

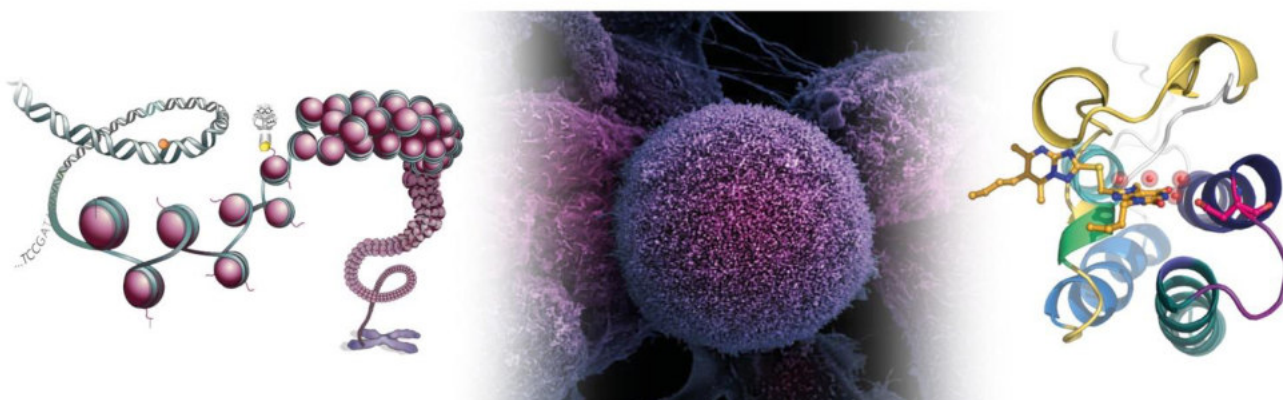
Aix-Marseille Université - École doctorale des sciences de la vie et de la santé
Centre de Recherches en Cancérologie de Marseille



Thèse Confidentielle

Pour l'obtention du grade de docteur en spécialité Biochimie Structurale
Soutenue le 6 novembre 2017

*Développement d'inhibiteurs d'interaction protéine-protéine ciblant
les protéines à bromodomaines :
implications en épigénétique et dans le développement de cancers*



Présentée par **Brigitt RAUX**
Sous la direction du **Dr. Xavier MORELLI**

Devant les membres du jury :

Dr. Françoise **OCHSENBEIN**, Rapporteur
Pr. Jean-François **GUICHOU**, Rapporteur
Pr. Jean-Claude **GUILLEMOT**, Examineur
Dr. Xavier **MORELLI**, Directeur de thèse

4^e année de thèse financée par la **Fondation pour la Recherche Médicale**





Tel un Phénix qui renait de ses cendres

Remerciements

Remerciements général

Aux membres du jury

J'exprime toute ma reconnaissance aux membres de mon jury d'avoir accepté d'évaluer mon travail : le **Dr. Françoise Ochsenbein** et le **Pr. Jean-François Guichou**, pour leurs relectures, l'intérêt qu'ils ont porté à mon manuscrit et leur regard critique de rapporteurs sur ces travaux de recherche ; ainsi que le **Pr. Jean-Claude Guillemot** pour avoir accepté d'être examinateur et de juger mon doctorat.

Au directeur de l'institut et directeurs de l'équipe d'accueil

Je remercie le **Pr. Jean-Paul Borg** de m'avoir permis de poursuivre ma thèse au sein du laboratoire « Centre de Recherche de Cancérologie de Marseille » et d'avoir soutenu mes demandes de financement, ainsi que les **Drs. Xavier Morelli** et **Yves Collette** de m'avoir accueillie dans leur équipe à l'issue de ma première année de thèse.

À la Fondation pour la Recherche Médicale

Au même titre, je remercie la **Fondation pour la Recherche Médicale (FRM)** pour son soutien financier et la confiance qu'elle a accordée à mes travaux de recherche, me permettant de réaliser ma quatrième année de thèse.

À Etienne Decroly et Karine Alvarez

Je suis également reconnaissante envers les **Drs. Etienne Decroly** et **Karine Alvarez** ma marraine de thèse de m'avoir soutenue et aidée depuis plus de cinq ans et plus particulièrement lors de mon changement de laboratoire. Sans vous je n'aurais sans doute pas continué ma thèse. Vous avez cru en moi et m'avez permis d'écrire une nouvelle page de ma vie.

À mes collaborateurs

Mes remerciements vont aussi à nos collaborateurs expérimentaux, à savoir le Consortium de **Génomique Structurale (SGC)** d'Oxford : les **Drs. Stefan Knapp**, **Susanne Müller-Knapp** et **Catherine Rogers** et la plateforme TrGET du CRCM : les **Drs. Yves Collette** et **Rémy Catellano**, ainsi qu'**Audrey Restouin** et **Camille Montersino**. Un grand merci à Audrey qui a réalisé les expériences de cellulaire au début du projet et qui a pris le temps de me former par la suite. Un grand merci également à Camille pour sa patience et tous ses conseils en l'absence d'Audrey. Je ne pensais pas un jour franchir l'étape de « Biologie Cellulaire », mais je peux maintenant dire je l'ai fait !

Remerciements à l'équipe

Ma plus profonde gratitude va à mon directeur de thèse Dr. Xavier Morelli.

À Xavier : Tu m'as fait confiance et tu as pris le risque de m'accueillir à l'issue de ma première année de thèse. L'intégration de ton équipe et l'opportunité que tu m'as donnée de travailler sur le projet « Bromodomaine » m'ont aidé à m'épanouir scientifiquement durant ces trois années passées à vos côtés. Tu as toujours été présent quand j'avais une question, un doute... Ton enthousiasme permanent et ta richesse scientifique a grandement contribué à la réussite de ce projet. J'ai pu participer à nombreuses formations et congrès et tu m'as également soutenue dans mes demandes de financement. Je te remercie pour toutes ces raisons, sans oublier pour ta patience et la relecture de mon manuscrit. J'espère réellement avoir été digne de l'investissement que tu as mis dans ma thèse et de la confiance que tu m'as accordée.

Je tiens également à remercier le **Dr. Yves Collette**, co-directeur de l'équipe pour son implication dans le projet de recherche, pour son regard scientifique et critique sur l'aspect « biologie » de ce projet de recherche.

À mes collègues

J'adresse toute ma gratitude à mes collègues qui ont participé à ce projet de recherche. Le **Dr. Sabine Milhas** qui a initié ce projet de recherche. Les chimistes à savoir le **Dr. Jean-Michel Brunel**, et plus particulièrement le **Dr. Sébastien Combes** et **Yulia Voitovich** qui ont synthétisés près de 500 molécules en trois ans. Les chémo-informaticiens à savoir les **Drs. Philippe Roche, Laurent Hoffer** et **Christophe Muller** (*Alias le commandant ou tu préférerais sans doute que je dise le BG. Que dire... sympas, drôle, taquin... Tu fais partie de la « team troll ».*), qui ont participé à l'optimisation de nos « hits ». Les Biochimistes et structuralistes de l'équipe à savoir **Carine Derviaux** pour l'HTRF, le **Dr. Stéphane Betzi** pour l'ITC et le **Dr. Etienne Rebuffet** pour l'aspect résolution des structures cristallographiques aux RX. Merci à vous tous de ne pas avoir baissé les bras, bien que les résultats tant espérés ne fussent pas souvent au rendez-vous. Tous, ensemble nous avons réussi à développer des molécules s'approchant des critères de sonde épigénétique définis par le SGC !! J'espère que ces molécules pourront aller plus loin dans des études de biologie plus approfondie et pourquoi pas en essais cliniques. Je remercie plus particulièrement Etienne, Laurent, Sébastien, Stéphane et Xavier pour la relecture de différentes parties de mon manuscrit.

Je tiens à également à remercier les autres membres de mon équipe, plus particulièrement le **Dr. Agnès Amouric** pour sa contribution en biologie moléculaire et la relecture du chapitre correspondant à cette discipline (*Attention c'est un dragon dans l'âme qui n'aime pas du tout le chocolat !!!!! Mais qui adore les kinder !!??*). Le **Dr. Hélène Launay** pour les discussions scientifiques, ou non, très enrichissantes, ses conseils et la relecture de mon « introduction bibliographique », et le **Dr Saez-Ayala** pour la relecture de mon introduction épigénétique.

Sans oublier, **Marie-Jeanne Basse** et le **Dr. Véronique Receveur**, ainsi que les derniers ayant rejoint notre équipe, le **Dr. Karine Barral**, le **Dr. Eddy Pasquier** (*Que dire... tu manges lentement à la cantine... tu as de nombreux papier... tu es destiné à devenir DR... À côté de toi on se sent tout petit... Bon courage pour la poursuite de mon projet de thèse avec Kendall, j'espère que le CHIP va donner des résultats intéressants*), **Jeremy Arieu-Bonnet** (*Tu es quelqu'un de sympas même si tu as des préjugés sur les tatouages lol. Tout comme Christophe, tu fais partie de la « team troll »*). *Bon courage pour ta thèse, je n'ai pas de doute avec Eddy tu vas avoir de supers papiers et tu vas tout déchirer !!*), le **Dr. Mikaël Feracci** (*Gentil, drôle, toujours prêt à aider et surtout tu as toi aussi des talents de pâtissier... miam. Tu fais également parti de la « team troll +++ »*), le **Dr. Bremond Paul** et **Kendall Carrasco** qui a débuté sa thèse en octobre (*Je te laisse mon bébé « bromodomaine » prend en grand soin... Bon courage pour ta thèse qui je pense devrait être toute aussi bonne voir meilleure que la mienne... Je compte sur toi pour que nos « pseudo-sondes » deviennent des « sondes » et surtout pour remotiver les troupes quand les résultats ne sont pas au rendez-vous*). Enfin **Marie-Amélie Maillard** stagiaire de passage dans l'équipe (*Tu es une fille adorable, discrète... Je te souhaite bon courage pour la suite de ton internat et ton job de maman avec ton nouveau petit bout à venir.*) J'espère n'avoir oublié personne, si tel est le cas je m'en excuse par avance.

À vous lecteurs, je vous remercie et j'espère que vous prendrez plaisir à lire ce manuscrit de thèse.

Spéciales dédicaces à la « team bioch »

Je tiens à remercier ceux qui sont devenus avec le temps plus que des collègues à savoir Stéphane, Carine, Elsa, Etienne et Magali. Sachez que j'ai passé à vos côtés trois merveilleuses années à vos côtés. Vous avez réussi à me faire oublier ma première année de thèse chaotique et à me sentir moins seule loin de Rémy. Vous resterez gravé à jamais dans mon cœur... j'espère ne pas trop déprimer à Toulouse.

Merci pour les discussions diverses et variées, les fous rires, les expéditions au relais H, les apéros ... Mojito sur la plage, suivi d'un restaurant et d'un coucher de soleil.... Cocktails dans les arbres... (Merci à Zezette pour cette magnifique découverte), repas au Royaume de la Chantilly.... Sans oublié que vous m'avez également initié à « Game of Thrones » !!!! ☺

À Stéphane Betzi « ITC Man » : **Stéphanou** comme te dirait Rémy, tu es mon « petit chef ». Je pensais être au point sur les présentations ppt avant mon arrivée dans l'équipe, mais ça c'était avant de te rencontrer. J'espère que la présentation de ma soutenance de thèse sera à la hauteur de tes exigences. ☺ Tu as été là pour m'accompagner, me guider dans les manips de manière générale et plus particulièrement l'ITC, avant que je prenne mon envol. Tu m'as souvent embêté pour Linux et l'organisation des data de cristallographie... Je tiens à te remercier pour ta patience car on ne se comprend pas toujours, tout ce que tu as pu m'enseigner, tes conseils pour améliorer mes figures. ☺ Un grand merci pour avoir relu mon roman...euh ma thèse, ce n'était pas évident mais heureusement qu'il y avait le post-it pour te rappeler de penser à moi. ☺ *Un souvenir : « Nos soirées apéro-dîner chez toi et la dégustation de bon vieux rhum de Martinique bien sûr ! ».*

À Carine Derviaux « HTRF Woman » : **Caribou** tu es une femme qui a du tempérament si bien qu'au début tu m'impressionnais et je me suis demandé si tu allais m'apprécier. En tout cas je suis ravie d'avoir fait ta connaissance ! Je pense que tu es celle avec qui j'ai passé le plus de temps au labo. Je te remercie de m'avoir formée à l'HTRF et d'avoir géré la programmation de ton super robot « Echo® ». L'HTRF c'est le pilier de ma thèse ! Merci pour ta patience, tes conseils et surtout pour ta bonne humeur et ton rire inoubliable. Ton absence a été longue, saches que tu nous a tous manqué !!! Il a donc fallu que je programme moi-même le robot et c'est bon maintenant je peux dire que j'ai su dompter l'Echo. ☺ *Un souvenir : « Quand tu nous as annoncé au Black Stone, un verre de jus de fruit à la main, que tu étais enceinte d'une petite Nora. C'était tellement drôle car j'avais déjà deviné depuis de quelques semaines déjà et qu'il commençait à avoir des rumeurs dans le bureau 1 ».* ☺

À Elsa Drivet « Cell Woman » : ZeZette tu étais la benjamine de l'équipe, comme une petite sœur. Ta joie de vivre se reflète sur l'équipe, tu mets du bonheur dans l'air. Toujours un air de musique à la bouche comme « Despasito 🎵🎵☀️ », pour ne pas dire un verre à la main. ☺ On n'était pas du même monde scientifique, toi « biologiste cellulaire » et moi « structuraliste »... Je disais « non la bio cell ce n'est pas pour moi » et pour finir cette année j'ai fait pas mal de manip de cellulaire, suite à l'absence d'Audrey, comme quoi. Merci pour ton petit cours sur la biologie cellulaire et sur les bonnes pratiques ou pas de manipulation en cellulaire. *Un souvenir : « Une bonne balade avec Stéphane, après une longue journée au labo au jardin de la Magalou pour une chasse aux Pokémons. »* ☺

À Etienne Rebuffet « Cristallo Man » : Titi ou Titou tu es un homme très discret qui fonctionne parfois sous forme d'onde sonar ☺ et qui a un talent pour la pâtisserie. Il n'est pas facile d'acquiescer ton savoir, il faut souvent te tirer les vers du nez pour comprendre ce que tu fais sur Linux, mais à force on apprend à te cerner. Je te remercie pour tes conseils cristallo et surtout de m'avoir enseigné les rudiments pour résoudre une structure RX. Linux et moi ça faisait deux... maintenant 1,5 lol. *Un souvenir : « Les pommes feuilletées en bouton de rose que tu nous a fait lorsque tu nous as annoncé que tu allais être papa d'une petite Charlotte ».*

À Magali Saez-Ayala « Super Woman » : Magalou tu es une espagnole au grand cœur mais qu'il ne faut surtout pas énerver. « Cálmesse Magalou ! » ☺ Je tiens à te remercier pour m'avoir écouté et soutenue pendant les nombreux moments difficiles rencontrés au cours de ma thèse. Tu pourras toujours compter sur moi-même dans les meilleurs et les pires moments. Saches que tu es une grande sœur pour moi, un modèle de travail. *Un souvenir : « Notre mission synchrotron qui a duré 24h... oui oui 24h..., cela nous a permis de créer une très bonne complicité ».*

Remerciements à la Famille

Plus personnellement,

À Rémy : mon amour cela fait treize ans que tu es entré dans ma vie et que tu me supportes. Tu as toujours été là pour moi et surtout dans les pires moments de ma vie, où tu savais toujours me faire retrouver le sourire. Je te remercie de m'avoir fait confiance et de m'avoir laissé réaliser mon rêve au prix de nombreux sacrifices dont celui de mettre en pause notre vie de couple pendant quatre ans. La distance entre nous, se voir le temps d'un week-end au maximum une fois par mois, le stress de rattraper une année de thèse perdue et d'obtenir des papiers, la fatigue accumulée, ont parfois généré des tensions entre nous, mais fort heureusement nous sommes toujours ensemble. Je pense que cette expérience aura renforcé notre couple. Saches que l'envie de tout lâcher pour rentrer à la maison à tes côtés m'a souvent traversé l'esprit, mais je voulais que tu sois fier de moi et de ne pas avoir fait des sacrifices pour rien c'est pourquoi j'ai tenu bon. # Mariage quand tu veux !

À ma famille : Sandra, Natacha, Alexandre, leurs conjoints Marc, Yann, Élodie et leurs enfants Océane, Lilly-Rose, Coline et Timothy. Merci d'avoir cru en moi, même dans les moments où moi-même je n'y croyais pas forcément. On ne se voit pas souvent, mais sachez que chaque instant passé avec vous me permet de recharger mes batteries en joie et en amour. Malgré la distance et les préoccupations, j'ai toujours pensé à vous. Je tiens tout particulièrement à remercier Chacha pour avoir été là pour me soutenir et m'encourager dans les différents moments difficiles que j'ai rencontrés au cours de ma thèse. Tu m'as aidée à ne pas perdre pieds et à me reconstruire. Merci également pour l'impression des invitations pour ma soutenance de thèse et de mon manuscrit. Les soirées passées au téléphone pour résoudre des problèmes de mise en page. Je suis fière de mon manuscrit et je le trouve très beau. Ainsi que mes défunts grand parents Gérard, Jacqueline, René et Hugnette, partie quand j'étais jeune et qui j'espère auraient été fiers de moi.

À ma belle-famille : Jean-Michel, Sylvie, Marine, Matthieu et la dernière arrivée dans la famille Mathilde, vous êtes plus qu'une belle-famille. Les triples M, bien sûr vous ne remplacez mes frères et sœurs, mais je vous considère tout comme. Spéciale dédicace à Marine, qui elle aussi m'a beaucoup écoutée et encouragée durant ma thèse. Merci d'avoir relu mon introduction et d'avoir corrigé certaines fautes...oups... Sylvie et Jean-Michel, dans mon cœur vous êtes mes parents d'adoption. Contrairement à mes propres parents, vous avez toujours été là pour moi dans les moments de joie ou de tristesse, vous m'avez permis de poursuivre mes études et m'avez toujours encouragé. Sans oublier mes grands-parents d'adoption Roger, ainsi que les défunts Georgette, Jeanne et Michel qui nous ont tous les trois quittés l'un après l'autre au cours de cette dernière année de thèse.

Sommaire

Table des matières

SOMMAIRE	13
TABLE DES MATIÈRES	15
LISTE DES ABRÉVIATIONS ET ACRONYMES	19
LISTE DES PUBLICATIONS, BREVETS ET STRUCTURES	23
AVANT-PROPOS	25
CHAPITRE 1 : INTRODUCTION	27
Partie 1 : Épigenétique & Cancers	31
I. Épigenétique	33
II. Cancers	47
III. Enjeux scientifiques	53
Partie 2 : Les Bromodomains, implication dans l'épigenétique et le développement de cancers	55
I. Les bromodomains	57
II. La famille BET « Bromo et Extra Terminal domain »	63
OBJECTIFS	99
CHAPITRE 2 : RÉSULTATS	105
Partie 1 : Criblage de la chimiothèque 2P2I3D : Identification et validation d'inhibiteurs des BETs	109
I. Développement de la chimiothèque 2P2I3D	111
II. Identification et validation d'inhibiteurs de PPIs	111
III. Conclusions & Perspectives	118
ARTICLE 1	119
Partie 2 : Compréhension du mécanisme d'action de la molécule BRD4(1) sélective « 5A14 » par SAR	125
I. Identification et Validation de la molécule 5A14	127
II. Caractérisation du mode de liaison de 5A14 à BRD4(1)	128
III. Validation de notre hypothèse	129
IV. Conclusions et perspectives	135
ARTICLE 2	137
Partie 3 : Optimisation par SAR de la molécule 5A14 : Sonde épigénétique sélective ?	143
I. Méthode d'optimisation originale : « S4MPLE	145
II. Premier dérivé sulfonamide issu de S4MPLE : « 5292 »	147
III. Deuxième série de dérivés sulfonamides de 5292 issues de S4MPLE	154

CHAPITRE 3 : CONCLUSIONS GÉNÉRALES & PERSPECTIVES	163
CHAPITRE 4 : PRINCIPES & PROTOCOLES EXPÉRIMENTAUX	175
Partie 1 : Biologie Moléculaire : Clonage & Transformation bactérienne	179
I. Plasmides utilisés	181
II. Clonage Gateway	181
III. Clonage par Assemblage Gibson	182
IV. Transformation des bactéries	183
Partie 2 : Production & Purification de protéines recombinantes	185
I. Production de protéines recombinantes	187
II. Purification de protéines recombinantes	188
III. Quantification de protéines recombinantes	195
Partie 3 : Équipements Scientifiques	199
I. Plateforme DOSynth	201
II. Plateforme IPC _{DD}	201
Partie 4 : Chimiothèques de composés	205
I. Chimiothèque de composés commerciaux	207
II. Composés synthétisés au sein de l'équipe	208
Partie 4 : Identification de « touches » par la technique d'Homogeneous Time Resolved Fluorescence (HTRF)	209
I. Principe de l'Homogeneous Time Resolved Fluorescence	211
II. Avantages et inconvénients de l'HTRF	213
III. Réactifs utilisés pour l'HTRF	214
IV. Protocole d'HTRF (manuel)	215
V. Automatisation des essais d'HTRF	216
VI. Analyse des résultats	221
Partie 6 : Validation orthogonale de « touches » par la technique de Thermal Shift Assay (TSA)	223
I. Principe du Thermal Shift Assay	225
II. Avantages et inconvénients du TSA	227
III. Protocole de TSA (manuel)	227
IV. Protocole de TSA au robot	228
V. Protocole de TSA (Bromoscan)	230
Partie 7 : Validation orthogonal de « touches » par la technique d'Isothermal Titration Calorimetry (ITC)	233
I. Principe de l'Isothermal Titration Calorimetry	235
II. Avantages et inconvénients de l'ITC	237
III. Protocole d'ITC (titration inverse)	237

Partie 8 : Viabilité cellulaire & Expression de c-Myc	239
I. Culture des lignées cellulaires	241
II. Viabilité cellulaire ou cytotoxicité (manuel)	242
III. Viabilité cellulaire ou cytotoxicité (au robot)	243
IV. Expression de c-Myc (western-blot)	243
V. Expression de c-Myc (HTRF)	244
Partie 9: Validation Cellulaire : Fluorescence Recovery After Photobleaching (FRAP)	247
I. Principe du Fluorescence Recovery After Photobleaching	249
II. Protocole du FRAP	249
Partie 10 : Caractérisation de complexes protéine-ligand par cristallographie aux rayons X	251
I. Définitions	253
II. Principe de la cristallogénèse	253
III. Criblage de condition de cristallisation	255
IV. Optimisation des conditions de cristallisation	256
V. Cristallisation des complexes protéine/ligand	257
VI. Avantages et inconvénients de la cristallographie	260
VII. Cryoprotection des cristaux	260
VIII. Collecte de jeux de diffractions aux RX	261
IX. Traitement des données de diffraction aux RX	262
X. Résolution et affinement des structures cristallographiques	263
XI. Analyse des contacts et génération de figures	265
XII. Dépôt des structures dans la PDB	266
LISTE DES FIGURES	267
LISTE DES TABLEAUX	269
ANNEXES	271
RÉFÉRENCES BIBLIOGRAPHIQUES	275

Liste des abréviations et acronymes

ADN : Acide désoxyribonucléique

ADP : Adénosine diphosphate

ADMETox : Absorption, Distribution, Métabolisation, Excrétion, Toxicité

AFMB : Laboratoire d'Architecture et Fonction des Macromolécules Biologiques de Marseille

AMM : Autorisation de mise sur le marché

ATP : Adénosine triphosphate

AML : Leucémie myéloïde aiguë

ARN : Acide ribonucléique

BET : « Bromodomain and Extraterminal-domain »

BCA : « Bicinchoninic acid assay »

BCP : « Bromodomain-containing protein »

BD : Module bromodomaine (domaine protéique)

BRET : « Bioluminescence Resonance Energy Transfer »

BSA : Albumine bovine sérique

CBS : Centre de Biochimie Structurale de Montpellier

CHIP-seq : Immunoprécipitation de la chromatine et séquençage

CMR : Cancérogène, mutagène, reprotoxique

CRCM : Centre de Recherche en Cancérologie de Marseille

CTD : « C-terminal domain »

DMSO : Diméthylsulfoxyde

DNMT : ADN méthyle transférase

DNdMT : ADN déméthyle transférase

DO : Densité optique

DOSynth : Diversity Oriented Synthesis

DTT : Dithiothréitol

EBV : Virus Epstein-Barr

EC₅₀ : Concentration efficace médiane

EDTA : Acide éthylène-diamine-tétra-acétique

ELISA : « Enzyme-linked immunosorbent assay »

EIPL: « Enzymology Interfacial & Physiology of Lipolysis »

ER : Récepteur à l'œstrogène
ESRF : « European Synchrotron Radiation Facility »
ET : « Extraterminal-domain »
FCS : Sérum de Veau Fœtal
FDA : « Food & Drug Administration »
FLAG : Étiquette avec peptide séquence DYKDDDK
FPLC : « Fast protein liquid chromatography »
FRET : « Förster ou Fluorescence Resonance Energy »
FRAP : « Fluorescence Recovery After Photobleaching »
FRM : Fondation pour la Recherche Médicale
GFP : « Green fluorescent protein »
GST : Glutathion S-transférase
GSH : Glutathion réduit
GSK : Glaxo Smith Kline
HAT : Histone acétylase
HDAC : Histone désacétylase
HIV : Virus de l'immunodéficience humaine
HKMT : Histone lysine méthyle transférase
HKDM : Histone lysine déméthylase
HMT : Histone méthyle transférase
HTRF : « Homogeneous Time Resolved Fluorescence »
HPV : Papillomavirus humain
IC₅₀ : Concentration inhibitrice médiane
IPC_{DD} : Institut Paoli Calmette Drug-Discovery
IPTG : Isopropyl β-D-1-thiogalactopyranoside
iSCB : « integrative Structural & Chemical Biology »
ITC : « Isothermal Titration assay »
KDM : Lysine déméthylase
KSHV : Kaposi's Sarcoma-associated Herpes Virus
LB : Luria Broth
LR reaction : Réaction de recombinaison homologue
MACE : « Major Adverse Cardiac Events »

MBD : « Methyl-CpG binding domain »
MBP: « Maltose-Binding Protein »
MBT : « Malignant brain tumor »
MLL : Leucémie myéloïde lymphoblastique
NMC : « Nut Midline Carcinoma »
NSCLC : « Non-small Cell Lung Cancer »
Pan-BET : l'ensemble des membres de la famille BET (8 BDs, 4BCPs)
PBS : Tampon phosphate salin
PDB : « Protein Data Bank »
PHD : « Plant homo-domain »
PMSF : Fluorure de phénylméthylsulfonyle
PPI : Interactions protéine-protéine
PRMT : Arginine méthyle transférase
PTM : Modifications post-traductionnelles
RMN : Résonance magnétique nucléaire
S4MPLE : « Sampler for Multiple Protein-Ligand Entities »
SAR : « Structure Activity Relationship » - Relation structure/fonction
SDS-PAGE : « sodium dodecyl sulfate polyacrylamide gel electrophoresis »
SGC : « Structural Genomic Consortium »
SIDA : Syndrome d'immunodéficience acquise
SLIC : « Sequence- and ligation-independent cloning »
SPR : « Surface Plasmon Resonance »
TA : Température ambiante
TBST : tampon salin tris et Tween 20
TEV : « Tobacco etch virus »
TSA : « Thermal shift assay »
TTD : « Tadem tudor-domain »
UV : Ultraviolet

Liste des publications, brevets et structures

Publications

Article1 : Protein-Protein Interaction Inhibition (2P2I)-Oriented Chemical Library Accelerates Hit Discovery.

Milhas S, **Raux B**, Betzi S, Derviaux C, Roche P, Restouin A, Basse MJ, Rebuffet E, Lugari A, Badol M, Kashyap R, Lissitzky JC, Eydoux C, Hamon V, Gourdel ME, Combes S, Zimmermann P, Aurrand-Lions M¹, Roux T, Rogers C, Müller S, Knapp S, Trinquet E, Collette Y, Guillemot JC, Morelli X.

ACS Chem Biol. 2016 Aug 19;11(8):2140-8.

Article2 : Exploring Selective Inhibition of the First Bromodomain of the Human Bromodomain and Extra-terminal Domain (BET) Proteins.

Raux B, Voitovich Y, Derviaux C, Lugari A, Rebuffet E, Milhas S, Priet S, Roux T, Trinquet E, Guillemot JC, Knapp S, Brunel JM, Fedorov AY, Collette Y, Roche P, Betzi S, Combes S, Morelli X.

J Med Chem. 2016 Feb 25;59(4):1634-41.

Article 3 : En cours de rédaction

Brevet

INSERM : EP 15 307 163.4

CNRS : 09276-022

Structures

5DLX, 5DLZ, 5EGU, 5EI4, 5EIS + une vingtaines de structures à déposer dans la PDB

Articles non présentés dans ce manuscrit

Article 4 : Supported inhibitor for fishing lipases in complex biological media and mass spectrometry identification.

Delorme V, **Raux B**, Puppo R, Leclaire J, Cavalier JF, Marc S, Kamarajugadda PK, Buono G, Fotiadu F, Canaan S, Carrière F.

Biochimie. 2014 Dec;107 Pt A:124-34.

Article 5 : Slowing down fat digestion and absorption by an oxadiazolone inhibitor targeting selectively gastric lipolysis.

Point V, Bénarouche A, Zarrillo J, Guy A, Magnez R, Fonseca L, **Raux B**, Leclaire J, Buono G, Fotiadu F, Durand T, Carrière F, Vaysse C, Couëdelo L, Cavalier JF.

Eur J Med Chem. 2016 Aug 8;123:834-848

Avant-propos

Les travaux de recherches présentés dans ce manuscrit ont été réalisés dans le laboratoire Centre de Recherche en Cancérologie de Marseille (CRCM) dirigé par le Pr. Jean-Paul Borg et dans l'équipe intégrative Structural & Chemical Biology (iSCB) codirigée par les Drs. Yves Collette et Xavier Morelli.

L'équipe, s'intéresse plus particulièrement à l'étude des interactions de type protéine-protéines (PPIs) et travaille au développement de petites molécules inhibitrices de PPIs utilisant des approches combinées de biochimie, biophysique, biologie structurale et biologie cellulaire. La force de notre équipe pluridisciplinaire repose sur le fait qu'elle est composée de chimistes, chémo-informaticiens, biochimistes-biologistes et structuralistes et que nous disposons de nombreux outils nécessaires au développement d'inhibiteurs PPIs.

Mes travaux de thèse s'inscrivent dans le projet « Bromodomaine » développé dans l'équipe. Le sujet d'étude de ma thèse portait sur *l'identification et la caractérisation d'inhibiteurs sélectifs des protéines à bromodomains de la famille BET*, à l'aide des divers outils disponibles au laboratoire.

Ce projet de recherche a été réalisé en étroite collaboration avec les chimistes (les Drs. Sébastien Combes et Jean-Michel Brunel, ainsi qu'une étudiante en thèse Yulia Voitovich), les chémo-informaticiens (les Drs. Philippe Roche, Laurent Hoffer et Christophe Muller) et les structuralistes de l'équipe (les Drs. Stéphane Betzi et Etienne Rebuffet). Nous avons également collaboré pour certaines études cellulaires avec la plateforme TrGET du CRCM (Dr Rémy Castellano, Audrey Restouin et Camille Montersino), ainsi que les Drs. Stefan Knapp, Susanne Müller-Knapp et Catherine Rogers du Consortium de Génomique Structurale (SGC, Oxford).

Mon travail au sein de l'équipe a consisté en **l'identification** par **HTRF (Homogeneous Time Resolved Fluorescence)**, la **validation orthogonale** par **ITC (Isothermal Titration Calorimetry)** et **TSA (Thermal Shift Assay)**, la **validation cellulaire** (viabilité cellulaire de lignées cancéreuses, expression de l'oncogène c-Myc et **FRAP Fluorescence Recovery After Photobleaching**) et la **caractérisation structurale** par **cristallographie aux rayons X d'inhibiteurs sélectifs des protéines à bromodomains de la famille BET**.

Cette thèse a été financée dans un premier temps par une bourse ministérielle de l'école Doctorale des Sciences Chimiques de Marseille (ED250) qui a été transférée à l'école Doctorale des Sciences de la Vie et de la Santé de Marseille (ED62) lors de mon changement de laboratoire d'accueil et de sujet de recherche la fin de ma première année de thèse. La quatrième année a quant à elle été financée par la **Fondation pour la Recherche Médicale (FRM)**.

Chapitre 1 : Introduction

Ce chapitre a pour but de replacer mon projet de thèse intitulé « **Développement d'inhibiteurs de bromodomaines (BDs) : modules de reconnaissance impliqués dans l'épigénétique et le développement de cancers** » dans son contexte.

La première partie intitulée « **Épigénétiques & Cancers** », présentera les différents facteurs qui participent à la régulation de l'expression des gènes, qui sont héréditaires, mais qui peuvent être modifiés au cours de la vie. La perte de régulation de ces facteurs conduit à des maladies graves comme le cancer. Dans un premier temps je décrirai les facteurs épigénétiques qui régulent l'activation et la répression de la transcription des gènes. Parmi les nombreuses pathologies qui résultent de la dérégulation de ces facteurs épigénétiques, on trouve les cancers sous différentes formes que je présenterai par la suite.

La seconde partie intitulée « **Les Bromodomaines : implication en épigénétique et le développement de cancers** », s'intéressera aux protéines à bromodomaines qui sont des régulateurs épigénétiques, dont la surexpression conduit au développement de diverses pathologies et notamment les cancers. Dans un premier temps, je présenterai de manière générale les protéines à bromodomaines qui sont classifiées en huit familles distinctes avant de me focaliser sur l'une d'entre elles, « la famille BET », pour laquelle je détaillerai le rôle de ces protéines dans la régulation épigénétique et l'impact de la dérégulation de leur expression dans le développement de maladies. Par la suite, je m'intéresserai à l'architecture des protéines BETs, avant de me focaliser sur leurs motifs « bromodomaines » ou BDs. Enfin, je présenterai de manière non exhaustive différentes classes d'inhibiteurs de BETs publiées et en essais cliniques.

Partie 1 : Épigénétique & Cancers

I. Épigénétique

1. Introduction

Le corps humain est constitué d'environ 10^{13} cellules (selon la littérature entre 10^{12} et 10^{16} ; (1)). Chacune d'entre elles contient en son noyau la même **information génétique** qui est codée par l'**ADN** (acide désoxyribonucléique) et pourtant, ces cellules remplissent différentes fonctions (exemples : cellules épidermiques, musculaires, neuronales, osseuses, sanguines, souches...). L'information génétique qui définit l'ensemble des caractères conduisant aux **génotypes** et **phénotypes** d'un individu est héréditaire et transmissible à la descendance. Par ailleurs, les **facteurs environnementaux** sont capables d'induire des **mutations de la séquence d'ADN ou des modifications des protéines associées à l'ADN, « les histones »**, plus ou moins importantes altérant le phénotype d'un individu.

En 1942, **Waddington** définit pour la première fois le terme « **Épigénétique** », comme étant « l'étude des relations de cause à effet entre les gènes et leur produits, qui font apparaître un phénotype » (2). Depuis, cette définition est en perpétuelle évolution du fait des avancées scientifiques dans le domaine de l'épigénétique. L'épigénétique a été décrit comme une « *frontière entre le génotype et le phénotype* » par **Goldberg** et **Allis** (3). Aujourd'hui, ce terme est défini comme « *l'étude des changements héréditaires portés par les gènes qui conduisent à un nouveau phénotype sans que la séquence primaire d'ADN ne soit altérée* » (4). Ces caractères sont déterminés par des **modifications chimiques covalentes réversibles, voir reprogrammables selon le type de cellule**. Ces modifications se produisent en dehors de la séquence d'ADN et contrôlent l'expression et la fonction des gènes (4).

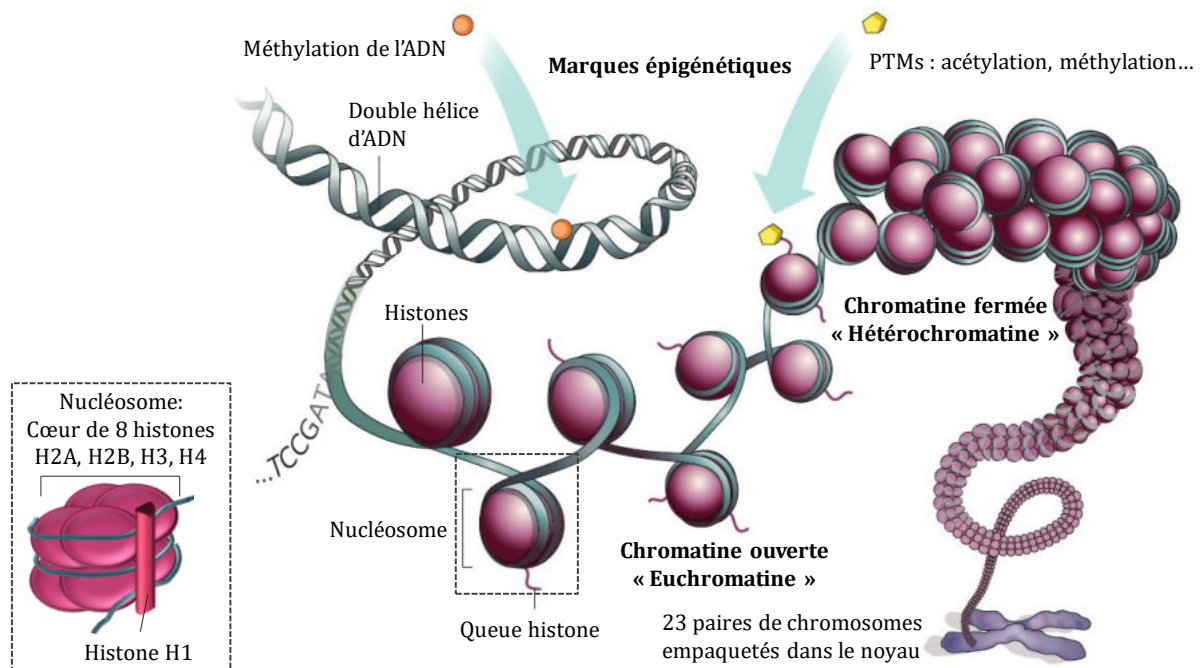
Pour mieux appréhender comment l'épigénétique est impliquée dans la régulation de l'expression de gènes tout au long de la vie, il est important de comprendre l'architecture de la chromatine.

La double hélice d'ADN découverte en 1953 par **Watson** et **Crick** (5), représente environ six milliards de paires de bases, soit deux mètres d'ADN étiré. Pour pouvoir être contenue dans le noyau des cellules qui ne représente qu'une centaine de μm , celle-ci doit être extrêmement compactée sous forme chromatine. Le super-enroulement de l'ADN est réalisé par des topoisomérases (6) et les protéines histones permettent quant à elles le maintien de l'assemblage de la chromatine (7). La chromatine est constituée de plusieurs unités répétées de nucléosome (8) formant ainsi le premier niveau de structuration de la chromatine (11 nm). La structure du nucléosome a été résolue (7 Å) pour la première fois en 1984 par **Richmond** (9), puis plus récemment à plus haute résolution (2,8 Å) par **Luger** (PDB : **1AOI**) (8).

Chaque nucléosome est formé d'un octamère d'histones autour duquel s'enroule 147 paires de bases de la double hélice d'ADN, soit environ deux tours de la double hélice (10-12). Les histones sont des petites protéines très riches en acides aminés basiques (arginine et lysine) chargés positivement à pH physiologique ce qui facilite leur interaction avec l'ADN qui est quant à lui chargé négativement par les groupements phosphates. Il existe cinq types d'histones : H1, H2A, H2B, H3 et H4. L'histone H1 permet l'association des nucléosomes entre eux (13), et la compaction chromatine (14). Les histones H2A, H2B, H3, H4 forment quant à elles l'octamère d'histones qui est le cœur du nucléosome. Les domaines globulaires des histones H2A, H2B, H3, H4 sont capables d'interagir entre eux pour former des hétérodimères H2A-H2B et H3-H4, qui vont ensuite s'associer pour former des tétramères pour enfin aboutir à l'octamère (8, 10, 11). Le nucléosome est maintenu grâce à des interactions électrostatiques, des liaisons hydrogènes entre les histones et les groupements phosphates de l'ADN, ainsi que par des contacts non polaires entre les histones et le sucre désoxyribose de l'ADN (15). La queue N-terminale des histones située à l'extérieur du nucléosome n'est pas structurée et elle est sujette à des modifications chimiques (15).

L'accessibilité de l'ADN aux facteurs de transcription est déterminée par les différents niveaux d'organisation de la chromatine : de la séquence primaire faisant 2 nm, au chromosome faisant 1,4 μ M (15, 16). La transcription des gènes est liée aux changements de conformation de la chromatine (formes condensée/relâchée). Le relâchement de la chromatine sous forme « euchromatine » permet de rendre l'ADN accessible aux facteurs de transcription et de ce fait d'activer la transcription des gènes. À l'inverse, la compaction de la chromatine sous forme « hétérochromatine » empêche les facteurs de transcription d'accéder à l'ADN ce qui induit la répression de la transcription des gènes (17, 18). Ce changement de conformation (formes condensée inactive et relâchée active) est un processus très dynamique rendu possible grâce à des **modifications chimiques covalentes réversibles, de l'ADN et des protéines structurales de la chromatine « les histones »**. Ces modifications chimiques au niveau de la chromatine sont appelées « **marques épigénétiques** » (Figure 1).

Figure 1 : Compaction de l'ADN.



Légende : Figure adaptée de Marx et al, (17). Le nucléosome est constitué d'un octamère d'histones (H2A, H2B, H3 et H4, en deux copies chacune) autour duquel s'enroule l'ADN (147 pb) et d'une histone H1 qui se lie aux extrémités 3' et 5' terminales de l'ADN. La méthylation de l'ADN, ainsi que l'ajout de marques épigénétiques au niveau des protéines histones permet le passage de la chromatine de la forme fermée (condensée) dite « hétérochromatine » à la forme ouverte (relâchée) dite « euchromatine ». Ce changement de conformation joue un rôle très important dans la régulation de la transcription des gènes.

2. Les marques épigénétiques

Nous avons vu que l'expression des gènes est déterminée par les différents états de compaction de la chromatine, qui sont rendu possibles grâce aux marques épigénétiques apportées à l'ADN et aux protéines histones. Cette partie présentera brièvement et de manière non exhaustive les deux grands types de modifications épigénétiques : **i) la méthylation de certaines bases azotées de l'ADN** et **ii) les modifications post-traductionnelles de certains résidus d'acides aminés des histones**, comme la méthylation ou l'acétylation des lysines.

a. La méthylation de l'ADN

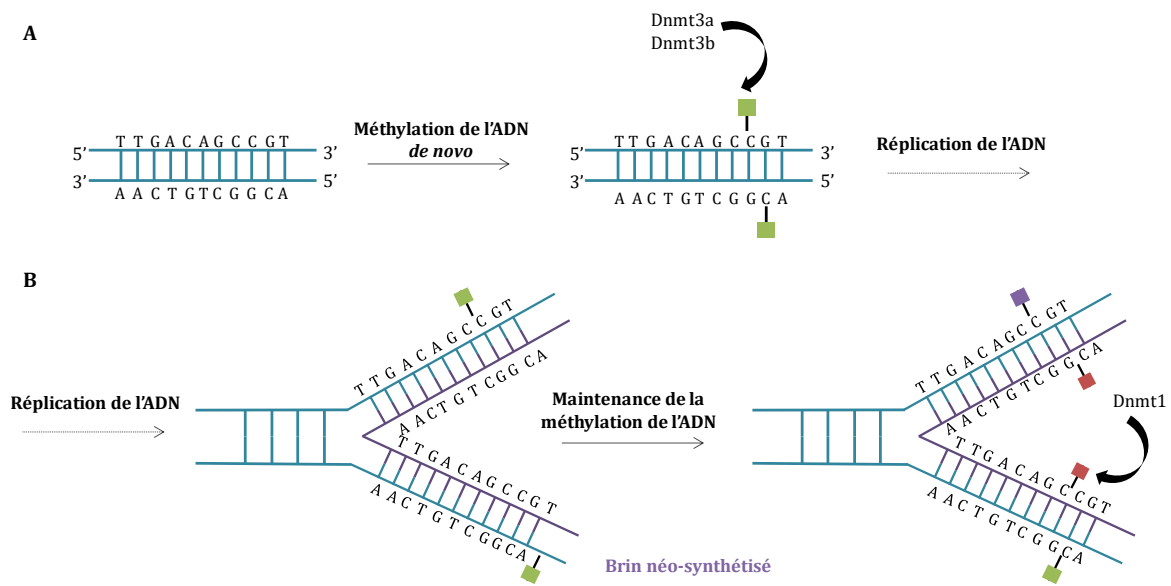
La méthylation de l'ADN consiste à introduire un groupement méthyle (CH_3) sur le carbone 5 des résidus de cytosine (5-mC). Cette modification épigénétique joue un rôle important dans le développement embryonnaire, la transcription, l'inactivation de l'un des deux chromosomes X chez la femme et l'empreinte génétique (19).

Chez les mammifères, la méthylation de l'ADN se produit principalement dans les régions riches en dinucléotides palindromiques 5'-CpG-3' ou îlots CpG, comme la région promotrice des gènes (extrémités 5') et/ou le premier exon. La méthylation de l'ADN est réalisée par l'ADN méthyltransférase (**DNMTs**) de manière dépendante à la S-adénosyl-méthionine (20). Il existe deux types de DNMTs (**Figure 2**) : i) Les méthyltransférases de novo DNMT3a et DNMT3b, permettent la méthylation d'un ADN « nu » avant le processus de réplication, notamment lors de l'embryogenèse et le développement des cellules germinales (21). ii) La méthyltransférase de maintien DNMT1, permet à l'issue de la réplication de passer d'un ADN hémi-méthylé à un ADN méthylé à l'issue de la réplication (19, 22, 23). La méthylation de l'ADN est un moyen pour la machinerie cellulaire de contrôler de la fidélité de réplication. En effet lors de la réplication, le brin « fille » est néo-synthétisé, puis il est contrôlé pour vérifier qu'il n'y a pas d'erreurs de réplication ou des mésappariements et enfin il est méthylé.

La méthylation des bases nucléotides des îlots CpG sur les promoteurs des gènes est reconnue par des protéines à motif methyl-CpG binding domains (**MBDs**). Ces protéines participent au maintien de la répression de la transcription *via* le recrutement de complexes enzymatiques contenant des histones méthyltransférases (**HMTs**) et des histones désacétylases (**HDACs**) dont l'activité conduit à la condensation de la chromatine (24). Une hyperméthylation de la région promotrice des gènes est donc associée à la répression de la transcription et inversement un gène hypométhylé sera transcrit activement. Pour que l'ADN soit accessible aux facteurs de transcription, les déméthylases (**DNdMTases**) telles que TET1, TET2 et TET3 doivent supprimer les modifications 5-mC. La protéine TET1 joue un rôle dans l'activation et la répression de la transcription, tandis que TET2 intervient dans la synthèse de gènes suppresseurs de tumeurs et TET3 permet la reprogrammation de la méthylation de l'ADN (25).

D'autres modifications des bases azotées cytosines ont été décrites dans la littérature : la 5-hydroxyméthylcytosine qui est impliqué dans la transcription et les 5-formylcytosine et 5-carboxylcytosine pour lesquelles les fonctions ne sont pas encore connues (26).

Figure 2 : Méthylation de l'ADN.



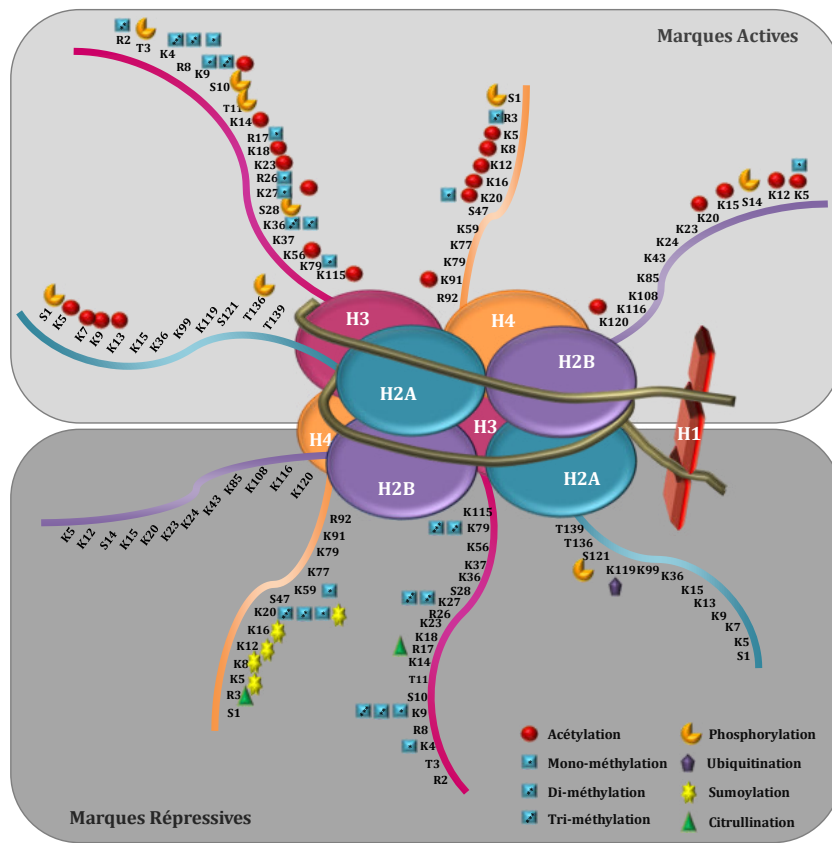
Légende : A- La méthylation (carré vert) de novo de l'ADN « nu » (brin bleu) est réalisée par les méthyltransférases DNMT3a et DNMT3b. B- Lors de la réplication, le brin d'ADN néo synthétisé (brin violet) est méthylé (carré rouge) par la méthyltransférase DNMT1 afin de maintenir la méthylation de l'ADN.

b. Les modifications post-traductionnelles des histones

À ce jour, plus de 519.327 modifications post-traductionnelles (PTMs) différentes ont été répertoriées (le 22/08/2017, <https://www.phosphosite.org/homeAction.action>) (27). Parmi les plus décrites dans la littérature on retrouve l'acétylation (Ac) des lysines, la méthylation (Me) des lysines et arginines, la phosphorylation (P) des sérines, thréonines et tyrosines et l'ubiquitination (Ub) des lysines. Notons qu'il existe d'autres PTMs, à savoir l'ADP-ribosylation, la biotinylation, la citrullination, la crotonylation, l'isomérisation cis-trans des prolines, la SUMOylation... Grâce à la spectrométrie de masse, le nombre de PTMs identifiées ne cesse d'augmenter (28-30). Les PTMs des histones sont impliquées dans la régulation de la transcription des gènes *via* le remodelage de la chromatine, ainsi que la réplication de l'ADN et la réparation des dommages à l'ADN (15, 31-34).

La figure ci-dessous (**Figure 3**) présente diverses PTMs retrouvées au niveau des queues des histones permettant d'activer ou de réprimer la transcription des gènes. Ces modifications sont réversibles grâce à l'action de plusieurs couples d'enzymes dont l'action agoniste ou antagoniste, est contrôlée par des protéines de type arrestine et des complexes protéiques coactivateurs ou corépresseurs de la transcription.

Figure 3 : Modifications post-traductionnelles des protéines histones.



Légende : Inspiré de Sawan et al., (35). Cette figure liste certaines modifications épigénétiques telles que l'acétylation, la méthylation, la phosphorylation, l'ubiquitination, la SUMOylation et la citrullination, impliquées dans l'activation (panneau du haut) et la répression (panneau du bas) de la transcription de gènes.

i. Le code « histone »

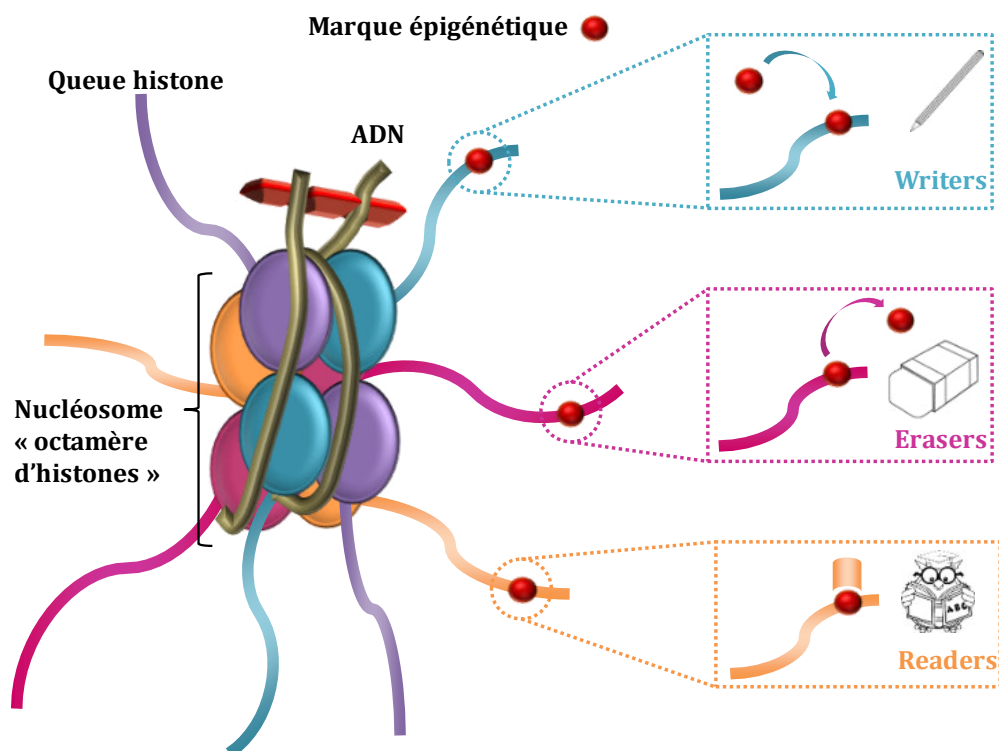
La modification post-traductionnelle des histones ne se produit pas individuellement, mais de manière combinée et cette combinaison constitue ce qu'on appelle le « code histone » (36). L'ensemble des PTMs est lu et interprété par la machinerie cellulaire qui va activer une cascade de réactions biologiques. Le « code histone » est très complexe pour différentes raisons (37) : i) Le nombre de combinaison possible au sein d'un nucléosome est évalué à plus d'un milliard. ii) La modification d'un résidu peut activer ou bien réprimer la modification de résidus voisins. Par exemple, la phosphorylation d'une sérine a pour effet secondaire d'inhiber la méthylation d'une lysine située en amont de la séquence et d'activer l'acétylation d'une lysine située en aval. iii) Une même histone peut subir différentes modifications en fonction de son emplacement (régions non codantes 5' ou 3' et région codante). iv) Les modifications des histones induisent le recrutement de protéines spécifiques qui vont interagir avec un résidu modifié bien précis, par exemples les protéines Trithorax et Polycomb qui respectivement, activent et répriment la transcription de gènes (31, 36, 38-40).

ii. Acteurs épigénétiques

Il existe différents types de régulateurs (41) permettant d'ajouter, de supprimer ou de lire les marques épigénétiques au niveau des queues des protéines histones (**Figure 4**).

- ❖ Les **writers** permettent de lier de manière covalente les marques épigénétiques, par exemple les histones acétyltransférases (HATs) sont responsables de l'acétylation des résidus lysines des queues des histones.
- ❖ Les **erasers** suppriment les marques épigénétiques, par exemple les histones désacétylases (HDACs) retirent les groupements acétyles des queues des histones.
- ❖ Les **readers** reconnaissent quant à eux spécifiquement les marques épigénétiques, par exemple les protéines à bromodomains sont capables de reconnaître les lysines acétylées des queues des histones et de recruter des facteurs de transcriptions.


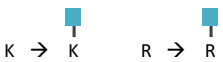





Figure 4 : Régulateurs épigénétiques : « writers », « erasers » et « readers ».



Légende : Il existe trois types de régulateurs épigénétiques : les « writers » qui ajoutent des marques épigénétiques, les « erasers » qui les suppriment et les « readers » qui les reconnaissent et sont capables de recruter des partenaires tels que les facteurs de transcriptions.

Parmi les writers, on trouve les histones acétyltransférases (**HATs**) et les histones méthyltransférases (**HMTs**) qui réalisent respectivement l'acétylation des lysines et la méthylation des lysines et arginines, ainsi que les kinases et les phosphorylases qui réalisent la phosphorylation des résidus sérines, thréonine et tyrosines. Parmi les erasers, on trouve les histones désacétylase (**HDAC**) et les histones déméthylase (**HKDMs**) qui suppriment respectivement les groupements acétyles et méthyles des résidus lysines, ainsi que les phosphatases qui suppriment les groupements phosphates. Enfin, parmi les readers, on trouve méthyl-lysine transférases et les methyl-arginine binding domain containing protein qui reconnaissent les lysines et arginines méthylées des histones, les PHD-containing proteins (**PHD**) qui reconnaissent les lysines et arginines méthylées des protéines histones et non-histones et les Bromodomain-containing protein (**BCP**) qui reconnaissent spécifiquement les lysines acétylées des histones. Le tableau ci-dessous (**Tableau 1**), liste de manière non exhaustive, divers régulateurs épigénétiques et leur fonction (41).

Tableau 1 : Régulateurs épigénétiques.

Famille	Activité	Nombre de protéine	Classes de protéine et fonctions
Writers			
Histone acétyltransférases (HATs)		8	MYST : dommages à l'ADN et translocation oncogénique GNAT : signalisation EGF, progression cycle cellulaire EP300 : divers processus cellulaires
Protein méthyltransférases (HMTs)		60	Domain SET : méthylation des lysines des protéines histones et non-histones PRMTs : méthylation des arginines des protéines histones et non-histones PRDMs : domaine SET like, tissu spécifique
Erasers			
Histone déacétylases (HDACs)		17	Classe I, IIb et IV : Substrat protéines histones et non-histones, répression de gènes Classe IIa : repliement des protéines Classe III (Sirtuins) : activités de déacétylation et d'ADP-ribosylation
Lysine déméthylases (KDMTs)		25	Demethylase lysine spécifique : régulation de la transcription au cours du développement Domaine Jumonji
Readers			
Bromodomain-containing protein (BCPs)		61	Se lie à la chromatine de manière spécifique aux lysines acétylées
Methyl-lysine transferase Ou methyl-arginine binding domain containing protein		95	Domaines Tudor : se lie aux lysines di- et tri-méthylées et aux arginines di-méthylées Domaines MBT : se lie aux lysines mono- et di-méthylées, avec une faible spécificité de séquence Chromodomains : se lie aux lysines tri-méthylées, avec une spécificité de séquence Domaines PWWP : se lie aux lysines tri-méthylées et à l'ADN
PHD-containing proteins		104	De nombreuses familles agissent sur divers substrats

Légende : Tableau adapté d'Arrowsmith et al, (41).

iii. L'acétylation des histones

L'**acétylation** des histones consiste à introduire un groupement acétyle (COCH_3) sur le groupe ϵ -amino des résidus lysines. Ce processus a été décrit pour la première fois en 1964 par **Allfrey et al** (42). C'est l'une des modifications post-traductionnelles les plus fréquentes, avec plus de 38.121 sites répertoriés à ce jour (le 22/08/2017, <https://www.phosphosite.org/homeAction.action>) (27).

Elle joue un rôle important dans la modulation de la structure de la chromatine, la régulation de la transcription des gènes via le recrutement de facteur de transcription et la réparation de l'ADN. En outre, elle régule diverses fonctions biologiques telles que la différenciation cellulaire et l'apoptose (31, 43, 44).

Le niveau cellulaire d'acétylation est maintenu grâce à l'action de deux types d'enzymes : les Histones Acétyle Transférase (**HATs**, writers) qui les ajoutent et les Histones DésAcétylase (**HDACs**, erasers) qui les suppriment.

Les dix-huit HAT codées par le génome humain (45) sont classées en trois familles : les protéines de la famille MYST (exemple : la protéine MOZ-Ybf2/Sas3-Sas2&Tip60 »-related HATs), sont impliquées dans la réparation des dommages à l'ADN et de la translocation oncogénique. Les protéines de la famille GNAT (exemple : la protéine GCN5 related N acetyltransférase) sont, quant à elles, impliquées dans la signalisation d'EGF et la progression du cycle cellulaire. Enfin, les protéines de la famille EP300/CREBBP, jouent également un rôle dans divers processus cellulaires (41).

L'acétylation permet de neutraliser la charge positive des résidus lysines des histones et induit un encombrement stérique. La diminution des interactions électrostatiques entre les résidus lysines et l'ADN, permet à la chromatine d'être plus flexible (euchromatine), de s'ouvrir pour rendre l'ADN plus accessible aux facteurs de transcription (41, 46-50).

Il existe de nombreux sites d'acétylation au niveau des histones et chaque modification est liée à un rôle biologique spécifique. Par exemples, l'acétylation de la lysine en position 12 de l'histone H4 (H4K12_{ac}) joue un rôle lors du développement embryonnaire et est impliquée dans la perte de mémoire liée à l'âge (51). L'acétylation de la lysine en position 14 de l'histone H3 (H3K14_{ac}) joue un rôle dans l'architecture de la chromatine et dans la réparation des dommages à l'ADN (52-55). L'acétylation de la lysine en position 16 de l'histone H4 (H4K16_{ac}) joue un rôle important dans le changement de conformation de la chromatine de la forme inactive dite « hétérochromatine » à la forme active dite « euchromatine » (55).

Le génome humain code dix-sept HDAC différentes (45) qui sont divisées en quatre classes (I à IV). Les protéines de la classe I, IIb et IV sont capables de désacétyler les protéines histones et non-histones et jouent un rôle dans la répression de la transcription des gènes. Les protéines de la famille IIa sont impliquées dans le repliement des protéines. Enfin, les protéines de la famille III ont une double activité enzymatique, elles permettent la désacétylation et l'ADP-ribosylation. La désacétylation a pour effet de restaurer la charge positive des lysines, ce qui stabilise localement la structure de la chromatine sous forme condensée (hétérochromatine) réprimant ainsi la transcription (38, 40, 41, 56).

Il existe des protéines capables de reconnaître spécifiquement les histones et les protéines de transcription acétylées. C'est le cas des Bromodomain-Containing Protein (**BCP**, reader) (57).

iv. La méthylation des histones

La méthylation des histones consiste à introduire un groupement méthyle (CH₃) sur les groupes ε-amino des résidus arginines et lysine. C'est l'une des PTMs les plus complexes, car ces résidus peuvent être mono ou poly-méthylés.

Cette modification épigénétique est impliquée l'activation et la répression de l'expression des gènes, en fonction de la position de la méthylation (58, 59). L'activation de la transcription est liée à la tri-méthylation des lysines en position 4 et 36 de l'histone H3 (H3K4_{me3} et H3K36_{me3}) et à la di-méthylation de l'arginine en position 17 de l'histone H3 (H3R17_{me2}) (60). La répression de la transcription est liée à la méthylation de la lysine en position 9 de l'histone H3 (H3K9_{me}) et la tri-méthylation de la lysine en position 27 de l'histone H3 (H3K27_{me3}) (61) et la di-méthylation de l'arginine en position 2 de l'histone H3 (H3R2_{me2a}) (62, 63). Par ailleurs, elle est souvent couplée à la désacétylation de ces mêmes lysines.

La méthylation des histones est effectuée par deux types d'Histones Méthyl Transférases (**HMTs**, writer) : les Histones Lysines MéthylTransférases (**HKMTs**) ou protéine à domaine SET qui permettent la mono-, di- ou tri-méthylation des lysines des protéines histones et non-histones et les Protéines Arginine MéthylTransférases (**PRMTs**) qui permettent la mono- et di-méthylation des arginines des protéines histones et non-histones (64, 65). La mono-méthylation des résidus histidines a également été décrite, mais elle reste une modification rare (66).

La déméthylation des lysines est réalisée par les histones lysines déméthylases (**HKDMs**, erasers) (65). Aucune enzyme déméthylase d'arginines n'est à ce jour connue, cependant la citrullination ou déimination semble être corrélée avec la disparition des arginines méthylées (31).

Certaines protéines readers sont capables de reconnaître les lysines et/ou les arginines méthylées. Les Malignant Brain Tumor (**MBT**) se fixent sur les lysines mono- et di-acétylée(s) avec une faible spécificité de séquence, tandis que les **Chromodomaines** se fixent sur les lysines tri-méthylées au niveau de séquences spécifiques. Les **domaines Tudor** sont capables de se fixer à la fois sur les lysines di- et tri-méthylées et sur les arginine di-méthylées. Les Plant HomoDomain (**PHD**) et les répétitions **WD40** sont également capables de reconnaître les lysines et les arginines méthylées. Enfin les **domaines PWWP** sont capables de se fixer sur les lysines tri-méthylées et sur l'ADN (41, 67).

v. *La phosphorylation des histones*

La phosphorylation des histones consiste à introduire un groupement phosphate (PO_3^{2-} , provenant de l'ATP) sur les groupes hydroxyle (OH) des résidus sérines, thréonines et tyrosines. La phosphorylation des histones joue un rôle dans l'activation de la transcription (68), probablement parce qu'elle crée une force répulsive entre les charges négatives des résidus phosphorylés et celles de l'ADN. Cette répulsion électrostatique permet la décondensation de la chromatine et donc l'accessibilité de l'ADN aux facteurs de transcription. La phosphorylation des histones permet également la condensation des chromosomes mitotiques (exemple : la phosphorylation de la serine en position 10 de l'histone H3 ou H3S10_p (69), la réparation de l'ADN (exemple : la phosphorylation de la serine en position 129 de l'histone H2A ou H2AS129_p) ou encore l'apoptose des cellules (exemple : la phosphorylation de la serine en position 14 de l'histone H2B ou H2BS14_p) (70).

vi. *L'ubiquitination des histones*

L'ubiquitination des histones consiste à ajouter une ou plusieurs ubiquitine(s) sur les groupes ϵ -amino (NH_2) des résidus lysines. L'ubiquitine est un peptide constitué de soixante-seize acides aminés. Contrairement à la poly-ubiquitination (poly-Ub) des protéines qui conduit à leur dégradation par le protéasome (71), la mono-ubiquitination (mono-Ub) des histones est impliquée dans la régulation de la transcription (exemples : H2A_{ub} induit la répression et H2B_{ub} induit l'activation) (72). Elle est également impliquée dans la réparation de l'ADN (72-74). Par ailleurs, il semblerait que cette ubiquitination soit une condition préalable à la méthylation des histones (75).

vii. *Les autres modifications post-traductionnelles des histones*

L'ADP-ribosylation des histones est une modification chimique réversible qui ajoute un ou plusieurs ADP-ribose sur un résidu d'acide glutamique. L'ADP-ribosylation joue un rôle dans l'expression génétique, la signalisation cellulaire, la réparation de l'ADN (*via* par exemple l'enzyme PARP-1) et l'apoptose. Certaines formes de cancer ont été associées à une ADP-ribosylation incorrecte et il existe à ce jour un anticancéreux qui inhibe la Poly-(ADP-ribose) polymérase 1 (PARP-1) (76).

La biotinylation des histones est une modification chimique qui ajoute une biotine (Vitamine B8) sur le groupe ϵ -amino des résidus lysines. L'étude de cette modification épigénétique est plutôt récente. Pour l'instant, il a été montré : i) Au cours du cycle cellulaire le niveau de biotinylation augmente en phase G1, tandis qu'il diminue durant les phases S, G2 et M. ii) Le niveau de biotinylation diminue en réponse à la prolifération des lymphocytes. iii) La biotinylation des lysines en position 4 ou 9 ou 18 de l'histone H3 est impliquée dans l'activation de gènes (76, 77).

La citrullination ou **déimination** des histones consiste en la conversion des résidus arginines en citrulline. La citrullination des arginines des histones H3 et H4 est liée au relâchement de la chromatine (78). Cette modification épigénétique intéresse plus particulièrement les chercheurs puisqu'elle joue un rôle important notamment dans le système immunitaire et le système nerveux central. Une citrullination incorrecte aurait un rôle dans l'apparition de certaines maladies telles qu'Alzheimer ou la sclérose en plaque et serait impliquée dans la tumorigenèse (76).

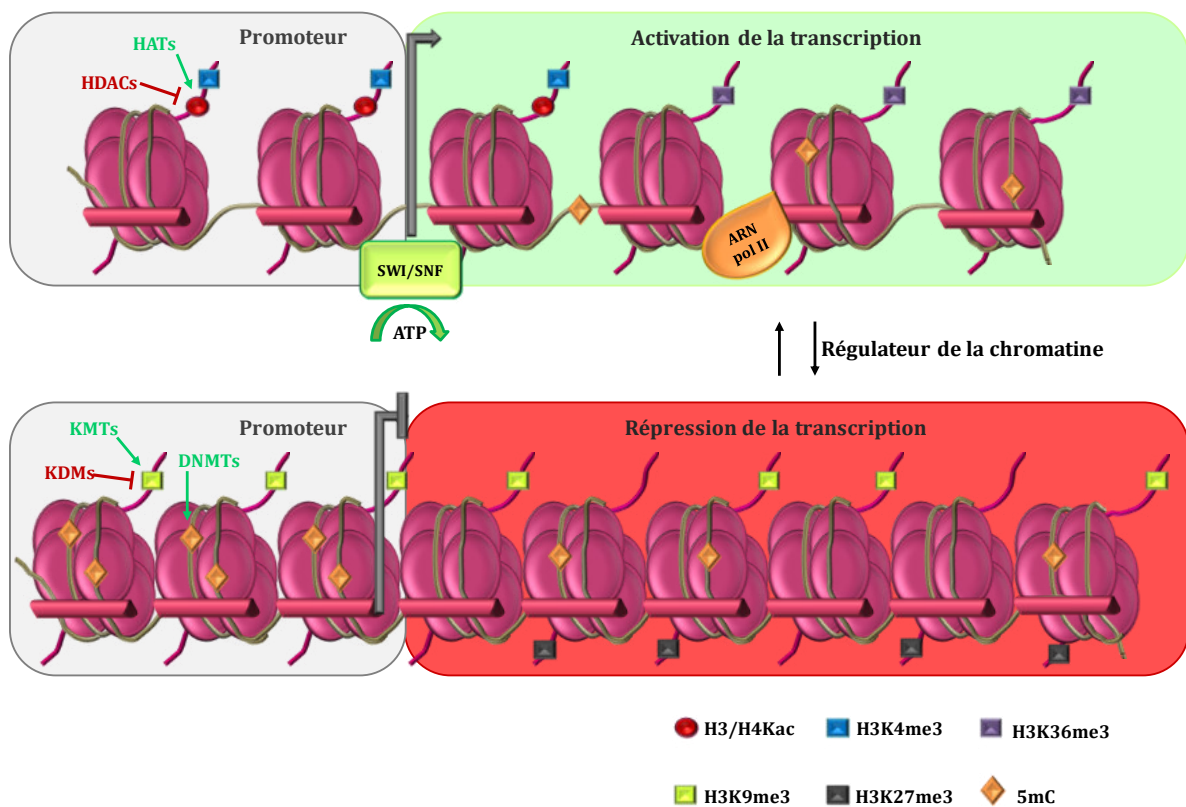
La crotonylation des histones est une modification chimique qui ajoute un groupement crotonyl sur le groupe ϵ -amino des résidus lysines. La crotonylation joue un rôle dans l'expression des gènes. Par exemple la protéine EP300 est capable de catalyser l'acétylation et la crotonylation de la lysine 18 de l'histone H3 (76).

L'isomérisation cis-trans des résidus prolines des histones est réversible. Elle joue un rôle dans le repliement des protéines (76).

La SUMOylation des histones est une modification chimique qui ajoute une ou plusieurs protéine(s) SUMO sur le groupe ϵ -amino des résidus lysines. La SUMOylation est impliquée dans le trafic intracellulaire des protéines, dans le mécanisme d'interaction de certaines protéines avec d'autres protéines ou avec l'ADN et la régulation de la transcription. La SUMOylation de la lysine en position 12 de l'histone H4 est impliquée dans la répression de gènes (76, 79).

La transcription des gènes est un processus régulé par le remodelage de la chromatine. Le relâchement de la chromatine permet d'activer la transcription des gènes, tandis que sa compaction induit la répression de celle-ci. La méthylation de l'ADN et des protéines histones, ainsi que l'acétylation des histones jouent un rôle important dans le remodelage de la chromatine et par conséquent, dans la régulation de la transcription des gènes (Figure 5).

Figure 5 : Régulation de la transcription.



Légende : Lorsque la chromatine est sous forme relaxée dite « euchromatine », la transcription des gènes est activée via l'acétylation des histones H3 et H4 (acétylations H3/H4Kac) et la tri-méthylation de l'histone H3 qui se lie à l'ADN au niveau du promoteur (méthylation H3K4me) et de l'histone H3 qui se lie le long du gène (méthylation H3K36me3). La position du nucléosome au niveau du promoteur est quant à elle régulée par le remodeleur chromatinien ATP-dépendant « SWI/SNF ». La répression de la transcription des gènes est associée à la structure dite « hétérochromatine », via la méthylation de l'ADN (méthylation 5-mC) et la tri-méthylation des histones H3 sur les lysines 9 et 27 (méthylations H3K9me3 et H3K27me3) (80).

3. Dérégulation épigénétiques

Le processus de modification épigénétique est un mécanisme très dynamique et réversible, qui est en perpétuel évolution au cours de la différenciation cellulaire et du développement. De plus, les facteurs environnementaux (alimentation, tabac, alcool, produits chimiques...) jouent un rôle non négligeable sur la fréquence d'apparition ou de disparition de certaines de ces marques épigénétiques (81).

L'apparition de mutations, ou l'altération de l'expression des facteurs épigénétiques (par les enzymes « writers », « erasers », « readers ») modifient le niveau de marques épigénétiques présentes au niveau de la chromatine et donc le profil épigénétique. C'est pourquoi, une dérégulation épigénétique conduit à la perte de la régulation de la transcription et modifie le nombre de gènes exprimés (up- ou down- régulation). C'est notamment le cas pour les gènes suppresseurs de tumeurs dont l'expression est réprimée dans le cas de cancers. Il n'est pas étonnant que les dérégulations épigénétiques soient associées à de nombreuses pathologies humaines telles que les maladies inflammatoires (sclérose en plaque), cardio-vasculaires, dégénérative du système nerveux central (Alzheimer et Parkinson), psychiatriques (schizophrénie), du métabolisme (diabète) et plus particulièrement les **cancers** (80-83).

Depuis quelques années, les protéines associées aux modifications épigénétiques sont devenues des cibles d'étude privilégiées car la compréhension de leurs mécanismes d'action et de régulation devrait permettre le développement de nouvelles thérapies, notamment contre le cancer. À ce jour, plusieurs molécules ciblant les régulateurs épigénétiques sont en essais cliniques (41), certains ont été validées par la FDA, pour le traitement de divers cancers ce qui montre que les thérapies épigénétiques fonctionnent.

II. Cancers

1. Introduction

Le mot latin cancer est originaire du mot grec karkinos, qui signifie crabe. C'est **Hippocrate** (460 - 377 avant J.C), le « *père de la médecine* » qui fut le premier à comparer l'aspect des tumeurs du sein à cet animal. Ce mot prend vraiment le sens de tumeur maligne à la fin du XV^e siècle. Aujourd'hui, ce terme général « **Cancer** » regroupe *toutes les maladies se caractérisant par une prolifération et une propagation anarchique de cellules anormales ayant subi des mutations au cœur d'un tissu sain de l'organisme.*

Actuellement, le cancer est la première cause de mortalité en France selon l'Institut National du **Cancer (INCa)**. En 2015, on comptait en France, 385.000 nouveaux cas de cancers et 150.000 personnes décédées suite à un cancer. De plus, 14,1 millions de nouveaux cas de cancers ont été diagnostiqués dans le monde en 2014.

Il existe près de deux cent types de cancer différents et seulement 5 à 10% sont génétiques. Les autres cancers sont dus à des facteurs environnementaux tels que le tabagisme (25 à 30%), l'alimentation (surpoids) et en particulier l'alcool (30 à 35%), l'infection par des virus (15 à 20%), ou encore les agents physiques, chimiques et biologiques (10 à 15 %) (84). Ce qui montre bien que les facteurs environnementaux ont un impact non négligeable sur la régulation épigénétique (84).

2. Gènes impliqués dans le développement de cancers

Il existe trois types de gènes impliqués dans le développement de cancers : les gènes dont l'expression favorise la survenue de cancers (**oncogènes**), les gènes dont l'expression favorisent la suppression des tumeurs (**suppresseur de tumeurs**) et les gènes dont l'expression favorisent la **réparation de l'ADN**.

Les proto-oncogènes (*gène normal*) sont des régulateurs positifs (stimulation) de la prolifération cellulaire, qui permettent d'induire transitoirement la prolifération des cellules avant qu'elles ne soient réprimées. À ce jour, plus d'une centaine de proto-oncogènes ont été identifiés et parmi les plus connus on retrouve myc et Abelson leukemia (*abl*). Ils jouent un rôle clef dans la croissance des tissus embryonnaires et de l'enfant, dans le renouvellement cellulaire au cours de la vie et dans la réparation en cas de dommages à l'ADN. Les proto-oncogènes peuvent être mutés sous l'influence de facteurs mutagènes, tels que les rayons UV, les rayons X, les produits chimiques cancérogène, mutagène et reprotoxique (CMR), comme l'amiante, la poussière de bois ou encore certains virus (HIV, HPV, EBV...), ainsi que le tabac et l'alcool...

Il suffit qu'un seul allèle soit muté pour qu'une maladie se manifeste, on parle de mutation dominante. Lorsqu'ils sont mutés on les appelle **oncogènes** ou **c-onc** (*gène anormal*) et dans le cas des cancers ils sont hyperactifs.

Les **gènes suppresseurs de tumeurs** (ou anti-oncogène) sont quant à eux des régulateurs négatifs (répression) de la prolifération cellulaire, autrement dit, ils la freinent. Dans le cas de cancer, les gènes suppresseurs de tumeurs ont leurs deux allèles inactivés (mutation récessive) ce qui induit la perte de leurs fonctions. Leur inactivation est impliquée dans le processus de transformation maligne. Les gènes qui codent les protéines : du rétinoblastome (pRb) et p53, sont inactivés dans la majorité des tumeurs. Or la protéine p53 est impliquée dans le maintien de la stabilité du génome et le processus d'apoptose et la protéine pRb joue un rôle dans le cycle cellulaire.

