSCC) in 24-deepwell plates, TPP® tubes and ambr250® bioreactors was evaluated by a statistical analysis, including assessment of normal distribution of data, and ANOVA. Assessment of normal distribution was performed on studied ln-transformed response residuals and normality was demonstrated only for 24-deepwell plate dataset (data not shown). Comparison of mean process specific productivity between long cell line development process (MP-CL) and short cell line development process (DSCC) is shown on Figure 64. Mean process specific productivity in 24-deepwell plates is significantly lower with short cell line development process (DSCC) compared to long cell line development process (MP-CL) (p-value below $\alpha_{5\%}$). Normality was not demonstrated for TPP® tube and ambr250® bioreactor datasets; therefore, non-parametric assay was performed (Wilcoxon exact test). Mean process specific productivity in TPP® tubes and ambr250® bioreactors is significantly lower with short cell line development process (DSCC) compared to long cell line development process (MP-CL) (p-value below $\alpha_{5\%}$).

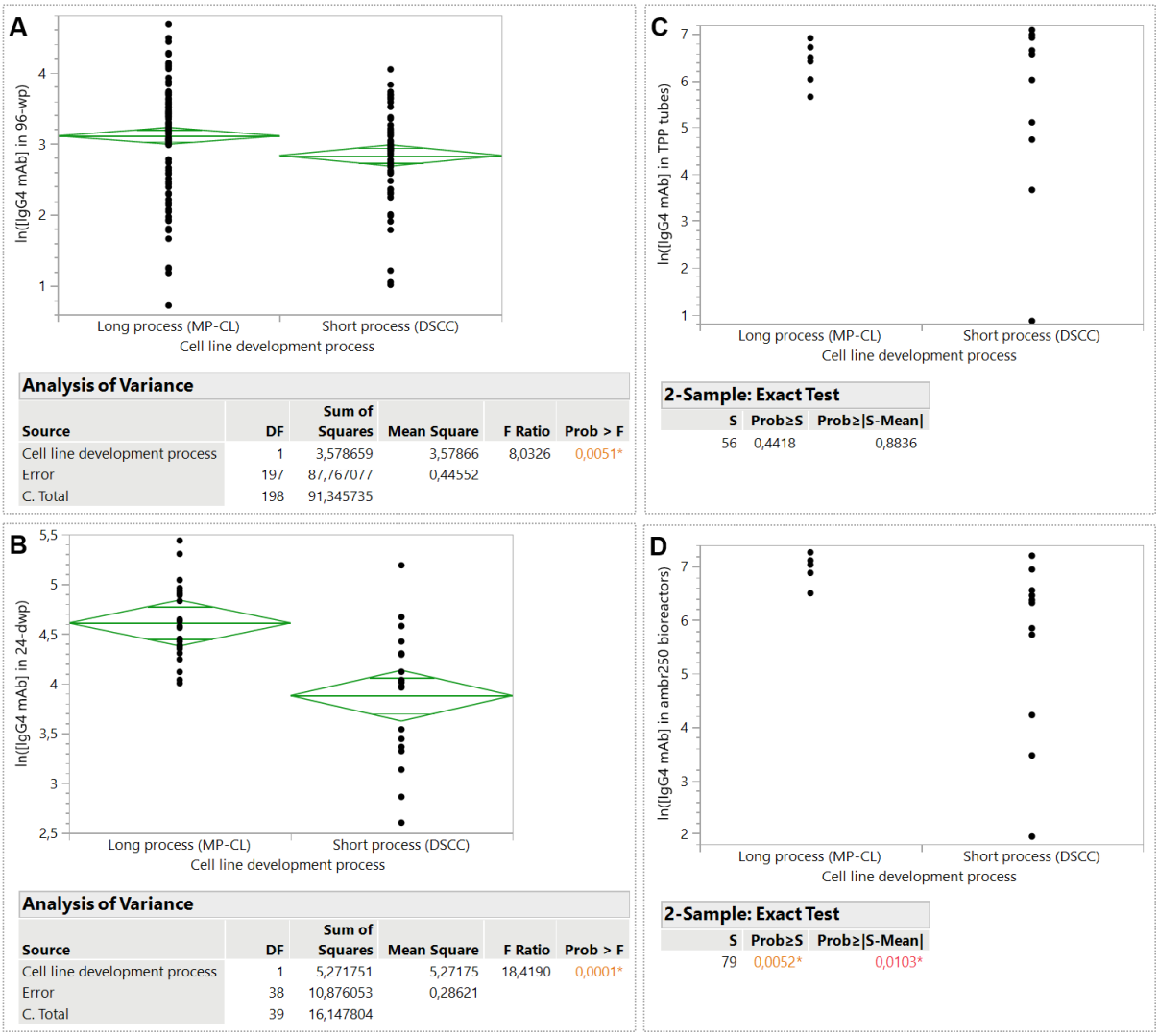
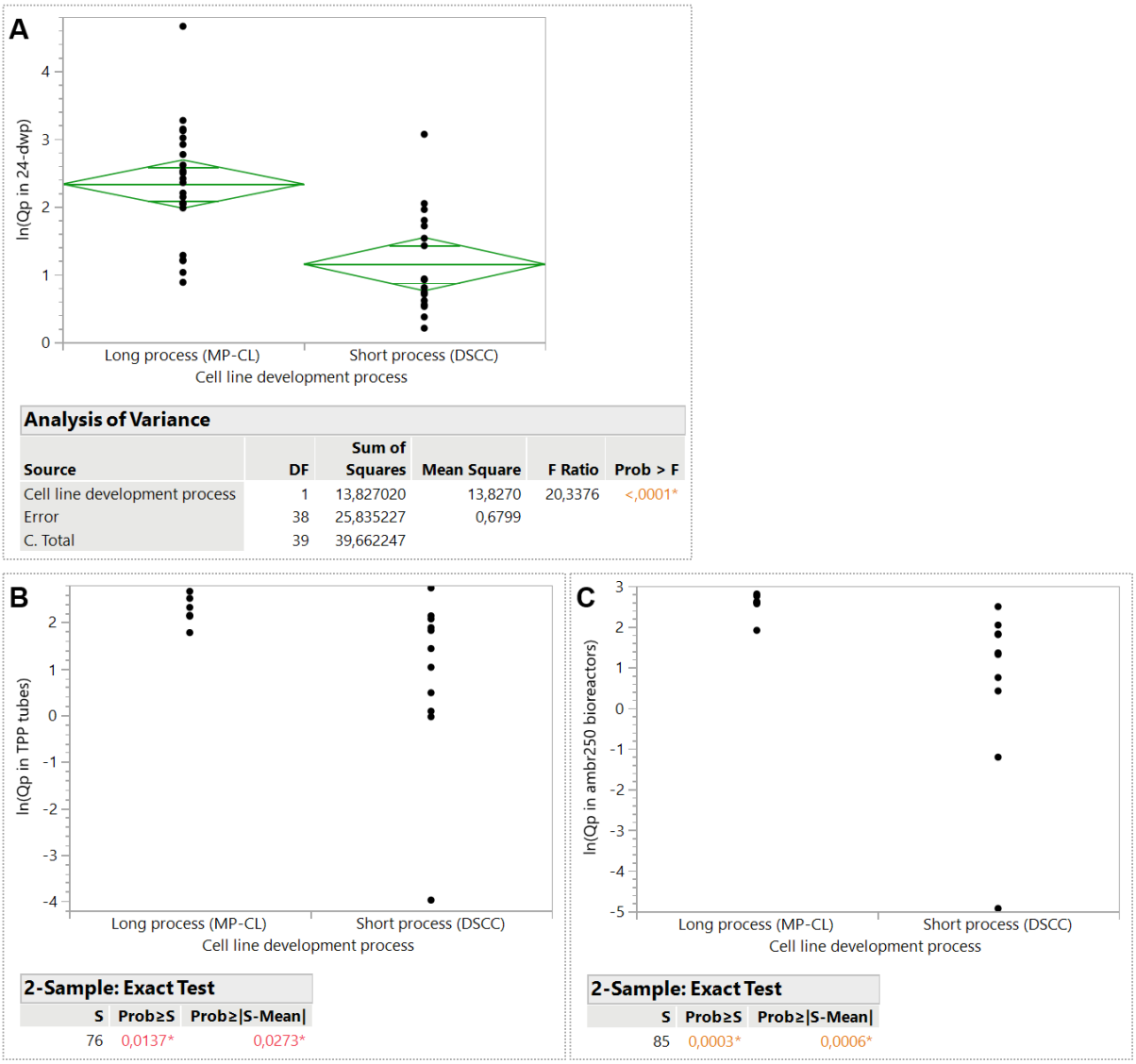


Figure 63: Statistical comparison of group mean IgG4 mAb concentration in clone's supernatants after productivity measurement in 96-well plates, and productivity assays in 24-deepwell plates, TPP® tubes and ambr250® bioreactors

(A), ***(B)*** *Analysis of variance results on mean ln-transformed IgG4 mAb concentration. Significance of mean titer difference between long cell line development process (MP-CL) and short cell line development process (DSCC) in 96-well plates (A) and 24-deepwell plates (B) was evaluated by a statistical analysis, including assessment of normal distribution and one-way analysis of variance (ANOVA). Green lines indicate the mean of each condition. X-axis size is not proportional to sample size.* ***(C)***, ***(D)*** *Analysis of variance results on mean ln-transformed IgG4 mAb concentration. Significance of mean titer difference between long cell line development process (MP-CL) and short cell line development process (DSCC) in TPP® tubes (C) and ambr250® bioreactors (D) was evaluated by a statistical analysis. Assumption of normality was not verified; therefore, non-parametric assay was performed (Wilcoxon exact test). X-axis size is not proportional to sample size. Data analysis was performed on JMP® 15 software.*



Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio	Prob > F
Cell line development process	1	13,827020	13,8270	20,3376	<,0001*
Error	38	25,835227	0,6799		
C. Total	39	39,662247			

2-Sample: Exact Test

S	Prob≥S	Prob≥|S-Mean|
76	0,0137*	0,0273*

2-Sample: Exact Test

S	Prob≥S	Prob≥|S-Mean|
85	0,0003*	0,0006*

Figure 64: Statistical comparison of group mean specific productivity after productivity assays in 24-deepwell plates, TPP® tubes and ambr250® bioreactors

(A) *Analysis of variance results on mean ln-transformed specific productivity. Significance of mean specific productivity difference between long cell line development process (MP-CL) and short cell line development process (DSCC) was evaluated by a statistical analysis, including assessment of normal distribution and one-way analysis of variance (ANOVA). Green lines indicate the mean of each condition. X-axis size is not proportional to sample size.* ***(B)****,* ***(C)*** *Analysis of variance results on mean ln-transformed specific productivity. Significance of mean specific productivity difference between long cell line development process (MP-CL) and short cell line development process (DSCC) in TPP® tubes (B) and ambr250® bioreactors (C) was evaluated by a statistical analysis. Assumption of normality was not verified; therefore, non-parametric assay was performed (Wilcoxon exact test). X-axis size is not proportional to sample size. Data analysis was performed on JMP® 15 software.*

Supplementary data on clone stability

Supplementary data related to 3.2.4.1.3 Impact of cell line development process on clone productivity is detailed hereinafter.

Table 32: IgG4 mAb titer variations between PDL_{20}, PDL_{40}, and PDL_{60} (clones generated with long or short CLD process)

Variations between PDL were calculated on the last common day of culture between the three productivity assays.

Clone ID	**PDL**	**Last production day (PD)**	**[IgG4] on last PD [mg/L]**	**Last common production day (PD)**	**[IgG4] on last common PD [mg/L]**	**Δ[IgG4] between PDL_{20}-PDL_{40}**	**Δ[IgG4] between PDL_{40}-PDL_{60}**	**Δ[IgG4] between PDL_{20}-PDL_{60}**	**Conclusion on expression level stability**
CL 15D08 (Long process)	PDL_{20}	PD_{12}	1247.3	PD_{9}	846.6	-4.3%		-32.4%	Unstable
	PDL_{40}	PD_{12}	1150	PD_{9}	809.9		-29.4%		
	PDL_{60}	PD_{9}	572.0	PD_{9}	572.0				
CL 16E06 (Long process)	PDL_{20}	PD_{9}	715.6	PD_{7}	381.5	9.4%		2.1%	**Stable**
	PDL_{40}	PD_{12}	1230.9	PD_{7}	417.3		-6.7%		
	PDL_{60}	PD_{7}	389.4	PD_{7}	389.4				
CL 38A06 (Short process)	PDL_{20}	PD_{11}	586.4	PD_{9}	346.4	-0.4%		-8.2%	**Stable**
	PDL_{40}	PD_{11}	672.1	PD_{9}	345.1		-7.8%		
	PDL_{60}	PD_{9}	318.1	PD_{9}	318.1				
CL 38C02 (Short process)	PDL_{20}	PD_{12}	1108.8	PD_{9}	522.9	2.0%		-35.6%	Unstable
	PDL_{40}	PD_{11}	1011.1	PD_{9}	533.6		-36.8%		
	PDL_{60}	PD_{9}	336.8	PD_{9}	336.8				

Table 33: Mean process Q_P variations between PDL_{20}, PDL_{40}, and PDL_{60} (clones generated with long or short CLD process)

Variations between PDL were calculated on the last common day of culture between the three productivity assays.

Clone ID	PDL	Last production day (PD)	Mean process Q_p on last PD [pg/cell/day]	Last common production day (PD)	Mean process Q_p on last common PD [pg/cell/day]	ΔQ_p between PDL_{20}-PDL_{40}	ΔQ_p between PDL_{40}-PDL_{60}	ΔQ_p between PDL_{20}-PDL_{60}	Conclusion on mean process Q_p stability
CL 15D08 (Long process)	PDL_{20}	PD_{12}	8.4	PD_9	8.9	-19.1%		-19.1	**Stable**
	PDL_{40}	PD_{12}	6.4	PD_9	7.2		0%		
	PDL_{60}	PD_9	7.2	PD_9	7.2				
CL 16E06 (Long process)	PDL_{20}	PD_9	14.5	PD_7	12.1	0.8%		8.3%	**Stable**
	PDL_{40}	PD_{12}	17.3	PD_7	12.2		7.3%		
	PDL_{60}	PD_7	13.1	PD_7	13.1				
CL 38A06 (Short process)	PDL_{20}	PD_{11}	5.5	PD_9	4.4	6.8%		18.2%	**Stable**
	PDL_{40}	PD_{11}	6.4	PD_9	4.7		10.6%		
	PDL_{60}	PD_9	5.2	PD_9	5.2				
CL 38C02 (Short process)	PDL_{20}	PD_{12}	9.0	PD_9	6.3	1.6%		-20.6%	**Stable**
	PDL_{40}	PD_{11}	8.5	PD_9	6.4		-21.9%		
	PDL_{60}	PD_9	5.0	PD_9	5.0				

Appendix 3: Supplementary data related to results chapter 3

Details on statistical analysis on promoters' transcriptional strength

Statistical analysis related to 3.3.5.1 Candidate promoters for *pac* expression have a lower transcriptional strength than SV40E wild type promoter is detailed hereinafter.

Significance of transcription level difference between candidate and SV40E wild type promoters was evaluated by a statistical analysis. Assessment of normal distribution of data is required to identify adequate statistical tests. Then one-way analysis of variance (ANOVA), Dunnett's and Tukey-Kramer tests were performed.

Analysis of experimental data distribution is shown on Figure 65-A. Analysis of experimental data distribution shows that data are not normally distributed. Indeed, visual inspection of quantile-quantile plot indicates a limited adequacy of the fit between data distribution and theoretical normal distribution model. In addition, kurtosis coefficient is higher than 2, indicating that the experimental data distribution is less flattened than a normal density. Therefore, experimental data were mathematically ln-transformed (natural logarithm). Analysis of ln-transformed experimental data distribution is shown on Figure 65-B. Analysis of ln-transformed experimental data distribution shows that data are normally distributed. Indeed, visual inspection of the probability density and quantile-quantile plots respectively shows that frequency histogram better fits to normal distribution curve and that there is a better adequation of the fit between data distribution and theoretical normal distribution. Lastly, calculated mean and median are in reasonable agreement, and skewness and kurtosis coefficient are comprised between ± 2. Parametric tests (one-way ANOVA and Dunnett's test) can therefore be used for statistical analysis on ln-transformed data.