De nombreux gènes codent les protéines qui sont capables de détecter les lésions au niveau de l'ADN et d'activer le processus de réparation afin de maintenir l'intégrité du génome, on les appelle les **gènes de réparation de l'ADN**. Par exemple, la protéine p53 bloque le cycle cellulaire en phase G1/S afin de permettre la réparation de l'ADN *via* l'activation de la transcription du gène codant la protéine cycline dépendent kinase inhibitor 1 (CIP/WAF1). Lorsque la mutation est trop importante pour être réparée, la protéine p53 induit l'apoptose *via* l'activation de la transcription du gène Bcl2-associated X protein (*bax*). Or nous avons vu précédemment que dans le cas de cancers, la protéine p53 est inactivée. Par ailleurs, les gènes *brca1* et *brca2* qui permettent la réparation de l'ADN par recombinaison homologue sont mutés dans le cas de cancer du sein. De la même manière, il existe des mutations sur les gènes qui codent les enzymes de réparations des mésappariements, par excision de nucléotide ou de base.

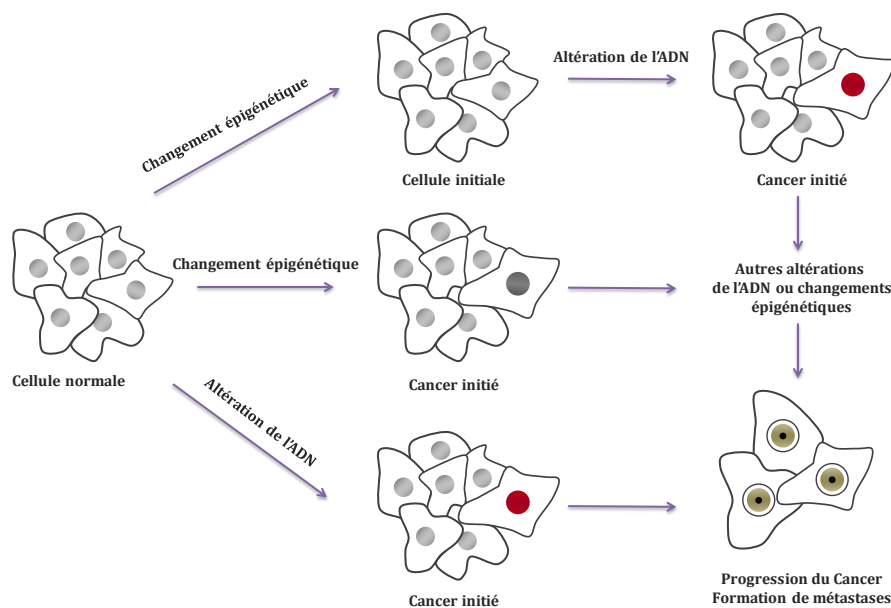
3. Altérations génétiques & épigénétiques conduisant au cancer

Les **anomalies génétiques** sont dues à des facteurs exogènes ou endogènes qui vont induire des lésions au niveau de l'ADN. Les facteurs exogènes sont par exemple, les rayonnements ionisants, l'alimentation, les produits chimiques. Les facteurs endogènes proviennent quant à eux de l'excès d'espèces réactives de l'oxygène issu du métabolisme. Les mutations de l'ADN (substitutions, délétions ou insertions) peuvent se produire accidentellement au cours de la réplication ou en dehors de la réplication, à cause d'agents physiques, chimiques, biologiques. L'accumulation de ces anomalies va induire d'une part l'activation des **oncogènes**, et d'autre part la répression des **gènes suppresseurs** de tumeurs, ainsi que des **gènes de réparation de l'ADN**.

Les **anomalies épigénétiques** sont des altérations du profil épigénétique de l'ADN et des protéines histones qui conduisent à la répression de la transcription des **gènes suppresseurs de tumeurs et de réparations de l'ADN**. L'hyperméthylation de l'ADN au niveau des promoteurs de gènes suppresseurs de tumeurs est liée à l'initiation et au développement de cancers. Par ailleurs, elle conduit à la modification aberrante des histones, comme par exemple l'acétylation de la lysine en position 16 de l'histone H4 (H4K16_{ac}), qui contribue également au développement de cancers. L'hypométhylation globale de l'ADN joue aussi un rôle dans la cancérisation. Elle contribue grandement au développement et à la progression de cancer (25, 85).

C'est donc l'accumulation de plusieurs **altérations génétiques** et/ou **épigénétiques**, rares et indépendantes qui conduit à l'apparition d'une cellule initiatrice de cancer, puis à la formation d'une tumeur et enfin au développement d'un cancer (**Figure 6**).

Figure 6 : Facteurs génétiques et épigénétiques conduisant au cancer.



Légende : Une cellule normale peut subir des changements épigénétiques et des altérations de son ADN. Le changement épigénétique conduit généralement à une cellule initiatrice de cancer (cellule initiale), parfois quand la modification épigénétique ou la mutation de l'ADN sont trop importantes cela conduit à la formation d'une cellule cancéreuse (cancer initié). L'accumulation d'autres altérations au niveau de l'ADN ou des protéines histones (modifications épigénétiques) conduit à la progression du cancer et à la formation de métastases.

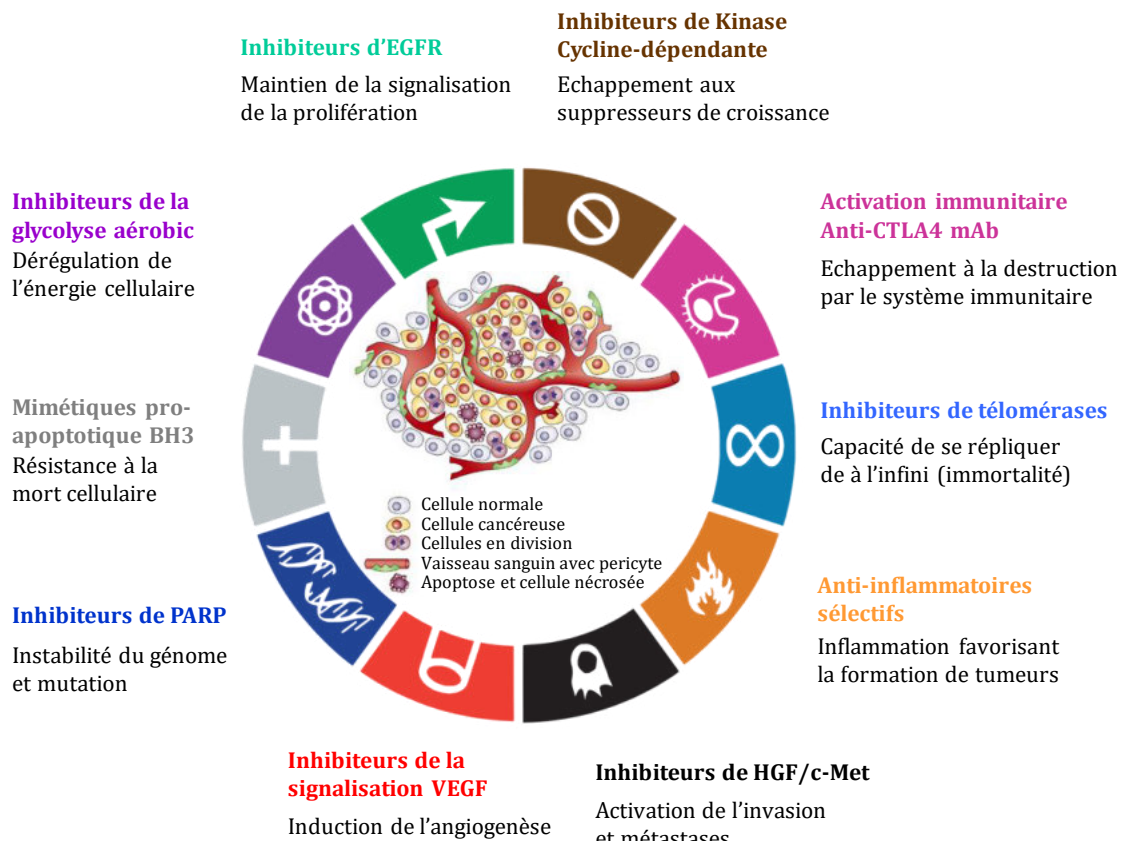
4. Caractéristiques d'une cellule cancéreuse

Pour rappel, le corps humain est composé de quelques milliards de cellules qui se répartissent dans les différents organes et chaque cellule contient en son noyau sa propre information génétique.

Une cellule normale possède des systèmes de contrôles lui permettant de vérifier l'intégrité de son génome. Dès lors qu'une mutation génétique apparaît au sein d'une cellule normale, un mécanisme de réparation de l'ADN est mis en place et lorsque celui-ci nécessite une réparation trop importante pour être effectuée, l'apoptose est déclenchée par des voies de signalisation cellulaires internes. Une cellule normale se multiplie par division cellulaire en vue de remplir une fonction biologique bien déterminée. Après un nombre de cycle prédéfini, elle devient sénescence et meurt de façon programmée par apoptose. Ainsi, chaque jour, chez un individu normal, un milliard de cellules meurent et sont aussitôt renouvelées.

Les cellules cancéreuses dérivent toutes d'un même clone, la cellule initiatrice de cancer qui, suite à l'accumulation d'une série de mutations, a acquis certaines caractéristiques qui lui permettent de se diviser de manière incontrôlée et de former des métastases (tumeurs malignes) (86, 87). En 2000, **Hanahan** et **Weinberg** suggèrent qu'une cellule cancéreuse est caractérisée par plusieurs propriétés (**Figure 7**).

Figure 7 : Caractéristiques d'une cellule cancéreuse et inhibiteurs connus.



*Légende : La figure est inspirée des figures issues des papiers d'**Hanahan et Weinberg** (87) & **Gabriele Bergers et Laura E. Benjamin** (88). Elle présente les différentes propriétés caractérisant une cellule cancéreuse décrites par **Hanahan et Weinberg** et divers inhibiteurs permettant de bloquer les cellules cancéreuses.*

- Une **indépendance vis-à-vis des signaux de prolifération**. La croissance mitotique des cellules normales est dépendante de signaux de prolifération. Or de nombreux oncogènes sont capables de mimer ces signaux. Les cellules cancéreuses produisent elles même leur signaux prolifératifs et sont indépendant des signaux exogènes.

- Une **insensibilité aux signaux antiprolifératifs**. Lorsqu'une cellule a fini de se diviser des signaux antiprolifératifs sont envoyés à celle-ci afin de bloquer son cycle cellulaire et inhiber sa prolifération (exemples : les protéines p53 et pRb stimulent la prolifération). Les cellules cancéreuses sont insensibles à ces signaux, car très souvent les protéines p53 et pRb sont mutées dans ces cellules.

- Une **résistance à l'apoptose**. Lorsqu'une cellule normale subit un stress ou que son ADN est trop endommagé pour être réparé, elle se suicide par apoptose. Les cellules cancéreuses ont accumulées des anomalies génétiques et/ou épigénétiques qui leur permettent d'échapper au mécanisme de réparation de l'ADN, ainsi qu'à l'apoptose.

- Une **prolifération illimitée**. Le nombre de division cellulaire d'une cellule normale est limité à une cinquantaine de cycles (limite de Hayflick), c'est ce qu'on appelle la sénescence. À chaque division cellulaire, une partie des télomères des chromosomes est raccourcie. Ainsi lorsque le télomère est trop court, un message indiquant l'état de sénescence est envoyé ce qui induit l'apoptose de la cellule. Les cellules cancéreuses sont capables de régénérer ce fragment d'ADN manquant au niveau des télomères *via* l'activation des télomérases et ainsi d'échapper à l'apoptose.

- Une **capacité à induire l'angiogenèse**. Chez l'adulte, le système de vascularisation, établit lors de l'embryogénèse est quiescents ou transitoirement actif lors du cycle menstruel et en cas de blessure. Pour survivre, les cellules tumorales ont besoin d'être alimentées en nutriments et en oxygène, ainsi que d'éliminer leur déchets métabolique tel que le CO₂. C'est pourquoï, elles induisent la synthèse de nouveaux vaisseaux sanguins (angiogenèse).

- Une **capacité d'invasion tissulaire et de diffusion métastatique**. Les cellules cancéreuses sont capables de se détacher des cellules voisines, de passer dans le sang et les vaisseaux lymphatiques afin de se nicher dans un tissu distant pour croître et former une nouvelle tumeur (métastase).

- Une **capacité d'induire une inflammation locale**. L'inflammation est un processus immunitaire qui permet l'élimination d'une cellule anormale ou infectée ou encore d'un pathogène. Ce processus fait intervenir des chémokines, des cytokines et d'autres facteurs

comme les facteurs de croissance ou de survie. Les cellules cancéreuses induisent une inflammation locale pro-tumorale qui facilite l'angiogenèse et l'invasion tumorale.

- **Un métabolisme énergétique adapté.** Les cellules normales tirent leur énergie de la glycolyse de manière dépendante à la présence d'oxygène. Les cellules tumorales sont quant à elles capables d'induire la glycolyse indépendamment de la présence l'oxygène, favorisant ainsi l'expression d'oncogènes, la production de macromolécules et la formation d'organelles nécessaires à la création de nouvelles cellules.

À ce jour, il existe de nombreuses molécules en essais cliniques (<https://clinicaltrials.gov/>) pour pallier les différentes propriétés acquises par les cellules cancéreuses (**Figure 7**). Par exemples, le **Palbociclib** est un inhibiteur de la kinase cycline-dépendante (CDK4/6) en essais clinique en phase III notamment dans le traitement du cancer du sein, qui permet de bloquer la croissance cellulaire. Cette molécule a déjà reçu son autorisation de mise sur le marché en 2016 (IBRANCE®, Pfizer). L'**Ipilimumab**, est un anticorps monoclonal dirigé contre la protéine 4 cytotoxique associée aux lymphocytes T (anti- CTLa4), qui a été autorisé par l'U.S FDA (**FDA** pour **F**ood and **D**rug **A**dministration) en 2011 pour le traitement de myélomes et qui permet de réactiver le système immunitaire (YERVOY®, Bistol-Myers Squibb). L'**Imetelstat**, est un inhibiteur de télomérases en phase III d'essai clinique indiqué pour les syndromes myélodysplasiques qui permet de bloquer la réplication. Des **anti-inflammatoires spécifiques** sont également administrés pour lutter contre les inflammations qui favorisent le développement de tumeurs. Le **Foretinib** (GSK089), est inhibiteurs double du facteur de croissance des hépatocytes et de la tyrosine kinase Met (HGF/c-Met) qui était en essai clinique pour le traitement de divers cancers (sein, estomac, foie, poumons, rein), qui permet de bloquer l'invasion et la formation de métastases. Le **Lucitanib**, est un inhibiteur la signalisation impliquant le facteur de croissance de l'endothélium vasculaire (VEGF) en essai clinique pour le traitement de tumeurs solides incluant le cancer du sein métastatique, qui permet de bloquer l'angiogenèse. La **molécule S63845** (Servier), est un mimétique du domaine d'interaction BH3 pro-apoptotique en essai clinique dans le traitement d'hémopathies malignes (leucémie aiguë myéloïde ou AML), tumeurs solides (mélanomes), cancers du poumon et du sein, qui induit l'apoptose. Les cellules cancéreuses utilisent la glycolyse pour la production d'ATP comme source d'énergie, ainsi le développement d'**inhibiteurs de la glycolyse** est une cible thérapeutique également étudiée. Par ailleurs, l'**Erlotinib**, est un inhibiteur du récepteur du facteur de croissance de l'épiderme (EGFR) donné dans le traitement de cancers bronchique et du pancréas, qui permet de bloquer le signal de transduction et la prolifération cellulaire.

III. Enjeux scientifiques

Les modifications épigénétiques sont importantes pour l'activation de nombreux gènes dont les gènes suppresseurs de tumeurs, c'est pourquoi leur dérégulation est impliquée dans le développement de cancers. Bien qu'il existe de nombreuses molécules capables de pallier aux différentes propriétés acquises par les cellules cancéreuses (prolifération, invasion..), il n'existe pas de protocoles thérapeutiques pour tous les cancers. Par ailleurs, de plus en plus de résistances aux traitements anti-cancéreux apparaissent (89, 90). C'est pourquoi, aujourd'hui les « writers », « erasers » et « readers » sont des cibles d'étude privilégiées dans la recherche contre le cancer (26, 91).

À l'origine les chercheurs s'intéressaient uniquement aux « writers » et « erasers », qui respectivement ajoutent et suppriment les marques épigénétiques. Ce qui explique que les premières molécules inhibitrices de ces régulateurs épigénétiques ont déjà été commercialisées il y a une quinzaine d'années, pour le traitement de diverses pathologies. En voici trois exemples connus : i) L'imatinib (Glivec®, Norvartis) un inhibiteur de tyrosines kinases (« writers ») (92) autorisé par l'HAS (Haute Autorité de Santé) en 2001 et utilisé dans le traitement de la leucémie myéloïde chronique (93, 94). ii) L'azacitidine (Vidaza®, Pharmion Corporation) un inhibiteur d'ADN méthyle-transférases (« writers ») (95) autorisé en 2004 par la FDA et indiqué dans le traitement de syndromes myélodysplasiques (96). iii) Le vorinostat (Zolinca®, Merck) un inhibiteur d'histones désacétylases (« erasers ») (97) autorisé en 2006 par la FDA et employé dans le traitement de lymphomes cutanés à cellules T (98).

Depuis plus d'une dizaine d'années, la communauté scientifique s'est aperçue que les « readers » étaient également l'un des piliers de la régulation épigénétique au même titre que les « writers » et « erasers ». En effet, les « readers » sont capables de reconnaître les marques épigénétiques, mais également de recruter des facteurs de transcription.

Dans ce contexte scientifique, mon projet de recherche s'intéresse aux protéines à bromodomaines (BCPs) qui sont des « readers », surexprimés dans le cas de cancers (dérégulation épigénétique).

*Partie 2 : Les Bromodomains,
implication dans l'épigénétique et
le développement de cancers*

I. Les bromodomains

1. Généralités

Le motif bromodomaine (**BD**) a été identifié pour la première fois en 1992, au niveau des gènes *brahma (brm)* et *female sterile homeostic (fsh)* de la drosophile *Drosophila melanogaster* (99). Il a été montré que ces gènes jouent un rôle clef dans le remodelage de la chromatine et la régulation de la transcription, suggérant ainsi que les motifs bromodomains seraient impliqués dans ces deux processus biologiques. Depuis, de nombreuses études ont clairement montré l'implication des bromodomains dans la régulation de la transcription de gènes (régulation épigénétique) et dans le développement de diverses maladies (inflammatoires, cardiovasculaires, cancers...) (25).

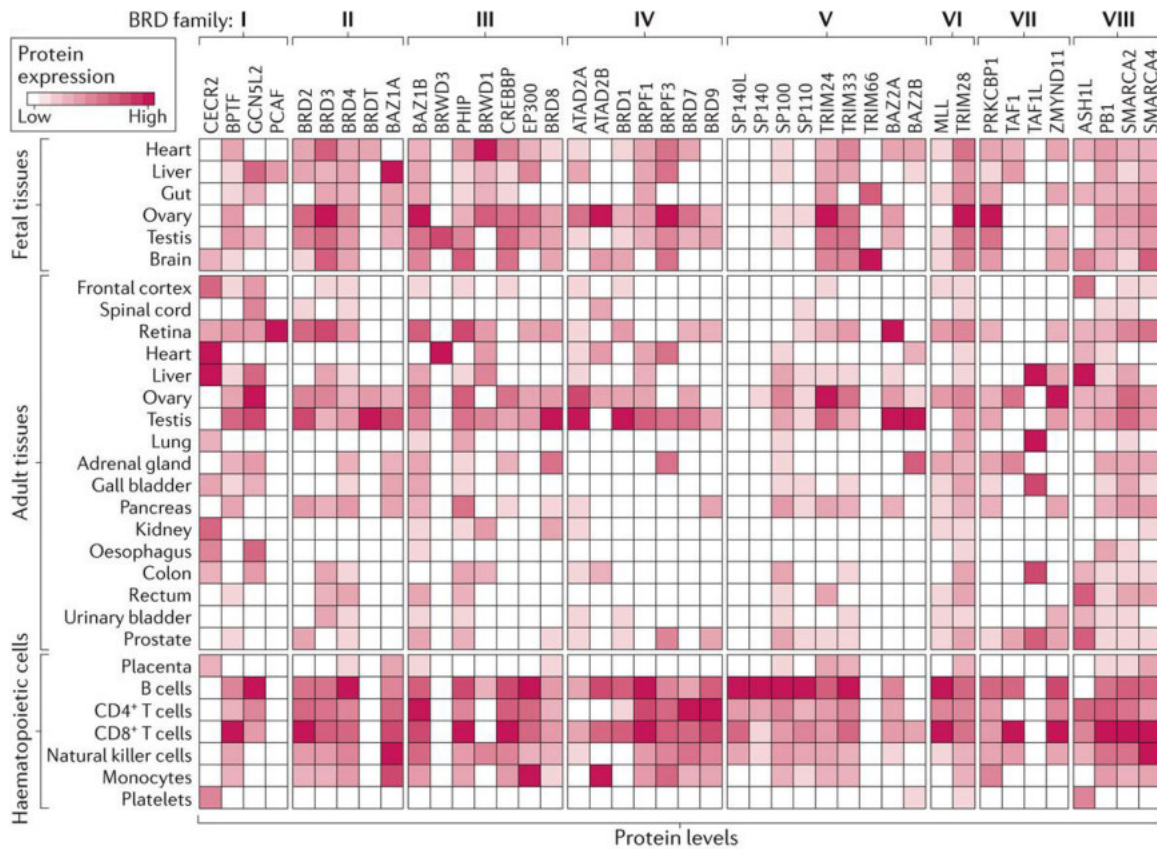
Les motifs **BDs sont des modules de reconnaissance épigénétique** « readers », capables de reconnaître spécifiquement les lysines ϵ -N-acétylées (K_{ac}) des queues des histones, ainsi que des protéines non histones qui sont associées à la transcription telles que les facteurs de transcription et les corégulateurs transcriptionnels. La spécificité de cette reconnaissance s'explique par les interactions protéine-protéine (ou PPI pour « protein-protein interaction ») qu'établissent ces protéines avec les BDs. Ces motifs sont conservés au cours de l'évolution aussi bien chez la levure (*Saccharomyces cerevisiae*), la plante (*Arabidopsis thaliana*), que l'animal (*Drosophila melanogaster*, *Homo sapiens*). Ils sont également retrouvés chez les bactéries (*Treponema denticola*) et les virus (*Escherichia phage Eps7*) (Serveur InterPro : www.ebi.ac.uk/interpro) (100).

2. Les protéines à bromodomains (BCPs)

Le protéome humain, code **quarante-deux protéines** à bromodomains différentes (**BCPs**), pour la plupart nucléaires et pour certaines également cytoplasmiques. Des isoformes protéiques issues d'épissages alternatifs au niveau des extrémités 3' de leurs gènes sont également retrouvés, notamment pour la protéine à bromodomaine 4 (BRD4) qui en présente trois. Il faut noter une ambiguïté sur la nomenclature de certaines de ces BCPs, qui ont été nommées protéines à bromodomains 1 à 9 et T (BRD1-9, BRDT), et qu'il ne faut pas confondre avec les motifs bromodomains (BDs) dont il est question dans la suite de ce chapitre.

De récentes études de protéomique ont montré que les BCPs sont exprimées dans divers tissus du corps humains (101) et pour la plupart de manière ubiquitaire (**Tableau 2**).

Tableau 2 : Expression tissulaire des différents BCPs.



Légende : Cette figure est issues de l'article de **Fujisawa et al**, (101). La localisation tissulaire des quarante-deux BCPs humaines a été déterminée au cours d'une vaste étude protéomique. La plupart de ces protéines sont exprimées de manière ubiquitaire, excepté les protéines suivantes : PCAF est exprimée principalement dans la rétine et un peu dans le foie ; BRDT est exprimée principalement dans les testicules, mais également dans les ovaires et le cœur ; BRWD3 est exprimée dans les testicules et le cœur ; ou encore SP140 et SP140L sont exprimées dans les cellules sanguines.

3. Fonctions des BCPs

Les BCPs humaines sont retrouvés dans de nombreuses classes fonctionnelles, d'où leurs implications dans de nombreux processus cellulaires (102). Le tableau ci-dessous (**Tableau 3**), présente les différentes fonctions biologiques des BCPs.

- **les histones acétyle-transférases HAT** : BRPF3, CREBBP, EP300, GCN5L2, PCAF, TAF1, TAF1L.
- **les histones méthyl-transférases HMT** : ASH1L, MLL.
- **les hélicases** : SMARCA2, SMARCA4.
- **les facteurs associés à EP300-CBP** : PCAF, BRD9.
- **les facteurs de remodelage de la chromatine ATP dépendant** : BAZ1B, BAZ1A, BRDT, BRWD3, CERCR2, PB1, SMARCA2, SMARCA4, BRWD1.
- **les protéines nucléaires d'échafaudage** : PB1.

- **les coactivateurs de la transcription** : BRPF1, GCN5L2, TAF1, TAF1L.
- **les régulateurs de la transcription ou répresseurs** : ATAD2A, BAZ2A, BPTF, BRD1, BRD2, BRD3, BRD4, BRD7, BRD8, BRD9, PKCB1, SP100, SP110, SP140, TRIM24, TRIM28, TRIM33, TRIM66, ZMYND11.

Tableau 3 : Fonctions biologiques et localisations subcellulaires des BCPs.

Protéine	Famille	Nom	Alias	Fonction	Localisation Subcellulaire	ID UniProt
ASH1L	VIII	Ash1 (absent, small, or homeotic) like	ASH1, KMT2H	histone methyltransferase	N, C	Q9NR48
ATAD2	IV	ATPase, Two AAA domain containing protein 2	ANCCA, CT137	transcriptional regulator	N	Q6PL18
ATAD2B	IV	ATPase, Two AAA domain containing protein 2B	KIAA1240	histone-binding factor	N	Q9UL10
BAZ1A	III	Bromodomain adjacent to zinc finger domain, 1A	ACF1, WCRF180, WALp1	chromatin remodeling factor	N	Q9NRL2
BAZ1B	III	Bromodomain adjacent to zinc finger domain, 1B	WSTF, WBSR9, WBSR10	transcriptional regulator	N	Q9UIG0
BAZ2A	V	Bromodomain adjacent to zinc finger domain, 2A	TIP5, WALp3	transcriptional repressor	N, C	Q9UIF9
BAZ2B	V	Bromodomain adjacent to zinc finger domain, 2B	WALp4	non défini	N, C	Q9UIF8
BPTF / FALZ	I	Bromodomain PHD finger transcription factor / Fetal Alzheimer antigen	FAC1, FALZ, BPTF	transcriptional regulator	N	Q12830
BRD1	IV	Bromodomain-containing protein 1	BRL, BRPF2	transcriptional regulator	N	Q95696
BRD2	II	Bromodomain-containing protein 2	FSH, RING3	transcriptional regulator	N	P25440
BRD3	II	Bromodomain-containing protein 3	ORFX, RING3L	transcriptional regulator	N	Q15059
BRD4	II	Bromodomain-containing protein 4	CAP, MCAP, HUNK1	transcriptional regulator	N	O60885
BRD7	IV	Bromodomain-containing protein 7	BP75, NAG4, CELTIX1	transcriptional regulator	N	Q9NPI1
BRD8	III	Bromodomain-containing protein 8B	SMAP, SMAP2	transcriptional regulator	N	Q9H0E9-2
BRD9	IV	Bromodomain-containing protein 9	-	transcriptional regulator	N, C	Q9H8M2
BRDT	II	Bromodomain containing protein, testis specific	BRD6	chromatin remodeling factor	N	Q58F21
BRPF1	IV	Bromodomain- and PHD finger-containing protein 1A	BR140, Peregrin	transcriptional activator	N, C	P55201-1
BRPF3	IV	Bromodomain- and PHD finger-containing protein 3	-	component of MOZ/MORF H3	N	Q9ULD4
BRWD1 / WDR9	III, VII	WD repeat domain 9	BRWD1	acetyltransferase complex	N	Q9NS16
BRWD3	III, VII	Bromodomain and WD repeat domain containing 3 disrupted in leukemia	BRODL	chromatin remodeling factor	N, C	Q6RI45
CREBBP / CBP	III	CREB-binding protein	CBP, KAT3A	histone acetyl transferase	N	Q92793
CECR2	I	Cat eye syndrome chromosome region	-	chromatin remodeling factor	N	Q9BXF3
EP300	III	E1A-binding protein p300	p300, KAT3B, RSTS2	histone acetyl transferase	N	Q09472
GCN5L2 / KAT2A	I	General control of amino acid synthesis 5-like 2; lysine acetyltransferase 2A	GCN5, GCN5L2, KAT2A	histone acetyl transferase	N	Q92830
PCAF / KAT2B	I	P300/CBP-associated factor	KAT2B	transcriptional activator	N	Q92831
MLL	VI	Myeloid/lymphoid or Mixed-lineage leukemia	HRX, TRX1, CXXC7, ALL-1	histone acetyl transferase	N	Q03164
PB1 / PBRM1	VIII	Polybromo 1	PB1, BAF180, PBRM1	histone methyl transferase	N, C	Q86U86
PHIP	III, VII	Pleckstrin homology domain-interacting protein	WDR11, BRWD2, DCAF14, ndrp	chromatin remodeling factor	N	Q8WWQ0
PKCB1 / PRKCBP1	VII	Protein kinase C-binding protein 1	ZMYND8, RACK7	insulin signaling	N	Q9ULU4
SMARCA2 / SMAC2	VIII	SWI/SNF related, matrix associated, actin-dependent regulator of chromatin, subfamily a, member 2	BRM, SNF2L2	transcriptional regulator	N	P51531
SMARCA4 / SMAC4	VIII	SWI/SNF related, matrix associated, actin-dependent regulator of chromatin, subfamily a, member 4	BRG1, SNF2L4	chromatin remodeling factor	N	P51532
SP100	V	Nuclear antigen Sp100	-	transcriptional regulator	N, C	P23497
SP110	V	Nuclear antigen Sp110A, Sp110C	IPR1, VOD1	transcriptional regulator	N	Q9H58
SP140	V	Sp140 nuclear body protein	LYSP100	transcriptional regulator	N, C	Q13342
SP140L	V	Sp140 nuclear body protein like	-	transcriptional regulator	-	Q13342
TAF1	VII	TAF1 RNA polymerase II, TATA box binding protein (TBP)-associated factor	TAF2A, TAF(II)250	transcription initiation	N	P21675
TAF1L	VII	TAF1 RNA polymerase II, TATA box binding protein (TBP)-associated factor like	TAF2A2, TAF(II)210	transcription initiation	N	Q8IZX4
TRIM24 / TIF1A	V	Tripartite motif containing 24	TIF1 α , PTC6, RNF82	transcription regulator	N, C	O15164
TRIM28 / TIF1B	VI	Tripartite motif containing 28	TIF1 β , KAP1, RNF96	transcription regulator	N, C	Q13263
TRIM33	V	Tripartite motif containing 33	TIF1 γ , ECTO, PTC7, RFG7	control of transcription elongation	N	Q9UPN9
TRIM66	V	Tripartite motif containing 66	TIF1 δ	transcriptional repressor	N	O15016
ZMYND11	VII	zinc finger, MYND-type containing 11	BS69, BRAM1, MRD30	transcriptional repressor	N	Q15326

Légende : Les informations listées dans ce tableau sont issues du recoupement des articles de **Filippakopoulos et al (103)** et de **Zhang et al (48)** et du site Uniprot (<http://www.uniprot.org/>). La localisation subcellulaire de ces protéines est Nucléaire (N) et/ou Cytoplasmique (C)

4. Implication des BCPs dans des pathologies humaines

Les **BCPs** sont très souvent **surexprimées dans les cellules cancéreuses** et dans le cas de maladies inflammatoires. Voici quelques exemples de protéines retrouvées surexprimées dans certaines pathologies : ATAD2A est surexprimée dans plus de 70% des cancers du sein, TRIM24 dans le cancer du sein, BRDT dans les cancers du poumon Non-Small Cell Lung Cancer (NSCLC);

EP300/CREBBP, EP300 et PCAF dans les maladies inflammatoires ; TRIM28 SMARCA2, BRD1, CREBBP, PCAF et EP300 dans les maladies neurologiques (104). Par ailleurs, il existe divers réarrangements chromosomiques oncogéniques avec les BDs qui sont liés à des formes agressives de tumeurs. C'est le cas pour les protéines BRD4 dans les **NUT Midline Carcinoma (NMC)**, BRD3 dans les leucémies AML et **Myeloïde Lymphoblastic Leukemia (MLL)** et CREBBP dans les MLL. Il a également été montré que les BCPs étaient impliquées dans le développement du **SIDA (Syndrome d'Immunodéficience Acquise)** *via* la transcription du virus du HIV (105, 106). Les BCPs sont liées au développement de nombreuses pathologies, à savoir les maladies inflammatoires, cardiovasculaires, du métabolisme, du système nerveux, de diverses cancers... (48) (Tableau 4).

Tableau 4 : Maladies associées aux BCPs.

Protéine	Famille	Implication pathologies
ASH1L	VIII	Cancer du poulmon, lié à la dystrophie musculaire facio-scapulo-humérale
ATAD2	IV	Surexprimée dans de nombreuses tumeurs
ATAD2B	IV	Exprimée dans différents type de tumeurs neuronales
BAZ1A	III	Rôle dans la spermatogenèse et possibilité d'infertilité masculine
BAZ1B	III	Surexprimée dans les tumeurs coloréctales, délétée dans le syndrome de William-Beuren
BAZ2A	V	Surexprimée dans le cancer de la prostate
BAZ2B	V	Associée à la mort subite cardiaque
BPTF / FALZ	I	Rôle dans la progression de myélome et de maladies neurodégénératives
BRD1	IV	Rôle dans les leucémie lymphoblastique aigüe
BRD2	II	Altérée dans la lymphomagenèse des cellules B
BRD3	II	Impliquée dans le carcinome NUT (NMC) et dans de nombreuses leucémies agressives
BRD4	II	Impliquée dans le carcinome NUT (NMC) et dans de nombreuses leucémies agressives, dans de nombreux cancers et dans des maladies inflammatoires ou du système immunitaire
BRD7	IV	Associée à de nombreux type de cancer
BRD8	III	Liée au cancer coloréctale
BRD9	IV	Rôle possible dans les fonctions immunitaires et le développement de multiple type de cancers
BRDT	II	Exprimée dans le cancer du poulmon
BRPF1	IV	-
BRPF3	IV	-
BRWD1 / WDR9	III, VII	-
BRWD3	III, VII	Exprimée de manière abérante dans la leucémie lymphoblastique chronique (LLC) à cellules B
CREBBP / CBP	III	Impliquée dans la leucémie myéloïde aigüe (AML), la leucémie lymphoblastique aigüe (ALL), lymphome des cellules B, mutée dans le syndrôme de Rubinstein-Taybi
CECR2	I	-
EP300	III	Impliquée dans la leucémie myéloïde aigüe (AML) et dans d'autres types de cancer
GCN5L2 / KAT2A	I	Impliquée le cancer du poulmon pas à cellules petites (NSCLC)
PCAF / KAT2B	I	Impliquée dans les voies de prolifération des cellules cancéreuses
MLL	VI	Translocations avec des gènes impliqués dans le développement de leucémies multiples
PB1 / PBRM1	VIII	Mutée dans le carcinome des cellules rénales
PHIP	III, VII	Exprimée dans les myélomes et les carcinome des cellules de l'épiderme
PKCB1 / PRKCBP1	VII	Surexprimée dans les cancers du col de l'utérus
SMARCA2 / SMAC2	VIII	Associée aux carcinomes des cellules de la peau du cou et de la tête, altérée dans le carcinome hepatocellulaire
SMARCA4 / SMAC4	VIII	Altérée dans le carcinome hepatocellulaire, mutée dans de nombreux autres types de cancers, incluant celui du poulmon, du sein et du pancréas
SP100	V	Rétro-contrôle négatif dans le cancer du larynx
SP110	V	Mutations associées à l'immunodéficience, les maladies hépatique véno-occlusive, changement du nombre de copie dans les myélomes desmoplastique et les tumeurs maligne de la gaine des nerfs périphériques
SP140	V	Liée à la leucémie lymphoblastique chronique (LLC)
SP140L	V	-
TAF1	VII	Associée au cancer de la prostate
TAF1L	VII	-
TRIM24 / TIF1A	V	Altérée dans de nombreux type de cancer, incluant le cancer papillaire de la thyroïde et le cancer du sein
TRIM28 / TIF1B	VI	Associée au cancer gastrique
TRIM33	V	Liée à la leucémie myélonocyttaire chronique (LLMC)
TRIM66	V	-
ZMYND11	VII	Associée à la répression de la croissance des cellules et au cancer du sein

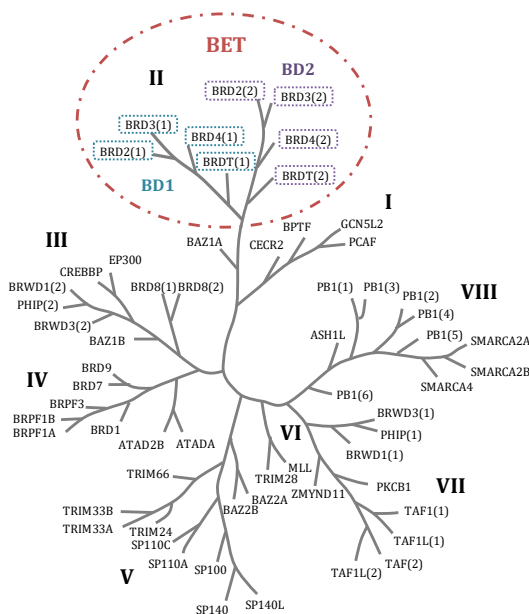
Légende : Les informations listées dans ce tableau sont issus du recoupement des divers revues lu et principalement de l'article de Zhang et al (48).

5. Classifications des BDs

Soixante-et-un motifs BDs ont été répertoriés à ce jour chez l'homme, ce qui implique que les protéines BCPs sont constituées d'un ou plusieurs BDs. La majorité des BCPs possèdent un unique motif BD, un tiers en possède deux exemplaires ce qui est le cas de la protéine à bromodomains 4 ou BRD4 (famille II) et seule la protéine à bromodomaine PB1 (famille VIII) en possède six exemplaires.

En 2012, l'équipe du Dr. **Knapp** a établi le premier arbre phylogénétique classifiant ces soixante-et-un BDs en s'appuyant sur leur degré de similarité de séquence et de structure (103). Pour cela, **Filippakopoulos et al** se sont basés d'une part sur l'alignement de séquences de l'ensemble des BDs humains et d'autre part sur l'alignement de structure tridimensionnelle sur la base de prédictions de structures secondaires ou des structures expérimentales disponibles dans la banque de données protéiques (**PDB : Protein Data Bank**, www.rcsb.org). Les structures disponibles dans la PDB ont été obtenues soit par cristallographie pour trente-quatre des bromodomains, soit par Résonance Magnétique Nucléaire (RMN) pour quatre d'entre eux (102, 103). Comme illustré dans la **Figure 8**, l'arbre phylogénétique répartit l'ensemble des BDs en **huit familles distinctes**, numérotées par des chiffres romains. Par ailleurs, la **famille II qui est aussi appelée famille « BET »** pour « Bromo and Extra-Terminal domain », est subdivisée en **deux sous-familles BD1 et BD2** (102).

Figure 8 : Arbre phylogénétique des BDs du protéome humains.



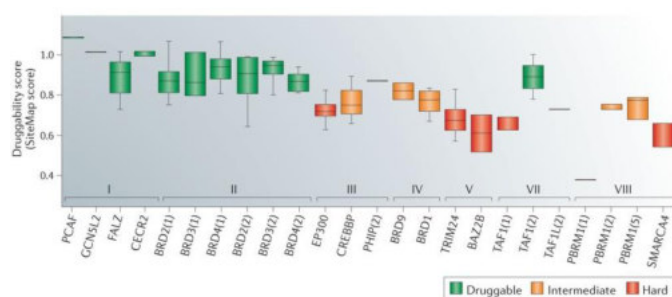
Légende : La figure est inspirée de **Filippakopoulos et al** (102). Les soixante-et-un BDs codés par le protéome humain ont été classifiés par **Knapp et al**, en huit familles distinctes (I à VIII) selon leur similarité de séquence et de structures. La famille II, aussi appelée famille « BET » comprend huit membres et est subdivisée en deux sous-familles BD1 et BD2.

6. Cible épigénétique et thérapeutique « druggable »

La **druggabilité** vient du terme « druggability » qui signifie « *la capacité pour une protéine cible à interagir avec une molécule pouvant aboutir au développement de médicaments (drug)* » (107, 108). Dans le cadre de l'étude des BDs, la druggabilité du complexe BD-peptide histone lysines acétylées, correspond à la faculté de trouver une petite molécule très affine pour la poche d'interaction K_{ac} des BDs, avec une constante de dissociation ou K_D de l'ordre du nanomolaire. Cette molécule doit être capable de moduler l'interaction PPI, soit en bloquant la formation de ce complexe, soit entrant en compétition avec le peptide pour la formation de ce complexe et ainsi avoir un impact thérapeutique.

En 2012, une étude de « druggability » a été menée par **Vidler et al** (109) sur vingt-quatre des soixante-et-un BDs, à l'aide du programme SiteMap. Les résultats de cette prédiction proposent que malgré un repliement très conservé au sein des BDs, la cavité d'interaction avec les K_{ac} semble ne pas être accessible de la même manière chez tous les BDs. De ce fait la probabilité de pouvoir inhiber l'interaction PPI, BDs-peptide histone lysines acétylées est très variable selon les familles de BDs. La **Figure 9**, indique en rouge les protéines qui sont des cibles faciles pour des petites molécules inhibitrices de la PPI comme les protéines de la famille BET ou PCAF, en jaune les protéines qui sont des cibles moyennes telles que les protéines BRD1, CREBBP ou encore PB1(5), et en blanc les protéines qui sont des cibles difficiles par exemple les protéines EP300 et SMARCA4. Cependant, depuis cette étude, un inhibiteur PFI-3, très afin de PB1(5) et de SMARCA4 a été identifié (K_D 48 nM et une concentration inhibitrice médiane ou IC_{50} 89 nM, respectivement) (110, 111). Ces résultats démontrent que, même si une protéine, SMARCA4 en l'occurrence, est prédite comme une cible difficile, il est tout de même possible de trouver un inhibiteur de sa PPI.

Figure 9 : Étude de "druggability" des BDs.



Légende : Cette figure est issue de l'article de **Filippakopoulos et al**, (112) et reprend les résultats obtenus par **Vidler et al**, (109). En vert sont indiqués les cibles faciles pour le développement d'inhibiteurs de PPI, en orange les cibles moyennes et en rouge les cibles difficiles. Bien que SMARCA4 ait été prédite comme une cible difficile, un inhibiteur (PFI-3) a été identifié en 2015 par **Fedorov et al**, (110).

Les BCPs sont des cibles d'étude très intéressantes pour le développement de molécules inhibitrices de la PPI entre les BDs et les peptides porteur de K_{ac} , d'une part afin de mieux comprendre le mécanisme d'action de chacune de ces protéines et d'autre part en vue de développer de potentiels médicaments.

Dans ce contexte, l'équipe des Drs. Xavier Morelli et Yves Collette, œuvre à l'identification et la caractérisation de petites molécules inhibitrices des PPI ciblant les BDs des protéines de la famille BET.

II. La famille BET « Bromo et Extra Terminal domain »

Chez l'homme, la famille BET comprend quatre membres: les protéines à bromodomaines BRD2, BRD3, BRD4 et BRDT. Chacune de ces BCPs est constituée de deux modules bromodomaines BDs positionnés en tandem (BD1 et BD2). C'est pour cette raison que les membres de la famille BET sont subdivisés en deux sous-familles, BD1 et BD2 (**Figure 8**). Une nouvelle nomenclature est utilisée pour distinguer les deux domaines d'une même protéine, par exemples BRD4(1) et BRD4(2), respectivement pour les domaines BD1 et BD2 de la protéine BRD4. Par ailleurs, la famille BET est la plus étudiée à ce jour parmi les sept autres.

1. Rôles biologiques des BETs

Les protéines de la famille BET sont impliquées dans de nombreux processus cellulaires (remodelage de la chromatine, régulation de l'expression de gènes cibles) (113), c'est pourquoi elles sont devenues des cibles pour la recherche d'inhibiteur de leur PPI très étudiées afin de comprendre de leur rôle biologique.

La protéine **BRDT** (Bromodomain Testis specific) est spécifique du tissu testiculaire, (114) bien qu'une étude protéomique récente ait montré qu'elle est également exprimée dans les ovaires et le cœur (101, 115). Cette localisation subcellulaire suggère que le rôle de cette protéine soit principalement lié à la spermatogenèse, *via* le remodelage de la chromatine (116, 117) et à la régulation de gènes essentiels (118). La délétion du premier domaine BD1 (ou BRDT(1)) de cette protéine est suffisante pour induire une stérilité chez la souris mâle homozygote (119) et des chercheurs ont ainsi émis l'hypothèse que le développement d'inhibiteurs ciblant la protéine BRDT pourrait être utilisé comme contraceptifs masculins (120). La protéine BRDT joue également un rôle dans l'épissage alternatif d'ARNm (121, 122).

Les protéines BRD2, BRD3 et BRD4 sont, quant à elles exprimées de manière ubiquitaire (102) et sont principalement impliquées dans la régulation de la transcription de gènes cibles (123, 124).

La protéine **BRD2**, est une serine/thréonine kinase nucléaire qui possède une activité de liaison à l'ADN, ainsi qu'une activité intrinsèque de chaperonne (123). Elle s'associe au facteur E2F1 (125), permettant alors le recrutement du facteur de transcription TATA-binding protein (TBP) (126), ainsi qu'avec le complexe de remodelage de la chromatine SWI/SNF régulant ainsi NF- κ B. Ce complexe est ensuite recruté au niveau de la région promotrice de certains gènes, telle que la cytokine pro-inflammatoire **TNF α** (Tumor Necrosis Factor alpha), ce qui stimule sa transcription. La protéine **BRD3**, interagit avec la protéine acétylée GATA1, un facteur de transcription essentiel à l'hématopoïèse, (127) ainsi qu'avec le complexe d'élongation PAF1 qui s'associe à son tour avec l'ARN polymérase au cours de la phase d'élongation de la transcription.

Enfin, la protéine **BRD4** est capable d'interagir avec différents régulateurs épigénétiques, tels que la bifonctionnelle arginine déméthylase et lysyl-hydroxylase (JMJD6), l'histone lysine N-méthyltransférase (NSD3) et le Chromodomain Helicase DNA Binding Protein4 (CHD4) qui joue un rôle dans la répression de la transcription (128). La double interaction de JMJD6 et de BRD4L (isoforme long de BRD4) avec le complexe de transcription P-TEFb permet l'activation du complexe P-TEFb et la libération de la pause de la régulation des gènes. NSD3, réalise la méthylation de l'histone H3 en position 36 (H3K36_{me3}) ce qui active la transcription. Lorsque les protéines BRD4 ou NSD3 sont délétées, le niveau de méthylation de l'histone H3 est réduit, ce qui suggère que le complexe BRD4/NSD3 jouerait un rôle dans cette modification épigénétique (128). Par ailleurs, les complexes BRD3/NSD3 et BRD4/NSD3 activent le recrutement du récepteur de l'estrogène alpha (ER α), ce qui favorise l'expression de ce gène. En interagissant avec les protéines CHD4, JMJD6 et NSD3, la protéine BRD4 est capable de réguler la transcription. La déphosphorylation de la sérine en position 10 de l'histone H3 (H3S10_P) médiée par la protéine phosphatase 1a (PP1 α) permet la désacétylation des lysines en position 5 et 8 des histones H4 (acétylations H4K5_{ac}K8_{ac}) du nucléosome par les histones désacétylases HDAC1/2/3. Le retrait de ces PTMs conduit au relargage de BRD4 liée à la chromatine, ce qui permet le recrutement de P-TEFb, et qui stimule l'expression de gènes inductibles. La protéine BRD4 est également capable d'interagir avec la sous-unité RelA du facteur NF- κ B, et ainsi participe à la régulation de ce gène. Il a été démontré que les protéines BRD2 et BRD4 sont essentielles au développement embryonnaires (ce que les auteurs ont démontré par délétion des gènes (129, 130)). De plus, au cours de la mitose, seules les protéines BRD2 et BRD4 restent liées à la chromatine alors que la majorité des facteurs nucléaires sont libérés dans le cytoplasme, ce qui suggère que ces protéines pourraient contribuer au caractère héréditaire du

profil épigénétique (131). Récemment il a été montré que la protéine BRD4 possédait une activité kinase et qu'elle était capable de phosphoryler le résidu Ser2 à l'extrémité C-terminal de l'ARN pol II (132).

2. Pathologies associées aux BETs

La dérégulation des protéines BETs affecte, entre autres, l'expression de nombreux gènes impliqués dans la croissance cellulaire ou qui codent des glycoprotéines pro-inflammatoires telles que les cytokines, ce qui conduit au développement de diverses pathologies, comme les maladies liées au métabolisme, inflammatoires ou encore une multitude de cancers... (25, 104) (Tableau 4).

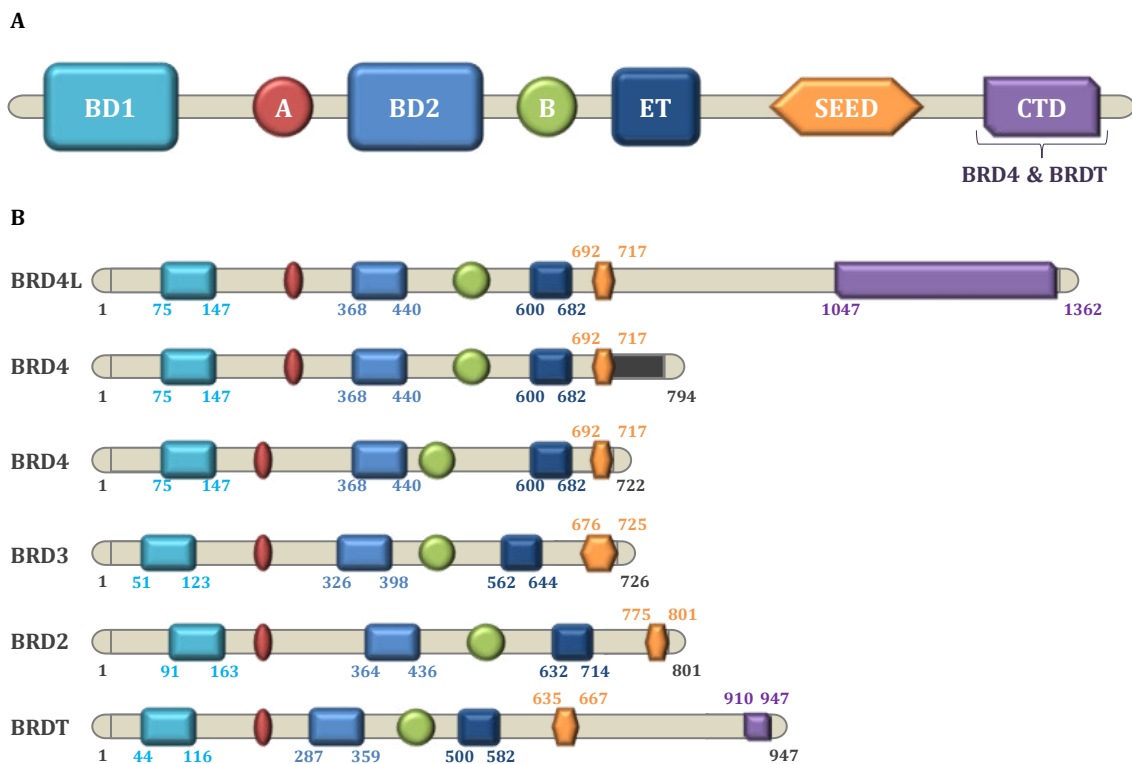
La protéine BRD2 est par exemple surexprimée dans le cancer du poumon de type « NSCLC » (133, 134). La perturbation de l'expression de la protéine BRD2 est quant à elle impliquée dans la dérégulation de l'expression de gènes qui contrôlent le métabolisme. Une sous-expression de BRD2 chez la souris conduit à une obésité extrême et à une hyper-insulinémie (134), sans induire le diabète de type 2. La protéine BRD2 joue aussi un rôle dans la réponse inflammatoire des macrophages *via* une interaction directe avec la région régulatrice de gènes qui codent les cytokines, de petites protéines immuno-modulateurs (135, 136). Elle serait également liée à des troubles épileptiques (137), bien qu'une étude plus récente n'ait montré aucun lien entre BRD2 et l'épilepsie photosensible (138). Par ailleurs, elle est surexprimée dans des lymphomes et des leucémies (139, 140). Des translocations chromosomiques dominantes entre NUT et les protéines BRD3 et BRD4 sont liées à des formes agressives de tumeurs, comme dans le cas du cancer NMC pour lequel il n'existe actuellement pas de traitement (104). Dans deux tiers des cas, le carcinome agressif NMC implique la fusion BRD4-NUT (141-145). Ces protéines de fusion BRD-NUT bloquent la différenciation cellulaire et favorise la croissance des tumeurs (142, 145). Des translocations chromosomiques impliquant ces protéines sont également responsables des leucémies AML et MLL (146, 147). Par ailleurs, l'oncogène c-Myc est souvent surexprimé dans ces leucémies et l'inhibition de la protéine BRD4 induit une diminution de l'expression de celui-ci (148).

3. Organisation structurale des BCPs de la famille BET

Lorsqu'une protéine cible a été identifiée dans une fonction biologique ou dans une pathologie, la première étape consiste à déterminer l'organisation de cette protéine, à savoir tous les domaines protéiques et toutes les séquences d'acides aminés formant des motifs conservés la constituant. L'étude dite par « domaine » est souvent privilégiée pour les protéines de grande taille en vue de décrypter leurs rôles individuels.

Comme indiqué dans la **Figure 10**, les protéines à bromodomains de la famille BET font entre 80 et 150 kDa et sont organisées en plusieurs domaines : à l'extrémité N-terminale se trouve deux domaines BDs (BD1 et BD2) positionnés en tandem, un motif « A » basique situé entre ces deux BDs et un motif « B » acide situé après le BD2, suivi d'un domaine ATP binding/kinase et à l'extrémité C-terminale se trouve un domaine Extra Terminal nommé « ET » et un motif « SEED » situé après celui-ci. Les protéines BRD4 et BRDT possèdent également un domaine C-terminal nommé « CTD » (149-152).

Figure 10 : Organisation des BCPs de la famille BET.



Légende : A- Architecture générale des BETs. L'ensemble des protéines de la famille BET sont constituées de : (i) Deux bromodomains en tandem (rectangles bleu clair pour le BD1 et bleu plus foncé pour le BD2), responsables de la reconnaissance des lysines acétylées des histones. (ii) Un motif Extra-Terminal « ET », carré bleu nuit, chargé d'établir des PPI notamment avec les facteurs de transcription. (iii) Trois autres motifs conservés « A, » en rouge, « B » en vert et « SEED » en orange, impliqués respectivement dans la localisation nucléaire, l' homo ou l'hétéro-dimérisation et l'interaction avec les facteurs de transcription de la famille E2F. (iv) Les protéines BRD4 et BRDT possèdent également un domaine « CTD », responsable de l'interaction avec des protéines virales et le facteur P-TEFb. B- Détail de l'architecture de chaque protéine. B- Le gène *brd4*, code pour trois isoformes protéiques, un isoforme long (BRD4L) comprenant 1362 résidus et deux isoformes courts comprenant 722 et 794 résidus respectivement. Ces deux isoformes courts diffèrent entre eux par la présence d'un exon codant de 75 résidus à l'extrémité C-terminale (gris foncé), décrit comme impliqué dans la réponse aux dommages à l'ADN (153). La forme la plus longue est celle retrouvée dans la majorité des processus biologiques. Pour les protéines BRD3, BRD2 et BRDT, seul l'isoforme le plus exprimé est représenté dans la figure.

Les domaines BDs :

Les deux modules BD1 et BD2 des protéines de la famille BET sont capables de reconnaître et d'interagir spécifiquement avec les lysines acétylées de protéines histones ou non histones. De manière générale, l'affinité de l'interaction BD-peptide lysines acétylées varie selon les BDs et leur partenaire. Les constantes thermodynamiques de dissociation (K_D) décrites dans la littérature oscillent entre des valeurs de l'ordre du millimolaire (mM) au micromolaire (μM) (102, 104).

L'article de **Filippakopoulos *et al*** (103) liste de nombreuses interactions BD-peptide identifiées, dont celles impliquant les membres de la famille BET (**Tableau 5**). Dans le noyau des cellules, les protéines de la famille BET peuvent se lier principalement aux histones H3 et/ou H4 avec des affinités différentes pour la chromatine. L'architecture en deux modules BDs placés en tandem permet d'augmenter l'affinité de l'interaction avec la chromatine. Les deux modules BD1 et BD2 de chacune de ces protéines sont capables de reconnaître les histones H3 mono- ou multi-acétylées avec une affinité allant de 67 μM à 301 μM et les histones H4 mono- ou multi-acétylées avec une affinité allant 2,7 μM à 3,6 mM. Les domaines BD1 présentent une affinité légèrement meilleure pour l'histone H4 comparé aux BD2. Par ailleurs, l'histone H4 tétra-acétylée en positions 5, 8, 12 et 16 (H4K5_{ac}K8_{ac}K12_{ac}K16_{ac}) présente la meilleure affinité que ce soit pour les BDs seuls ou en tandem. De plus, le module BD1 de la protéine BRDT peut aussi se lier l'histone H2B multi-acétylées avec une affinité de 157 μM . Bien que les domaines BD1 et BD2 soient tous deux capables de reconnaître des lysines acétylées des histones H4, les domaines BD1 semblent reconnaître une séquence spécifique contrairement au domaine BD2. De plus les domaines BD1 semblent être plus spécifiques des lysines acétylées des histones H4, tandis que les BD2 reconnaissent plus largement diverses histones. **Ceci suggère qu'il serait possible de cibler spécifiquement et indépendamment ces deux sous-familles.**

Les modules BDs des protéines de la famille BET sont également capables d'interagir avec des protéines acétylées non histones, telles que, la sous unité RelA du facteur nucléaire NF- κB , les facteurs de transcription GATA1, TWIST, E2F1, la cyclin-T1 du facteur d'élongation P-TEFb ou encore le complexe de remodelage de la chromatine SWI/SNF (104). Par exemple, le domaine BD1 de la protéine BRD3 interagit avec la protéine GATA1 acétylée sur les lysines en positions 312 et 315 (127, 154). Le domaine BD de la protéine BRD2 interagit avec la protéine acétylée E2F1 (123, 155) Les domaines BD1 et BD2 de la protéine BRD4 interagissent avec la sous-unité RelA du facteur NF- κB acétylé sur la lysine en position 310. Le domaine BD2 de cette protéine est également capable d'interagir avec la protéine TWIST acétylée sur les lysines en positions 73 et 76. Ce complexe est nécessaire à l'expression de WNT5A, le substrat utilisé dans

la voie de signalisation cellulaire Wnt. Cette interaction est impliquée dans l'invasion des cellules cancéreuses et la tumorigenèse dans le cas du cancer du sein (156). Ce domaine, interagit aussi avec un constituant du facteur d'élongation P-TEFb, à savoir la protéine acétylée cycline-T1 acétylée sur les lysines en positions 380, 386 et 390, ce qui active le facteur d'élongation P-TEFb.

Tableau 5 : Liste des PPI BDs-peptides histones acétylés connues pour les BETs.

Bromodomains	Peptides	Affinité (µM)	Techniques	Bromodomains	Peptides	Affinité (µM)	Techniques
BRD2(1/2)	H3K14	non quantifié	IP	BRD4(1)	H4K5K12	non quantifié	RMN
BRD2(1/2)	H4K5	non quantifié	IP	BRD4(1)	H4K8K12	27,4 +/- 0,9	ITC
BRD2(1/2)	H4K12	non quantifié	IP	BRD4(1)	H4K12K16	46,1 +/- 0,9	ITC
BRD2(1)	H4K5	130 +/- 4,4	ITC	BRD4(1)	H4K12K16K20	20,4 +/- 0,8	ITC
BRD2(1)	H4K5K8	13,5 +/- 0,3	ITC	BRD4(1)	H4K5K8K12K16	2,8 +/- 0,2	ITC
BRD2(1)	H4K5K12	360	SPR	BRD4(1)	RelA-K310	non quantifié	IP
BRD2(1)	H4K8K12	65,8 +/- 3,5	ITC	BRD4(2)	H3K4	260 +/- 32,2	ITC
BRD2(1)	H4K12K16	29,4 +/- 1,0	ITC	BRD4(2)	H3K9K14	non quantifié	RMN
BRD2(1)	H4K12K16K20	27,9 +/- 0,7	ITC	BRD4(2)	H4K5	1000 +/- 126	RMN
BRD2(1)	H4K5K8K12K16	3,7 +/- 0,1	ITC	BRD4(2)	H4K5	60,2 +/- 2,5	ITC
BRD2(2)	HAK5K8	13,5 +/- 0,3	ITC	BRD4(2)	H4K12	1350 +/- 78	RMN
BRD2(2)	H4K12	2900	RMN	BRD4(2)	H4K5K12	non quantifié	RMN
BRD2(2)	H4K8K12	65,8 +/- 3,5	ITC	BRD4(2)	H4K31	170 +/- 10,1	ITC
BRD2(2)	H4K12K16	29,4 +/- 1,0	ITC	BRD4(2)	H4K5K8	63,3 +/- 2,3	ITC
BRD2(2)	H4K12K16K20	27,9 +/- 0,7	ITC	BRD4(2)	H4K8K12	20,4 +/- 0,8	ITC
BRD2(2)	H4K5K8K12K16	3,7 +/- 0,1	ITC	BRD4(2)	H4K12K16	49,3 +/- 1,1	ITC
BRD3(1/2)	H3K14	non quantifié	IP	BRD4(2)	H4K12K16K20	22,7 +/- 0,8	ITC
BRD3(1/2)	H4K5	non quantifié	IP	BRD4(2)	H4K5K8K12K16	26,6 +/- 0,1	ITC
BRD3(1)	H4K8	211 +/- 5,2	ITC	BRD4(2)	RelA-K310	non quantifié	IP
BRD3(1)	H4K12K16K20	31,4 +/- 0,9	ITC	BRD4(2)	CyclinT1 K380K386K390	non quantifié	RMN
BRD3(1)	GATA1 K312K314K315K316	non quantifié	IP	BRDT(1)	H2BK12K15	157 +/- 15	ITC
BRD3(1)	GATA1 K314K315	non quantifié	RMN	BRDT(1)	H3K9K14K18K23	390 +/- 36	ITC
BRD3(1)	GATA1 K314K315K316	0,011 +/- 0,006	RMN	BRDT(1)	H4K12	150 +/- 14,6	ITC
BRD3(2)	H3K8	67,1 +/- 4,5	ITC	BRDT(1)	H4K5K8	21,9 +/- 1,7	ITC
BRD3(2)	H4K12	149 +/- 8,0	ITC	BRDT(1)	H4K8K12	193 +/- 22	ITC
BRD3(2)	H4K20	10,5 +/- 0,2	ITC	BRDT(1)	H4K12K16	117 +/- 22	ITC
BRD3(2)	H4K12K16K20	61,3 +/- 1,9	ITC	BRDT(1)	H4K5K12	340 +/- 67	ITC
BRD4(1/2)	HAK5K8K12K16	2,7 +/- 0,2	ITC	BRDT(1)	H4K5K8K12K16	28,0 +/- 0,6	ITC
BRD4(1/2)	RelA-K310	non quantifié	IP	BRDT(1)	H4K5K8K12K16	23,0 +/- 0,5	ITC
BRD4(1)	H3K9	301 +/- 40,9	ITC	BRDT(2)	H3K9K14K18K23	214 +/- 48	ITC
BRD4(1)	H3K9K14	non quantifié	RMN	BRDT(2)	H3K14K18	217 +/- 28	ITC
BRD4(1)	H4K5	810 +/- 57	RMN	BRDT(2)	H3K14K23	176 +/- 52	ITC
BRD4(1)	H4K8	84,7 +/- 8,2	ITC	BRDT(2)	H3K9K18	360 +/- 140	ITC
BRD4(1)	H4K12	650 +/- 11	RMN	BRDT(2)	H3K18	251 +/- 62	ITC
BRD4(1)	H4K5K8	6,8 +/- 0,1	ITC				

Légende : Tableau listant les interactions BET-Peptide, repris de l'article de **Filippakopoulos et al (102)**.

Motif A :

Le motif A serait un signal de localisation pour les protéines. Récemment, une région de douze acides aminés conservés (KGVKRKADTTTP) située à la fin de ce motif, a été identifiée. La mutation de celle-ci conduit à une mauvaise localisation des protéines (157).

Motif B :

Le motif B serait quant à lui responsable de l'homo et l'hétéro-dimérisation entre les BET, du moins dans le cas des protéines BRD2. La suppression de ce motif chez BRD2, entraîne la dissociation des chromosomes mitotiques (158).

Le domaine ET :

Le domaine Extra Terminal (159) est un domaine d'interaction multiprotéique. Il facilite les PPI avec d'autres partenaires, tels que le facteur de transcription TFIID ou les corégulateurs ou encore d'autres régulateurs épigénétiques (CHD4, JMJD6, NSD3...) (128, 160). En interagissant avec NSD3, il permet de stimuler la transcription de manière P-TEFb indépendante, notamment dans le cas de l'isoforme court de BRD4 qui ne possède pas de motif « CTD » (128). Il est également capable d'interagir avec le facteur nucléaire LANA (Latent Nuclear Antigen) du virus de Kaposi **KSHV** (**K**aposi's **S**arcoma-associated **H**erpes **V**irus) (161). Ce domaine supplémentaire, propre à cette famille, est à l'origine du surnom « BET ».

Le motif SEED :

Le motif SEED est riche en sérines, acide glutamiques et acide aspartiques (162) et il est nécessaire aux protéines BET pour la liaison de facteurs de transcription de la famille E2F (125).

Le domaine CTD :

Seules les protéines BRD4 et BRDT possèdent un domaine conservé « CTD » pour « C-terminal domain ». Le domaine « **CTD** » de la protéine BRD4 et BRDT est capable de recruter les sous unités CDK9 et CCNT1 du complexe P-TEFb (163), par ailleurs il permet le maintien de la structure de la chromatine (54). Actuellement, seule la structure d'une portion du motif CTD de la protéine BRD4 en complexe avec la protéine E2 du papillomavirus (HPV) a été élucidée, suggérant que ce motif serait impliqué dans l'interaction avec des protéines virales (164). De plus, cette interaction a pour effet de stimuler la transcription du virus (165) et de réprimer les onco-protéines virales E6 et E7 qui sont des agonistes des suppresseur de tumeurs p53 et pRB (166).

Mon projet de thèse s'intéresse plus particulièrement aux motifs BDs des protéines de la famille BET car ces domaines sont impliqués dans de nombreux processus biologiques (régulation de la transcription, le métabolisme et la signalisation cellulaire (167, 168) et dans le développement de nombreuses pathologies (maladies inflammatoires, cardiovasculaires, du métabolisme) et plus particulièrement les cancers (104). Le développement d'inhibiteurs de la PPI BD-peptide devrait permettre de décrypter le rôle de ces BDs et de développer de nouvelles thérapies.

4. Architecture BDs et plus particulièrement des BETs

a. Structure primaire : identité de séquence

Afin de comparer les identités de séquences entre les soixante-et-un motifs bromodomains, j'ai utilisé les séquences en acides aminés de ces BDs accessibles sur la base de données Uniprot (<http://www.uniprot.org/>) et je me suis servi de l'outil SIAS disponible sur le serveur <http://imed.med.ucm.es/Tools/sias.html>.

Il s'avère que les soixante-et-un BDs ne partagent que 21% d'identité de séquence (acide aminés) entre eux. L'analyse comparative pour les huit motifs bromodomains de la famille BET, à savoir BRD4(1), BRD4(2), BRD3(1), BRD3(2), BRD2(1), BRD2(2), BRDT(1) et BRDT(2), est présentée dans le **Tableau 6**. Ce travail m'a permis de constater que les huit domaines ne partagent entre eux que 43 à 54% d'identité de séquence (acide aminés). En revanche, au sein d'une même sous-famille BD1 ou BD2, les quatre domaines de chaque sous-famille présentent un pourcentage d'identité de séquence beaucoup plus élevé. Prenons l'exemple de la sous-famille BD1, le domaine BRD4(1) partage entre 74 et 92% d'identité de séquence (acide aminés) avec les modules BRD3(1) et BRD2(1) et BRDT(1). Les domaines de la sous-famille BD2 partagent quant à eux entre 71 et 90% d'identité de séquence (acide aminés).

Tableau 6 : Identité de séquence au sein de la famille BET.

	BRD2(1)	BRD3(1)	BRD4(1)	BRDT(1)	BRD2(2)	BRD3(2)	BRD4(2)	BRDT(2)	
BRD2(1)	100	91,8	87,7	74,0	53,4	46,6	50,7	43,8	Identité de séquence <div style="border: 1px solid black; padding: 2px; margin: 2px;">BD1 74% à 92%</div> <div style="border: 1px solid black; padding: 2px; margin: 2px;">BD2 71% à 90%</div> <div style="border: 1px solid black; padding: 2px; margin: 2px;">BD1 vs BD2 43 à 54%</div>
BRD3(1)	91,8	100	89,0	73,3	52,0	46,6	50,7	45,2	
BRD4(1)	87,7	89,0	100	79,5	50,7	45,2	50,7	43,8	
BRDT(1)	74,0	73,3	79,5	100	50,7	45,2	47,9	45,2	
BRD2(2)	53,4	52,0	50,7	50,7	100	89,0	89,9	71,2	
BRD3(2)	46,6	46,6	45,2	45,2	89,0	100	83,6	72,6	
BRD4(2)	50,7	50,7	50,7	47,9	89,9	83,6	100	79,4	
BRDT(2)	43,8	45,2	43,8	45,2	71,2	72,6	79,4	100	

Légende : L'identité de séquence en acide aminés entre les différents huit membres de la famille BET a été établie à l'aide de l'outil SIAS (<http://imed.med.ucm.es/Tools/sias.html>) en utilisant la matrice BLOSUM62.

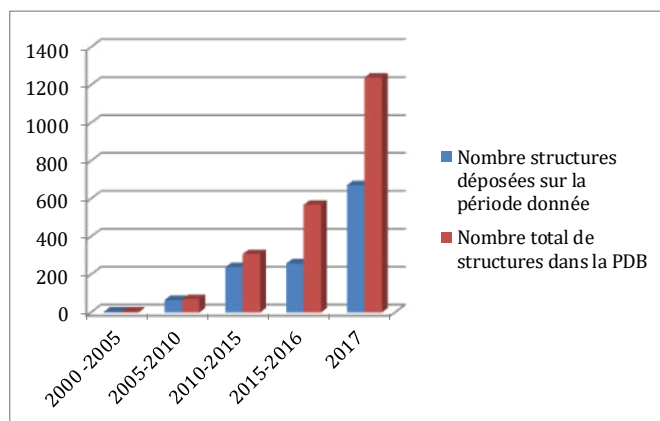
b. Structure tridimensionnelle

i. Historique

Pour rappel, le motif bromodomaine a été identifié pour la première fois en 1992 par **Haynes *et al*** (99), dans le gène brahma de la Drosophile. Quelques années plus tard la première structure tridimensionnelle d'un BD, celui de la protéine PCAF sous forme libre (169) (PDB : **1N72**), était résolue par RMN par **Zhou *et al***. Puis en 2002, cette même équipe a résolu la structure du BD de la protéine PCAF en complexe avec le peptide Tat AcK50 (170) (PDB : **1JM4**). En 2000, **Jacobson *et al*** publie la première structure cristallographique d'un BD, celui de la protéine TAFII (171) (PDB : **1EQF**). La première structure résolue d'un BD en interaction avec son peptide acétylé, fut celle de la protéine Gcn5p de *Saccharomyces cerevisiae* en présence d'un court peptide GGK_{ac}A (172) (PDB : **1E6I**). Publiée en 2000 par **Owen *et al***, cette structure cristallographique a permis de mettre en évidence un résidu clef responsable de l'interaction avec le peptide, une asparagine (N) conservée chez la plupart des bromodomains. Puis en 2005, l'équipe du **Dr. Zhou** a résolu les structures du BD de la protéine PCAF en complexe avec deux inhibiteurs de la PPI BD-peptide acétylé, les molécules NP1 et NP2 (173) (PDB : **1WUG** et **1WUM**). Ces inhibiteurs ont été identifiés par RMN par un criblage d'une librairie de petites molécules.

De nombreuses équipes à travers le monde s'intéressent aux protéines à bromodomains dans le contexte du développement de petites molécules inhibitrices, ce qui explique que le nombre de nouvelles structures de BDs en complexe avec des inhibiteurs soit en pleine expansion ces dernières années. Dans un article de 2016 (174), **Romero *et al***, présente une figure avec le nombre de structures de bromodomains publiées chaque année. J'ai actualisé ces données à partir des informations fournies par la PDB (**Figure 11**). À ce jour, 1236 structures RMN (2.9%) ou cristallographiques (97.1%) de BDs humains sont disponibles dans la PDB, avec au moins un représentant pour chaque famille (Recherche « human bromodomains » et « homosapiens only », dans la PDB <https://www.rcsb.org/pdb/home/home.do> le 27/08/2017). Une grande partie de ces structures, a été résolue au sein du Consortium de Génomique Structurale. Par ailleurs, 2017 est une année record, avec 692 structures déjà déposées dans la PDB depuis janvier. Certaines familles sont plus étudiées que d'autres, c'est notamment le cas de la famille BET avec 262 structures (dont 260 BDs et 2 ET), et dont 161 pour le domaine BD1 de la protéine BRD4. (Le 27/08/2017 : 217 structures pour BRD4).

Figure 11 : Évolution du nombre de structures de BDs dans la PDB.



Légende : Depuis 2000, le nombre de structure de bromodomains humains seuls ou en complexes avec un ligand (peptide ou inhibiteur) n'a cessé d'augmenter. L'année 2017, est une année record, avec plus de 690 structures déposées dans la PDB.

ii. Repliement et résidus clefs conservés

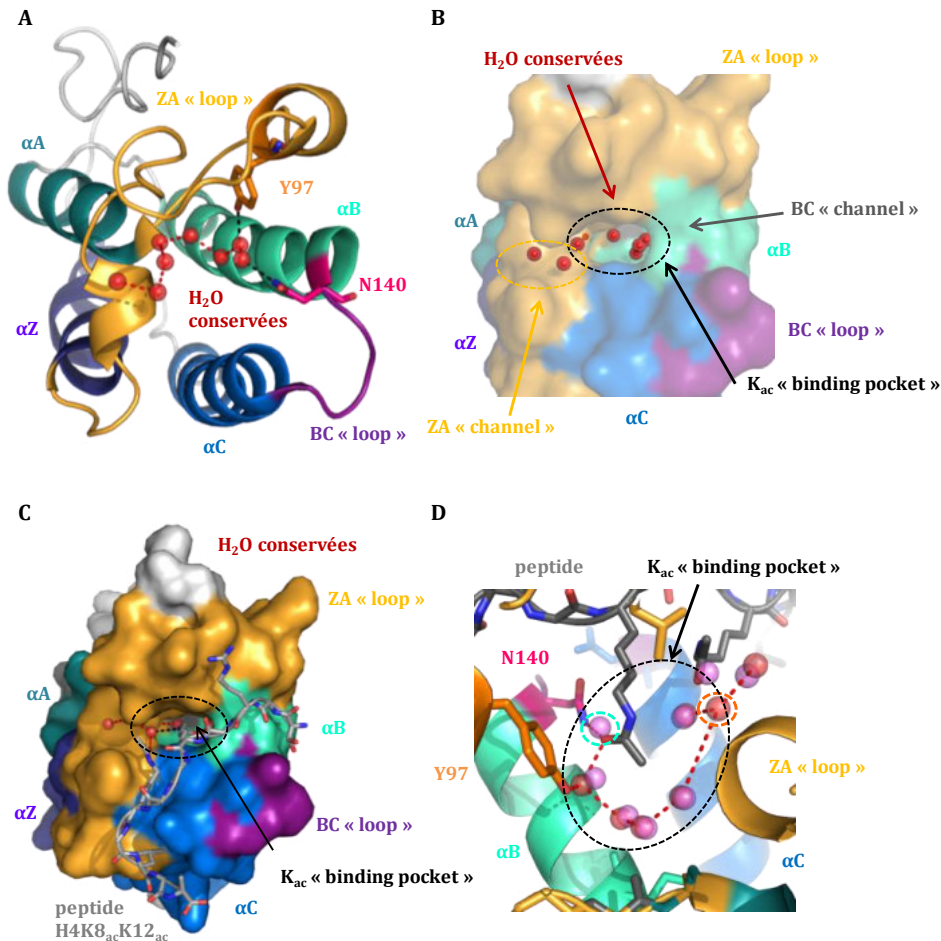
Bien qu'uniquement 21% d'identité de séquence soit observé, les modules BDs possèdent un repliement conservé au cours de l'évolution, aussi bien chez la drosophile (Fsh de *Drosophila*), la levure (Bdf1 et Bdf2 de *Saccharomyces cerevisiae*) et les mammifères (175, 176).

Le module BD comprend environ 110 acides aminés, dont 60 sont très conservés (176). Comme indiqué dans la **Figure 12-A**, du point de vue structural leur repliement protéique est formé de **quatre hélices α antiparallèles** (α_Z , α_A , α_B , α_C) et de **deux boucles d'insertion flexibles et variables (BC et ZA loops)** (57, 103, 177). Les résidus de ces boucles forment les canaux appelés « BC et ZA channels » respectivement. Les quatre hélices α forment quant à elles, une cavité hydrophobe profonde très caractéristique des BDs, aussi appelée « **K_{ac} binding pocket** » dans laquelle se trouve des molécules d'eau enfouies et où vient se fixer le peptide histone lysines acétylées.

Les motifs BDs présentent deux résidus clefs conservés une **asparagine** et une **tyrosine** respectivement en position 140 et 97 chez le domaine BD1 de la protéine BRD4 (ou BRD4(1)), ainsi que des molécules d'eau située dans la cavité K_{ac}, comme indiqué **Figure 12-B**. Cette asparagine N140 est responsable de la reconnaissance des histones acétylées chez les BDs « canoniques ». Le résidu N140, est situé au début de boucle BC à l'entrée de la poche de fixation K_{ac}. Ce mode de liaison a été décrit pour la première fois par **Dhalluin et al** pour la structure du domaine PCAF (169). La reconnaissance du peptide histone acétylé (**Figure 12-C & D**) se fait d'une part *via* l'établissement d'une interaction directe de type liaison hydrogène entre l'atome d'oxygène du groupement acétyle du peptide et le groupe amine (NH₂) du groupement amide

du résidu d'asparagine conservé (48). D'autre part, *via* une interaction indirecte de type liaison hydrogène entre le groupe hydroxyle (OH) du groupement phénol de la tyrosine et une des molécules d'eau conservées dans la poche K_{ac} .

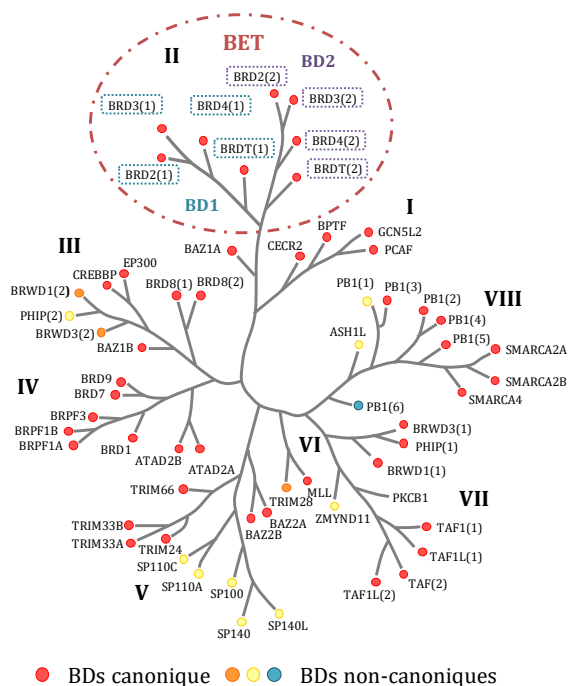
Figure 12 : Architecture des BDs et mode de liaison avec les histones acétylées.



Légende : Structure cristallographique du domaine BD1 de BRD4 ou BRD4(1) sous forme libre (PDB : **2OSS**) en **A & B** ou en complexe avec le peptide H4K8_{ac}K12_{ac} (PDB : **3UVW**) en **C & D**. **A**- Représentation « cartoon ». Les BDs sont formés de quatre hélices α (α_A , α_Z , α_B , α_C ; bleu) et de deux boucles ZA (jaune) et BC (violet). Les résidus Y97 (orange) et N140 (rose) sont représentés en « sticks ». Les molécules d'eau conservées (sphères rouges) qui établissent un réseau d'interaction de type liaisons hydrogènes entre elles (lignes en pointillés rouge ---). Les résidus conservés N140 et Y97 du domaine BRD4(1) établissent tous deux une liaison hydrogène (ligne pointillés noir ---) avec l'une de ces molécules d'eau. **B**- Zoom sur la cavité K_{ac} . Représentation de BRD4(1) en « surface ». Le même code couleur est utilisé. Le repliement de ces hélices reliées entre elles par les deux boucles, forment une cavité profonde caractéristique, appelée « K_{ac} binding pocket » (entourée en pointillés noir ---) dans laquelle sont enfouies pour la plupart des bromodomaines cinq molécules d'eau. Les membres de la famille BET présentent également deux molécules d'eau situées dans le « ZA channel » (entourées en pointillés orange ---). **C**- Représentation en « surface » pour BRD4(1) et en « stick » gris pour le peptide histone di-acétylé H4K8_{ac}K12_{ac}. Le peptide se fixe à l'intérieur de la poche K_{ac} , avec son extrémité N-terminale au niveau du « BC channel » et son extrémité C-terminale entre la boucle ZA et hélice α_C . **D**- Zoom sur la cavité K_{ac} , représentation « cartoon ». Les molécules d'eau présentes dans la forme libre de BRD4(1) sont représentées en violet, celles présentes au sein du complexe BRD4(1)-H4K8_{ac}K12_{ac} sont représentées en rouge. Dans la structure du complexe BRD4(1)-H4K8_{ac}K12_{ac}, une des molécules d'eau (entouré en pointillés bleu ---) proche de l'hélice α_B (bleu) est absente et une autre (entouré en pointillés orange ---) proche de la boucle ZA (orange) est déplacée, par rapport à la forme libre. Le peptide établit des interactions (lignes pointillées rouge ---) : directe avec l'asparagine conservée N140 et indirecte médiée par une molécule d'eau, avec la tyrosine conservée Y97 et une autre liaison avec l'une des molécules d'eau conservées située à l'entrée de la poche de fixation K_{ac} .

Bien que la plupart des bromodomains aient un repliement canonique, douze sont « non-canoniques » (178), présentent à la place de cette asparagine (N), soit un résidu thréonine (T) : pour BRWD1(2) BRDW3(2), TRIM28, soit un résidu tyrosine (Y) : pour ASHL1, PB1(1), PHIP(2), SP100, SP110, SP140, SP140L, ZMYND11 ou encore un résidu d'acide aspartique (D) pour PB1(6). La **Figure 13**, indique les BDs canoniques et non-canoniques. Tous les membres des familles I, II, IV sont canoniques. Les familles III, V, VI, VII et VIII présentent quant à elles des BDs canoniques et non-canoniques.

Figure 13 : BDs canoniques et non-canoniques.



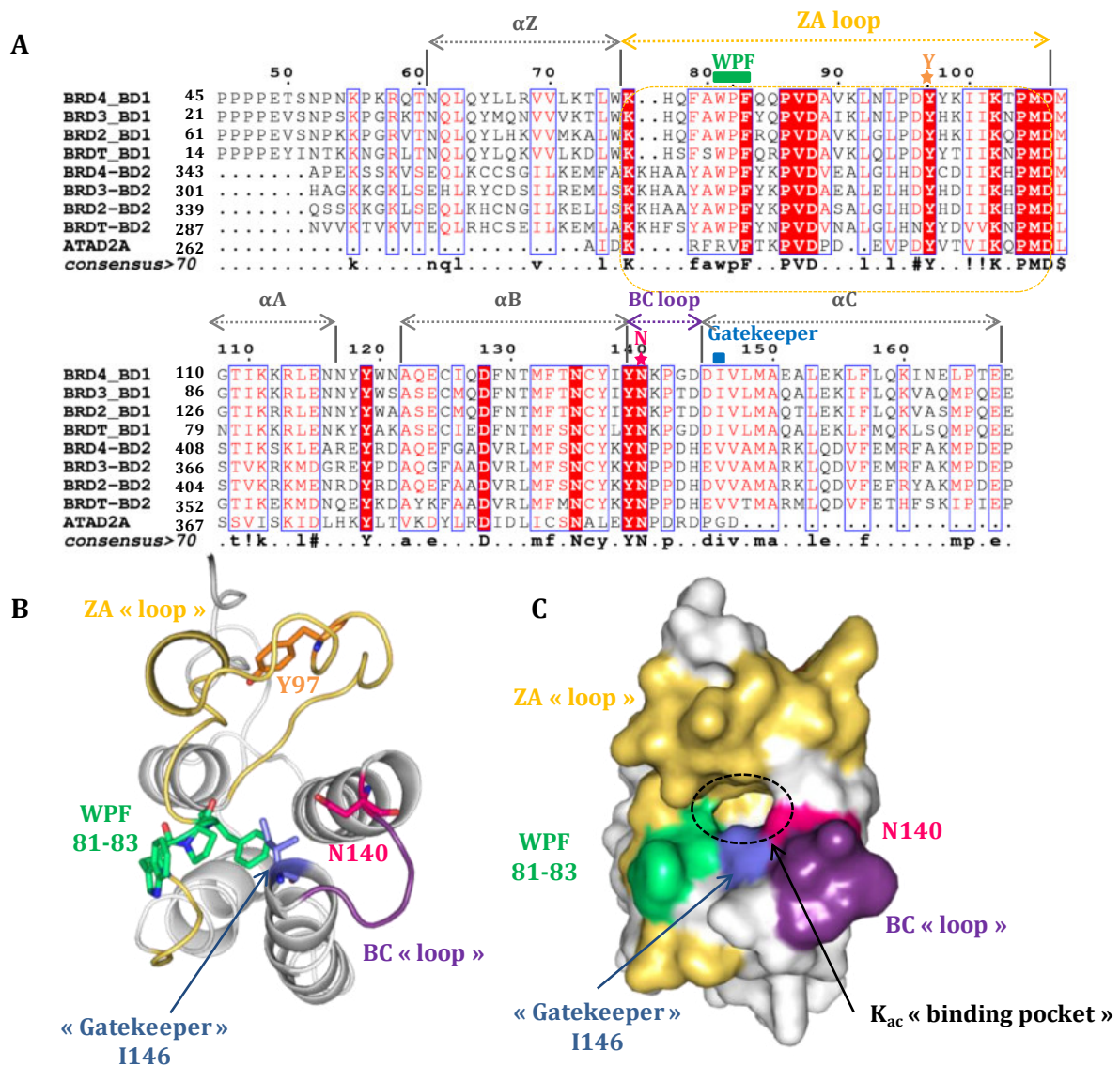
Légende : Les BDs canoniques sont indiqués par des cercles rouges (N conservée). Les BDs non-canoniques sont indiqués par des cercles oranges (N remplacée par T), jaunes (N remplacée par Y), et bleus (N remplacée par D).

Jusqu'à 2016, aucune étude n'avait montré que ce résidu clef puisse avoir un impact sur la fixation du peptide acétylé, mais **Meslamani et al**, décrivent ces protéines non-canoniques comme incapables de lier les lysines des histones acétylées (178). Le résidu Y97 est situé tout proche de la cavité K_{ac} , à l'extrémité C-terminale de la boucle ZA et il est lui aussi impliqué dans cette interaction protéine-ligand. Enfin, cinq molécules d'eau conservées (109) dans la plupart des BDs sont situées au fond de la poche de fixation des K_{ac} et une autre se situe à l'entrée de celle-ci. Ces molécules d'eau jouent un rôle essentiel dans la stabilisation du complexe BD-ligand. De plus, le réseau de liaisons hydrogènes s'étend jusqu'au « ZA channel » avec deux molécules d'eau supplémentaires chez les membres de la famille BET, ainsi que chez la protéine CREBBP (179, 180).

iii. Différences structurales

La vaste étude structurale menée en 2012, par l'équipe du **Dr. Knapp**, a permis d'identifier les différences structurales entre les familles distinctes de BD (103, 181). La différence notable observée est la présence d'une boucle insertion en épingle à cheveux « hairpin » localisée entre l'hélice α_z et la boucle ZA. Ce type d'insertion est fréquemment trouvé au niveau de la boucle ZA, ce qui entraîne des différences importantes en termes de taille, d'accessibilité et de charge de la poche de fixation K_{ac} . **Fillipakopoulos et al**, ont montré que le potentiel électrostatique autour du site de liaison K_{ac} est très variable, allant d'une valeur chargée très positivement à très négativement comme indiqué dans la **Figure 14**, ce qui suggère que les BDs sont capables de reconnaître une large gamme de séquences. Cependant, la surface du domaine 3 de la protéine PB1 ou PB1(3), est chargée positivement, ce qui n'est normalement pas compatible avec la fixation d'histones lysines acétylées qui sont très basiques. **La présence du site de liaison K_{ac} à proximité de cette boucle, suggère que la boucle ZA pourrait jouer un rôle dans le recrutement de partenaires d'interaction acétylés et donc dans la fonction biologique.** Par ailleurs, **Fillipakopoulos et al**, ont suggéré que les résidus hydrophobes situés au niveau de cette boucle ZA, pourraient contribuer à l'instabilité des BDs et de ce fait au faible taux de réussite de leur cristallisation, notamment pour les BDs des familles VI et V. Les extrémités N- et C- terminales sont également très divergentes et peuvent contenir une hélice supplémentaire qui prolonge le repliement canonique des BDs, comme c'est le cas pour les six BDs de la protéine PB1, ou encore une hélice « twisted » qui se prolonge largement au niveau des extrémités N- ou C- terminales, pour les protéines TAF1L ou ATAD2A, par exemples.

Figure 15 : Architecture et résidus clés des BDs de la famille BET.



Légende : **A-** Alignement des séquences des huit membres de la famille BET et d'ATAD2A, un non-BET (famille IV), réalisé sur le serveur Clustal Omega (<http://www.ebi.ac.uk/Tools/msa/clustalo/>). Graphique généré sur le serveur Espript 3.0 (<http://espript.ibcp.fr/ESPrIPT/ESPrIPT/>), puis complété à la main. Les flèches en pointillés représentent, en gris les quatre hélices α , en orange la boucle ZA et en violet la boucle BC. Les deux résidus conservés, asparagine (N140 pour BRD4(1), rose) et tyrosine (Y97 pour BRD4(1), orange) sont indiqués par une étoile. Le motif « WPF » (81-83, vert) et le résidu « gatekeeper » (I146, bleu) sont indiqués par des rectangles. **B-** Représentation « cartoon » du domaine BRD4(1) (PDB : 20SS). Les boucles ZA et BC sont colorée en jaune et violet, respectivement. Les résidus clés sont représentés en « sticks » avec le même code couleur. **C-** Représentation « surface » du domaine BRD4(1). Le même code couleur que précédemment est utilisé. La poche Kac formée par le repliement des hélices α et des boucles est indiquée par un cercle en pointillés noir.

Le résidu « gatekeeper » joue un rôle important dans la fixation des ligands, puisqu'il définit la taille de l'accès à la poche K_{ac} . Le « gatekeeper » est généralement une isoleucine (I), leucine (L) ou valine (V), cependant certains BDs, comme par exemple celui de la protéine GCN5L2, présentent à la place un résidu plus volumineux, une tyrosine (Y) qui peut empêcher le ligand d'accéder à la poche K_{ac} en bloquant l'entrée du « WPF shelf » (182). Les domaines BD1 des protéines à bromodomains de la famille BET présentent une isoleucine (I), tandis que les domaines BD2 présentent une valine (V) (Tableau 7).

Tableau 7 : Alignements du motif « WPF » et du résidu « gatekeeper ».

Famille	Bromodomains	Résidus du motif "WPF"			Gatekeeper
I	CECR2	W	P	F	Y
I	PCAF	W	P	F	Y
II	BRD4(1)	W	P	F	I
II	BRD4(2)	W	P	F	V
III	CREBBP	L	P	F	V
IV	ATAD2A	R	V	F	I
IV	BRD9	G	F	F	Y
IV	BRPF1	N	I	F	F
IV	BRPF3	H	I	F	F
V	BAZ2B	W	P	F	I
V	TRIM24	L	A	F	V
VII	TRIM28	R	P	L	V
VIII	SMARCA4	Q	V	F	I

Légende : Figure adapté de Romero et al (174), complété avec TRIM28 (famille VII).

Le challenge actuel est de développer des inhibiteurs très affins de l'ordre du nanomolaire (nM) et sélectifs de l'une des sous-familles BD1 et BD2 ou plus idéalement d'un seul domaine de la famille BET, afin de mieux appréhender le rôle biologique associés aux BD1 et aux BD2 et plus idéalement à chaque BDs. Dans ce contexte, j'ai été amenée à analyser les séquences en acides aminés et les structures tridimensionnelles de l'ensemble des BDs de la famille BET.

v. Étude comparative séquences / structures 3D entre BDs de la famille BET

Dans un premier temps, j'ai effectué trois types d'alignements de séquences (acides aminés), à savoir, les quatre BDs de la sous-famille BD1, ceux de la sous-famille BD2 et enfin les huit BDs de la famille BET, sur le serveur Clustal Omega (<http://www.ebi.ac.uk/Tools/msa/clustalo/>). Puis j'ai utilisé les fichiers générés par Clustal Omega pour réaliser mes figures d'alignement de séquences (acides aminés) via le serveur Esript 3.0 (<http://esript.ibcp.fr/ESPript/ESPript/>).

Dans un second temps, j'ai téléchargé dans le logiciel de visualisation Pymol, les structures suivantes : BRD4(1) (PDB : **2OSS**), BRD4(2) (PDB : **2OUO**) ; BRD3(1) (PDB : **2NXB**), BRD3(2) (PDB : **2001**), BRD2(1) (PDB : **1X0J**), BRD2(2) (PDB : **5IBN**) et BRDT(1) (PDB : **2RFJ**) , aucune structure cristallographique n'est disponible pour le domaine BRDT(2) et j'ai superposé les différentes structures sur une même figure.

Enfin, j'ai défini vingt-trois résidus situés autour de la poche de fixation K_{ac} , d'après les résidus décrits dans la littérature et ceux que j'ai pu identifier en comparant les structures du domaine BRD4(1) sous forme libre (PDB : **2OSS**) et en complexe avec différents peptides acétylés (PDB : **3UVW**, **3UVY**, **3UVX**, **3UW9**). Seize d'entre eux, à savoir les résidus W81, P82, F83, Q85, Y97, M105, M132, N135, C136, Y139, N140, K141, D144, D145, I146 et M149, se situent à 3,5 Å du peptide H4K16_{ac}K20_{ac} (PDB : **3UVY**) et les sept autres à savoir les résidus F79, P86, V87, D88, L92, L94, D106 se trouvent à 4,5 Å de ce peptide.

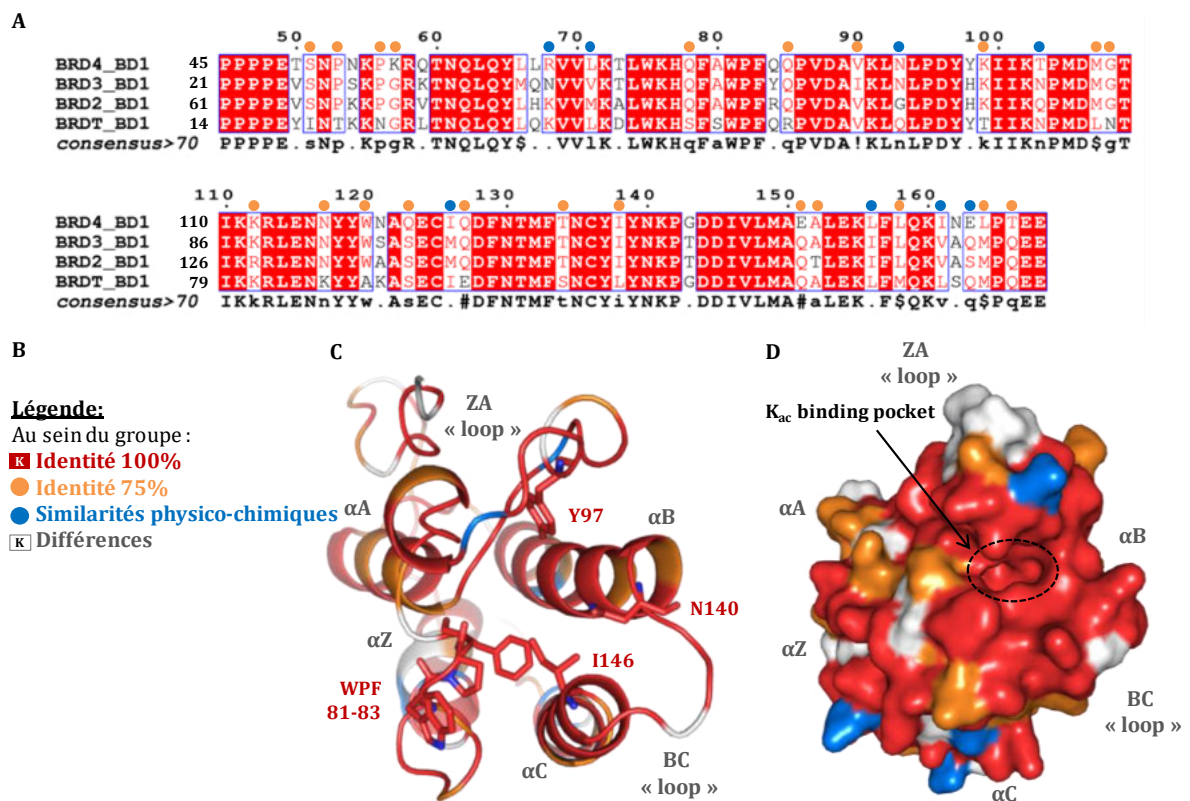
Les observations que j'ai pu déduire de la comparaison des alignements de séquences (acide aminés) et des structures cristallographiques sont décrites dans les sections suivantes. Dans un premier temps j'ai confronté les données pour les quatre BDs de la sous-famille BD1 (**Figure 16**), puis celles pour les quatre BDs de la sous-famille BD2 (**Figure 17**), et enfin celles pour les huit BDs de la famille BET (**Figure 18**).

Comparaison des BD1

La **Figure 16** présente la comparaison que j'ai faite pour les quatre domaines BD1 de la famille BET. Au sein de la sous-famille BD1, j'ai pu constater que les résidus autour de la cavité K_{ac} sont tous identiques, excepté un qui diffère pour la protéine BRDT(1). Au niveau de la boucle ZA, les résidus sont pour la plupart identiques à 100% (23 résidus sur 31, soit 74,2%), sauf quatre résidus qui ne le sont qu'à 75%, deux autres qui présentent des propriétés physico-chimiques similaires et enfin deux qui diffèrent entre les protéines.

En regardant plus précisément au niveau du « ZA channel », j'ai observé deux résidus conservés à 100%, un conservé à 75% et un qui diffère entre les protéines, **ce qui laisse supposer qu'on pourrait cibler certaines protéines des BD1 dans cette zone**. Par ailleurs, la boucle ZA qui est très flexible, est basculée vers l'avant ce qui referme la cavité sur elle-même et rend la poche de fixation K_{ac} des BD1 moins accessible. Enfin, au niveau du « WPF shelf » formé par la boucle ZA et le début de l'hélice α_c , les résidus sont identiques à 100% (six premiers résidus de l'hélice α_c).

Figure 16 : Comparaison des BDs au sein de la sous-famille BD1.

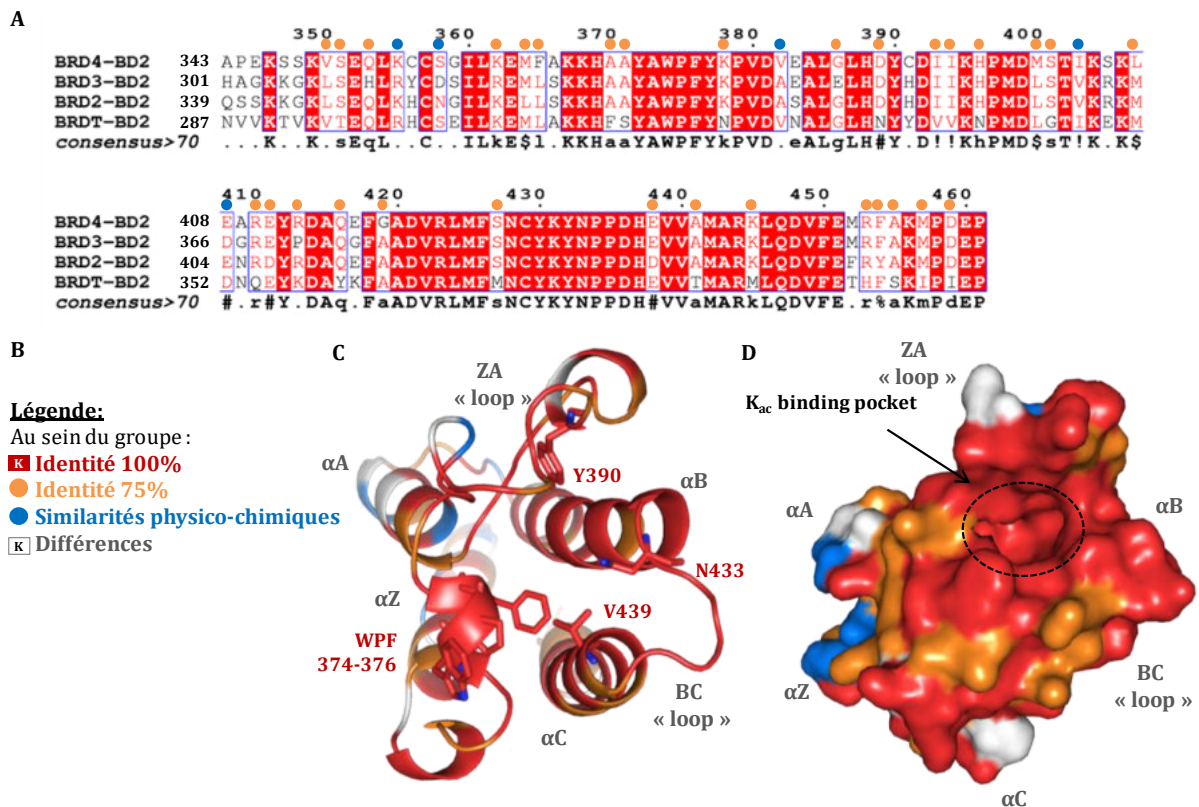


Légende : A- Alignement des séquences des quatre domaines de la sous-famille BD1, avec la légende correspondante en B. Les résidus conservés à 100% sont indiqués en blanc dans des cadres rouges et les non conservés sont indiqués en noir. Les résidus conservés à 75%, c'est-à-dire pour trois des BD1 sont indiqués par un rond orange. Les résidus différents mais présentant des propriétés physico-chimiques similaires sont indiqués par un rond bleu. C- Structure du domaine BRD4(1) (PDB : 20SS) en représentation « cartoon ». Les résidus clefs, à savoir l'asparagine (N140), la tyrosine (Y97), le « WPF » (81-83) et le « gatekeeper » (I146) sont représentés en « stick ». La structure est colorée avec le même code couleur que la légende. D- Structure du domaine BRD4(1) (PDB : 20SS) en représentation « surface ». Le même code couleur que précédemment est utilisé. La cavité K_{ac} est représentée par un cercle noir en pointillés.

Comparaison des BD2

La **Figure 17** présente la comparaison que j'ai faite pour les quatre BD2 de la famille BET. Au sein de la sous-famille BD2, j'ai pu constater que les résidus autour de la cavité K_{ac} sont eux aussi tous identiques. Au niveau de la boucle ZA, ils le sont également pour la plupart à 100% (22 résidus sur 33, soit 66,7%) excepté huit résidus qui sont identiques à 75%, un résidu qui présentent des propriétés physico-chimiques similaires et deux résidus qui diffèrent entre les protéines. Cette boucle semble donc moins conservée chez les BD2 (66,7% identité) que chez les BD1 (74,2% identité), ce qui porte à croire **qu'il serait peut-être plus facile de cibler sélectivement les domaines spécifiques au sein de la sous-famille BD2**. Enfin, au niveau du canal formé par le motif WPF ou « WPF shelf » quatre résidus sont identiques à 100% et deux sont conservés à 75% (six premiers résidus de l'hélice α_C).

Figure 17 : Comparaison des BDs au sein de la sous-famille BD2.

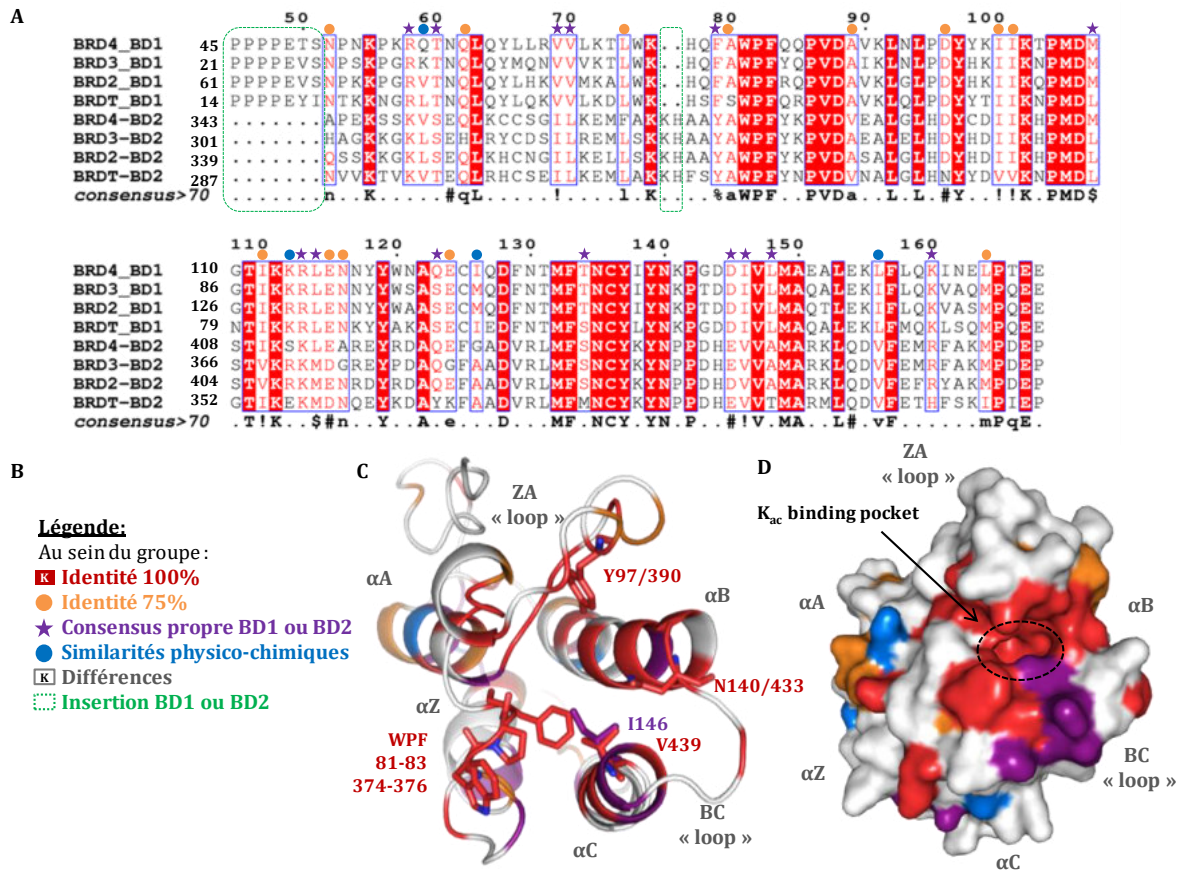


Légende : **A-** Alignement des séquences des quatre domaines de la sous-famille BD2, avec la légende correspondante en **B**. Les résidus conservés à 100% sont indiqués en blanc dans des cadres rouges et les non conservés sont indiqués en noir. Les résidus conservés à 75%, c'est-à-dire pour trois des BD2 sont indiqués par un rond orange. Les résidus différents mais présentant des propriétés physico-chimiques similaires sont indiqués par un rond bleu. **C-** Structure du domaine BRD4(2) (PDB : **20U0**) en représentation « cartoon ». Les résidus clefs, à savoir l'asparagine (N433), la tyrosine (Y390), le « WPF » (374-376) et le « gatekeeper » (V439) sont représentés en « stick ». La structure est colorée avec le même code couleur que la légende. **D-** Structure du domaine BRD4(2) (PDB : **20U0**) en représentation « surface ». Le même code couleur que précédemment est utilisé. La cavité K_{ac} est représentée par un cercle noir en pointillés.

Comparaison des BETs

La **Figure 18** présente la comparaison que j'ai faite entre les huit BDs de la famille BET. Tout d'abord, les domaines BD1 présentent sept résidus supplémentaires, dont un motif proline-proline, à l'extrémité N-terminale. Parmi les résidus impliqués dans l'interaction avec les lysines acétylées, j'ai observé des différences entre les BD1 et les BD2. Au niveau de la boucle ZA, un résidu glutamine (Q) pour les BD1 est remplacé par un résidu lysine (K) pour les BD2 (Q85 pour BRD4(1) versus K378 pour BRD4(2)) et un résidu phénylalanine (F) pour les BD1 est remplacé par un résidu tyrosine (Y) pour les BD2 (F79 pour BRD4(1) versus Y372 pour BRD4(2)). Au niveau de la boucle BC, un résidu acide aspartique (D) pour les BD1 est remplacé par un résidu histidine (H) pour les BD2 (D144 pour BRD4(1) versus H437 pour BRD4(2)) et un résidu acide aspartique (D) pour les BD1 est remplacé par un résidu acide glutamique (E) pour les BD2 (D145 pour BRD4(1) versus E438 pour BRD4(2)). Enfin au niveau de l'hélice α , le résidu « gatekeeper » est une isoleucine (I) pour les BD1 et une valine (V) pour les BD2, (I146 BRD4(1) versus V439 pour BRD4(2)). Certains de ces résidus (Q85, D144, I146 pour BRD4(1)) ont été décrit par **Galdeano *et al***, comme trois résidus clefs qui pourraient servir au développement de molécules sélectives des BD1 ou des BD2 (183). Au niveau de la boucle ZA, j'ai pu constater qu'il y a une insertion de deux résidus successifs une lysine et une histidine (KH) pour les BD2 et situés au début de cette boucle. Lorsqu'on compare les alignements de séquence et les structures à ce niveau constate que cette insertion induit un décalage d'un résidu supplémentaire pour l'alignement des structures des BD1 sur le BD2 par rapport à l'alignement de leur séquence d'acides aminés. Les BD2 pourraient éventuellement être ciblés *via* ces deux résidus. De plus, cette boucle est très flexible et diffère entre les BD1 et les BD2 (42% d'identité). En effet, seul quatorze résidus sont identiques à 100%, et cinq résidus identiques à 75%. Par ailleurs, la flexibilité de cette boucle permet de réguler l'ouverture et la fermeture de la cavité K_{ac} . Dans le cas des BD1 l'entrée de la poche K_{ac} est plus étroite que celles des BD2. Enfin au niveau du « WPF shelf », j'ai identifié un motif consensus situé au début de l'hélice α_c , pour les BD1 ce motif est « DIVLMA » et pour les BD2 c'est « EVVAMA », ce qui conduit à penser que **l'on pourrait cibler sélectivement soit les BD1 soit les BD2 au niveau de ce « WPF shelf »**.

Figure 18 : Comparaison des BD1 et BD2 au sein de la famille BET.



Légende : A- Alignement des séquences des huit BDs de la famille BET, avec la légende correspondante en B. Les résidus conservés à 100% sont indiqués en blanc dans des cadres rouges et les non conservés sont indiqués en noir. Les résidus conservés à 75% c'est-à-dire conservés pour au moins six des BDs de la famille BET, sont indiqués par un rond orange. Les résidus formant « consensus », pour l'une des sous-familles BD1 versus BD2, sont indiqués par une étoile violette. Les résidus différents mais présentant des propriétés physico-chimiques similaires sont indiqués par un rond bleu. Les domaines BD1 présentent sept résidus supplémentaires, dont un motif proline-proline, à l'extrémité N-terminale, tandis que les domaines BD2 présentent une insertion de deux résidus au début de la boucle ZA. C- Structure du domaine BRD4(1) (PDB : 20SS) en représentation « cartoon ». Les résidus clefs, à savoir les asparagines (N140/N433), les tyrosines (Y97/Y390), les « gatekeepers » (I146 (violet)/V439 (rouge)) sont représentés en « stick ». La structure est représentée avec le même code couleur que la légende. D- Structure du domaine BRD4(1) (PDB : 20SS) en représentation « surface ». Le même code couleur que précédemment est utilisé. La cavité K_{ac} est représentée par un cercle noir en pointillés.

5. Inhibiteurs de BETs

Le développement d'inhibiteurs ciblant les protéines de la famille BET peut s'appuyer sur les connaissances des structures tridimensionnelles de ces protéines, ainsi que de leurs modes de liaison avec les histones acétylées qui ont été développées ces dix dernières années. Un nombre important d'articles publiés et de structures déposées dans les bases de données dédiées (PDB) permettent de guider le développement de nouvelles molécules.

La recherche d'inhibiteurs des PPI BDs-peptide acétylé a deux objectifs majeurs. Le premier est de développer des « *sondes chimiques* » capables de moduler la fonction des protéines afin d'étudier leur mécanisme d'action et d'élucider leur rôle biologique par exemples dans les processus d'activation ou répression de la transcription et de signalisation cellulaire. La base de données ChEpiM0d (<http://chepimod.org/>), répertorie les ligands (modulateurs) des « readers » épigénétiques, aussi bien pour les bromodomains, chromo-domaines, tudor-domaines, que pour les domaines MBT, PHD, PWWP. Le second objectif est d'identifier des « *candidats médicaments* » pouvant être utilisés dans le traitement des divers pathologies associées à une dérégulation épigénétique telles que les maladies inflammatoires, auto-immunes et du métabolisme ou encore les cancers (184, 185).

A. Définitions :

Une **sonde chimique** ou « chemical probes » est définie comme « *une petite molécule chimique* » capable de moduler la fonction d'une protéine ». Le plus souvent ce sont des inhibiteurs qui permettent d'étudier le mécanisme d'action et le phénotype d'une cible moléculaire au travers d'études biochimiques, biophysiques, cellulaires ou en modèle animal (184, 185).

Le terme **médicament** est quant à lui défini depuis 1967 par l'article L.5111-1 du code de la santé publique Français comme « *toute substance ou composition présentée comme possédant des propriétés curatives ou préventives à l'égard des maladies humaines ou animales, ainsi que tout produit pouvant être administré à l'homme ou à l'animal, en vue d'établir un diagnostic médical ou de restaurer, corriger, modifier leurs fonctions physiologiques en exerçant une action pharmacologique, immunologique ou métabolique* ».

B. Différences entre un « médicaments » et une « sonde chimique »

Un médicament doit répondre avant tout à des critères ADMETox de pharmacodynamique (ADMETox = Absorption, Distribution, Métabolisation, Excrétion, Toxicité) et de biodisponibilité. Il doit être efficace mais non toxique pour l'homme. Son mécanisme d'action n'est pas forcément identifié et il n'est pas nécessairement sélectif vis-à-vis de sa cible thérapeutique. En effet de nombreux médicament sont multi-cibles (poly-pharmacologie).

En revanche, **une sonde chimique** n'est pas obligée de répondre aux critères ADMETox et de biodisponibilité. Par contre, son mécanisme d'action doit être déterminé et elle doit présenter à la fois une très bonne affinité et une certaine sélectivité pour sa cible.

C'est pourquoi une bonne sonde chimique ne devient pas forcément un bon médicament et inversement un bon médicament n'est pas obligatoirement une bonne sonde chimique. Le site <http://www.chemicalprobes.org/> répertorie un grand nombre de sondes chimiques classées en fonction de la protéine qu'elles ciblent.

C. Sonde épigénétique :

La qualité d'une sonde est difficile à évaluer, néanmoins le Consortium de Génomique Structurale (<http://www.thesgc.org/>) en collaboration avec l'industrie pharmaceutique, ont établi des critères de « qualité » pour les sondes ciblant les protéines épigénétiques. Ainsi, une bonne sonde épigénétique doit répondre aux cinq critères suivants :

- Elle doit présenter une activité *in vitro* sur sa cible inférieure à 100 nM.
- Elle doit présenter un profil de sélectivité supérieure à 30 fois pour sa cible comparée aux autres protéines appartenant à la même famille.
- Elle doit avoir été testée sur d'autres familles de protéines pertinentes.
- Elle ne doit pas présenter d'effet « off-targets ». Les expériences de RNAi permettent notamment de déterminer si l'effet de la sonde est dû à la structure « scaffolding », ou à l'inhibition de l'activité catalytique ou encore à l'inhibition de l'interaction protéique.
- Elle doit présenter un effet cellulaire « on-target » inférieur à 1 μ M.

Par ailleurs, en plus de ces critères le SGC recommande de tester des analogues inactifs de la sonde identifiée comme contrôle négatif.

a. Inhibiteurs publiés

Depuis 2005, de nombreux inhibiteurs de BDs ont été caractérisés et plus particulièrement envers la famille BET (48, 178, 186). À ce jour, plus de deux cent vingt structures de domaines BDs de la famille BET en complexe avec des inhibiteurs sont disponibles dans la PDB. La plupart de ces molécules sont des **mimétiques des lysines acétylées**, certaines d'entre elles sont des **inhibiteurs doubles de kinases et de BDs « dual-kinase »** et quelques-unes sont des **non-mimétiques des lysines acétylées**. La suite de ce paragraphe a pour but de présenter de manière non-exhaustive les différents « cœurs chimiques » retrouvés parmi les inhibiteurs de BETs (Figure 19), ainsi que leur mode de fixation (Figure 20).

Figure 19 : Différents "cœurs chimiques" retrouvés parmi les inhibiteurs de BETs.

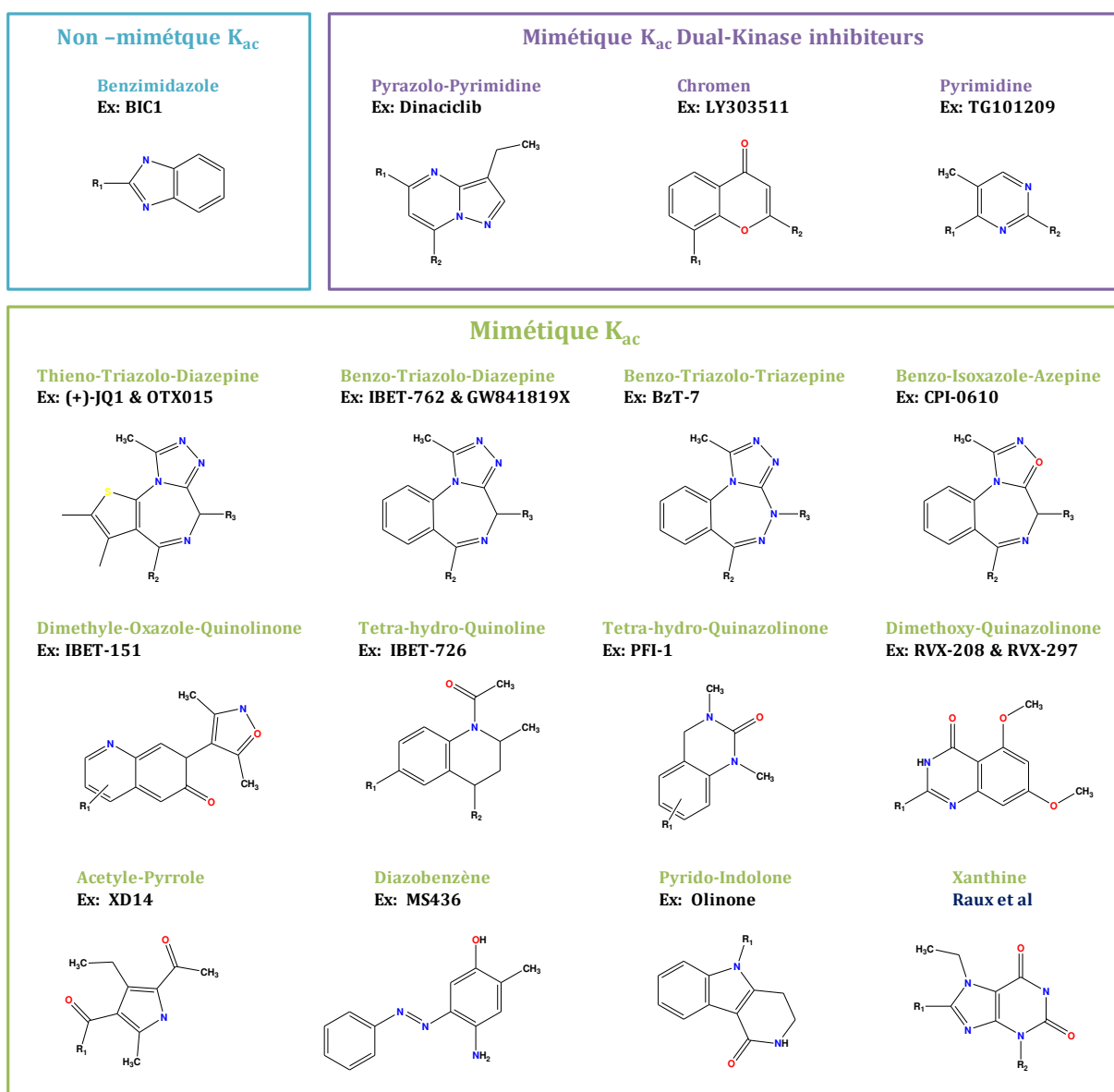
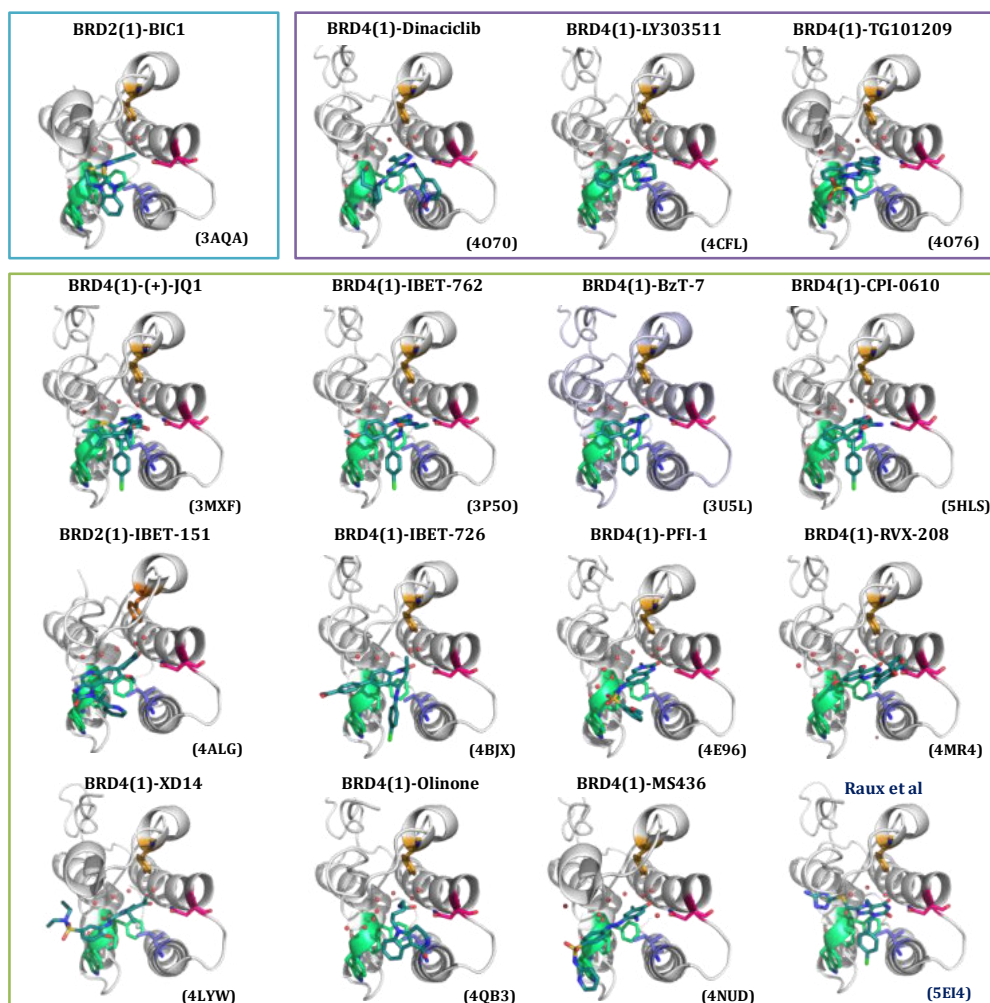


Figure 20 : Mode de liaison de différents inhibiteurs de BETs.



i. Inhibiteurs de BETs non-mimétiques

Le premier inhibiteur de BDs à avoir été caractérisé, la molécule **NP1** est un inhibiteur non-mimétique (173). Un autre inhibiteur de ce type a également été identifié, la molécule **BIC1** qui inhibe tous les domaines de la famille BET, ce qu'on appelle un « pan-BET inhibiteur ». Celle-ci inhibe notamment le domaine BRD2(1) avec un K_D de 28 μ M et bloque l'activité de transcription de cette protéine. La structure du complexe BRD2(1)/BIC1 (187) (PDB : **3AQA**), montre que le cœur chimique « mercapto-benzimidazole » entre dans la cavité K_{ac} , mais n'établit aucune interactions avec l'asparagine conservée N156, ce qui explique la faible activité de cette molécule. Une première interaction implique l'atome d'azote du groupe « benzimidazole » situé à 3,6 Å d'une molécule d'eau conservée dans la poche qui établit une interaction avec la glutamine Q101. Une deuxième interaction concerne l'atome de soufre en position R₁ du benzimidazole situé à 3,0 Å du résidu proline P98. Une troisième interaction fait intervenir le groupe phényle du second « benzimidazole » situé à 3,2 Å du résidu tryptophane W97.

ii. Inhibiteurs de BETS « dual-kinase »

La protéine BRD4 est une kinase atypique capable de phosphoryler le résidu Ser2 à l'extrémité C-terminal de l'ARN pol II (132), c'est pourquoi certains inhibiteurs de kinases touchent également les bromodomains (132, 188). Voici quelques exemples d'inhibiteurs de kinases qui sont de type pan-BET : La molécule **Dinaciclib**, un inhibiteur de kinases cycline-dépendante (ex : CDK2), présente une IC₅₀ de 19 µM sur le domaine BRD4(1) (189). La structure du complexe BRD4(1)/Dinaciclib (190) (PDB : **4O70**), montre que le cœur chimique « pyrazolopyrimidine » entre légèrement dans la poche de fixation K_{ac}, mais n'établit aucune interaction avec les molécules d'eau conservées au fond de celle-ci. En revanche, l'asparagine conservée N140 établit des interactions avec l'un des atomes d'azote du « pyrazolo-pyrimidine » ainsi qu'avec l'atome d'azote situé en position R₂. Le manque d'interaction avec les molécules d'eau dans la cavité permettant de stabiliser le complexe pourrait expliquer la faible affinité de l'inhibiteur avec une constante thermodynamique de dissociation de l'ordre du micromolaire. La molécule **LY30351**, un inhibiteur de la kinase PI3K (phosphoinositide 3-kinase), présente une IC₅₀ comprise entre 7 et 12 µM pour les BET et semble être un inhibiteur sélectif pour les domaines BD1. Par ailleurs, cette molécule a montré un effet immunosuppresseur sur des lignées cellulaires cancéreuses MV4-11. La structure du complexe BRD4(1)/LY30351 (191) (PDB : **4CFL**), montre que l'atome d'oxygène du cœur chimique « chromen » établit des liaisons hydrogènes avec l'asparagine conservée N140, ainsi qu'avec la tyrosine conservée Y97 par le biais d'une molécule d'eau conservée dans la cavité K_{ac}. De plus le groupe phenyl en R₁ se situe à 4,0 Å des résidus W81 et P82 du « WPF ». Enfin la molécule **TG101209**, un inhibiteur de la kinase PLK1 (polo-like kinase), présente une IC₅₀ de 130 nM pour le domaine BRD4(1). La structure du complexe BRD4(1)/TG101209 (190) (PDB : **4O76**), montre que les deux atomes d'azote du cœur chimique « pyrimidine » interagissent avec l'asparagine conservée N140. Cependant ce composé ne rentre pas au fond de la poche de fixation K_{ac} et n'établit aucune interaction avec les molécules d'eau conservées dans celle-ci. Par ailleurs, l'atome d'azote en position R₁ se situe à 3,6 Å du résidu proline P82 du « WPF ».

iii. Inhibiteurs de BETS mimétiques

De manière générale, les inhibiteurs sélectifs et affins de bromodomains « canoniques » sont mimétiques des lysines acétylées. Ils sont développés afin de mimer les résidus lysines acétylées des histones de manière à occuper le site d'interaction des histones acétylées des bromodomains. C'est pourquoi les inhibiteurs mimétiques sont développés en vue de conserver le mode de reconnaissance et de fixation des peptides acétylés par les BDs. Cette reconnaissance se fait au travers de divers interactions de type liaisons hydrogènes, à savoir une interaction

directe entre le peptide et un résidu asparagine conservé (N140 du domaine BRD4(1)), une interaction indirecte entre le peptide et un résidu tyrosine conservé (Y97 chez BRD4(1)) médiée par une molécule d'eau conservée dans la cavité K_{ac} . Par ailleurs, un réseau d'interactions entre les cinq molécules d'eau conservées dans la poche de fixation s'établit avec la protéine et le peptide afin de stabiliser la structure du complexe. Parmi les inhibiteurs de ce type, les plus représentés sont les analogues azépines, oxazoles, quinolines et pyrroles, néanmoins de nouveaux « cœurs chimiques » ont émergés récemment.

Dérivés azepines

Parmi les analogues azépines se trouvent les thienotriazolodiazépines (TTDs) et les benzotriazolodiazépines (BZDs). Des molécules de ce type sont actuellement utilisées comme anesthésiant, anxiolytique, immunosuppresseur et dans le traitement de l'insomnie ou de l'anxiété. Le premier inhibiteur de type pan-BET caractérisé en 2010, est la molécule **(+)-JQ1**, un dérivé de TTDs. Cette molécule présente un K_D de 49 et 90 nM pour les domaines BD1 et BD2 de la protéine BRD4 respectivement. De plus, l'énantiomère (+)-JQ1 présente une affinité cent fois meilleure que (-)-JQ1. Il a été montré que JQ1 induit l'apoptose en ciblant la protéine de fusion BRD4-NUT dans les cellules humaines de carcinomes NMC et diminue la mortalité dans les xénogreffe. Cette sonde chimique a permis d'identifier la protéine BRD4 comme cible dans le traitement de leucémies AML (146), dans la suppression de la croissance tumorale dans le cancer ovarien et dans l'atténuation de la signalisation des cellules souches dans les tumeurs cérébrales présentant une surexpression de l'oncogène c-Myc (192). La structure du complexe BRD4(1)/(+)-JQ1 (PDB : **3MXF**), montre que le cœur chimique « triazole » est en contact avec l'asparagine N140 et établit également des liaisons hydrophobes avec la boucle ZA, et notamment les résidus suivants : W81, P82, V87, L92, L94, Y97, Y139, C136 et M149. L'énantiomère (-)-JQ1 induit quant à lui une gêne stérique avec les résidus leucines L92 et L94, ce qui explique la perte d'affinité (57). Cette molécule ayant un temps de demi-vie court dans l'organisme n'a pas pu être évaluée en essais cliniques. En revanche la molécule **OTX015**, un analogue de JQ1 présentant une IC_{50} comprise entre 92 et 112 nM et une concentration efficace médiane ou EC_{50} comprise entre 10 et 19 nM (193) a pu être évaluée dans plusieurs essais cliniques. Cette molécule a montré un effet antiprolifératif sur de nombreuses lignées cellulaires cancéreuses humaines, notamment les tumeurs hématologiques (194), ainsi qu'un effet sur la croissance des tumeurs solides en induisant leur senescence et apoptose et une diminution de l'expression de c-Myc dans des cellules de leucémies aiguës (195).

Des inhibiteurs dérivés de BZDs ont également été développés pour cibler les BETs et, à ce jour, ils sont tous pan-BET. C'est le cas de la molécule **IBET-762** qui présente une valeur de K_D comprise entre 50 et 61 nM et une IC_{50} comprise entre 32 et 42 nM pour les BETs. Il a été montré que cette molécule inhibe la synthèse de lipopolysaccharide (LPS) qui stimule les chimiokines et cytokines dans les macrophages. Tout comme JQ1, elle induit l'arrêt du cycle cellulaire et conduit à un effet antiprolifératif dans le cas de myélomes (196). Cette molécule est actuellement en essais clinique. La structure du complexe BRD4(1)/IBET-762 (197) (PDB : **3P50**), montre que le cœur chimique « triazolo » entre profondément dans la cavité K_{ac} et est en contact avec l'asparagine N140 ainsi qu'avec le réseau de molécules d'eau situées dans la poche. L'affinité de cette molécule serait expliquée par l'interaction qu'elle établit avec les résidus hydrophobes du « ZA channel », ainsi qu'avec le « WPF » et le résidu « gatekeeper » I146. Une autre molécule, **GW841819X**, qui est un inducteur de la protéine Apolipoprotein A1 (ApoA1) touche également les protéines de la famille BET (K_D de 50 nM pour BRD2). Elle présente une activité anti-inflammatoire en inhibant la production d'interleukine 6 (IL-6) et bloque l'interaction entre BRD3 et le peptide GATA1. Des essais précliniques pour l'utilisation de GW841819X sont en cours, cependant à cause de sa faible biodisponibilité due à sa dégradation rapide à pH acide, elle ne peut être évaluée en essais clinique. La structure du complexe BRD4(1)/ GW841819X (154) (PDB : **2YEL**), montre que le groupe carbamate est en contact avec la cavité K_{ac} . La molécule **BzT-7** est quant à elle un dérivé benzotriazepine, qui diffère des BZDs par l'ajout d'un groupe nitrogène à la place du carbone chiral. Cette molécule pan-BET, présente est un inhibiteur de BRD4(1) avec un K_D de 640 nM. La structure du complexe BRD4(1)/BzT-z (198) (PDB : **3U5L**) montre que l'atome d'azote du cœur chimique « triazolo » est en contact avec l'asparagine N140, ainsi que la tyrosine Y97 *via* une molécule d'eau située dans la poche, cependant il ne pénètre pas au fond de celle-ci. L'atome d'azote situé à la place du carbone chiral des diazépines se trouve à 4,2 Å de l'asparagine N140. Par ailleurs, cette molécule conserve quelques interactions hydrophobes avec le motif « WPF ».

Dérivés oxazoles

Parmi les dérivés oxazole, tous de type pan-BET se trouvent les molécules suivantes : La molécule **CPI-0610** est un dérivé benzo-isoxazole-azepine qui inhibe la protéine BRD4(1) avec une IC_{50} de 39 nM. Elle présente également un effet sur l'expression de l'oncogène c-Myc avec une EC_{50} de 180 nM. Cette molécule qui a démontré un effet antiprolifératif sur des lignées cancéreuse MV4-11 et dans des modèles de souris xénotransgéniques est actuellement en essais cliniques pour des cancers hématologiques. La structure du complexe BRD4(1)-CPI-0610 (199) (PDB : **5HLS**), montre que l'atome d'oxygène du cœur chimique « isoxazole » est en contact avec

l'asparagine N140 ainsi que l'atome d'azote en position R₃ qui crée également des liaisons également avec la tyrosine Y97 *via* une molécule d'eau situé dans la cavité K_{ac}. La molécule **IBET-151** est, quant à elle, un dérivé oxazole quinolinone qui présente une IC₅₀ comprise entre 250 et 790 nM, un K_D compris entre 20 et 100 nM. Cette molécule empêche la formation du complexe Super Elongation complex - Polymerase Association Factor Complex (SEC-PAF_c) en bloquant l'interaction avec les protéines BRD3 et BRD4 (147, 200), ce qui bloque la transcription des gènes codant BCL-2, c-Myc, CDK6. Par ailleurs, elle a été évaluée dans divers lignées cellulaires de leucémies (l'IC₅₀ est comprise entre 15 et 192 nM), dont des cellules portant la translocation MLL (**MLL** pour **Mixed Lineage Leukemia**). Dans les cellules portant cette translocation MLL, elle induit l'apoptose, ainsi que l'arrêt du cycle cellulaire en phase G0/G1. De plus, elle présente de meilleures propriétés pharmacocinétique que JQ1. Cette molécule est actuellement en essais clinique. La structure du complexe BRD4(1)/IBET151 (147) (PDB : **3ZYU**), montre que l'atome d'oxygène du cœur chimique « oxazole » est en contact avec l'asparagine N140, tandis que l'atome d'azote se trouve à proximité de la tyrosine Y97 *via* une molécule d'eau située dans la poche K_{ac}. Le groupe pyridine quant à lui se trouve à proximité du motif « WPF ».

Dérivés quinolines

Parmi les dérivés quinolines qui sont pour la plupart de type pan-BET inhibiteurs se trouvent les molécules suivantes : la molécule **IBET-726**, qui présente une IC₅₀ comprise entre 22 et 41 nM avec une efficacité d'inhibition supérieure d'un facteur mille par rapport à la protéine CREBBP. Ce composé a été évalué dans divers lignées cellulaires de neuroblastomes (Une inhibition de croissance médiane ou GI₅₀ 75 nM) et présente une activité cinq fois meilleure que la molécule IBET-151 (201). De surcroit, cette molécule inhibe l'expression des gènes codant N-Myc et BCL2 (201) et a montré un effet sur la croissance tumorale des myélomes dans des xénogreffes (196). La structure du complexe BRD4(1)/IBET726 (201) (PDB : **4BJX**), montre qu'il se fixe de manière similaire à JQ1, et que le groupe acide benzoïque se prolonge vers le motif « WPF ». La molécule **PFI-1**, présente une IC₅₀ de 180 nM pour la protéine BRD4(1), ainsi qu'une légère inhibition de la protéine CREBBP avec une IC₅₀ de 49 µM. Cette molécule inhibe la production d'interleukine 6 (IL6) dans les cellules PBMCs (**PBMC** pour **Primary Peripheral Blood Mononuclear Cells**) avec une EC₅₀ de 1,9 µM. Son temps de demi-vie dans l'organisme est d'une heure en injection intraveineuse et de 2h en sous-cutané (202). La structure du complexe BRD4(1)/PFI-1 (203) (PDB : **4E96**) montre une interaction avec l'asparagine conservée N140 et que le groupe sulfonamide s'oriente vers la région hydrophobe du motif « WPF ». En 2013, est publié le premier inhibiteur sélectif des domaines BD2 de la famille BET, et pan-BD2, à savoir la

molécule **RVX-208**. Elle inhibe préférentiellement le domaine BD2 des protéines de la famille BET, dont celui de la protéine BRD3 avec un K_D de 200 nM pour le BD2 versus 4,1 μ M pour le BD1 (ratio BD2/BD1 : 20 fois) (204). Cette molécule augmente le taux d'expression d'ApoA1 (205) et le bon cholestérol HDL(206), ce qui en fait une molécule prometteuse pour le traitement de maladies cardiovasculaires. Cette molécule a été évaluée dans plusieurs essais cliniques. La comparaison des structures des complexes BRD2(2)/RVX-208 et BRD4(1)/RVX-208 (204) (PDB : **4MR6** & **4MR4**) montre que le groupement amide du cœur chimique « quinazolinone » entre en contact avec l'asparagine N140, ainsi qu'avec les résidus tyrosine Y97 et glutamine Q85 médiée par des molécules d'eau. Le groupe éthylène glycol sort de la cavité et crée de légers contacts avec les résidus de surface. Contrairement aux autres inhibiteurs de bromodomains ce composé n'interagit pas avec le motif « WPF » ce qui diminue l'affinité pour les domaines BD1. Par ailleurs dans le cas du domaine BD2, le groupe phényle établit des interactions supplémentaires de type π - π avec le résidu histidine H433 qui est une acide aspartique D144 chez le domaine BD1 de la protéine BRD4, ce qui explique la sélectivité. En 2016, un analogue de celle-ci est publié, la molécule **RVX-297**. Cette molécule est également BD2 sélective (207) (PDB : **5DW2** & **5DW1**).

Dérivés pyrroles

Plus d'une dizaine de dérivés acétyle-pyrrole de type pan-BET ont été publiés et parmi eux se trouve la molécule **XD14** qui inhibe le domaine BRD4(1) avec une valeur de K_D de 237 nM. Cette molécule présente une très bonne affinité pour tous les domaines BETs excepté ceux de BRDT, mais elle a également une affinité élevée pour les BDs des protéines CREBBP et EP300. Cette molécule montre une activité sur les cellules de leucémies HL60 et SR, cependant elle n'a aucun antiprolifératif sur les cellules cancéreuses HeLa même à 50 μ M (208). La structure du complexe BRD4(1)/XD14 (208) (PDB : **4LYW**), montre que le cœur chimique « pyrrole » pénètre profondément dans la cavité K_{ac} et que l'azote N_1 crée une liaison hydrogène avec le carbonyle de résidu proline P82 du motif « WPF ». Le sulfonamide se positionne à proximité du « ZA channel » de manière à pouvoir établir des interactions CH- π -avec le résidu W81. Cette molécule forme également un contact avec le résidu leucine L92.

Nouveaux « cœurs chimiques »

Récemment, des nouveaux inhibiteurs présentant une préférence pour les domaines BD1, qui sont des molécules de type pan-BD1, ont vu le jour. C'est notamment le cas des molécules suivantes : la molécule **MS436** présente une sélectivité de dix fois pour le BD1 de la protéine BRD4 par rapport au BD2. Elle inhibe, dans les macrophages murins, l'activité de NF- κ B qui régule la production de cytokines pro-inflammatoires IL-6. La structure du complexe BRD4(1)/MS436 (209) (PDB : **4NUD**), montre un contact de type π - π (π stacking) entre le groupe pyridine et le tryptophane W81. Le contact de MS436 avec le résidu lysine K91, médiée par une molécule d'eau serait à l'origine de cette sélectivité observée, puisque le BD2 présente à la place de cette lysine une alanine A384 qui n'est pas orientée dans la bonne direction pour établir cette interaction. La très bonne affinité de cette molécule s'explique par le fait qu'elle établit d'autres contacts médiés par des molécules d'eau, avec les résidus suivants : P82, Q85, Y97 et N140. Un analogue, **MS611** présente quant à lui une sélectivité pour BD1 avec une affinité de quarante fois supérieure par rapport au BD2. MS611 stimule la progression des progéniteurs oligodendrocytaires primaires vers la différenciation, alors que les molécules pan-BET au contraire bloquent cette différenciation (210). La molécule **Olinone** a les mêmes effets pharmacologiques chez la souris. Elle inhibe la protéine BRD4, avec une préférence pour le domaine BD1 (K_D de 3,4 μ M) contre une affinité quatre-vingt fois inférieure pour BD2 (300 μ M). La même préférence a été observée pour les domaines BD1 des protéines BRD3 et BRD2 (210). La structure du complexe BRD4(1)-Olinone (210)(PDB : **4QB3**), montre que la chaîne alkyle se fixe dans la poche K_{ac} et entre en interaction avec l'asparagine N140 et que le triple cycle de l'Olinone est en contact avec l'acide aspartique D144. Dans le domaine BRD4(2), ce résidu est remplacé par un résidu beaucoup plus large, une histidine H437 qui induit une gêne stérique avec le triple cycle de l'Olinone, ce qui explique la perte d'affinité pour les BD2. Cette molécule présente également une affinité moyenne à basse pour les BDs des protéines CREBBP (K_D de 33 μ M) et PHIP (K_D de 103 μ M), mais n'est pas un inhibiteur des autres familles de bromodomains. Enfin l'une des molécules que j'ai identifiée et caractérisée au cours de ma thèse (211) (PDB : **5E14**) présente également une préférence pour les domaines BD1 de la famille BET.

b. Inhibiteurs de BETs en essais préclinique & cliniques

i. Généralités

Présentation d'un essai préclinique

Les essais précliniques sont réalisés dans un premier temps *in vitro* sur des modèles cellulaires pertinents. La cytotoxicité de la molécule sélectionnée est évaluée sur une ou plusieurs lignées cellulaires. Il a été montré que la protéine BRD4 était impliquée dans le développement de cancers, c'est pourquoi des inhibiteurs de cette protéine sont évalués sur des lignées cancéreuses. Par ailleurs, lorsque la protéine cible est impliquée dans l'expression d'un quelconque facteur, la molécule sera également testée pour voir l'effet sur l'expression de celui-ci. La protéine BRD4 est par exemple impliquée dans l'expression de l'oncogène c-Myc, l'expression de celui-ci sera donc déterminée en présence et en absence d'inhibiteur de cette protéine.

Présentation d'un essai clinique

Les essais cliniques chez l'homme ont pour but de confirmer (ou d'infirmer) les données obtenues *in vitro* et chez l'animal. Les paramètres pharmacocinétiques, pharmacodynamiques et thérapeutiques du futur médicament sont alors vérifiés chez l'homme. Les essais cliniques comprennent trois étapes successives (phases I, II, III) détaillées ci-dessous. Chaque « candidat médicament » doit passer et réussir ces trois étapes avant de pouvoir obtenir une autorisation de mise sur le marché (**AMM**) et une commercialisation.

✓ La Phase I : première administration chez l'homme sain

Cette étude a pour objectifs de déterminer la toxicité à court terme d'une substance active et ainsi la dose tolérée, mais aussi d'évaluer sa pharmacocinétique chez l'homme et d'identifier les métabolites potentiels de la molécule. La molécule est testée sur un nombre limité d'individus volontaires sains (20 à 80 personnes), excepté en oncologie où elle est testée chez des malades. L'administration du produit se fait en dose unique et avec des doses progressivement croissantes jusqu'à l'apparition des premiers signes d'intolérances cliniques et biologiques. En général, la première dose correspond à 1/20^e ou 1/10^e de la dose maximale tolérée chez l'animal (en mg/kg). Cette phase dure entre six et douze mois et environ 70% des composés testés peuvent poursuivre en phase II.

✓ **Phase II : administration chez l'homme malade**

Cette seconde étape a pour but d'établir la relation dose/effet de la molécule, de déterminer les paramètres pharmacocinétique des patients malades, de définir les conditions optimales d'utilisations telles que la dose, le rythme et la durée d'administration et la surveillance nécessaire. La molécule est administrée à un plus grand nombre d'individus et uniquement à des patients malades (100 à 300 personnes). Plusieurs doses (3 à 10 doses) sont alors administrées jusqu'à l'observation de l'effet optimal. Cette phase dure entre douze et vingt-quatre mois et un tiers des substances administrées passent en phase III.

✓ **Phase III : étude comparative du médicament versus un placebo**

Cette troisième étape dite « essais thérapeutique » a pour objectif de démontrer l'efficacité et la tolérance du produit dans les conditions d'utilisation élargies tel que l'administration prolongée (durée minimale un à deux ans). Cette étude est réalisée sur un très grand effectif, représentatif de patients souffrant une pathologie que le médicament doit traiter (1000 à 3000 personnes). Les patients sont séparés en deux groupes homogènes par tirage au sort. Le premier groupe recevra le produit à tester et l'autre recevra un placebo ou un médicament existant déjà pour la pathologie à soigner. Cette troisième phase d'essai clinique qui dure entre dix-huit et vingt-quatre mois a un coût élevé et un enjeu industriel très important puisque les résultats de cette étude seront déterminants pour l'obtention de l'AMM et la commercialisation ou non du produit. Entre 70 à 90 % des médicaments testés seront retenus pour une demande d'AMM.

✓ **Phase IV : pharmacovigilance post-commercialisation**

Cette quatrième étape dite de « Surveillance post-AMM », a pour but d'évaluer les bénéfices thérapeutiques dans des conditions réelles d'utilisation, c'est-à-dire sur des patients qui n'ont pas été sélectionnés pour les études précédentes. Des données sur la qualité de vie, la morbidité, la mortalité, les interactions médicamenteuses potentielles et les effets indésirables rares non détectés lors des phases précédentes seront recensées pour plusieurs dizaine de millier de patients. Par ailleurs, l'effet du nouveau produit pourra être comparé à celui d'un autre médicament donné pour la même pathologie (ex : les génériques). La pharmacovigilance se fait sur une période de dix à vingt ans et la réévaluation de l'AMM se fait tous les 5 ans.

ii. Inhibiteurs de BETs en essais précliniques

Actuellement, une dizaine d'inhibiteurs (**Tableau 8**) qui inhibent l'interaction PPI entre les bromodomains de la famille BET et les lysines acétylées des histones, sont en essais préclinique et sont évalués pour une utilisation dans le traitement des cancers, des maladies inflammatoires et des désordres du métabolisme des lipides (186).

Tableau 8 : Inhibiteurs de BETs en essais précliniques.

	Compagnie	Molécule	Trial ID	Pathologie(s) indiquée(s)
Essais précliniques	Aurigene Discovery Technologies	-	-	Cancer
	Bayer Healthcare	BET-BAY-002	-	Cancer
	BeiGene Ltd	-	-	Hématologie, Neoplasme, Tumeurs solides
	Calleron Therapeutics	-	-	Cancer
	CRELUX/4SC	-	-	-
	Dybyl	-	-	Maladies inflammatoires
	Epigenetix	-	-	Cancer
	GlaxoSmithKline	GW-841819X	-	Désordre du métabolisme des lipides
	Nuevolution	-	-	Cancer
	OncoFusion Therapeutics; Medivation	-	-	Cancer
	SignalRx Pharmaceuticals	SF2523	-	Cancer
	SignalRx Pharmaceuticals	SF2535	-	Cancer
	University of Oxford; Pfizer	-	-	Cancer, Désordre du métabolisme, Syndrome de Rubini-stein-Taybi

iii. Inhibiteurs de BETs en essais cliniques

À ce jour, le site <https://clinicaltrials.gov/>, répertorie plus d'une vingtaine d'essais cliniques en cours pour une quinzaine de molécules différentes (15 juin 2017, **Tableau 9**). La majorité de ces inhibiteurs sont évalués en phase I, un quart d'entre eux sont passés en phase II et un seul est arrivé en phase III, il s'agit du composé RVX208. Toutes ces molécules sont indiquées dans le traitement de divers cancers (<https://www.cancer.gov/>), exceptée RVX208 qui est évaluée chez des patients atteints de maladies cardiovasculaires ou de diabète de type 2.

Le composé **RVX-208** (RVX000222 ou apabetalone) développé par Resverlogix Corp, est la molécule la plus avancée en essai clinique à ce jour. Au total, sept essais cliniques ont été initiés avec cette drogue dans le traitement de l'athérosclérose, de maladies cardiovasculaires et du diabète. En phase II, cette molécule a montré des effets bénéfiques sur les patients, à savoir l'augmentation du bon cholestérol (HDL), ainsi que du niveau d'expression de l'Apo lipoprotéine A1 et une diminution des événements cardiaques indésirables ou **MACE** (**M**ajor **A**dverse **C**ardiac **E**vents). Ce qui a permis d'initier une phase III depuis l'automne 2015. Les phases I et II sont à présent toutes terminées, seule la phase III est toujours en cours.

Trois molécules sont en phase II, à savoir **BMS-9861158**, **INCB057643**, **INCB54329**. La molécule BMS-9861158 (structures non publiée) développée par Bristol Meyers Squibb est indiquée pour de multiples cancers et des tumeurs solides. Les molécules INCB057643, INCB54329, développées par Incyte Pharmaceutical sont quant à elles, indiquées pour des tumeurs solides, lymphomes ou cancer hématologiques.

Quatre essais cliniques ont été initiés pour la molécule **OTX015** développée par OncoEthix et Merck, dans le traitement de tumeurs solides, de leucémie aiguë, d'hémopathies malignes (en phases I) et de glioblastomes multiformes (en phase II). Ces essais cliniques sont à présent tous terminés. Actuellement, un dernier essai clinique en phase I est en cours pour le traitement de leucémies myéloïdes aiguës, la phase II (dans lequel OTX015 était administré en combinaison avec l'azacitidine) initié pour ce type de patients ayant été stoppée prématurément.

Onze molécules sont actuellement en phase I, pour des leucémies AML et ALL, des myélomes, les cancers du sein, de la prostate, du poumon, des tumeurs solides... Un ou plusieurs essais cliniques sont en cours d'évaluations pour ces molécules. La phase I initiée pour la molécule **BAY1238097** (structure non publiée) est maintenant terminée.

Bien que l'inhibiteur (+)-JQ1, sélectif des BDs de la famille BET, « pan-BET », ait été testé dans des essais précliniques sur de nombreux type de cancers, incluant le carcinome NMC, cette molécule n'a pu être évaluée sur l'homme du fait de son temps de demi-vie trop court. [Mitsubishi Tanabe Pharma [WO/2009/084693]. **Aucun inhibiteur sélectif pour la sous-famille BD1 ou un seul BD de la famille BET n'a fait l'objet d'un essai clinique à ce jour.**

Tableau 9 : Inhibiteurs de BETs en essais cliniques.

	Compagnie	Molécule	Trial ID	Pathologie(s) indiquée(s)
Essais cliniques - Phases de développement	AbbVie	ABBV-075	NCT02391480	Cancers: Sein, Poumon (à petites cellules SCLC ou non NSCLC), Leucémie Myéloïde Aiguë (AML), Myélome multiple, Prostate, Lymphome non Hodgkinien
	Constellation Pharmaceuticals	CPI-0610	NCT02158858, NCT01949883, NCT02157636	Leucémie Myélocytaire Aiguë, Syndrome Myélodésplasique (MDS), Néoplasme Myélodésplasique et Myéloprolifératif, Myélofibrose / Lymphome / Myélome multiple
	Daiichi Sankyo	PLX51107	NCT02683395	Tumeurs solides, AML, MDS
	FORMA Therapeutics; Celgene	FT-1101	NCT02543879	AML, MDS
	Gilead Sciences	GS-5829	NCT02392611	Tumeurs solides, Lymphome diffus à cellules B et non Hodgkinien
	GlaxoSmithKline	GSK1210151A (I-BET151)	NCT02630251	Néoplasme hématologique Cancers, tumeurs solides
	GlaxoSmithKline	GSK525762 (I-BET762)	NCT01587703	NMC
	SignalRx Pharmaceuticals	SF1126	NCT03059147	Cancer hépatocellulaire
	Hoffmann-La Roche	TEN-010 / RO6870810	NCT03068351, NCT01987362, NCT02308761	Myélomes multiples / Tumeurs solides / AML, syndromes myéloblastique
	Zenith Epigenetics	ZEN-003694	NCT02711956, NCT02705469	cancer métastatique de la prostate résistant à la castration
	Merck; OncoEthix GmbH	OTX-008	NCT01724320	Tumeurs solides
	Merck; Mitsubishi Tanabe	OTX-015 / MK-8628	NCT02698189, NCT02698176, NCT01713582, NCT02259114,	NMC, Cancer du sein triple négatif (TNBC), Cancer de la Prostate résistant à la Castration (CRPC), Adénocarcinome canalaire pancréatique, cancer du poumon (NSCLC)AML incluant AML de novo et AML secondaire à MDS et DLBCL / NMC, TNBC, NSCLC, CRPC / AML, Leucémie Lymphoblastique Aiguë (ALL), Lymphome diffus à cellules B, myélomes multiples /
	Zenith Epigenetics	ZEN-3365	NCT02238522	AML, Tumeurs lymphoproliférative
	Bayer Healthcare	BAY 1238097	NCT02369029	Néoplasmes (tumeurs solides, lymphomes, NUT carcinome (NMC), myélome)
	Bristol-Myers Squibb	BMS-986158	NCT02419417	Tumeurs solides, Cancer multiple indications
Incyte Pharmaceuticals	INCB057643	NCT02711137	Tumeurs solides, lymphomes et cancers hématologique	
Incyte Pharmaceuticals	INCB54329	NCT02431260	Cancers hématologiques	
Merck; OncoEthix GmbH	OTX-015 / MK-8628	NCT02296476	Glioblastomes multiformes	
Merck; OncoEthix GmbH	tOTX015	NCT02303782	AML	
Resverlogix	RVX208 / RVX000222	NCT01058018, NCT01067820, NCT01863225, NCT01423188, NCT01728467, NCT00768274	Atherosclérose, maladies cardiovasculaires / Maladies de l'artère coronaire / Dyslipidémie, Maladies de l'artère coronaire / Dyslipidémie, Maladies de l'artère coronaire / Diabète / Dyslipidémie, Atherosclérose, Syndrome Coronarien Aiguë, maladies cardiovasculaires	
University of Texas Southwestern Medical Center	CPI-0610	NCT02986919	Tumeurs du système nerveux périphérique	
Resverlogix	RVX208 / RVX000222	NCT02586155	Diabète sucré de type 2, Maladies de l'artère coronaire, maladies cardiovasculaires	

Objectifs

L'inhibition de l'interaction PPI BD-peptide acétylé fait aujourd'hui l'objet d'un nombre important de recherche (1453 articles dans PubMed. Recherche « bromodomain » le 31/08/2017 <https://www.ncbi.nlm.nih.gov/pubmed/>) aussi bien au niveau académique par exemple au sien du Consortium de Génomique Structurale, qu'au niveau industriel comme c'est le cas dans la société pharmaceutique Glaxo Smith Kline (**GSK**) (174, 179, 180). De plus, la famille BET et plus particulièrement le domaine BRD4(1), qui font l'objet de ma thèse, sont les BDs les plus étudiés parmi l'ensemble des motifs BDs. De ce fait, nous disposons de nombreuses données de comparaisons pour le développement de nouveaux inhibiteurs.