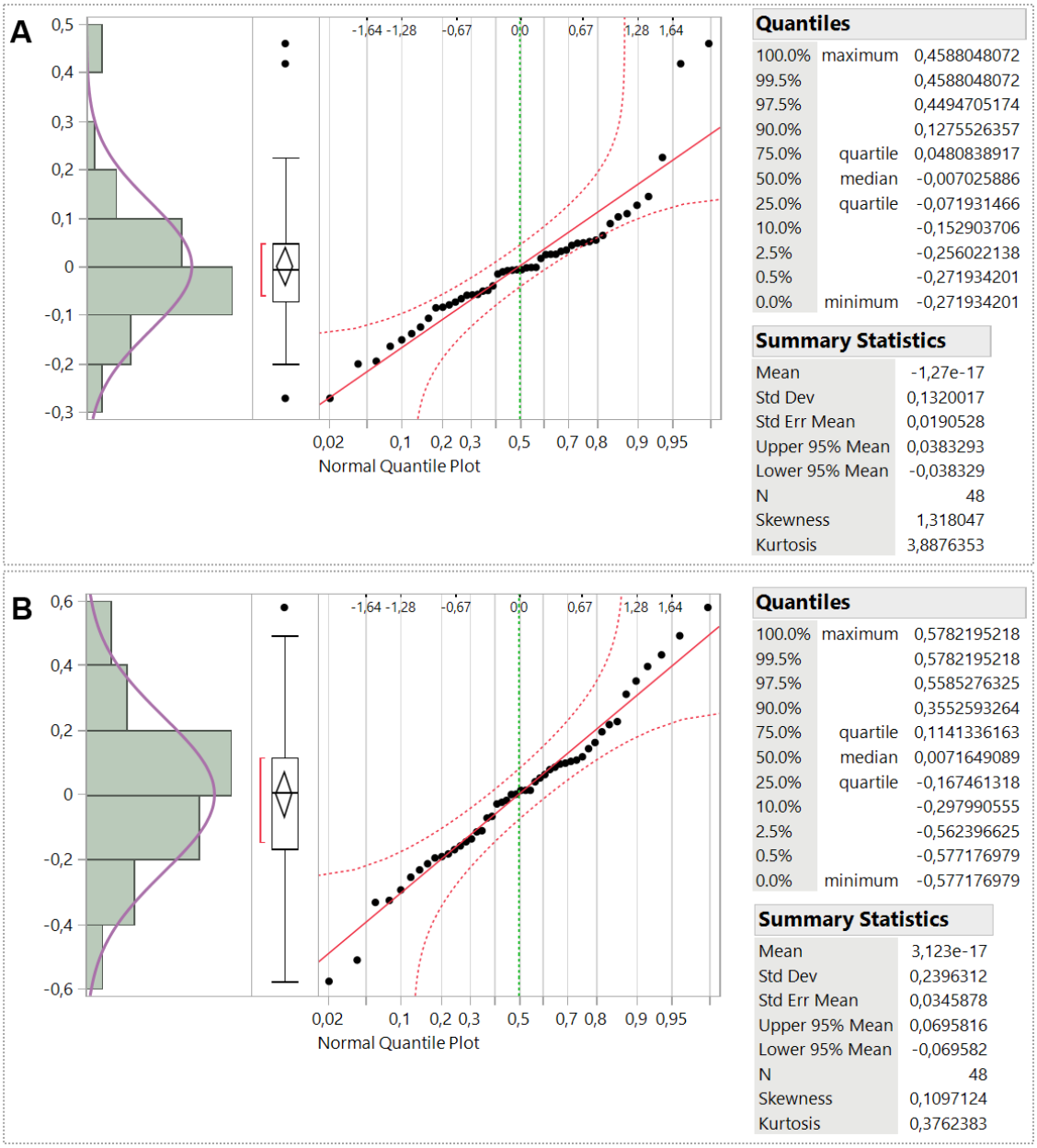


Figure 65: Assessment of normal distribution of RT-qPCR data

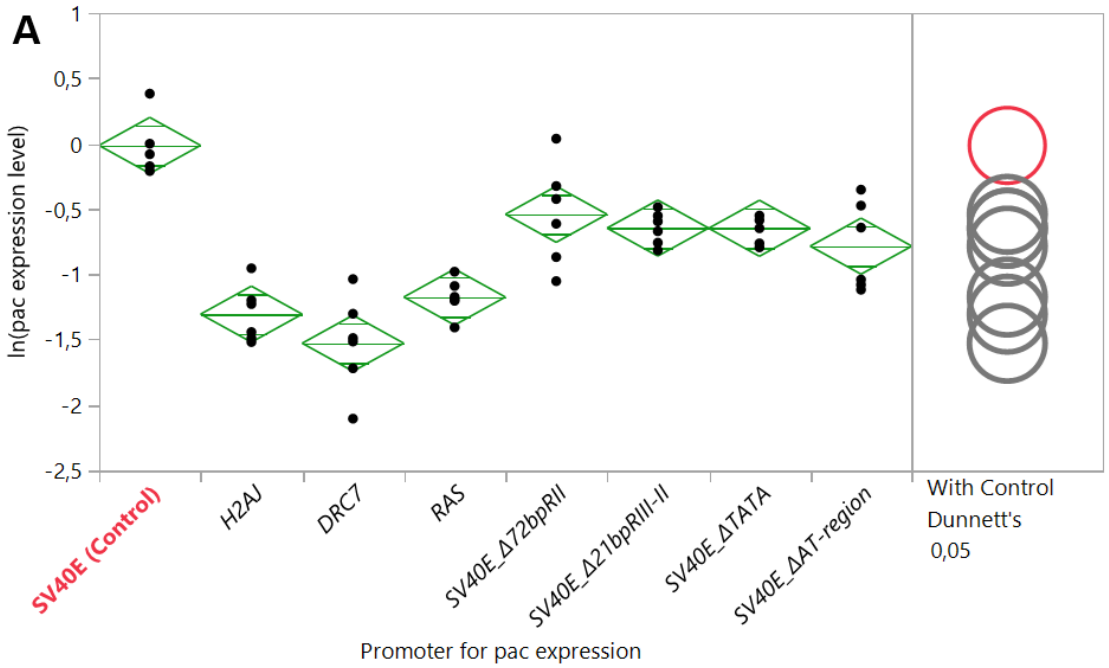
Assessment of normal distribution was performed on residuals, by visual inspection of the probability distribution plot, visual inspection of the quantile-quantile plot, verification of mean and median values and verification of skewness and kurtosis values. ***(A)*** *Assessment of normal distribution on residuals from pac transcription levels normalized with geometric mean of rituHC and rituLC relative quantities.* ***(B)*** *Assessment of normal distribution on residuals from natural logarithm (ln) of pac transcription levels normalized with geometric mean of rituHC and rituLC relative quantities. Data analysis was performed on JMP® 15 software.*

Statistical analyses are run to determine if (i) there is a significant statistical different mean transcription level between candidate and SV40E wild type promoters and if (ii) there is significant statistical different mean transcription between candidate promoters. A two test-based approach is chosen to answer those questions. A one-way ANOVA and a Dunnett's test are run to answer the first question, on the whole dataset of natural logarithm of *pac* transcription levels normalized with geometric mean of rituHC and

rituLC relative quantities. A one-way ANOVA and a Tukey-Kramer test are run to answer the second question, on the dataset restricted to natural logarithm of *pac* transcription levels normalized with geometric mean of rituHC and rituLC relative quantities for candidate promoters only. The Dunnett's test is indeed more relevant and powerful than Tukey-Kramer test for comparison of a set of means against the mean of a control group, because of reduced number of tests compared to all pairwise comparisons (Lee & Lee, 2018). Nonetheless, as the second question is also relevant, pairwise comparisons of means is performed with a Tukey-Kramer test, by deleting data from group (comparison to control having been made with Dunnett's test).

Comparison of mean transcription level between candidate and SV40E wild type promoters is shown on Figure 66-A. One-way analysis of variance indicates that at least one mean is different (p-value below $\alpha_{5\%}$). Analysis is further investigated with a Dunnett's test to know which candidate promoter has a significantly different mean *pac* transcription level than the SV40E wild type promoter. All candidate promoters show a significantly reduced mean transcriptional activity compared to SV40E wild type promoter (all p-values below $\alpha_{5\%}$ (in LSD threshold matrix)).

Comparison of mean transcription levels between all candidate promoters (mean transcription level of SV40E wild type being excluded) is shown on Figure 66-B. Univariate analysis of variance (ANOVA) indicates that at least one mean is different (p-value below $\alpha_{5\%}$). Analysis is further investigated with a Tukey-Kramer test to know which pairs of means are significantly different. Deleted SV40 promoters (SV40E_Δ72bpRII, SV40E_Δ21bpRIII-II, SV40E_ΔTATA, and SV40E_ΔAT-region) display similar promoter transcriptional activities (p-values above $\alpha_{5\%}$ (in ordered differences report)). The same observation is made for CHO K1 endogenous promoters (H2AJ, DRC7 and RAS). Interestingly, SV40E_ΔAT-region and RAS transcriptional activities are similar. All other comparisons (SV40E_Δ72bpRII /SV40E_Δ21bpRIII-II /SV40E_ΔTATA *vs* H2AJ /DRC7 /RAS, and SV40E_ΔAT-region *vs* H2AJ /DRC7) show that promoter have statistically different transcriptional activities.



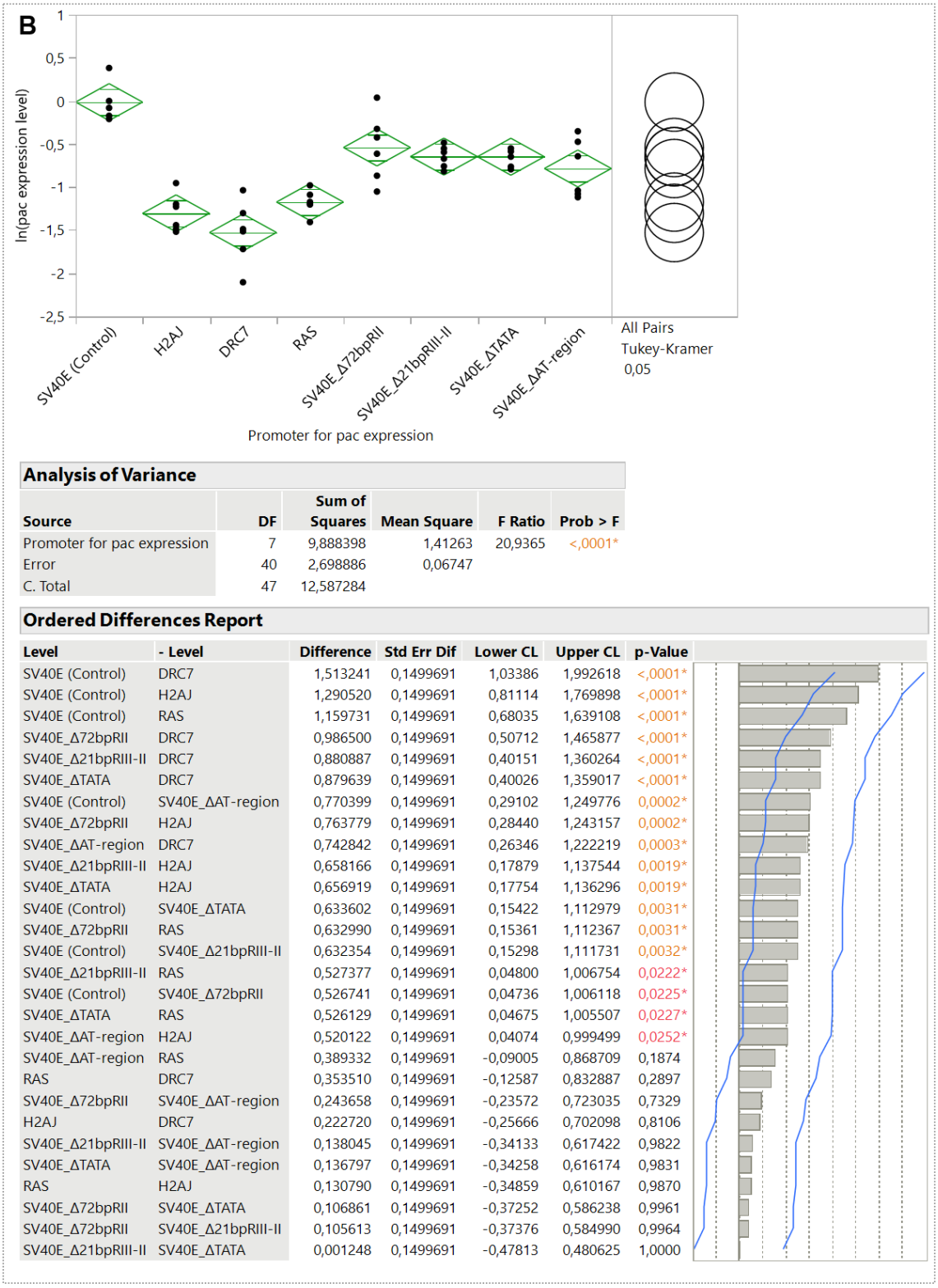
Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio	Prob > F
Promoter for pac expression	7	9,888398	1,41263	20,9365	<,0001*
Error	40	2,698886	0,06747		
C. Total	47	12,587284			

LSD Threshold Matrix

Level	Abs(Dif)-LSD	p-Value
SV40E (Control)	-0,41	1,0000
SV40E_Δ72bpRII	0,117	0,0067*
SV40E_Δ21bpRIII-II	0,223	0,0009*
SV40E_ΔTATA	0,224	0,0009*
SV40E_ΔAT-region	0,361	<,0001*
RAS	0,75	<,0001*
H2AJ	0,881	<,0001*
DRC7	1,104	<,0001*

Positive values show pairs of means that are significantly different.



Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio	Prob > F
Promoter for pac expression	7	9,888398	1,41263	20,9365	<,0001*
Error	40	2,698886	0,06747		
C. Total	47	12,587284			

Ordered Differences Report

Level	- Level	Difference	Std Err Dif	Lower CL	Upper CL	p-Value
SV40E (Control)	DRC7	1,513241	0,1499691	1,03386	1,992618	<,0001*
SV40E (Control)	H2AJ	1,290520	0,1499691	0,81114	1,769898	<,0001*
SV40E (Control)	RAS	1,159731	0,1499691	0,68035	1,639108	<,0001*
SV40E_Δ72bpRII	DRC7	0,986500	0,1499691	0,50712	1,465877	<,0001*
SV40E_Δ21bpRIII-II	DRC7	0,880887	0,1499691	0,40151	1,360264	<,0001*
SV40E_ΔTATA	DRC7	0,879639	0,1499691	0,40026	1,359017	<,0001*
SV40E (Control)	SV40E_ΔAT-region	0,770399	0,1499691	0,29102	1,249776	0,0002*
SV40E_Δ72bpRII	H2AJ	0,763779	0,1499691	0,28440	1,243157	0,0002*
SV40E_ΔAT-region	DRC7	0,742842	0,1499691	0,26346	1,222219	0,0003*
SV40E_Δ21bpRIII-II	H2AJ	0,658166	0,1499691	0,17879	1,137544	0,0019*
SV40E_ΔTATA	H2AJ	0,656919	0,1499691	0,17754	1,136296	0,0019*
SV40E (Control)	SV40E_ΔTATA	0,633602	0,1499691	0,15422	1,112979	0,0031*
SV40E_Δ72bpRII	RAS	0,632990	0,1499691	0,15361	1,112367	0,0031*
SV40E (Control)	SV40E_Δ21bpRIII-II	0,632354	0,1499691	0,15298	1,111731	0,0032*
SV40E_Δ21bpRIII-II	RAS	0,527377	0,1499691	0,04800	1,006754	0,0222*
SV40E (Control)	SV40E_Δ72bpRII	0,526741	0,1499691	0,04736	1,006118	0,0225*
SV40E_ΔTATA	RAS	0,526129	0,1499691	0,04675	1,005507	0,0227*
SV40E_ΔAT-region	H2AJ	0,520122	0,1499691	0,04074	0,999499	0,0252*
SV40E_ΔAT-region	RAS	0,389332	0,1499691	-0,09005	0,868709	0,1874
RAS	DRC7	0,353510	0,1499691	-0,12587	0,832887	0,2897
SV40E_Δ72bpRII	SV40E_ΔAT-region	0,243658	0,1499691	-0,23572	0,723035	0,7329
H2AJ	DRC7	0,222720	0,1499691	-0,25666	0,702098	0,8106
SV40E_Δ21bpRIII-II	SV40E_ΔAT-region	0,138045	0,1499691	-0,34133	0,617422	0,9822
SV40E_ΔTATA	SV40E_ΔAT-region	0,136797	0,1499691	-0,34258	0,616174	0,9831
RAS	H2AJ	0,130790	0,1499691	-0,34859	0,610167	0,9870
SV40E_Δ72bpRII	SV40E_ΔTATA	0,106861	0,1499691	-0,37252	0,586238	0,9961
SV40E_Δ72bpRII	SV40E_Δ21bpRIII-II	0,105613	0,1499691	-0,37376	0,584990	0,9964
SV40E_Δ21bpRIII-II	SV40E_ΔTATA	0,001248	0,1499691	-0,47813	0,480625	1,0000