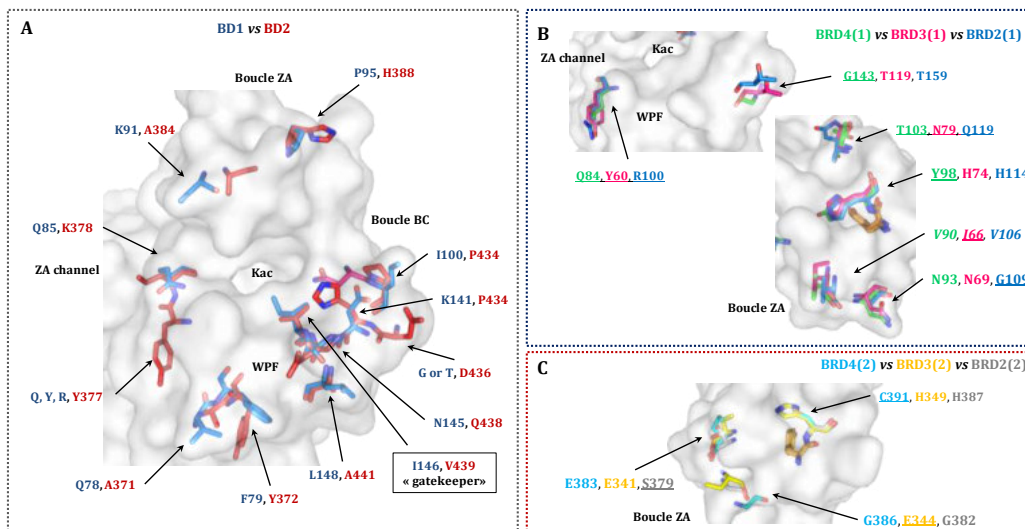
Depuis la découverte en 2010 de la première molécule sonde, JQ1, un inhibiteur de type pan-BET (57), plus de 160 structures du domaine BRD4(1) en complexe avec des inhibiteurs ont été résolues et sont disponibles dans la PDB. Par ailleurs, la majorité de ces molécules sont de type pan-BET inhibitrices et une quinzaine d'entre elles sont actuellement en essais cliniques (112, 212) (<https://clinicaltrials.gov/>). Cependant, il s'avère que l'inhibition de type pan-BET peut potentiellement conduire à des effets secondaires, du fait que ces protéines soient impliquées dans divers rôles au niveau cellulaire (213), ainsi qu'à l'apparition de résistances (214-217). Cette résistance pourrait être expliquée par le fait que cette molécule n'est pas sélective. Le développement d'inhibiteurs spécifiques des BD1 devrait pallier ce problème.

Dans ce contexte, les objectifs de ma thèse étaient d'identifier et de caractériser des inhibiteurs sélectifs envers la sous-famille BD1 ou BD2 et plus idéalement envers un seul BD de la famille BET, telle que BRD4(1). Le but étant de développer i) des « sondes épigénétiques » afin de décrypter le rôle individuel de chaque protéine et BDs dans les divers processus biologiques connus (régulation de la transcription, métabolisme, signalisation cellulaire ...) et ii) des « candidats médicaments » dans l'optique d'une part de développer de nouvelles thérapies (maladies inflammatoires, cardiovasculaires, du système nerveux, du métabolisme, les cancers...), d'autre part de pallier les problèmes de résistances liés à l'inhibition pan-BET récemment mises en évidence dans les cancers ovariens et du sein (214-217) et enfin d'avoir des traitements plus spécifiques.

Comment cibler sélectivement les BETs : BD1 versus BD2 versus un seul BD

Nous avons vu précédemment que : i) les domaines BD1 et BD2 présentent des différences de substrats (103), ii) les BD1 et les BD2 partagent plus de 70% d'identité de séquence au sein de leur sous-famille, tandis qu'à l'intérieur de la famille BET, ils partagent environ 50% (43-54%) d'identité de séquence. Ceci suggère que les membres des deux sous-familles ont évolué différemment et qu'on pourrait ainsi les cibler sélectivement. Bien que la poche de fixation K_{ac} des BDs de la famille BET soit identique à 95% pour tous les BDs, la comparaison des structures des BD1 et des BD2 montre que trois résidus clés pour la fixation de ligand différent et pourraient être ciblé pour développer des inhibiteurs sélectifs BD1 versus BD2. Ces résidus sont les suivants : i) le résidu glutamine des BD1 (Q85 chez BRD4(1)) est remplacé par un résidu lysine pour les BD2 (K378 chez BRD4(2)) ; ii) le résidu acide aspartique des BD1 (D144 chez BRD4(1)) est remplacé par un résidu histidine pour les BD2 (H437 chez BRD4(2)) ; et iii) le résidu « gatekeeper » est une isoleucine pour les BD1 (I146 chez BRD4(1)) et une valine pour les BD2 (V439 chez BRD4(2)) (183). Par ailleurs, pour pouvoir cibler sélectivement chaque BDs, il faudrait explorer les autres résidus de la cavité K_{ac} qui diffèrent entre les huit BDs (183), ceux-ci sont indiqués dans la figure et le tableau ci-après (**Figure 21** et **Tableau 10**).

Figure 21 : Exploration des résidus autour de la poche K_{ac} des BETs.



Légende : **A- Analyse de la sélectivité au sein des BD1 versus BD2 :** Résidus pouvant être explorés en vue de développer des inhibiteurs sélectifs des domaines BD1 (en bleu) ou BD2 (en rouge). **B- Analyse de la sélectivité au sein des BD1 :** Les BD1 présentent plus de divergence entre eux que les BD2. Les résidus situés au niveau des boucles ZA et BC pourraient être explorés pour développer des inhibiteurs sélectifs de BRD4(1) en vert, BRD3(1) en rose et BRD2(1) en bleu. **C- Analyse de la sélectivité au sein des BD2 :** Seuls quelques résidus situés au niveau de la boucle ZA pourraient être explorés pour développer des inhibiteurs sélectifs de BRD4(2) en bleu, BRD3(2) en jaune et BRD2(2) en gris. Les structures de BRDT(1) et BRDT(2) n'ont pas été incluses car celle pour BRDT(1) est très divergente des BD1 et celle de BRDT(2) n'est pas disponible dans la PDB.

Tableau 10 : Récapitulatif des résidus autour de la poche Kac des BETs.

Motif	BD1				BD2			
	BRD2(1)	BRD3(1)	BRD4(1)	BRDT(1)	BRD2(2)	BRD3(2)	BRD4(2)	BRDT(2)
hélice αZ					K366	K326	K368	K287
	K92	K52	K76	K45	H365	H327	H369	H288
	H93	H53	H77	H46	A366	A328	A370	F289
	Q94	Q54	Q78	S47	A367	A329	A371	S290
	F95	F55	F79	F48	Y368	Y330	Y372	Y291
	A96	A56	A80	S49	A369	A331	A373	A292
WPF shelf	W97	W57	W81	W50	W370	W332	W374	W293
WPF	P98	P58	P82	P51	P371	P333	P375	P294
WPF	F99	F59	F83	F52	F372	F334	F376	F295
ZA channel	R100	Y60	Q84	Q53	Y373	Y335	Y377	Y296
ZA channel	Q101	Q61	Q85	R54	K374	K336	K378	N297
ZA channel	P102	P92	P86	P55	P375	P337	P379	P298
ZA channel	V103	V63	V87	V56	V376	V338	V380	V299
boucle ZA	D104	D64	D88	D57	D377	D339	D381	D300
boucle ZA	A105	A65	A89	A58	A378	A340	V382	V301
boucle ZA	V106	I66	V90	V59	S379	E341	E383	N302
boucle ZA	K107	K67	K91	K60	A380	A342	A384	A303
boucle ZA	L108	L68	L92	L61	L381	L343	L385	L304
boucle ZA	G109	N69	N93	Q62	G382	E344	G386	G305
boucle ZA	L110	L70	L94	L63	L383	L343	L387	L306
	P111	P71	P95	P64	H384	H344	H388	H307
	D112	D72	D96	D65	D385	D345	D389	N308
Tyr conservée	Y113	Y73	Y97	Y66	Y386	Y346	Y390	Y309
	H114	H74	Y98	Y67	H387	H347	C391	Y310
	K115	K75	K99	T68	D388	D348	D392	D311
	I116	I76	I100	I69	I389	I349	I393	V312
	I117	I77	I101	I70	I390	I350	I394	V313
	K118	K78	K102	K71	K391	K351	K395	K314
	Q119	N79	T103	N72	H392	H352	H396	N315
	P120	P80	P104	P73	P393	P353	P397	P316
	M121	M81	M105	M74	M394	M354	M398	M317
	D122	D82	D106	D75	D395	D355	D399	D318
hélice αA								
hélice αB								
	Y155	Y115	Y139	Y108	Y428	Y390	Y432	Y351
Asn conservée	N156	N116	N140	N109	N429	N391	N433	N352
	K157	K117	K141	K110	P430	P392	P434	P353
	P158	P118	P142	P111	P431	P393	P435	P354
boucle BC	T159	T119	G143	G112	D332	D394	D436	D355
boucle BC	D160	D120	D144	D113	H433	H395	H437	H356
boucle BC	D161	D121	D145	D114	D434	E396	E438	E357
Gatekeeper	I162	I122	I146	I115	V435	V397	V439	V358
	V163	V123	V147	V116	V436	V398	V440	V359
	L164	L124	L148	L117	A437	A399	A441	H360
hélice αC	M165	M125	M149	M118	M438	M400	M442	M361

Légende : Les résidus indiqués en noir sont communs pour les BETs, en bleus sont spécifiques des BD1, en violet spécifiques des BD2 et en rouge différent au sein de la famille BETs. Les résidus indiqués en bleu et violet pourraient être utilisés pour développer des inhibiteurs sélectifs des sous-familles BD1 et BD2. Ceux indiqués en rouge pourraient être utilisés pour développer des inhibiteurs sélectifs d'un seul BD parmi les BETs.

Chapitre 2 : Résultats

Dans ce chapitre j'introduirai dans un premier temps les résultats préliminaires qui sont à l'origine de mon projet de recherches. D'une part, l'identification et la validation orthogonale de sept molécules inhibitrices ciblant les BDs de la famille BET, identification obtenue au cours du criblage de notre chimiothèque 2P2I_{3D} focalisée sur les interactions PPIs : ce travail a été réalisé par Sabine Milhas avant mon arrivée dans l'équipe et a fait l'objet de son manuscrit de thèse. D'autre part, les caractérisations structurales que j'ai effectuées pour certains complexes BD-ligand obtenus, et qui ont servis de point de départ pour l'optimisation des molécules les plus intéressantes en termes d'affinité ou de sélectivité. Ces travaux ont fait l'objet d'une publication en 2016 dans le journal ACS Chemical Biology (218).

Dans un deuxième temps, je me focaliserai sur l'une des molécules identifiées au cours de ce criblage, baptisée **5A14**, présentant un profil de sélectivité original, à savoir « BRD4(1) sélective ». Je présenterai en premier le travail de **SAR (Structure Activity Relationship)** qui nous a permis de déterminer l'origine de ce profil de sélectivité observé. Ce travail a fait l'objet d'une publication en 2016 dans le journal Medicinal Chemistry (211) et d'un dépôt de brevet (**EP 15 307 163.4**). Je développerai ensuite le travail d'optimisation de cette molécule tout en tentant de conserver son profil de sélectivité. Ces travaux feront l'objet d'une publication et d'un dépôt de brevet en 2018.

Partie 1 : Criblage de la chimiothèque
2P2J3D : Identification et
validation d'inhibiteurs des BETs

I. Développement de la chimiothèque 2P2I_{3D}

Les chémo-informaticiens de l'équipe ont développé une chimiothèque, nommée « 2P2I_{3D} », composée de 1664 molécules (219). Cette chimiothèque, focalisée sur les interactions protéine-protéine (PPIs), a été développée à partir des informations extraites de la base de données 2P2I_{DB} (220-222) qui répertorie l'ensemble des inhibiteurs orthostériques de PPIs dont les structures tridimensionnelles de complexes protéine-protéine et/ou protéine-ligand ont été validées. L'analyse des propriétés chimiques et des structures tridimensionnelles des inhibiteurs de PPIs disponibles dans la PDB a permis de définir des descripteurs spécifiques pour les interactions de type PPI, à savoir le poids moléculaire > 400 Da, le nombre de cycle > 4, le nombre de donneur et accepteur d'hydrogène > 4 et le LogP > 4. Un algorithme tenant compte de l'ensemble de ces descripteurs combiné à l'utilisation de critères de diversité moléculaire a ensuite été utilisé pour filtrer *in silico* 8,3 millions de molécules commerciales. Ce processus a permis de sélectionner 1664 petites molécules présentant des propriétés compatibles avec une inhibition de type « PPI » qui ont été commandées et mises en plaques pour des expériences *in vitro*.

II. Identification et validation d'inhibiteurs de PPIs

Dans un premier temps, cette chimiothèque 2P2I_{3D} a été validée expérimentalement sur une série de complexes protéine-protéine ou protéine-peptide présentant des surfaces d'interactions aux caractéristiques diverses, afin de confirmer les taux d'enrichissement prédits par ces approches prédictives. Les complexes évalués sont les suivants : Nef/HCK du virus du HIV, NS3/NS5 du virus de la Dengue, domaine PDZ/peptide des protéines GRASP55 et Syntenine avec leur peptide respectif, domaine SH3/peptide de la protéine Fyn et enfin l'interaction BRD4/peptide lysines acétylées.

Mon projet se focalisant sur l'identification d'inhibiteurs sélectifs de bromodomains, la suite de cette partie s'intéressera uniquement aux résultats obtenus pour le complexe BRD4(1)-peptide lysines acétylées.

1. Identification de « touches » ciblant les BETs

Les 1664 molécules de la chimiothèque 2P2I_{3D} ont été évaluées sur le complexe BRD4(1)-peptide lysines acétylées (H3K5_{ac}K8_{ac}K12_{ac}K16_{ac}) *via* la technique d'HTRF. Ce criblage réalisé à l'aide d'une concentration en inhibiteur de 20 µM, a permis d'identifier quarante-quatre « touches » ciblant ce complexe avec un pourcentage d'inhibition supérieur à 40%. Parmi celles-ci, dix-sept molécules, soit 1% de la chimiothèque, ont par la suite été validées par HTRF au cours d'un contre criblage (en présence uniquement de GST-biotine et en absence du BRD4(1)), puis d'un criblage secondaire avec la mesure des valeurs IC₅₀. Quinze de ces molécules se sont avérées spécifiques du complexe BD-peptide, tandis que les deux autres présentaient également un effet sur le complexe NS3/NS5 (**Tableau 11**). Les valeurs d'IC₅₀ pour ces molécules sont comprises entre 200 nM et 38 µM et ont été confirmées sur un nouveau lot de composés.

Tableau 11 : Criblage de la chimiothèque 2P2I_{3D}.

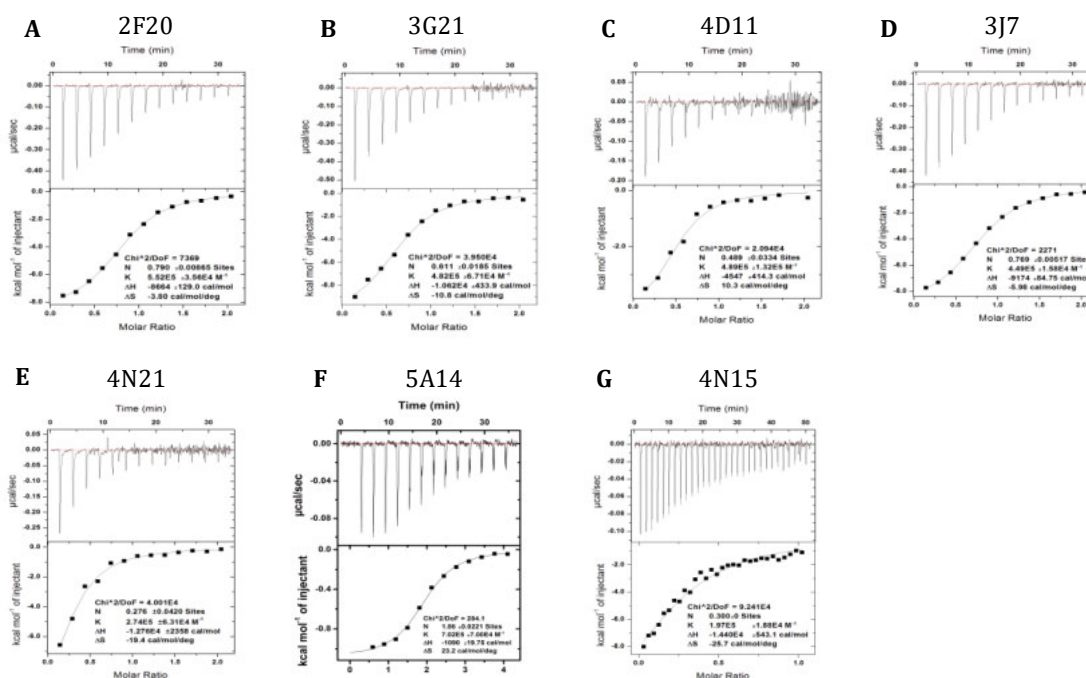
Partenaires	BRD4(1) / H3K5 _{ac} K8 _{ac} K12 _{ac} K16 _{ac}
Bromodomaine / Peptide histone lysines acétylées	
Type d'interface	Protéine / Peptide
K _D (µM)*	2,8
HTRF IC ₅₀	3,0
Z'	0,78
no. Hits non-auto fluorescent (inhibition >40%)	44
no. Molécules Validées (criblage secondaire)	17 (1%)
no. IC ₅₀ mesurées par HTRF	17
no. Molécules uniques	15
no. Sensible au TSA	10
no. Interactions validées par ITC	7

Légende : Le premier criblage a été réalisé à la concentration de 20 µM et a permis d'identifier quarante-quatre touches présentant une inhibition supérieure à 40%. Le criblage secondaire et le contre-criblage ont permis de valider dix-sept de ces touches dont quinze spécifiques du complexe BD-peptide. Les expériences TSA (ratio BD-ligand : 2 µM/2 mM) ont permis de valider dix d'entre elles et celles d'ITC (ratio BD-ligand : dans la Figure 22) ont permis d'en valider sept. La valeur de K_D (*) est issue de l'article de **Filippakopoulos et al** (103).

2. Validation biophysique de ces « touches »

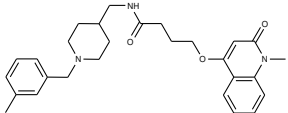
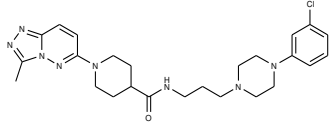
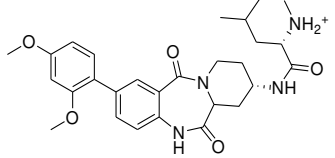
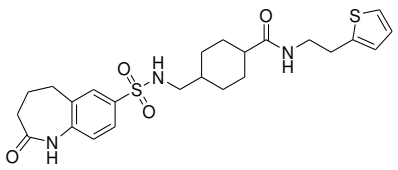
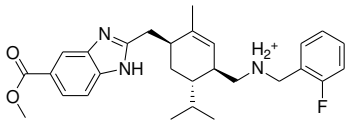
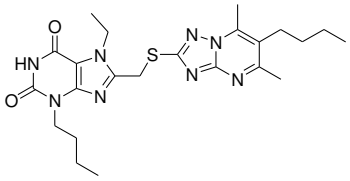
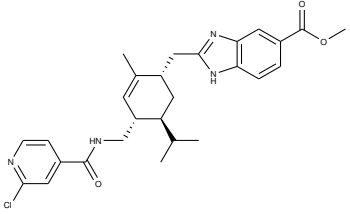
Ces quinze molécules ont ensuite été évaluées par deux types de techniques orthogonales, le TSA et l'ITC. Les expériences de TSA (ratio BD-ligand : 2 μM /2 mM) ont permis de valider dix de ces touches présentant des ΔTm compris entre $-1,5^\circ\text{C}$ et $+1,9^\circ\text{C}$. Les expériences d'ITC (ratio BD-ligand : dans la légende de la **Figure 22**) ont permis de confirmer la liaison de sept d'entre elles sur le domaine BRD4(1) avec des affinités (K_D) comprises entre 1,8 μM et 4,7 μM (**Figure 22**). Le **Tableau 12**, récapitule les structures chimiques, ainsi que les valeurs d' IC_{50} , de ΔTm et de K_D déterminées pour les sept molécules validées.

Figure 22 : Validation orthogonale par ITC des « touches » identifiées.



Légende : Les expériences d'ITC ont été réalisées en titration inverse avec le domaine BRD4(1) dans la seringue et l'inhibiteur dans la cellule de mesure à 15°C excepté pour la molécule 5A14 qui a été mesurée à 35°C . Le rapport en μM , BD-ligand était en μM de 20 : 200 pour les molécules 2F20, 3G21, 3J7, 4D11, 4N21, de 20 : 100 pour 4N15 et de 15 : 300 pour 5A14.

Tableau 12 : Récapitulatifs des résultats pour les sept molécules validées.

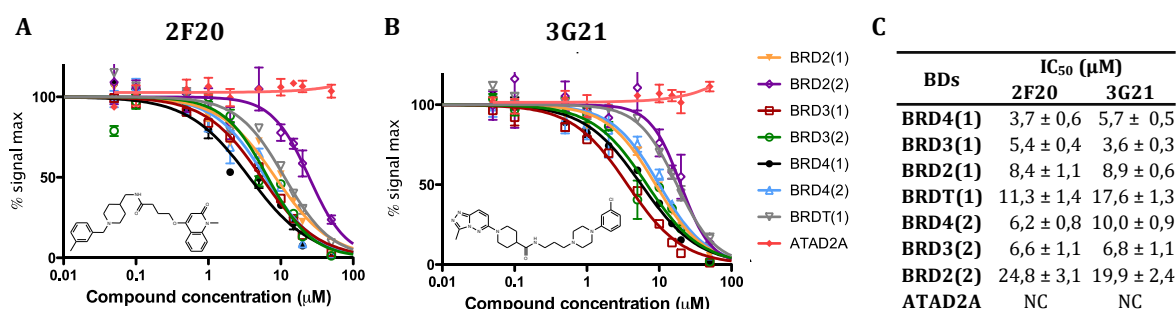
Nom CRCM	Composés		HTRF IC ₅₀ (μM)	BRD4(BD1) TSA ΔTm (°C)	ITC K _D (μM)
	Structures chimiques				
2F20			8,4	1,9	1,8
3G21			7,8	1,5	2,0
4D11			12,9	1,8	2,0
3J7			8,9	1,5	2,2
4N21			22,8	-1,4*	3,6
5A14			5,0	-1,5	1,4
4N15			0,2	1,1	4,7

Légende : Les expériences d'HTRF ont été réalisées à une concentration finale de 1% en DMSO. Celles de TSA ont été réalisées en présence du fluorophore dye 1X se fixant sur les régions hydrophobes de la protéine, une concentration finale en DMSO de 2,5%, et un rapport en BD : ligand de 2 : 50 (μM) L'Astérix (*) indique que la molécule 4N21 seule dans le tampon avec le fluorophore présentait un effet sur la fluorescence. Les expériences d'ITC ont quant à elles été réalisées comme décrites dans la légende de la **Figure 22**.

3. Détermination du profil de sélectivité des inhibiteurs validés

Les profils de sélectivité de ces sept molécules ont été évalués au sein de la famille BET et ATAD2A (famille IV) comme contrôle. Parmi celle-ci, cinq étaient de type pan-BET (**2F20**, **3G21**, **4D11**, **4N21**, **4N15** ; numérotées dans l'article #1, #2, #3, #5 et #7), inhibant l'ensemble des membres de la famille BET et deux présentaient des profils de sélectivité assez originaux, à savoir sélectives du domaine BD1 de la protéine BRD4 ou BRD4(1) pour **5A14** (#6) et des domaines BD1 de la famille BET ou BET-BD1 pour **3J7** (#4). Étant donné que nous souhaitons initier des programmes d'optimisation par SAR pour la meilleure molécule pan-BET (**4N15**) présentant des **critères proches d'une sonde épigénétique** (page 85), et pour les deux molécules sélectives (**5A14** et **3J7**), nous n'avons publié dans l'article *Milhas et al* (218), que les données (sélectivité, étude cellulaire, structures cristallographiques) pour les composés **2F20** et **3G21**. Les valeurs d'IC₅₀ pour les BETs pour **2F20** sont comprises entre 3,7 et 25 µM, celles pour **3G21** sont comprises entre 3,6 et 20 µM, par ailleurs leur affinité pour le domaine ATAD2A est supérieure à 50 µM (**Figure 23**).

Figure 23 : Détermination des profils de sélectivité de 2F20 et 3G21.



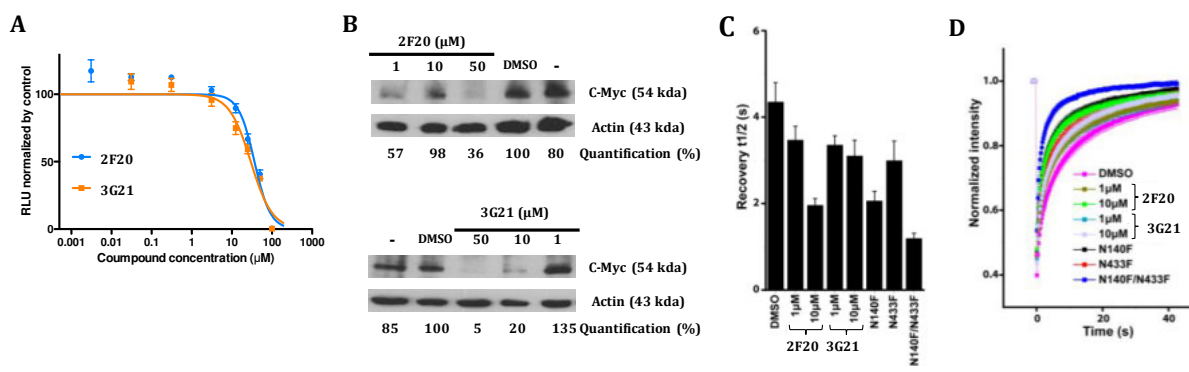
Légende : Les profils de sélectivité ont été déterminés par HTRF. **A-** Profil de sélectivité pour la molécule 2F20. **B-** Profil de sélectivité pour la molécule 3G21. Les courbes sont colorées en fonction du BD comme indiqué dans la légende à droite du panneau B. **C-** Tableau récapitulatif des valeurs d'IC₅₀ mesurées pour les molécules 2F20 et 3G21.

4. Validation cellulaire

L'effet de ces molécules sur la viabilité cellulaire des lignées leucémiques Jurkat a été évalué et les valeurs d'EC₅₀ mesurées sont de 31 et 21 µM respectivement pour les molécules **2F20** et **3G21** (**Figure 24**). De plus, ces molécules induisent une répression de l'expression (down-régulation) de l'oncogène c-Myc. Par ailleurs, des expériences de FRAP réalisées dans des cellules épithéliales d'ostéosarcome osseux humain (U2OS), ont quant à elles permis de montrer que ces deux molécules étaient capables d'accélérer de manière significative le temps de

semi-récupération suite au photoblanchiment des cellules à 10 μM , suggérant qu'elles sont capables de déplacer le domaine BRD4(1) de la chromatine.

Figure 24 : Études cellulaires pour les composés 2F20 et 3G21.



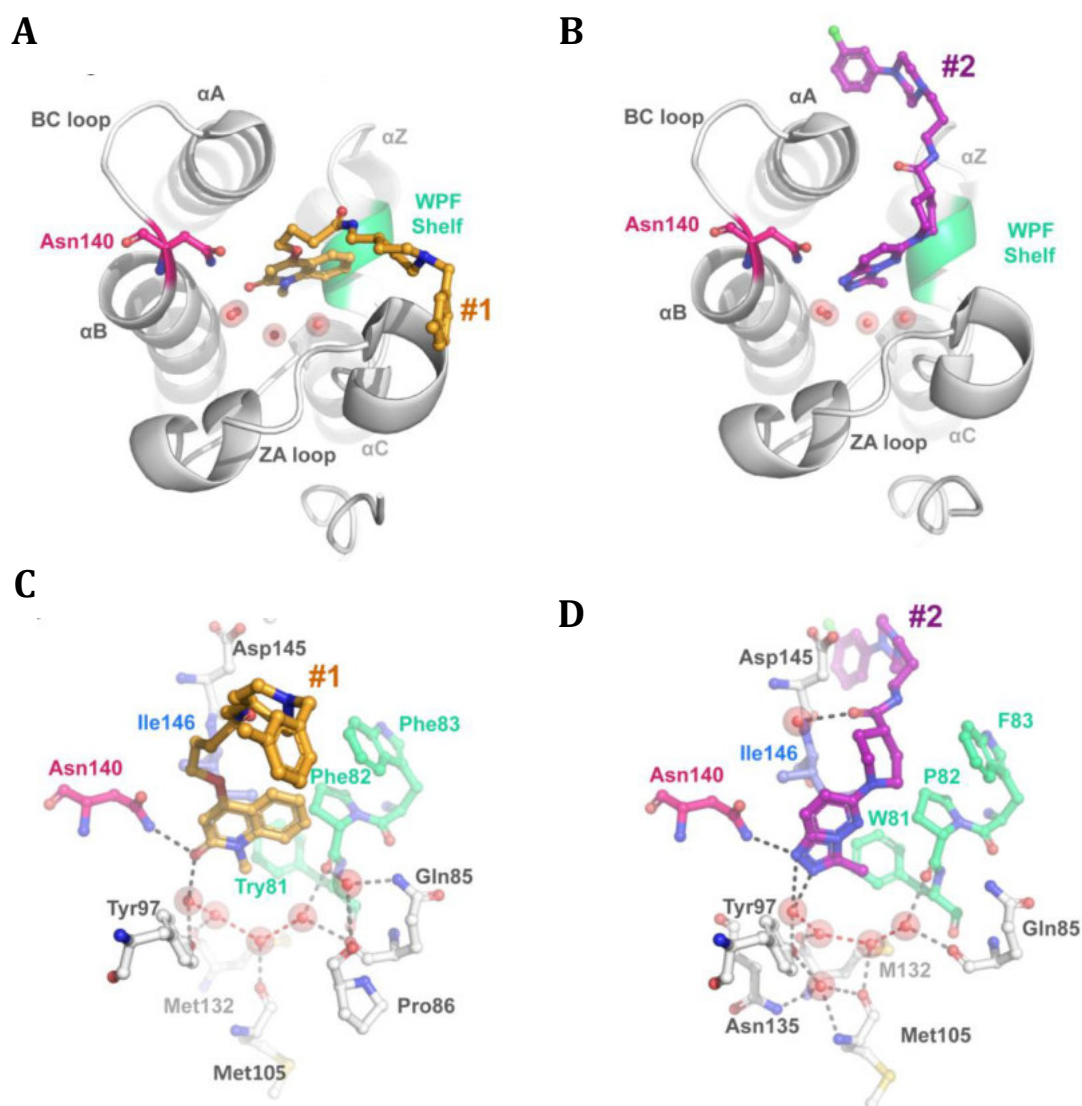
Légende : **A-** Effet des composés 2F20 en bleu et 3G21 en orange sur la viabilité cellulaire de la lignée leucémique Jurkat. **B-** Effet des composés 2F20 en haut et 3G21 en bas sur la régulation de l'oncogène c-Myc (western blot). **C-** Expériences de FRAP réalisées dans des cellules U2OS transfectées avec la protéine GPF-BRD4 et traitées avec les composés 2F20 et 3G21. L'ordonnée indique la demi-période de récupération pour au moins quinze cellules. Le temps de la récupération de fluorescence est également indiqué pour les cellules GFP-BRD4 dont l'asparagine conservée est mutée pour le domaine BD1 (N140F), le domaine BD2 (N433F) et les deux domaines (N140F/N433F). **D-** Les courbes représentent les moyennes à chaque instant pour au moins quinze cellules.

5. Caractérisation du mode de liaison

J'ai résolue les structures cristallographique des complexes BD-ligand impliquant le domaine BRD4(1) et les composés **2F20** (PDB : **5DLZ** - 1,7 Å) et **3G21** (PDB : **5DLX** - 1,9 Å). Les modes de liaison de ces deux molécules au domaine BRD4(1) sont présentés dans la **Figure 25**. Le groupement triazolopyrimidazinyle du composé **3G21**, pénètre dans la cavité K_{ac} et établit une liaison hydrogène avec l'asparagine conservée N140, de la même manière que les lysines acétylées des queues des histones, comme pour le complexe BRD4(1)/H4K5acK8ac (PDB : **3UVW**). Il établit également une liaison avec la tyrosine Y97, médiée par l'une des molécules d'eau enfouies dans cette poche K_{ac} , stabilisant ainsi la structure du complexe. Le réseau d'interaction avec les cinq molécules d'eau conservées dans cette cavité K_{ac} est par ailleurs maintenu. La chaîne latérale carbamoylpiperidinyl *N*-substituée se positionne quant à elle entre la boucle BC et le « WPF shelf », de la même manière que la deuxième lysine acétylée dans la structure du complexe BRD4(1)/H4K5_{ac}K8_{ac} (PDB : **3UVW**). De plus, le groupe carbonyle établit une interaction avec le résidu « gatekeeper » I146, médiée par une molécule d'eau. Le groupe oxoquinolinyle du composé **2F20**, pénètre dans la cavité K_{ac} et établit lui aussi une liaison hydrogène avec l'asparagine N140. Le réseau d'interaction avec les molécules d'eau est également conservé. Pour cette molécule, la densité électronique est bien définie au niveau de ce groupement chimique, cependant le reste de la molécule l'est beaucoup moins (OMIT map situé

dans le supplementary data de l'article), c'est pourquoi nous n'avons pu établir avec certitudes d'autres interactions.

Figure 25 : Structures de BRD4(1) en complexe avec les composés 2F20 et 3G21.



Légende : Les conditions de cristallisation et les tables de statistiques sont en annexe. **A & B-** Vue d'ensemble de la fixation des composés 2F20 et 3G21 au domaine BRD4(1). Représentation du domaine BRD4(1) en « cartoon » de couleur grise. Le domaine BRD4(1) est composé de quatre hélices α (α A, α B, α C, α Z) et deux boucles (BC and ZA). Les quatre molécules d'eau conservées et enfouies dans la cavité K_{ac} sont représentées par des sphères de couleur rouge. L'asparagine conservée N140 est représentée en « balls and sticks » de couleur rose. Le « WPF shelf » est coloré en vert. Les composés sont représentés en « balls and sticks » de couleur orange en pour 2F20 (PDB : **5DLZ** - 1,7 Å) et de couleur violet en pour 3G21 (PDB : **5DLX** - 1,9 Å). **C & D-** Zoom sur la cavité K_{ac} . Les résidus de BRD4(1) situés autour de cette poche de fixation sont représentés en « balls and sticks ». **C-** Le groupement oxoquinoline du composé 2F20 entre dans la cavité K_{ac} et interagit (lignes en pointillés noirs ---) avec l'asparagine N140 et la tyrosine Y97 via une molécule d'eau conservée dans la poche. Le réseau d'interaction entre les molécules d'eau conservées dans la cavité K_{ac} , stabilisant le complexe est représenté par des lignes en pointillés rouges ---. **D-** Le groupement triazolopyridazine du composé 3G21 entre dans la cavité K_{ac} et interagit comme le composé 2F20 avec l'asparagine N140 et la tyrosine Y97. Le reste de la molécule est proche du « WPF shelf » (vert) et le groupement carbonyle établit une liaison hydrogène avec le groupement NH du résidu « gatekeeper » I146 (bleu).

III. Conclusions & Perspectives

Le criblage de la chimiothèque 2P2I_{3D} nous a permis d'identifier sept inhibiteurs ciblant les BDs de la famille BET. Cinq d'entre eux présentent un profil de sélectivité pan-BET classique, c'est le cas des composés **2F20**, **3G21**, **4D11**, **4N21** et **4N15**. Seule la molécule **4N15** présentait une affinité ($IC_{50} = 200$ nM) proche des **critères d'une sonde épigénétique** (page 85). Les deux autres molécules, avec une activité micromolaire, présentaient des profils de sélectivité originaux à savoir BRD4(1) sélective pour la molécule **5A14** et BET-BD1 sélectives pour la molécule **3J7**. À noter qu'en septembre 2014, lorsque j'ai commencé ma thèse dans l'équipe, il n'existait aucune molécule sonde sélective décrite dans la littérature, envers le domaine BD1 des BETs, ou envers un unique BD de cette famille répondant aux critères de sonde épigénétique (page 85). Seule la molécule RVX208 était décrite comme sélective du domaine BD2 des membres de la famille BET. À ce jour, d'autres molécules ont été décrites à savoir RVX-297 qui est également BET-BD2 sélective et les molécules Olinone (210) (PDB : **4QB3**) et MS436 (209) (PDB : **4NUD**) qui sont quant à elles BET-BD1 sélectives, mais aucune ne passe les critères de sonde. L'ensemble de ce travail a fait l'objet de la publication (218) située ci-après.

À l'issue de ce travail, nos perspectives étaient de répondre aux questions suivantes :

- (1) Quels sont les groupements chimiques ou « cœurs chimiques » responsables respectivement de la sélectivité observée et de l'affinité ?
- (2) Quels sont les mécanismes structuraux impliqués dans cette sélectivité ?
- (3) Est-il possible d'optimiser ces molécules pour en faire des « sondes épigénétiques » ou des « candidats médicaments » ?

Pour répondre à l'ensemble de ces questions nous avons initié des programmes de SAR basés sur la comparaison des données d'activité (IC_{50} , K_D , EC_{50}) et les informations obtenues des structures cristallographiques pour la meilleure molécule **pan-BET** (**4N15**), la molécule **BRD4(1) sélective** (**5A14**) et la molécule **BET-BD1 sélective** (**3J7**). Nous avons dans un premier temps fragmenté ces molécules afin d'identifier les « cœurs chimiques » responsables de l'affinité et des profils de sélectivité observés, puis nous avons modifié certains des groupements chimiques en vue d'augmenter l'affinité du micromolaire au nanomolaire tout en tentant de conserver la sélectivité. Dans ce manuscrit, je ne présenterai que les résultats du programme de SAR initié pour la molécule **5A14**.

Article 1

Article N°1

Protein-Protein Interaction Inhibition (2P2I)-Oriented Chemical Library Accelerates Hit Discovery.

Milhas S, ***Raux B***, Betzi S, Derviaux C, Roche P, Restouin A, Basse MJ, Rebuffet E, Lugari A, Badol M, Kashyap R, Lissitzky JC, Eydoux C, Hamon V, Gourdel ME, Combes S, Zimmermann P, Aurrand-Lions M, Roux T, Rogers C, Müller S, Knapp S, Trinquet E, Collette Y, Guillemot JC, Morelli X.

ACS Chem Biol, 2016 Aug 19;11(8):2140-8.

Abstract:

Protein–protein interactions (PPIs) represent an enormous source of opportunity for therapeutic intervention. We and others have recently pinpointed key rules that will help in identifying the next generation of innovative drugs to tackle this challenging class of targets within the next decade. We used these rules to design an oriented chemical library corresponding to a set of diverse “PPI-like” modulators with cores identified as privileged structures in therapeutics. In this work, we purchased the resulting 1664 structurally diverse compounds and evaluated them on a series of representative protein–protein interfaces with distinct “druggability” potential using homogeneous time-resolved fluorescence (HTRF) technology. For certain PPI classes, analysis of the hit rates revealed up to 100 enrichment factors compared with nonoriented chemical libraries. This observation correlates with the predicted “druggability” of the targets. A specific focus on selectivity profiles, the three-dimensional (3D) molecular modes of action resolved by X-ray crystallography, and the biological activities of identified hits targeting the well-defined “druggable” bromodomains of the bromo and extraterminal (BET) family are presented as a proof-of-concept. Overall, our present study illustrates the potency of machine learning-based oriented chemical libraries to accelerate the identification of hits targeting PPIs. A generalization of this method to a larger set of compounds will accelerate the discovery of original and potent probes for this challenging class of targets.

PMID: 2721984

Protein–Protein Interaction Inhibition (2P2I)-Oriented Chemical Library Accelerates Hit Discovery

Sabine Milhas,^{†,‡} Brigitt Raux,[†] Stéphane Betzi,[†] Carine Derviaux,[†] Philippe Roche,[†] Audrey Restouin,[†] Marie-Jeanne Basse,[†] Etienne Rebuffet,[†] Adrien Lugari,^{†,○} Marion Badol,[§] Rudra Kashyap,^{†,||} Jean-Claude Lissitzky,[†] Cécilia Eydoux,[‡] Véronique Hamon,[†] Marie-Edith Gourdel,[⊥] Sébastien Combes,[†] Pascale Zimmermann,^{†,||} Michel Aurrand-Lions,[†] Thomas Roux,[§] Catherine Rogers,^{#,∇} Susanne Müller,^{#,∇} Stefan Knapp,^{#,∇,¶} Eric Trinquet,[§] Yves Collette,[†] Jean-Claude Guillemot,[‡] and Xavier Morelli^{*,†,‡}

[†]CNRS, INSERM, Aix-Marseille Université, Institut Paoli-Calmettes, Centre de Recherche en Cancérologie de Marseille, CS30059, 13273 Marseille Cedex 9, France

[‡]CNRS, Aix-Marseille Université, Screening Platform AD2P, AFMB UMR7257, 13288, Marseille, France

[§]Cisbio Bioassays, R&D, Parc Marcel Boiteux, BP 84175, 30200 Codolet, France

^{||}Department of Human Genetics, KU Leuven, B-3000 Leuven, Belgium

[⊥]Hybrigenics Services, 3-5 Impasse Reille, 75014 Paris, France

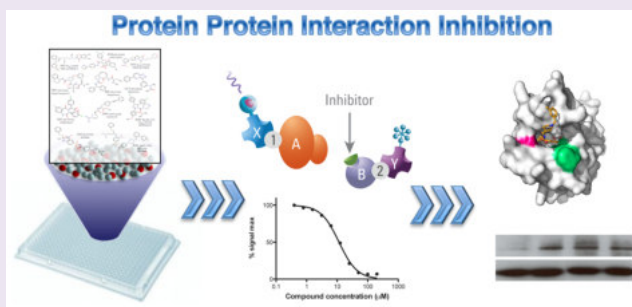
[#]Target Discovery Institute, University of Oxford, NDM Research Building, Roosevelt Drive, Oxford OX3 7FZ, U.K.

[∇]Structural Genomics Consortium, University of Oxford, Old Road Campus Research Building, Roosevelt Drive, Oxford OX3 7DQ, U.K.

[¶]Goethe-University, Institute for Pharmaceutical Chemistry and Buchmann Institute for Life Science, Campus Riedberg, Max-von-Laue Str. 9, 60438 Frankfurt am Main, Germany

Supporting Information

ABSTRACT: Protein–protein interactions (PPIs) represent an enormous source of opportunity for therapeutic intervention. We and others have recently pinpointed key rules that will help in identifying the next generation of innovative drugs to tackle this challenging class of targets within the next decade. We used these rules to design an oriented chemical library corresponding to a set of diverse “PPI-like” modulators with cores identified as privileged structures in therapeutics. In this work, we purchased the resulting 1664 structurally diverse compounds and evaluated them on a series of representative protein–protein interfaces with distinct “druggability” potential using homogeneous time-resolved fluorescence (HTRF) technology. For certain PPI classes, analysis of the hit rates revealed up to 100 enrichment factors compared with nonoriented chemical libraries. This observation correlates with the predicted “druggability” of the targets. A specific focus on selectivity profiles, the three-dimensional (3D) molecular modes of action resolved by X-ray crystallography, and the biological activities of identified hits targeting the well-defined “druggable” bromodomains of the bromo and extraterminal (BET) family are presented as a proof-of-concept. Overall, our present study illustrates the potency of machine learning-based oriented chemical libraries to accelerate the identification of hits targeting PPIs. A generalization of this method to a larger set of compounds will accelerate the discovery of original and potent probes for this challenging class of targets.



The modulation of protein–protein complexes using small molecule compounds is becoming more and more popular for the development of probes that can result in new therapeutic drugs or original tools to decipher biological processes,^{1–5} particularly because of the multiple and complex implications of protein–protein interactions (PPIs) in cellular signaling networks. Over the last 15 years, the number of success stories of small molecular inhibitors of PPIs has grown considerably. There are currently thousands of small molecule compounds targeting more than 50 protein–protein complexes (please refer to <http://www.ipidb.cdthem.fr/>,^{6,7} <https://www.ebi.ac.uk/chembl/>,⁸ or

<http://mordred.bioc.cam.ac.uk/timbal>^{9,10} for more details). Therefore, a wealth of information is available to define what makes a protein–protein interaction a good target for drug discovery (referred to as the “druggability” problem) and to characterize the potential profile of PPI modulators. Despite being a popular strategy both in academic laboratories and in pharmaceutical companies, experimental screening of large

Received: March 30, 2016

Accepted: May 24, 2016

Published: May 24, 2016

collections of diverse compounds has delivered lower hit rates than expected when targeting protein–protein complexes.^{11–14} One of the reasons for these low success rates is thought to result from the inadequacy of the libraries used to find the “hidden gem”. It is therefore crucial to improve the quality and adequacy of the chemical libraries used to search for active PPI compounds to improve success rates and lower the costs of experimental screening campaigns. To this end, we have recently developed 2P2I_{DB}, a hand-curated structural database dedicated to PPIs with known orthosteric inhibitors (<http://2p2idb.cnrs-mrs.fr>).^{15–17} Analysis of the small molecule inhibitors present in 2P2I_{DB} led us to propose the “rule-of-four” as a guideline for characterizing PPI inhibitors.¹⁸ Using dedicated support vector machine approaches, we developed 2P2I_{HUNTER}, a tool that allows potential orthosteric PPI modulators to be filtered from large collections of compounds.¹⁹ This innovative tool has been applied to a set of 8.3 million compounds from the “big vendors” to design several *in silico* PPI-oriented chemical libraries.²⁰ Compounds corresponding to medicinally important privileged structures identified as core structures in therapeutics^{21,22} were prioritized in a medicinally oriented version of the library, 2P2I_{PRIV}. The library was subsequently filtered for diversity and with the carbon bond saturation index (Fsp3) to escape from flatland, which resulted in a structurally diverse and medicinally oriented chemical library composed of 1664 compounds named 2P2I_{3D}.²⁰ Notably, these compounds retain reasonable drug-like properties, including Lipinski’s “rule of five”, Veber’s rules, and other oral bioavailability rules (please refer to Zhang et al.⁵ for more details). In the present study, the ability of this PPI-oriented library to enhance hit rates in screening campaigns was experimentally evaluated using homogeneous time-resolved fluorescence (HTRF) assay, an established screening technology. Each molecule of the chemical library was evaluated against a representative set of protein–protein interfaces. This set included (i) a protein–protein large-sized globular interface (Dengue virus NS3/NS5 PPI complex), which represented the globular type I protein–protein interface, as defined by Basse et al.,¹⁶ (ii) a medium-sized protein–domain interface (the HIV virus Nef protein in complex with the SH3 domain from Hck kinase), a type I protein/domain interface (with several screens already referenced in PubChem and PubMed),^{23–25} (iii), a smaller protein–peptide interface (a bromodomain in complex with its acetylated histone peptide), which

represents a “druggable” protein/peptide interface as exemplified by several compounds in preclinical and clinical development,^{26,27} and (iv) two smaller protein–peptide interfaces (two PDZ domain-containing proteins in complex with their cognate peptides), which represent “poorly druggable” or challenging protein–peptide interfaces.^{28,29} The physicochemical properties of the validated hits are presented and discussed as key factors to accelerate the identification of hits targeting PPIs. For example, the selectivity profiles, three-dimensional molecular modes of action and biological activities of original validated hits identified from the primary screening targeting the bromodomains of BET family members are revealed. The present work using our 2P2I_{3D} “PPI-oriented” chemical library consolidates the proof-of-concept that the design of oriented chemical libraries through support vector machine-based approaches will significantly accelerate the identification of probes targeting PPIs.

RESULTS AND DISCUSSION

In Vitro (HTRF) Evaluation of the “2P2I_{3D}” Library. The autofluorescence of each compound in the library was evaluated before screening on a representative set of protein–protein interfaces. When 20 μ M of each compound was used for screening, we observed that 3.2% and 2.3% of the compounds were significantly autofluorescent at the HTRF detection wavelengths of 665 and 620 nm, respectively (Figure S1). This rate may be higher than the rate obtained with classic “nonoriented” chemical libraries but remains low and is thus compatible with HTRF experiments.

In addition to the ability of HTRF to limit compound interferences in high-throughput screening (HTS),³⁰ it is also flexible in the design of screening assays. Because HTRF can accommodate the use of a wide range of tag/antitag reagents, the design of robust assays to screen the 2P2I_{3D} library was noticeably simplified, and we could establish robust assays based on the four representative protein–protein interfaces (Figure S1a and Table 1). Of note, each complex has been primarily validated with competition, when available, in the presence of an untagged protein partner (refer to row 3, Table 1 for the validated IC₅₀’s and to Figure S3 for curves) or in the presence of a positive control (i.e., JQ1, a well-known probe for the BRD4/acetylated peptide complex highlighted a similar IC₅₀ for BRD4(BD1)

Table 1. High Throughput Screening Results^a

protein partners	NS3/NS5 (Dengue virus)	Nef/SH3-Hck (HIV virus)	bromodomain/H4 (K _{acs} /8/12/16)	Grasp55/JamB (PDZ domain)	syntenin/syndecan (PDZ domain)
interface type	protein/protein	protein/domain	protein/peptide	protein/peptide	protein/peptide
K _D (μ M)	1.0 ³⁹	1.5 ⁴⁰	2.8 ⁴¹	1 ^b	1.8 ⁴²
HTRF IC ₅₀ value from competition with untagged protein (μ M)	0.4	N/A	3.0 ^d	8.2	9.2
Z'	0.70	0.88	0.78	0.81	0.92
no. hits non-autofluorescent (inhibition >40%)	39	16	44	44	19
no. validated molecules (secondary screening)	26 (1.6%)	2 (0.1%)	17 (1%)	2 (0.1%)	3 (0.2%)
no. IC ₅₀ measured	26	2	17	2	3
no. unique molecules	22	2	15	2	1
no. thermal shift sensitive	10	0	10	1	c

^aSummary of the results obtained during the screening process of the 2P2I_{3D} chemical library on various PPI targets. The first row indicates the type of interface, and the second and third rows compare K_D’s extracted from the literature with IC₅₀’s measured by HTRF (this work) in the presence of the untagged protein. The quality of the screening procedure is provided in the fourth row, in which the Z' value is specified for each complex. The sixth row corresponds to the number of compounds validated after the counter screen and the secondary screening, and the seventh row indicates the number of compounds validated by IC₅₀’s measurements. Finally, the number of unique compounds, those never validated for another screening, are shown in the eighth row, and the number of thermal shift sensitive molecules, i.e., directly interacting with one of the 2 proteins, is specified in the last row. ^bInternal data. ^cNot applicable. ^dInhibitory constant (K_i) value calculated with Cheng and Prusoff equation.

(Figure S3), 100 nM, compared with the value of 76.9 nM previously observed in the literature (<http://www.thesgc.org/chemical-probes/JQ1>)).

The primary screening of our 2P2I_{3D} library was then performed at 20 μM against each of the above-described PPIs. Compounds exceeding 40% of the HTRF signal inhibition were then tested in a counterscreen assay setup with fusion protein composed only of the tags used to link the proteins of interest with the HTRF probes (for example, GST-biotin or GST-His, Figure S1a). This counter screen was used to eliminate compounds exhibiting a nonspecific inhibition profile for the two proteins of interest. Compounds exhibiting less than 15% inhibition in the counterscreen assay were then validated as potential inhibitors in a secondary screening assay. Finally, the IC₅₀'s of the confirmed compounds were precisely determined using fresh powder reordered to the corresponding provider and tested in the presence of diverse buffers, including bovine serum albumin (BSA) and DTT or Tween, to reduce the selection of false positives (refer to Figure S2 for more details regarding the screening cascade).

The Z'-factors of the primary screens measured for these five complexes ranged from 0.7 to 0.9, underlining the quality of the screening procedure³¹ (Figure S1b and Figure S3). Hit rates, defined as the percentage of compounds reaching pre-established activity criteria after primary and secondary screenings (i.e., inhibition higher than 40%, less than 15% inhibition in the counterscreen and a dose–response profile in the secondary screening, and IC₅₀'s measured in triplicate), ranged from 0.2% for the PDZ complexes to 1.6% for the protein–protein complex involving the NS3 and NS5 proteins from the Dengue virus (Table 1).

The experimental hit rates measured in the screening campaign presented large enrichment factors (at least 10 to 100×) compared with nonoriented libraries evaluated against PPIs.⁴ Notably, these primary hit rates also reflected the “druggability” of the targets. We observed a high validation rate of 17 hits, among which 15 were unique in this evaluation study (i.e., 1% hit rate of at least 10 to 50× compared with nonoriented libraries) when screening the chemical library against the BRD4(1)/acetylated peptide complex, a well-known “druggable” protein–peptide complex. By contrast, the hit rate value dropped down to 0.2% for the PDZ/C-terminal peptide complexes evaluated here. This PPI interface has been categorized as a challenging PPI interface; several research articles have reported on failures to identify validated hits in a large HTS campaign involving more than 100 000 compounds²⁸ or to identify few “mild” PDZ binders in another HTS program.²⁹ Our protocol, if applied on a larger scale of approximately 10 000 to 100 000 compounds, would dramatically increase the chance of discovering a true probe amenable to medicinal chemistry programs. Finally, the Nef/SH3-Hck protein/domain complex has been presented as a potential target in several seminal papers. Among them, Emert-Sedlak et al.²⁴ presented a hit rate of 0.01% (three different libraries from ChemDiv, Inc., were screened at 10 μM), whereas Breuer et al. observed a similar ratio when screening the LOPAC library.²⁵ We observed a hit rate of 0.1% for the same interface using our 2P2I_{3D} library as a template when screening at 20 μM. Although it is difficult to calculate the enrichment factor for this complex because these different experiments were not performed in parallel, a 5–10-fold enrichment ratio could be predicted.

Most validated hits were specific to the complex of interest and not identified as hits when screened against other PPI complexes. This result demonstrates the quality, diversity, and selectivity of our smaller chemical library (Table 1, number of unique

molecules identified for each complex). However, these numbers could potentially drop when increasing the number of complexes to be screened. Furthermore, we decided to perform cross-validation experiments of these unique molecules identified by HTRF (i.e., on another PPI complex defined as a “control” experiment, here the “BRD4/H4” PPI complex) and by differential scanning fluorimetry (DSF) on the predicted targeted protein. We report two different cases according to the studied complexes (Figure 1). In the first case scenario, for the compounds identified for the PDZ-Grasp55/JamB peptide and the NS3/NS5 complex, the identified hits did not inhibit the PPI “control” (Figure 1a,c); they influenced the DSF curve of the protein target, Grasp55 or NS3 (Figure 1b,d), including those in the presence of DTT or Tween. This result demonstrated that these compounds were not acting by promiscuous, aggregation-based inhibition^{32,33} (Figure 1b). In the second scenario, for the two compounds identified for the Nef/SH3-Hck complex and the one identified for the syntenin/syndecan complex, similar IC₅₀ values were obtained with the PPI “control” used for cross-validation, and therefore these compounds were not further studied in the present work. We thus decided to further investigate some of these unique hits to more deeply evaluate their selectivity profiles and biological activities (please refer to Table 2 and S1 for more details regarding the 2D chemical structures, IC₅₀'s, or provider's information for all unique, validated compounds). In particular, we performed supplementary orthogonal validation using isothermal titration calorimetry (ITC), X-ray crystallography, and cellular evaluation for one target of interest, the bromodomain BRD4/H4 complex.

Focusing on One Target of Interest - Bromodomains.

Inhibitors identified during the screening of the bromodomain/acetylated lysine-containing peptide target complex were evaluated more thoroughly for three main reasons: (i) the high hit rate obtained for this complex (1%), (ii) its unexpected high specificity (15 of the 17 compounds were not identified in any of the other PPI complexes evaluated in this program), and (iii) bromodomains have recently received much attention because they represent a new class of targets with broad interest in epigenetics and cancer, and a large quantity of data are thus available in the literature to compare with our identified hits.

Bromodomains (BRDs) are protein–protein interaction modules that are selectively recruited by N-acetylated lysine-containing sequences, particularly histone tails. For this reason, BRD-containing proteins function as “readers” of histone lysine acetylation, a component of the proposed “histone code”, and thereby act as scaffolds for the recruitment of transcription factors and chromatin organizers required in transcription initiation and elongation. BRD-containing proteins have been involved in the development of a large variety of diseases, and their highly “druggable” acetyl-lysine binding pocket architectures make them attractive targets for the development of potent and specific inhibitors (please refer to ref 26 or 27 for more details regarding current specific inhibitors developed in the clinic). In particular, BRD2, BRD3, BRD4, and the testis-specific BRDT proteins, which form the bromo and extraterminal (BET) subfamily, have recently received much attention after the discovery of highly potent and cell-active pan-BET inhibitors. These inhibitors display strong phenotypic effects in several cell lines and affect a range of cancers *in vivo* (for more details please visit the SGC consortium Web site: <http://www.thesgc.org/>). The recent developments with this class of targets have garnered considerable interest in the epigenetic community to discover new and potent compounds specific for each of the 61 known modules present in

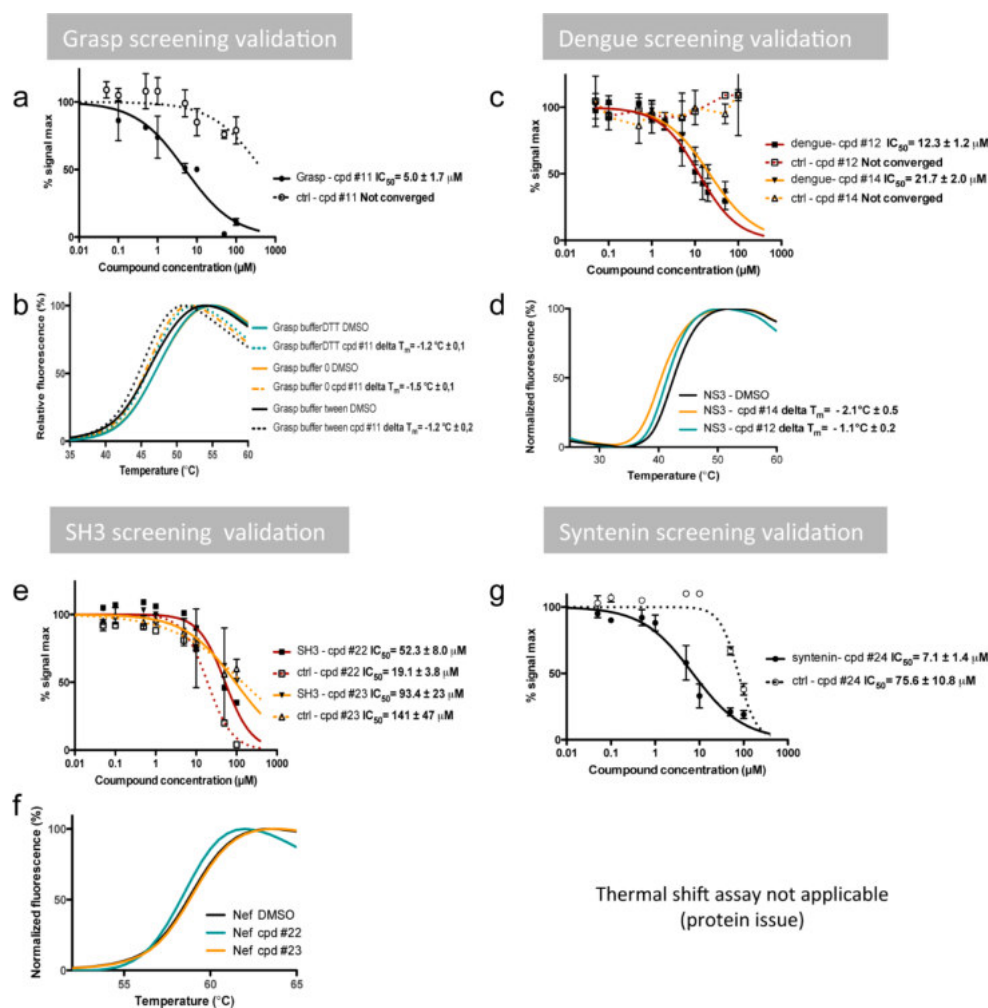


Figure 1. Cross-validation experiments for each identified hit. Examples of IC_{50} cross-validation experiments performed by HTRF on an “irrelevant/control” assay (BRD4/H4 complex) to assess the specificity of each identified hit. In a second experiment, each potential hit was evaluated by differential scanning fluorimetry (DSF) experiments to validate direct binding with an orthogonal method. (a) HTRF IC_{50} assays performed for compound 11 on Grasp55/JamB (plain curve) compared with the “irrelevant/control” assay (dashed curve). (b) DSF experiments performed for compound 11 on Grasp55. The bold curves represent Grasp55 alone in the presence of 2.5% DMSO, whereas dashed curves represent the same experiment with a 25-fold excess of compound 11. Experiments were performed in “0” buffer (HEPES, pH 7.5, 150 mM NaCl), DTT buffer (buffer 0 + 2 mM DTT), and Tween buffer (buffer 0 + 0.012% Tween20). (c) Examples of HTRF IC_{50} assays performed for two potential hits (compounds 12 and 14) on NS3/NS5 (plain curve) compared with the “irrelevant/control” assay (dashed curve). (d) DSF experiments performed for the same two compounds on NS3. The black curve represents NS3 alone in the presence of 2.5% DMSO, whereas the green and orange curves represent the same experiment with compounds 12 and 14. (e) HTRF IC_{50} assays performed for the two HTRF-validated compounds (named compounds 22 and 23 on Nef/SH3 (plain curve) compared with the “irrelevant/control” assay (dashed curves). (f) DSF experiments performed for compounds 22 and 23 on Nef. The black curve represents Nef alone in the presence of 2.5% DMSO, whereas the green and orange curves represent the same experiment with compounds 22 and 23. (g) HTRF IC_{50} assays performed for the HTRF validated compound 24 on syntenin/syndecan compared with the “irrelevant/control” assay (dashed curves). DSF experiment could not be achieved using syntenin as a target.

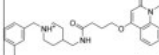

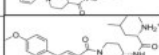
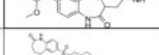
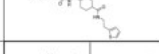
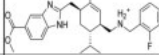
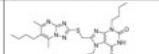
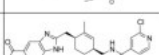
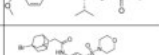
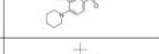
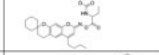
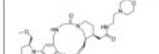
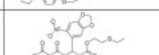
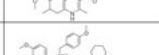
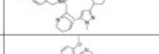
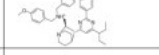
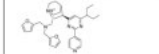
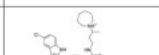
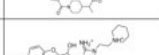
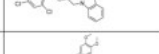
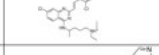
46 human proteins.³⁴ The current challenge still lies in the identification of subfamily specific or bromodomain-specific modulators. To assess the quality of our established library and screening results, we thus further evaluated the 15 hits targeting the BRD4(1)/acetylated peptide complex in an orthogonal validation assay using ITC and determined their BRD selectivity profiles.

Orthogonal Validation and Selectivity Profile. ITC was chosen for the orthogonal validation because it is direct and in solution. Of the 15 unique HTRF-validated hits, we confirmed the interaction of 7 molecules with the BRD4 bromodomain with K_D values ranging from 1.8 to 4.7 μ M (Table 3 and Figure S4). The large overlap observed between the number of hits for the bromodomain/acetylated lysine-containing peptide complex and the number of compounds confirmed by the orthogonal

assay (ITC) demonstrate the robustness of our HTRF screening procedure (in particular, the use of counter screens assays to reject compounds interfering with the detection system is mandatory to limit false positives).

The two most potent compounds (Figure S5a), a quinolone derivative, compound 1 (CID no. 20862085), and a triazolopyridazine derivative, compound 2 (CID no. 46326697), exhibited ITC thermograms fitted to 1.8 and 2.0 μ M K_D , respectively (Figure S5 and Table 2). DSF experiments revealed a ΔT_m of +1.9 and +1.5 $^{\circ}$ C, respectively (at 1:25 protein/compound excess) (Figure S5c). Both experiments validated the direct binding of the compounds on the BRD4(1) domain. A selectivity profile of the best two compounds on the BET family members displayed a “pan-inhibitor” profile with IC_{50} 's ranging

Table 2. Two-Dimensional Chemical Structure and Features of the Thermal Shift Sensitive Validated Compounds^a

Target	Compound #	Structure	IC ₅₀ (μM)	PAINS prediction	Potential aggregator	Molecular weight (g.mol ⁻¹)	PubChem Compound Identifier (CID)	Smiles
Bromodomain	1		8.4	0	no	461.6	20862085	<chem>C1=CC=C(C=C1)C(=O)CC(=O)OCC1=CC=C(C=C1)C(=O)N1C=CC=C(N1)C</chem>
Bromodomain	2		7.8	0	yes	497.0	46326697	<chem>C1=CC=C(C=C1)C(=O)CC(=O)OCC1=CC=C(C=C1)C(=O)N1C=CC=C(N1)C</chem>
Bromodomain	3		12.9	0	no	509.6	7134939	<chem>N1C=CC=C(C=C1)C(=O)CC(=O)OCC1=CC=C(C=C1)C(=O)N1C=CC=C(N1)C</chem>
Bromodomain	4		8.9	0	yes	489.7	20937233	<chem>C1=CC=C(C=C1)C(=O)CC(=O)OCC1=CC=C(C=C1)C(=O)N1C=CC=C(N1)C</chem>
Bromodomain	5		22.8	0	no	463.6	71694343	<chem>C1=CC=C(C=C1)C(=O)CC(=O)OCC1=CC=C(C=C1)C(=O)N1C=CC=C(N1)C</chem>
Bromodomain	6		18.0	0	no	484.6	60498769	<chem>C1=CC=C(C=C1)C(=O)CC(=O)OCC1=CC=C(C=C1)C(=O)N1C=CC=C(N1)C</chem>
Bromodomain	7		0.2	0	no	495.0	71693175	<chem>C1=CC=C(C=C1)C(=O)CC(=O)OCC1=CC=C(C=C1)C(=O)N1C=CC=C(N1)C</chem>
Bromodomain	8		37.9	0	yes	580.6	4658136	<chem>C1=CC=C(C=C1)C(=O)CC(=O)OCC1=CC=C(C=C1)C(=O)N1C=CC=C(N1)C</chem>
Bromodomain	9		25.2	0	no	526.7	16416191	<chem>C1=CC=C(C=C1)C(=O)CC(=O)OCC1=CC=C(C=C1)C(=O)N1C=CC=C(N1)C</chem>
Bromodomain	10		13.6	1	no	569.7	71694343	<chem>C1=CC=C(C=C1)C(=O)CC(=O)OCC1=CC=C(C=C1)C(=O)N1C=CC=C(N1)C</chem>
Grasp	11		10.3	0	no	532.6	3324458	<chem>C1=CC=C(C=C1)C(=O)CC(=O)OCC1=CC=C(C=C1)C(=O)N1C=CC=C(N1)C</chem>
NS3 NS5	12		12.3	0	no	542.8	44715900	<chem>C1=CC=C(C=C1)C(=O)CC(=O)OCC1=CC=C(C=C1)C(=O)N1C=CC=C(N1)C</chem>
NS3 NS5	13		7.1	0	no	605.8	45360407	<chem>C1=CC=C(C=C1)C(=O)CC(=O)OCC1=CC=C(C=C1)C(=O)N1C=CC=C(N1)C</chem>
NS3 NS5	14		20.6	0	no	525.7	45360408	<chem>C1=CC=C(C=C1)C(=O)CC(=O)OCC1=CC=C(C=C1)C(=O)N1C=CC=C(N1)C</chem>
NS3 NS5	15		25.5	0	no	487.1	50816819	<chem>C1=CC=C(C=C1)C(=O)CC(=O)OCC1=CC=C(C=C1)C(=O)N1C=CC=C(N1)C</chem>
NS3 NS5	16		20.6	0	no	463.4	1306476	<chem>C1=CC=C(C=C1)C(=O)CC(=O)OCC1=CC=C(C=C1)C(=O)N1C=CC=C(N1)C</chem>
NS3 NS5	17		18.8	0	no	517.5	5793057	<chem>C1=CC=C(C=C1)C(=O)CC(=O)OCC1=CC=C(C=C1)C(=O)N1C=CC=C(N1)C</chem>
NS3 NS5	18		21.8	0	no	526.7	40777317	<chem>C1=CC=C(C=C1)C(=O)CC(=O)OCC1=CC=C(C=C1)C(=O)N1C=CC=C(N1)C</chem>
NS3 NS5	19		12.4	1	no	579.6	40776845	<chem>C1=CC=C(C=C1)C(=O)CC(=O)OCC1=CC=C(C=C1)C(=O)N1C=CC=C(N1)C</chem>
NS3 NS5	20		21.1	0	no	477.7	25313771	<chem>C1=CC=C(C=C1)C(=O)CC(=O)OCC1=CC=C(C=C1)C(=O)N1C=CC=C(N1)C</chem>
NS3 NS5	21		13.1	1	no	536.0	44716462	<chem>C1=CC=C(C=C1)C(=O)CC(=O)OCC1=CC=C(C=C1)C(=O)N1C=CC=C(N1)C</chem>

^aSummary of the 2D chemical structure, molecular weight, and IC₅₀ values obtained with “thermal shift sensitive” inhibitors. Pan-assay interference compounds (“PAINS”) and aggregator predictions are indicated based on FAFDrugs³ and Aggregator advisor (<http://advisor.bkslab.org/>), respectively. Enlarged 2D chemical structures are provided in Table S1.

from 3.7 to 25 μM for compound 1 (Figure S5a,c) and from 3.6 to 20 μM for compound 2 (Figure S6b,c). The selectivity window was more than 10 times higher than ATAD2, a non-BET family member, thus representing an excellent starting point for the development of more potent BET bromodomain leads.

Binding Modes Determined by X-ray Crystallography. To further demonstrate the direct binding of the inhibitor, we set up

a cocrystallization assay and elucidated the 3D structure for both compounds 1 (PDB code SDLZ) and 2 (PDB code 5DLX) in complex with BRD4(1). BRD4(1) readily crystallized in the presence of the inhibitors, and crystallographic data collection and refinement statistics are described in Table S2. The structures of these complexes with compound 1 and compound 2 were resolved at 1.7 and 1.9 Å resolution, respectively (Figure 2).

Table 3. ITC Thermodynamic Parameters, IC_{50} , and Ligand Efficiency Values^a

compd no.	ITC			MW (Da)	ligand efficiency
	K_D ITC (μM)	ΔH (cal/mol)	ΔG (kcal/mol)		
1	1.8	-8664 ± 129	-7.56 ± 0.04	462.7	0.22
2	2.0	-7996 ± 84.8	-7.48 ± 0.07	498.1	0.21
3	2.0	-4547 ± 414.3	-7.49 ± 0.14	509.7	0.20
4	2.2	-9174 ± 84.8	-7.44 ± 0.02	489.7	0.23
5	3.6	$-1.276 \times 10^4 \pm 2358$	-7.16 ± 0.12	464.6	0.21
6	4.5	-3915 ± 868.6	-7.04 ± 0.29	484.7	0.21
7	4.7	$-1.288 \times 10^4 \pm 3172$	-6.97 ± 0.05	495.1	0.20

^aBinding affinities and thermodynamic parameters measured by ITC during the orthogonal validation process of the hits identified by HTRF on BRD4(1). Ligand efficiency (LE) values are indicated based on the ΔG values obtained from the ITC experiments.

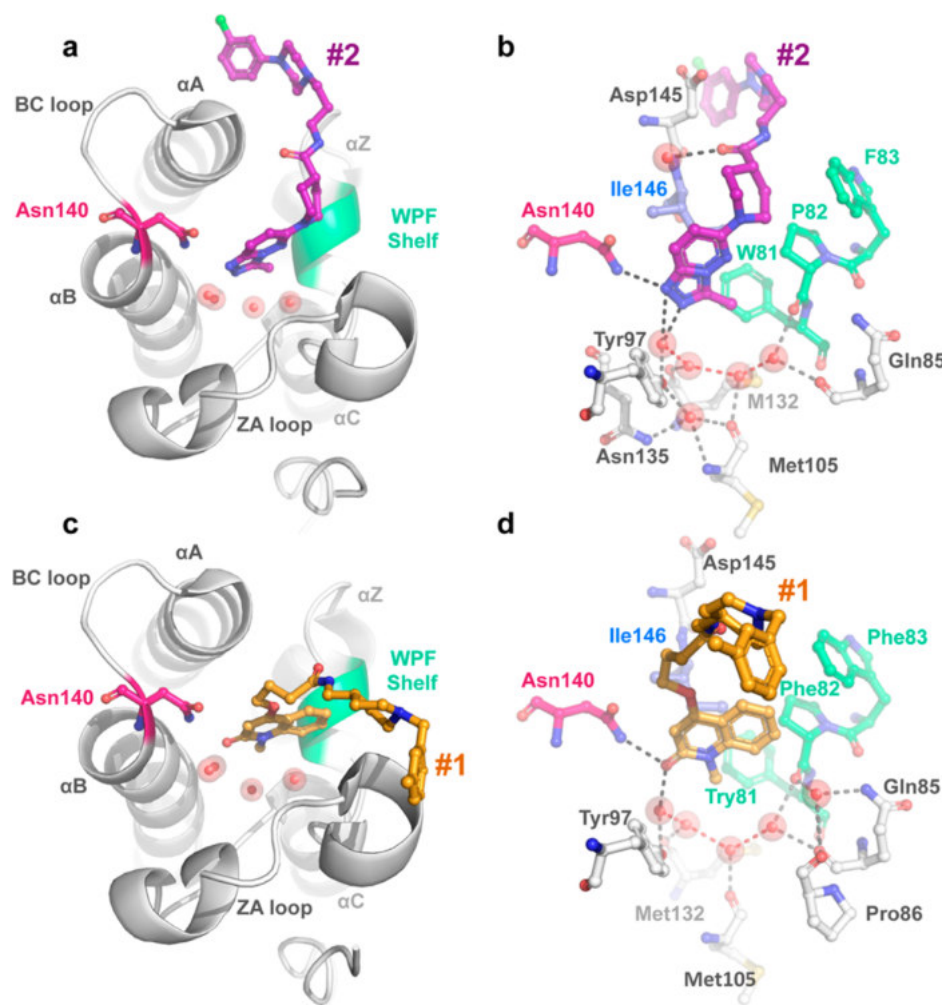


Figure 2. Structures of the BRD4(1) complexes with compounds 1 and 2. (a) Overall folding of BRD4(1) (gray, cartoon representation) in complex with compound 2 (purple ball and stick representation; PDB code SDLX). BRD4(1) is folded into four α helices (αA , αB , αC , αZ) and two loops (BC and ZA). Compound 2 (purple) binds in the N-acetylated-lysine binding pocket through interactions with the conserved asparagine, Asn140 residue (pink), and four conserved water molecules deep in the cavity (red spheres). (b) The triazolopyridazine moiety of compound 2 occupies the N-acetylated-lysine pocket and initiates a direct hydrogen bond interaction with Asn140 (pink). Four water molecules are located at the bottom of the BRD4(1) cavity (red spheres), as seen in the N-acetylated lysine peptide complex structure (PDB ID 3UVW). The water molecule hydrogen bond network (red dash lines) contributes to stabilize the ligand through an additional hydrogen bond. The rest of the molecule is close to the BRD4(1) WPF shelf residues (green), and the carbonyl moiety establishes a water bridge hydrogen-bond interaction with the NH amide of the gatekeeper residue Ile146 (blue). (c) Overall folding of BRD4(1) in complex with compound 1 (orange ball and stick representation; PDB code SDLZ). (d) The oxoquinoline moiety of compound 1 occupies the N-acetylated-lysine pocket by contacts with the Asn140 and the conserved water molecules.

The $2F_o - F_c$ map at 1σ around compound 2 was well-defined for the entire compound (Figure S7). The triazolopyridazinyl moiety of this compound penetrated deeply into the N-acetylated lysine pocket and occupied the same position as the acetyl

headgroup of the acetylated lysine present in the histone tail peptides, initiating a direct hydrogen bond with the conserved asparagine (Asn140) (Figure 2a). Four water molecules were located at the bottom of the BRD4(1) cavity, as observed in

the N-acetylated lysine peptide complex structure (PDB ID 3UVW) and in other BRD4(1) inhibitor complexes. The water molecule hydrogen bond network contributes to stabilizing the ligand through additional hydrogen bonds (Figure 2b). The N-substituted carbamoylpiperidinyl side chain occupied the same localization as the second acetylated lysine, as observed in the peptide complex (PDB ID 3UVW), engaging the shelf between the BC loop and the WPF motif composed of the amino acids Trp81, Pro82, and Phe83 (Figure 2a and Figure S8). The carbonyl group establishes a water bridged hydrogen bond interaction with the NH amide of the gatekeeper residue Ile146 stabilizing the complex. The $2F_o - F_c$ map at 1σ around the oxoquinolinyl moiety of compound 1 was well-defined (Figure S7), whereas a weaker electron density was observed for the rest of the molecule outside of the N-acetylated binding pocket. Similar to compound 2, the oxoquinolinyl ring of the molecule penetrated deeply into the N-acetylated lysine pocket. The four water molecules, which contribute to stabilizing the ligand through an additional hydrogen bonds, were also conserved (Figure 2c,d and Figure S8).

Cellular Evaluation. BETs are transcriptional regulators that control the expression of genes that play key regulatory roles in cellular proliferation, cell cycle progression, and apoptosis. BRD4 is known to be a key regulator of cell cycle control and the transcriptional elongation of growth-promoting genes, including *c-myc* and its target genes.³⁵ Down-regulation of *c-myc* is a unifying, yet not exclusive, mechanistic theme reported for the various developed BET inhibitors.

We thus evaluated the antiproliferative capacities of compounds 1 and 2 and their effects on the levels of *c-myc* expression in tumor cell lines. Both compounds 1 and 2 reduced the viability of Jurkat T cells in a dose-dependent manner (Figure 3a), with EC_{50} values of 31 ± 1.9 and $21 \pm 0.5 \mu\text{M}$, respectively. Both compounds also induced dose-dependent down-regulation of *c-myc* protein levels at 1, 10, and $50 \mu\text{M}$, as determined by immunoblotting (Figure 3b), at a similar effective dose observed in cell viability studies. To assess whether compounds 1 and 2 engage with their target protein BRD4 in live cells and are able to displace BRD4 from chromatin binding, fluorescence recovery after photobleaching (FRAP) experiments were performed in U2OS cells overexpressing GFP-tagged BRD4. After photobleaching, the diffusion of unbleached protein back into the bleached region is retarded by protein binding to chromatin and chromatin-associated complexes and is therefore slower compared with freely diffusible molecules. Both compounds were able to significantly accelerate half-recovery times in FRAP experiments at $10 \mu\text{M}$ (Figure 3c), suggesting that the inhibitors effectively displaced BRD4(1) from chromatin. These data are thus consistent with the *in vitro* potencies of these compounds and their observed antiproliferative effects.

The present work demonstrates the potential to exploit PPI-oriented chemical libraries using HTRF in a general screening cascade to accelerate hit identification targeting PPIs.

For example, the screening of our 2P2I_{3D} over the bromodomain target identified 15 hits in the low micromolar range. We used ITC as an orthogonal assay and validated direct binding for 7 of the 17 compounds tested (Table 3), with ligand efficiency (LE) values ranging from 0.20 to 0.23. The best compounds (K_D of 1.8 and $2 \mu\text{M}$), a quinolone derivative (compound 1) and a triazolopyridazine derivative (compound 2), both compatible with the “rule of five” filter,³⁶ demonstrated pan-inhibition profiles, with selective inhibition of the BET family proteins versus non-BET members. Finally, demonstrating that

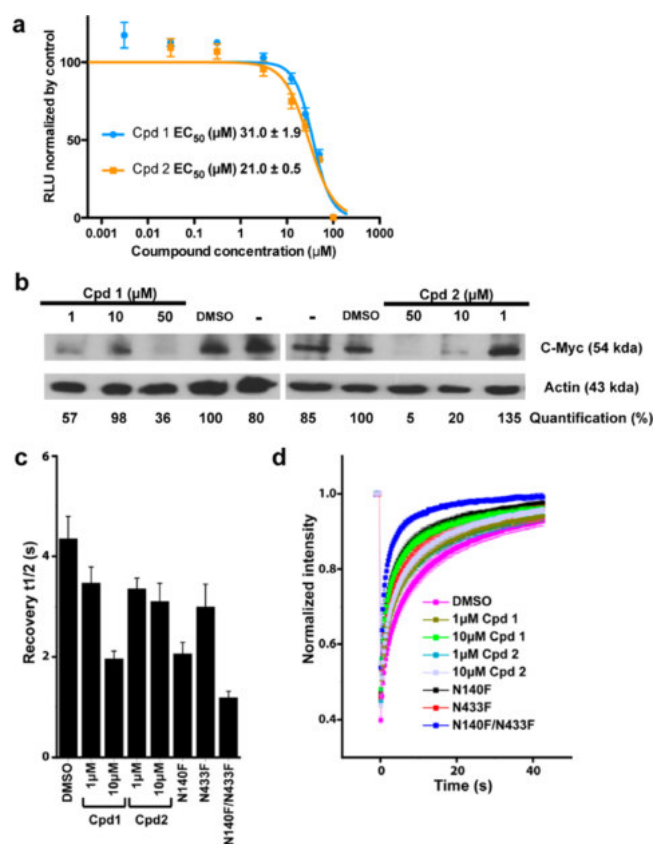


Figure 3. Cell viability and FRAP assays. (a) Viability curves and EC_{50} values (\pm SEM, $N = 3$) of Jurkat cell lines treated with compound 1 (blue curve) or compound 2 (orange curve). (b) Western blot of *c-Myc* expression in Jurkat cells treated with compound 1 or 2 or DMSO as a control. Quantifications were performed using the actin expression level as reference and the DMSO condition corresponding to 100%. (c) FRAP experiments in U2OS cells transfected with GFP-tagged BRD4 and treated with compound 1 or 2 or DMSO as a control. Shown are the half times of recovery of at least 15 cells (left panel) and the time dependence of fluorescence recovery in the bleached area of cells expressing wt or mutant GFP-BRD4 with the corresponding treatment. (d) Curves represent the means at each time point of at least 15 cells in each group.

the identified compounds are indeed orthosteric inhibitors of the targeted interface requires a structural biology approach in which the target is cocrystallized in the presence of the identified hits. However, 3D structures obtained from primary screening are relatively rare (3D structures generally require a first hit to lead optimization to obtain sufficient affinity for the target). Both hits exhibit typical acetyl-lysine mimetic head groups, and our cocrystal structures indeed revealed the canonical acetyl-lysine mimetic binding mode that forms a hydrogen bond with the conserved asparagine (Asn140). Furthermore, direct visualization of an on-target effect in the nucleus of live cells using FRAP experiments emphasized the quality of these hits identified from a primary screening. Quinolone derivatives were recently described as ATAD2 inhibitors³⁷ and have been developed into potent probes targeting BRD9.³⁸ The identified hits may serve as chemical starting points for the development of inhibitors targeting non-BET bromodomains, an area of chemical biology that has not yet been extensively explored.

Overall, we here functionally validated the 2P2I_{3D} chemical library developed by a support vector machine (SVM) approach based on current identified orthosteric inhibitors present in our

2P2I_{DB} database. We demonstrate that the 2P2I_{3D} chemical library is enriched in biologically active PPI modulators that present sufficient basic properties to be proposed for medicinal chemistry programs. Our report thus provides a “proof of concept” for the development of similar approaches applied to larger and more complex libraries that will innovate and accelerate the identification of hits for this challenging class of targets.

METHODS

An extended methods section describing the molecular biology, protein expression, preparation, crystallization, structure determination, cell biology (culture conditions, reagents, cytotoxicity assays and Western blot studies), and biochemical (such as HTRF) assays can be found in the Methods section of the [Supporting Information](#).

ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: [10.1021/acschembio.6b00286](https://doi.org/10.1021/acschembio.6b00286).

Supporting methods, principle of the HTRF assays and characterization of the 2P2I_{3D} chemical database, general screening cascade, overall quality-control (distribution) of the HTRF assay, ITC thermograms for the binding of “thermal shift sensitive” compounds 3–7 to BRD4(1) protein, chemical structures and biophysical characterization of compounds 1 and 2, selectivity of compounds 1 and 2, $2F_o - F_c$ map and omit map around compounds 1 and 2, binding mode of (S)JQ1 compared with compounds 1 and 2, 2D chemical structures of the “thermal shift sensitive” compounds, X-ray data collection and refinement statistics, and HTRF assay pipetting procedures ([PDF](#))

Accession Codes

The coordinates and structure factors of refined BRD4(1) complexed with compounds 1 and 2 have been deposited in the Protein Data Bank with accessions code 5DLZ and 5DLX.

AUTHOR INFORMATION

Corresponding Author

*Xavier Morelli. Phone: +33 (0)486 977 331. E-mail: xavier.morelli@inserm.fr.

Present Address

[○]A.L.: Bioaster; 40 Avenue Tony Garnier 69007 Lyon, France.

Author Contributions

J.-C.G. and X.M. contributed equally to this work as senior authors. J.-C.G. and X.M. designed the experiments. S.M., B.R., S.B., C.D., A.R., E.R., A.L., M.B., R.K., J.-C.L., C.E., and C.R. performed the experiments. S.M., B.R., S.B., C.D., P.R., A.R., M.J. B., E.R., A.L., R.K., J.-C.L., M.E.G., S.C., P.Z., M.A.-L., T.R., S.M., S.K., E.T., Y.C., J.-C.G., and X.M. analyzed and interpreted the data. V.H. and P.R. performed molecular modeling and cheminformatics. The manuscript was written by X.M. with input from all authors.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

We thank B. Canard for access to the facilities at UMR7257 CNRS-AMU, Marseille, M. Choley for technical support in Hybrigenics and V. Depraetere-Ferrier for editing the manuscript. This work was supported by grants from Aix-Marseille

University (A*MIDEX program, No. ANR-11-IDEX-0001-02), the ‘Institut National du Cancer’ - INCa (Program INCa PLBIO Nos. 2012-128 and 2013-105), the Fund for Scientific Research Flanders (FWO, G.0479.12 and G.0846.15), the Belgian Foundation Against Cancer (STK, FA/2014/294), and the concerted action program of the KU Leuven (GOA/12/016). This research also received support from the Innovative Medicines Initiative Joint Undertaking under Grant Agreement No. 115766, the resources of which are composed of financial contributions from the European Union’s Seventh Framework Program (FP7/2007-2013) and EFPIA companies’ kind contribution. R.K. was the recipient of a six-month fellowship from La Ligue French Foundation for Cancer Research. S.M. was supported by the Ph.D. program from the “Région PACA” and a 6-month fellowship from La Fondation pour la Recherche Medicale FRM (FDT20150532618).

REFERENCES

- (1) Morelli, X., and Hupp, T. (2012) Searching for the Holy Grail; protein-protein interaction analysis and modulation. *EMBO Rep.* 13, 877–879.
- (2) Arkin, M. R., Tang, Y., and Wells, J. A. (2014) Small-molecule inhibitors of protein-protein interactions: progressing toward the reality. *Chem. Biol.* 21, 1102–1114.
- (3) Kuenemann, M. A., Sperandio, O., Labbé, C. M., Lagorce, D., Miteva, M. A., and Villoutreix, B. O. (2015) In silico design of low molecular weight protein-protein interaction inhibitors: overall concept and recent advances. *Prog. Biophys. Mol. Biol.* 119, 20–32.
- (4) Villoutreix, B. O., Kuenemann, M. A., Poyet, J., Bruzzoni-Giovanelli, H., Labbé, C., Lagorce, D., Sperandio, O., and Miteva, M. A. (2014) Drug-like protein-protein interaction modulators: challenges and opportunities for Drug discovery and chemical biology. *Mol. Inf.* 33, 414–437.
- (5) Zhang, X., Betzi, S., Morelli, X., and Roche, P. (2014) Focused chemical libraries—design and enrichment: an example of protein-protein interaction chemical space. *Future Med. Chem.* 6, 1291–1307.
- (6) Labbé, C. M., Laconde, G., Kuenemann, M. A., Villoutreix, B. O., and Sperandio, O. (2013) iPPI-DB: a manually curated and interactive database of small non-peptide inhibitors of protein-protein interactions. *Drug Discovery Today* 18, 958–968.
- (7) Labbé, C. M., Kuenemann, M. A., Zarzycka, B., Vriend, G., Nicolaes, G. A., Lagorce, D., Miteva, M. A., Villoutreix, B. O., and Sperandio, O. (2016) iPPI-DB: an online database of modulators of protein-protein interactions. *Nucleic Acids Res.* 44, D542–D547.
- (8) Gaulton, A., Bellis, L. J., Bento, A. P., Chambers, J., Davies, M., Hersey, A., Light, Y., McGlinchey, S., Michalovich, D., Al-Lazikani, B., and Overington, J. P. (2012) ChEMBL: a large-scale bioactivity database for drug discovery. *Nucleic Acids Res.* 40, D1100–D1107.
- (9) Higuieruelo, A. P., Schreyer, A., Bickerton, G. R., Pitt, W. R., Groom, C. R., and Blundell, T. L. (2009) Atomic interactions and profile of small molecules disrupting protein-protein interfaces: the TIMBAL database. *Chem. Biol. Drug Des.* 74, 457–467.
- (10) Higuieruelo, A. P., Jubb, H., and Blundell, T. L. (2013) TIMBAL v2: update of a database holding small molecules modulating protein-protein interactions. *Database* 2013, bat039.
- (11) Gupta, U. C., Bhatia, S., Garg, A., Sharma, A., and Choudhary, V. (2011) *New Drug Development, Drug Discovery and Clinical Research*, pp 1–135, JayPee Brothers Medical Publishers Ltd, New Dehli.
- (12) Arkin, M. R., and Wells, J. A. (2004) Small-molecule inhibitors of protein–protein interactions: progressing towards the dream. *Nat. Rev. Drug Discovery* 3, 301–317.
- (13) Macarron, R., Banks, M. N., Bojanic, D., Burns, D. J., Cirovic, D. A., Garyantes, T., Green, D. V., Hertzberg, R. P., Janzen, W. P., Paslay, J. W., Schopfer, U., and Sittampalam, G. S. (2011) Impact of high-throughput screening in biomedical research. *Nat. Rev. Drug Discovery* 10, 188–195.

- (14) Davies, J. W., Glick, M., and Jenkins, J. L. (2006) Streamlining lead discovery by aligning in silico and high-throughput screening. *Curr. Opin. Chem. Biol.* 10, 343–351.
- (15) Bourgeas, R., Basse, M.-J., Morelli, X., and Roche, P. (2010) Atomic analysis of protein-protein interfaces with known inhibitors: the 2P2I database. *PLoS One* 5, e9598.
- (16) Basse, M. J., Betzi, S., Bourgeas, R., Bouzidi, S., Chetrit, B., Hamon, V., Morelli, X., and Roche, P. (2013) 2P2Idb: a structural database dedicated to orthosteric modulation of protein-protein interactions. *Nucleic Acids Res.* 41, D824–D827.
- (17) Basse, M.-J., Betzi, S., Morelli, X., and Roche, P. (2016) 2P2Idb v2: update of a structural database dedicated to orthosteric modulation of protein-protein interactions. *Database* 2016, baw007.
- (18) Morelli, X., Bourgeas, R., and Roche, P. (2011) Chemical and structural lessons from recent successes in protein-protein interaction inhibition (2P2I). *Curr. Opin. Chem. Biol.* 15, 475–481.
- (19) Hamon, V., Bourgeas, R., Ducrot, P., Theret, I., Xuereb, L., Basse, M. J., Brunel, J. M., Combes, S., Morelli, X., and Roche, P. (2014) 2P2I hunter: a tool for filtering orthosteric protein-protein interaction modulators via a dedicated support vector machine. *J. R. Soc., Interface* 11, 20130860.
- (20) Hamon, V., Brunel, J. M., Combes, S., Basse, M. J., Roche, P., and Morelli, X. (2013) 2P2Ichem: focused chemical libraries dedicated to orthosteric modulation of protein-protein interactions. *MedChemComm* 4, 797–809.
- (21) Horton, D. A., Bourne, G. T., and Smythe, M. L. (2003) The combinatorial synthesis of bicyclic privileged structures or privileged substructures. *Chem. Rev.* 103, 893–930.
- (22) Welsch, M. E., Snyder, S. A., and Stockwell, B. R. (2010) Privileged scaffolds for library design and drug discovery. *Curr. Opin. Chem. Biol.* 14, 347–361.
- (23) Betzi, S., Restouin, A., Opi, S., Arold, S. T., Parrot, I., Guerlesquin, F., Morelli, X., and Collette, Y. (2007) Protein protein interaction inhibition (2P2I) combining high throughput and virtual screening: application to the HIV-1 Nef protein. *Proc. Natl. Acad. Sci. U. S. A.* 104, 19256–19261.
- (24) Emert-Sedlak, L., Kodama, T., Lerner, E. C., Dai, W., Foster, C., Day, B. W., Lazo, J. S., and Smithgall, T. E. (2009) Chemical library screens targeting an HIV-1 accessory factor/host cell kinase complex identify novel antiretroviral compounds. *ACS Chem. Biol.* 4, 939–947.
- (25) Breuer, S., Espinola, S., Morelli, X., Torbett, B. E., Arold, S. T., and Engels, I. H. (2013) A biochemical/biophysical assay dyad for HTS-Compatible triaging of inhibitors of the HIV-1 Nef/Hck SH3 interaction. *Curr. Chem. Genomics Transl. Med.* 7, 16–20.
- (26) Filippakopoulos, P., and Knapp, S. (2014) Targeting bromodomains: epigenetic readers of lysine acetylation. *Nat. Rev. Drug Discovery* 13, 337–356.
- (27) Brand, M., Measures, A. M., Wilson, B. G., Cortopassi, W. A., Alexander, R., Höss, M., Hewings, D. S., Rooney, T. P., Paton, R. S., and Conway, S. J. (2015) Small molecule inhibitors of bromodomain-acetyl-lysine interactions. *ACS Chem. Biol.* 10, 22–39.
- (28) Chen, X., Longgood, J. C., Michnoff, C., Wei, S., Frantz, D. E., and Bezprozvanny, L. (2007) High-throughput screen for small molecule inhibitors of Mint1-PDZ domains. *Assay Drug Dev. Technol.* 5, 769–783.
- (29) S. Thorsen, T., L. Madsen, K. L., Dyhring, T., Bach, A., Peters, D., Strömgaard, K., Rønne, L. C., and Gether, U. (2011) A fluorescence polarization based screening assay for identification of small molecule inhibitors of the PICK1 PDZ domain. *Comb. Chem. High Throughput Screening* 14, 590–600.
- (30) Degorce, F., Card, A., Soh, S., Trinquet, E., Knapik, G. P., and Xie, B. (2009) HTRF: A technology tailored for drug discovery - a review of theoretical aspects and recent applications. *Curr. Chem. Genomics* 3, 22–32.
- (31) Zhang, J.-H., Chung, T. D., and Oldenburg, K. R. (1999) A simple statistical parameter for use in evaluation and validation of high throughput screening assays. *J. Biomol. Screening* 4, 67–73.
- (32) Ryan, A. J., Gray, N. M., Lowe, P. N., and Chung, C. W. (2003) Effect of detergent on “promiscuous” inhibitors. *J. Med. Chem.* 46, 3448–3451.
- (33) Feng, B. Y., and Shoichet, B. K. (2006) A detergent-based assay for the detection of promiscuous inhibitors. *Nat. Protoc.* 1, 550–553.
- (34) Filippakopoulos, P., Picaud, S., Mangos, M., Keates, T., Lambert, J.-P., Barsyte-Lovejoy, D., Felletar, I., Volkmer, R., Müller, S., Pawson, T., Gingras, A.-C., Arrowsmith, C. H., and Knapp, S. (2012) Histone recognition and large-scale structural analysis of the human bromodomain family. *Cell* 149, 214–231.
- (35) Rahl, P. B., Lin, C. Y., Seila, A. C., Flynn, R. A., McCuine, S., Burge, C. B., Sharp, P. A., and Young, R. A. (2010) c-Myc regulates transcriptional pause release. *Cell* 141, 432–445.
- (36) Lipinski, C. A., Lombardo, F., Dominy, B. W., and Feeney, P. J. (2001) Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv. Drug Delivery Rev.* 46, 3–26.
- (37) Chaikuad, A., Petros, A. M., Fedorov, O., Xu, J., and Knapp, S. (2014) Structure-based approaches towards identification of fragments for the low-druggability ATAD2 bromodomain. *MedChemComm* 5, 1843–1848.
- (38) Clark, P. G., Vieira, L. C., Tallant, C., Fedorov, O., Singleton, D. C., Rogers, C. M., Monteiro, O. P., Bennett, J. M., Baronio, R., Müller, S., Daniels, D. L., Méndez, J., Knapp, S., Brennan, P. E., and Dixon, D. J. (2015) LP99: discovery and synthesis of the first selective BRD7/9 Bromodomain inhibitor. *Angew. Chem., Int. Ed.* 54, 6217–6221.
- (39) Zou, G., et al. (2011) Functional analysis of two cavities in flavivirus NS5 polymerase. *J. Biol. Chem.* 286 (16), 14362–14372.
- (40) Arold, S., et al. (1998) RT loop flexibility enhances the specificity of Src family SH3 domains for HIV-1 Nef. *Biochemistry* 37 (42), 14683–14691.
- (41) Filippakopoulos, P., et al. (2012) Histone recognition and large-scale structural analysis of the human bromodomain family. *Cell* 149 (1), 214–231.
- (42) Baietti, M. F., et al. (2012) Syndecan-syntenin-ALIX regulates the biogenesis of exosomes. *Nat. Cell Biol.* 14 (7), 677–685.

Supplementary Information - METHODS

Library

The library screened contained 1664 compounds, arrayed in 96-well plates (Thermo Scientific™ Nunc™ 96-Well Polystyrene Conical Bottom) as single compounds at 10 mM in DMSO, sealed and stored at -20 °C (for more details, refer to Hamon *et al.* (1)). The quality of all compounds was assured by the vendors as greater than 90% pure with provided quality control data.

HTRF Setup Assays

Titration curves were optimized, keeping donor and donor-coupled proteins at constant concentration and increasing acceptor/acceptor-coupled proteins concentrations to avoid the 'hook effect' (2). Screening conditions were selected at 80% of the maximum observed signal when assessing the titration curve. This protocol ensures the optimization of a sufficient signal to noise ratio to perform HTS with adequate Z'. This optimization is crucial to achieve optimal sensitivity required for the detection of even weak small-molecule inhibitors of each protein-protein interaction. An estimated minimum of 2 to 3 value for signal to noise is necessary to ensure reliable HTS. Such signal to noise ratio was observed for each PPI complex at 80% of the maximum measured signal. The estimated time to reach equilibrium, the period of stability of the measured signal and the DMSO tolerance were evaluated to ensure stability of the signal during HTS (data not shown).

HTRF screen

HTRF assays were performed in white 384 Well Small Volume™ HiBase Polystyrene Microplates (Greiner) with a total working volume of 20 µL. Compounds were dispensed, with either 400 nL per well (2% final DMSO), from a concentration stock of 1mM in 100% DMSO for the primary screen, or with serial DMSO dilutions for secondary screening, counter screen and IC₅₀ measurement assays, using a Mosquito Crystal pipetting robot platform (TTP labtech). For bromodomain selectivity assays, 200 nL of compounds were dispensed to obtain a 1% final concentration of DMSO. Primary screening assays have been performed in monoplicate while secondary screenings, counter screens and IC₅₀ measurements were performed in triplicates.

All HTRF reagents were purchased from CisBio Bioassays and reconstituted according to the supplier protocols. For each assay 14.7 µL of mix1 (Table S3) is added in the assay wells, containing previously dispensed inhibitors, according to the final concentration and buffer described in Table S3, using a Biomek NX MC pipetting robot (Beckman). After 1 h incubation, 4.9 µL of protein B is added. HTRF signals were measured, after a final incubation, using a PHERAstar FS (BMG Labtech) with an excitation filter at 337 nm and

fluorescence wavelength measurement at 620 and 665 nm, an integration delay of 60 μ s and an integration time of 400 μ s. Results were analyzed with a two-wavelengths signal ratio: [intensity (665 nm)/intensity (620 nm)]*10⁴. Percentage of inhibition was calculated using the following equation: % inhibition = [(compound signal) - (min signal)] / [(max signal) - (min signal)] * 100, where 'max signal' is the signal ratio with the compound vehicle alone (DMSO) and 'min signal' the signal ratio without protein B (Table S3). The Z' factor is calculated using the following equation: Z' = 1 - [3(SD of max) + 3(min SD)] / [(mean max signal) - (mean min signal)], where SD is the standard deviation. For IC₅₀ measurements, values were normalized and fitted with Prism (GraphPad software) using the following equation: $Y = 100 / (1 + ((X / IC_{50})^{Hill\ slope}))$.

Autofluorescence experiment

Autofluorescence experiments were performed at 20 μ M final inhibitor concentration with 2% final concentration of DMSO. 400 nL of compounds were dispensed in 384 well plates as previously described, and 19.6 μ L of buffer (Hepes 50 mM pH 7.5, 150 mM NaCl, 0.1% BSA (Sigma)) were then added. After overnight incubation at 4 °C HTRF signals were measured. Analyses were performed directly comparing the 620 and 665 nm fluorescence values with the fluorescence values at the same wavelength of the buffer + vehicle (DMSO). Thresholds have been set up at 10 times the control value measured at 665 nm and 50 times the control value at 620 nm.

SNAP-tag® labeling

The HTRF assay for the NS3/NS5 complex has been set-up using a SNAP-tag construct. This 20 kDa tag is based on a family of enzymes called « suicide enzymes » as they react covalently with their substrate which blocks the enzyme and prevents any subsequent enzymatic activity (3). SNAP-tag is a mutant version of the DNA repair protein O⁶-alkylguanine-DNA alkyltransferase that reacts specifically and rapidly with benzylguanine (BG) derivatives, leading to an irreversible covalent bond. A large variety of BG derivatives have been engineered to label proteins of interest with fluorophores compatible with *in vivo* imaging, microscopy or even immobilization on a specific support. We used one of these derivatives, 'BG Red', as acceptor for our HTRF assay. This type of labeling presents a main advantage compared to « classic » labeling since it does not require secondary labeled-antibodies to label the protein of interest with a HTRF compatible fluorophore.

SNAP-tag tagged protein were incubated at a final concentration of 20 μ M in presence of 3 times excess of Tag-lite ® SNAP Red (CisBio Bioassays) in Hepes 50 mM pH 7.5, 150 mM NaCl overnight at 4 °C in the dark in low absorbent tubes. Labeled protein and unreacted dye were then separated using Illustra NAP-5 Columns (GE healthcare). Protein concentration and labeling efficiency was evaluated by measuring protein and dye absorbance,

respectively at 280 and 650 nm using a NanoDrop 2000 (Thermo scientific). Proteins were stored at -80 °C after addition of BSA at a final concentration of 0.1%.

Molecular biology

Dengue Virus Proteins

Dengue virus proteins were amplified from a synthetic gene (GeneArt, Life technologies) of dengue serotype 3 of Singapore. Non-structural protein 3 (NS3) fused, through G4-S-G4 linker, on the N-terminal part with hydrophilic core of 18 amino acid (43-60) of NS2B protein (NS3-NS2B) of dengue virus serotype 3 (DV3), amplified from the synthetic gene (forward primer :

GGGGACAAGTTTGTACAAAAAAGCAGGCTTAGAAAACCTGTACTTCCAGGGTGCAGATC
TGACCGTTGAAAAGC, reverse primer:

GGGGACCACTTTGTACAAGAAAGCTGGGTCTATTACCAAATTGCGCCTTCGCTTTC),

have been cloned in a pETG-30A (Arie Geerlof, EMBL, Hamburg, Germany) by Gateway® recombination (Life technologies) for GST fusion. For the cloning of methyltransferase domain of non-structural protein 5 (MTase NS5) we used a modified Gateway® pDEST™ 17 Vector (Life technologies) (insertion of an XhoI restriction site between His tag and attR1 site by site directed mutagenesis). SNAP-tag previously amplified from pSNAP-tag® (T7)-2 Vector (New England Biolabs) (forward primer: GGG GCT CGA GTT ATG GAC AAA GAT TGC GAA ATG AAA C, reverse primer: ATA TAC TCG AGT CCC AGA CCC GGT TTA CCC AG) was inserted by restriction ligation in XhoI site. The coding sequence of MTaseNS5 DV3 amplified (forward primer:

GGGGACAAGTTTGTACAAAAAAGCAGGCTTAGAAAACCTGTACTTCCAGGGTGGTACAG
GTAGCCAGGGTG, reverse primer:

GGGGACCACTTTGTACAAGAAAGCTGGGTCTATTAATGACGGGTGCCTGCACCCAG)

from the synthetic gene was then sub cloned in this vector by Gateway® recombination (Life technologies) resulting in a MTase NS5 fused with His and SNAP-tag on the N-terminal part.

GRASP55 PDZ domains

To generate GRASP55 expression constructs, the full-length mouse GORASP2 cDNA encoding the mGRASP55 protein was amplified by polymerase chain reaction (PCR) using the oligonucleotides CTCGAGATGGGCTCCTCGCAGAGC and GGATCCCCAGAAGGCTCTGAAGCATCTGC.

The amplification product was cloned in the pGEM - T Easy vector (Promega). Then, this mouse GORASP2 cDNA was used as template for PCR amplification of the open reading frame (ORF) for mGRASP55 PDZ12 (aa 2 - 208) and mGRASP55 Full - length (FL) (aa 2 - 451) using forward and reverse oligonucleotides flanked by attb1 and attb2 recombination sites. The following primers pairs were used:

the forward primer was: GGGGACAAGTTTGTACAAAAAAGCAGGCTTCGGCTCCTCGCAGAGCGTCGAGAT and the reverse primers were : GGGGACCACTTTGTACAAGAAAGCTGGGTCTTATTCAAAGGGGCGTGTAGGTATTCGGT GCA and GGGGACCACTTTGTACAAGAAAGCTGGGTCTTAAGAAGGCTCTGAAGCATCTGCATCAG AC for mGRASP55 PDZ12 and mGrasp55 FL respectively. The amplicons were cloned by the BP reaction into pDONRZeo to produce the corresponding entry vectors. The coding sequences were transferred by LR cloning in pDESTTM15 prokaryotic expression vectors intended to produce the corresponding Nterminal GST-tagged fusion protein.

Syntenin PDZ domains

Human syntenin1 cDNA (4) was cloned in 3' into the pGEX-5X vector (Amersham Pharmacia Biotech AB) for the expression of recombinant GST-syntenin1 full length.

Protein expression and purification

Dengue Virus Proteins

The dengue virus proteins were expressed in T7 Express Competent *E. coli* (NEB). Cells were grown in LB medium (Euromedex) at 37 °C until OD₆₀₀ = 0.6 was reached and then induced using 0.5 mM isopropylβD-thiogalactopyranoside (IPTG) and cultured overnight at 17 °C for NS2B-NS3-GST and 5 hrs at 25 °C for MTaseNS5 SNAP-tag fusion. Cells were pelleted by centrifugation (10 min at 7,000g), and pellets were re-suspended in 20 mM HEPES (pH 7.5), 500 mM NaCl, 10% glycerol, 5 mM Imidazole, 2 mM 2-Mercaptoethanol, 1 mg/mL lysozyme, 22 µg/mL DNase I and 0.2 mM of benzamidine. After sonication and centrifugation (30 min at 45,000g), the supernatant was purified on HisPurTM Cobalt or GST agarose resin (Thermo Scientific). Elution fractions containing protein were then loaded onto a Superdex 75 or 200 gel filtration column (GE Healthcare) and eluted with 50 mM Hepes (pH 7.5), 300 mM NaCl, 1 mM DTT. Purified proteins were further concentrated and stored at -80 °C.

GRASP55 PDZ domains

The production of N-terminal GST-tagged GRASP fusion proteins were accomplished by induction for 3 h at 37 °C or 18 h at 25 °C with 0.2 mM IPTG in *E. coli* BL21 (DE3) bacteria cells transformed with the purified pDESTTM15 GRASP expression plasmids. Fusion proteins were recovered from the cell lysates by conventional affinity chromatography on Glutathione resin.

Syntenin PDZ domain

E. coli ER2566 cells were transformed with the human syntenin1 pGEX-5X expression vector.

Expression of N-terminally GST tagged syntenin 1 full length was induced at 30 °C and by the addition of IPTG (Fisher Scientific). Protein was purified using GSTrap4B columns 28-4017-45 (GE Healthcare) using an Akta Explorer system (GE Healthcare).

Nef protein & SH3 domain from Hck

For the Hck SH3 domain, the prokaryotic expression vector and the GST affinity chromatography purification protocol have been previously described in (5).

For the HIV-1 Nef WT protein, the prokaryotic expression vector and the His affinity chromatography purification protocol have been kindly provided by Dr. Sebastian Breuer as described in (6). For both Hck-SH3 and HIV-1 Nef WT, protein peak fractions were further purified using a Superdex 75 10/300 GL column (GE Healthcare) in 20 mM TRIS pH 8.0, 150 mM NaCl Buffer.

BRD4

For isothermal titration Calorimetry (ITC), thermal shift assay (TSA) and crystallogenesi, BRD4(1) was produced and purified using a histidine tag affinity chromatography as described by Filipakopoulos *et al.* (7). For these experiments, a pNIC28-BSA4 expression vector containing BRD4(1) and a Tobacco Etch Virus (TEV) protease cleavage site have kindly been provided by Stefan Knapp laboratory from the SGC at the University of Oxford. After size exclusion chromatography, the fractions presenting pure BRD4 after TEV cleavage of the histidine tag were pooled and concentrated to 25 mg/mL for crystallogenesi. For ITC assays, the protein was concentrated up to at 6 mg/mL and the DTT was removed using a buffer exchange column (PD10 from GE healthcare) equilibrated with 10 mM HEPES pH 7.5, 150 mM NaCl. For HTRF experiments, a BRD4(1) synthetic gene that includes a TEV cleavage site was purchased from LifeTechnology in a pDONR transport vector before cloning into a pDESTTM15 expression vector for GST affinity purification. Protein production and purification was carried out using similar protocols and buffers as used for the His-BRD4(1) system. Purification was carried on GST affinity resin (Thermo Scientific) and reduced glutathione was used for protein release. GST-BRD4(1) was further purified by size exclusion chromatography on a Superdex 16/60 Hiload column (GE Healthcare) using 20 mM TRIS pH 8.0, 150 mM NaCl Buffer.

Bromodomain selectivity Profiles

Selectivity profiles of bromodomain inhibitors were performed as described in HTRF screen and HTRF Setup Assays sections. Concentration of histone peptide was optimized to ensure sufficient signal to noise ratio, sufficient sensitivity for detection of weak inhibitors and comparable data from one bromodomain to another. HTRF detection reagents

(EPIgeneous™ Binding Domain kits) were purchased from Cisbio Bioassays and used according to supplier's protocol. GST tagged bromodomain proteins were purchased from BPS Bioscience and histone peptide from Anaspec.

Fusion protein, counter screen assays

GST-biotinylation: Purified Glutathione S-transferase was biotinylated with Sulfo-NHS-LC-biotin (ThermoFischer) at a 10 : 1 biotin : protein molar ratio in Phosphate Buffer Saline pH 8.3 for 40 min on ice followed by quenching with 100 mM glycine pH 8.3 and removal of free biotin by desalting on PD10 column.

GST-SNAP: GST-Snap fusion protein was provided as custom reagent from Cisbio Bioassays and labeled with BG-Red as previously described.

GST-His: GST-His fusion protein has been produced via Gateway plasmid pETG30A (Arie Geerlof, EMBL, Hamburg, Germany) in *E. coli* as previously described for dengue virus protein production.

Cells and Cell Culture

The human leukemia cell line Jurkat (ATCC® TIB-152) was maintained in RPMI-1640 medium supplemented with 10% FBS at 37 °C and 5% CO₂. Human osteosarcoma cell line (U2OS, ATCC® HTB-96™) was maintained in DMEM supplemented with 10% FBS at 37 °C and 5% CO₂.

Western blot

Non-treated Jurkat cells and Jurkat cells treated during 24 hours with compound #1, compound #2 or DMSO were lysed in RIPA lysis buffer (Tris HCl pH 7.5 50 mM, NaCl 150 mM, Triton 1%, SDS 0.1%, sodium deoxycholate 1%) supplemented with protease inhibitor cocktail (P8340, Sigma-Aldrich) and 1 mM PhenylMethylSulfonyl Fluoride (Sigma-Aldrich). 50 µg of protein were loaded onto 10% acrylamide SDS/PAGE and then transferred to nitrocellulose Hybond C-extra membranes, 45 micron (GE Healthcare). The membranes were saturated with 5% (wt/vol) skim milk in TBST [Tris-buffered saline/0.1% (vol/vol) Tween 20] 1 hour at room temperature and incubated with anti-myc antibody (clone 9E10, sc40 Santa Cruz) at a 1 : 500 dilution in 0.5% (wt/vol) skimmed milk in TBST overnight at 4 °C. Membranes were then washed with TBST, incubated with an HRP-conjugated anti-mouse secondary antibody (polyclonal goat anti mouse P0447, Pierce) at a 1 : 20,000 dilution in TBST 1 hour at room temperature. Immunoreactive bands were revealed using SuperSignal™ West Dura Extended Duration Substrate (Pierce) detection reagents. Quantification was performed using ImageJ software.

Fluorescence Recovery After Photobleaching (FRAP)

FRAP studies were performed essentially as described (8). In brief, U2OS cells were transfected (Fugene HD; Roche) with mammalian over-expression constructs encoding GFP fused to the N-terminus of human full-length BRD4. The FRAP and imaging system consisted of a Zeiss LSM 710 laser-scanning and control system (Zeiss) coupled to an inverted Zeiss Axio Observer.Z1 microscope equipped with a high-numerical-aperture (N. A. 1.3) 40 x oil immersion objective (Zeiss). Samples were placed in an incubator chamber capable of maintaining temperature and humidity. FRAP and GFP fluorescence imaging were both carried out with an argon-ion laser (488 nm) and with a PMT detector set to detect fluorescence between 500-550 nm. Once an initial scan had been taken, a region of interest corresponding to approximately 50% of the entire GFP positive nucleus was empirically selected for bleaching. A time lapse series was then taken to record GFP recovery using 1% of the power used for bleaching. The image datasets and fluorescence recovery data were exported from ZEN 2009, the microscope control software, into Microsoft Excel to determine the average half-time for full recovery for 10-20 cells per treatment point. Where appropriate, inhibitor was added 1 hour before imaging, which was carried out 24 hours after transfection.

Isothermal titration calorimetry

ITC was used to evaluate the thermodynamics parameters of the binding between BRD4(1) and the selected compounds, using ITC conditions previously described by Filippakopoulos *et al.* (7). Purified BRD4(1) was extensively dialyzed in the ITC buffer containing 10 mM HEPES pH 7.5 and 150 mM NaCl. Compounds were diluted directly in the last protein dialysate prior to experiments. Titrations were carried out on a MicroCal ITC200 microcalorimeter (GE Healthcare, Piscataway, NJ). Each experiment was designed using a titrant concentration (protein in the syringe) set 10 to 15 times the analyte concentration (compound in the cell generally between 10 and 30 μ M) and generally using 13 to 17 injections at 15 °C. A first small injection (generally 0.2 μ L) was included in the titration protocol in order to remove air bubbles trapped in the syringe prior titration. Raw data were scaled to zero by subtracting a negative control experiment (buffer injected in the cell containing the same compound concentration) or after setting the zero to the titration saturation heat value when possible. Integrated raw ITC data were fitted to a one site non-linear least squares fit model using the MicroCal Origin plugin as implemented in Origin 9.1 (Origin Lab). Finally, ΔG and $T\Delta S$ values were calculated from the fitted ΔH and K_A values using the equations $\Delta G = -R.T.\ln K_A$ and $\Delta G = \Delta H - T\Delta S$.

Differential scanning fluorimetry

Differential scanning fluorimetry (DSF) or Thermal Shift Assay (TSA) was performed as

described previously (9). A protein/Sypro® orange mix containing 2 μM BRD4(1) and a 1 : 1,000 dilution of dye in DMSO (as supplied by Life Technologies) was prepared just before plate setup in 50 mM HEPES pH 7.5, 200 mM NaCl. 19.5 μL of the protein/Sypro orange mix was aliquoted into a 96-well plate and 0.5 μL of compounds (2 mM 100% DMSO, final concentration of 50 μM 2.5% DMSO) or DMSO for control were dispensed. For experiment performed with DTT and tween, 5mM of DTT or 0.012% (0.3 CMC) were added to previously described buffer. After sealing with optical tape, the plates were centrifuged at 1,000 rpm for 1 min at room temperature. Each experiment was performed in triplicate. Thermal melting experiments were carried out using a CFX96 Real Time PCR machine (Biorad). The plates were first equilibrated at 25 °C for 2 min in the PCR machine before starting the thermal melting experiment. The plates were heated by 1 °C per 60 sec from 25 to 95 °C and the fluorescence intensity was recorded at each temperature step using the built-in FRET filter. Raw fluorescence data were treated using Excel files adapted from (9) and T_m measured using a Boltzmann fit equation from GraphPad Prism software.

Crystallography

BRD4(1)-Inhibitor co-crystallization was performed at 19 °C (292K) using the hanging drop vapor diffusion method. For the complex with 1 mM of compound #2, 12 mg/mL of BRD4(1) preparation was mixed at a 1 : 1 ratio with the precipitant solution (0.3 M Sodium Formate, 0.1 M NaCl, 22% (w/v) PEG 3350, 10% (w/v) ethylene glycol) and crystals grew to diffracting quality within 5-7 days. Co-crystallization experiments with 5 mM of compound #1 were carried using the same protocol (Sodium Formate was replaced by 0.3 M NaNO_3) and crystals grew to diffracting quality within 3 weeks. Crystals were cryo-protected using the precipitant solution supplemented with 10% glycerol and were flash frozen in liquid nitrogen. Data for compound #1 and compound #2 were respectively collected at the ESRF beamlines ID23-2 and ID30A-3. Indexing, integration and scaling were performed using XDS (10). Initial phases were calculated by molecular replacement with Phaser MR (CCP4 suite) (11) using a model of the first domain of BRD4 (extracted from the Protein Data Bank accession code: 2OSS). Initial models for the protein and the ligands were built in COOT (12). The cycles of refinement were carried out with Refmac5 (CCP4 suite (11)). Data collection and refinement statistics can be found in Table S2. The models and structure factors have been deposited with Protein Data Bank accession code for compound #1: 5DLZ and compound #2: 5DLX.

“PAINS” and Aggregators predictions

Pan-assay interference compounds (“PAINS”) predictions are indicated based on FAFDrugs³ (<http://fafdrugs3.mti.univ-paris-diderot.fr/index.html>) (13–15). Default parameters and “Filter Pan Assay Interference Compounds (PAINS) Filter A” have been chosen for analysis.

Aggregator predictions are indicated based on Aggregator advisor website (<http://advisor.bkslab.org/>) (16).

ACCESSION CODES

Coordinates and structure factors of refined BRD4(1) complexed with compound 1 and 2 have been deposited in the Protein Data Bank with accession code 5DLZ and 5DLX.

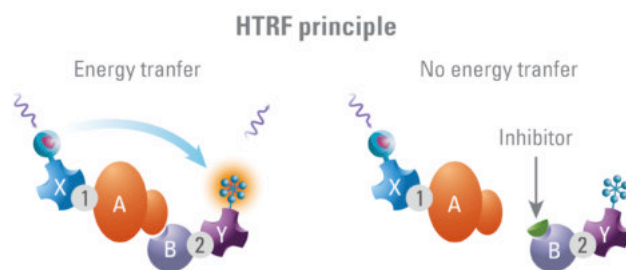
SI References

1. Hamon V, et al. (2013) 2P2Ichem: focused chemical libraries dedicated to orthosteric modulation of protein-protein interactions. *Med Chem Commun* (4):797–809.
2. Arkin MR, Glicksman MA, Fu H, Havel JJ, Du Y (2004) Inhibition of Protein-Protein Interactions: Non-Cellular Assay Formats. *Assay Guidance Manual*, eds Sittampalam GS, et al. (Eli Lilly & Company and the National Center for Advancing Translational Sciences, Bethesda (MD)). Available at: <http://www.ncbi.nlm.nih.gov/books/NBK92000/> [Accessed September 15, 2015].
3. Keppler A, et al. (2003) A general method for the covalent labeling of fusion proteins with small molecules in vivo. *Nat Biotechnol* 21(1):86–89.
4. Zimmermann P, et al. (2001) Characterization of syntenin, a syndecan-binding PDZ protein, as a component of cell adhesion sites and microfilaments. *Mol Biol Cell* 12(2):339–350.
5. Arold S, et al. (1997) The crystal structure of HIV-1 Nef protein bound to the Fyn kinase SH3 domain suggests a role for this complex in altered T cell receptor signaling. *Struct Lond Engl* 1993 5(10):1361–1372.
6. Breuer S, et al. (2006) Biochemical indication for myristoylation-dependent conformational changes in HIV-1 Nef. *Biochemistry (Mosc)* 45(7):2339–2349.
7. Filippakopoulos P, et al. (2010) Selective inhibition of BET bromodomains. *Nature* 468(7327):1067–1073.
8. Philpott M, et al. (2014) Assessing cellular efficacy of bromodomain inhibitors using fluorescence recovery after photobleaching. *Epigenetics Chromatin* 7:14.
9. Niesen FH, Berglund H, Vedadi M (2007) The use of differential scanning fluorimetry to detect ligand interactions that promote protein stability. *Nat Protoc* 2(9):2212–2221.
10. Kabsch W (2010) XDS. *Acta Crystallogr D Biol Crystallogr* 66(Pt 2):125–132.
11. Collaborative Computational Project, Number 4 (1994) The CCP4 suite: programs for protein crystallography. *Acta Crystallogr D Biol Crystallogr* 50(Pt 5):760–763.
12. Emsley P, Lohkamp B, Scott WG, Cowtan K (2010) Features and development of Coot. *Acta Crystallogr D Biol Crystallogr* 66(Pt 4):486–501.
13. Miteva MA, et al. (2006) FAF-Drugs: free ADME/tox filtering of compound collections. *Nucleic Acids Res* 34(Web Server issue):W738–744.
14. Lagorce D, et al. (2011) The FAF-Drugs2 server: a multistep engine to prepare electronic chemical compound collections. *Bioinforma Oxf Engl* 27(14):2018–2020.
15. Lagorce D, Sperandio O, Baell JB, Miteva MA, Villoutreix BO (2015) FAF-Drugs3: a web server for compound property calculation and chemical library design. *Nucleic Acids Res* 43(W1):W200–207.

16. Irwin JJ, et al. (2015) An Aggregation Advisor for Ligand Discovery. *J Med Chem* 58(17):7076–7087.

Supplementary Information - Figures

a



b

Protein/Protein Complex	Interface type	Kd (μM)	Protein 1 MW (Including tag, kDa)	Proteins 2 MW (Including tag, kDa)	Signal to noise	Tag 1	Tags 2	HTRF combinaison donnor/ acceptor	Length and temperature of incubation	Z'
NS3/NS5 dengue virus	Protein/protein	1.0	100	50	6	GST	SNAP Tag	Tb-Bg Red	4h RT	0.70
Nef /SH3-Hck (HIV virus)	Protein/domain	1.5	35	33	44	GST	His	KEu-d2	ON 4°C	0.88
Bromodomain/H4 (K ₅₂ S ₉₁ /12/16)	Protein/peptide	2.8	42	3	18	GST	Biotin	KEu-d2	4h RT	0.78
Grasp55/ JamB (PDZ domain)	Protein/peptide	1.0	48	3	26	GST	Biotin	Tb- d2	ON 4°C	0.81
Syntenin/ Syndecan (PDZ domain)	Protein peptide	1.8	60	2	22	GST	Biotin	Tb- d2	ON 4°C	0.92

c

HTRF Autofluorescence (%)	
665nm (10x blank buffer value)	620nm (50x blank buffer value)
3.18%	2.34%

Figure S1: Principle of the HTRF assays and characterization of the 2P2I_{3D} chemical database. a) HTRF assay principle, b) HTRF assays parameters, c) Summary of the autofluorescence experiments. Compounds have been considered as autofluorescent, when fluorescence values were respectively higher than 10 and 50 times the values of the buffer alone at 665 and 620 nm.

General Screening Cascade

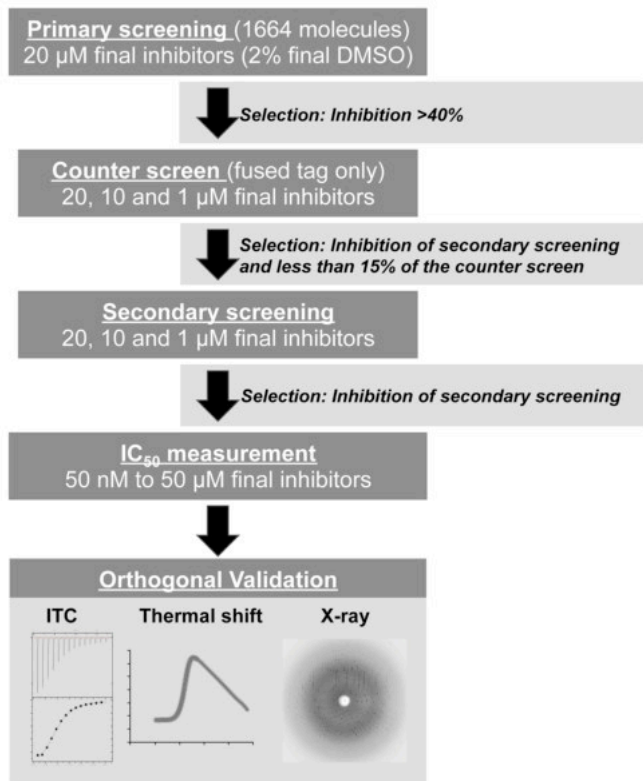


Figure S2: General screening cascade. Primary screenings were performed at 20 μM final concentration of each compounds. We selected compounds with a percentage of inhibition higher than 40% for the second step (secondary screenings and counter screens). We selected for the subsequent IC₅₀ measurements, only compounds inhibiting the secondary screening and less than 15% of the counter screen signal. Compounds with suitable dose response curves were validated by orthogonal experiments (isothermal titration calorimetry, thermal shift assays and X-ray crystallography).

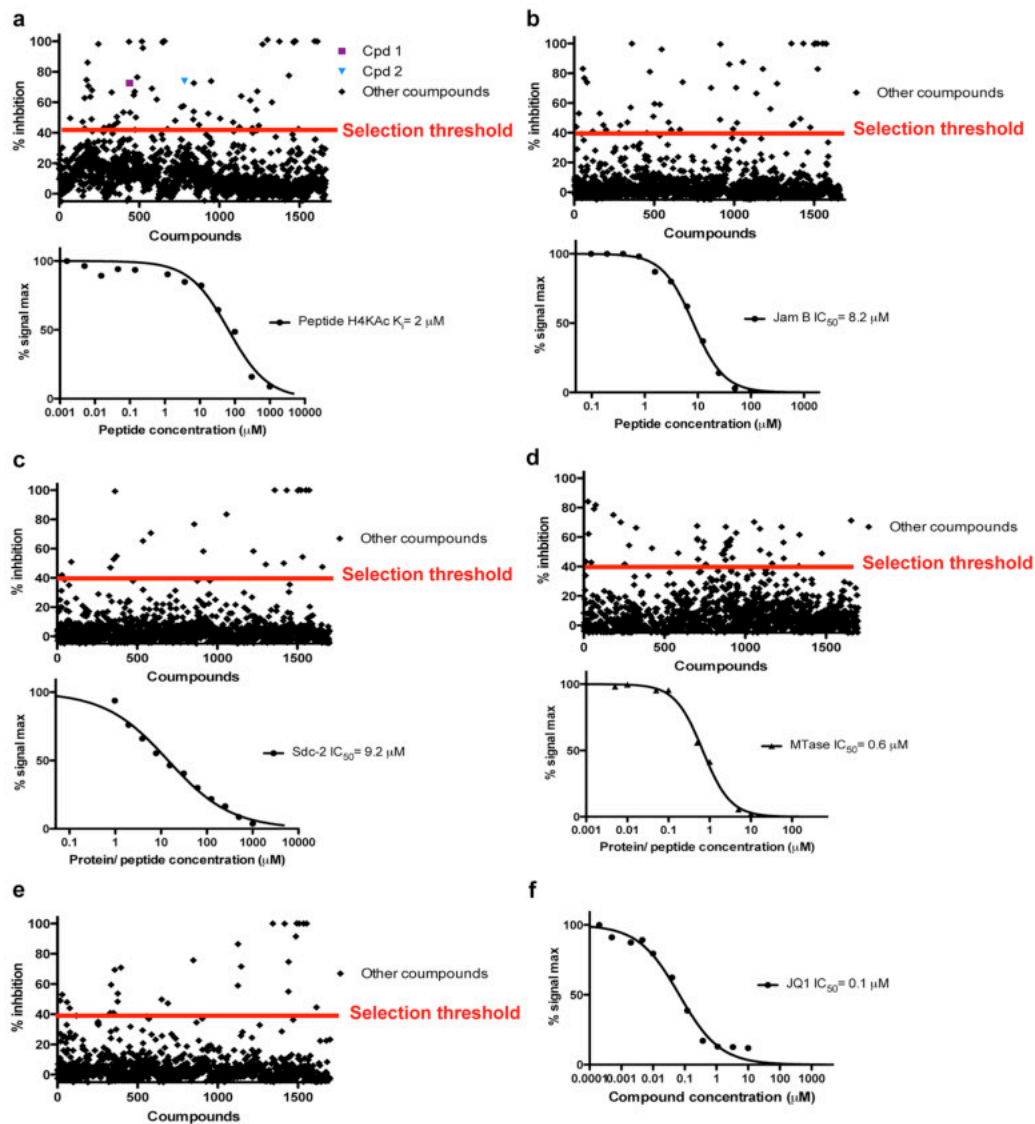


Figure S3: Overall quality-control (distribution) of the HTRF assay. Distribution of the percentage of inhibition values obtained during the primary screenings and competition curve obtained with untagged protein of peptide of HTRF assay.

a) Bromodomain screening, Compound #1 is highlighted in purple and compound #2 in blue. b) Grasp screening, c) Syntenin screening, d) Dengue screening and e) Nef screening, competition curves are not available with HCK peptide. f) competition curve of BRD4/H4 peptide assay with (+) JQ1 compound.

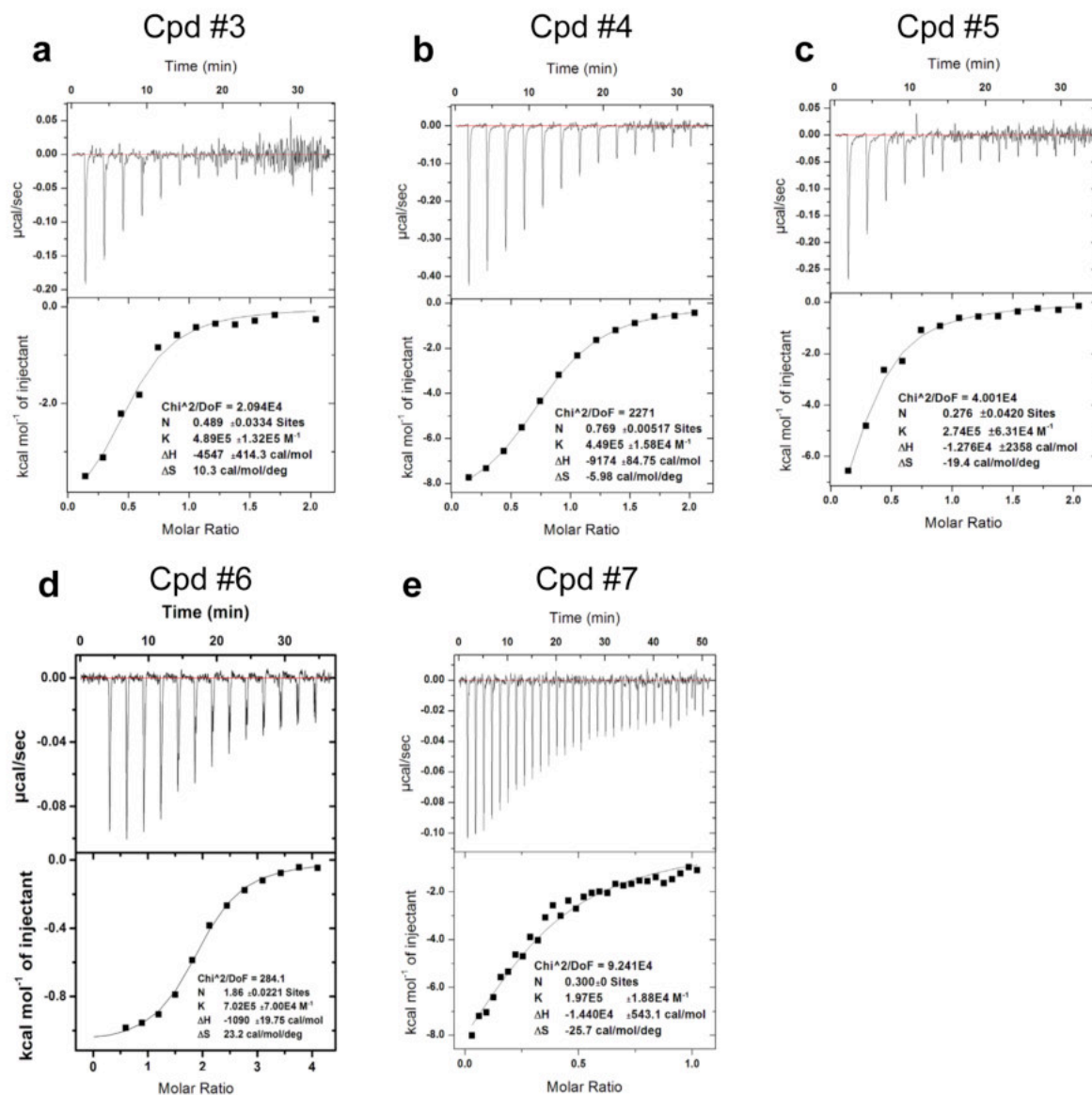


Figure S4: ITC thermograms for the binding of “thermal shift sensitive” compounds #3 (a) to #7 (e) to BRD4(1) protein

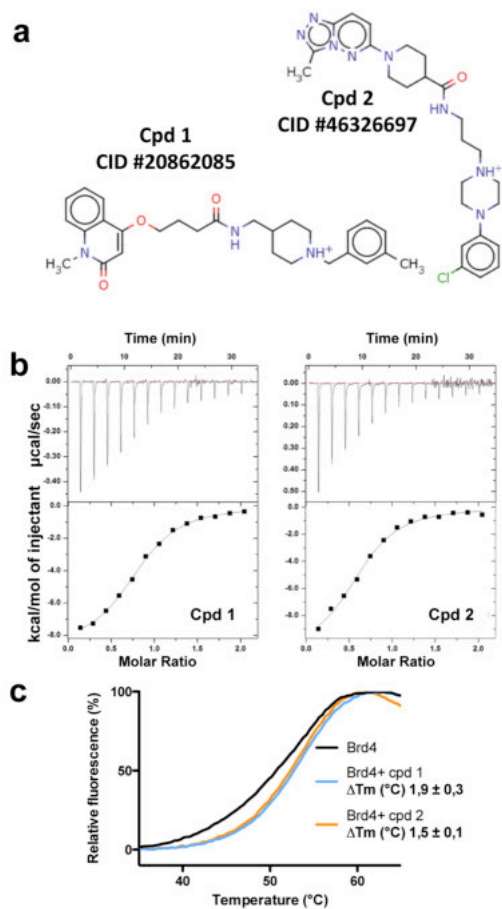


Figure S5: Chemical structures and biophysical characterization of compounds #1 and #2. a) Chemical structures of compounds #1 and #2. b) ITC thermograms for the binding of compounds #1 ($K_D = 1,8 \mu\text{M}$) and #2 ($K_D = 2.0 \mu\text{M}$) to BRD4(1). c) Thermal shift experiments performed as 3 independent experiments with compounds #1 and #2 on bromodomain BRD4(1). The black curve represents BRD4(1) alone in presence of 2.5% DMSO, the blue curve BRD4(1) with a 25 fold excess of compound 1 ($50 \mu\text{M}$ 2.5% DMSO, $\Delta T_m = 1.9 \text{ }^\circ\text{C}$) and orange curve with compound 2 ($\Delta T_m = 1.5 \text{ }^\circ\text{C}$).

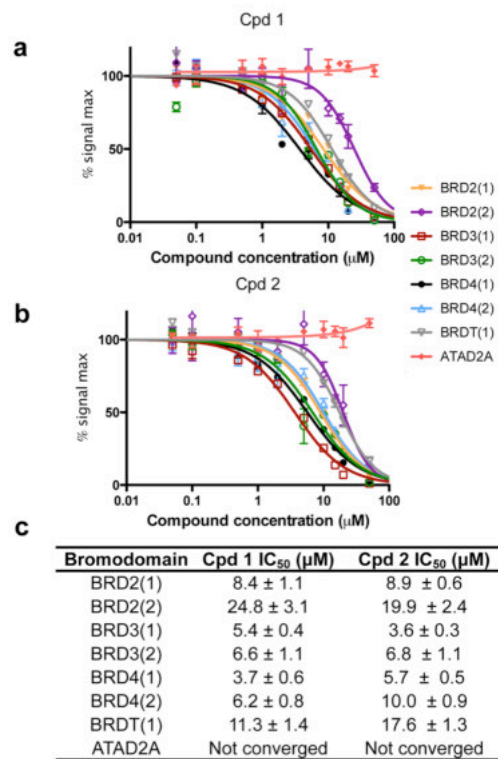


Figure S6: Selectivity of compounds #1 and #2. a) Selectivity profiles of compounds #1 and b) #2 on BET bromodomains, BRD2(1) (yellow), BRD2(2) (purple), BRD3(1) (red), BRD3(2) (green), BRD4(1) (black), BRD4(2) (blue), BRDT(1) (grey), ATAD2A (orange), using HTRF assays
 c) IC₅₀ values (± SEM, N=3) of bromodomain selectivity experiments with compound #1 and compound #2.

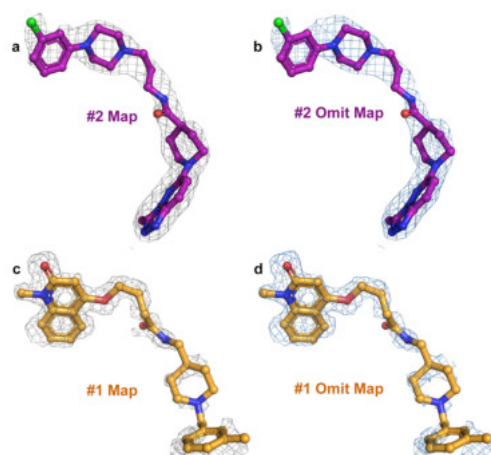


Figure S7: 2Fo-Fc map and omit map around compounds #1 and #2.

- a) 2Fo-Fc map contoured at 1σ around compound **#2**. This map is very well defined for the entire molecule.
- b) 2Fo-Fc omit map contoured at 1σ around the compound **#2**.
- c) 2Fo-Fc map around compound **#1**. This map indicated a lack of electron density for a part of the compound located outside of the N-acetylated binding pocket.
- d) 2Fo-Fc omit map contoured at 1σ sigma around compound **#1**.

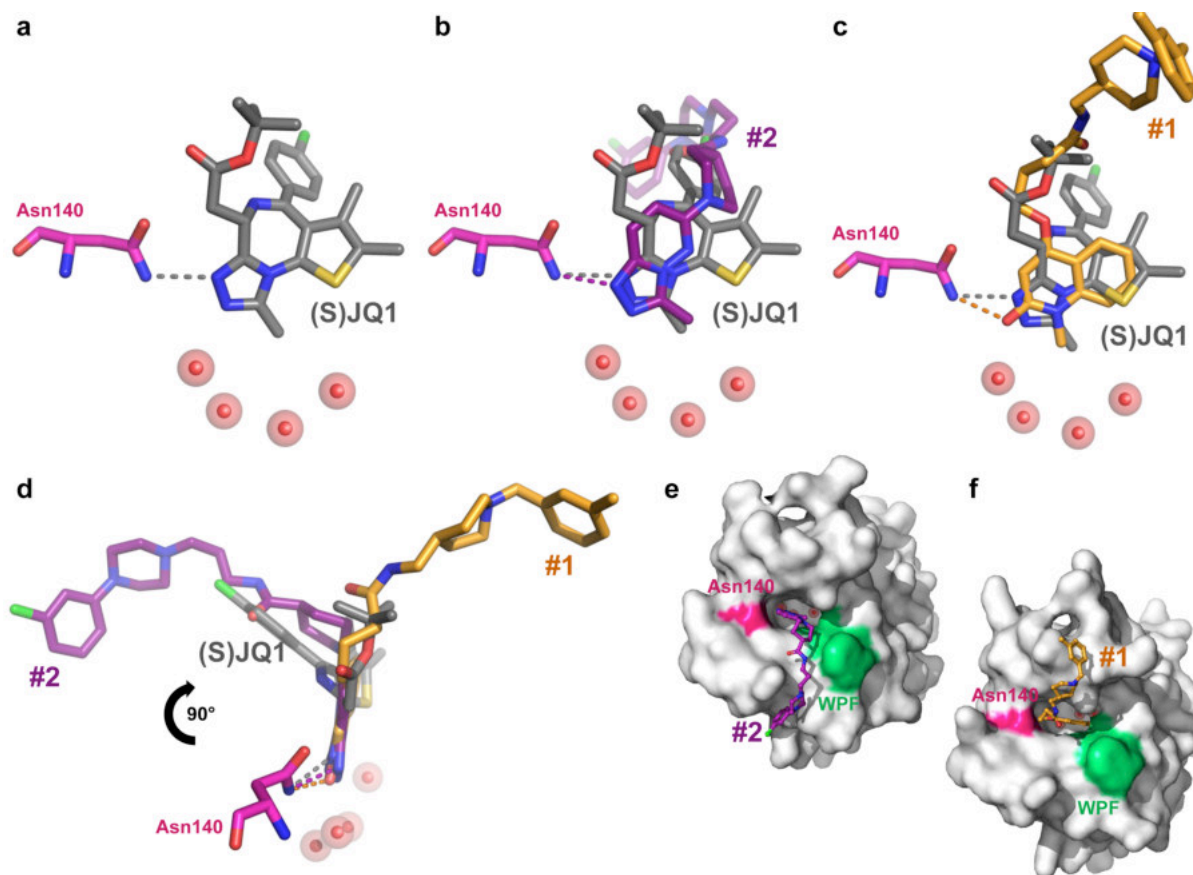


Figure S8: Binding mode of (S)JQ1 compared to compounds #1 and #2.

a) (S)JQ1, a pan-BET inhibitor (grey) in complex with BRD4(1) and detailed view of the the N140 residue (pink) hydrogen bond interaction (dash line). The four highly conserved water molecules (red spheres) are located at the bottom of the cavity.

b) Superimposition of (S)JQ1 with compound #2 (purple). The triazolo moieties of these two compounds are superimposed, with a slight shift.

c) Superimposition of (S)JQ1 (grey) with compound #1 (orange). The oxoquinoline moiety of compound #1 is in the same plane as the thieno-triazolo-diazepine moiety of (S)JQ1.

d) Superimposition of (S)JQ1, compound #1 and compound #2 with a 90° rotation. The chlorophenyl-piperazinyl-piperidinecarboxamide substituents of compound #1 are oriented in the same direction as the (S)JQ1 chlorophenyl moiety, engaging the shelf between the BC loop and the WPF motif. The methylphenyl-piperidinyl-butanamide substituents of compound #2 are oriented in the opposite direction.

e) Surface representations of BRD4(1) in complex with compound #2 and f) with compound #1. The surface of N140 is colored in pink while the surface of the WPF shelf is highlighted in green.

Supplementary Information – Tables

Table S1: 2D chemical structure of “thermal shift sensitive” compounds

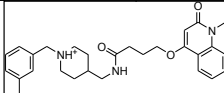
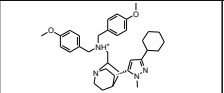
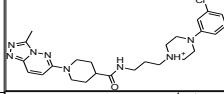
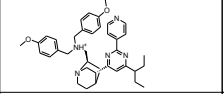
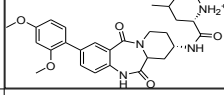
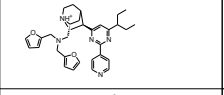
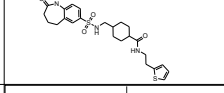
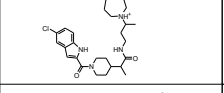
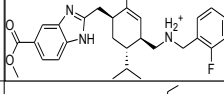
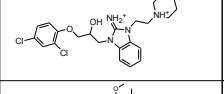
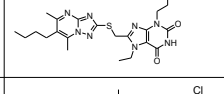
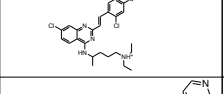
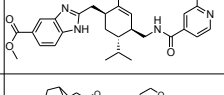
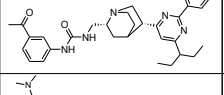
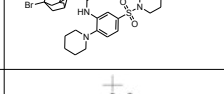
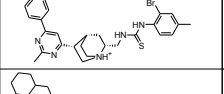
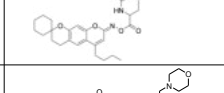
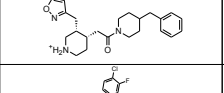
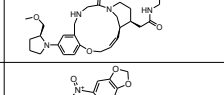
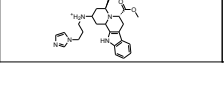
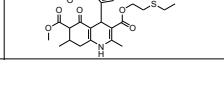
Compound #	Structure	PubChem Compound Identifier (CID)	Compound #	Structure	PubChem Compound Identifier (CID)
1		20862085	12		44715900
2		46326697	13		45360407
3		7134939	14		45360408
4		20937233	15		50816819
5		71694343	16		1306476
6		60498769	17		5793057
7		71693175	18		40777317
8		4658136	19		40776845
9		16416191	20		25313771
10		71694343	21		44716462
11		3324458			

Table S2: X-ray data collection and refinement statistics

	Cpd 1 (pdb: 5DLZ)	Cpd 2 (pdb: 5DLX)
Crystal parameters		
Space group	P2 ₁ 2 ₁ 2 ₁	P2 ₁ 2 ₁ 2 ₁
Unit-cell dimensions		
a, b, c (Å)	37.26, 44.24, 78.44	41.72, 48.51, 58.09
α, β, γ (°)	90.00, 90.00, 90.00	90.00, 90.00, 90.00
Data collection		
Beamline	ESRF – ID23-2	ESRF – ID30A-3
Wavelength (Å)	0.8726	0.9677
Resolution range (Å)	50 - 1.70	50 - 1.90
Total reflections	64144 (14342)	46717 (6904)
Unique reflections	9741 (2221)	9664 (1505)
Completeness (%)	96.3 (95.2)	99.3 (97.8)
Average I/σ	7.54 (4.15)	20.09 (3.73)
Rfactor obs (%)	14.9 (42.1)	4.2 (37.9)
Redundancy	4.47 (4.38)	4.83 (4.59)
Structure Refinement		
Starting model for Molecular Replacement	2OSS	2OSS
Resolution range (Å)	39.22 - 1.70	37.25 - 1.80
Rfree	0.3339	0.20707
R value		
Working + test set	0.25529	0.12959
Working set	0.25122	0.12547
Solvent (%)	42.92	37.31
R.m.s.d.		
Bond lengths (Å)	0.028	0.028
Angles (°)	2.329	2.316

Table S3: HTRF assay pipetting procedures

Protein A		Protein B		HTRF reagent A (CisBio)		HTRF reagent B (CisBio)		Buffer assay	Final incubation	Final DMSO concentration	N between partners
Name	Final Concentration (nM)	Name	Final Concentration (nM)	Name	Final Concentration (nM)	Name	Final Concentration (nM)				
Mix1		Mix 2		Mix 1		Mix 2					
Syntenin1FL-GST	2.5	Syndecan 2 peptide-Biot	47.5	MAB Anti GST-Tb	0.8	Streptavidin d2	10	Buffer A	ON 4°C	2%	19
GRASP55-GST	1.6	JamB peptide-Biot	5.9	MAB Anti GST-Tb	0.8	Streptavidin d2	1.25	Buffer A	ON 4°C	2%	3.7
NS3-GST	5	NS5Mtease-SNAP	15	MAB Anti GST-Tb	0.8	SNAP Red	-	Buffer A	4h RT	2%	3
NEF FL-His	2.5	GST-Hck-SH3	2.5	MAB Anti His-K	1	GST-d2	1.25	Buffer B	ON 4°C	2%	1
Mix 1		Mix 2		Mix 1		Mix 2					
Bromodomain BRD4(1)-GST	5	H4 Kac 5/8/12/16 peptide	100	MAB Anti GST-K	0.5	Streptavidin d2	12.5	Buffer B	3h RT	2%	20
BRD4(1)-GST	5	H4 KAc 5/8/12/16 peptide	25	MAB Anti GST-K	0.5	Streptavidin d2	3.1	Buffer B	3h RT	1%	5
BRD4(2)-GST	5	H4 KAc 5/8/12/16 peptide	40	MAB Anti GST-K	0.5	Streptavidin XL665	5	Buffer B	3h RT	1%	8
BRD2(1)-GST	5	H4 KAc 5/8/12/16 peptide	15	MAB Anti GST-K	0.5	Streptavidin d2	1.9	Buffer B	3h RT	1%	3
BRD2(2)-GST	5	H4 KAc 5/8/12/16 peptide	15	MAB Anti GST-K	0.5	Streptavidin XL665	1.9	Buffer B	3h RT	1%	3
BRD3(1)-GST	5	H4 KAc 5/8/12/16 peptide	10	MAB Anti GST-K	0.5	Streptavidin d2	1.25	Buffer B	3h RT	1%	2
BRD3(2)-GST	5	H4 KAc 5/8/12/16 peptide	200	MAB Anti GST-K	0.5	Streptavidin XL665	25	Buffer B	3h RT	1%	40
BRDT(1)-GST	5	H4 KAc 5/8/12/16 peptide	20	MAB Anti GST-K	0.5	Streptavidin d2	2.5	Buffer B	3h RT	1%	4
ATAD2A-GST	5	H4 KAc 5/8/12/16 peptide	80	MAB Anti GST-Tb	0.5	Streptavidin XL665	10	Buffer B	3h RT	1%	16
GST-Biotin	6.25	-		MAB Anti GST-Tb	0.8	Streptavidin d2	5	Buffer A	ON 4°C	2%	-
GST-SNAP	10	-		MAB Anti GST-Tb	0.8	SNAP Red	-	Buffer A	4h RT	2%	-
GST-His	2.5	-		MAB Anti His-K	1	GST-d2	1.25	Buffer B	ON 4°C	2%	-

Buffer A 50 mM HEPES, pH 7.5, 150 mM NaCl, 0.1% BSA

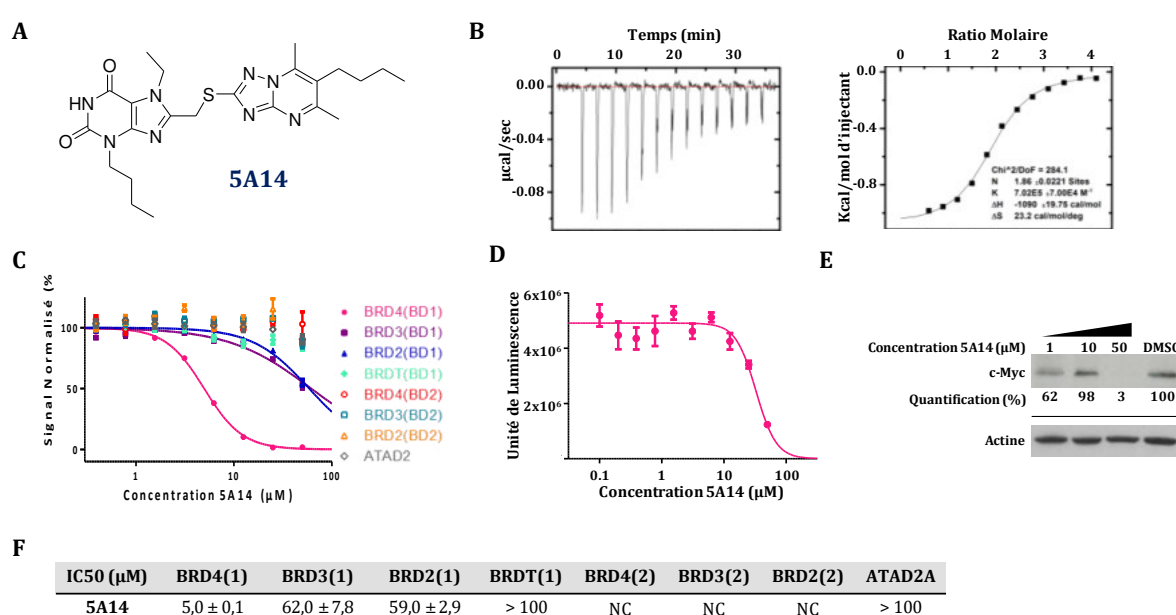
Buffer B 50mM KPO4, pH7, BSA 0.1%, 0.1 M KF

*Partie 2 : Compréhension du
mécanisme d'action de la molécule
BRD4(1) sélective « 5A14 »
par SAR*

I. Identification et Validation de la molécule 5A14

L'ensemble des résultats d'identification et de validation orthogonale de **5A14** est présenté dans la **Figure 26**. La molécule **5A14** a été identifiée comme inhibiteur du domaine BRD4(1) à l'issue du criblage de la chimiothèque 2P2I_{3D} en utilisant la méthode d'HTRF. Par la suite, j'ai déterminé le profil de sélectivité de cette molécule au sein de la famille BET. L'affinité (IC₅₀) du composé **5A14** pour le domaine BRD4(1) est de 5 µM, tandis qu'elle est supérieure à 50 µM pour les sept autres BDs. La molécule **5A14** est donc **sélective du domaine BD1 de la protéine BRD4 avec un rapport d'au moins dix fois en comparaison aux autres BDs de la famille BET**. J'ai confirmé cette affinité pour le domaine BRD4(1) par la mesure des valeurs du K_D de 1,4 µM par la méthode d'ITC et du ΔT_m de -1,5°C par la technique de TSA (ratio BD-ligand : 2 µM / 2 mM). La cytotoxicité de **5A14** a également été évaluée dans des cellules leucémiques Jurkat et la valeur d'EC₅₀ mesurée est de 27 µM. Cette valeur est semblable à celle obtenue sur les lignées KE37, MOLM14 et MV4-11. Par ailleurs, **5A14** a montré un effet de répression (down-régulation) de l'expression de l'oncogène c-Myc de manière dose dépendante.