Figure 66: Statistical analysis on promoters' transcriptional strength

(A) *Analysis of variance and Dunnett's test results on natural logarithm of pac transcription levels normalized to geometrical mean of rituHC and rituLC relative quantities. Significance of transcription level difference between candidate and wild type SV40 promoters was evaluated by a statistical analysis, including one-way analysis of variance (ANOVA) and Dunnett's test (α 5%).* ***(B)*** *Analysis of variance and Tukey-Kramer test results on natural logarithm of pac transcription levels normalized to geometrical mean of rituHC and rituLC relative quantities. Significance of transcription level difference*

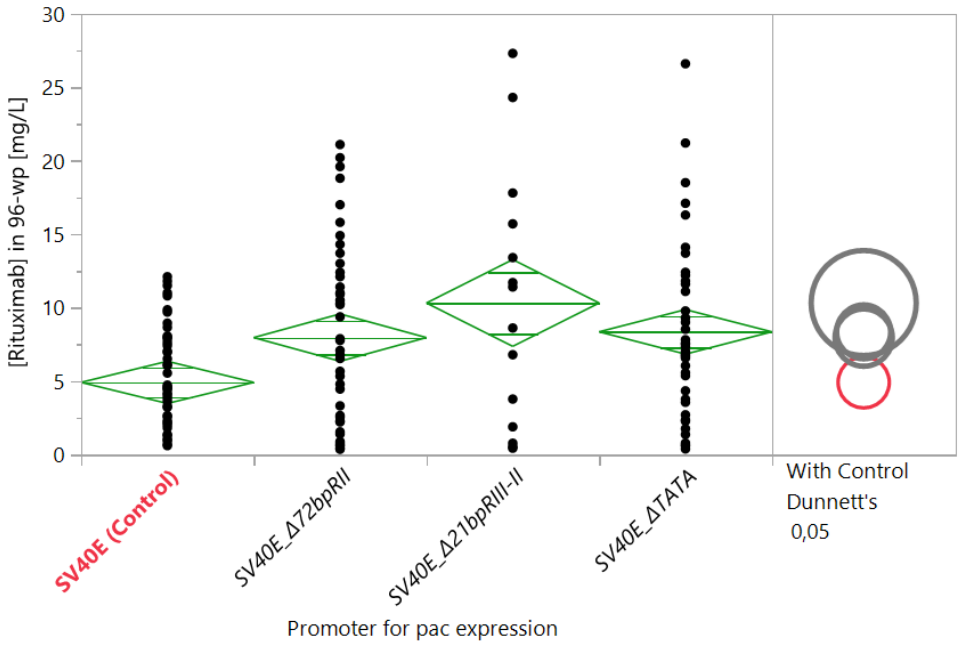
between candidate promoters (SV40E wild type promoter excluded) was evaluated by a statistical analysis, including one-way analysis of variance (ANOVA) and Tukey-Kramer test (α 5%). Data analysis was performed on JMP® 15 software.

Details on statistical analysis on mini-pools' productivity

Statistical analysis related to 3.3.5.2.2 Impact of promoter strength on mini-pool productivity is detailed hereinafter.

Significance of mean titer difference between candidate and SV40E wild type promoters was evaluated by a statistical analysis, including assessment of normal distribution of data, one-way analysis of variance (ANOVA), and Dunnett's test. It should be noticed that analysis was restricted to SV40E wild type, SV40E_Δ72bpRII, SV40E_Δ21bpRIII-II, and SV40E_ΔTATA promoters, because of too small sample size for other conditions (RAS and SV40E_ΔAT-region, due to highly decreased mini-pool reemergence for both conditions). Furthermore, at this stage of the evaluation of the various candidate promoters for *pac* expression, pairwise comparison of group means through a Tukey-Kramer test is not necessary. Therefore, analysis was limited to a Dunnett's test.

Assessment of normal distribution was performed on studied response residuals (remaining variations once potential group-related effects have been removed) and normality was demonstrated (data not shown, as an example of normal distribution verification has previously been detailed). Comparison of mean titer between candidate and SV40E wild type promoters is shown on Figure 67. Univariate analysis of variance (ANOVA) indicates that at least one mean is different (p-value below $\alpha_{5\%}$). Analysis is further investigated with a Dunnett's test to know which candidate promoter has a significantly different mean titer than SV40E wild type promoter. All candidate promoters show a significantly increased mean titer compared to SV40E wild type promoter (all p-values below $\alpha_{5\%}$ (in LSD threshold matrix)).



Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio	Prob > F
Promoter for pac expression	3	542,1653	180,722	5,7737	0,0009*
Error	169	5289,8343	31,301		
C. Total	172	5831,9996			

LSD Threshold Matrix

Level	Abs(Dif)-LSD	p-Value
SV40E_Δ21bpRIII-II	1,428	0,0041*
SV40E_ΔTATA	0,909	0,0041*
SV40E_Δ72bpRII	0,433	0,0171*
SV40E (Control)	-2,46	1,0000

Positive values show pairs of means that are significantly different.

Figure 67: Statistical analysis on rituximab concentration in mini-pools' supernatants after batch productivity assay in 96-well plate

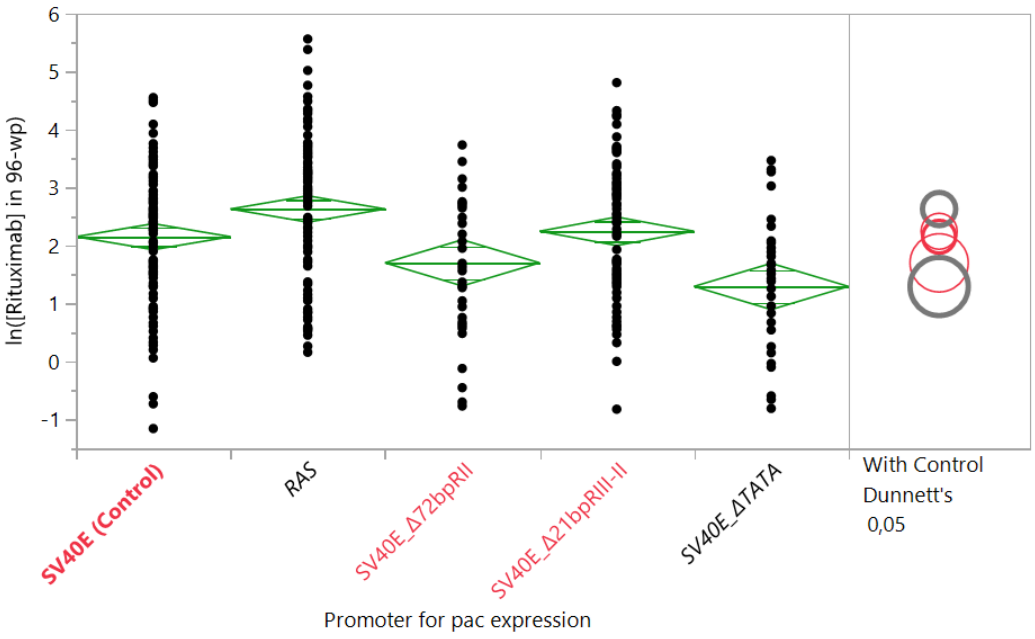
Analysis of variance and Dunnett's test results on rituximab concentration. Normality of data distribution was previously assessed (data not shown). Significance of mean titer difference between candidate and wild type SV40 promoters was evaluated by a statistical analysis, including one-way analysis of variance (ANOVA) and Dunnett's test (α 5%). This analysis was limited to conditions having more than 10 samples (exclusion of RAS and SV40E_ΔAT-region conditions). Green lines indicate the mean of each condition. X-axis size is not proportional to sample size (SV40E (Control): 59 MP, SV40E_Δ72bpRII: 47 MP, SV40E_Δ21bpRIII-II: 14 MP, SV40E_ΔTATA: 53 MP). Data analysis was performed on JMP® 15 software.

Details on statistical analyses on clones' productivity

Clone productivity (titer) in 96-well plates

Statistical analysis related to 3.3.5.4.2.1 Clone productivity (titer) in 96-well plates is detailed hereinafter.

Significance of mean titer difference between candidate and SV40E wild type promoters was evaluated by a statistical analysis, including assessment of normal distribution of data, ANOVA, and Dunnett's test. Assessment of normal distribution was performed on studied ln-transformed response residuals and normality was demonstrated (data not shown). Comparison of mean titer between candidate and SV40E wild type promoters is shown on Figure 68. ANOVA indicates that at least one mean is different (p-value below $\alpha_{5\%}$). Analysis is further investigated with a Dunnett's test to know which candidate promoter has a significantly different mean titer than SV40E wild type promoter. RAS promoter only shows a significantly increased mean titer compared to SV40E wild type promoter (p-value below $\alpha_{5\%}$).



Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio	Prob > F
Promoter for pac expression	4	56,00996	14,0025	10,0309	<,0001*
Error	363	506,72621	1,3959		
C. Total	367	562,73617			

LSD Threshold Matrix

Level	Abs(Dif)-LSD	p-Value
RAS	0,075	0,0134*
SV40E_Δ21bpRIII-II	-0,33	0,9600
SV40E (Control)	-0,4	1,0000
SV40E_Δ72bpRII	-0,13	0,1796
SV40E_ΔTATA	0,281	0,0010*

Positive values show pairs of means that are significantly different.

Figure 68: Rituximab concentration in clone's supernatants after productivity measurement in 96-well plate and statistical comparison of group means

Analysis of variance and Dunnett's test results on ln-transformed rituximab concentration. Normality of data distribution was previously assessed (data not shown). Significance of mean titer difference between candidate and wild type SV40 promoters was evaluated by a statistical analysis, including one-way analysis of variance (ANOVA) and Dunnett's test (α 5%). Green lines indicate the mean of each condition. X-axis size is not proportional to sample size. Data analysis was performed on JMP® 15 software.

Clone productivity (titer and specific productivity) in 24-deepwell plates

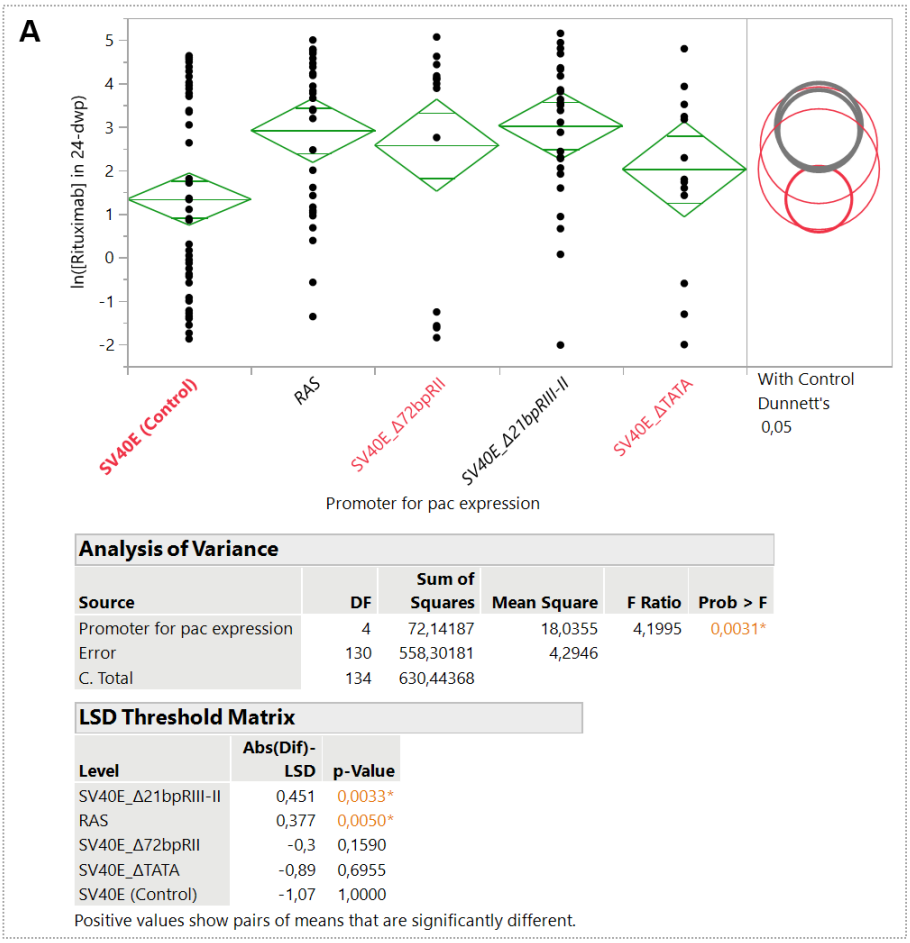
Statistical analysis related to 3.3.5.4.2.2 Clone productivity (titer and specific productivity) in 24-deepwell plates is detailed hereinafter.

Significance of mean titer and specific productivity differences in 24-deepwell plates between candidate and SV40E wild type promoters was evaluated by a statistical analysis, including assessment of normal distribution of data, ANOVA, and Dunnett's test. Assessment of normal distribution was performed on

studied ln-transformed response residuals and normality was demonstrated (data not shown). Comparison of mean titer between candidate and SV40E wild type promoters is shown on Figure 69-A. ANOVA indicates that at least one mean is different (p-value below $\alpha_{5\%}$). Analysis is further investigated with a Dunnett's test to know which candidate promoter has a significantly different mean titer than SV40E wild type promoter. RAS and SV40E_Δ21bprIII-II promoters show a significantly increased mean titer compared to SV40E wild type promoter (p-value below $\alpha_{5\%}$).

The analysis is repeated by removing the control group, therefore focusing only on candidate promoters. Comparison of mean titers between all candidate promoters (mean titer of SV40E wild type being excluded) is shown on Figure 69-B. ANOVA indicates that there is no statistical difference on mean titer (p-value below $\alpha_{5\%}$) between candidate promoters.

The same methodology of statistical analysis is performed on specific productivity data. Comparison of mean specific productivities between candidate and SV40E wild type promoters is shown on Figure 72-A and comparison of mean specific productivities between candidate (mean specific productivity of SV40E wild type being excluded) is shown on Figure 72-B. RAS and SV40E_Δ21bprIII-II promoters show a significantly increased mean specific productivity compared to SV40E wild type promoter (p-value below $\alpha_{5\%}$). However, there is no statistical difference on mean specific productivity (p-value below $\alpha_{5\%}$) between candidate promoters.



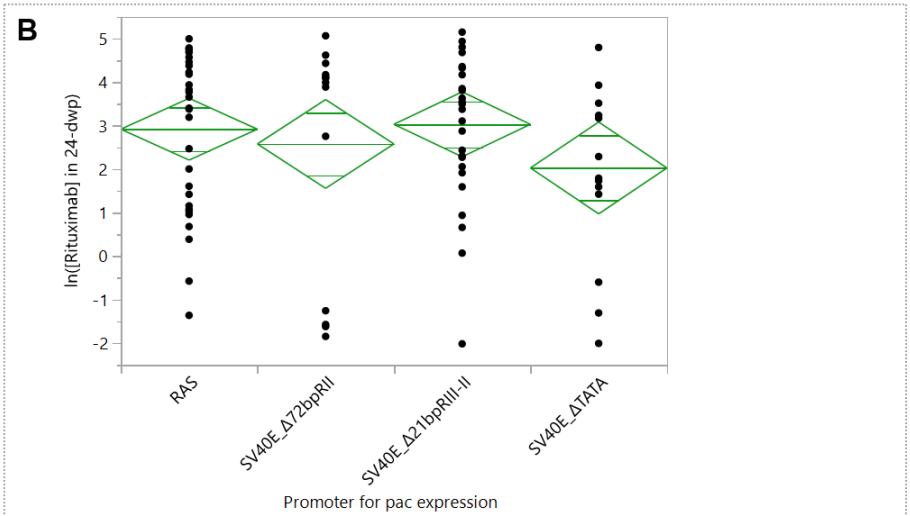
Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio	Prob > F
Promoter for pac expression	4	72,14187	18,0355	4,1995	0,0031*
Error	130	558,30181	4,2946		
C. Total	134	630,44368			

LSD Threshold Matrix

Level	Abs(Dif)-LSD	p-Value
SV40E_Δ21bpRIII-II	0,451	0,0033*
RAS	0,377	0,0050*
SV40E_Δ72bpRII	-0,3	0,1590
SV40E_ΔTATA	-0,89	0,6955
SV40E (Control)	-1,07	1,0000

Positive values show pairs of means that are significantly different.