Figure 26 : Identification et validation orthogonale de la molécule 5A14.

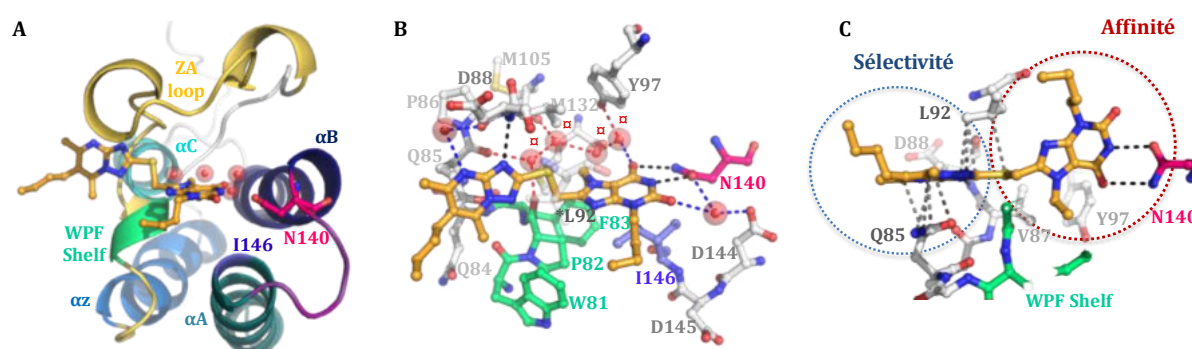


Légende : **A-** Structure chimique de la molécule 5A14 (3-butyl-8-(6-butyl-5,7-diméthyl-[1,2,4]triazolo[1,5-a]pyrimidin-2-ylsulfanylméthyl)-7-éthylxanthine). **B-** Validation orthogonale de 5A14 par ITC : à gauche thermogramme et à droite l'intégration des données brutes. La titration inverse a été réalisée à 35°C (35 µM de ligand 5A14 dans la seringue et 300 µM de BRD4(1) dans la cellule de mesure). Le K_D mesuré est de 1,4 µM. **C-** Évaluation par HTRF de 5A14 sur l'ensemble des membres de la famille BET, ainsi que sur le domaine ATAD2A comme contrôle négatif. Les courbes sont colorées en fonction du BD, comme indiqué à droite des courbes. L'IC₅₀ mesurée pour le domaine BRD4(1) est de 5 µM. **D-** Effet de 5A14 sur la viabilité cellulaire sur la lignée en fonction de la concentration et après incubation pendant 72h à 37°C. L'EC₅₀ mesurée est de 27 µM. **E-** Effet de 5A14 à 1, 10 et 50 µM sur la régulation de l'expression de l'oncogène c-Myc après incubation pendant 24h. L'inhibition de l'expression c-Myc est complète à 50 µM. **F-** Valeurs d'IC₅₀ mesurée par HTRF pour 5A14.

II. Caractérisation du mode de liaison de 5A14 à BRD4(1)

J'ai résolu la structure cristallographique du complexe BRD4(1)-5A14. Le mode de liaison de la molécule 5A14 sur le domaine BRD4(1), présenté dans la **Figure 27**, a été déterminé par cristallographie aux rayons X (PDB : **5EGU**, 2,2 Å). Le cœur xanthine forme une liaison hydrogène avec l'asparagine conservée N140, ainsi qu'une liaison avec la tyrosine Y97 par le biais d'une des molécules d'eau conservées dans la cavité K_{ac}. Dans cette poche sont enfouies cinq molécules d'eau conservées qui forment un réseau d'interaction permettant de stabiliser la structure du complexe. Nous avons pu constater que le groupement triazolopyrimidinyle se plaque contre le « ZA channel ».

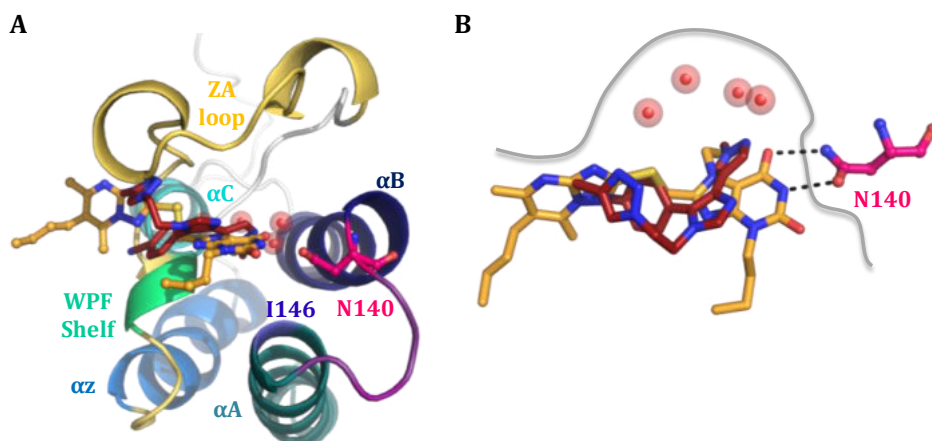
Figure 27 : Structure cristallographique du complexe BRD4(1)-5A14.



Légende : Les conditions de cristallisation et les tables de statistiques sont en annexe. **A-** Structure du complexe BRD4(1)-5A14 (PDB : **5EGU**, 2,2 Å). Le domaine BRD4(1) est représentée en « cartoon », les quatre hélices (αA , αB , αC , αZ) sont colorées en bleu (dégradé), les boucles ZA et BC sont représentées en jaune et violet respectivement. Les molécules d'eau conservées dans le site actif de BRD4(1) sont représentées par des sphères roses. Le résidu asparagine N140 conservé est représenté en « sticks and balls » rose. Le « gatekeeper » I146 est coloré en bleu et le « WPF shelf » est coloré en vert. La molécule 5A14 est quant à elle représentée en « sticks and balls » orange. **B-** Les résidus clés impliqués dans le réseau d'interaction avec les molécules d'eau, BRD4(1) et le ligand sont représentés en « stick and balls » gris excepté pour le « WPF » (vert), « gatekeeper » (bleu), N140 (rose). Les liaisons hydrogènes sont représentées par des lignes en pointillés de couleurs rouge --- pour les molécules d'eau entre elles et noire --- pour les interactions entre la molécule 5A14 et BRD4(1), et bleue --- entre la molécule 5A14 et les molécules d'eau. **C-** En pointillés noirs --- sont représentés les liaisons hydrogènes entre la molécule 5A14 et le résidu conservé N140 de BRD4(1), responsable de l'affinité de la molécule pour BRD4(1). En pointillés gris sont représentés les interactions de type van der Waals entre la molécule et BRD4(1). Ces interactions sont impliquées dans le profil de sélectivité observé.

Ce mode de liaison a récemment été décrit pour un inhibiteur de type benzimidazolone ciblant le domaine BAZ2B (**Figure 28**) (223-225). Le groupement triazolo établit quant à lui une liaison hydrogène avec le résidu d'acide aspartique D88 conservé chez les BETs. Le triazolopyrimidinyle présent dans le « ZA channel » établit des contacts de type van der Waals avec les résidus leucine L92 et glutamine Q85. Par ailleurs le groupement pyrimidine est très proche de la glutamine Q85. Ce résidu est conservé chez les BD1 excepté le domaine BRDT(1) et il est remplacé par un résidu plus long arginine ou lysine chez les BD2.

Figure 28 : Comparaison du mode de fixation de 5A14 avec l'inhibiteur de BAZ2B.



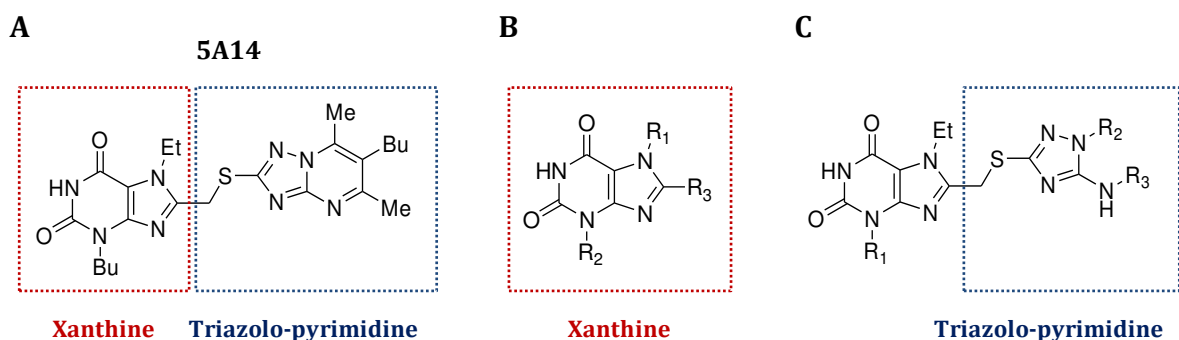
*Légende : A- Superposition des molécules 5A14 (représentation « balls and sticks » de couleur orange, PDB : **5EGU**) et de l'inhibiteur de BAZ2B (représentation « sticks » de couleur rouge, PDB : **4XUB**) dans la cavité K_{ac} du domaine BRD4(1). BRD4(1) est représentée en comme précédemment. B- Vue détaillée du mode de fixation des deux molécules. Le trait gris représente la poche K_{ac}.*

Ces informations structurales nous ont conduits à émettre l'hypothèse que le cœur xanthine serait responsable de l'affinité de la molécule 5A14 et que le groupement triazolopyrimidine serait impliqué dans le profil de sélectivité observé. Afin de valider cette hypothèse et de mieux comprendre l'origine de ce profil de sélectivité nous avons initié un programme de SAR pour cette molécule.

III. Validation de notre hypothèse

Comme indiqué dans la **Figure 29-A**, la molécule **5A14** est formée des deux cœurs chimiques, une « xanthine » (purine-2,6-dione) et une « triazolo-pyrimidine » ([1,2,4]triazolo[1,5-a]pyrimidine), qui d'après la structure du complexe BRD4(1)-**5A14** décrite précédemment (PDB : **5EGU**), se fixent respectivement dans la cavité K_{ac} et le ZA channel du domaine BRD4(1). En vue de valider nos postulats : (1) le cœur « xanthine » serait responsable de l'affinité de la molécule **5A14** et (2) le cœur « triazolopyrimidine » serait quant à lui responsable de la sélectivité de celle-ci, nous avons modulé les positions R₁, R₂ et R₃ de ces deux cœurs chimiques respectifs (**Figure 29-B & C**). Au total 130 analogues ont été synthétisés et évalués.

Figure 29 : Fragmentation de la molécule 5A14.



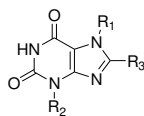
Légende : **A-** Structure chimique de la molécule 5A14, où sont encadrés en rouge le cœur xanthine (1H-purine-2,6(3H,7H)-dione) et en bleu le cœur triazolo-pyrimidine ((6-butyl-2-(ethylthio)-5,7-dimethyl-[1,2,4]triazolo[1,5-a]pyrimidine). **B-** Modulation du cœur xanthine aux positions R_1 , R_2 et R_3 . **C-** Modulation du cœur triazolopyrimidine aux positions R_1 , R_2 et R_3 .

1. Modulation du cœur « xanthine »

Dans un premier temps, nous avons modulé le cœur xanthine (**Figure 29-B**). J'ai évalué l'ensemble de ces molécules par HTRF (IC_{50}), puis quand cela était possible je les ai validé par ITC (K_D) et caractérisé par cristallographie. Dans l'article **Raux et al (211)** situé à la section « article 2 », nous avons publié les résultats pour les quatre principales molécules, présentées dans le **Tableau 13**. Ainsi, la xanthine elle-même (1H-purine-2,6(3H,7H)-dione ; composé **1882**) ne présente aucune activité sur BRD4(1), de même que la xanthine substituée en position R_1 par un éthyle (7-ethyl-1H-purine-2,6(3H,7H)-dione ; composé **1888**). Par contre les xanthines disubstituées en positions R_1 par un éthyle et R_2 par un butyle (composé **1915**) ou un 4-chlorobenzyle (composé **1916**), présentent respectivement des activités (IC_{50}) de 13 μM et 2,6 μM .

Tableau 13 : Évaluation des analogues « xanthine » par HTRF et ITC.

Cœur chimique

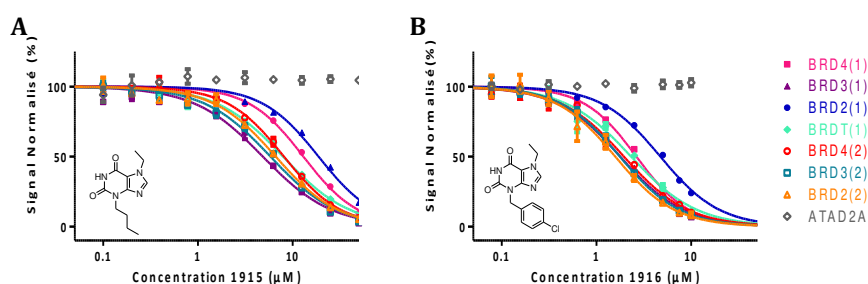


Composés	Substituants		BRD4(BD1)		
	R ₁	R ₂	IC ₅₀ (μM)	K _D (μM)	LE
1882	H	H	NA	-	NA
1888	Et	H	NA	-	NA
1915	Et	Bu	13,0 ± 0,3	-	0,40
1916	Et	4-ClBn	2,6 ± 0,1	1.8	0,37

Légende : L'ensemble de ces molécules ont été évaluées sur le domaine BRD4(1). Les valeurs mesurées (IC₅₀ et K_D) sont indiquées en micromolaire. Les IC₅₀ ont été mesurées par HTRF et le K_D a été mesuré par ITC. Le Ligand Efficiency (LE) est calculé à partir de la formule $LE = (\Delta G)/N = 1,4 (-\log (IC_{50}))/N$ (226).

J'ai déterminé les profils de sélectivité des molécules **1915** et **1916** au sein de la famille BET (**Figure 30**). Comme nous l'avions prédit ces deux composés sont des inhibiteurs de type pan-BET avec des IC₅₀ comprises entre 4,8 et 18,8 μM pour **1915** et entre 1,6 et 4,7 μM pour **1916**, ATAD2A (contrôle négatif) n'étant pas inhibé à 50 μM par ces deux molécules.

Figure 30 : Profils de sélectivité des molécules 1915 et 1916.



C

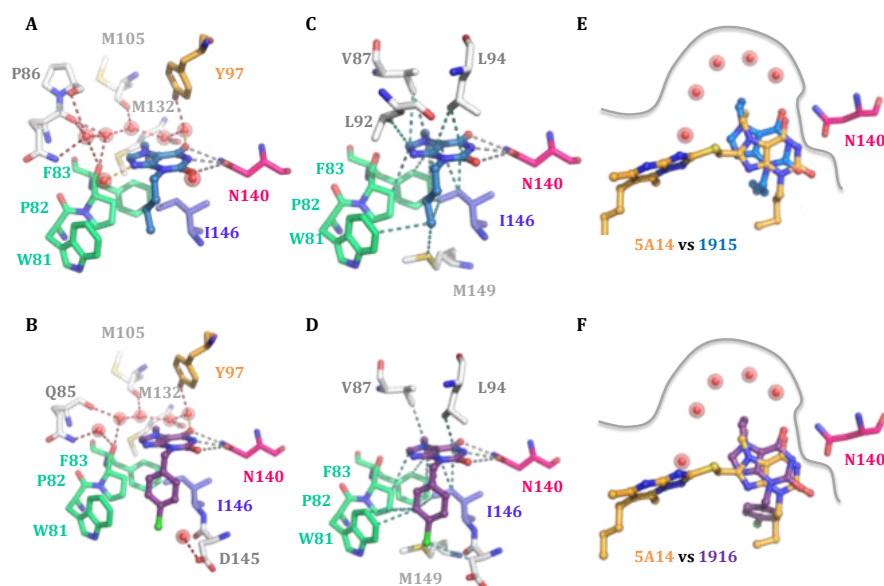
IC ₅₀ (μM)	BRD4(1)	BRD3(1)	BRD2(1)	BRD1(1)	BRD4(2)	BRD3(2)	BRD2(2)	ATAD2A
1915	13,0 ± 0,3	4,8 ± 0,3	18,8 ± 0,7	7,9 ± 0,4	8,1 ± 0,3	5,8 ± 0,2	6,6 ± 0,3	NC
1916	2,6 ± 0,1	1,8 ± 1,1	4,7 ± 0,1	2,3 ± 0,2	1,9 ± 0,1	1,8 ± 0,2	1,6 ± 0,2	NC

Légende : Les profils de sélectivité ont été déterminés par HTRF. **A-** Profil de sélectivité de la molécule 1915. **B-** Profil de sélectivité de la molécule 1916. Les courbes sont colorées en fonction du BD, comme indiqué à droite du panneau **B**. **C-** Valeurs d'IC₅₀ pour les molécules 1915 et 1916.

J'ai résolu les structures cristallographiques de plusieurs molécules en complexes avec BRD4(1). Les modes de liaison des molécules **1915** et **1916** au domaine BRD4(1) sont présentés dans la **Figure 31**. J'ai comparé leurs modes de liaison à celui de **5A14**, afin de déterminer les interactions impliquées dans l'affinité. Comme observé pour la structure du complexe BRD4(1)-**5A14** (PDB : **5EGU**), les molécules **1915** (PDB : **À déposer**, 1,3 Å) et **1916** (PDB : **5EIS**, 1,6 Å) établissent des liaisons hydrogènes avec les résidus asparagine N140 et tyrosine Y97 et des contacts de type van der Waals avec les résidus de la boucles ZA : W81, P82, F83, V87 et L94 et de l'hélice α_c : M149, dont le « gatekeeper » I146. L'atome de chlore de la molécule **1916** établit un contact supplémentaire avec le résidu acide aspartique D145 de l'hélice α_c .

En superposant les structures de ces trois complexes, nous avons pu constater que les molécules **1915** et **1916** pénètrent plus profondément dans la cavité K_{ac} que **5A14**. De plus, la molécule d'eau située à l'entrée du « ZA channel » pour les deux complexes (PDB : **À déposer**, **5EIS**) qui est conservée dans de nombreuses structures de complexes BRD4(1)-ligand, est absente dans la structure du complexe avec la molécule **5A14**. Enfin, l'orientation du groupement xanthine est légèrement modifiée pour ces deux molécules par rapport à **5A14** et le groupement butyle du composé **1915** est plaqué contre le « WPF channel », contrairement à celui de **5A14**.

Figure 31 : Structures de BRD4(1)-1915 et BRD4(1)-1916 vs BRD4(1)-5A14.



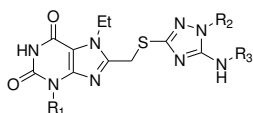
Légende : Les conditions de cristallisation et les tables de statistiques sont en annexe. **A & B-** Mode de fixation des molécules 1915 (PDB : **À déposer**, 1,3 Å) (bleu) et 1916 (PDB : **5EIS**, 1,6 Å) (violet). Les liaisons hydrogènes établies dans la cavité K_{ac} sont représentées par des lignes en pointillés, de couleurs gris pour les liaisons directes entre BRD4(1) et l'inhibiteur, en orange --- pour les liaisons indirectes entre l'inhibiteur, une molécule d'eau et la protéine et en rouge --- pour le réseau d'interactions entre les molécules d'eau. **C & D-** Contacts de van der Waals établis entre BRD4(1) et les molécules 1915 (bleu) et 1916 (violet) dans la cavité K_{ac} du domaine BRD4(1) sont représentées par des lignes en pointillés de couleur bleue ---. **E & F-** Superposition la molécule 5A14 (orange) avec les composés 1915 (bleu) et 1916 (violet). Le trait gris représente la cavité K_{ac} .

2. Modulation du cœur « triazolopyrimidine »

Dans un deuxième temps nous avons modulé le cœur triazolopyrimidine (**Figure 29-C**). J'ai évalué l'ensemble de ces molécules par HTRF (IC_{50}), et quand cela était possible je les ai validé par ITC (K_D) et caractérisé par cristallographie. Dans l'article **Raux et al** (211), situé à la section « article 2 », nous avons publié les résultats pour les trois principales molécules, présentés dans le **Tableau 14**. La molécule **1913** est quatre fois moins active que **5A14**. Inversement à ce qu'on a pu observer pour les molécules **1915** et **1916**, le composé substitué en position N-3 par un groupement 4-chlorobenzyle est moins actif que l'analogue butyle. Les molécules **1928** et **5024** sont elles aussi moins actives que **5A14** et de nouveau celle portant le groupement 4-chlorobenzyle est plus active que celle avec le groupement butyle.

Tableau 14 : Évaluation des analogues « triazolo-pyrimidine » par HTRF et ITC.

Cœur chimique

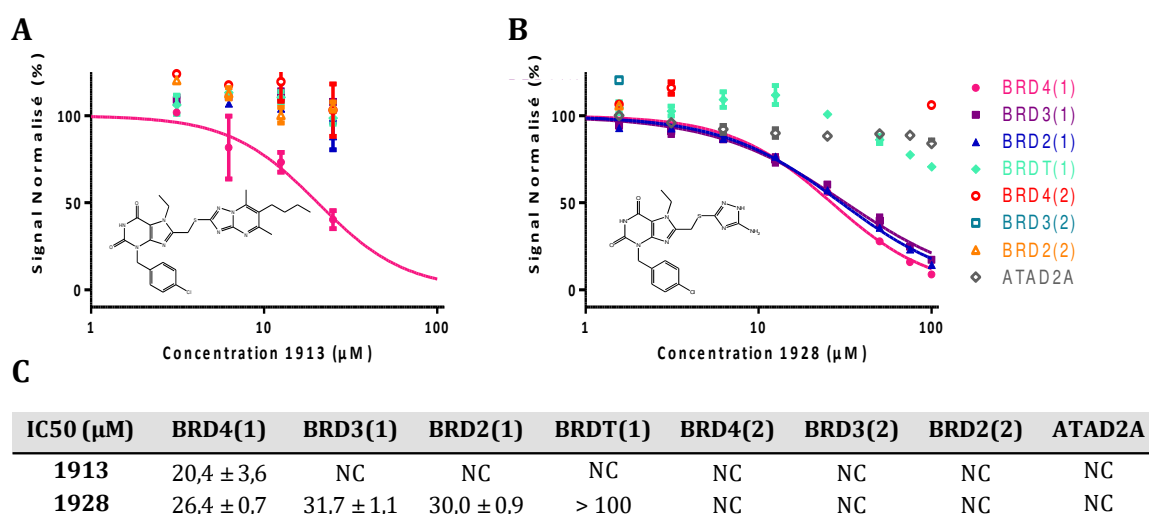


Composés	Substituants			BRD4(BD1)		
	R ₁	R ₂	R ₃	IC ₅₀ (μM)	K _D (μM)	LE
5A14	Bu			5,0 ± 0,1	1,4	0,22
1913	4-ClBn			20,4 ± 3,6	-	0,15
5024	Bu	H	H	> 50	-	NA
1928	4-ClBn			26,4 ± 0,7	-	0,17

Légende : L'ensemble de ces molécules ont été évaluées sur BRD4(1). Les valeurs mesurées (IC_{50} et K_D) sont indiquées en micromolaire. Les IC_{50} ont été mesurées par HTRF et le K_D a été mesuré par ITC. Le Ligand Efficiency (LE) est calculé à partir de la formule $LE = (\Delta G)/N = 1,4 (-\log (IC_{50}))/N$ (226).

J'ai déterminé les profils de sélectivité des deux meilleures molécules, à savoir **1913** et **1928**, au sein de la famille BET, présentés dans la **Figure 32**. Tout comme **5A14**, la molécule **1913** semble être BRD4(1) sélective, les valeurs IC_{50} des autres BDs sont supérieures à 50 μM. La molécule **1928** semble quant à elle BET-BD1 sélective, avec des IC_{50} de l'ordre de 30 μM pour les BET-BD1 excepté BRDT(1) qui présente une valeur d' IC_{50} supérieure à 50 μM tout comme les BET-BD2 et ATAD2A (contrôle négatif).

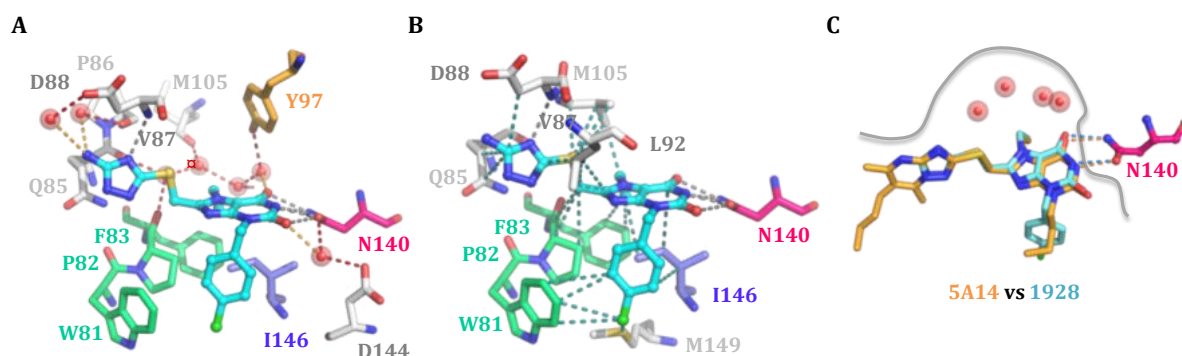
Figure 32 : Profils de sélectivité des molécules 1913 et 1928.



Légende : Les profils de sélectivité ont été déterminés par HTRF. **A-** Profil de sélectivité de la molécule 1913. **B-** Profil de sélectivité de la molécule 1928. Les courbes sont colorées en fonction du BD, comme indiqué à droite du panneau **B**. **C-** Valeurs d'IC₅₀ pour les molécules 1913 et 1928.

Le mode de liaison de la molécule **1928** au domaine BRD4(1), présenté dans la **Figure 33**, a été caractérisé par cristallographie et comparé à celui des molécules **5A14** et **1916**. Comme observé dans la structure du complexe BRD4(1)-**5A14** (PDB : **5E14**, 1,0 Å), le réseau d'interaction avec les molécules d'eau est conservé. De plus, le cœur xanthine forme des liaisons, hydrogène, directe avec l'asparagine conservée N140 et indirecte avec la tyrosine Y97 par le biais d'une des molécules d'eau conservées dans la cavité K_{ac}. Enfin, le groupement triazolo établit des liaisons avec les résidus du « ZA channel », directe avec le résidu d'acide aspartique D88 et indirecte *via* des molécules d'eau avec les résidus proline P86 et glutamine Q85. Concernant les contacts de type van der Waals, la molécule **1928** interagit *via* la xanthine avec les résidus du « ZA channel » : F83 et V87 et le « gatekeeper » I146 comme pour les composés **5A14** et **1916**, ainsi qu'avec le résidu leucine L92, versus L94 pour **5A14** et versus P82 et L94 pour **1916**. Le groupe chlorophényle interagit comme le composé **1916**, avec les résidus du « ZA channel » W81 et de l'hélice α_c : M149, dont le « gatekeeper » I146, mais pas avec les résidus proline P82 et acide aspartique D145. Le triazolo interagit quant à lui, comme **5A14**, avec les résidus Q85, D88 et L92.

Figure 33 : Structure de BRD4(1)-1928 et comparaison avec BRD4(1)-5A14.



Légende : Les conditions de cristallisation et les tables de statistiques sont en annexe. **A-** Mode de fixation des molécules 1928 (bleu clair) (PDB : **5E14**, 1,0 Å). Comme précédemment, les liaisons hydrogènes établies dans la cavité K_{ac} sont représentées par des lignes en pointillés, de couleurs grise -- pour les liaisons directes entre BRD4(1) et l'inhibiteur, en orange --- pour les liaisons indirectes entre l'inhibiteur, une molécule d'eau et BRD4(1) et en rouge --- pour le réseau d'interactions entre les molécules d'eau. **B-** Contacts de van der Waals établis entre BRD4(1) et 1928 sont représentées par des lignes en pointillés de couleur bleue ---. **C-** Superposition la molécule 5A14 (orange) avec les composés 1928 (bleu clair). Le trait gris représente la cavité K_{ac} .

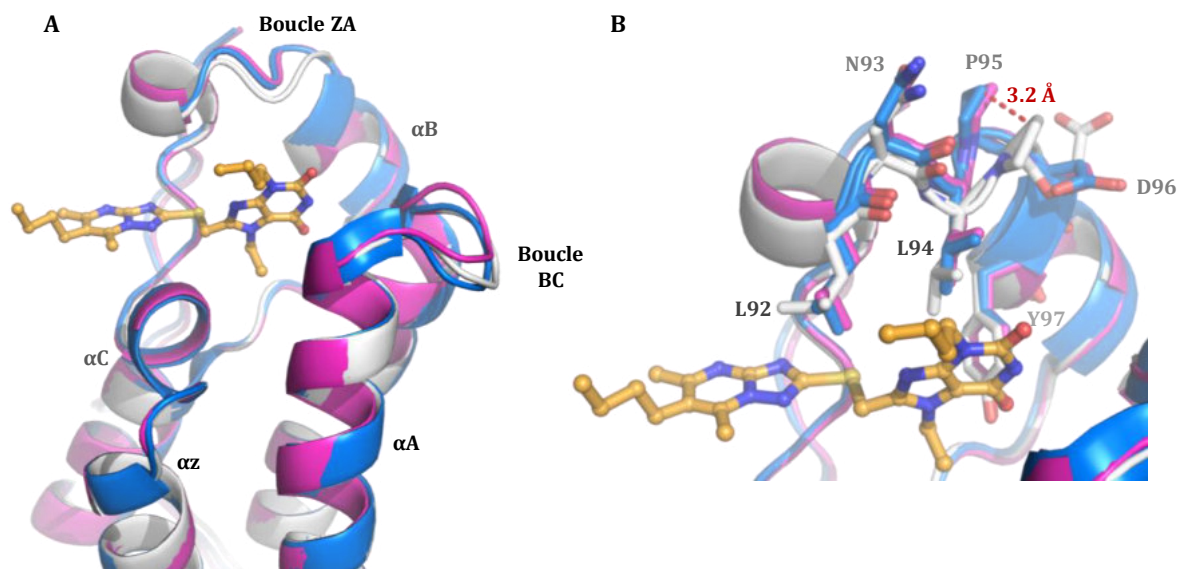
IV. Conclusions et perspectives

Le programme de SAR que nous avons initié pour la molécule 5A14, nous a permis d'identifier les cœurs chimiques impliqués dans les profils sélectivité observés pour les différents analogues synthétisés : pan-BET, BET-BD1 et BRD4(1) sélectifs.

En effet, l'ensemble des résultats suggèrent que : i) Le cœur xanthine N_3,N_7 dialkylé est responsable de l'inhibition pan-BET (d'après les profils de sélectivité de **1915** et **1916**) et que cette inhibition se fait *via* une liaison hydrogène avec l'asparagine N140 et une interaction indirecte avec la tyrosine Y97 par le biais d'une molécule d'eau (d'après les structures de BRD4(1) en complexe avec **1915** et **1916**). ii) La condensation du groupe triazolo conduit à une inhibition préférentielle des domaines BD1 des protéines BET « BET-BD1 » (d'après le profil de sélectivité de **1928**), *via* la formation d'une liaison hydrogène avec le résidu conservé d'acide aspartique D88 (d'après la structure de BRD4(1) en complexe avec **1928**). iii) La condensation du groupe pyrimidinyle est quant à elle impliquée dans la sélectivité BRD4(1) de la molécule **5A14** (d'après le profil de sélectivité de **5A14**), par le biais de l'établissement de contacts de type van der Waals avec le résidu glutamine Q85. En effet, un encombrement stérique avec les résidus arginine (R) de BRDT(1) et lysine (K) des domaines BD2 des protéines BET, empêche cette interaction (d'après la structure de BRD4(1) en complexe avec **5A14**).

Par ailleurs, la liaison hydrogène entre le groupe triazolo et l'acide aspartique D88 conservée pour l'ensemble des BET, pourrait être stabilisée pour certaines de ces BDs du fait de la dynamique de la boucle ZA. Plusieurs articles ont montré que la dynamique de cette boucle était cruciale pour la fixation d'inhibiteurs de BETs (57, 193, 227). Nous avons pu constater cette dynamique de la boucle ZA, en superposant les structures du domaine BRD4(1) sous forme libre ou en complexe avec les molécules **5A14** et **1915**. Comme le montre la **Figure 34**, les boucles de la forme libre et en complexe avec le composé **1915** se superposent très bien, alors que celle de la forme en complexe avec la molécule **5A14** est déplacée au niveau des résidus leucine L94, proline P95 et acide aspartique D96. Le résidu proline P95 est notamment décalé de 3,2 Å par rapport aux deux autres structures.

Figure 34 : Comparaison de la forme libre et complexée de BRD4(1).



Légende : Superpositions BRD4(1) sous forme libre ou en complexe avec 5A14 et 1915. Les structures de BRD4(1) sont représentées en « cartoon » et colorées en blanc pour la forme libre (PDB : **2OSS**), en bleu pour le complexe avec 1915 (PDB : **5EIS**) et en rose pour le complexe avec 5A14 (PDB : **5EGU**) A- Vu d'ensemble de la fixation de la molécule 5A14 (représenté en « balls and sticks » de couleur orange). B- Zoom sur le réarrangement de la boucle ZA au niveau des résidus L94, P95 et D96.

À l'issue de ce travail, nous nous sommes fixés pour objectif d'améliorer l'affinité de cette molécule au nanomolaire afin d'obtenir une sonde épigénétique nous permettant de mener une étude cellulaire plus approfondie afin d'étudier le rôle individuel du domaine BRD4(1) dans la prolifération de cellules cancéreuses et le processus biologique lié à sa fixation à la chromatine.

Article 2

Article N°2

Exploring Selective Inhibition of the First Bromodomain of the Human Bromodomain and Extra-terminal Domain (BET) Proteins.

Raux B, Voitovich Y, Derviaux C, Lugari A, Rebuffet E, Milhas S, Priet S, Roux T, Trinquet E, Guillemot JC, Knapp S, Brunel JM, Fedorov AY, Collette Y, Roche P, Betzi S, Combes S1, Morelli X.

J Med Chem. 2016 Feb 25;59(4):1634-41.

Patent: EP 15 307 163.4

Abstract:

A midthroughput screening follow-up program targeting the first bromodomain of the human BRD4 protein, BRD4(BD1), identified an acetylated-mimic xanthine derivative inhibitor. This compound binds with an affinity in the low micromolar range yet exerts suitable unexpected selectivity in vitro against the other members of the bromodomain and extra-terminal domain (BET) family. A structure-based program pinpointed a role of the ZA loop, paving the way for the development of potent and selective BET-BRD_i probes.

PMID: 26735842

Exploring Selective Inhibition of the First Bromodomain of the Human Bromodomain and Extra-terminal Domain (BET) Proteins

Brigitt Raux,[†] Yuliia Voitovich,^{†,‡} Carine Derviaux,[†] Adrien Lugari,^{†,○} Etienne Rebuffet,[†] Sabine Milhas,^{†,§} Stéphane Priet,[§] Thomas Roux,^{||} Eric Trinquet,^{||} Jean-Claude Guillemot,[§] Stefan Knapp,^{⊥,¶,▽} Jean-Michel Brunel,[†] Alexey Yu. Fedorov,[‡] Yves Collette,[†] Philippe Roche,[†] Stéphane Betzi,[†] Sébastien Combes,^{*,†} and Xavier Morelli^{*,†}

[†]Centre National de la Recherche Scientifique (CNRS), Centre de Recherche en Cancérologie de Marseille (CRCM), UMR 7258; INSERM U1068; Institut Paoli-Calmettes; Aix-Marseille Université, UM105, 13273 Marseille, France

[‡]Department of Organic Chemistry, Lobachevsky State University of Nizhni Novgorod, Gagarina av. 23, Nizhni Novgorod 603950, Russia

[§]Screening Platform AD2P, CNRS, AFMB UMR 7257, Aix-Marseille Université, 13288 Marseille, France

^{||}Cisbio Bioassays, R&D, Parc Marcel Boiteux, BP 84175, 30200 Codolet, France

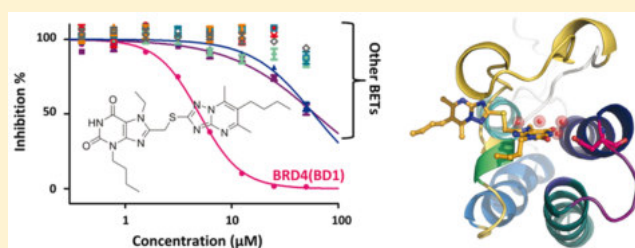
[⊥]Target Discovery Institute, University of Oxford, NDM Research Building, Roosevelt Drive, Oxford OX3 7FZ, U.K.

[¶]Structural Genomics Consortium, University of Oxford, Old Road Campus Research Building, Roosevelt Drive, Oxford OX3 7DQ, U.K.

[▽]Goethe-University, Institute for Pharmaceutical Chemistry and Buchmann Institute for Life Science, Campus Riedberg, Max-von Laue Str. 9, 60438 Frankfurt am Main, Germany

S Supporting Information

ABSTRACT: A midthroughput screening follow-up program targeting the first bromodomain of the human BRD4 protein, BRD4(BD1), identified an acetylated-mimic xanthine derivative inhibitor. This compound binds with an affinity in the low micromolar range yet exerts suitable unexpected selectivity *in vitro* against the other members of the bromodomain and extra-terminal domain (BET) family. A structure-based program pinpointed a role of the ZA loop, paving the way for the development of potent and selective BET-BRD_i probes.

**■ INTRODUCTION**

Bromodomains (BRD) are protein interaction modules that preferentially bind ϵ -N-acetylated lysine residues through structurally well-defined pockets. BRDs are found in eight protein families, which include a total of 46 nuclear or cytoplasmic proteins in human with diverse structures and functions, including chromatin-modifying enzymes, helicases, chromatin remodelers, transcriptional coactivators and mediators, and the bromodomain and extra-terminal domain (BET) family of proteins.^{1–3} BET proteins (BRD2, BRD3, BRD4, and the testis-specific BRDT) have a conserved modular architecture including two N-terminal tandem BRDs (BD1 and BD2). The BETs play a central role in chromatin biology by acting as tissue-specific recruitment platforms that tether complexes to acetylated histones and chromatin, facilitating the assembly of the transcriptional machinery and controlling gene expression in inflammation, viral infection, and cancer biology. For example, BRD2 is specifically recruited to acetylated histones H3 and H4, and this interaction is linked to active transcription and mitosis.^{4,5} BRD2 and BRD3 are required for permissive RNA polymerase II transcription through acetylated

nucleosomes,⁶ and it has been suggested that BRD4 binds acetylated histones using primarily its first bromodomain (BD1).^{7,8} The BD2 domain also recognizes and interacts with the acetylated region of cyclin T1, which forms a complex with the positive transcription elongation factor b and is crucial for the sustained presence of Pol II in active genes and for transcription initiation and elongation,⁹ thereby regulating the expression of cell proliferation supporting genes, including c-Myc and its target genes.¹⁰

BET proteins are often deregulated in disease, their transcription-regulating activity being altered and thus affecting numerous growth-promoting genes and cytokines. BET proteins are known to be deregulated in cancer,¹¹ and the recent disclosure of pan-BET inhibitors that attenuate BRD function has allowed the validation of these drug targets, shedding light on their roles in disease. Interestingly, BRD4

Special Issue: Epigenetics

Received: November 2, 2015

Published: January 6, 2016

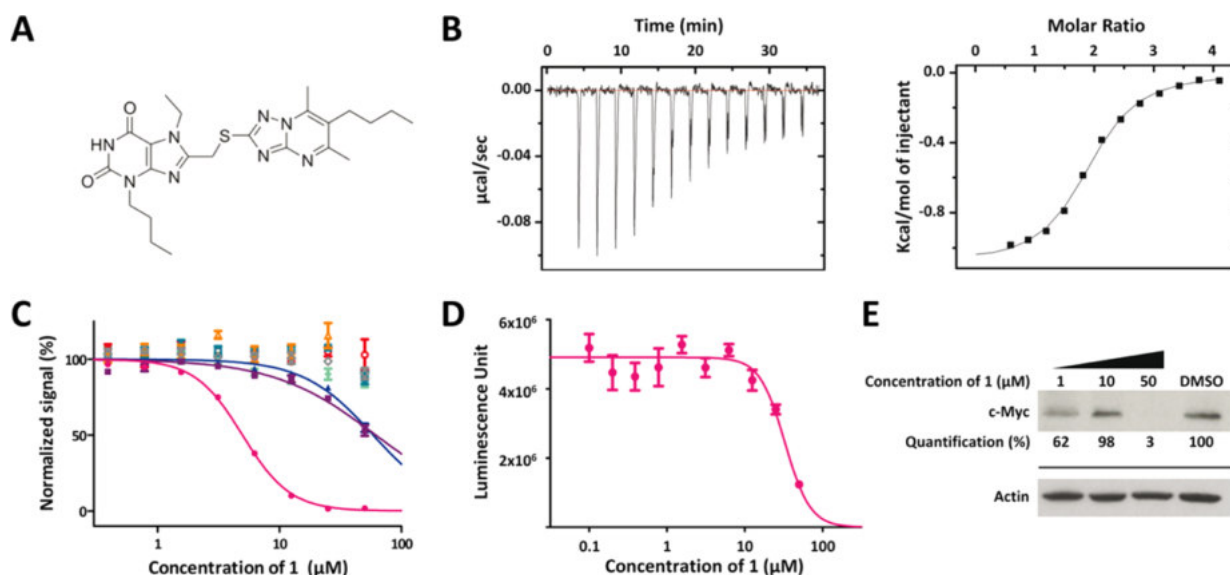


Figure 1. Identification of a selective BRD4(BD1) acetylated-mimic xanthine inhibitor **1**. (A) Structure of 3-butyl-8-(6-butyl-5,7-dimethyl-[1,2,4]triazolo[1,5-*a*]pyrimidin-2-ylsulfanylmethyl)-7-ethylxanthine, **1**. (B) Isothermal titration calorimetry (ITC) thermogram (left) and nonlinear least-squares fit model of the integrated data (right) at 35 °C (300 μM BRD4(BD1) vs 35 μM **1**) as a function of the molar ratio of ligand (cell) to the protein (syringe). (C) Homogeneous time resolved fluorescence (HTRF) selectivity assay of **1** on all BET members and ATAD2 as a non-BET control after excitation at 337 nm, energy transfer at 620 nm, and fluorescence emission at 665 nm. Fluorescence data are normalized and plotted as a function of the ligand concentration. Curves are colored by bromodomains (BRD4(BD1), pink \bullet ; BRD3(BD1), purple \blacksquare ; BRD2(BD1), blue \blacktriangle ; BRDT(BD1), cyan \blacklozenge ; BRD4(BD2), red \circ ; BRD3(BD2), gray \square ; BRD2(BD2), orange \triangle ; ATAD2, gray \diamond). (D) Effect of **1** on Jurkat cell viability as a function of the compound concentration after 72 h incubation at 37 °C. (E) C-myc pro-oncogene downregulation profile in the presence of different concentrations of **1** at 0.5% DMSO after 24 h incubation (50 μg of loaded protein).

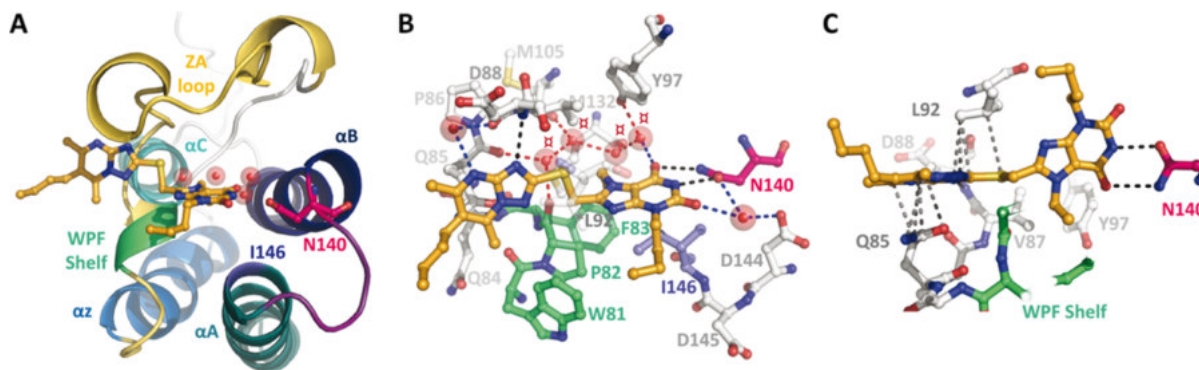
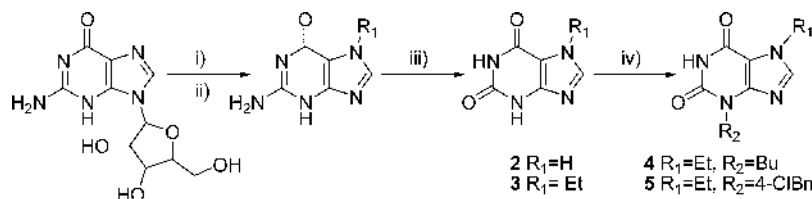


Figure 2. Molecular mode of action of **1** in the N-acetylated binding pocket of BRD4(BD1). (A) Three dimensional crystallographic structure of **1** (orange stick and ball representation) complexed with BRD4(BD1) obtained at 2.2 Å resolution (as a colored cartoon labeled by secondary structures) showing the conserved N140 residue (pink), WPF shelf (green), I146 gatekeeper residue (blue), and four highly conserved water molecules (red spheres) located at the bottom of the cavity. (B) Detailed view of **1**/BRD4(BD1) interaction. BRD4(BD1) is shown as ball and stick representation. The four conserved water molecules at the bottom of the cavity (red spheres marked with a π) are involved in an extensive hydrogen bond network with BRD4(BD1) (red dashed lines). Direct hydrogen bond interactions between **1** and BRD4(BD1) are shown as black dashed lines, while indirect water mediated interactions are shown as blue dashed lines. The L92 is displayed in transparency for clarity. (C) Detailed view of the van der Waals interactions (gray dashed lines) of **1** stacked between Q85 and L92 of BRD4(BD1).

occupies “superenhancers” and its inhibition leads to significant reduction of the transcript levels of only a few hundred genes,¹² in a cell-, disease-, and context-specific manner. Preclinical targeting of BETs has had initial successes, particularly in oncology. For example, in a phase I acute leukemia study, (6*S*)-4-(4-chlorophenyl)-*N*-(4-hydroxyphenyl)-2,3,9-trimethyl-6*H*-thieno[3,2-*f*][1,2,4]triazolo[4,3-*a*][1,4]diazepine-6-acetamide (OTX015),^{13–16} a thienodiazepine, induced remissions, including complete remission in two patients with refractory disease, and 2-[4-(2-hydroxyethoxy)-3,5-dimethylphenyl]-5,7-dimethoxy-4(3*H*)-quinazolinone (RVX-208),^{8,17} a quinazolinone derivative of resveratrol that binds preferentially to the BD2

domain of BRD2 and BRD3, has already been tested in hundreds of patients in phase II clinical trials for the treatment of atherosclerosis, providing proof-of-concept that selective inhibition within the BET family is feasible. However, pan-BET inhibition might remain an issue regarding the impact on numerous transcriptional pathways and the individual tissue specific functions of BET members. The selective targeting of individual BET and the discrimination between BD1 and BD2 present one opportunity to achieve more selective transcriptional effects. Here, we follow-up on a previous midthroughput screen (MTS) using the 2P2I_{3D} chemical library, a structurally diverse “protein–protein interaction inhibition (2P2I)-ori-

Scheme 1. Synthesis of Substituted Xanthines^a

^aReagents and conditions: (i) EtI, DMA, 60 °C, 12 h; (ii) HCl_{aq} 3 N, reflux, 1 h, then rt NH_{3aq} 30%; (iii) NaNO₂, H₂O–AcOH, 50 °C, 1 h; (iv) R₂X, K₂CO₃, DMF, rt, 2 days.

ented” collection of compounds,^{18–23} by focusing on a selective BRD4(BD1) acetylated-mimic xanthine inhibitor. Among the 17 hits that were identified following this screen, one xanthine derivative was found to present a low micromolar IC₅₀ measured by homogeneous time-resolved fluorescence (HTRF) and an unforeseen selectivity profile among the BET family members. A structure-based investigation program around xanthine scaffold containing compounds was the starting point of this observed selectivity with a key role identified for the bromodomain ZA loop.

RESULTS AND DISCUSSION

A MTS of the 2P2I_{3D} chemical library against BRD4(BD1) identified 17 hits with affinities within the low micromolar range. We setup isothermal titration calorimetry (ITC) as an orthogonal assay and were able to validate direct binding for 7 out of the 17 compounds tested, with ligand efficiency (LE) values ranging from 0.20 up to 0.23. Among them, **1** (Figure 1A), a xanthine derivative, presented a *K_d* of 1.4 μM, driven by a large favorable entropy term ($-T\Delta S$ of -7.1 kcal/mol) as demonstrated by ITC (Figure 1B). This biophysical result suggested conformational changes and water rearrangement upon binding of the compound. Moreover, an HTRF experiment against seven out of the eight BD1 and BD2 BET domains (BRDT(BD2) was not available as single bromodomain from the company Cisbio) pinpointed a 10-fold lower IC₅₀ for **1** toward BRD4(BD1) (IC₅₀ of 5 μM) compared with its relative's BD1 (IC₅₀ > 50 μM), and no detectable inhibition of the BD2 counterparts (Figure 1C and Table 3). In cell-based assays, **1** reduced Jurkat T cell viability in a dose-dependent manner with an EC₅₀ of 27 μM. Also, **1** down-regulated c-Myc, a pro-oncogene contributing to the pathogenesis of numerous human cancers, in the same range of concentration (Figure 1D,E). We next solved the crystal structure of **1** in complex with BRD4(BD1). The cocrystal revealed the globular domain organization for BRD4(BD1)²⁴ and a well-defined electron density for **1** with the expected binding mode of an acetyl-lysine mimetic (Figure 2A,B and see also Supporting Information Figure S1),²⁵ forming the canonical hydrogen bond with the conserved asparagine N140 and a water mediated hydrogen bond with Y97 also linking the inhibitor to the conserved water network at the bottom of the binding pocket. Surprisingly, the triazolopyrimidinyl moiety stacked against the ZA channel, occupying the space at the rim of the acetyl-lysine binding site, a binding mode that has also recently been reported for benzimidazolone derivative inhibitors targeting BAZ2B bromodomain (Supporting Information Figure S2).^{26–28} This triazolo moiety establishes a hydrogen bond interaction with the main chain (NH amido group) of D88, a residue conserved throughout the BET family. This interaction orientates the triazolopyrimidinyl fragment in the ZA channel, while its ring

system is locked from one side by van der Waals contacts with L92 and from the other side with Q85; the pyrimidine ring sits tightly on this glutamine, which is bent by 90° to adopt this complementary orientation (Figure 2C). This BD1-conserved amino acid (except in BRDT(BD1)) is replaced by larger residues such as arginine or lysine in the BD2 subfamily (Supporting Information Figure S3). This difference could be responsible for the observed selectivity profile of **1** toward BRD4-, BRD3-, and BRD2(BD1) vs all the other BET proteins tested. Crystallographic data collection statistics are summarized in Supporting Information Table S1.

In order to gain better understanding of this selectivity profile, we undertook a structure-based program aiming to define the structure/affinity and structure/selectivity relationships by decomposing this model compound. First, the xanthine **2** and the corresponding mono- and unsymmetrically dialkylated derivatives **3–5** were prepared from guanosine, taking advantage of an oriented N7-alkylation as previously described in literature (Scheme 1).²⁹

These compounds were further analyzed by HTRF, ITC, and X-ray crystallography, when available (Table 1, Figure 3).

Table 1. Effect of Various Substituted Xanthine on BRD4(BD1) Activity

compd	substituent		BRD4(BD1)		
	R ₁	R ₂	IC ₅₀ ^a (μM)	<i>K_d</i> ^b (μM)	LE ^c
2	H	H	<i>d</i>		<i>d</i>
3	Et	H	<i>d</i>		<i>d</i>
4	Et	Bu	13.0 ± 0.3		0.40
5	Et	4-ClBn	2.6 ± 0.1	1.8	0.37

^aDrug concentration that inhibits protein–protein interaction by 50%. Data are the mean ± standard deviation (SD) of three experiments. ^b*K_d* determined by isothermal titration calorimetry. ^cLigand Efficiency (LE) defined as the ratio of the log of the IC₅₀ to the number (N) of non-hydrogen atoms of the compound LE = 1.4(-log(IC₅₀)/N). ^dNot applicable.

Chemical variations of the xanthine mimicking the reference probe (6S)-4-(4-chlorophenyl)-2,3,9-trimethyl-6H-thieno[3,2-*f*][1,2,4]triazolo[4,3-*a*][1,4]diazepine-6-acetic acid 1,1-dimethyl ethyl ester (JQ1, PDB accession code 3MXF),²⁴ such as a substitution of the butyl group (compound **4**) by a chlorobenzyl one (compound **5**), allowed an optimization of the core with an IC₅₀ of 2.6 μM as measured by HTRF (2-fold increased potency) and a *K_d* of 1.8 μM as measured by ITC. This optimized core presented a favorable enthalpy change of -6.3 kcal/mol (Figure 3A). More importantly, the selectivity profile of this optimized core **5** displayed a pan-BET inhibition profile, similarly to **4** (Figure 3B). Analysis of the cocrystal structures confirmed a characteristic recognition of the pocket

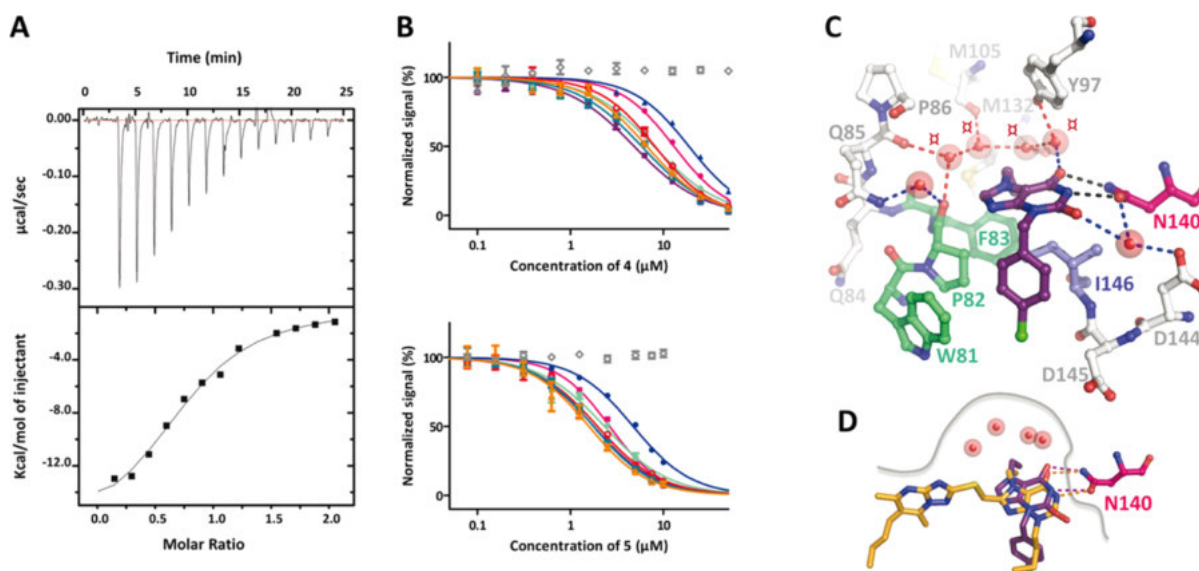
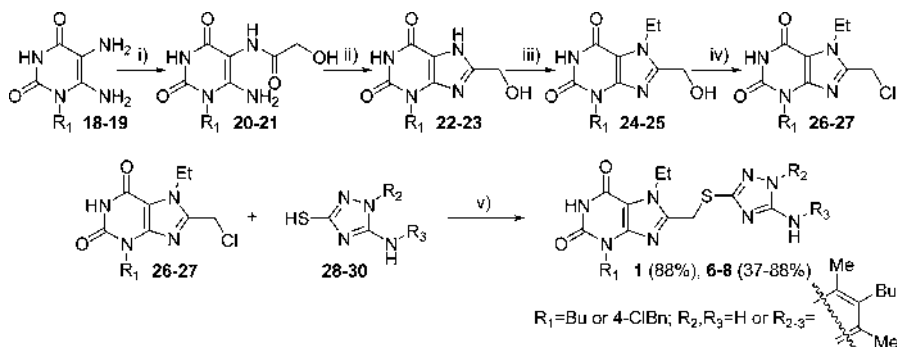


Figure 3. Studies of disubstituted xanthine derivatives as potential selective bromodomain inhibitors. (A) ITC thermogram (top) and integrated data (bottom) at 35 °C of BRD4(BD1) (100 μ M in the syringe) vs 5 (10 μ M in the cell) as a function of the molar ratio of ligand to protein. (B) HTRF selectivity assay of 4 (top) and 5 (bottom). The curve symbols and color scheme use the same pattern as in Figure 1. (C) Three dimensional detailed view of 5 (purple)/BRD4(BD1) molecular mode of action obtained from the crystal structure of the complex (obtained at 1.60 Å resolution) as a ball and stick representation as used for Figure 2. (D) Superimposed detailed view of the binding of 1 (orange) and 5 (purple) in the N-acetylated lysine pocket. The gray trace illustrates the cavity surface profile. Hydrogen bond interactions with the N140 are shown in orange and purple for 1 and 5, respectively.

Scheme 2. Design and Synthesis of Triazolosulfanylmethyl-Substituted Xanthine-Based BRD4(BD1) Inhibitors 1 and 6–8^a



^aReagents and conditions: (i) glycolic acid, 120 °C, 30 min; (ii) NaOHaq, H₂O–EtOH, 85 °C, 2 h; (iii) EtI, K₂CO₃, DMF, 0 °C to rt, 12 h; (iv) SOCl₂, CH₂Cl₂, reflux, 15 min; (v) DIPEA, DMF, rt, 30 min.

by this core, as exemplified by the superimposition of 1 with 5 (Figure 3C,D). The core was thus not responsible (as expected) for the observed selectivity profiles of 1.

We next questioned whether the observed selectivity profile of 1 was due to the triazolo or to the pyrimidinyl moiety of the core extension. To address this purpose, a series of *N*-3, *N*-7 disubstituted triazoloxanthine derivatives 1 and 6–8 were synthesized in ten steps following the synthetic route illustrated in Scheme 2.


The 3-alkyl-8-chloromethyl-7-ethylxanthine substrates 26 and 27 for reaction were prepared from corresponding 5,6-diaminouracils,³⁰ 18 and 19, by condensation with glycolic acid followed by base induced ring closure acylation. Starting from alkylamine, the condensation with potassium cyanate afforded alkylurea 10 and 11 converted into corresponding 6-amino-uracil 14 and 15 by treating with cyanoacetic acid and sodium hydroxide successively (Supporting Information Scheme S1). Subsequent nitrosylation with sodium nitrite in aqueous acetic acid, then reduction with sodium dithionite led to expected

diaminouracils 18 and 19 in good overall yields (42% and 20%, respectively).

After evaluation, compound 7 revealed an IC₅₀ of 26.4 μ M in HTRF, which represents a 10-fold decrease in potency compared with 5 (Tables 2 and 3). Compound 7, however, exhibited a shift of selectivity toward BET-BD1 domains (Figure 4A and Table 3). The introduction of the triazolo fragment thus brought a decrease of potency yet drove a commencement of selectivity among BET bromodomains (BD1 vs BD2).

A potential hydrogen bond between the triazolo fragment and D88, conserved in all the BET bromodomains, could be stabilized with certain bromodomains and not with others, due to differences in the dynamics of the ZA loop, which has been demonstrated to be crucial in the binding kinetics of several BET-BRDs.^{24,31,32} Superimposition of BRD4 with or without 1 illustrates structural rearrangements with displacement of the ZA loop up to 2.7 Å (Supporting Information Figure S4). These rearrangements are not observed with the “pan inhibitor”

Table 2. Effect of Various Triazolo-Substituted Xanthine Derivatives on BRD4(BD1) Activity

Compd	Substituent			BRD4(BD1)		
	R ₁	R ₂	R ₃	IC ₅₀ (μM) ^a	K _D (μM) ^b	LE ^c
1	Bu			5.0 ± 0.1	1.4	0.22
6	4-ClBn			20.4 ± 3.6	-	0.15
7	4-ClBn	H	H	26.4 ± 0.7	-	0.17
8	Bu			> 50	-	NA

^aDrug concentration that inhibits protein–protein interaction by 50%. Data are the mean ± standard deviation (SD) of three experiments. ^bK_D determined by isothermal titration calorimetry. ^cLigand efficiency (LE) defined as the ratio of the IC₅₀ to the number (N) of non-hydrogen atoms of the compound LE = 1.4(−log IC₅₀)/N.

5, thus validating this assumption. The role of this loop and its associated variable dynamics would also explain the IC₅₀ observed for **1** toward BRD4(BD1) vs BRD3(BD1) and BRD2(BD1), this hydrogen bond being more stable with BRD4(BD1). Analysis of the protein–ligand(s) interactions including hydrophobic and hydrogen-bonding contacts was carried out using LigPlot+,³³ as shown in Supporting Information Figure S5. This weak interaction prevented us from solving a nonambiguous 3D structure of **7** in complex with BRD4(BD1). However, crystals with 50% occupancy of **7** highlighted that this compound was still interacting with BRD4(BD1) and that hydrogen bond between the triazolo moiety and the main chain (NH amido group) of D88 was present, with BRD4(BD1) (Figure 4B,C). The triazolo was thus explaining only partly the observed selectivity profiles of **1**. An explanation for the selectivity profile of **1** was finally confirmed with the synthesis of **6**, the *N*-chlorobenzyl analogue of **1**. Of note, this compound was found to be less active than the original compound **1** (IC₅₀ values of 20 μM vs 5 μM, respectively). However, the selectivity profile of this compound was similar to that of **1** (Figure 4A), that is, selective toward BRD4(BD1). Altogether, these data suggest that the *N*₃,*N*₇-dialkylated xanthine core (see **4** and **5**) is responsible for the pan-BET inhibition, while condensed with triazolo moiety (see **7**) it allows the preferential inhibition of BET-BD1 domains forming a hydrogen bond with the conserved D88 residue. Further condensation with pyrimidinyl moiety drives the corresponding triazolopyrimidinyl xanthine derivatives (**1** and **6**) to be selective toward BRD4(BD1) due to van der Waals

interaction with the side chain of Q85. A steric clash prevents this interaction with the RS4, its equivalent residue in BRD3(BD1) numbering, or the K present in the other BRDX(BD2) bromodomains. These findings suggest a potential mechanism for the selectivity of BET bromodomains. However, a more comprehensive evaluation of the role of the first bromodomain in cancer cell proliferation and chromatin biology will require the development of more selective probes with improved selectivity.

CONCLUSION

A permanent wavering in drug discovery is related to the development of pan- (multitargeted) or selective (single targeted) inhibitors. In this study, we have investigated the key structural feature responsible for the selectivity of a xanthine-based BRD4(BD1) inhibitor identified through MTS. This compound represents the first low micromolar selective inhibitor targeting BRD4(BD1) with a >10-fold ratio in binding affinity toward any other BET bromodomain tested, yet presenting low but dose–response down regulation of c-Myc levels in cell-based assay. The cocrystal structure revealed an original orientation of the triazolopyrimidinyl moiety, expanding into the ZA channel, setting the basis of a structure–selectivity relationship inside the BET family. Even though it is conserved throughout the BET family, the different dynamic behavior of the ZA loop could be exploited to tune the selectivity of BET inhibitors. Optimization of the selective BRD4(BD1) inhibitor to a probe has not been achieved yet. However, the structural features identified in this study will allow further optimization of more selective BRD4(BD1) inhibitors. The generation of such probes will allow a genuine evaluation of the biological role for each bromodomain and a validation (or not) of the interest toward the development of selective vs pan-BET inhibitors in the clinic.

EXPERIMENTAL SECTION

General Methods. All solvents were purified according to reported procedures, and the commercially available reagents were used as received. Separation by column chromatography was performed using SDS Kieselgel (70–230 mesh). Petroleum refers to the fraction with distillation range 40–65 °C. ¹H and ¹³C NMR spectra were recorded using a Bruker AC400 or AC250 spectrometer. Chemical shifts, δ, are reported in ppm and coupling values, *J*, in hertz. Abbreviations for peaks are br = broad, s = singlet, d = doublet, t = triplet, q = quadruplet, quint = quintuplet, sex = sextuplet, and m = multiplet. Reaction monitoring and purity of compounds were recorded by using analytical Agilent Infinity high performance liquid

Table 3. Selectivity Profiles of Xanthine Derivatives toward BET Family Members

bromodomain	IC ₅₀ ^a (μM)				
	1	4	5	6 ^c	7
BRD4(BD1)	5.0 ± 0.1	13.0 ± 0.3	2.6 ± 0.1	20.4 ± 3.6	26.4 ± 0.7
BRD3(BD1)	62.0 ± 7.8	4.8 ± 0.3	1.8 ± 1.1	<i>d</i>	31.7 ± 1.1
BRD2(BD1)	59.0 ± 2.9	18.8 ± 0.7	4.7 ± 0.1	<i>d</i>	30.0 ± 0.9
BRDT(BD1)	>100	7.9 ± 0.4	2.3 ± 0.2	<i>d</i>	>100
BRD4(BD2)	<i>d</i>	8.1 ± 0.3	1.9 ± 0.1	<i>d</i>	<i>d,e</i>
BRD3(BD2)	<i>d</i>	5.8 ± 0.2	1.8 ± 0.2	<i>d</i>	<i>d,e</i>
BRD2(BD2)	<i>d</i>	6.6 ± 0.3	1.6 ± 0.2	<i>d</i>	<i>d,e</i>
ATAD2 ^b	>100	<i>d</i>	<i>d</i>	<i>d</i>	<i>d</i>

^aDrug concentration that inhibits protein–protein interaction by 50%. Data are the mean ± standard deviation (SD) of three experiments. ^bATAD2 is used as a non-BET family member bromodomain control. ^cData above 25 μM have been excluded due to precipitation and fluorescence interference at 620 nm. ^dNot converged. ^eFluorescence interference at 620 nm with BD2.

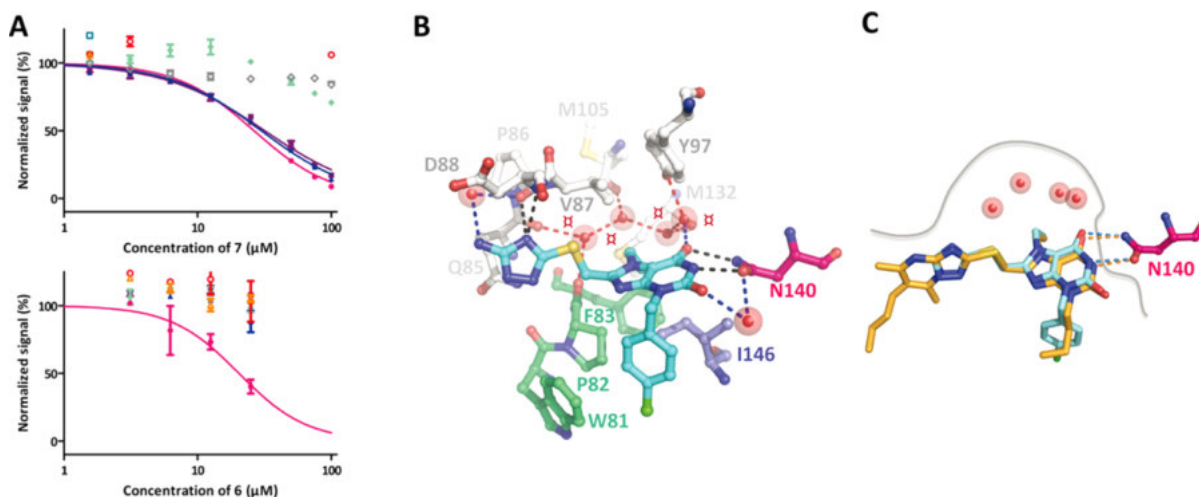


Figure 4. Structure–activity relationship of some triazolo-substituted xanthine derivatives. (A) HTRF selectivity assay results of **7** (top) and **6** (bottom). The curve symbols and color scheme use the same pattern as in Figure 1. For compound **6**, data above 25 μM have been excluded due to precipitation and fluorescence interference at 620 nm. (B) Three dimensional detailed view of **7** (cyan)/BRD4(BD1) molecular mode of action obtained from the crystal structure of the complex (1.05 Å resolution) as a ball and stick representation as used for Figure 2. (C) Superimposed detailed view of the binding of **1** (orange) and **7** (cyan) in the N-acetylated lysine pocket. Hydrogen bond interactions with the N140 are shown in orange and cyan for **1** and **7**, respectively.

chromatography (Column Zorbax SB-C18 1.8 μM (2.1 mm \times 50 mm); mobile phase A 0.1% FA H_2O , B 0.1% FA MeCN, time/%B 0/10, 4/90, 7/90, 9/10, 10/10; flow rate 0.3 mL/min; diluent MeOH) with DAD at 230 nm. All tested compounds yielded data consistent with a purity of $\geq 95\%$. Low-resolution mass spectra were obtained with Agilent SQ G6120B mass spectrometer in positive electrospray mode. High-resolution mass spectra were performed at the Spectropole Analytical Laboratory of the University Paul Cezanne, Marseille.

General Procedure for the Synthesis of 8-Triazolosulfanylmethylxanthine Derivatives 1 and 6–8. To a solution of mercapto triazole derivative (0.15 mmol) in dimethylformamide (2 mL) was injected diisopropylethylamine (26 μL , 1 equiv). After 5 min under stirring, a solution of appropriate 3-alkyl-8-chloromethyl-7-ethyl-3,7-dihydro-purine-2,6-dione (1 equiv) in dimethylformamide (2 mL) was added dropwise. The resulting mixture was stirred at room temperature (30 $^\circ\text{C}$) for 30 min. The solvent was distilled off under reduced pressure, and the residue was purified by column chromatography (CH_2Cl_2 –MeOH) to afford the corresponding 3-alkyl-8-([1,2,4]triazolo-3-yl-sulfanylmethyl)-7-ethyl-3,7-dihydro-purine-2,6-dione.

3-Butyl-8-(6-butyl-5,7-dimethyl-[1,2,4]triazolo[1,5-a]pyrimidin-2-ylsulfanylmethyl)-7-ethyl-3,7-dihydro-purine-2,6-dione, 1. Yield 88% as white solid. $R_f = 0.29$ (CH_2Cl_2 –MeOH 10:0.5). ^1H NMR (250 MHz, CDCl_3) δ 7.88 (1H, sbroad), 4.73 (2H, s), 4.49 (2H, q, $J = 7.1$ Hz), 4.04 (2H, t, $J = 7.4$ Hz), 2.76 (3H, s), 2.73–2.69 (2H, m), 2.69 (3H, s), 1.73 (2H, quint, $J = 7.4$ Hz), 1.56–1.34 (6H, m), 1.40–1.28 (2H, m), 1.45 (3H, t, $J = 7.4$ Hz), 1.01 (3H, t, $J = 7.1$ Hz) and 0.95 (3H, t, $J = 7.4$ Hz). ^{13}C NMR (100 MHz, CDCl_3) δ 165.0, 164.1, 154.1, 153.8, 150.6, 150.1, 149.2, 143.5, 121.8, 42.7, 41.4, 31.9, 30.2, 28.0, 26.8, 23.8, 22.9, 20.0, 16.6, 14.0, 13.9, and 13.8. LCMS $\text{C}_{23}\text{H}_{32}\text{N}_8\text{O}_2\text{S}$ $R_t = 6.755$ min, $m/z = 484.6$, purity $>99\%$. HRMS (ESI+) for $\text{C}_{23}\text{H}_{33}\text{N}_8\text{O}_2\text{S}$ (M + H) calcd, 485.2442; found, 485.2442.

8-(6-Butyl-5,7-dimethyl-[1,2,4]triazolo[1,5-a]pyrimidin-2-ylsulfanylmethyl)-3-(4-chlorobenzyl)-7-ethyl-3,7-dihydro-purine-2,6-dione, 6. Yield 37% as light yellow solid. $R_f = 0.35$ (EtOAc). ^1H NMR (400 MHz, CDCl_3) δ 7.48 (2H, d, $J = 8.5$ Hz), 7.23 (2H, d, $J = 8.5$ Hz), 5.15 (2H, s), 4.71 (2H, s), 4.47 (2H, q, $J = 7.2$ Hz), 2.75 (3H, s), 2.73–2.70 (2H, m), 2.69 (3H, s), 1.54–1.49 (4H, m), 1.45 (3H, t, $J = 7.2$ Hz) and 1.00 (3H, t, $J = 7.2$ Hz). ^{13}C NMR (100 MHz, CDCl_3) δ 164.8, 164.0, 153.8, 153.7, 150.4, 149.5, 149.2, 143.4, 134.7, 133.8, 130.7, 128.6, 121.8, 107.6, 45.1, 41.3, 31.8, 27.9, 26.7, 23.7, 22.8, 16.4, 13.9, and 13.8. LCMS $\text{C}_{26}\text{H}_{29}\text{ClN}_8\text{O}_2\text{S}$ $R_t = 6.728$ min, $m/z = 552.5$, purity $>99\%$. HRMS (ESI+) for $\text{C}_{26}\text{H}_{30}\text{ClN}_8\text{O}_2\text{S}$ (M + H) calcd, 553.1895; found, 553.1898.

8-(5-Amino-1H-[1,2,4]triazol-3-ylsulfanylmethyl)-3-(4-chlorobenzyl)-7-ethyl-3,7-dihydro-purine-2,6-dione, 7. Yield 62% as yellow solid. $R_f = 0.18$ (EtOAc). ^1H NMR (250 MHz, MeOD) δ 7.38 (2H, d, $J = 8.5$ Hz), 7.29 (2H, d, $J = 8.5$ Hz), 5.12 (2H, s), 4.42 (2H, s), 4.38 (2H, q, $J = 7.2$ Hz), and 1.42 (3H, t, $J = 7.1$ Hz). ^{13}C NMR (63 MHz, MeOD) δ 156.3, 152.7, 151.5, 150.7, 136.8, 134.5, 131.1, 129.6, 108.8, 45.9, 42.2, 31.7, 16.6. LCMS $\text{C}_{17}\text{H}_{17}\text{ClN}_8\text{O}_2\text{S}$ $R_t = 4.838$ min, $m/z = 432.5$, purity $>96\%$. HRMS (ESI+) for $\text{C}_{17}\text{H}_{18}\text{ClN}_8\text{O}_2\text{S}$ (M + H) calcd, 433.0956; found, 433.0956.

8-(5-Amino-1H-[1,2,4]triazol-3-ylsulfanylmethyl)-3-butyl-7-ethyl-3,7-dihydro-purine-2,6-dione, 8. Yield 82% as light yellow powder. $R_f = 0.40$ (EtOAc–MeOH 6:1). ^1H NMR (250 MHz, DMSO- d_6) δ 12.08 (1H, sbroad), 11.07 (1H, sbroad), 6.15 (2H, sbroad), 4.41 (2H, s), 4.28 (2H, q, $J = 7.1$ Hz), 3.85 (2H, t, $J = 7.3$ Hz), 1.58 (2H, quint, $J = 7.3$ Hz), 1.34–1.19 (5H, m) and 0.89 (3H, t, $J = 7.3$ Hz). ^{13}C NMR (100 MHz, DMSO- d_6) δ 154.6, 150.8, 149.6, 149.1, 106.9, 41.6, 40.6, 29.9, 26.8, 19.6, 16.3, and 13.9. LCMS $\text{C}_{14}\text{H}_{20}\text{N}_8\text{O}_2\text{S}$ $R_t = 4.841$ min, $m/z = 364.6$, purity $>96\%$. HRMS (ESI+) for $\text{C}_{14}\text{H}_{21}\text{N}_8\text{O}_2\text{S}$ (M + H) calcd, 365.1503; found, 365.1512.

■ ASSOCIATED CONTENT

📄 Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.jmedchem.5b01708.

Additional figures, schemes, and tables and detailed experimental procedures for biochemistry, biology, and chemistry including spectral data and structures of synthesized compounds (PDF)

SMILES representations of compounds with associated data (CSV)

Accession Codes

PDB codes for BRD4(BD1) with bound **1**, **5** and **7**, respectively, are SEGU, SEIS, and SEI4.

■ AUTHOR INFORMATION

Corresponding Authors

*S.C. E-mail: sebastien.combes@univ-amu.fr. Phone: (+33) 491-835-551.

*X.M. E-mail: xavier.morelli@inserm.fr. Phone: (+33)486-977-331. Mailing address: Centre de Recherche en Cancérologie de Marseille (CRCM); INSERM U1068; CNRS UMR7258; Aix-Marseille Université UM105 & Institut Paoli Calmettes, 27 Boulevard Leï Roure CS30059, 13273 Marseille Cedex 9, France.

Present Address

○A.L.: Bioaster; 40 Avenue Tony Garnier, 69007 Lyon, France.

Author Contributions

B.R. and Y.V. contributed equally to this work as first author. S.C. and X.M. contributed equally to this work as senior author.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

This work was supported by grants from the French Infrastructure for Integrated Structural Biology (FRISBI) ANR-10-INSB-05-01, from the “Region PACA”, and from “fond FEDER” 1860-38025 CPER 2007-2013. Sabine Milhas was the recipient of a Ph.D. fellowship from Region PACA and from the “Fondation pour la Recherche Medicale” (FRM). Yuliia Voitovich was a recipient of a “Metchnikov” Ph.D. fellowship from French government. The authors acknowledge the European Synchrotron Radiation Facility for provision of synchrotron radiation facilities (under the beam time application MX1696) and thank Antoine Royant and David Flot for assistance in using beamline ID29 and ID23-2, respectively. We also thank Bruno Canard for access to facilities at UMR7257 CNRS-AMU, Marseille, and Valérie Depraetere-Ferrier for manuscript editing.