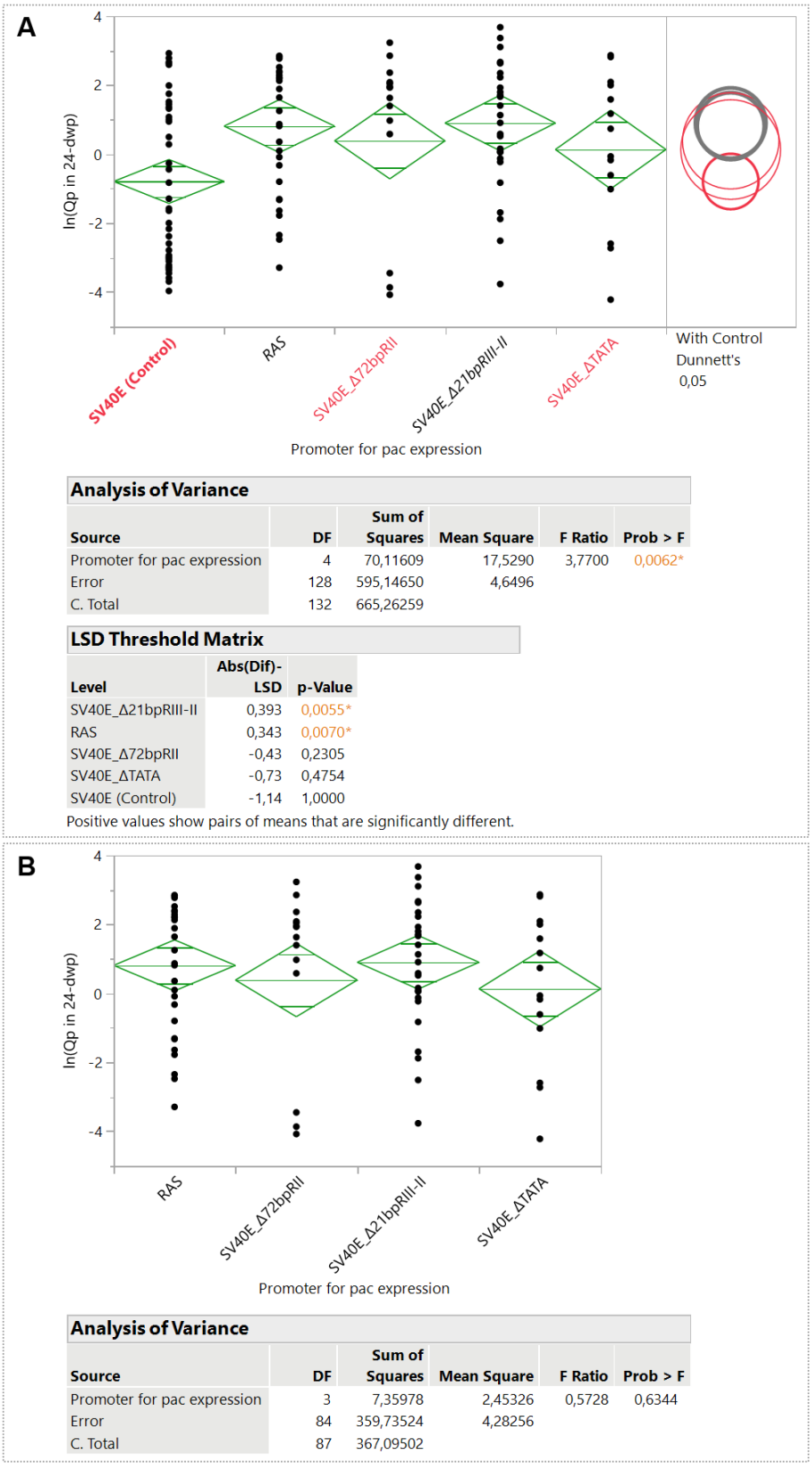
Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio	Prob > F
Promoter for pac expression	3	10,82061	3,60687	0,9135	0,4381
Error	84	331,66067	3,94834		
C. Total	87	342,48129			

Figure 69: Statistical comparison of group mean rituximab concentration in clone's supernatants after batch productivity assay in 24-deepwell plates

(A) *Analysis of variance and Dunnett's test results on mean ln-transformed rituximab concentration. Normality of data distribution was previously assessed (data not shown). Significance of mean titer*

difference between candidate and wild type SV40 promoters was evaluated by a statistical analysis, including one-way analysis of variance (ANOVA) and Dunnett's test (α 5%). Green lines indicate the mean of each condition. X-axis size is not proportional to sample size. ***(B)*** *Analysis of variance and Tukey-Kramer test results on mean ln-transformed rituximab concentration, for candidate promoters only (data from SV40E wild type being excluded). Normality of data distribution was previously assessed (data not shown). Significance of mean titer difference between all candidate promoters was evaluated by a statistical analysis, including one-way analysis of variance (ANOVA) and Tukey-Kramer test (α 5%). Green lines indicate the mean of each condition. X-axis size is not proportional to sample size. Data analysis was performed on JMP® 15 software.*



Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio	Prob > F
Promoter for pac expression	4	70,11609	17,5290	3,7700	0,0062*
Error	128	595,14650	4,6496		
C. Total	132	665,26259			

LSD Threshold Matrix

Level	Abs(Dif)-LSD	p-Value
SV40E_Δ21bpRIII-II	0,393	0,0055*
RAS	0,343	0,0070*
SV40E_Δ72bpRII	-0,43	0,2305
SV40E_ΔTATA	-0,73	0,4754
SV40E (Control)	-1,14	1,0000

Positive values show pairs of means that are significantly different.

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio	Prob > F
Promoter for pac expression	3	7,35978	2,45326	0,5728	0,6344
Error	84	359,73524	4,28256		
C. Total	87	367,09502			

Figure 70: Statistical comparison of group mean specific productivities after batch productivity assay in 24-deepwell plates

(A) *Analysis of variance results on mean ln-transformed Q_P. Normality of data distribution was previously assessed (data not shown). Significance of mean Q_P difference between candidate and wild*

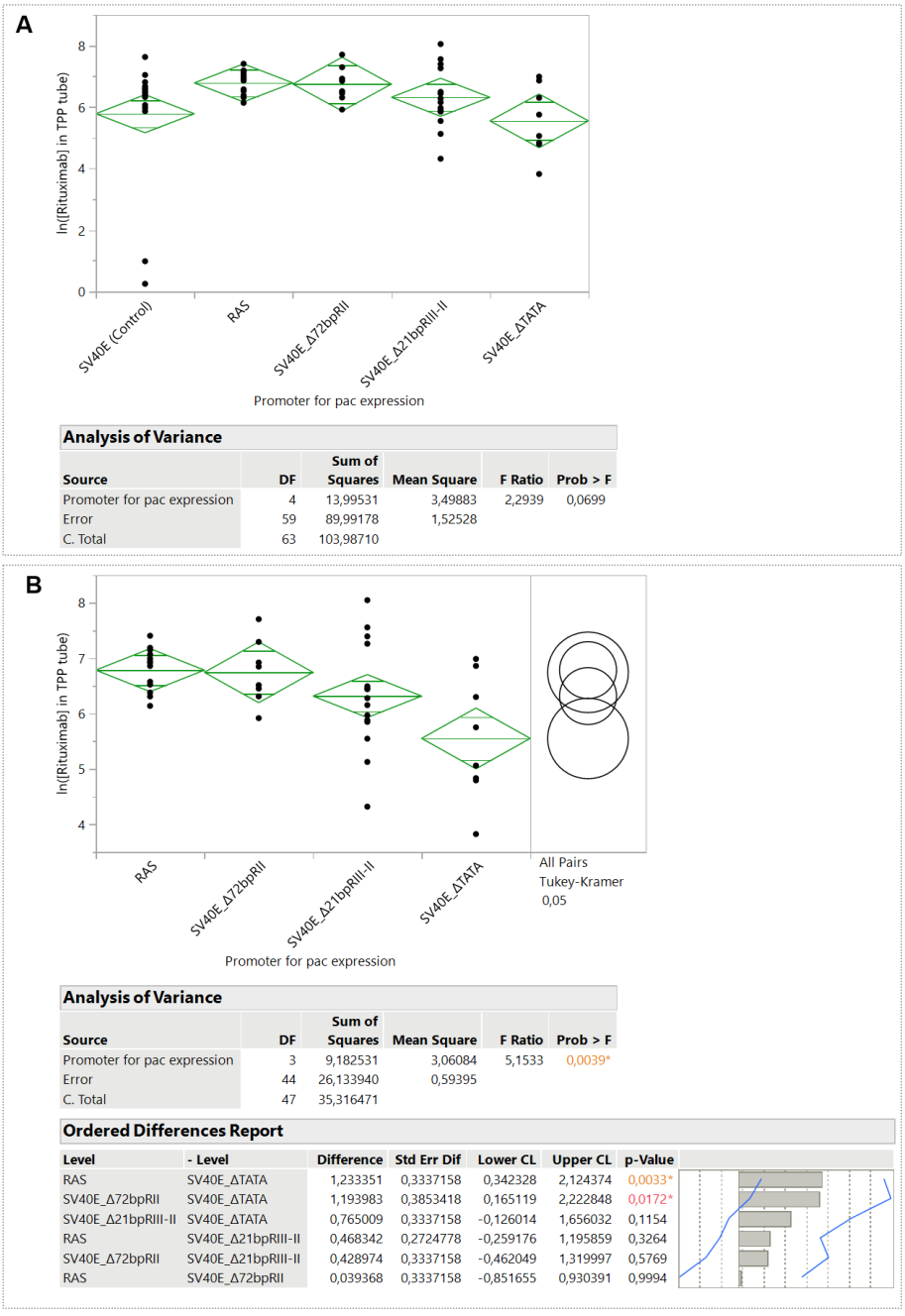
type SV40 promoters was evaluated by a statistical analysis, including one-way analysis of variance (ANOVA) (α 5%). Green lines indicate the mean of each condition. X-axis size is not proportional to sample size. ***(B)*** *Analysis of variance results on mean ln-transformed Q_P, for candidate promoters only (data from SV40E wild type being excluded). Normality of data distribution was previously assessed (data not shown). Significance of mean Q_P difference between all candidate promoters was evaluated by a statistical analysis, including one-way analysis of variance (ANOVA) (α 5%). Green lines indicate the mean of each condition. X-axis size is not proportional to sample size. Data analysis was performed on JMP® 15 software.*

Clone productivity (titer and specific productivity) in TPP® tubes

Statistical analysis related to 3.3.5.4.2.3 Clone productivity (titer and specific productivity) in TPP® tubes is detailed hereinafter.

Significance of mean titer and specific productivity differences in TPP® tubes between candidate and SV40E wild type promoters was evaluated by a statistical analysis, including assessment of normal distribution of data, ANOVA, and Dunnett's test. Assessment of normal distribution was performed on studied ln-transformed response residuals and normality was demonstrated (data not shown). Comparison of mean titer between candidate and SV40E wild type promoters is shown on Figure 71-A. ANOVA indicates that there is no statistical difference on mean titer (p-value above $\alpha_{5\%}$). The analysis is repeated by removing the control group, therefore focusing only on candidate promoters. Comparison of mean titers between all candidate promoters (mean titer of SV40E wild type being excluded) is shown on Figure 71-B. ANOVA indicates that at least one mean is different (p-value below $\alpha_{5\%}$). Analysis is further investigated with a Tukey-Kramer test to know which candidate promoter pairs have a significantly different mean titer. SV40E_ΔTATA shows a significantly lower mean titer than RAS and SV40E_Δ72bpRII conditions.

The same methodology of statistical analysis is performed on specific productivity data. Comparison of mean specific productivities between candidate and SV40E wild type promoters is shown on Figure 72-A and comparison of mean titer between candidate (mean specific productivity of SV40E wild type being excluded) is shown on Figure 72-B. However, analysis is limited to ANOVA, as the two ANOVAs performed indicate that there is no statistical difference on mean specific productivity (p-value above $\alpha_{5\%}$).



Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio	Prob > F
Promoter for pac expression	4	13,99531	3,49883	2,2939	0,0699
Error	59	89,99178	1,52528		
C. Total	63	103,98710			

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio	Prob > F
Promoter for pac expression	3	9,182531	3,06084	5,1533	0,0039*
Error	44	26,133940	0,59395		
C. Total	47	35,316471			

Ordered Differences Report

Level	- Level	Difference	Std Err Dif	Lower CL	Upper CL	p-Value
RAS	SV40E_ΔTATA	1,233351	0,3337158	0,342328	2,124374	0,0033*
SV40E_Δ72bpRII	SV40E_ΔTATA	1,193983	0,3853418	0,165119	2,222848	0,0172*
SV40E_Δ21bpRIII-II	SV40E_ΔTATA	0,765009	0,3337158	-0,126014	1,656032	0,1154
RAS	SV40E_Δ21bpRIII-II	0,468342	0,2724778	-0,259176	1,195859	0,3264
SV40E_Δ72bpRII	SV40E_Δ21bpRIII-II	0,428974	0,3337158	-0,462049	1,319997	0,5769
RAS	SV40E_Δ72bpRII	0,039368	0,3337158	-0,851655	0,930391	0,9994

Figure 71 : Statistical comparison of group mean rituximab concentration in clone's supernatants after fed-batch productivity assay in TPP® tubes

(A) *Analysis of variance and Dunnett's test results on mean ln-transformed rituximab concentration. Normality of data distribution was previously assessed (data not shown). Significance of mean titer difference between candidate and wild type SV40 promoters was evaluated by a statistical analysis, including one-way analysis of variance (ANOVA) and Dunnett's test (α 5%). Green lines indicate the mean of each condition. X-axis size is not proportional to sample size.* ***(B)*** *Analysis of variance and Tukey-Kramer test results on mean ln-transformed rituximab concentration, for candidate promoters only (data from SV40E wild type being excluded). Normality of data distribution was previously assessed (data not shown). Significance of mean titer difference between all candidate promoters was evaluated by a statistical analysis, including one-way analysis of variance (ANOVA) and Tukey-Kramer test (α 5%). Green lines indicate the mean of each condition. X-axis size is not proportional to sample size. Data analysis was performed on JMP® 15 software.*

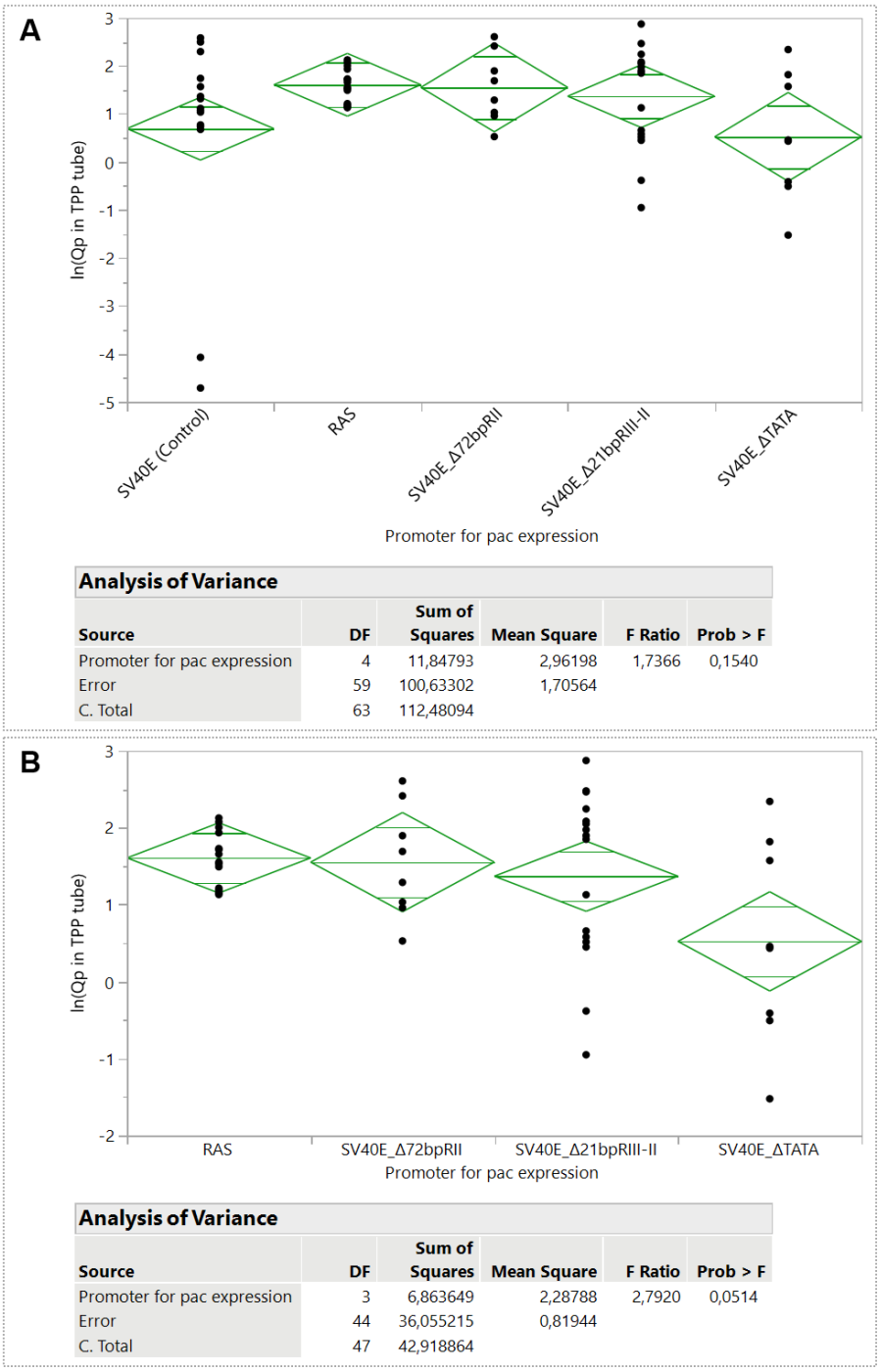


Figure 72 : Statistical comparison of group mean specific productivities after fed-batch productivity assay in TPP® tubes

(A) *Analysis of variance results on mean ln-transformed Q_P. Normality of data distribution was previously assessed (data not shown). Significance of mean Q_P difference between candidate and wild type SV40 promoters was evaluated by a statistical analysis, including one-way analysis of variance (ANOVA) (α 5%). Green lines indicate the mean of each condition. X-axis size is not proportional to sample size.* ***(B)*** *Analysis of variance results on mean ln-transformed Q_P, for candidate promoters only (data from SV40E wild type being excluded). Normality of data distribution was previously assessed (data not shown). Significance of mean Q_P difference between all candidate promoters was evaluated by a statistical analysis, including one-way analysis of variance (ANOVA) (α 5%). Green lines indicate the mean of each condition. X-axis size is not proportional to sample size. Data analysis was performed on JMP® 15 software.*

Supplementary data and details on statistical analyses on clone stability

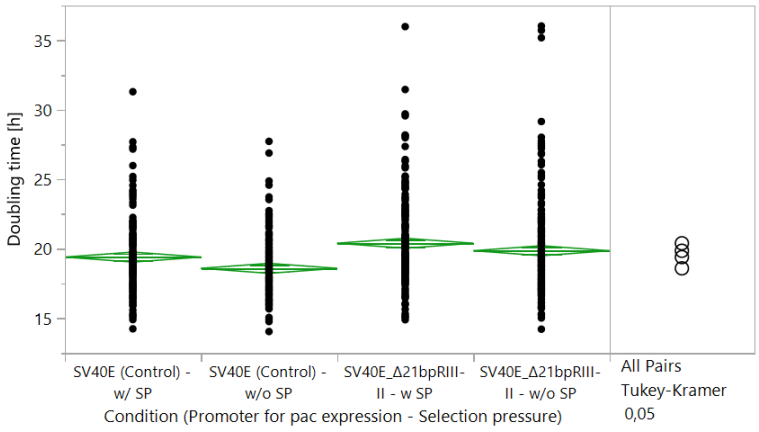
Supplementary data and statistical analysis related to 3.3.5.5 Expression level stability over 60 PDL is similar between SV40 and SV40E_Δ21bpRIII-II promoters is detailed hereinafter.

Effects of long-term cultivation on growth

Statistical analysis related to 3.3.5.5.1 Effects of long-term cultivation on growth is detailed hereinafter.

Significance of mean doubling time difference was evaluated by a statistical analysis, including assessment of normal distribution of data, ANOVA, and Tukey-Kramer test. Assessment of normal distribution was performed on ln-transformed studied response residuals and normality was demonstrated (data not shown). Comparison of mean doubling time differences is shown on Figure 73. One-way analysis of variance indicates that at least one mean is different (p-value below $\alpha_{5\%}$). Analysis is further investigated with a Tukey-Kramer test for pairwise comparisons of means. This comparison indicates that mean doubling time of clones cultured with selection pressure is significantly higher than mean doubling time of clones cultured without selection pressure (p-value below $\alpha_{5\%}$ (in Ordered Differences Report)). Besides, it indicates that mean doubling time of clones generated with SV40E_Δ21bpRIII-II promoter is significantly higher than mean doubling time of clones generated with SV40E wild type promoter (p-value below $\alpha_{5\%}$ (in Ordered Differences Report)). Nonetheless, doubling

time of clones generated with SV40E_Δ21bpRIII-II promoter ranges from 18 to 22 hours, which is perfectly suitable for further process development steps and industrial scale-up.



Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio	Prob > F
Condition (Promoter for pac expression - Selection pressure)	3	392,1072	130,702	17,4135	<,0001*
Error	892	6695,1984	7,506		
C. Total	895	7087,3056			

Ordered Differences Report

Level	- Level	Difference	Std Err Dif	Lower CL	Upper CL	p-Value
SV40E_Δ21bpRIII-II - w SP	SV40E (Control) - w/o SP	1,804121	0,2588751	1,13779	2,470449	<,0001*
SV40E_Δ21bpRIII-II - w/o SP	SV40E (Control) - w/o SP	1,266238	0,2588751	0,59991	1,932567	<,0001*
SV40E_Δ21bpRIII-II - w SP	SV40E (Control) - w/ SP	0,995566	0,2588751	0,32924	1,661894	0,0007*
SV40E (Control) - w/ SP	SV40E (Control) - w/o SP	0,808555	0,2588751	0,14223	1,474883	0,0100*
SV40E_Δ21bpRIII-II - w SP	SV40E_Δ21bpRIII-II - w/o SP	0,537882	0,2588751	-0,12845	1,204211	0,1612
SV40E_Δ21bpRIII-II - w/o SP	SV40E (Control) - w/ SP	0,457684	0,2588751	-0,20864	1,124012	0,2895

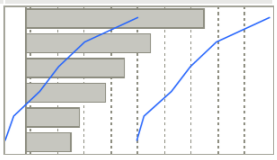


Figure 73: Statistical comparison of group mean doubling time

Analysis of variance and Tukey Kramer test results on doubling time. Significance of mean doubling time difference between promoters (SV40E wild type vs SV40E_Δ21bpRIII II) and presence/absence of selection pressure (SP) was evaluated by a statistical analysis, including assessment of normal distribution, one way analysis of variance (ANOVA) and Tukey Kramer test (α 5%). Green lines indicate the mean of each condition. X-axis size is not proportional to sample size. SP: selection pressure; w/: with; w/o: without. Data analysis was performed on JMP® 15 software.

Effects of long-term cultivation on productivity

Supplementary data and statistical analysis related to **Erreur ! Source du renvoi introuvable. Erreur ! Source du renvoi introuvable.** is detailed hereinafter.

Rituximab titer variations between PDL_{20}, PDL_{40}, and PDL_{60} are detailed in Table 34, Table 35, Table 36, and Table 37. Mean process specific productivity variations between PDL_{20}, PDL_{40}, and PDL_{60} are detailed in Table 38, Table 39, Table 40, and Table 41.

Analysis of clone stability, irrespectively of the promoter used for *pac* expression

Supplementary data and statistical analysis related to 3.3.5.5.2.1 Analysis of clone stability, irrespectively of the promoter used for *pac* expression is detailed hereinafter.

Significance of mean titer variability and mean specific productivity variability differences between condition with puromycin and condition without puromycin during sub-culturing (irrespective of the promoter used for *pac* expression) was evaluated by a statistical analysis, including assessment of normal distribution of data and ANOVA. Assessment of normal distribution was performed on studied response residuals and normality was demonstrated (data not shown). Comparison of mean titer variability between conditions with and without selection pressure during sub-culturing is shown on Figure 74-A. This comparison indicates that mean titer variability in condition without selection pressure during sub-culturing is higher and significantly different from condition with selection pressure during sub-culturing. Comparison of mean specific productivity variability between conditions with and without selection pressure during sub-culturing is shown on Figure 75-A. This comparison indicates that mean specific productivity variability in condition without selection pressure during sub-culturing is higher and significantly different from condition with selection pressure during sub-culturing.

Analysis of clone stability, depending on promoter used for *pac* expression

Supplementary data and statistical analysis related to 3.3.5.5.2.2 Analysis of clone stability, depending on promoter used for *pac* expression is detailed hereinafter.

Significance of mean titer variability and mean specific productivity variability differences between SV40E wild type and SV40E_Δ21bpRIII-II promoters was evaluated by a statistical analysis, including assessment of normal distribution of data, ANOVA, and Tukey-Kramer test. Assessment of normal distribution was performed on studied response residuals and normality was demonstrated (data not shown). Comparison of mean titer variability between SV40E wild type and SV40E_Δ21bpRIII-II promoters is shown on Figure 74-B. ANOVA indicates that there is no statistical difference on mean titer variability (p-value above $\alpha_{5\%}$) between conditions. Comparison of mean specific productivity variability between SV40E wild type and SV40E_Δ21bpRIII-II promoters is shown on Figure 75-B. ANOVA indicates that is at least one statistical difference on mean specific productivity variability (p-value below $\alpha_{5\%}$). Analysis is further investigated with a Tukey-Kramer test. Clones generated with SV40E_Δ21bpRIII-II and cultivated without selection pressure show a significantly higher variability than clones generated with SV40E and cultivated with selection pressure. However, this comparison does not seem relevant, as both promoter for *pac* expression and selection pressure vary. All other comparisons of interest, namely SV40E wild type promoter *versus* SV40E_Δ21bpRIII-II promoter with selection pressure during sub-culturing, and SV40E wild type promoter *versus* SV40E_Δ21bpRIII-II promoter without selection pressure during sub-culturing, show no statistically significant difference.

Table 34: Rituximab titer variations between PDL_{20}, PDL_{40}, and PDL_{60} (clones with SV40E wild type promoter, w/o selection pressure during sub-culturing)

Variations between PDL were calculated on the last common day of culture between the three productivity assays.

Clone ID	PDL	Last production day (PD)	[Rituximab] on last PD [mg/L]	Last common production day (PD)	[Rituximab] on last common PD [mg/L]	Δ[Rituximab] between PDL_{20}-PDL_{40}	Δ[Rituximab] between PDL_{40}-PDL_{60}	Δ[Rituximab] between PDL_{20}-PDL_{60}	Conclusion on expression level stability
CL 01H09 w/o SP	PDL_{20}	PD_{12}	1061.6	PD_{10}	706.5	15.5%		-0.1%	**Stable**
	PDL_{40}	PD_{10}	816.4	PD_{10}	816.4		-13.5%		
	PDL_{60}	PD_{10}	706.0	PD_{10}	706.0				
CL 02B09 w/o SP	PDL_{20}	PD_{12}	603.0	PD_{10}	620.5	-7.9%		-23.4%	**Stable**
	PDL_{40}	PD_{12}	798.1	PD_{10}	571.6		-16.8%		
	PDL_{60}	PD_{10}	475.3	PD_{10}	475.3				
CL 02D11 w/o SP	PDL_{20}	PD_{10}	764.3	PD_{10}	764.3	21.1%		23.1%	**Stable**
	PDL_{40}	PD_{10}	925.3	PD_{10}	925.3		1.7%		
	PDL_{60}	PD_{10}	941.1	PD_{10}	941.1				
CL 02E06 w/o SP	PDL_{20}	PD_{12}	868.2	PD_{10}	666.9	6.3%		35.3%	Unstable
	PDL_{40}	PD_{10}	708.8	PD_{10}	708.8		27.3%		
	PDL_{60}	PD_{10}	902.6	PD_{10}	902.6				
CL 02F05 w/o SP	PDL_{20}	PD_{12}	1028.7	PD_{12}	1028.7	-6.3%		0.4%	**Stable**
	PDL_{40}	PD_{12}	964.2	PD_{12}	964.2		7.1%		
	PDL_{60}	PD_{12}	1032.9	PD_{12}	1032.9				
CL 02H09 w/o SP	PDL_{20}	PD_{12}	731.3	PD_{10}	532.0	-19.6%		-30.4%	Unstable
	PDL_{40}	PD_{10}	427.6	PD_{10}	427.6		-13.4%		
	PDL_{60}	PD_{10}	370.2	PD_{10}	370.2				
CL 03E05 w/o SP	PDL_{20}	PD_{12}	1579.0	PD_{12}	1579.0	-61.7%		-91.8%	Unstable
	PDL_{40}	PD_{12}	605.5	PD_{12}	605.5		-78.7%		
	PDL_{60}	PD_{12}	128.8	PD_{12}	128.8				
CL 05E12 w/o SP	PDL_{20}	PD_{12}	666.4	PD_{12}	666.4	-6.6%		-43.3%	Unstable
	PDL_{40}	PD_{12}	622.2	PD_{12}	622.2		-39.3%		
	PDL_{60}	PD_{12}	377.9	PD_{12}	377.9				

Table 35: Rituximab titer variations between PDL_{20}, PDL_{40}, and PDL_{60} (clones with SV40E wild type promoter, w/ selection pressure during sub-culturing)

Variations between PDL were calculated on the last common day of culture between the three productivity assays.

Clone ID	PDL	Last production day (PD)	[Rituximab] on last PD [mg/L]	Last common production day (PD)	[Rituximab] on last common PD [mg/L]	Δ[Rituximab] between PDL_{20}-PDL_{40}	Δ[Rituximab] between PDL_{40}-PDL_{60}	Δ[Rituximab] between PDL_{20}-PDL_{60}	Conclusion on expression level stability
CL 01H09 w SP	PDL_{20}	PD_{12}	1124.8	PD_{10}	865.8	6.4%		23.1%	**Stable**
	PDL_{40}	PD_{10}	921.6	PD_{10}	921.6		15.7%		
	PDL_{60}	PD_{12}	840.7	PD_{10}	1066.1				
CL 02B09 w SP	PDL_{20}	PD_{12}	851.2	PD_{10}	700.2	17.5%		25.5%	**Stable**
	PDL_{40}	PD_{12}	988.7	PD_{10}	822.7		6.8%		
	PDL_{60}	PD_{10}	878.5	PD_{10}	878.5				
CL 02D11 w SP	PDL_{20}	PD_{7}	368.1	PD_{7}	368.1	28.5%		51.9%	Unstable
	PDL_{40}	PD_{10}	751.9	PD_{7}	473.0		18.2%		
	PDL_{60}	PD_{10}	976.7	PD_{7}	559.0				
CL 02E06 w SP	PDL_{20}	PD_{12}	603.5	PD_{12}	603.5	42.5%		30.1%	Unstable
	PDL_{40}	PD_{12}	859.7	PD_{12}	859.7		-8.7%		
	PDL_{60}	PD_{12}	784.9	PD_{12}	784.9				
CL 02F05 w SP	PDL_{20}	PD_{12}	1106.5	PD_{12}	1106.5	14.9%		1.4%	**Stable**
	PDL_{40}	PD_{12}	1271.5	PD_{12}	1271.5		-11.8%		
	PDL_{60}	PD_{12}	1122.0	PD_{12}	1122.0				
CL 02H09 w SP	PDL_{20}	PD_{12}	884.4	PD_{12}	884.4	8.9%		3.5%	**Stable**
	PDL_{40}	PD_{12}	963.4	PD_{12}	963.4		-5.0%		
	PDL_{60}	PD_{12}	915.0	PD_{12}	915.0				
CL 03E05 w SP	PDL_{20}	PD_{10}	1245.9	PD_{10}	1245.9	-22.0%		-22.0%	**Stable**
	PDL_{40}	PD_{12}	1412.1	PD_{10}	972.1		-0.1%		
	PDL_{60}	PD_{10}	971.2	PD_{10}	971.2				
CL 05E12 w SP	PDL_{20}	PD_{12}	750.8	PD_{12}	750.8	8.7%		-20.8%	**Stable**
	PDL_{40}	PD_{12}	816.0	PD_{12}	816.0		-27.1%		
	PDL_{60}	PD_{12}	594.6	PD_{12}	594.6				

Table 36: Rituximab titer variations between PDL_{20}, PDL_{40}, and PDL_{60} (clones with SV40E_Δ21bpRIII-II promoter, w/o selection pressure during sub-culturing)

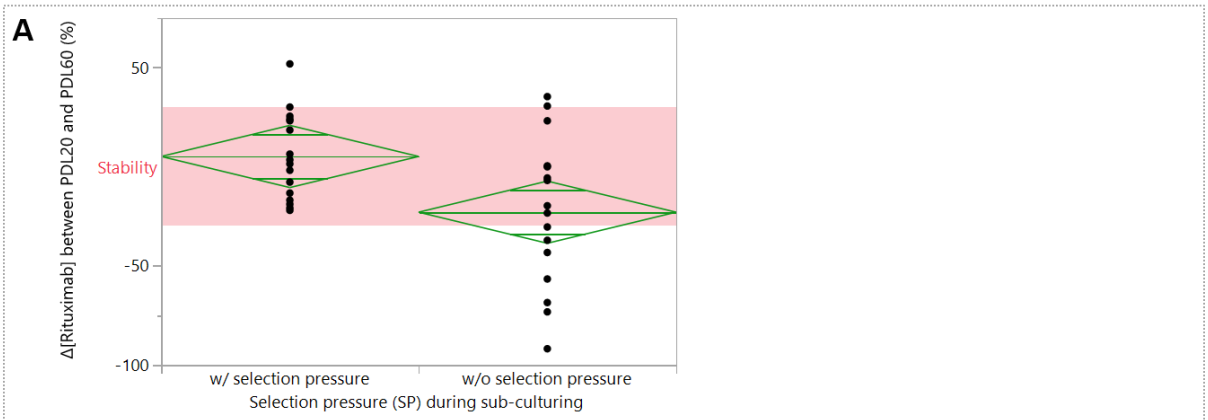
Variations between PDL were calculated on the last common day of culture between the three productivity assays.