ABBREVIATIONS

BRD, bromodomain; BET, bromodomain and extra-terminal domain; MTS, midthroughput screening; HTRF, homogeneous time-resolved fluorescence; ITC, isothermal titration calorimetry

REFERENCES

- (1) Filippakopoulos, P.; Knapp, S. The bromodomain interaction module. *FEBS Lett.* **2012**, *586*, 2692–2704.
- (2) Filippakopoulos, P.; Picaud, S.; Mangos, M.; Keates, T.; Lambert, J.-P.; Baryshte-Lovejoy, D.; Felletar, I.; Volkmer, R.; Muller, S.; Pawson, T.; Gingras, A.-C.; Arrowsmith, C. H.; Knapp, S. Histone recognition and large-scale structural analysis of the human bromodomain family. *Cell* **2012**, *149*, 214–231.
- (3) Hewings, D. S.; Rooney, T. P.; Jennings, L. E.; Hay, D. A.; Schofield, C. J.; Brennan, P. E.; Knapp, S.; Conway, S. J. Progress in the development and application of small molecule inhibitors of bromodomain-acetyl-lysine interactions. *J. Med. Chem.* **2012**, *55*, 9393–9413.
- (4) Kanno, T.; Kanno, Y.; Siegel, R. M.; Jang, M. K.; Lenardo, M. J.; Ozato, K. Selective recognition of acetylated histones by bromodomain proteins visualized in living cells. *Mol. Cell* **2004**, *13*, 33–43.
- (5) Huang, H.; Zhang, J.; Shen, W.; Wang, X.; Wu, J.; Shi, Y. Solution structure of the second bromodomain of Brd2 and its specific interaction with acetylated histone tails. *BMC Struct. Biol.* **2007**, *7*, 57.
- (6) LeRoy, G.; Rickards, B.; Flint, S. J. The double bromodomain proteins Brd2 and Brd3 couple histone acetylation to transcription. *Mol. Cell* **2008**, *30*, 51–60.
- (7) Baud, M. G.; Lin-Shiao, E.; Cardote, T.; Tallant, C.; Pschibul, A.; Chan, K. H.; Zengerle, M.; Garcia, J. R.; Kwan, T. T.; Ferguson, F. M.; Ciulli, A. Chemical biology. A bump-and-hole approach to engineer controlled selectivity of BET bromodomain chemical probes. *Science* **2014**, *346*, 638–641.
- (8) Picaud, S.; Wells, C.; Felletar, I.; Brotherton, D.; Martin, S.; Savitsky, P.; Diez-Dacal, B.; Philpott, M.; Bountra, C.; Lingard, H.; Fedorov, O.; Müller, S.; Brennan, P. E.; Knapp, S.; Filippakopoulos, P. RVX-208, an inhibitor of BET transcriptional regulators with selectivity for the second bromodomain. *Proc. Natl. Acad. Sci. U. S. A.* **2013**, *110*, 19754–19759.
- (9) Schröder, S.; Cho, S.; Zeng, L.; Zhang, Q.; Kaehlcke, K.; Mak, L.; Lau, J.; Biggrove, D.; Schnölzer, M.; Verdin, E.; Zhou, M.-M.; Ott, M. Two-pronged binding with bromodomain-containing protein 4 liberates positive transcription elongation factor b from inactive ribonucleoprotein complexes. *J. Biol. Chem.* **2012**, *287*, 1090–1099.
- (10) Rahl, P. B.; Lin, C. Y.; Seila, A. C.; Flynn, R. A.; McCuine, S.; Burge, C. B.; Sharp, P. A.; Young, R. A. c-Myc regulates transcriptional pause release. *Cell* **2010**, *141*, 432–445.
- (11) Belkina, A. C.; Denis, G. V. BET domain co-regulators in obesity, inflammation and cancer. *Nat. Rev. Cancer* **2012**, *12*, 465–477.
- (12) Lovén, J.; Hoke, H. A.; Lin, C. Y.; Lau, A.; Orlando, D. A.; Vakoc, C. R.; Bradner, J. E.; Lee, T. I.; Young, R. A. Selective inhibition of tumor oncogenes by disruption of super-enhancers. *Cell* **2013**, *153*, 320–334.
- (13) Boi, M.; Bonetti, P.; Ponzoni, M.; Tibiletti, M. G.; Stathis, A.; Cvitkovic, E.; Inghirami, G.; Zucca, E.; Bertoni, F. The brd-Inhibitor OTX015 shows pre-clinical activity in anaplastic large T-cell lymphoma (ALCL). *ASH Annu. Meet. Abstr., Blood* **2012**, *120*, 4872.
- (14) Bonetti, P.; Boi, M.; Ponzoni, M.; Tibiletti, M. G.; Stathis, A.; Inghirami, G.; Noel, K.; Zucca, E.; Bertoni, F. *ASH Annu. Meet. Abstr., Blood* **2012**, *120*, 1657.
- (15) Braun, T.; Coude, M.; Berrou, J.; Bertrand, S.; Riveiro, E.; Herait, P.; Baruchel, A.; Dombret, H.; Gardin, C. Preclinical study of the bromodomain inhibitor OTX015 in acute myeloid (AML) and lymphoid (ALL) leukemias. *ASH Annu. Meet. Abstr., Blood* **2013**, *122*, 4218.
- (16) Noel, J. K.; Iwata, K.; Ooike, S.; Sugahara, K.; Nakamura, H.; Daibata, M. Development of the BET bromodomain inhibitor OTX015. *Mol. Cancer Ther.* **2013**, *12*, C244.
- (17) Jahagirdar, R.; Zhang, H.; Azhar, S.; Tobin, J.; Attwell, S.; Yu, R.; Wu, J.; McLure, K. G.; Hansen, H. C.; Wagner, G. S.; Young, P. R.; Srivastava, R. A. K.; Wong, N. C. W.; Johansson, J. A novel BET bromodomain inhibitor, RVX-208, shows reduction of atherosclerosis in hyperlipidemic ApoE deficient mice. *Atherosclerosis* **2014**, *236*, 91–100.
- (18) Bourgeas, R.; Basse, M.-J.; Morelli, X.; Roche, P. Atomic analysis of protein-protein interfaces with known inhibitors: the 2P2I database. *PLoS One* **2010**, *5*, e9598.
- (19) Basse, M.-J.; Betzi, S.; Bourgeas, R.; Bouzidi, S.; Chetrit, B.; Hamon, V.; Morelli, X.; Roche, P. 2P2Idb: a structural database dedicated to orthosteric modulation of protein-protein interactions. *Nucleic Acids Res.* **2013**, *41*, D824–327, <http://2p2idb.cnrs-mrs.fr>.
- (20) Morelli, X.; Bourgeas, R.; Roche, P. Chemical and structural lessons from recent successes in protein-protein interaction inhibition (2P2I). *Curr. Opin. Chem. Biol.* **2011**, *15*, 475–481.
- (21) Hamon, V.; Bourgeas, R.; Ducrot, P.; Theret, I.; Xuereb, L.; Basse, M.-J.; Brunel, J.-M.; Combes, S.; Morelli, X.; Roche, P. 2P2I HUNTER: a tool for filtering orthosteric protein-protein interaction modulators via a dedicated support vector machine. *J. R. Soc., Interface* **2014**, *11*, 20130860.
- (22) Hamon, V.; Brunel, J. M.; Combes, S.; Basse, J.-M.; Roche, P.; Morelli, X. 2P2Ichem: focused chemical libraries dedicated to orthosteric modulation of protein-protein interactions. *MedChem-Comm* **2013**, *4*, 797–809.
- (23) Zhang, X.; Betzi, S.; Morelli, X.; Roche, P. Focused chemical libraries—design and enrichment: an example of protein-protein interaction chemical space. *Future Med. Chem.* **2014**, *6*, 1291–1307.
- (24) Filippakopoulos, P.; Qi, J.; Picaud, S.; Shen, Y.; Smith, W. B.; Fedorov, O.; Morse, E. M.; Keates, T.; Hickman, T. T.; Felletar, I.; Philpott, M.; Munro, S.; McKeown, M. R.; Wang, Y.; Christie, A. L.; West, N.; Cameron, M. J.; Schwartz, B.; Heightman, T. D.; La Thangue, N.; French, C. A.; Wiest, O.; Kung, A. L.; Knapp, S.;

Bradner, J. E. Selective inhibition of BET bromodomains. *Nature* **2010**, *468*, 1067–1073.

(25) Vollmuth, F.; Blankenfeldt, W.; Geyer, M. Structures of the dual bromodomains of the P-TEFb-activating protein Brd4 at atomic resolution. *J. Biol. Chem.* **2009**, *284*, 36547–36556.

(26) Drouin, L.; McGrath, S.; Vidler, L. R.; Chaikuad, A.; Monteiro, O.; Tallant, C.; Philpott, M.; Rogers, C.; Fedorov, O.; Liu, M.; Akhtar, W.; Hayes, A.; Raynaud, F.; Müller, S.; Knapp, S.; Hoelder, S. Structure enabled design of BAZ2-ICR, a chemical probe targeting the bromodomains of BAZ2A and BAZ2B. *J. Med. Chem.* **2015**, *58*, 2553–2559.

(27) Bennett, J.; Fedorov, O.; Tallant, C.; Monteiro, O.; Meier, J.; Gamble, V.; Savitsky, P.; Nunez-Alonso, G. A.; Haendler, B.; Rogers, C.; Brennan, P. E.; Müller, S.; Knapp, S. Discovery of a chemical tool inhibitor targeting the bromodomains of TRIM24 and BRPF. *J. Med. Chem.* **2015**, DOI: [10.1021/acs.jmedchem.5b00458](https://doi.org/10.1021/acs.jmedchem.5b00458).

(28) Palmer, W. S.; Poncet-Montange, G.; Liu, G.; Petrocchi, A.; Reyna, N.; Subramanian, G.; Theroff, J.; Yau, A.; Kost-Alimova, M.; Bardenhagen, J. P.; Leo, E.; Shepard, H. E.; Tieu, T. N.; Shi, X.; Zhan, Y.; Zhao, S.; Barton, M. C.; Draetta, G.; Toniatti, C.; Jones, P.; Geck Do, M.; Andersen, J. N. Structure-guided design of IACS-9571, a selective high-affinity dual TRIM24-BRPF1 bromodomain inhibitor. *J. Med. Chem.* **2015**, DOI: [10.1021/acs.jmedchem.5b00405](https://doi.org/10.1021/acs.jmedchem.5b00405).

(29) Vidal, A.; Giraud, I.; Madelmont, J.-C. Improved synthesis of 7-(alkyl/aryl)guanine. *Synth. Commun.* **2004**, *34*, 3359–3365.

(30) Kalla, R. V.; Elzein, E.; Perry, T.; Li, X.; Gimbel, A.; Yang, M.; Zeng, D.; Zablocki, J. Selective, high affinity A2B adenosine receptor antagonists: N-1 monosubstituted 8-(pyrazol-4-yl)xanthines. *Bioorg. Med. Chem. Lett.* **2008**, *18*, 1397–1401.

(31) Kuang, M.; Zhou, J.; Wang, L.; Liu, Z.; Guo, J.; Wu, R. Binding Kinetics versus Affinities in BRD4 Inhibition. *J. Chem. Inf. Model.* **2015**, *55*, 1926–1935.

(32) Ran, T.; Zhang, Z.; Liu, K.; Lu, Y.; Li, H.; Xu, J.; Xiong, X.; Zhang, Y.; Xu, A.; Lu, S.; Liu, H.; Lu, T.; Chen, Y. Insight into the key interactions of bromodomain inhibitors based on molecular docking, interaction fingerprinting, molecular dynamics and binding free energy calculation. *Mol. BioSyst.* **2015**, *11*, 1295–1304.

(33) Laskowski, R. A.; Swindells, M. B. LigPlot+: multiple ligand-protein interaction diagrams for drug discovery. *J. Chem. Inf. Model.* **2011**, *51*, 2778–2786.

Exploring Selectivity of the First Bromo and Extra-Terminal (BET) Bromodomain Inhibition

Brigitt Raux, Yuliia Voitovich, Carine Derviaux, Adrien Lugari, Etienne Rebuffet, Sabine Milhas, Stéphane Priet, Thomas Roux, Eric Trinquet, Jean-Claude Guillemot, Stefan Knapp, Jean-Michel Brunel, Alexey Yu. Fedorov, Yves Collette, Philippe Roche, Stéphane Betzi, Sébastien Combes and Xavier Morelli

TABLE OF CONTENTS

Experimental procedures and compounds characterization	2
HTRF screen	4
Protein expression and purification	4
Bromodomain selectivity profiles	4
Cells and cell culture	4
Cytotoxicity experiments	4
Western blot	4
Isothermal titration calorimetry	5
Crystallography	5
¹H and ¹³C NMR spectra of triazolo-substituted xanthine-based BET inhibitors	6
Schemes S1-S2	10
Figures S1-S5	11
Tables S1-S3	15
References	18

Experimental procedures and compounds characterization

General Methods for Chemistry. All solvents were purified according to reported procedures, and the reagents commercially available were used as received. Separation by column chromatography was performed using SDS Kieselgel (70-230 mesh). Petroleum refers to the fraction with distillation range 40-65 °C. ¹H and ¹³C NMR spectra were recorded by using a Bruker AC400 or AC250 spectrometer. Chemical shifts, δ are reported in ppm and coupling values, J , in hertz. Abbreviations for peaks are, br: broad, s: singlet, d: doublet, t: triplet, q: quadruplet, quint: quintuplet, sex: sextuplet and m: multiplet. Reaction monitoring and purity of compounds were recorded by using analytical Agilent Infinity high performance liquid chromatography (Column Zorbax SB-C18 1.8 μ M (2.1x50 mm); Mobile phase (A: 0.1% FA H₂O, B: 0.1% FA MeCN, Time/%B 0/10, 4/90, 7/90, 9/10, 10/10); Flow rate 0.3 mL/min.; Diluent MeOH) with DAD at 230 nM. All tested compounds yielded data consistent with a purity of \geq 95%. Low-resolution mass spectra were obtained with Agilent SQ G6120B mass spectrometer in positive electrospray mode. High-resolution mass spectra were performed at the Spectropole Analytical Laboratory of the University Paul Cezanne, Marseille.

General procedure for the synthesis of xanthines 3 to 5. *N*-7-mono- and *N*-3,*N*-7-dialkylated xanthines **3** and **4-5**, respectively, were prepared according to a modified procedure previously described in the literature.¹ Under argon, a suspension of guanosine (10 g, 35.5 mmol) and ethyl iodide (6.8 mL, 2.4eq) in dimethylacetamide (100 mL) was heated at 60 °C overnight. Then reaction mixture was diluted with aqueous 3N HCl (90 mL) and heated at 110 °C for 1 h. The resulting solution was then treated with aqueous 10% NH₃ until pH>8 to allow the precipitation. The solid was collected by filtration and washed with Et₂O to afford 7-ethylguanine² (6.16 g, 95%) as light brown powder. ¹H NMR (250 MHz, DMSO-*d*₆) δ 7.95 (1H, s), 4.17 (2H, q, J = 7.0 Hz) and 1.32 (3H, t, J = 7.0 Hz); ¹³C NMR (63 MHz, DMSO-*d*₆) δ 155.6, 151.5, 149.7, 142.1, 106.2, 41.5 and 16.4. **7-Ethylxanthine 3.** At 50 °C, to a solution of 7-ethylguanine (6.2 g, 34.6 mmol) in a mixture of acetic acid (50 mL) and water (5 mL) was added dropwise a solution of sodium nitrite (9.6 g, 4eq) in water (15 mL) over 15 min. After addition the reaction mixture was stirred for 1 h then concentrated under reduced pressure. The residue was poured into cold H₂O (50 mL) and the precipitate collected by filtration was air dried to afford 7-ethylxanthine **3** (5.53 g, 89%) as colorless solid. ¹H NMR (250 MHz, DMSO-*d*₆) δ 11.54 (1H, sbroad), 10.86 (1H, sbroad), 7.97 (1H, s), 4.20 (2H, q, J = 7.1 Hz) and 1.37 (3H, t, J = 7.1 Hz). HRMS (ESI+) for C₇H₈N₄O₂ (M+H) calcd, 181.0720; found, 181.0720. A suspension of **3** (3.61 g, 20 mmol), potassium carbonate (3.04 g, 1.1eq) and appropriate alkyl iodide or bromide (1eq) in dimethylformamide (300 mL) was stirred at room temperature for 5h. The solvent was distilled off under reduced pressure and the residue was successively diluted with H₂O (50 mL), acidified with aqueous 1N HCl until pH=6, and extracted with CH₂Cl₂ (3x30 mL). The combined organic layers were dried over Na₂SO₄ and concentrated under reduced pressure until few volume (15 mL). Et₂O (100 mL) was added to the residue to allow crystallization and the precipitate was collected by filtration affording the expected compound.

3-Butyl-7-ethyl xanthine 4 (55%) as white solid. ¹H NMR (400 MHz, CDCl₃) δ 9.14 (1H, sbroad), 7.60 (1H, s), 4.33 (2H, q, J = 7.2 Hz), 4.08 (2H, t, J = 7.5 Hz), 1.75 (2H, quint, J = 7.5 Hz), 1.53 (3H, t, J = 7.2 Hz), 1.90 (2H, sex, J = 7.5 Hz) and 0.95 (3H, t, J = 7.5 Hz); ¹³C NMR (100 MHz, CDCl₃) δ 154.7, 150.9, 150.7, 140.7, 107.3, 42.7, 42.4, 30.2, 19.9, 16.4 and 13.7. HRMS (ESI+) for C₁₁H₁₇N₄O₂ (M+H) calcd, 237.1346; found, 237.1345.

3-(4-Chlorobenzyl)-7-ethyl xanthine 5 (83%) as light yellow solid. ¹H NMR (250 MHz, CDCl₃) δ 8.46 (1H, sbroad), 7.59 (1H, s), 7.47 (2H, d, J = 8.4 Hz) 7.27 (2H, d, J = 8.4 Hz), 5.21 (2H, s), 4.32 (2H, q, J = 7.0 Hz) and 1.53 (3H, t, J = 7.0 Hz); ¹³C NMR (100 MHz, CDCl₃) δ 154.2, 150.7, 150.4, 140.7, 134.7, 133.8, 130.3, 128.7, 107.3, 45.2, 42.5 and 16.4. HRMS (ESI+) for C₁₄H₁₄N₄O₂Cl (M+H) calcd, 305.0800; found, 305.0800.

General procedure for the synthesis of 5,6-diaminouracils 18 and 19. 5,6-diaminouracil **18** and **19** were prepared according to a modified procedure previously described in the literature.^{3,4} **Butylurea 10.** At 0 °C, to butylamine (10.73 g, 147 mmol) was added concentrated hydrochloric acid (14.7 ml, 1.2eq) and the mixture was poured into hot ethanol (150 mL). The resulting solution was then added to a solution of potassium cyanate (14.61 g, 1.2eq) in water (150 mL) and stirring was maintained at room temperature overnight. The reaction mixture was concentrated under reduced pressure until few volume to allow butylurea **10**⁵ to crystallize (15 g, 88%) as white plates. ¹H NMR (250 MHz, DMSO-*d*₆) δ 6.02 (1H, t, J = 6.7 Hz), 5.41 (2H, sbroad), 2.92 (2H, q, J = 6.7 Hz), 1.34-1.19 (4H, m) and 0.84 (3H, t, J = 6.7 Hz); ¹³C NMR (63 MHz, DMSO-*d*₆) δ 160.0, 39.5, 32.4, 20.1 and 14.3. ***N*-Butyl-*N'*-cyanoacetylurea 12.** At 0 °C, to a solution of cyanoacetic acid (10.18 g, 120 mmol) in acetic anhydride (22.3 mL) was added **10** (13.88 g, 1eq) by solid fraction over 5 min. The mixture was successively stirred at 0 °C for 15 min., allowed to warm at room temperature, heated at 60 °C for 30 min, until completion of the reaction, then allowed to cool at room temperature. To the resulting precipitate was added Et₂O (60 mL) and the solid was collected by filtration to afford *N*-butyl-*N'*-cyanoacetylurea **12** (18.5 g, 84%) as fine white needles. ¹H NMR (250 MHz, DMSO-*d*₆) δ 10.53 (1H, sbroad), 7.96 (1H, t, J = 7.4 Hz), 3.91 (2H, s), 3.14 (2H, q, J = 7.4 Hz), 1.44 (2H, quint, J = 7.4 Hz), 1.28 (2H, sex, J = 7.4 Hz), and 0.88 (3H, t, J = 7.4 Hz). **6-Amino-1-butyluracil 14.** At 85 °C, to a suspension of **12** (13.5 g, 73.8 mmol) in a mixture of water (30 mL) and ethanol (50 mL) was added dropwise a 10% aqueous sodium hydroxide (10 mL) over a period of 15 min. After addition, the resulting mixture was stirred at 85 °C for 45 min. The reaction solution was concentrated under reduced pressure to allow crystallization. The solid was collected by filtration and fully washed with Et₂O to afford 6-amino-1-butyluracil **14**⁶ (13 g, 96%) as white powder. ¹H NMR (250 MHz, DMSO-*d*₆) δ 10.29 (1H, sbroad), 6.77 (2H, sbroad), 4.52 (1H, s), 3.71 (2H, t, J = 7.3 Hz), 1.46 (2H, quint, J = 7.3 Hz), 1.27 (2H, sex, J = 7.3 Hz) and 0.88 (3H, t, J = 7.3 Hz). **6-Amino-5-nitroso-1-butyluracil 16.** At 70 °C, to a solution of **14** (7.7 g, 42mmol) in a mixture of *N,N*-dimethylformamide (70 mL) and water (25 mL) was added sodium nitrite (1.27 g, 69 mmol) then concentrated hydrochloric acid

(5 mL). After 10 min. under stirring at 70 °C, the mixture was allowed to cool at room temperature and kept at 4 °C overnight. The precipitate was filtered off, washed with H₂O, and dried under reduced pressure, to afford 6-amino-5-nitroso-1-butyluracil **16** (7.68 g, 86%) as purple powder. ¹H NMR (250 MHz, DMSO-d₆) δ 12.62 (2H, sbroad), 10.89 (1H, sbroad), 3.76 (2H, t, *J* = 7.4 Hz), 1.48 (2H, quint, *J* = 7.4 Hz), 1.29 (2H, sex, *J* = 7.4 Hz) and 0.91 (3H, t, *J* = 7.4 Hz). **5,6-Diamino-1-butyluracil 18**. At 70 °C, to a solution of **16** (7.68 g, 36 mmol) in 12% aqueous solution of ammonia (150 mL) was added sodium hydrosulfite (19 g, 3eq) in solid fraction over 15 min. After addition, the reaction mixture was allowed to cool at room temperature and kept at 4 °C for 1 h. The precipitate was collected by filtration, washed with H₂O, and dried under reduced pressure to afford 5,6-diamino-1-butyluracil **18** (4.9 g, 68%) as light green powder. ¹H NMR (250 MHz, DMSO-d₆) δ 10.51 (1H, sbroad), 6.13 (2H, sbroad), 3.76 (2H, t, *J* = 7.3 Hz), 2.85 (2H, sbroad), 1.48 (2H, quint, *J* = 7.3 Hz), 1.28 (2H, sex, *J* = 7.3 Hz) and 0.88 (3H, t, *J* = 7.3 Hz). ¹³C NMR (63 MHz, MeOD) δ 159.3, 152.9, 152.3, 124.4, 44.5, 30.8, 20.7 and 14.1. According to the general procedure described above and starting from commercially available 4-chlorobenzylamine, the 5,6-diamino-1-(4-chlorobenzyl)uracil **19** was prepared in 20% overall yield. ¹H NMR (400 MHz, MeOD) δ 7.32 (2H, d, *J* = 8.6 Hz), 7.27 (2H, d, *J* = 8.6 Hz) and 4.28 (2H, s).

General procedure for the synthesis of 3-alkyl-8-hydroxymethylxanthine 22 and 23. The 3-Alkyl-8-hydroxymethylxanthine were prepared from corresponding 5,6-diaminouracil by condensation with glycolic acid followed by base induced ring closure acylation. *N*-(6-Amino-1-butyl-2,4-dioxo-1,2,3,4-tetrahydropyrimidin-5-yl)-2-hydroxyacetamide **20**. Under stirring and solvent free condition, a mixture of diaminouracil **18** (2.75 g, 13.9 mmol) and glycolic acid (2.16 g, 2eq) was heated at 120 °C for 30 min. To the resulting solid, allowed to cool at room temperature, was successively added EtOH (20 mL), H₂O (4 mL) and Et₂O (60 mL). The precipitate was filtrated off and washed with Et₂O to afford *N*-(6-Amino-1-butyl-2,4-dioxo-1,2,3,4-tetrahydropyrimidin-5-yl)-2-hydroxyacetamide **20** (3.29 g, 93%) as light yellow powder. ¹H NMR (250 MHz, DMSO-d₆) δ 10.57 (1H, sbroad), 8.09 (1H, s), 6.57 (2H, s), 5.34 (1H, sbroad), 3.93 (2H, s), 3.78 (2H, t, *J* = 7.2 Hz), 1.49 (2H, quint, *J* = 7.2 Hz), 1.30 (2H, sex, *J* = 7.2 Hz) and 0.89 (3H, t, *J* = 7.2 Hz). **3-Butyl-8-hydroxymethylxanthine 22**. At 85 °C, to a solution of **20** (1.83 g, 7.1 mmol) in a mixture of ethanol (10 mL) and water (5 mL) was added dropwise a 10% aqueous sodium hydroxide (5 mL) over 30 min. After addition the reaction mixture was stirred at 85 °C for 2 h, allowed to cool at room temperature, diluted with H₂O (10 mL) and then acidified with aqueous 1N HCl until pH=1. After cooling in ice bath for few min. the precipitate was collected by filtration and generously washed with H₂O. The solid was dried overnight in oven at 80 °C affording 3-butyl-8-hydroxymethylxanthine **22** (1.45 g, 81%) as a light yellow powder. ¹H NMR (250 MHz, DMSO-d₆) δ 4.49 (2H, s), 3.89 (2H, t, *J* = 7.3 Hz), 1.61 (2H, quint, *J* = 7.3 Hz), 1.28 (2H, sex, *J* = 7.3 Hz) and 0.90 (3H, t, *J* = 7.3 Hz). ¹³C NMR (63 MHz, DMSO-d₆) δ 154.5, 153.9, 150.9, 149.4, 106.8, 56.9, 41.7, 19.5 and 13.0.

According to the general procedure described above and starting from **19**, the 3-(4-chlorobenzyl)-8-hydroxymethylxanthine **23** was prepared in 43% overall yield. ¹H NMR (250 MHz, MeOD) δ 7.40 (2H, d, *J* = 8.5 Hz), 7.31 (2H, d, *J* = 8.5 Hz), 5.19 (2H, s) and 4.69 (2H, s).

General procedure for the synthesis of 3-alkyl-7-ethyl-8-hydroxymethylxanthine 24 and 25. 3,7-Dialkyl-8-hydroxymethylxanthine **24** and **25** were prepared according to a modified procedure previously described in the literature.⁷ **3-Butyl-7-ethyl-8-hydroxymethylxanthine 24**. Under argon, at 0 °C, to a mixture of **22** (238 mg, 1 mmol) and potassium carbonate (152 mg, 1.1eq) in dimethylformamide (20 mL) was injected a solution of ethyl iodide (80 mL, 1eq) in dimethylformamide (5 mL). The resulting mixture was allowed to warm at room temperature and stirred at this temperature overnight. The solvent was distilled off under reduce pressure and the residue successively diluted with H₂O (10 mL), acidified with aqueous 10% HCl (10 mL) and extracted with CH₂Cl₂ (3x5 mL). The combined organic layers were dried over Na₂SO₄ and the solvent was distilled off to afford 3-butyl-7-ethyl-8-hydroxymethylxanthine **24** (182 mg, 72%) as a light yellow solid. ¹H NMR (250 MHz, MeOD) δ 4.72 (2H, s), 4.40 (2H, q, *J* = 7.2 Hz), 3.98 (2H, t, *J* = 7.3 Hz), 1.69 (2H, quint, *J* = 7.3 Hz), 1.45 (3H, t, *J* = 7.2 Hz) 1.37 (2H, sex, *J* = 7.3 Hz) and 0.93 (3H, t, *J* = 7.3 Hz); ¹³C NMR (63 MHz, MeOD) δ 156.4, 153.7, 152.7, 150.4, 109.0, 56.9, 43.4, 42.1, 31.2, 20.9, 16.7 and 14.1.

According to the general procedure described above and starting from **23**, the 3-(4-chlorobenzyl)-7-ethyl-8-hydroxymethylxanthine **25** was prepared in 69% yield. ¹H NMR (250 MHz, MeOD) δ 7.38 (2H, d, *J* = 8.5 Hz), 7.26 (2H, d, *J* = 8.5 Hz), 5.15 (2H, s), 4.38 (2H, q, *J* = 7.0 Hz), 4.70 (2H, s) and 1.51 (3H, t, *J* = 7.0 Hz).

General procedure for the synthesis of 3-alkyl-8-chloromethyl-7-ethylxanthine 26 and 27. **3-Butyl-8-chloromethyl-7-ethylxanthine 26**. To a solution of **24** (345 mg, 1.3 mmol) in dichloromethane (5 mL) was injected thionyl chloride (283 μL, 3eq). The reaction mixture was refluxed 15 min. then the solvent was distilled off under reduced pressure. The residue was extended with Et₂O (20 mL) to allow crystallization and the precipitate was collected by filtration affording 3-butyl-8-chloromethyl-7-ethylxanthine **26** (325 mg, 88%) as a light yellow powder. ¹H NMR (250 MHz, MeOD) δ 4.77 (2H, s), 4.38 (2H, q, *J* = 7.1 Hz), 4.00 (2H, t, *J* = 7.3 Hz), 1.70 (2H, quint, *J* = 7.3 Hz), 1.50 (3H, t, *J* = 7.1 Hz) 1.39 (2H, sex, *J* = 7.3 Hz) and 0.94 (3H, t, *J* = 7.3 Hz). According to the general procedure described above and starting from **25**, the 3-(4-chlorobenzyl)-7-ethyl-8-chloromethylxanthine **27** was prepared in quantitative yield and use for next step without further purification.

Synthesis of triazolopyrimidine-2-thiol 29 and 30. [1,2,4]Triazolopyrimidine-2-thiol derivatives were prepared according to the reported method.⁸ **5,7-Dimethyl-[1,2,4]triazolo[1,5-a]pyrimidine-2-thiol 29** (94%) as yellow powder. **6-Butyl-5,7-Dimethyl-[1,2,4]triazolo[1,5-a]pyrimidine-2-thiol 30** (86%) as light green powder.

HTRF screen.

HTRF assays were performed in white 384 Well Small Volume™ HiBase Polystyrene Microplates (Greiner) with a total working volume of 20 µL. Compounds were dispensed, with 200 nL per well (1% final DMSO), from a concentration stock of 1 mM in 100% DMSO and with serial DMSO dilutions, using a Mosquito Crystal pipetting robot platform (TTP labtech). The IC₅₀ measurements were performed in triplicates. All HTRF reagents were purchased from CisBio Bioassays and reconstituted according to the supplier protocols. For each assay 14.7 µL of mix 1 (protein + peptide) is added in the assay wells, containing previously dispensed inhibitors, according to the final concentration and buffer described in Supporting Information Table S2, using a Biomek NX MC pipetting robot (Beckman). Then, 5.1 µL of mix 2 (donnor + acceptor) is added. HTRF signals were measured, after a final incubation (overnight at room temperature), using a PHERAstar FS (BMG Labtech) with an excitation filter at 337 nm and fluorescence wavelength measurement at 620 and 665 nm, an integration delay of 60 µs and an integration time of 500 µs. Results were analyzed with a two-wavelengths signal ratio: [intensity (665 nm)/intensity (620 nm)]*10⁴. Percentage of inhibition was calculated using the following equation: % inhibition = [(compound signal) - (min signal)] / [(max signal) - (min signal)] * 100, where 'max signal' is the signal ratio with the compound vehicle alone (DMSO) and 'min signal' the signal ratio without peptide. For IC₅₀ measurements, values were normalized and fitted with Prism (GraphPad software) using the following equation: $Y = 100 / (1 + ((X / IC_{50})^{Hill\ slope}))$.

Protein expression and purification.

For Isothermal Titration Calorimetry (ITC) and crystallogensis, BRD₄(BD₁) was produced and purified using a histidine tag affinity chromatography as described by Filipakopoulos *et al.*⁹ For these experiments, a pNIC28-BSA₄ expression vector containing BRD₄(BD₁) and a Tobacco Etch Virus (TEV) protease cleavage site have kindly been provided by Stefan Knapp laboratory from the SGC at the University of Oxford. After size exclusion chromatography, the fractions presenting pure BRD₄(BD₁) after TEV cleavage of the histidine tag were pooled and concentrated to 25 mg/mL for crystallogensis. For ITC assays, the protein was concentrated up to 6 mg/mL and the DTT was removed using a buffer exchange column (PD₁₀ from GE healthcare) equilibrated with 10 mM HEPES pH=7.5, 150 mM NaCl. For Homogeneous Time Resolved Fluorimetry (HTRF) experiments, a BRD₄(BD₁) synthetic gene that includes a TEV cleavage site was purchased from LifeTechnology in a pDONR transport vector before cloning into a pDESTTM₁₅ expression vector for GST affinity purification. Protein production and purification was carried out using similar protocols and buffers as used for the His-BRD₄(BD₁) system. Purification was carried on GST affinity resin (Thermo Scientific) and reduced glutathione was used for protein release. GST-BRD₄(BD₁) was further purified by size exclusion chromatography on a Superdex 16/60 Hiload column (GE Healthcare) using 20 mM TRIS pH=8.0, 150 mM NaCl Buffer.

Bromodomain selectivity profiles.

Selectivity profiles of bromodomain inhibitors were performed as described in HTRF screen section. Concentration of histone peptide was optimized to ensure sufficient signal to noise ratio, sufficient sensitivity for detection of weak inhibitors and comparable data from one bromodomain to another. HTRF detection reagents (EPIgeneousTM Binding Domain kits) were purchased from Cisbio Bioassays and used according to supplier's protocol. GST tagged bromodomain proteins were purchased from BPS Bioscience and histone peptide from Anaspec.

Cells and cell culture.

The human leukemia cell line Jurkat (ATCC[®] TIB-152) was maintained in RPMI-1640 medium supplemented with 10% FBS at 37 °C and 5% CO₂. Human osteosarcoma cell line (U2OS, ATCC[®] HTB-96TM) was maintained in DMEM supplemented with 10% FBS at 37 °C and 5% CO₂.

Cytotoxicity experiments.

In antiproliferative assays, compounds were assayed for their growth inhibiting activity towards the described cancer cell lines using the Cell Titer-Glo Luminescent Cell Viability Assay as described by the manufacturer (Promega Corporation). Briefly, 10⁴ cells were plated onto 96 well- plates (white with clear bottom (3610, Corning Costar)) in 100 µL media per well immediately before assay. Compounds were added at different concentrations (ranging from 100 to 0.05 µM) to each well and cell cultures were incubated 37 °C during 72 h. Vehicle (DMSO) was used as control and all compounds were tested in constant percentage of DMSO (1%). After addition of 50 µL Cell Titer-GLO, Luminescence was measured using a Centro luminometer LB960 (Berthold). Dose-response curves were generated and effective dose 50 values (EC₅₀) were calculated using non-linear regression analysis (Graph Pad Prism).

Western blot.

Non-treated Jurkat cells and Jurkat cells treated during 24 h with compound 1, or DMSO were lysed in RIPA lysis buffer (Tris HCl pH=7.5 50 mM, NaCl 150 mM, Triton 1%, SDS 0.1%, sodium deoxycholate 1%) supplemented with protease inhibitor cocktail (P8340, Sigma-Aldrich) and 1 mM PhenylMethylSulfonyl Fluoride (Sigma-Aldrich). 75 µg of protein were loaded onto 10% acrylamide SDS/PAGE and then transferred to nitrocellulose Hybond C-extra membranes, 45 micron (GE Healthcare). The membranes were saturated with 5% (wt/vol) skim milk in TBST [Tris-buffered saline/0.1% (vol/vol) Tween 20] 1 h at room

temperature and incubated with anti-myc antibody (clone 9E10, sc40 Santa Cruz) at a 1 : 500 dilution in 0.5% (wt/vol) skimmed milk in TBST overnight at 4 °C. Membranes were then washed with TBST, incubated with an HRP-conjugated anti-mouse secondary antibody (polyclonal goat anti mouse Po447, Pierce) at a 1 : 20,000 dilution in TBST 1 h at room temperature. Immunoreactive bands were revealed using SuperSignal™ West Pico Extended Duration Substrate (Pierce) detection reagents. Quantification was performed using ImageJ software.

Isothermal titration calorimetry.

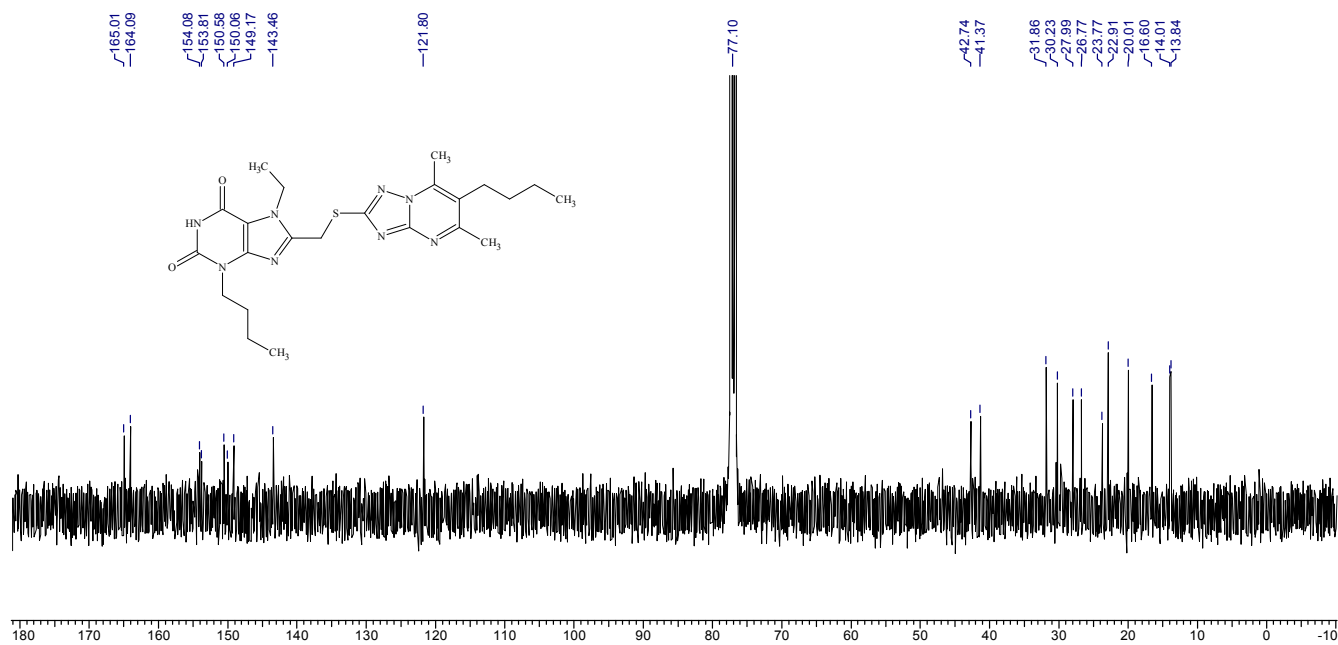
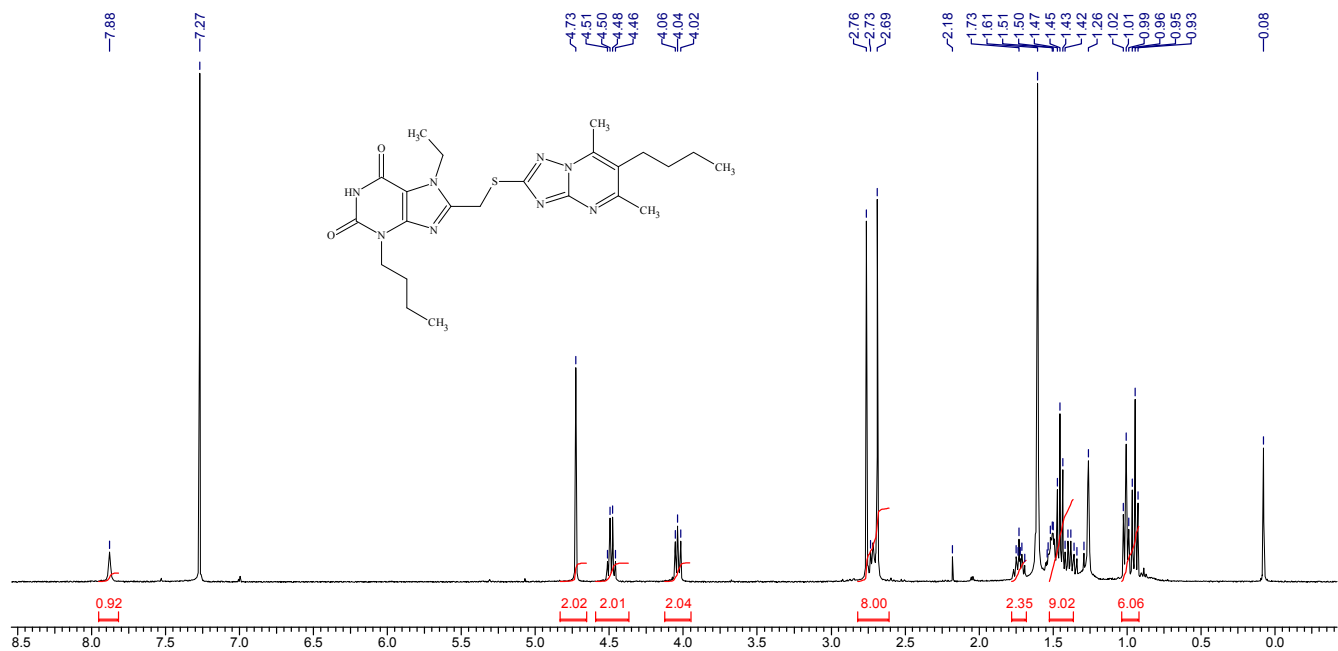
ITC was used to evaluate the thermodynamics parameters of the binding between BRD₄(BD₁) and the selected compounds, using ITC conditions previously described by Filippakopoulos *et al.*⁸ Purified BRD₄(BD₁) was extensively dialyzed in the ITC buffer containing 10mM Hepes pH=7.5 and 150 mM NaCl. Compounds were diluted directly in the last protein dialysate prior to experiments. Titrations were carried out on a MicroCal ITC200 microcalorimeter (GE Healthcare, Piscataway, NJ). Each experiment was designed using a titrant concentration (protein in the syringe) set 10 to 15 times the analyte concentration (compound in the cell generally between 10 and 35 μM) and using 13 injections at 35 °C. A first small injection (generally 0.2 μL) was included in the titration protocol in order to remove air bubbles trapped in the syringe prior titration. Raw data were scaled after setting the zero to the titration saturation heat value. Integrated raw ITC data were fitted to a one site non-linear least squares fit model using the MicroCal Origin plugin as implemented in Origin 7 (Origin Lab). Finally, ΔG and TΔS values were calculated from the fitted ΔH and K_A values using the equations $\Delta G = -R.T.\ln K_A$ and $\Delta G = \Delta H - T\Delta S$.

Crystallography.

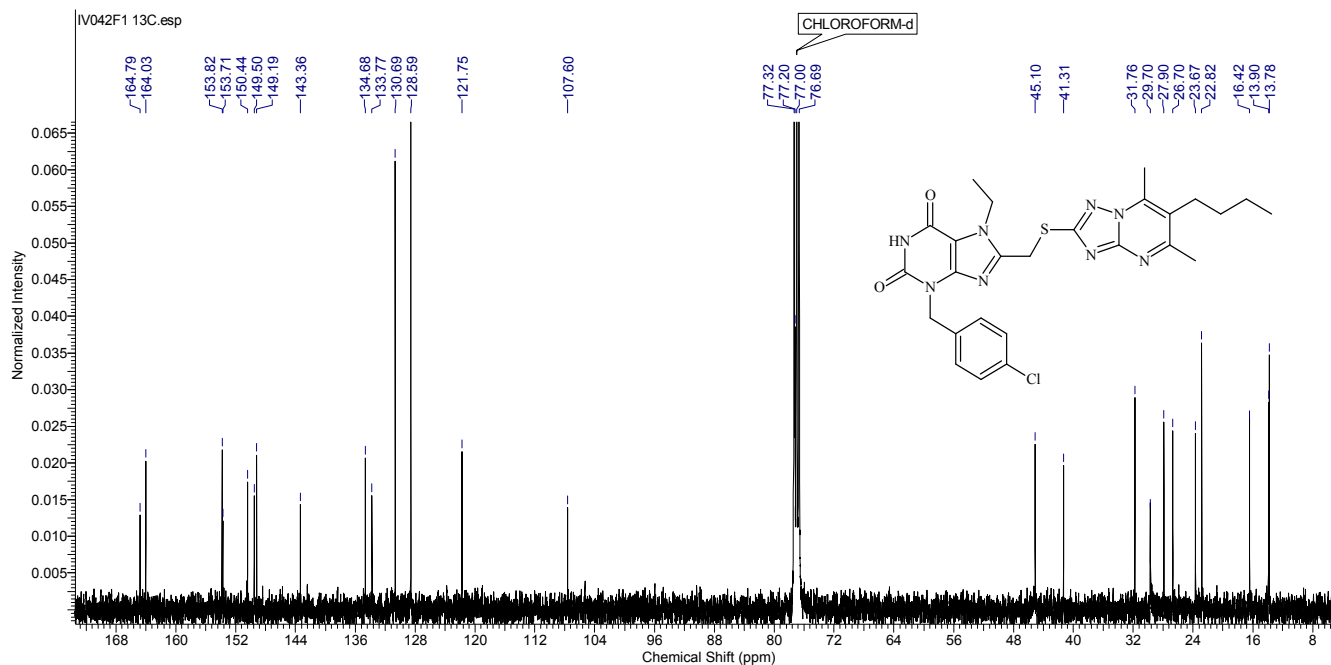
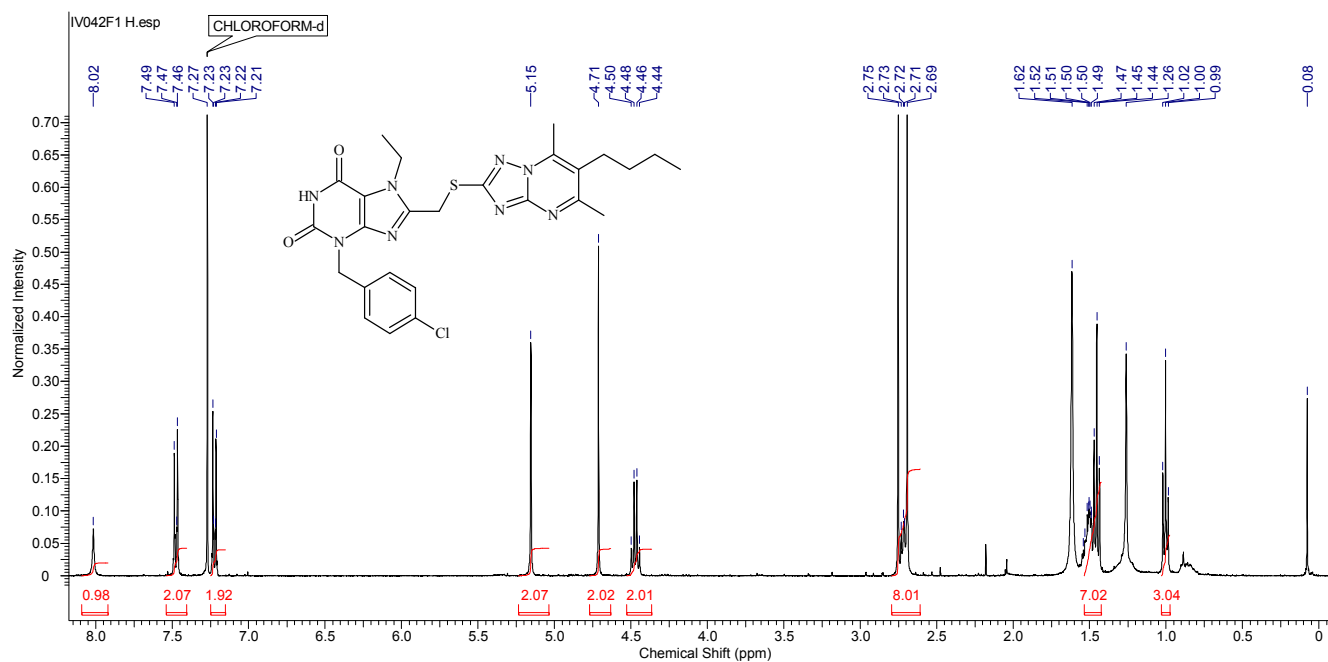
BRD₄(BD₁)-Inhibitor co-crystallization was performed at 19 °C (292K) using the hanging drop vapor diffusion method. For the complex with 1 mM of compound **1**, 25 mg/mL of BRD₄(BD₁) preparation was mixed at a 1 : 1 ratio with the precipitant solution (200 mM ammonium sulfate, 100 mM Tris pH=8.5, 25% PEG₃₃₅₀) and crystals grew to diffracting quality within 3-5 days. For the complex with 2.5 mM of **5**, 16 mg/mL of BRD₄(BD₁) preparation was mixed at a 1 : 1 ratio with the precipitant solution (250 mM ammonium sulfate, 100 mM Tris pH=8.5, 19% PEG₃₃₅₀) and crystals grew to diffracting quality within 15 days. For the complex with **7**, 2.5 mM of compound and 12 mg/mL of BRD₄(BD₁) preparation was mixed at a 1 : 1 ratio with the precipitant solution (300 mM Na Formate, 100 mM NaCl, 22% PEG₃₃₅₀, 10% Ethylene glycol). Crystals grew to diffracting quality within 1month. Crystals were cryo-protected using the precipitant solution supplemented with 10% ethylene glycol for **1** and **5** or with 10% glycerol for **7** and were flash frozen in liquid nitrogen. Data for **1**, **5** and **7** were respectively collected at the AFMB laboratory - Marseilles (Bruker AXS MICROSTAR equipped with a Mar345 detector), and on the ESRF beamlines ID23-2 and ID29. Indexing, integration and scaling were performed using XDS.¹⁰ Initial phases were calculated by molecular replacement with Phaser MR (CCP4 suite)¹¹ using a model of the first domain of BRD₄ (extracted from the Protein Data Bank accession code: 2OSS). Initial models for the protein and the ligands were built in COOT.¹² The cycles of refinement were carried out with Refmac5 (CCP4 suite). Data collection and refinement statistics can be found in Supplementary Information Table S1. The models and structures factors have been deposited with Protein Data Bank accession code for compound **1**: 5EGU, compound **5**: 5EIS and compound **7**: 5E14.

¹H and ¹³C NMR spectra of triazolo-substituted xanthine-based BET inhibitors

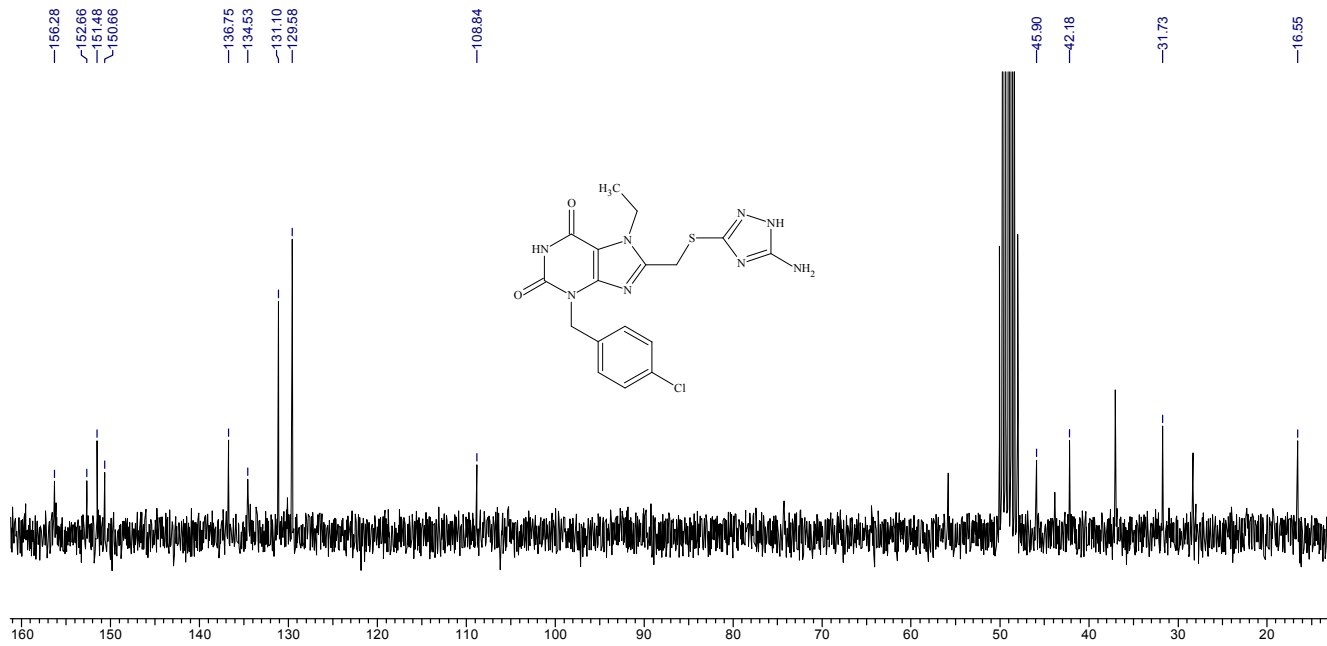
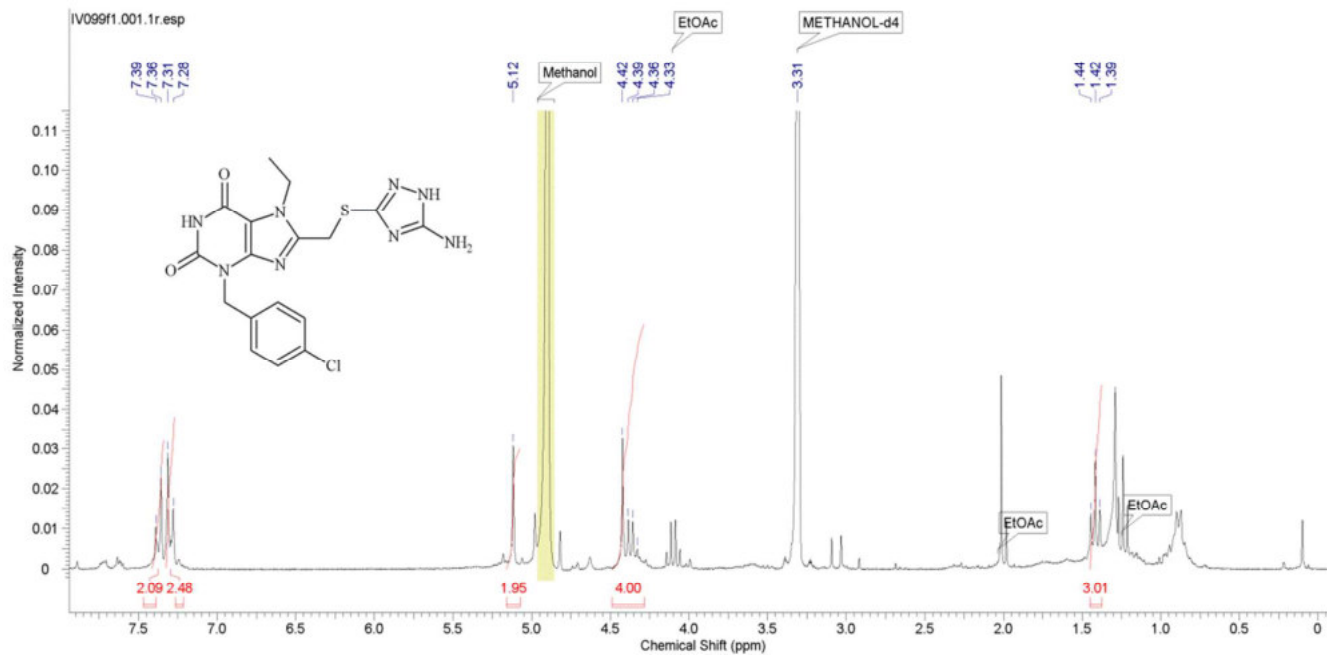
3-Butyl-8-(6-butyl-5,7-dimethyl-[1,2,4]triazolo[1,5-a]pyrimidin-2-ylsulfanylmethyl)-7-ethyl-3,7-dihydro-purine-2,6-dione **1**.



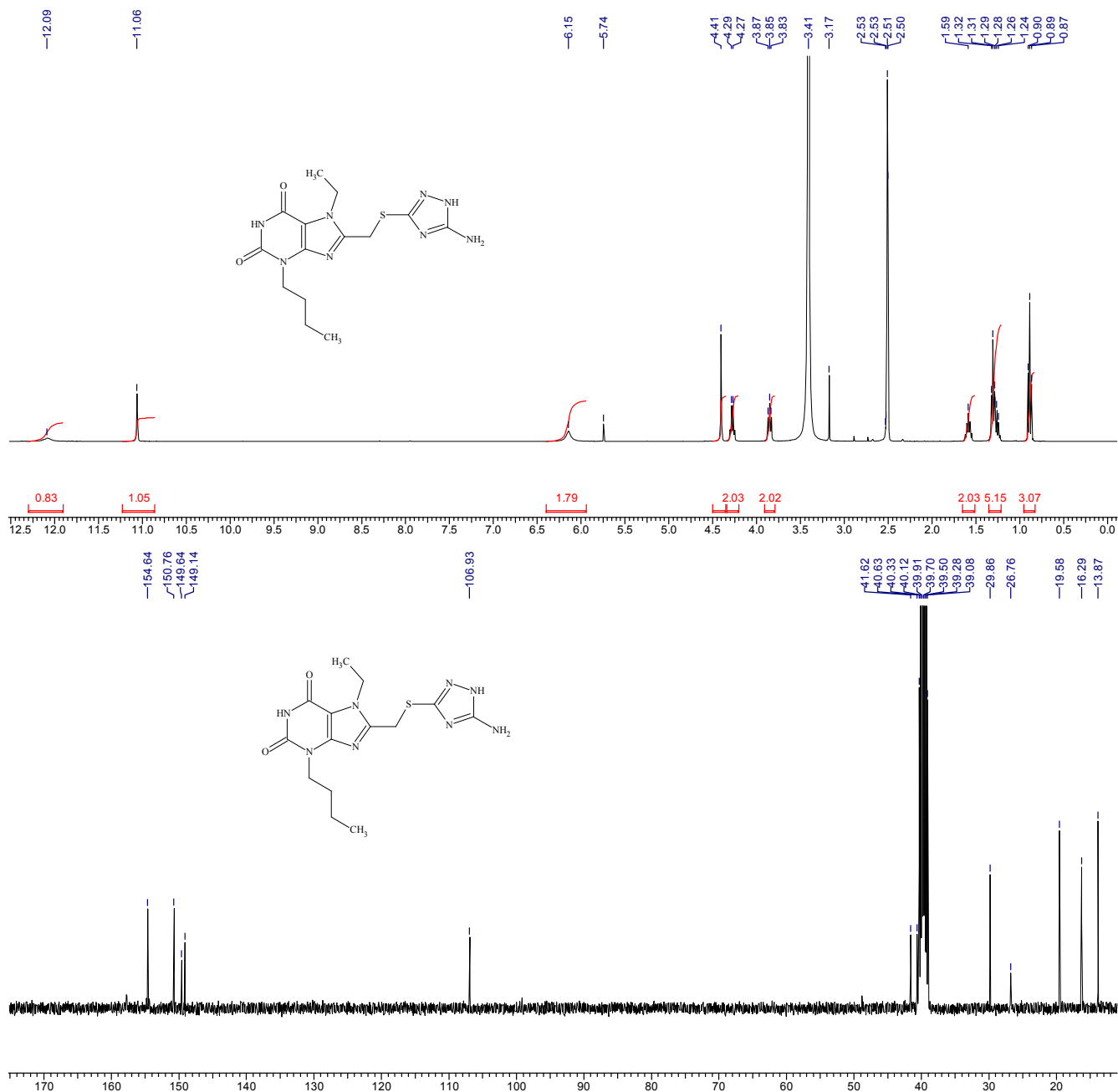
8-(6-Butyl-5,7-dimethyl-[1,2,4]triazolo[1,5-a]pyrimidin-2-ylsulfanylmethyl)-3-(4-chlorobenzyl)-7-ethyl-3,7-dihydropurine-2,6-dione 6.



8-(5-Amino-1H-[1,2,4]triazol-3-ylsulfanylmethyl)-3-(4-chlorobenzyl)-7-ethyl-3,7-dihydro-purine-2,6-dione 7.

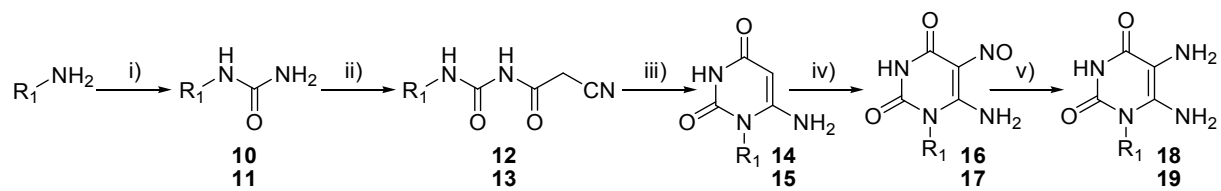


8-(5-Amino-1H-[1,2,4]triazol-3-ylsulfanylmethyl)-3-butyl-7-ethyl-3,7-dihydro-purine-2,6-dione 8.



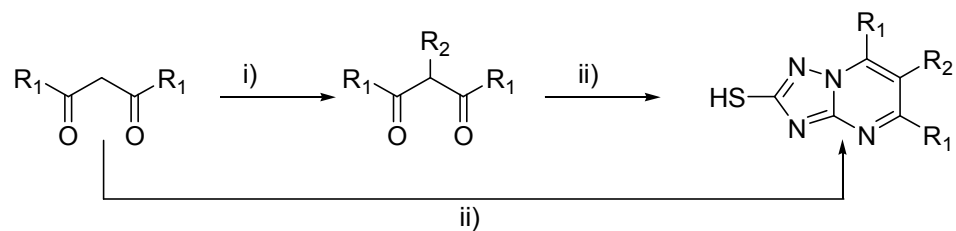
Schemes S1-S2

Scheme S1. Synthesis of 5,6-diaminouracil intermediates 18 and 19^a



^areagents and conditions : i) KOCN, HCl_{aq.}-EtOH, rt, 12h ii) cyanoacetic acid, Ac₂O, 0 to 60 °C, 30 min iii) NaOH, H₂O-EtOH, 85 °C, 45 min iv) NaNO₂, HCl_{aq.}, DMF-H₂O, 70 °C, 10 min v) Na₂S₂O₄, NH₃(aq), 70 °C, 15 min.

Scheme S2. Synthesis of mercapto triazolopyrimidine reagents^a



^areagents and conditions : i) BuI, C₆H₆, DBU, rt, 12h ii) 3-Amino-1,2,4-triazol-5-thiol, AcOH, piperidine, reflux, 24 h.

Figures S1-S5

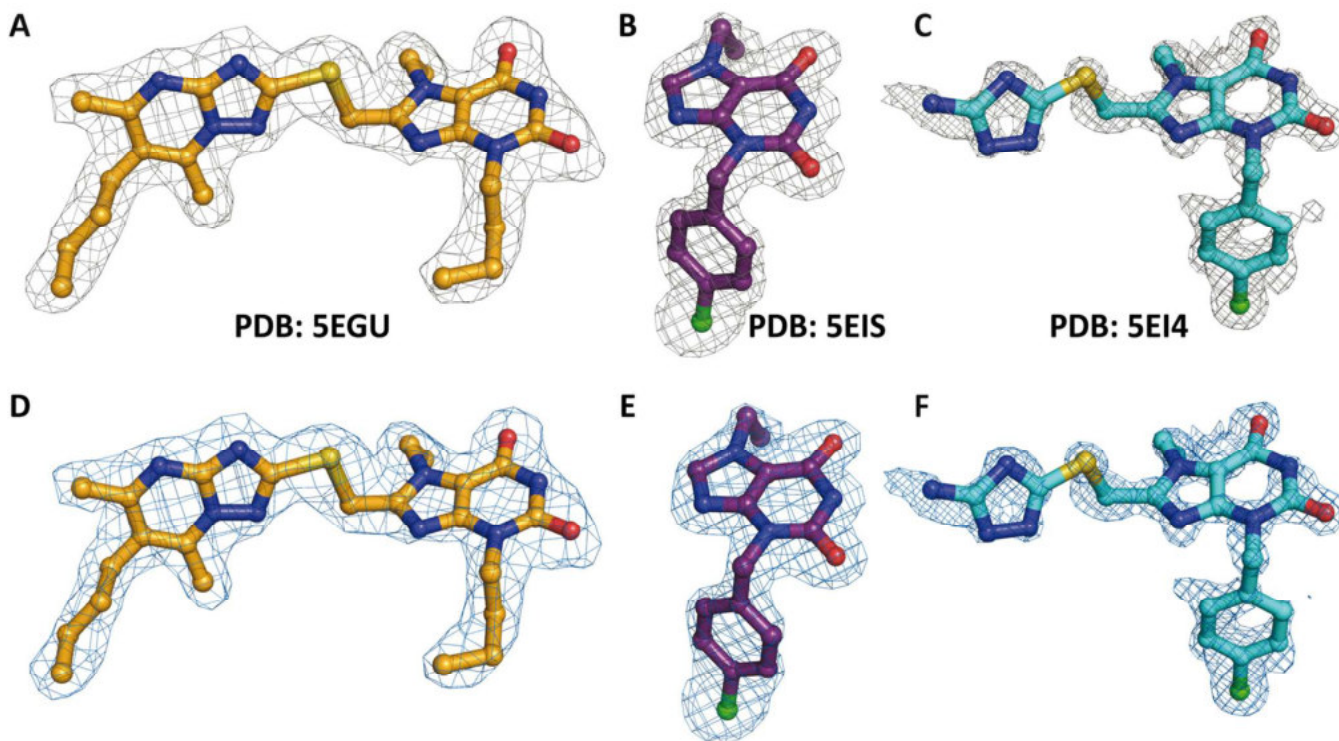


Figure S1. 2Fo-Fc map and omit map around the compounds. (A, B and C) 2Fo-Fc map contoured at 1σ around **1** (orange), **5** (purple) and **7** (cyan), respectively. For **1** and **5**, the maps are very well defined for the entire molecules, while **7** exhibits a weaker density due to its low occupancy in the crystal. (D, E and F) 2Fo-Fc map contoured at 1σ around **1** (orange), **5** (purple) and **7** (cyan) respectively, with the model compound omitted before refinement.

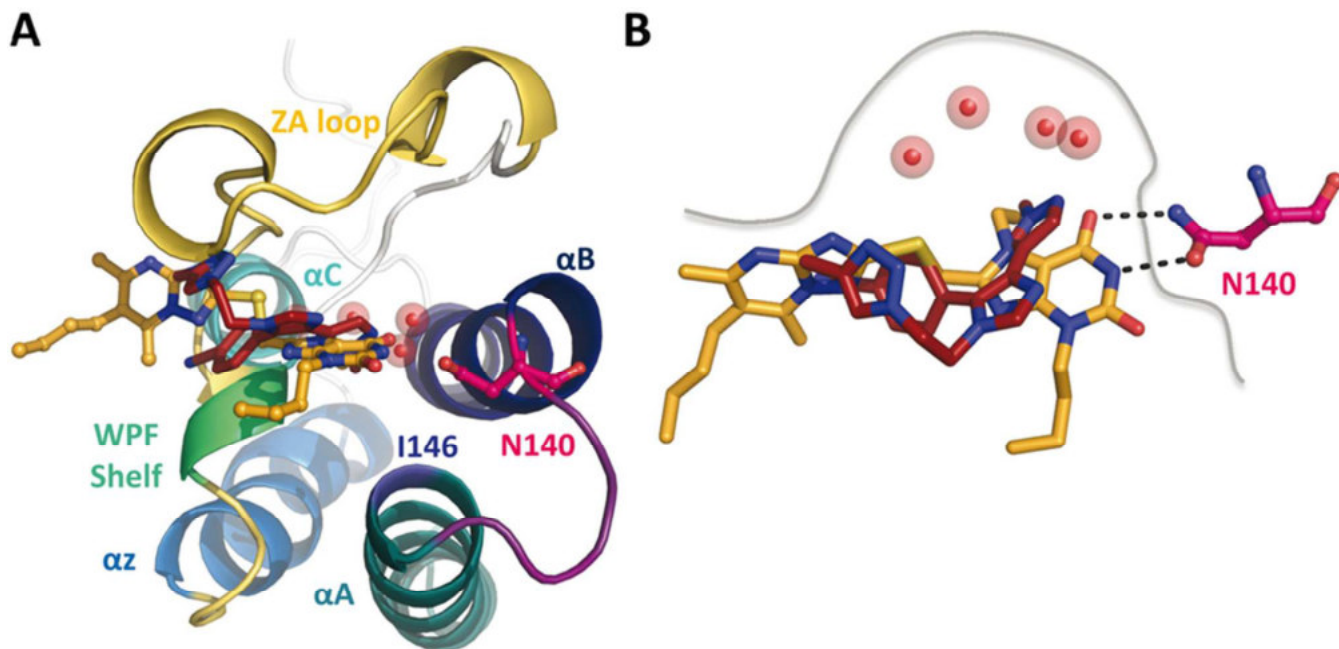


Figure S2: Superimposition of **1** (pdb: 5EGU) and a BAZ2B inhibitor (pdb:4XUB). (A) Comparison of the binding of **1** (orange) and **4XUB** (red) in the *N*-acetylated binding pocket of BRD4(BD1). (B) Detailed view of the molecular mode of action of **1** interaction with N140 superimposed with **4XUB**. The gray trace illustrates the cavity surface.

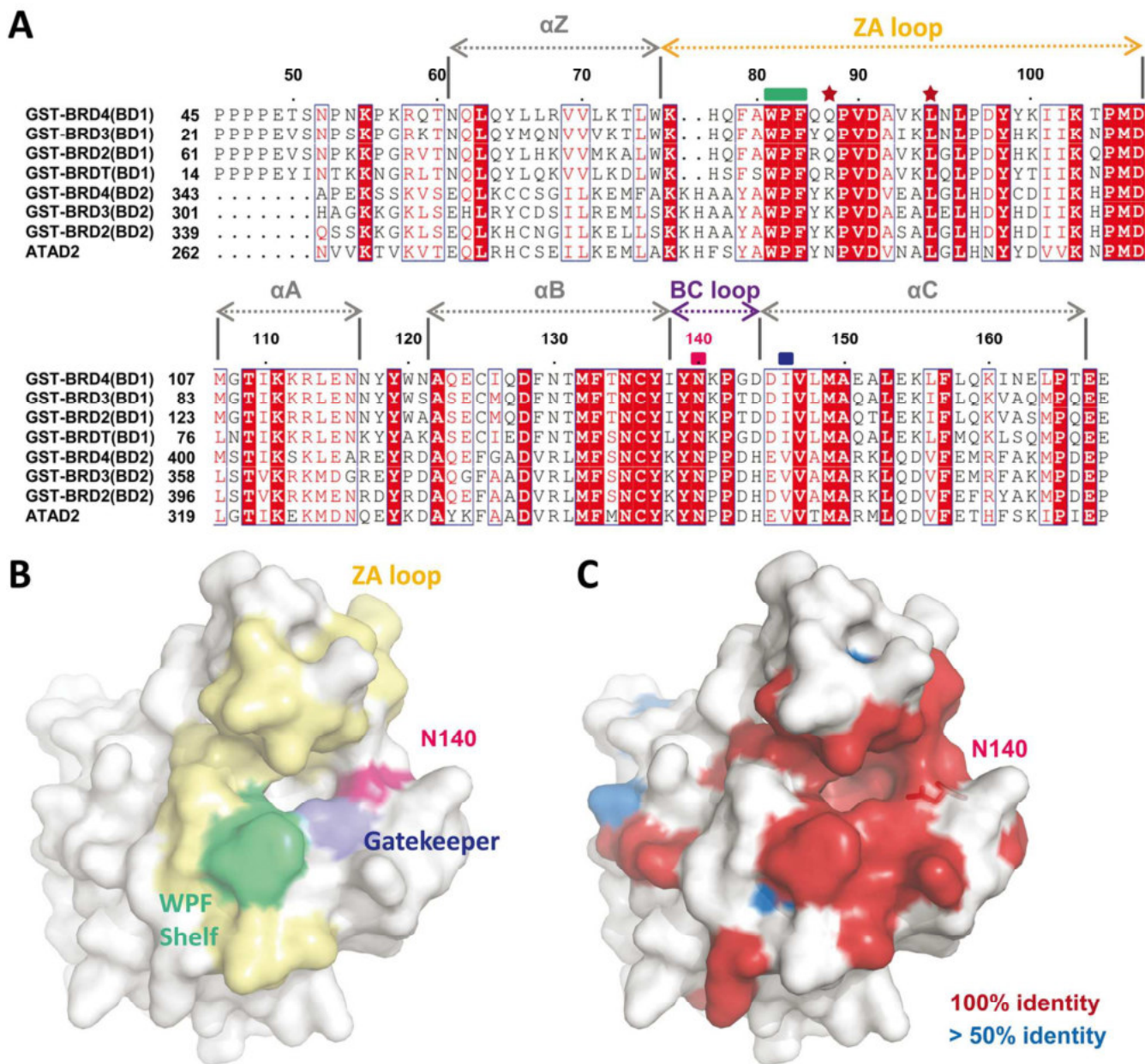


Figure S3. Sequences alignment and analysis of the BET family. (A) Multiple-sequence alignment of the BET family members (BRD4-, BRD3-, BRD2-, BRDT-, BD₁ and BD₂). Sequence alignment was generated and analyzed using ClustalW¹³ and ESPRIPT.¹⁴ Similar residues are shown in red text and similar regions are boxed in blue. Identical residues among the family are highlighted in red. The sequence is annotated with the BRDs described secondary structures as BET members are folded in four alpha helices (gray: α A, α B, α C, α AZ) and the two BC and ZA loops (orange and purple respectively). The conserved WPF shelf, N140 and I146 gatekeeper location are pinpointed with green, pink and blue boxes respectively on top of the sequence alignment. Red stars highlight the position of two residues (Q85 and L92) involved in the stacking of the triazolopyrimidinyl moiety of **1**. (B) Surface representation of BRD4(BD₁) colored in yellow for the ZA loop, green for the WPF shelf, blue for the gatekeeper I146 residue and pink for the conserved N140. (C) Surface representation colored according to the sequence identity among the BET family (red = 100%, 100% > blue > 50%, white < 50% identity).

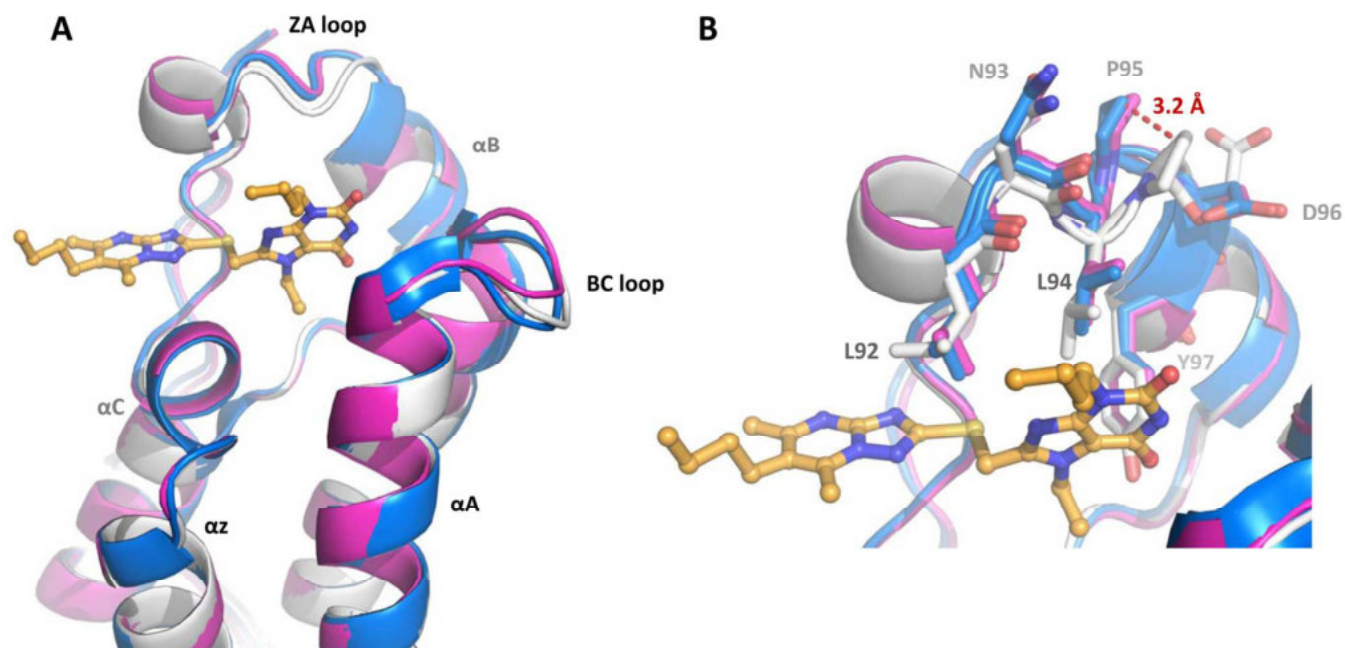


Figure S4: Structures superimposition of the ligand-free BRD₄-BD₁ and the complexes with **1** (pdb: 5EGU) and **4** (pdb: 5EIS). (A) Cartoon view of the superimposition of the three structures. **1** is displayed as orange sticks and balls and the proteins as cartoons (ligand-free: pink; **1**: white; **4**: blue). (B) Detailed view of the ZA loop rearrangements upon fixation of **1**. The red dashed line indicates the distance of 3.2 Å observed between the P95 in complex with **1** versus **4**. These rearrangements of the ZA loop at the residues positions L94, P95 and D96 are not the results of crystallographic packing contacts.