Clone ID	PDL	Last production day (PD)	[Rituximab] on last PD [mg/L]	Last common production day (PD)	[Rituximab] on last common PD [mg/L]	Δ[Rituximab] between PDL_{20}-PDL_{40}	Δ[Rituximab] between PDL_{40}-PDL_{60}	Δ[Rituximab] between PDL_{20}-PDL_{60}	Conclusion on expression level stability
CL 11F08 w/o SP	PDL_{20}	PD_{12}	1068.8	PD_{12}	1068.8	-25.8%		-56.7%	Unstable
	PDL_{40}	PD_{12}	793.0	PD_{12}	793.0		-41.6%		
	PDL_{60}	PD_{12}	463.2	PD_{12}	463.2				
CL 13D11 w/o SP	PDL_{20}	PD_{12}	976.1	PD_{12}	976.1	14.8%		-7.0%	**Stable**
	PDL_{40}	PD_{12}	1120.8	PD_{12}	1120.8		-19.0%		
	PDL_{60}	PD_{12}	908.2	PD_{12}	908.2				
CL 13E01 w/o SP	PDL_{20}	PD_{12}	1916.2	PD_{12}	1916.2	-45.2%		-68.6%	Unstable
	PDL_{40}	PD_{12}	1049.5	PD_{12}	1049.5		-42.6%		
	PDL_{60}	PD_{12}	602.4	PD_{12}	602.4				
CL 15A10 w/o SP	PDL_{20}	PD_{12}	1330.6	PD_{12}	1330.6	11.1%		-5.7%	**Stable**
	PDL_{40}	PD_{12}	1478.3	PD_{12}	1478.3		-15.1%		
	PDL_{60}	PD_{12}	1255.2	PD_{12}	1255.2				
CL 21H09 w/o SP	PDL_{20}	PD_{12}	3382.6	PD_{12}	3382.6	-4.8%		-37.1%	Unstable
	PDL_{40}	PD_{12}	3220.3	PD_{12}	3220.3		-34.0%		
	PDL_{60}	PD_{12}	2126.7	PD_{12}	2126.7				
CL 22G03 w/o SP	PDL_{20}	PD_{12}	810.1	PD_{10}	757.2	14.2%		30.6%	Unstable
	PDL_{40}	PD_{10}	864.5	PD_{10}	864.5		14.4%		
	PDL_{60}	PD_{12}	1001.0	PD_{10}	988.7				
CL 24B06 w/o SP	PDL_{20}	PD_{12}	1417.6	PD_{10}	1074.1	-59.2%		-73.3%	Unstable
	PDL_{40}	PD_{12}	445.7	PD_{10}	438.4		-34.5%		
	PDL_{60}	PD_{10}	287.2	PD_{10}	287.2				
CL 24H08 w/o SP	PDL_{20}	PD_{10}	736.1	PD_{7}	526.0	-12.8%		-19.7%	**Stable**
	PDL_{40}	PD_{7}	458.5	PD_{7}	458.5		-7.9%		
	PDL_{60}	PD_{10}	536.5	PD_{7}	422.3				

Table 37: Rituximab titer variations between PDL_{20}, PDL_{40}, and PDL_{60} (clones with SV40E_Δ21bpRIII-II promoter, w/ selection pressure during sub-culturing)

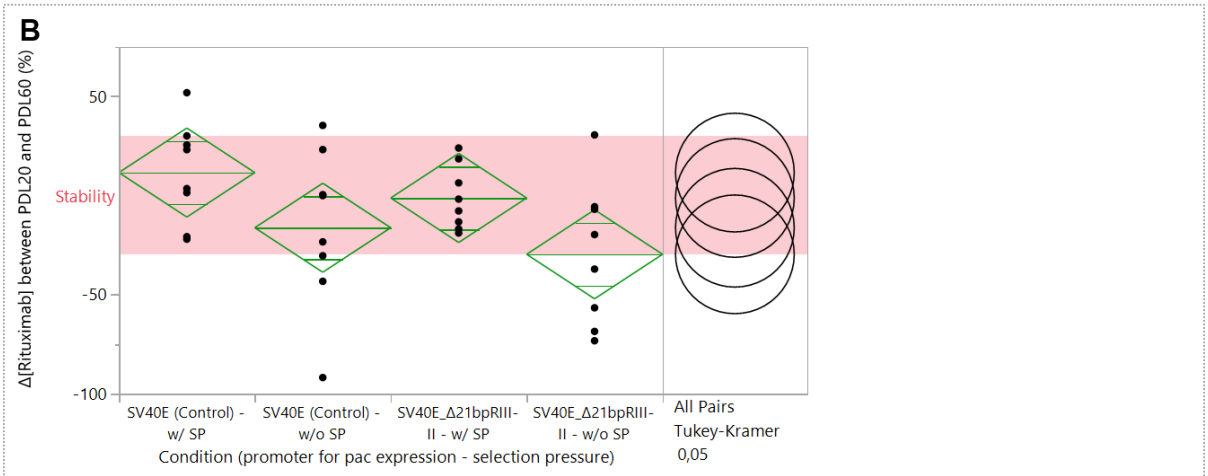
Variations between PDL were calculated on the last common day of culture between the three productivity assays.

Clone ID	PDL	Last production day (PD)	[Rituximab] on last PD [mg/L]	Last common production day (PD)	[Rituximab] on last common PD [mg/L]	Δ[Rituximab] between PDL_{20}-PDL_{40}	Δ[Rituximab] between PDL_{40}-PDL_{60}	Δ[Rituximab] between PDL_{20}-PDL_{60}	Conclusion on expression level stability
CL 11F08 w SP	PDL_{20}	PD_{12}	1292.4	PD_{12}	1292.4	-25.1%		-18.9%	**Stable**
	PDL_{40}	PD_{12}	968.3	PD_{12}	968.3		8.2%		
	PDL_{60}	PD_{12}	1047.9	PD_{12}	1047.9				
CL 13D11 w SP	PDL_{20}	PD_{12}	796.8	PD_{12}	796.8	-6.5%		-13.3%	**Stable**
	PDL_{40}	PD_{12}	745.3	PD_{12}	745.3		-7.3%		
	PDL_{60}	PD_{12}	690.7	PD_{12}	690.7				
CL 13E01 w SP	PDL_{20}	PD_{12}	1953.4	PD_{12}	1953.4	0.1%		6.4%	**Stable**
	PDL_{40}	PD_{12}	1954.5	PD_{12}	1954.5		6.3%		
	PDL_{60}	PD_{12}	2077.7	PD_{12}	2077.7				
CL 15A10 w SP	PDL_{20}	PD_{12}	1871.2	PD_{12}	1871.2	-35.2%		-7.8%	**Stable**
	PDL_{40}	PD_{12}	1213.3	PD_{12}	1213.3		42.2%		
	PDL_{60}	PD_{12}	1725.3	PD_{12}	1725.3				
CL 21H09 w SP	PDL_{20}	PD_{12}	2704.5	PD_{12}	2704.5	-8.0%		-16.9%	**Stable**
	PDL_{40}	PD_{12}	2487.4	PD_{12}	2487.4		-9.6%		
	PDL_{60}	PD_{12}	2247.4	PD_{12}	2247.4				
CL 22G03 w SP	PDL_{20}	PD_{10}	776.1	PD_{10}	776.1	1.3%		23.9%	**Stable**
	PDL_{40}	PD_{10}	786.2	PD_{10}	786.2		22.3%		
	PDL_{60}	PD_{12}	961.9	PD_{10}	961.9				
CL 24B06 w SP	PDL_{20}	PD_{12}	2104.3	PD_{12}	2104.3	3.3%		18.5%	**Stable**
	PDL_{40}	PD_{12}	2173.2	PD_{12}	2173.2		14.7%		
	PDL_{60}	PD_{12}	2492.8	PD_{12}	2492.8				
CL 24H08 w SP	PDL_{20}	PD_{10}	689.3	PD_{7}	480.5	16.8%		-1.8%	**Stable**
	PDL_{40}	PD_{7}	561.2	PD_{7}	561.2		-15.9%		
	PDL_{60}	PD_{10}	723.7	PD_{7}	471.8				



Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio	Prob > F
Selection pressure (SP) during sub-culturing	1	6330,970	6330,97	6,7088	0,0147*
Error	30	28310,498	943,68		
C. Total	31	34641,467			



Means for Oneway Anova

Level	Number	Mean	Std Error	Lower 95%	Upper 95%
SV40E (Control) - w/ SP	8	11,568	10,966	-10,89	34,03
SV40E (Control) - w/o SP	8	-16,269	10,966	-38,73	6,19
SV40E_Δ21bpRIII-II - w/ SP	8	-1,247	10,966	-23,71	21,21
SV40E_Δ21bpRIII-II - w/o SP	8	-29,672	10,966	-52,13	-7,21

Std Error uses a pooled estimate of error variance

Ordered Differences Report

Level	- Level	Difference	Std Err Dif	Lower CL	Upper CL	p-Value
SV40E (Control) - w/ SP	SV40E_Δ21bpRIII-II - w/o SP	41,24039	15,50778	-1,1007	83,58148	0,0584
SV40E_Δ21bpRIII-II - w/ SP	SV40E_Δ21bpRIII-II - w/o SP	28,42512	15,50778	-13,9160	70,76621	0,2797
SV40E (Control) - w/ SP	SV40E (Control) - w/o SP	27,83752	15,50778	-14,5036	70,17862	0,2968
SV40E_Δ21bpRIII-II - w/ SP	SV40E (Control) - w/o SP	15,02225	15,50778	-27,3188	57,36335	0,7681
SV40E (Control) - w/o SP	SV40E_Δ21bpRIII-II - w/o SP	13,40287	15,50778	-28,9382	55,74396	0,8230
SV40E (Control) - w/ SP	SV40E_Δ21bpRIII-II - w/ SP	12,81527	15,50778	-29,5258	55,15637	0,8415

Figure 74: Statistical comparison of group mean expression level variations between PDL_{20} and PDL_{60}

(A) *Analysis of variance results on mean rituximab titer variability. Normality of data distribution was previously assessed (data not shown). Significance of mean rituximab titer variability difference between conditions where selection pressure is maintained or not during sub-culturing (irrespectively of promoter used for pac expression) was evaluated by a statistical analysis, including assessment of normal distribution, one-way analysis of variance (ANOVA) (α 5%). Green lines indicate the mean of each condition. X-axis size is not proportional to sample size.* ***(B)*** *Analysis of variance results on mean*

rituximab titer variability. Normality of data distribution was previously assessed (data not shown). Significance of mean rituximab titer variability difference between conditions (promoter and selection pressure) was evaluated by a statistical analysis, including one-way analysis of variance (ANOVA) (α 5%). Green lines indicate the mean of each condition. X-axis size is not proportional to sample size. Data analysis was performed on JMP® 15 software.

Table 38: Q_P variations between PDL_{20}, PDL_{40}, and PDL_{60} (clones with SV40E wild type promoter, w/o selection pressure during sub-culturing)

Variations between PDL were calculated on the last common day of culture between the three productivity assays.

Clone ID	PDL	Last production day (PD)	Mean process Q_p on last PD [pg/cell/day]	Last common production day (PD)	Mean process Q_p on last common PD [pg/cell/day]	ΔQ_p between PDL_{20}-PDL_{40}	ΔQ_p between PDL_{40}-PDL_{60}	ΔQ_p between PDL_{20}-PDL_{60}	Conclusion on mean process Q_p stability
CL 01H09 w/o SP	PDL_{20}	PD_{12}	4.9	PD_{10}	4.0	15.7%		-18.1%	**Stable**
	PDL_{40}	PD_{10}	6.5	PD_{10}	4.6		-29.2%		
	PDL_{60}	PD_{10}	4.5	PD_{10}	3.3				
CL 02B09 w/o SP	PDL_{20}	PD_{12}	2.4	PD_{10}	2.8	-20.1%		-29.9%	**Stable**
	PDL_{40}	PD_{12}	2.3	PD_{10}	2.3		-12.3%		
	PDL_{60}	PD_{10}	2.6	PD_{10}	2.0				
CL 02D11 w/o SP	PDL_{20}	PD_{10}	6.1	PD_{10}	4.8	15.9%		6.2%	**Stable**
	PDL_{40}	PD_{10}	6.9	PD_{10}	5.6		-8.4%		
	PDL_{60}	PD_{10}	7.4	PD_{10}	5.1				
CL 02E06 w/o SP	PDL_{20}	PD_{12}	3.0	PD_{10}	3.3	2.6%		17.9%	**Stable**
	PDL_{40}	PD_{10}	2.4	PD_{10}	3.4		14.9%		
	PDL_{60}	PD_{10}	3.7	PD_{10}	3.9				
CL 02F05 w/o SP	PDL_{20}	PD_{12}	10.2	PD_{12}	8.9	-5.2%		-14.4%	**Stable**
	PDL_{40}	PD_{12}	5.2	PD_{12}	8.4		-9.7%		
	PDL_{60}	PD_{12}	4.9	PD_{12}	7.6				
CL 02H09 w/o SP	PDL_{20}	PD_{12}	3.9	PD_{10}	3.7	-23.9%		-38.9%	Unstable
	PDL_{40}	PD_{10}	3.5	PD_{10}	2.8		-19.6%		
	PDL_{60}	PD_{10}	3.2	PD_{10}	2.3				
CL 03E05 w/o SP	PDL_{20}	PD_{12}	22.0	PD_{12}	9.3	-67.9%		-93.5%	Unstable
	PDL_{40}	PD_{12}	4.6	PD_{12}	3.0		-79.9%		
	PDL_{60}	PD_{12}	0.9	PD_{12}	0.6				
CL 05E12 w/o SP	PDL_{20}	PD_{12}	1.3	PD_{12}	3.0	-11.3%		-49.5%	Unstable
	PDL_{40}	PD_{12}	1.0	PD_{12}	2.6		-43.1%		
	PDL_{60}	PD_{12}	0.9	PD_{12}	1.5				

Table 39: Q_P variations between PDL_{20}, PDL_{40}, and PDL_{60} (clones with SV40E wild type promoter, w/ selection pressure during sub-culturing)

Variations between PDL were calculated on the last common day of culture between the three productivity assays.

Clone ID	PDL	Last production day (PD)	Mean process Q_p on last PD [pg/cell/day]	Last common production day (PD)	Mean process Q_p on last common PD [pg/cell/day]	ΔQ_p between PDL_{20}-PDL_{40}	ΔQ_p between PDL_{40}-PDL_{60}	ΔQ_p between PDL_{20}-PDL_{60}	Conclusion on mean process Q_p stability
CL 01H09 w SP	PDL_{20}	PD_{12}	5.0	PD_{10}	4.3	11.9%		13.5%	**Stable**
	PDL_{40}	PD_{10}	7.1	PD_{10}	4.8		1.3%		
	PDL_{60}	PD_{12}	6.4	PD_{10}	4.8				
CL 02B09 w SP	PDL_{20}	PD_{12}	2.5	PD_{10}	3.5	11.4%		8.4%	**Stable**
	PDL_{40}	PD_{12}	4.2	PD_{10}	3.9		-2.7%		
	PDL_{60}	PD_{10}	3.9	PD_{10}	3.8				
CL 02D11 w SP	PDL_{20}	PD_{7}	3.5	PD_{7}	3.6	28.6%		53.7%	Unstable
	PDL_{40}	PD_{10}	6.1	PD_{7}	4.6		19.5%		
	PDL_{60}	PD_{10}	7.1	PD_{7}	5.6				
CL 02E06 w SP	PDL_{20}	PD_{12}	0.8	PD_{12}	2.7	28.2%		7.8%	**Stable**
	PDL_{40}	PD_{12}	2.8	PD_{12}	3.5		-15.9%		
	PDL_{60}	PD_{12}	3.3	PD_{12}	2.9				
CL 02F05 w SP	PDL_{20}	PD_{12}	17.5	PD_{12}	9.1	18.2%		-1.1%	**Stable**
	PDL_{40}	PD_{12}	17.9	PD_{12}	10.8		-16.4%		
	PDL_{60}	PD_{12}	9.5	PD_{12}	9.0				
CL 02H09 w SP	PDL_{20}	PD_{12}	11.5	PD_{12}	6.0	5.7%		-5.6%	**Stable**
	PDL_{40}	PD_{12}	10.6	PD_{12}	6.3		-10.6%		
	PDL_{60}	PD_{12}	1.3	PD_{12}	5.7				
CL 03E05 w SP	PDL_{20}	PD_{10}	12.2	PD_{10}	8.5	-14.3%		-12.4%	**Stable**
	PDL_{40}	PD_{12}	9.7	PD_{10}	7.3		2.2%		
	PDL_{60}	PD_{10}	9.0	PD_{10}	7.4				
CL 05E12 w SP	PDL_{20}	PD_{12}	2.7	PD_{12}	3.2	15.3%		-25.7%	**Stable**
	PDL_{40}	PD_{12}	4.3	PD_{12}	3.7		-35.5%		
	PDL_{60}	PD_{12}	11.3	PD_{12}	2.4				

Table 40: Q_P variations between PDL_{20}, PDL_{40}, and PDL_{60} (clones with SV40E_Δ21bpRIII-II promoter, w/o selection pressure during sub-culturing)

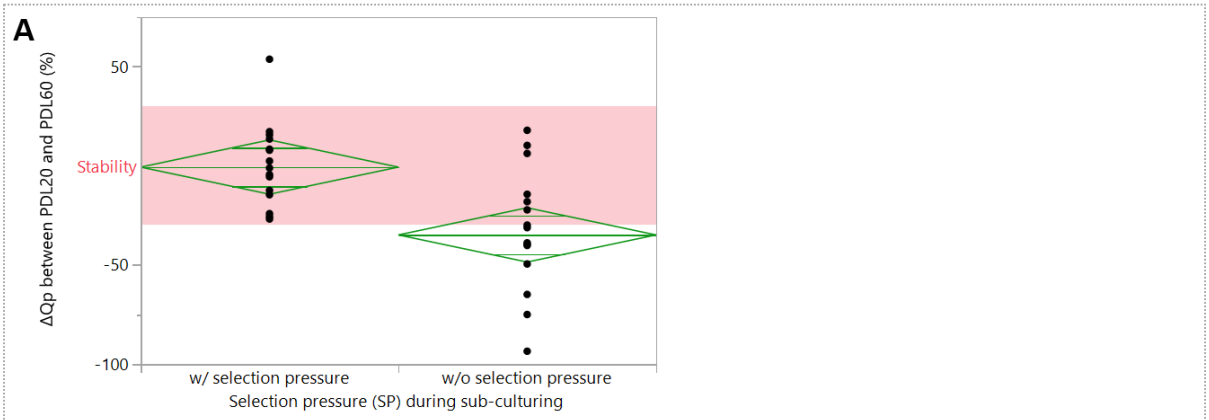
Variations between PDL were calculated on the last common day of culture between the three productivity assays.

Clone ID	PDL	Last production day (PD)	Mean process Q_p on last PD [pg/cell/day]	Last common production day (PD)	Mean process Q_p on last common PD [pg/cell/day]	ΔQ_p between PDL_{20}-PDL_{40}	ΔQ_p between PDL_{40}-PDL_{60}	ΔQ_p between PDL_{20}-PDL_{60}	Conclusion on mean process Q_p stability
CL 11F08 w/o SP	PDL_{20}	PD_{12}	4.2	PD_{12}	5.6	-34.3%	-46.5%	-64.9%	Unstable
	PDL_{40}	PD_{12}	4.5	PD_{12}	3.7				
	PDL_{60}	PD_{12}	0.8	PD_{12}	2.0				
CL 13D11 w/o SP	PDL_{20}	PD_{12}	11.8	PD_{12}	10.9	1.4%	-23.3%	-22.2%	**Stable**
	PDL_{40}	PD_{12}	5.3	PD_{12}	11.1				
	PDL_{60}	PD_{12}	13.8	PD_{12}	8.5				
CL 13E01 w/o SP	PDL_{20}	PD_{12}	10.3	PD_{12}	10.3	-54.7%	-44.9%	-75.0%	Unstable
	PDL_{40}	PD_{12}	6.9	PD_{12}	4.7				
	PDL_{60}	PD_{12}	3.5	PD_{12}	2.6				
CL 15A10 w/o SP	PDL_{20}	PD_{12}	7.8	PD_{12}	7.0	-17.3%	-27.1%	-39.7%	Unstable
	PDL_{40}	PD_{12}	7.8	PD_{12}	5.8				
	PDL_{60}	PD_{12}	3.8	PD_{12}	4.2				
CL 21H09 w/o SP	PDL_{20}	PD_{12}	12.9	PD_{12}	15.0	-11.2%	-32.7%	-40.2%	Unstable
	PDL_{40}	PD_{12}	11.2	PD_{12}	13.3				
	PDL_{60}	PD_{12}	14.0	PD_{12}	9.0				
CL 22G03 w/o SP	PDL_{20}	PD_{12}	4.1	PD_{10}	3.8	2.6%	7.5%	10.3%	**Stable**
	PDL_{40}	PD_{10}	5.0	PD_{10}	3.9				
	PDL_{60}	PD_{12}	4.4	PD_{10}	4.2				
CL 24B06 w/o SP	PDL_{20}	PD_{12}	9.1	PD_{10}	5.6	-61.8%	-34.1%	-74.8%	Unstable
	PDL_{40}	PD_{12}	2.8	PD_{10}	2.1				
	PDL_{60}	PD_{10}	1.9	PD_{10}	1.4				
CL 24H08 w/o SP	PDL_{20}	PD_{10}	11.7	PD_{7}	9.8	-20.7%	-13.2%	-31.2%	Unstable
	PDL_{40}	PD_{7}	7.4	PD_{7}	7.8				
	PDL_{60}	PD_{10}	9.2	PD_{7}	6.8				

Table 41: Q_P variations between PDL_{20}, PDL_{40}, and PDL_{60} (clones with SV40E_Δ21bpRIII-II promoter, w/ selection pressure during sub-culturing)

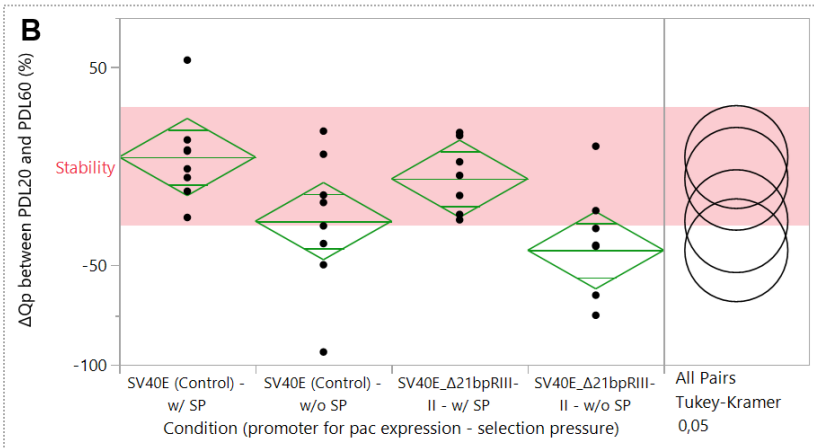
Variations between PDL were calculated on the last common day of culture between the three productivity assays.

Clone ID	PDL	Last production day (PD)	Mean process Q_p on last PD [pg/cell/day]	Last common production day (PD)	Mean process Q_p on last common PD [pg/cell/day]	ΔQ_p between PDL_{20}-PDL_{40}	ΔQ_p between PDL_{40}-PDL_{60}	ΔQ_p between PDL_{20}-PDL_{60}	Conclusion on mean process Q_p stability
CL 11F08 w SP	PDL_{20}	PD_{12}	17.5	PD_{12}	8.6	-10.5%		-14.6%	**Stable**
	PDL_{40}	PD_{12}	2.6	PD_{12}	7.7		-4.6%		
	PDL_{60}	PD_{12}	0.7	PD_{12}	7.4				
CL 13D11 w SP	PDL_{20}	PD_{12}	10.0	PD_{12}	8.1	-4.1%		-14.4%	**Stable**
	PDL_{40}	PD_{12}	3.1	PD_{12}	7.8		-10.7%		
	PDL_{60}	PD_{12}	5.0	PD_{12}	6.9				
CL 13E01 w SP	PDL_{20}	PD_{12}	15.5	PD_{12}	10.1	-9.7%		-4.5%	**Stable**
	PDL_{40}	PD_{12}	18.8	PD_{12}	9.1		5.8%		
	PDL_{60}	PD_{12}	28.8	PD_{12}	9.7				
CL 15A10 w SP	PDL_{20}	PD_{12}	15.4	PD_{12}	8.8	-36.0%		-24.1%	**Stable**
	PDL_{40}	PD_{12}	3.5	PD_{12}	5.7		18.6%		
	PDL_{60}	PD_{12}	9.0	PD_{12}	6.7				
CL 21H09 w SP	PDL_{20}	PD_{12}	3.3	PD_{12}	13.0	-20.3%		-26.9%	**Stable**
	PDL_{40}	PD_{12}	6.8	PD_{12}	10.3		-8.3%		
	PDL_{60}	PD_{12}	7.8	PD_{12}	9.5				
CL 22G03 w SP	PDL_{20}	PD_{10}	3.7	PD_{10}	3.6	11.1%		15.6%	**Stable**
	PDL_{40}	PD_{10}	4.7	PD_{10}	4.0		4.1%		
	PDL_{60}	PD_{12}	5.4	PD_{10}	4.2				
CL 24B06 w SP	PDL_{20}	PD_{12}	9.8	PD_{12}	8.7	4.2%		17.2%	**Stable**
	PDL_{40}	PD_{12}	5.9	PD_{12}	9.0		12.4%		
	PDL_{60}	PD_{12}	11.3	PD_{12}	10.2				
CL 24H08 w SP	PDL_{20}	PD_{10}	8.0	PD_7	7.5	20.9%		2.4%	**Stable**
	PDL_{40}	PD_7	7.5	PD_7	9.1		-15.3%		
	PDL_{60}	PD_{10}	8.9	PD_7	7.7				



Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio	Prob > F
Selection pressure (SP) during sub-culturing	1	9355,790	9355,79	13,0180	0,0011*
Error	30	21560,463	718,68		
C. Total	31	30916,253			



Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio	Prob > F
Condition (promoter for pac expression - selection pressure)	3	10700,086	3566,70	4,9400	0,0071*
Error	28	20216,167	722,01		
C. Total	31	30916,253			

Ordered Differences Report

Level	- Level	Difference	Std Err Dif	Lower CL	Upper CL	p-Value
SV40E (Control) - w/ SP	SV40E_Δ21bpRIII-II - w/o SP	47,02694	13,43508	10,3450	83,70893	0,0081*
SV40E_Δ21bpRIII-II - w/ SP	SV40E_Δ21bpRIII-II - w/o SP	36,05340	13,43508	-0,6286	72,73539	0,0554
SV40E (Control) - w/ SP	SV40E (Control) - w/o SP	32,34173	13,43508	-4,3402	69,02372	0,0989
SV40E_Δ21bpRIII-II - w/ SP	SV40E (Control) - w/o SP	21,36819	13,43508	-15,3138	58,05018	0,4001
SV40E (Control) - w/o SP	SV40E_Δ21bpRIII-II - w/o SP	14,68521	13,43508	-21,9968	51,36719	0,6966
SV40E (Control) - w/ SP	SV40E_Δ21bpRIII-II - w/ SP	10,97354	13,43508	-25,7084	47,65552	0,8460

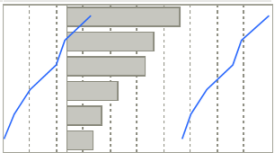


Figure 75:Statistical comparison of group mean specific productivity variations between PDL_{20} and PDL_{60}

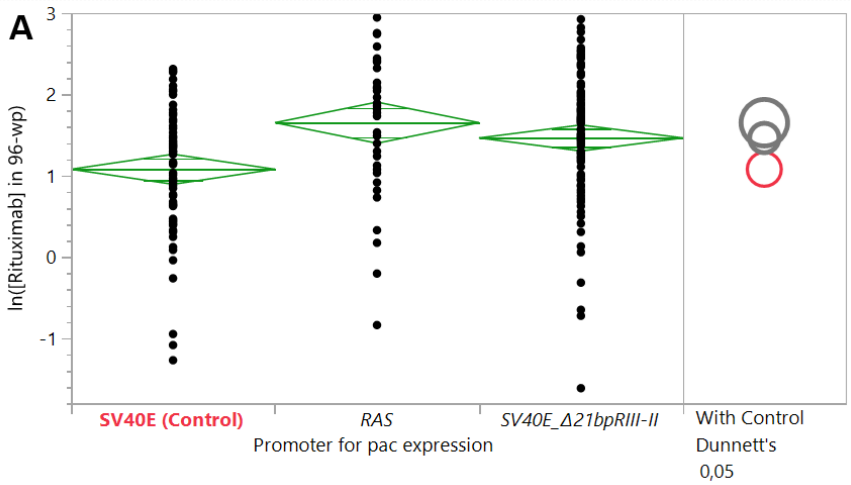
(A) *Analysis of variance results on mean specific productivity variability. Normality of data distribution was previously assessed (data not shown). Significance of mean specific productivity variability difference between conditions where selection pressure is maintained or not during sub-culturing (irrespectively of promoter used for pac expression) was evaluated by a statistical analysis, including assessment of normal distribution, one-way analysis of variance (ANOVA) (α 5%). Green lines indicate the mean of each condition. X-axis size is not proportional to sample size.* ***(B)*** *Analysis of variance results on mean specific productivity variability. Normality of data distribution was previously assessed*

(data not shown). Significance of mean specific productivity variability difference between conditions (promoter and selection pressure) was evaluated by a statistical analysis, including one-way analysis of variance (ANOVA) (α 5%). Green lines indicate the mean of each condition. X-axis size is not proportional to sample size. Data analysis was performed on JMP® 15 software.

Details on statistical analysis on mini-pools' productivity (confirmation on several molecules)

Statistical analysis related to 3.3.5.6 Productivity gain obtained with RAS and SV40E_Δ21bpRIII-II promoters is confirmed on other recombinant proteins is detailed hereinafter.

Significance of mean titer difference between candidate and SV40E wild type promoters was evaluated by a statistical analysis. Assessment of normal distribution was performed on studied ln-transformed response residuals and normality was demonstrated only for rituximab, trastuzumab, and pembrolizumab datasets (data not shown). Thus, parametric tests can be applied for rituximab, trastuzumab, and pembrolizumab datasets analysis, and non-parametric tests should be applied for nivolumab and EPO-F_C datasets analysis. Comparison of mean titer between candidate and SV40E wild type promoters is shown on Figure 76. RAS promoter shows a significantly higher mean titer than SV40E wild type promoter for rituximab, trastuzumab, and pembrolizumab expression. RAS promoter also shows a significantly higher mean titer than SV40E wild type promoter for nivolumab and EPO-F_C expression, though it should be kept in mind that non-parametric tests are less robust than parametric ones. SV40E_Δ21bpRIII-II promoter shows a significantly higher mean titer than SV40E wild type promoter for rituximab and trastuzumab. SV40E_Δ21bpRIII-II promoter also shows a significantly higher mean titer than SV40E wild type promoter for EPO-F_C expression (same consideration about non-parametric assay to be considered).



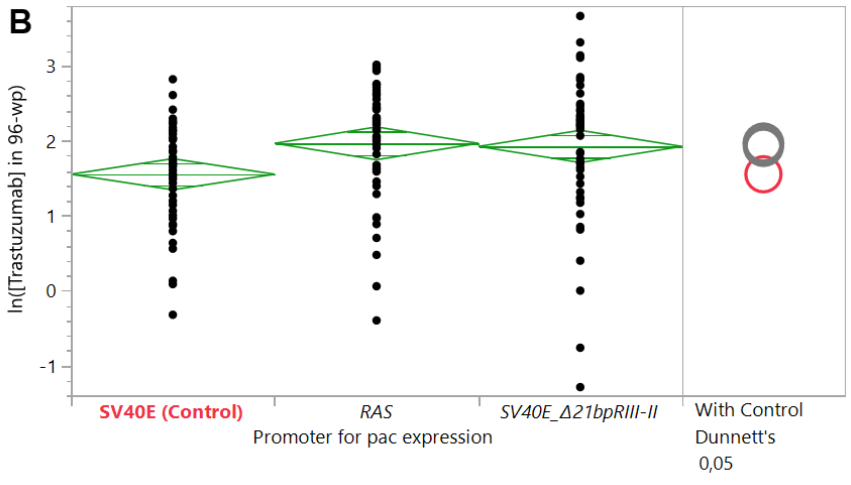
Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio	Prob > F
Promoter for pac expression	2	10,40347	5,20173	7,8943	0,0005*
Error	212	139,69140	0,65892		
C. Total	214	150,09487			

LSD Threshold Matrix

Level	Abs(Dif)-LSD	p-Value
RAS	0,218	0,0008*
SV40E_Δ21bpRIII-II	0,108	0,0042*
SV40E (Control)	-0,3	1,0000

Positive values show pairs of means that are significantly different.



Analysis of Variance

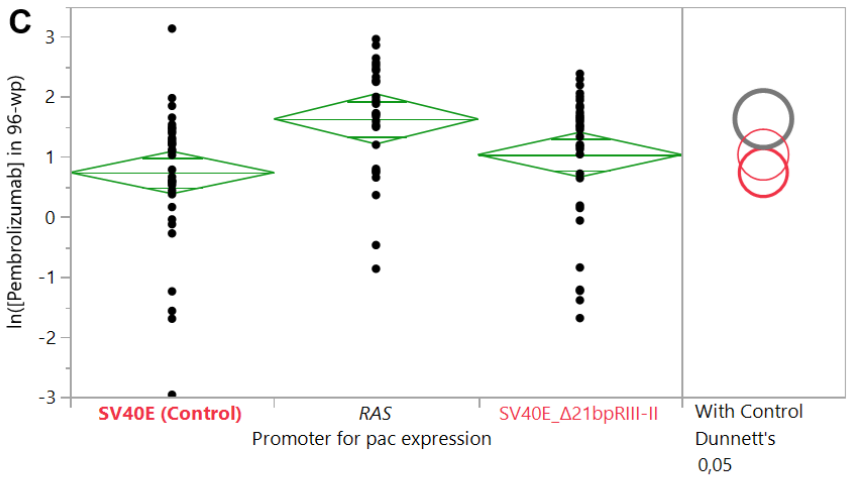
Source	DF	Sum of Squares	Mean Square	F Ratio	Prob > F
Promoter for pac expression	2	5,78674	2,89337	4,5147	0,0124*
Error	161	103,18125	0,64088		
C. Total	163	108,96799			

LSD Threshold Matrix

Level	Abs(Dif)-LSD	p-Value
RAS	0,071	0,0148*
SV40E_Δ21bpRIII-II	0,033	0,0290*
SV40E (Control)	-0,33	1,0000

Positive values show pairs of means that are significantly different.

C



Analysis of Variance

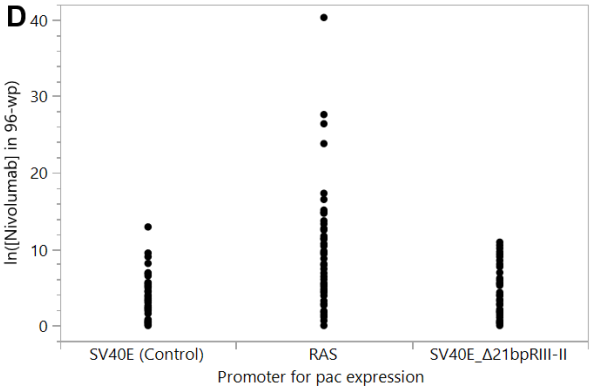
Source	DF	Sum of Squares	Mean Square	F Ratio	Prob > F
Promoter for pac expression	2	12,72106	6,36053	5,3109	0,0065*
Error	96	114,97409	1,19765		
C. Total	98	127,69515			

LSD Threshold Matrix

Level	Abs(Dif)-LSD	p-Value
RAS	0,273	0,0032*
SV40E_Δ21bpRIII-II	-0,28	0,4167
SV40E (Control)	-0,57	1,0000

Positive values show pairs of means that are significantly different.

D

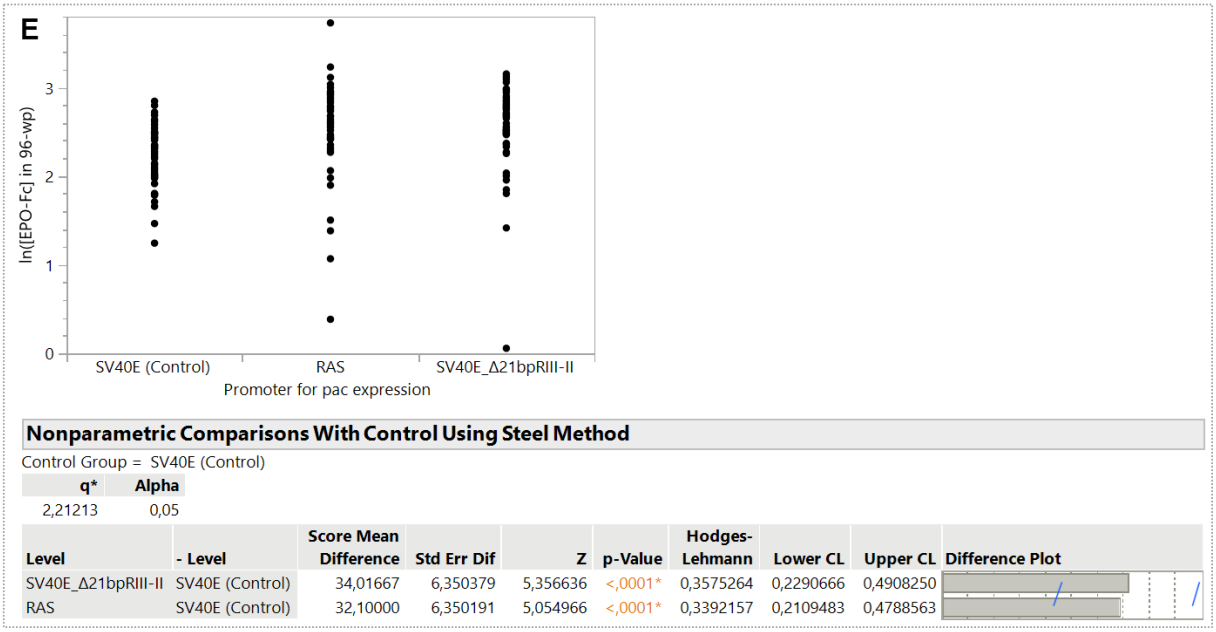


Nonparametric Comparisons With Control Using Steel Method

Control Group = SV40E (Control)

q*	Alpha
2,21285	0,05

Level	- Level	Score Mean Difference	Std Err Dif	Z	p-Value	Hodges-Lehmann	Lower CL	Upper CL	Difference Plot
RAS	SV40E (Control)	29,3868	6,268284	4,68818	<,0001*	4,53500	2,40000	6,850000	
SV40E_Δ21bpRIII-II	SV40E (Control)	-7,3500	6,301071	-1,16647	0,3990	-0,39430	-1,66000	0,375700	



Nonparametric Comparisons With Control Using Steel Method

Control Group = SV40E (Control)

q*	Alpha
2,21213	0,05

Level	- Level	Score Mean Difference	Std Err Dif	Z	p-Value	Hodges-Lehmann	Lower CL	Upper CL	Difference Plot
SV40E_Δ21bpRIII-II	SV40E (Control)	34,01667	6,350379	5,356636	<,0001*	0,3575264	0,2290666	0,4908250	
RAS	SV40E (Control)	32,10000	6,350191	5,054966	<,0001*	0,3392157	0,2109483	0,4788563	

Figure 76: Statistical comparison of group mean recombinant protein concentration in mini-pools' supernatants after batch productivity assay in 96-well plate

(A), ***(B)***, *and* ***(C)*** *Analysis of variance and Dunnett's test results on mean rituximab, trastuzumab, and pembrolizumab (respectively) concentration. Normality of data distribution was previously assessed (data not shown). Significance of mean titer difference between candidate and wild type SV40 promoters was evaluated by a statistical analysis, including assessment of normal distribution, one way analysis of variance (ANOVA) and Dunnett's test (α 5%). Green lines indicate the mean of each condition. X-axis size is not proportional to sample size.* ***(D)*** *and* ***(E)*** *Steel with control test result on mean nivolumab and EPO-FC (respectively) concentration. Normality of data distribution was previously assessed, and data distribution proved to be non-normally distributed (data not shown). Significance of mean titer difference between candidate and wild type SV40 promoters was evaluated with Steel with control test (non-parametric test). Data analysis was performed on JMP® 15 software.*

CHOZN® CHO K1 cell line development platform optimization for production of therapeutic recombinant proteins

Production of therapeutic recombinant proteins is a major development, which has made possible to respond to the limitations of production by natural sources and to develop new drugs. Monoclonal antibodies now dominate the market for recombinant proteins for therapeutic applications. The generation of a cell line is the first step in the development of a recombinant protein production process for therapeutic use. This step, which is decisive from many points of view, ensures the development of a stable clonal cell line that produces the recombinant protein.

A cell line generation platform is composed of a host cell line, an expression vector and a cell line generation process. The objective of this thesis was to understand the critical parameters for the expression of recombinant proteins in CHO (Chinese Hamster Ovary) cells, from the point of view of each of the platform components. This work was directly applied in an industrial context. Indeed, this work aims at optimizing the CHOZN® CHO K1 platform, proposed within the framework of Merck's CDMO (Contract Development and Manufacturing Organization) activities, in order to increase productivity of recombinant cell lines and to reduce the development timelines.

The characteristics and performance of the CHOZN® CHO K1 cell line generation platform were first defined in a process based on a mini-pool generation and a clone generation. This initial characterization, which led to the production of clones producing more than 1 g/L of a standard monoclonal antibody in 7.5 months, led to the identification of two areas of research: optimization of the clonal cell line generation process to reduce cell line development times, and optimization of the expression vector to increase the productivity of generated cell lines.

An accelerated cell line development process, called direct single-cell cloning, was evaluated. This process reduced the duration of the line development process by almost half, but led to a decrease in the productivity of the generated cell lines.

In addition, several attenuated promoters were evaluated for the expression of resistance to selection pressure, essential for the generation of stable and producing clones. Some promoters, used in a cell line generation process based on direct single-cell cloning previously evaluated, led to the identification in 4 months of clones producing more than 3 g/L of a standard monoclonal antibody. In general, the use of these promoters led to an increase of productivity of generated cell lines. This productivity increase was confirmed on several monoclonal antibodies and an Fc fusion protein.

This work has thus made it possible to increase the productivity of recombinant cell lines and to reduce cell line development timelines. These issues are crucial in a context where the process of bringing a therapeutic molecule to the market is highly competitive.

Key-words: Cell line development; CHO K1; Transfection; Single-cell cloning; Selection pressure; Monoclonal antibody

---

Merck Biodevelopment
1 Rue Jacques Monod, 33650 Martillac, France

Institut de Chimie et de Biologie des Membranes et des Nano-objets (CBMN) – UMR 5248 (CNRS, University of Bordeaux, Bordeaux INP) – Equipe Structure et Activités des Macromolécules Biologiques
146 Rue Léo Saignat, 33000 Bordeaux, France

Optimisation de la plateforme CHOZN® CHO K1 de génération de lignées cellulaires clonales pour la production de protéines recombinantes à usage thérapeutique

La production de protéines recombinantes thérapeutiques est une évolution majeure, qui a permis de répondre aux limitations de la production par des sources naturelles et de développer de nouveaux médicaments. Les anticorps monoclonaux dominent désormais le marché des protéines recombinantes à application thérapeutique. La génération d'une lignée cellulaire est la première étape du développement du procédé de production d'une protéine recombinante à usage thérapeutique. Cette étape, décisive à de nombreux points de vue, assure le développement d'une lignée cellulaire clonale, productrice de la protéine recombinante et stable.

Une plateforme de génération de lignées cellulaires est composée d'une lignée hôte, d'un vecteur d'expression et d'un procédé de génération de lignée. La thèse a pour objectif de comprendre les paramètres critiques pour l'expression de protéines recombinantes en cellules CHO (Chinese Hamster Ovary), du point de vue de chacun des éléments de la plateforme. Ces travaux ont directement été mis en application dans un contexte industriel. En effet, ils visent à optimiser la plateforme CHOZN® CHO K1, proposée dans le cadre des activités de CDMO (Contract Development and Manufacturing Organization) de Merck, afin d'augmenter la productivité des lignées générées et de réduire les délais de développement.

Les caractéristiques et la performance de la plateforme de génération de lignée cellulaires CHOZN® CHO K1 ont d'abord été définies au cours d'un procédé basé sur une génération de mini-pools et une génération de clones. Cette caractérisation initiale, ayant conduit à l'obtention en 7,5 mois de clones produisant plus de 1 g/L d'un anticorps monoclonal standard, a permis d'identifier deux axes de travaux de recherche : l'optimisation du procédé de génération de lignées cellulaires clonales pour la diminution des délais de développement de lignées, et l'optimisation du vecteur d'expression, pour l'augmentation de la productivité des lignées générées.

Un procédé de génération de lignées accéléré, appelé clonage direct, a été évalué. Ce procédé a permis de réduire de près de moitié la durée du procédé de développement de lignées, mais a, en revanche, conduit à la diminution de la productivité des lignées générées.

Par ailleurs, plusieurs promoteurs atténués ont été évalués pour l'expression de la résistance à la pression de sélection, indispensable pour la génération de clones stables et producteurs. Certains promoteurs, utilisés lors d'un procédé de génération de lignées basé sur le clonage direct précédemment évalué, ont conduit à l'obtention en 4 mois de clones produisant plus de 3 g/L d'un anticorps monoclonal standard. De façon générale, l'utilisation de ces promoteurs a conduit à l'augmentation de la productivité des lignées générées. L'augmentation de la productivité des lignées a été confirmée sur plusieurs anticorps monoclonaux et une protéine de fusion Fc.

Ces travaux ont donc permis d'augmenter la productivité des lignées recombinantes, et de diminuer les délais de développement de ces lignées. Ces enjeux sont cruciaux, dans un contexte où le processus de mise sur le marché d'une molécule thérapeutique est très concurrentiel.

Mots-clés : Développement de lignées cellulaires ; CHO K1 ; Transfection ; Clonage ; Pression de sélection ; Anticorps monoclonal

---

Merck Biodevelopment
1 Rue Jacques Monod, 33650 Martillac, France

Institut de Chimie et de Biologie des Membranes et des Nano-objets (CBMN) – UMR 5248 (CNRS, Université de Bordeaux, Bordeaux INP) – Equipe Structure et Activités des Macromolécules Biologiques
146 Rue Léo Saignat, 33000 Bordeaux, France