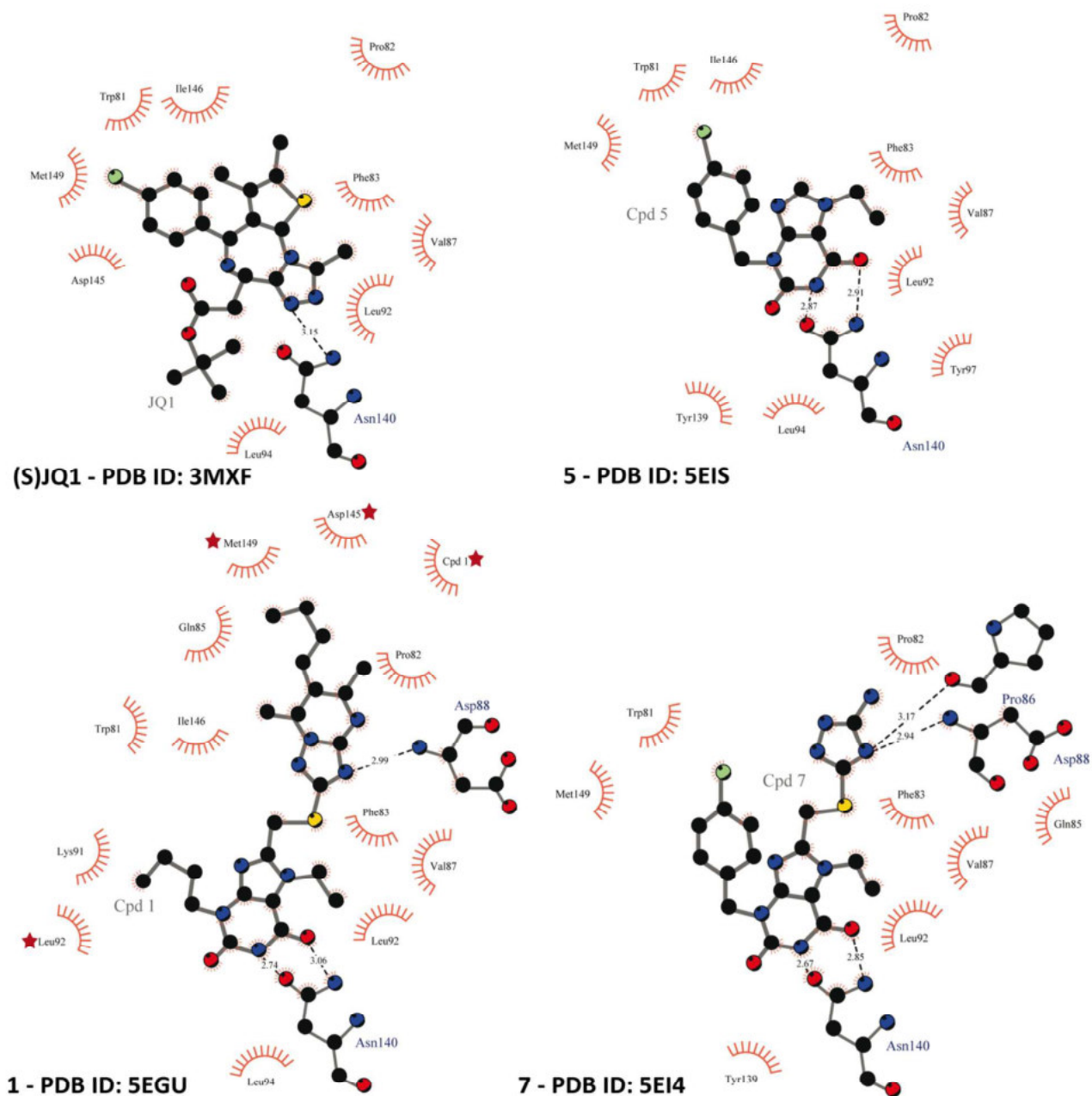


Figure S5. Ligplot contacts analysis for **1**, **5** and **7** compared to *S*(JQ₁). Ligplus/ligplot representation of the crystallographic structure contacts between the ligands (top left: *S*-JQ₁ – PDB ID: 3MXF, top right: **1**, bottom left: **5**, bottom right: **7**) and BRD₄(BD₁). Ligands are automatically positioned in the same orientation and are shown as a two dimensions balls and sticks representations surrounded by the residues involved in hydrogen bond interactions (black dashed lines). Potential van der Waals contact residues are presented as red oriented arc shapes around the ligands that point towards their contact points on the compound atoms. Only for **1**, van der Waals contacts established with other chains in the crystallographic asymmetry unit or with symmetry related protein are indicated as red stars.

Tables S1-S3
Table S1: X-ray data collection and refinement statistics

	1 (pdb: 5EGU)	5 (pdb: 5EIS)	7 (pdb: 5EI4)
Crystal parameters			
Space group	P2 ₁ 2 ₁ 2 ₁	P2 ₁ 2 ₁ 2 ₁	P2 ₁ 2 ₁ 2 ₁
Unit-cell dimensions			
a, b, c (Å)	69.31, 83.20, 106.75	35.93, 46.54, 77.66	37.35, 44.07, 78.27
α , β , γ (°)	90.0, 90.0, 90.0	90.0, 90.0, 90.0	90.0, 90.0, 90.0
Data collection			
Beamline	AFMB - IP2	ESRF - ID23-2	ESRF - ID29
Wavelength (Å)	1.541790	0.872600	0.972390
Resolution range (Å)	50.0 - 2.2 (2.21 - 2.35)	50.0 - 1.6 (1.70 - 1.60)	50.0 - 1.05 (1.11 - 1.05)
Total reflections	274579 (42137)	82522 (13318)	306790 (44637)
Unique reflections	31319 (4855)	17766 (2789)	60686 (9588)
Completeness (%)	97.9 (95.5)	99.4 (98.7)	99.5 (98.6)
Average I/σ	17.91 (4.73)	14.87 (2.73)	11.16 (3.10)
Rfactor obs (%)	11.0 (51.0)	5.3 (49.7)	7.4 (48.1)
Redundancy	8.8 ()	4.6 (4.8)	5.1 (4.7)
Structure Refinement			
Nmol/asym	4	1	1
Matthews Coefficient	2.59	2.15	2.16
Solvent (%)	52.59	42.90	43.19
Starting model for Molecular Replacement	2OSS	2OSS	2OSS
Resolution range	35.85 2.213 (2.292 - 2.213)	38.83 - 1.6 (1.657 - 1.6)	39.14 - 1.051 (1.088 - 1.051)
Reflections used in refinement	31316 (2977)	17741 (1740)	60684 (5875)
Reflections used for R-free	1565 (148)	888 (87)	3035 (294)
R-Work	0.1834 (0.2086)	0.1891 (0.3363)	0.1778 (0.2580)
R-Free	0.2316 (0.2797)	0.2284 (0.3514)	0.1990 (2.2605)
R.m.s deviation			
Bond lengths (Å)	0.019	0.019	0.030
Angles (°)	1.83	1.86	2.40
Average B-factor			
Protein	25.08	27.34	10.51
Ligand	22.06	26.00	10.82
Solvent	26.02	35.89	20.35
Ramachandran plot (%)			
Favored regions	98	98	99
Allowed regions	2.2	2.4	0.76
No. of non hydrogen atoms			
macromolecule	4355	1204	1312
ligand	3952	1071	1115
	140	29	33

Table S2. HTRF selectivity experimental procedures^a

Protein name	MIX 1		(nM)	MIX 2			Assay buffer ^b	DMSO
	(nM)	Peptide name		Donnor name	(nM)	Acceptor name (nM)		
GST-BRD ₄ (BD ₁)	5	H ₄ KAc 5/8/12/16 peptide	25	MAb Anti GST-Keu	0.5	Streptavidin d2 3.125	Buffer A	1%
GST-BRD ₃ (BD ₁)	5	H ₄ KAc 5/8/12/16 peptide	10	MAb Anti GST-Keu	0.5	Streptavidin d2 1.25	Buffer A	1%
GST-BRD ₂ (BD ₁)	5	H ₄ KAc 5/8/12/16 peptide	15	MAb Anti GST-Keu	0.5	Streptavidin d2 1.875	Buffer A	1%
GST-BRDT(BD ₁)	5	H ₄ KAc 5/8/12/16 peptide	50	MAb Anti GST-Keu	0.5	Streptavidin XL665 6.3	Buffer A	1%
GST-BRD ₄ (BD ₂)	5	H ₄ KAc 5/8/12/16 peptide	75	MAb Anti GST-Keu	0.5	Streptavidin XL665 9.375	Buffer A	1%
GST-BRD ₃ (BD ₂)	5	H ₄ KAc 5/8/12/16 peptide	150	MAb Anti GST-Keu	0.5	Streptavidin XL665 18.75	Buffer A	1%
GST-BRD ₂ (BD ₂)	5	H ₄ KAc 5/8/12/16 peptide	25	MAb Anti GST-Keu	0.5	Streptavidin XL665 3.125	Buffer A	1%
GST-ATAD ₂	5	H ₄ KAc 5/8/12/16 peptide	10	MAb Anti GST-KTb	0.5	Streptavidin XL665 10	Buffer B	1%

^aFor each bromodomain, concentrations of protein, peptide, donor and acceptor have been optimized and are presented in this summary table also indicating the donor and acceptor type as well as the final DMSO concentration and buffer composition.

^bBuffer A: 50 mM KPO₄, pH7, 100 mM KF, BSA 0.1%; Buffer B: 50 mM Hepes, pH7.5, 150 mM NaCl, BSA 0.1%.

Table S3. Biochemical and biophysical results summary^a

Cpd	SMILE	EC ₅₀ ^b (μ M)	HTRF IC ₅₀ (μ M)						ITC ^b						
			4(1)	4(2)	3(1)	3(2)	2(1)	2(2)	T(1)	ATAD2	K _d (μ M)	Δ G (Kcal/mol)	Δ H (Kcal/mol)	-T Δ S (Kcal/mol)	N ^c
2	O=C1C2=C(N=CN2)NC(N1)=O	-	NC	-	-	-	-	-	-	-	-	-	-	-	
3	O=C1C2=C(N=CN2CC)NC(N1)=O	-	NC	-	-	-	-	-	-	-	-	-	-	-	
4	O=C1C2=C(N=CN2CC)N(CCCCC)C(N1)=O	-	13.0 \pm 0.3	8.1 \pm 0.3	4.8 \pm 0.3	5.8 \pm 0.2	18.8 \pm 0.7	6.6 \pm 0.3	7.9 \pm 0.4	NC	1.8	-8.1 \pm 0.1	-17.2 \pm 1.1	9.2 \pm 1.1	0.8
5	O=C1C2=C(N=CN2CC)N(CC3=CC=C(Cl)C=C3)C(N1)=O	-	2.6 \pm 0.1	1.9 \pm 0.1	1.8 \pm 1.1	1.8 \pm 0.2	4.7 \pm 0.1	1.6 \pm 0.2	2.3 \pm 0.2	NC	-	-	-	-	-
1	O=C1C2=C(N=C(CSC3=NN4C(N=C(C)C(CC)=C4C)=N3)N2CC)N(CCCCC)C(N1)=O	27	5.0 \pm 0.1	NC	62.0 \pm 7.8	NC	59.0 \pm 2.9	NC	> 100	> 100	1.4	-8.2 \pm 0.1	-1.1 \pm 0.0	-7.1 \pm 0.1	1.9
6	O=C1C2=C(N=C(CSC3=NNC(N)=N3)N2CC)N(CC4=CC=C(Cl)C=C4)C(N1)=O	-	20.4 \pm 3.6 ^d	NC ^d	NC ^d	NC ^d	NC ^d	NC ^d	NC ^d	NC ^d	-	-	-	-	-
7	O=C1C2=C(N=C(CSC4=NNC(N)=N4)N2CC)N(CC3=CC=C(Cl)C=C3)C(N1)=O	-	26.4 \pm 0.7	NC ^e	31.7 \pm 1.1	NC ^e	30.0 \pm 0.9	NC ^e	> 100	NC	-	-	-	-	-
8	O=C1C2=C(N=C(CSC3=NNC(N)=N3)N2CC)N(CCCCC)C(N1)=O	-	> 50	-	-	-	-	-	-	-	-	-	-	-	-

^aFor clarity, bromodomain names have been indicated as the following abbreviations: 4(1) : BRD4-BD1 ; 4(2) : BRD4-BD2 ; 3(1) : BRD3-BD1 ; 3(2) : BRD3-BD2 ; 2(1) : BRD2-BD1 ; 2(2) : BRD2-BD2 ; T(1) : BRDT-BD1. NC abbreviation stands for Not Converged. ^b Experiment only performed on BRD4-BD1. ^c N = 0.8 indicates a 1 molecule stoichiometry ; N = 1.9 : indicates a 2 proteins for 1 molecule stoichiometry. ^dData above 25 μ M have been excluded due to precipitation and fluorescence interference at 620 nm. ^eFluorescence interference at 620 nm with BD2.

REFERENCES

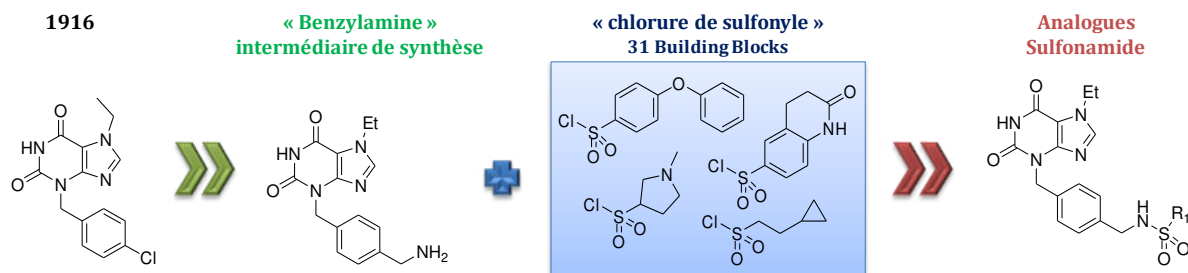
- (1) Pinto, I. L.; Rahman, S. S.; Nicholson, N. H. Preparation of xanthine derivatives as HM74A agonists. *PCT Int. Appl.* **2005**, WO 2005077950 A2.
- (2) Vidal, A.; Giraud, I.; Madelmont, J.-C. Improved synthesis of 7-(alkyl/aryl)guanine. *Synthetic comm.* **2004**, *34*, 3359-3365.
- (3) Cavalla, D.; Hofer, P.; Chasin, M. Preparation of aryl thioxanthines as PDE IV inhibitors. US patent **2000** 6066641.
- (4) Kalla, R. V.; Elzein, E.; Perry, T.; Li, X.; Palle, V.; Varkhedkar, V.; Gimbel, A.; Maa, T.; Zeng, D.; Zablocki, J. Novel 1,3-disubstituted 8-(1-benzyl-1H-pyrazol-4-yl) xanthines: high affinity and selective A₂B adenosine receptor antagonists. *J. Med. Chem.* **2006**, *49*, 3682-3692.
- (5) Manidhar, D. M.; Uma Maheswara Rao, K.; Suresh Reddy, C.; Syamasunder, C.; Adeppa, K.; Misra, K. An improved method for the preparation of alkyl/arylurea derivatives using chlorocarbonylsulfonyl chloride as carbonylating agent. *Res. Chem. Intermed.* **2012**, *38*, 2479-2489.
- (6) Chavan, S.; Shelke, R.; Degan, M. Carboxylic acid-catalyzed one-pot synthesis of cyanoacetylureas and their cyclization to 6-aminouracils in guanidine ionic liquid. *Monatsh. Chem.* **2013**, *144*, 399-403.
- (7) Baraldi, P. G.; Preti, D.; Tabrizi, M. A.; Fruttarolo, F.; Romagnoli, R.; Zaid, N. A.; Moorman, A. R.; Merighi, S.; Varani, K.; Borea, P. A. New pyrrolo[2,1-*f*]purine-2,4-dione and imidazo[2,1-*f*]purine-2,4-dione derivatives as potent and selective human A₃ adenosine receptor antagonists. *J. Med. Chem.* **2005**, *48*, 4697-4701.
- (8) Luo, Y.; Zhang, S.; Liu, Z.-J.; Chen, W.; Fu, J.; Zeng, Q.-F.; Zhu, H.-L. Synthesis and antimicrobial evaluation of a novel class of 1,3,4-thiadiazole: Derivatives bearing 1,2,4-triazolo[1,5-*a*]pyrimidine moiety. *Euro. J. Med. Chem.* **2013**, *64*, 54-61.
- (9) Filippakopoulos, P.; Knapp, S. The bromodomain interaction module. *FEBS Lett.* **2012**, *586*, 2692-2704.
- (10) Kabsch, W. XDS. *Acta Crystallogr. D. Biol. Crystallogr.* **2010**, *66*, 125-132.
- (11) Collaborative Computational Project, Number 4. The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D Biol. Crystallogr.* **1994**, *50*, 760-763.
- (12) Emsley, P.; Lohkamp, B.; Scott, W. G.; Cowtan, K. Features and development of Coot. *Acta Crystallogr. D. Biol. Crystallogr.* **2010**, *66*, 486-501.
- (13) Larkin, M. A.; Blackshields, G.; Brown, N. P.; Chenna, R.; McGettigan, P. A.; McWilliam, H.; Valentin, F.; Wallace, I. M.; Wilm, A.; Lopez, R.; Thompson, J. D.; Gibson, T. J.; Higgins, D. G. Clustal W and Clustal X version 2.0. *Bioinformatics* **2007**, *23*, 2947-2948.
- (14) Gouet, P.; Courcelle, E.; Stuart, D. I.; Metz, F. ESPript: analysis of multiple sequence alignments in PostScript. *Bioinformatics* **1999**, *15*, 305-308.

*Partie 3 : Optimisation par SAR
de la molécule 5A14 :
Sonde épigénétique sélective ?*

I. Méthode d'optimisation originale : « S4MPLE »

Lorsque nous avons initié le programme d'optimisation par SAR pour la molécule **5A14**, j'avais identifié par HTRF la molécule **1916** (3-(4-chlorobenzyl)-7-ethyl-1H-purine-2,6(3H,7H)-dione) présentant une activité au micromolaire. Notre stratégie d'optimisation (**Figure 35**) a été initiée à partir de ce « hit » dont le mode de liaison a été déterminé par cristallographie (PDB : **5EIS**). L'optimisation en position *para* du groupement phényle semblait naturelle, mais elle nécessitait d'une part l'ajout d'un « atome espaceur » pour des raisons de nature stérique (contraintes du site de liaison), et d'autre part l'introduction d'une nouvelle fonction réactive pour permettre la synthèse d'une série de dérivés. Les chimistes ont donc synthétisé le dérivé nucléophile de type « benzylamine » (3-(4-(aminométhyl)benzyl)-7-ethyl-1H-purine-2,6(3H,7H)-dione) qui a servi par la suite d'intermédiaire de synthèse. Par ailleurs, l'analyse de la structure cristallographique du complexe BRD4(1)-**1916** (PDB : **5EIS**) a suggéré une optimisation de type sulfonamide comme étant compatible avec les contraintes du site de liaison. Nous avons appliqué une méthode originale d'optimisation générique de composé (« hit-2-lead » et/ou « growing » de fragment) combinant une approche de modélisation moléculaire, une synthèse organique semi-automatisé (plateforme DOSynth, **page 201**) et un criblage HTRF automatisé (plateforme IPC_{DD}, **page 201**).

Figure 35 : Méthode S4MPLE pour l'optimisation de 5A14.



Légende : La molécule de départ 1916 a été modifiée afin d'obtenir un intermédiaire de synthèse de type « benzylamine ». Cette « benzylamine » a ensuite été couplé avec divers « building blocks » de type « chlorure de sulfonyle » pour conduire à des analogues de type « sulfonamide ».

Notre stratégie d'optimisation *in silico* repose sur deux étapes principales : Premièrement, la création d'une chimiothèque virtuelle diverse et composée de molécules satisfaisantes d'un point de vue de chimie médicinale. Dans un second temps, le criblage virtuel sous contrainte de cette chimiothèque en utilisant l'outil de modélisation moléculaire S4MPLE (228, 229) qui a été co-développé par le Dr. Laurent Hoffer (membre de notre équipe) au cours

de sa thèse dans le Laboratoire de Chémo-informatique de l'Université de Strasbourg sous la direction du Dr. Horvath.

Ce programme d'échantillonnage conformationnel repose sur un algorithme génétique de type Lamarckien et un champ de forces AMBER/GAFF pour les évaluations énergétiques.

En pratique, la chimiothèque virtuelle est composée de molécules théoriquement accessibles et a été générée à l'aide (i) de l'intermédiaire de synthèse incluant une amine terminale primaire, (ii) d'une règle de synthèse organique virtuelle de type sulfonamide et (iii) d'une banque de « Building-Blocks » (BB) commercialement disponibles et contenant une fonction chlorure de sulfonyle. Cette chimiothèque a ensuite été traitée afin de ne sélectionner que les composés satisfaisants d'un point de vue de chimie médicinale. Du fait de sa taille plutôt réduite (nombre limité de BB contenant une fonction chlorure de sulfonyle), un algorithme minimal de « clusterisation » a été appliqué afin d'extraire un sous-ensemble représentatif : environ 600 composés sur un total à ce stade de 800 molécules. Ce sous-ensemble a ensuite été criblé virtuellement avec S4MPLE en utilisant une contrainte sur le positionnement de la sous-structure xanthine dont le mode de liaison est connu (PDB : **5EIS**). Cette dernière interagit avec l'asparagine N140 du domaine BRD4(1) et l'on souhaite maintenir cette interaction fondamentale tout au long du processus d'optimisation. En effet, toute stratégie d'optimisation vise en premier lieu à augmenter l'affinité d'un composé tout en maintenant son mode de liaison initial. Le top 100 d'un point de vue énergétique a été considéré afin de sélectionner les composés à synthétiser. Au final, une trentaine de BB a été commandée à la suite d'une analyse visuelle des poses générées tout en affichant une certaine diversité chimique. Enfin, les composés cibles ont été synthétisés en parallèle à l'aide d'une plateforme robotique « ChemSpeed » (modèle Accelerator Synthetizer SLT100) et ils ont ensuite été évalués par HTRF à l'aide d'un robot « Labcyte Access/Echo® », sans étape intermédiaire de purification. Ces plateformes sont présentées en page 201.

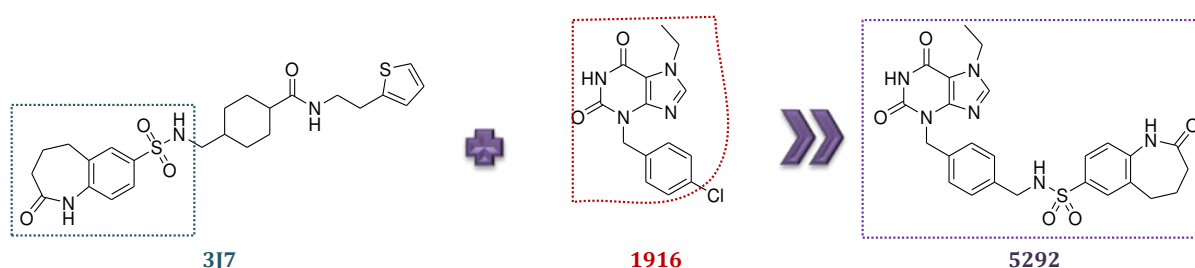
Cette phase d'optimisation a été réalisé en concertation entre les chémo-informaticiens (les Drs. Philippe Roche, Laurent Hoffer et Christophe Muller), les chimistes (Dr. Sébastien Combes et Yulia Voitovich) et les structuralistes de l'équipe (Dr. Stéphane Betzi et moi-même).

II. Premier dérivé sulfonamide issues de S4MPLE : « 5292 »

1. Identification et validation de la molécule 5292

Parmi les building-blocks identifiés dans le top 100, se trouvait une oxobenzodiazépine (benzazépine-2-one). Or les benzodiazépines sont des groupements chimiques privilégiés en chimie médicinale, ils sont par exemple retrouvés dans les médicaments anxiolytiques, tel que le bromazépan, Lexomil®. De plus, la molécule **3J7** (BET-BD1 sélective) que nous avons identifié lors du criblage de la chimiothèque 2P2I_{3D} contenait également ce groupement chimique. C'est pourquoi nous avons dans un premier synthétisé et évalué cette molécule présentée dans la **Figure 36** et baptisée **5292**.

Figure 36 : Structure chimique de 5292, un hybride entre 3J7 & 1916.

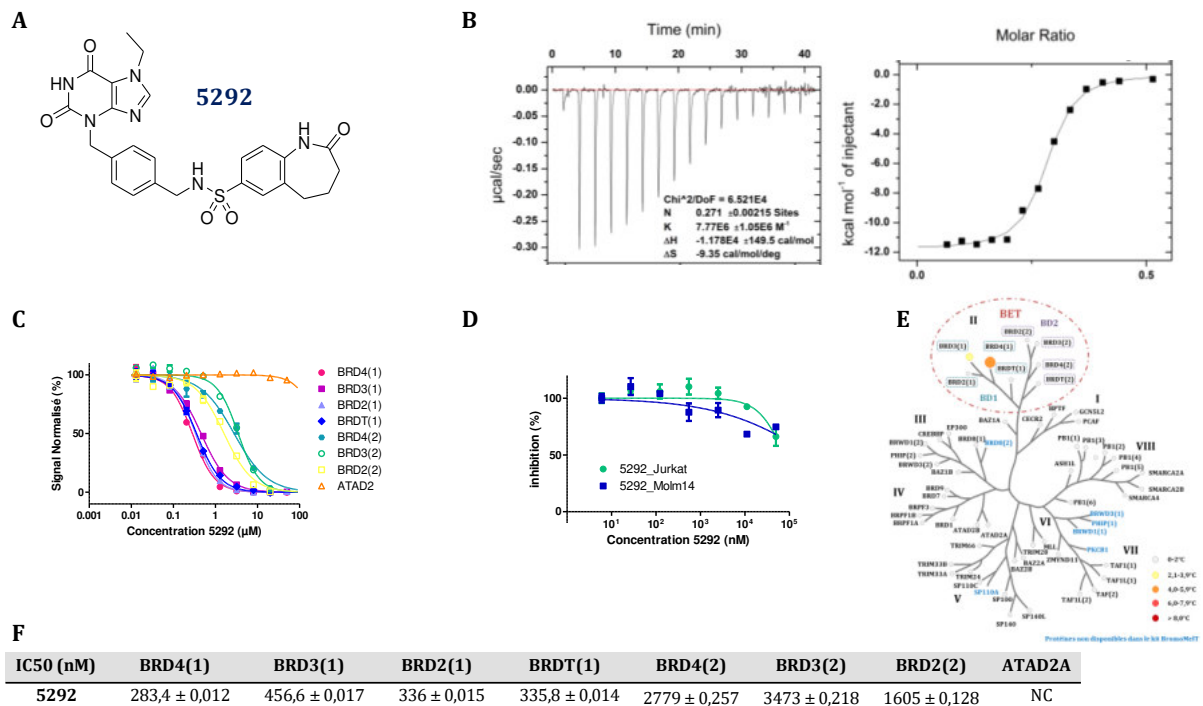


Légende : Le groupe encadré en bleu (oxobenzodiazépine avec le groupe sulfonamide) de la molécule 3J7 a été couplé avec le groupe encadré en rouge (xanthine avec le benzyle) de la molécule 1916, pour donner la molécule hybride 5292 encadrée en violet.

L'ensemble des résultats d'identification et validation orthogonale de **5292** sont présentés dans la **Figure 37**. J'ai identifié par HTRF, cette molécule comme inhibiteur de BRD4(1) avec une affinité (IC_{50}) de 283 nM. Par ailleurs, cette affinité est quasiment trente fois mieux que la molécule **3J7** (IC_{50} 7,5 μ M) et presque dix fois mieux que **1916** (IC_{50} 2,6 μ M). Afin de déterminer si cette molécule était BET-BD1 sélective comme **3J7** ou pan-BET comme **1916**, je l'ai évalué par HTRF sur les autres BDs de la famille BET. Le profil de sélectivité a montré que la molécule **5292** était plutôt pan-BET inhibitrice avec une sélectivité d'au moins un demi-log entre les BD1 et les BD2. L'affinité est comprise entre 283 et 457 nM pour les domaines BD1, entre 1,6 et 3,5 μ M pour les domaines BD2 des BETs et ne présente aucune activité d'inhibition sur ATAD2A. De plus, j'ai confirmé l'affinité pour le domaine BRD4(1) par la mesure de la valeur du K_D par la méthode d'ITC, celle-ci est de 128 nM. Enfin, j'ai évalué sa cytotoxicité dans des cellules leucémiques Jurkat et Molm14 et les valeurs d' EC_{50} étaient semblables pour les deux lignées et supérieures à 50 μ M, ce qui est surprenant en comparaison de son affinité mais pouvait être expliquée par sa faible solubilité en présence du milieu cellulaire. J'ai également

mesuré le ΔT_m induit par cette molécule pour quasiment tous les BDs à l'aide du kit BromoMelt et *via* la technique de TSA. J'ai fixé un seuil à $+2^\circ\text{C}$. Cette molécule stabilise uniquement les BDs suivants BRD3(1) avec $+3^\circ\text{C}$ et BRD4(1) avec $+4,5^\circ$ et le ΔT_m observé pour les domaines BD2 des protéines BETs est inférieur ou égal à $+1^\circ\text{C}$. Ceci reflète la préférence pour les domaines BD1 des protéines de la famille BET.

Figure 37 : Identification, Validations orthogonale et cellulaire de 5292.



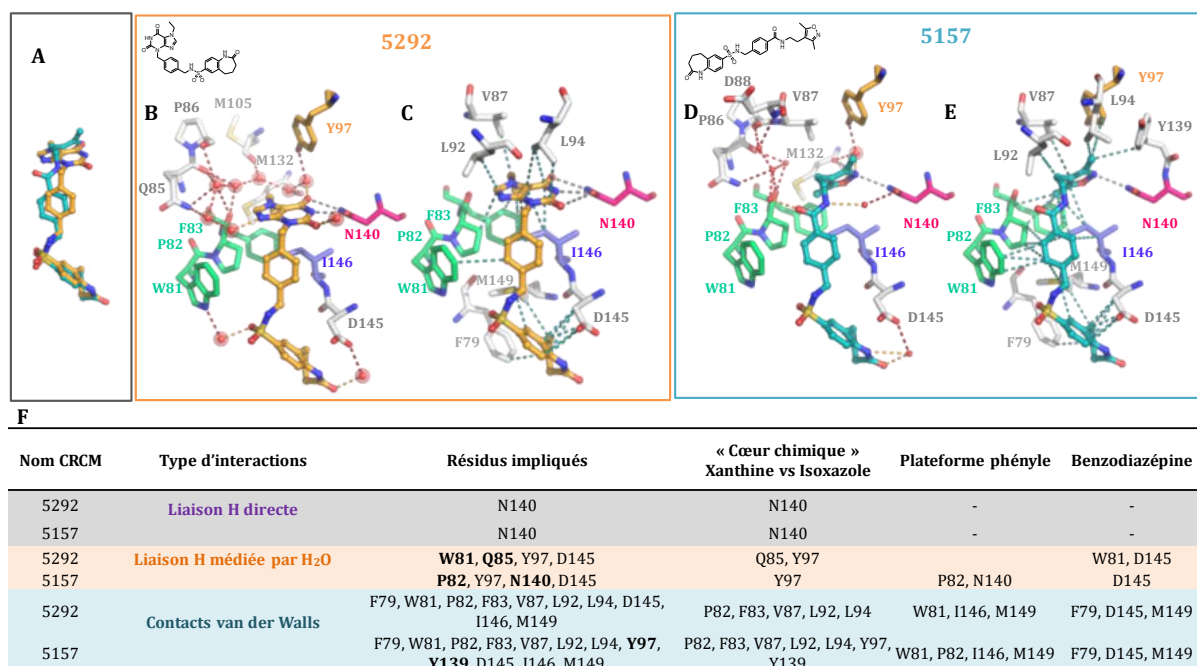
Légende : Évaluation de la molécule 5292. **A-** Structure chimique de 5292 (4-[(2-oxo-2,3,4,5-tetrahydro-1H-1-benzazepine-7-sulfonamido)méthyl]-3-benzyl-7-éthyl-1H-purine-2,6(3H,7H)-dione)). **B-** Validation orthogonale par ITC : à gauche thermogramme et à droite l'intégration des données brutes en suivant un modèle à un site et un ajustement des moindres carrés. La titration inversée a été réalisée à 15°C ($40\ \mu\text{M}$ de ligand 5292 dans la seringue et $100\ \mu\text{M}$ de BRD4(1) dans la cellule de mesure). Le K_D mesuré est de $283\ \text{nM}$. **C-** Évaluation par HTRF de 5292 a été évalué sur l'ensemble des membres de la famille BET, ainsi que sur le domaine ATAD2A (famille IV) comme contrôle. Les courbes sont colorées en fonction des BDs comme indiqué dans la figure. L' IC_{50} mesurée pour BRD4(1) est de $283\ \text{nM}$. **D-** Effet de 5292 sur la viabilité cellulaire sur les lignées leucémiques Jurkat et Molm14 en fonction de la concentration et après incubation pendant 72h à 37°C . Les EC_{50} mesurées sont supérieures à $50\ \mu\text{M}$. **E-** Les ΔT_m induit par la molécule 5292 ont été mesurés à l'aide du kit BromoMelt sur quasiment tout le BDs excepté ceux indiqué en bleu. Les valeurs du ΔT_m (ratio BD-ligand comme indiqué dans le kit 5X protéine et $25\ \mu\text{M}$ d'inhibiteur) sont indiquées par un code couleur comme indiqué dans la légende en bas à droite de la figure. Un seuil a été fixé à $+2^\circ\text{C}$. Cette molécule stabilise uniquement BRD3(1) avec $+3^\circ\text{C}$ et BRD4(1) avec $+4,5^\circ\text{C}$ et le ΔT_m observé pour BRD2(1) est de $+1,75^\circ\text{C}$ et pour les domaines BD2 des protéines BETs, ainsi que les autres BDs il est inférieur ou égal à $+1^\circ\text{C}$. **F-** Valeurs d' IC_{50} pour 5292 mesurées par HTRF.

2. Caractérisation du mode de liaison de la molécule 5292

Lors de l'optimisation par SAR de la molécule **3J7** (une soixantaine de molécules), j'avais identifié une molécule **5157**, présentant une affinité semblable à **3J7** (IC_{50} 9,9 μ M et K_D 5,8 μ M pour BRD4(1)) et un profil de sélectivité BET-BD1 (IC_{50} comprise entre 6,7 et 12,5 μ M pour les domaines BD1 de la famille BET et aucune inhibition à 50 μ M pour les domaines BD2 de la famille BET, ainsi que le domaine ATAD2A), pour laquelle j'avais résolue la structure cristallographie en complexe avec BRD4(1).

J'ai pu comparer les structures cristallographiques de BRD4(1) en complexe avec **5292** (PDB : **à déposer**, 0,95 Å) et **5157** (PDB : **à déposer**, 1,1 Å), afin d'identifier les interactions impliquées dans le profil de sélectivité BET-BD1 de **5157** et pan-BET de **5292** (**Figure 38**). Le groupement xanthine interagit comme attendu avec les résidus asparagine N140, tyrosine Y97 et glutamine Q85. Les cycles sulfonamidobenzodiazépine se superposent parfaitement. Par ailleurs, pour les deux molécules, on constate que l'atome d'oxygène du groupement sulfonamide interagit avec le tryptophane W81 du « WPF » *via* une molécule d'eau et l'atome d'oxygène de l'oxobenzodiazépine interagit avec l'acide aspartique D145 *via* une molécule d'eau. J'ai constaté deux interactions médiées par des molécules d'eau différentes entre **5292** et **5157**. En effet, le composé **5157** interagit avec les résidus proline P82 du « WPF » et asparagine N140, tandis que **5292** interagit avec les résidus tryptophane W81 du « WPF » et glutamine Q85 de la boucle ZA. De plus, **5157** établit deux contacts de type van der Waals supplémentaires avec les résidus tyrosines : Y97 et Y139 respectivement de la boucle ZA et de l'hélice α_c . **Les interactions spécifiques à la molécule 5157 pourraient être à l'origine de la sélectivité BD1-BET observée pour celle-ci.**

Figure 38 : Comparaison des modes de liaisons des molécules 5292 et 5157.



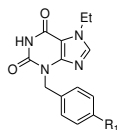
Légende : Les conditions de cristallisation et les tables de statistiques sont en annexe. Encadré en orange les interactions établies par 5292 et encadré en bleu celles établies par 5157. **A-** Superposition des molécules 5292 et 5157. **B & C-** Les liaisons hydrogènes établies dans la cavité K_{ac} sont représentées par des lignes en pointillés, de couleurs grise -- pour les liaisons directes entre la BRD4(1) et l'inhibiteur, en orange --- pour les liaisons indirectes entre l'inhibiteur, une molécule d'eau et BRD4(1) et en rouge --- pour le réseau d'interactions entre les molécules d'eau. **D & E-** Contacts de van der Walls établis entre BRD4(1) et les molécules (sont représentées par des lignes en pointillés de couleur bleue ---. **F-** Tableau récapitulatif des interactions établies par 5292 et 5157.

3. Optimisation de la molécule 5292 : « hit to lead »

L'affinité de la molécule **5292** pour les BDs de la famille BET ayant des critères proches d'une sonde, nous avons poursuivi nos efforts pour améliorer son activité et surtout sa sélectivité afin de la qualifier comme sonde épigénétique. Dans un premier temps, nous avons reconstruit la molécule **5292** (une dizaine de molécules), à partir du cœur « 3-benzyl-7-ethylxanthine » correspondant à la molécule **5327**. Comme indiqué dans le **Tableau 15**, tous les dérivés sont actifs et les valeurs d' IC_{50} que j'ai mesuré sont comprises entre 283 nM et 27,1 μ M. Par ailleurs, la molécule **5292** est la plus active de ces dérivés. Dans un second temps, nous avons greffé en position C-8 le groupement triazolopyrimidine de la molécule **5A14**, substitué ou non par un butyle (**5316** et **5318** respectivement). L'objectif était de pouvoir observer une affinité similaire au composé **5292** tout en conservant la sélectivité de **5A14** pour BRD4(1). J'ai déterminé les profils de sélectivité pour ces deux molécules, présentés dans le **Tableau 16**. Ces molécules ont une affinité (IC_{50}) pour BRD4(1) autour de 1 μ M, mais elles sont pan-BET inhibitrices avec une préférence d'un demi-Log pour BRD4(1).

Tableau 15 : Valeurs d'IC₅₀ mesurées par HTRF pour les analogues de 5292.

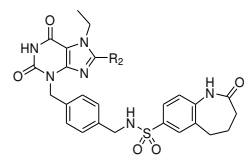
A Cœur chimique



Composés Nom CRCM	Substituants R ₁	BRD4(BD1)		
		IC ₅₀ (μM)	K _D (μM)	LE
1916	Cl	2,6 ± 0,1	1,8	0,37
1917	H	11,97 ± 0,6	ND	0,32
5327		27,1 ± 3,4	ND	0,29
5328		5,3 ± 0,3	ND	0,28
5329		1,1 ± 0,6	ND*	0,27
5362		0,687 ± 0,5		0,27
5331		0,492 ± 0,045	ND*	0,28
5330		1,3 ± 0,7	ND*	0,24
5292		0,283 ± 0,012	0,128	0,25

B Cœur chimique

« 5292 »



Composés Nom CRCM	Substituants R ₂	BRD4(BD1)		
		IC ₅₀ (μM)	K _D (μM)	LE
5318		0,964 ± 0,076	ND*	0,17
5316		1,1 ± 0,064	ND*	0,15

Légende : L'ensemble de ces molécules ont été évaluées sur BRD4(1). Les valeurs mesurées (IC₅₀ et K_D) sont indiquées en micromolaire. Les IC₅₀ ont été mesurées par HTRF et le K_D a été mesuré par ITC. Le Ligand Efficiency (LE) est calculé à partir de la formule $LE = (\Delta G)/N = 1,4 \cdot (-\log(IC_{50}))/N$ (226). A- Analogues issues de la reconstruction la molécule 5292. B- Analogues issues de la condensation de 5292 avec le groupement triazolopyrimidine de 5A14.

Tableau 16 : Profils de sélectivités des analogues de 5318 et 5316.

IC50 (μM)	BRD4(1)	BRD3(1)	BRD2(1)	BRDT(1)	BRD4(2)	BRD3(2)	BRD2(2)	ATAD2A
5318	0,964 ± 0,1	9,6 ± 0,5	4,4 ± 0,4	9,5 ± 0,9	28,0 ± 2,9	7,3 ± 0,5	3,9 ± 0,3	NC
5316	1,0 ± 0,1	4,4 ± 0,2	4,7 ± 0,4	18,9 ± 3,5	> 50	4,1 ± 0,5	3,3 ± 0,6	NC

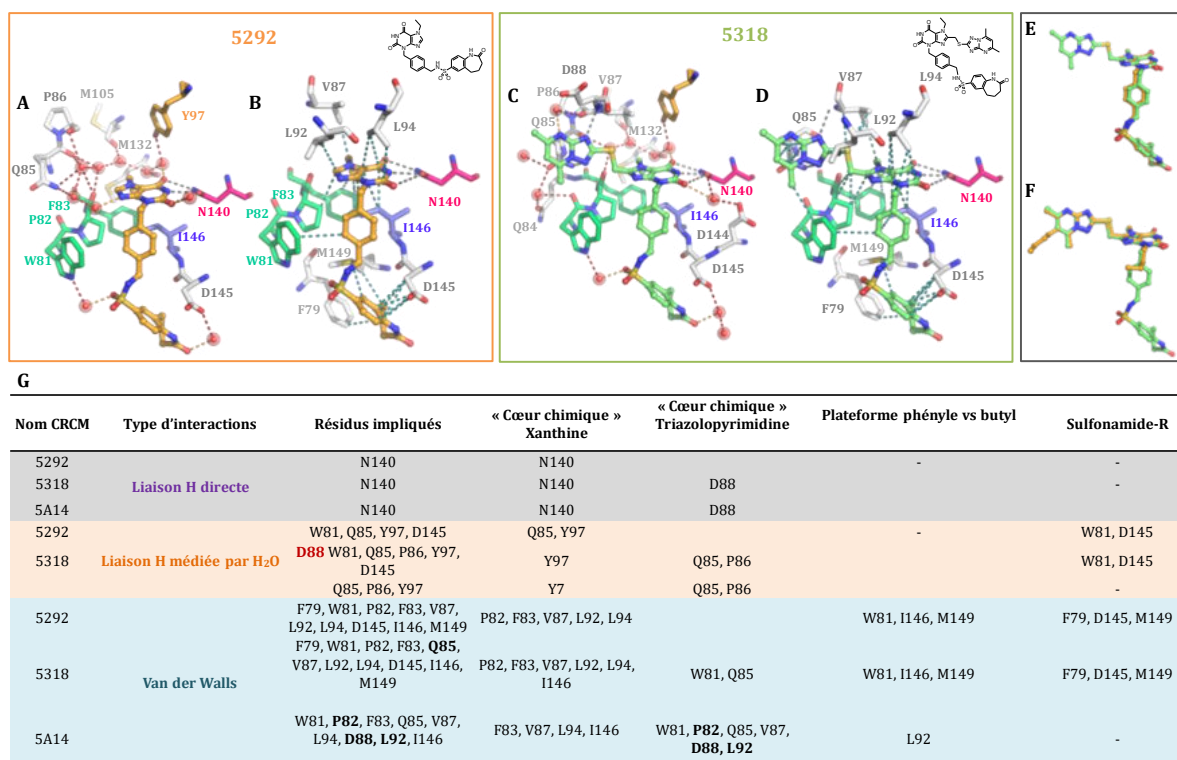
Légende : Les profils de sélectivité des molécules 5318 et 5316 ont été déterminés par HTRF. Les valeurs d'IC₅₀ sont indiquées en micromolaire (μM).

J'ai résolu la structure du domaine BRD4(1) en complexe avec la molécule **5318** (**Figure 39**). Comme observé pour la molécule **5A14**, **5292** établit des liaisons hydrogènes avec les résidus clefs asparagine N140 et tyrosine Y97 et les résidus de la boucle ZA : Q85, P86 et D88. D'autres liaisons hydrogènes indirectes, médiées par des molécules d'eau sont établies par les groupements sulfonamide avec le résidu tryptophane W81 et oxobenzodiazépine avec l'acide aspartique D145.

Le groupement xanthine établit aussi des contacts de type van der Waals avec les résidus de la boucle ZA : P82, F83, V87, L92, L94 et le résidu « gatekeeper » I146. Le benzyle établit des contacts avec les résidus W81 du « WPF » et de l'hélice α_c : M139 et le « gatekeeper » I146. L'oxobenzodiazépine établit quant à elle, des interactions avec les résidus F79 de la boucle ZA, D145 et M149 de l'hélice α_c. En comparant les structures des molécules **5292**, **5318** et **5A14** (**Figure 39**), on constate que la molécule **5318** établit une liaison hydrogène supplémentaire avec l'acide aspartique D88, tandis que la molécule **5292** établit une liaison supplémentaire médiée par une molécule d'eau avec la glutamine Q85.

La préférence des molécules 5318 et 5316 pour le domaine BD1 de la protéine BRD4 pourrait quant à elle être expliquée par la liaison directe supplémentaire avec l'acide aspartique D88. Par ailleurs, cette liaison a été observée dans la structure du complexe BRD4(1)-5A14, or la molécule 5A14 est BRD4(1) sélective.

Figure 39 : Comparaison de mode de liaison des molécules 5292, 5318 et 5A14.



Légende : Les conditions de cristallisation et les tables de statistiques sont en annexe. Encadré en orange les interactions établies par 5292 et encadré en vert celles établies par 5318. **A & B-** Comme précédemment, les liaisons hydrogènes établies dans la cavité K_{ac} sont représentées par des lignes en pointillés, de couleurs grise -- pour les liaisons directes entre la BRD4(1) et l'inhibiteur, en orange --- pour les liaisons indirectes entre l'inhibiteur, une molécule d'eau et la BRD4(1) et en rouge --- pour le réseau d'interactions entre les molécules d'eau. **B & D-** Contacts de van der Walls établis entre BRD4(1) et les molécules sont représentées par des lignes en pointillés de couleur bleue ---. **E-** superposition de 5318 (vert) et 5292(orange). **F-** superposition de 5318 (vert) et 5A14 (orange). **G-** Tableau récapitulatif des interactions établies par 5292 et 5318 et 5A14.

Le couplage des cœurs xanthine et oxobenzodiazépine des molécules 1916 (pan-BET) et 3J7 (BET-BD1 sélective) nous a permis d'améliorer l'affinité d'un facteur trente par rapport à 3J7 et presque de dix par rapport à 1916. Ainsi nous avons conçu une sonde épigénétique de type pan-BET. Le couplage de la molécule 1913 (correspondant à la molécule 5A14 BRD4(1) sélective pour laquelle le groupe butyle du cœur xanthine a été remplacé par un chlorobenzyle) et du cœur oxobenzodiazépine de 3J7 (BET-BD1 sélective) a permis d'augmenter l'affinité d'un facteur vingt par rapport à 1913. Contrairement à 5A14, la molécule 5316 est pan-BET inhibitrice avec une préférence d'un demi-log pour les domaines BD1. Il semblerait que dès lors qu'on améliore l'affinité de nos molécules on perd la sélectivité.

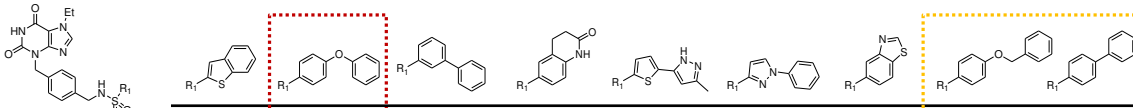
III. Deuxième série de dérivés sulfonamides de 5292 issues de S4MPLE

1. Évaluation et validation des autres dérivés

Une soixantaine d'analogues (des dérivés sulfonamides et amides) ont été synthétisés dans du tampon triéthylamine (TEA), utilisé comme base, en une seule étape à l'aide du robot de synthèse semi-automatisé Chemspeed et sans étape de purification. J'ai évalué par HTRF, l'ensemble de ces molécules non isolées, ainsi que le tampon TEA comme contrôle négatif sur les domaines BD1 et BD2 de la protéine BRD4 (BRD4(1) et BRD4(2)). Parmi ces molécules, quinze dérivés sulfonamides et un seul dérivé amide ont montré une affinité (IC_{50}) pour BRD4(1) inférieure à 1 μ M. Neuf des dérivés sulfonamides se sont avérés meilleurs pour le domaine BD1 de la protéine BRD4 d'un facteur allant de 15 à 62. J'ai confirmé par HTRF les résultats pour ces neuf molécules sur un nouveau lot de composés, cette fois-ci purifiés. J'ai ensuite déterminé leurs profils de sélectivité au sein de la famille BET. Comme l'indique le **Tableau 17**, les trois molécules répondant au critère de sonde épigénétique (ratio d'au moins 30 fois entre les domaines BD1 et BD2 de la protéine) sont **5386**, **5392** et **5393**. La molécule **5393** a été abandonnée car elle touchait également le domaine ATAD2A de manière non négligeable par rapport aux domaines BD2 de la famille BET.

En parallèle, j'ai déterminé l'effet de ces molécules non purifiées et purifiées sur la viabilité cellulaire sur les lignées leucémiques Jurkat et Molm14. Les résultats sont présentés dans **Tableau 17**. Pour toutes ces molécules excepté **5388**, les valeurs d' EC_{50} mesurées sont comprises entre 5,2 et 19,4 μ M, ce qui est surprenant au vu des valeurs d' IC_{50} obtenues. Par ailleurs, les valeurs mesurées sont semblables pour les molécules non purifiées et celles purifiées. Dans le cas de **5388**, la valeur d' IC_{50} de 50 μ M peut être expliquée par la propension de cette molécule à former des cristaux dans le milieu cellulaire. Il faut tout de même noter que la plupart de ces composés précipitaient aux fortes concentrations, de ce fait les valeurs d' EC_{50} mesurées sont probablement sous-estimées.

Tableau 17 : Évaluation HTRF & viabilité cellulaire des neuf dérivés sulfonamides.

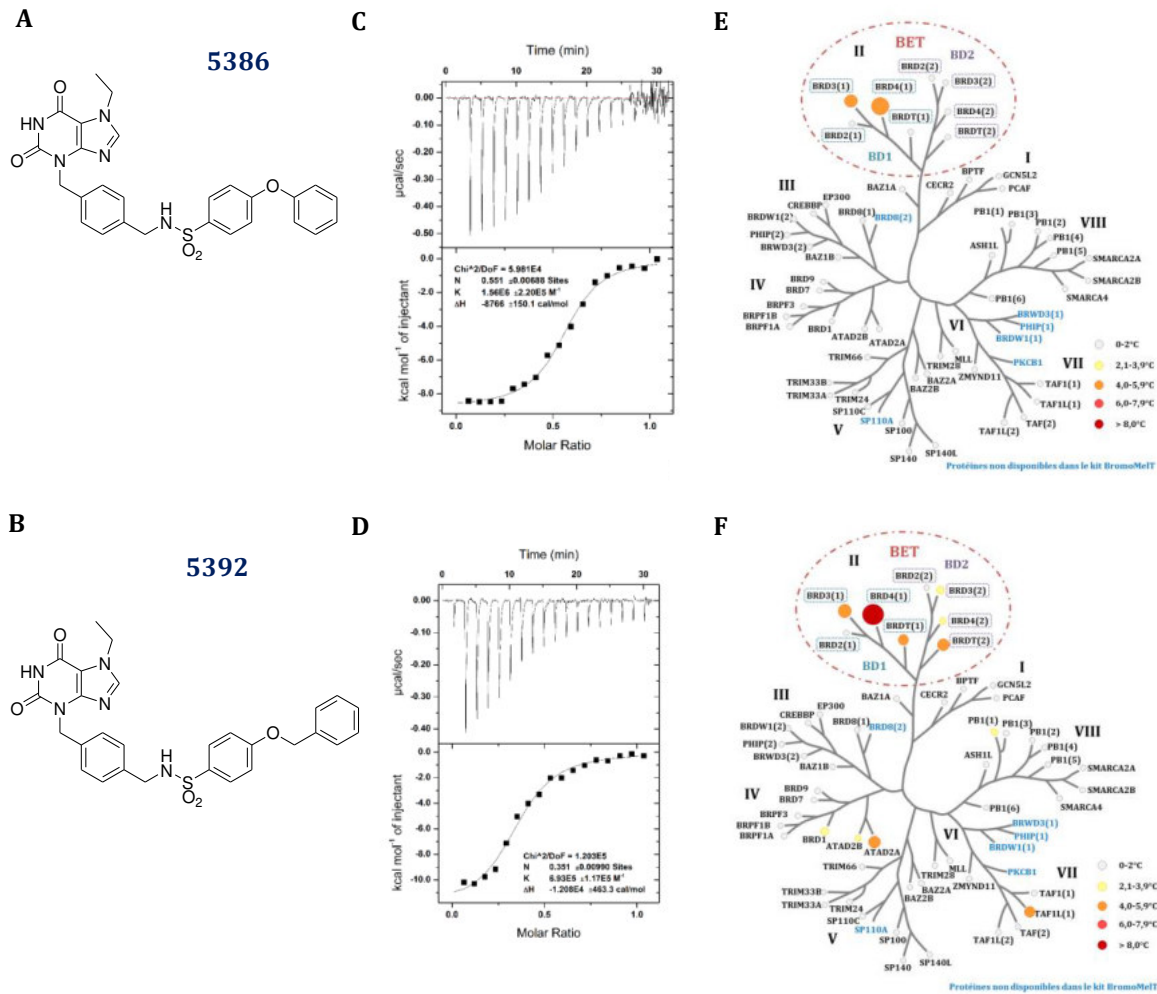


	5385	5386	5387	5388	5389	5390	5391	5392	5393
BDs	HTRF IC₅₀ (nM)								
BRD4(1)	779	379	814	766	284	846	1595*	355	450
BRD3(1)	1536	837	1596	940	507	1004	4767	596	897
BRD2(1)	1384	2263	5470	3488	2105	3334	4177	1456	3314
BRDT(1)	7437	> 50000	7440	2150	3471	6257	14043	8508	19659
BRD4(2)	12802	> 50000	> 12165*	8314	4054	5848	21185	9708	17723
BRD3(2)	12360	> 50000	18931	3031	3168	11982	17428	13462	13538
BRD2(2)	10486	> 50000	14698	5577	3198	6050	23514	12771	15988
ATAD2A	> 50000	> 50000	> 50000	> 50000	> 50000	> 50000	> 50000	> 50000	24647*
Sonde BRD4	Non	Oui	Non	Non	Non	Non	Non	Oui	Oui
Sonde BRD3	Non	Oui	Non	Non	Non	Non	Non	Non	Non
Sonde BRD2	Non	Non	Non	Non	Non	Non	Non	Non	Non
Lignées cellulaires	EC₅₀ (nM)								
Jurkat	5220	10228	6920	>50000*	13494	14174	6156	19422	26500
Molm14	11082	9212	8929	>50000*	9950	10147	14813	17941	10600
LE (BRD4(1))	0,25	0,24	0,23	0,24	0,25	0,24	0,24	0,23	0,24

Légende : Certaines de ces molécules interfèrent avec le signal d'HTRF à 50 µM, c'est de le cas de 5385, 5387, 5388, 5389 et 5392 et également à 20 µM pour 5388 et 5389. Les molécules sont qualifiées de sonde lorsque d'une part leur affinité est inférieure à 1 µM et d'autre part lorsque le rapport entre les affinités du domaine BD1 et BD2 est supérieur à 30. La molécule 5388 forme des cristaux dans le milieu cellulaire ce qui explique la valeur d'EC₅₀ de 50 µM mesurée. Les valeurs d'IC₅₀ et d'EC₅₀ sont indiquées en nanomolaire (nM). Le Ligand Efficiency (LE) est calculé à partir de la formule $LE = (\Delta G)/N = 1,4 (-\log (IC_{50}))/N$ (226).

J'ai par la suite confirmé les molécules **5386** et **5392**, par ITC et TSA (**Figure 40**). Les valeurs de K_D mesurées par ITC pour les molécules **5386** et **5392** sont respectivement de 641 nM et 1,4 µM. J'ai mesuré le ΔTm (ratio BD-ligand comme indiqué dans le kit 5X protéine et 25 µM d'inhibiteur) induit par ces molécules sur quasiment tous les BDs à l'aide du kit BromoMelt. La molécule **5386** est sélective et stabilise les domaines suivants : BRD4(1) et BRD3(1), respectivement avec des ΔTm de +5,5 °C et +4,0 °C. La molécule **5392** est moins sélective et stabilise les domaines suivants : BRD4(1) et BRD3(1), ATAD2A et TAF1L, respectivement avec des ΔTm de +8,0°C, +5,0°C, +4,5°C et +4,0°C.

Figure 40 : Validations orthogonales par ITC et TSA de 5386 et 5392.



Légende: A & B- Structures chimique des molécules 5386 en A et 5392 en B. C & D- Validation orthogonale par ITC de 5386 en C et 5392 en D. Les figures d'ITC présentent en haut le thermogramme et en bas l'intégration des données brutes en suivant un modèle à un site et un ajustement des moindres carrés. La titration inversée a été réalisée à 25°C (35 μM de ligand dans la seringue et 200 μM de BRD4(1) dans la cellule de mesure). Les valeurs de K_D mesurées sont respectivement pour 5386 et 5392 de 641 nM et 1,4 μM. E & F- Les ΔT_m induit par les molécules 5386 en E et 5392 en F, ont été mesurés à l'aide du kit BromoMeIT sur quasiment tout le BDs excepté ceux indiqué en bleu. Les valeurs du ΔT_m (ratio BD-ligand, comme indiqué dans le kit 5X protéine et 25 μM d'inhibiteur) sont indiquées par un code couleur comme indiqué dans la légende en bas à droite de la figure. La molécule 5386 est sélective et BRD4(1) et BRD3(1), respectivement avec des ΔT_m de +5,5 °C et +4,0 °C. La molécule 5392 est moins sélective et stabilise BRD4(1) et BRD3(1), ATAD2A et TAF1L, respectivement avec des ΔT_m de +8,0°C, +5,0°C, +4,5°C et +4,0°C.

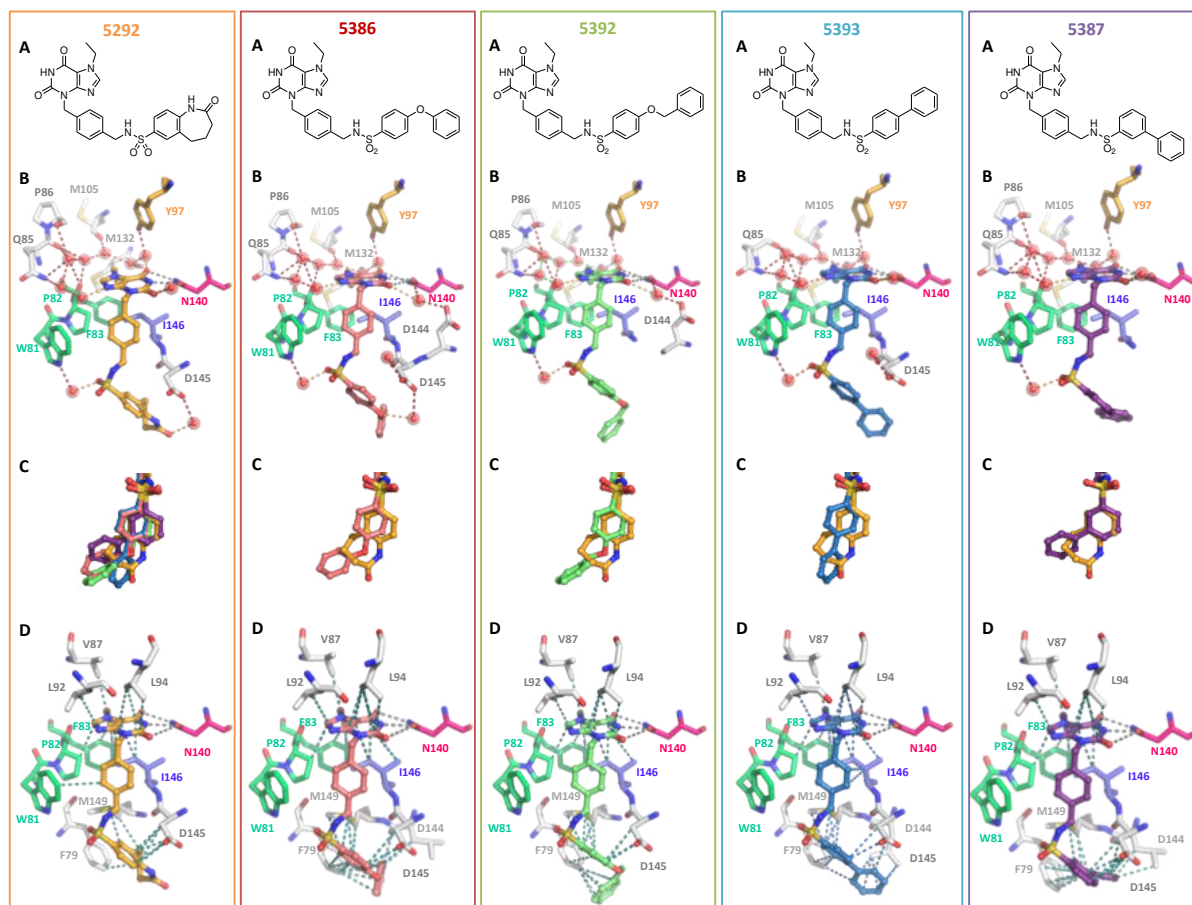
2. Caractérisation du mode de liaison de quatre dérivés sulfonamides

J'ai résolue les structures cristallographiques du domaine BRD4(1) en complexe avec les molécules **5386**, **5387**, **5392** et **5393** et je les ai comparées entre elles et avec celle que j'avais obtenue pour **5292**, afin de comprendre les différents profils de sélectivité observés (**Figure 41**). Les cœurs *N*-benzylxanthine de l'ensemble de ces molécules se superposent, par conséquent elles établissent toutes une interaction directe avec l'asparagine N140 et des interactions indirectes médiées par des molécules d'eau avec les résidus de la boucle ZA *via* le groupement sulfonamide avec le tryptophane W81, et *via* la xanthine avec les résidus glutamine Q85 et tyrosine Y97). La molécule **5386**, *via* groupe (phenoxy)methyl)benzene) établit une interaction indirecte médiée par une molécule d'eau *via* l'atome d'oxygène du groupe phenoxy, avec le résidu d'acide aspartique D145 de l'hélice α_c , tout comme **5292** l'établissait avec l'atome d'oxygène de la benzodiazépinone, ce qui n'est pas le cas des trois autres dérivés sulfonamides.

Ces quatre molécules établissent également des contacts de type van der Waals avec les résidus de la boucle ZA : P82, F83, V87, L92, L94 et le « gatekeeper » I146 *via* la xanthine. Le benzyle interagit quant à lui avec le résidu méthionine M149. Les groupes phénoxybenzène et bi-phényle interagissent quant à eux avec les résidus phénylalanine F79 de la boucle ZA et de l'hélice α_c : D145, M149 ainsi qu'avec la leucine L148 excepté pour les molécules **5292** et **5392**. En étudiant de plus près les structures au voisinage du groupement phénylsulfonamide, on s'aperçoit que le cycle benzénique des molécules **5386**, **5392** et **5393** se superposent, mais pas celui de **5387**. Par contre, aucuns des cycles benzéniques terminaux de ces molécules ne se superposent. Pour les molécules **5386**, **5387** et **5393**, ils sont orientés du côté du « WPF », alors que celui de **5392** reste dans le même axe que le cycle benzodiazépinone de **5292**. Concernant le nombre de contacts de van der Waals établit par chacun des groupements terminaux de ces molécules, on constate que **5386** établit 90% de ses contacts (16 interactions) avec les résidus de la boucle ZA versus 10% avec l'hélice α_c . Les molécules **5392** et **5393** établissent quant à elles 70-75% de leurs contacts (10 interactions) avec les résidus de la boucle ZA versus 25-30% avec l'hélice α_c . Enfin **5387** établit quant à elle autant de contacts (10 interactions) avec les résidus de la boucle ZA qu'avec ceux de l'hélice α_c .

Cette différence pourrait peut-être expliquer la sélectivité observée de la molécule 5386 pour les domaines BRD4(1) et BRD3(1) et des molécules 5387 et 5393 pour le domaine BRD4(1).

Figure 41 : Comparaison des modes de liaison de 5292, 5386, 5387, 5392 et 5393.

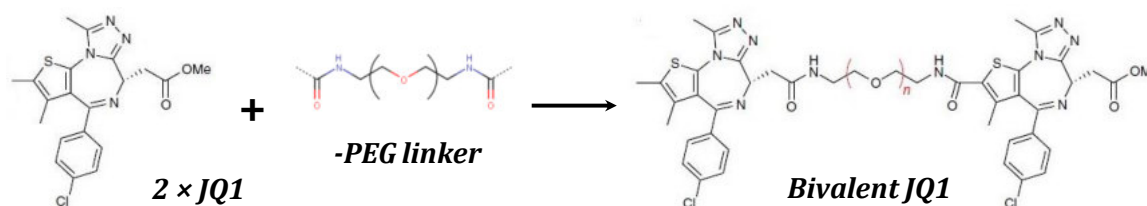


Légende : Les conditions de cristallisation et les tables de statistiques sont en annexe. Encadrés en orange les interactions établies par 5292 (orange ; PDB : **à déposer**, 0,95 Å), en rouge celles établies par 5386 (rose ; PDB : **à déposer**, 1,4 Å), en vert celles établies par 5392 (vert ; PDB : **à déposer**, 1,2 Å), en bleu celles établies par 5393 (bleu ; PDB : **à déposer**, 1,2 Å), et en violet celles établies par 5387 (violet ; PDB : **à déposer**, 1,2 Å). **A-** Structures chimiques des cinq molécules. **B-** Les liaisons hydrogènes établies dans la cavité K_{ac} sont représentées par des lignes en pointillés, de couleurs grise -- pour les liaisons directes entre la BRD4(1) et l'inhibiteur, en orange --- pour les liaisons indirectes entre l'inhibiteur, une molécule d'eau et BRD4(1) et en rouge --- pour le réseau d'interactions entre les molécules d'eau. **C-** Superposition de la partie « sulfonamide » de la molécule 5292 et des autres molécules. **D-** Contacts de van der Waals établis entre BRD4(1) et les molécules sont représentés par des lignes en pointillés de couleur bleue ---.

3. Seconde stratégie d'optimisation : analogues 5A14-PEG

Lors de la résolution de la structure cristallographique du complexe BRD4(1)-**5A14** (PDB : **5EGU**) nous avons pu constater que le domaine BRD4(1) formait un pseudo dimère protéique dans lequel les deux molécules **5A14** présentes dans leur poche K_{ac} respective, se faisaient face en position tête-bêche. *Ce qui nous a amené à imaginer la conception de molécules homodimère et hétérodimère par couplage de deux molécules 5A14 entre elles.* Par ailleurs, deux articles publiés en 2016, portant sur la caractérisation d'un nouvel inhibiteur JQ1 bivalent des bromodomains de la famille BET, nous ont confortés dans notre idée. Cet inhibiteur JQ1 bivalent (**Figure 42**), correspond au couplage de deux molécules de JQ1 entre elles à l'aide de chaînes de PEG de longueur variable. Des différences notables en termes d'activité ont été observées en fonction de la longueur de la chaîne PEG. Le meilleur dimère est le composé MT1 qui présente un effet cytotoxique sur des lignées leucémiques (EC_{50}) de 100 pM, soit cent fois meilleure que JQ1 (230, 231).

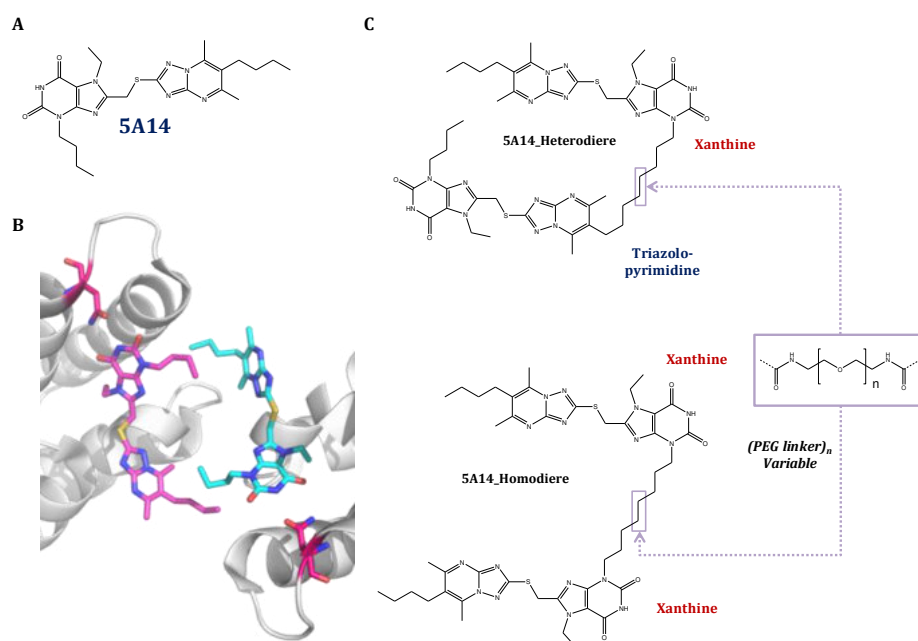
Figure 42 : Molécule JQ1 bivalente.



Légende : Deux molécules JQ1 ont été couplées entre elles via un espaceur de PEG_n de longueur n variable pour former une molécule JQ1 bivalente.

Au vu de ces résultats plutôt exceptionnels, nous avons décidé d'utiliser un espaceur PEG_n de longueur n variable pour la synthèse de nos molécules **5A14** bivalentes. En combinant les informations structurales des complexes BRD4(1)-**5A14** et les données d' IC_{50} , j'ai proposé aux chimistes de synthétiser une molécule **5A14** bivalente de deux manières différentes en utilisant la chaîne butyle soit du cœur xanthine créant ainsi un homodimère ou du cœur triazolopyrimidine créant alors un hétérodimère (**Figure 43**). Au vu de la structure cristallographique BRD4(1)-**5A14** (PDB : **5EGU**), la molécule **5A14** bivalente compatible avec cette structure cristallographique serait un hétérodimère avec un espaceur PEG_n d'une longueur n minimale de n=1. La synthèse d'homodimère de **5A14** est néanmoins plus simple à mettre en place que la synthèse d'hétérodimère.

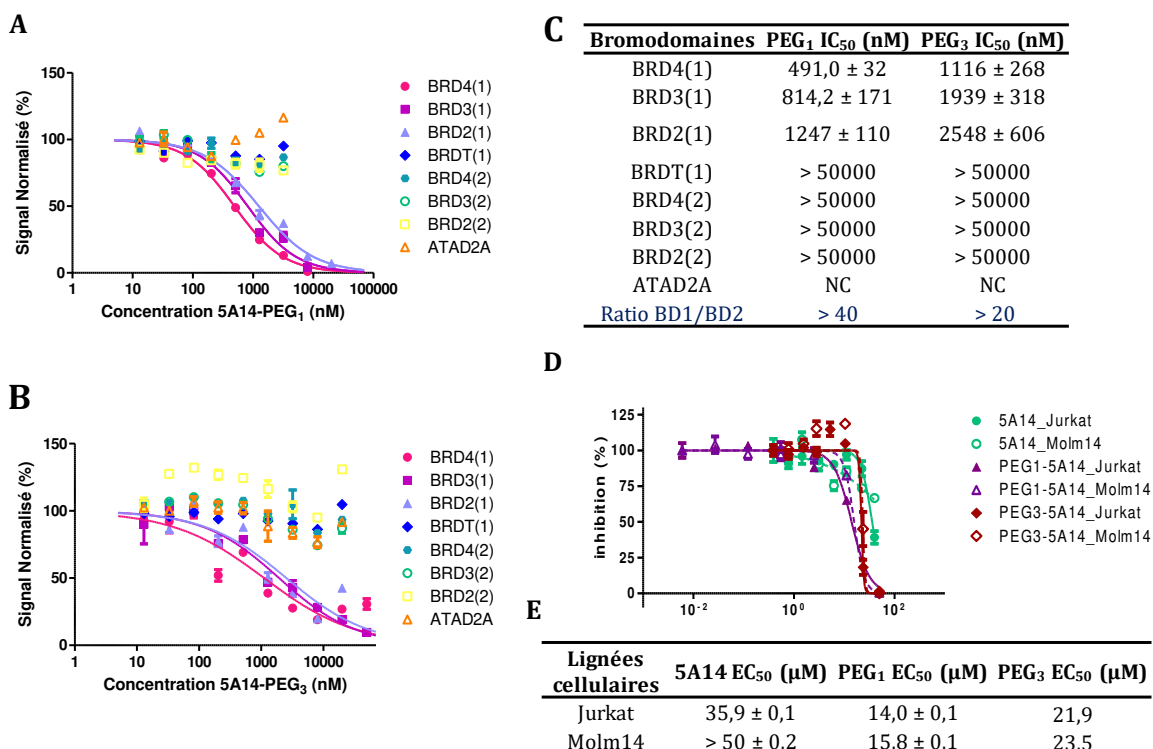
Figure 43 : Molécules 5A14 bivalentes.



Légende : A- Structure chimique de 5A14. B- Structure cristallographique du complexe BRD4(1)-5A14 (PDB : **5EGU**). C- Proposition de deux molécules 5A14 bivalentes, un homodimère et un hétérodimère reliés entre eux par un espaceur de type PEG_n de longueur n variable.

La préparation et la purification d'homodimère de **5A14** s'est avérée plus délicate que prévu en raison de problème de régiosélectivité des réactions de condensation et de stabilité des composés obtenus. À ce jour, j'ai malgré tout pu évaluer deux homodimère avec espaceur PEG_n de longueur n = 1 et 3 et une troisième molécule est en cours d'évaluation avec un espaceur PEG_n de longueur n = 5. L'ensemble des résultats préliminaires sont présentés dans la **Figure 44**. L'affinité de la molécule **5A14-PEG₁** (491 nM) est meilleure que celle de **5A14-PEG₃** (1,1 μM). J'ai également déterminé les profils de sélectivité de ces molécules. Celles-ci sont BET-BD1 sélectives avec un rapport d'au moins quarante fois pour le PEG₁ et vingt fois pour le PEG₃. Les expériences de TSA que j'ai réalisé sur la molécule PEG₁ ont montré un ΔT_m de -2,5°C pour la BRD4(1) versus inférieure à +1°C pour les autres BDs de la famille BET et CREBBP comme contrôle négatif. J'ai par la suite évalué l'effet cytotoxique de ces deux molécules sur les lignées cellulaires leucémiques Jurkat et Molm14. Celles-ci sont plus cytotoxiques que **5A14**. Les valeurs d'EC₅₀ mesurées sont respectivement sur les lignées Jurkat et Molm14, de 14,0 et 15,8 μM pour **5A14-PEG₁** et de 21,9 23,5 μM pour **5A14-PEG₃** versus 35,9 et supérieure à 50 μM pour **5A14**. Nous espérons pouvoir obtenir les structures de BRD4(1) en complexe avec les molécules **5A14** bivalentes, cependant comme pour le complexe BRD4(1)-**5A14** il s'avère que la cristallisation de ces complexes est difficile du fait de leur précipitation dans les gouttes de cristallisation.

Figure 44 : Évaluation des molécules 5A14-PEG_n.



Légende : A à B – Profil de sélectivité des homodimères 5A14-PEG_n avec n=1 et 3. C- Tableau récapitulatif avec les valeurs d'IC₅₀ pour les huit BDs évalués. NC pour non convergent. Les valeurs d'IC₅₀ sont indiquées en nanomolaire (nM). D- Courbes de cytotoxicité sur les lignées Jurkat et Molm14 pour 5A14 et les deux molécules bivalentes. E- Tableau récapitulatif des valeurs d'EC₅₀. Les valeurs d'EC₅₀ sont indiquées en micromolaire (µM).

Ces premiers résultats sont encourageants puisque nous avons réussi à améliorer l'affinité de la molécule 5A14 d'un facteur supérieur à dix, bien que le profil de sélectivité soit passé de BRD4(1) à BD1 sélectif. Nous envisageons de synthétiser d'autres homodimères avec des espaceur PEG_n de longueur n variable pour compléter la série et confirmer nos résultats préliminaires semblant indiquer que plus l'espaceur est long moins la molécule est affine. Comme a priori les hétérodimères auraient une conformation plus compatible avec la structure cristallographique du complexe BRD4(1)-5A14, nous envisageons de synthétiser les hétérodimères analogues afin de comparer les résultats avec ceux obtenus pour les homodimères.

*Chapitre 3 : Conclusions
générales & Perspectives*

Les régulateurs épigénétiques (writers, erasers et readers) jouent un rôle clef dans l'expression des gènes, ainsi que dans divers processus biologiques (réparation de l'ADN, différenciation cellulaire, apoptose...). Par ailleurs, leur dérégulation est impliquée dans le développement de nombreuses pathologies humaines dont les cancers. Depuis une dizaine d'années, les spécialistes du domaine de la régulation épigénétique se sont intéressés plus particulièrement aux « readers » qui étaient jusqu'alors peu étudiés. Les « readers » sont en effet capables de lire les marques épigénétiques, ainsi que de recruter des partenaires d'interaction comme les facteurs de transcription. Les protéines à bromodomains sont des « readers » épigénétiques capables de reconnaître les lysines acétylées des queues des histones. Les protéines de la famille BET et plus particulièrement BRD4 sont très souvent surexprimées dans les maladies inflammatoires et les cancers, c'est pourquoi de nombreuses équipes à travers le monde s'intéressent aux BDs dont ceux de la famille BET. À ce jour, une vingtaine d'essais cliniques sont en cours pour le traitement de différents types de cancers, plus particulièrement les leucémies, mais également les maladies inflammatoires. Ces molécules sont toutes pan-BET, inhibant les huit BDs de la famille BET : BRD4(1-2), BRD3(1-2), BRD2(1-2) et BRDT(1-2).

Le challenge actuel, est de décrypter le rôle individuel et le mécanisme d'action de chacun des BDs de la famille BET *via* le développement de « sondes épigénétiques », mais également de pouvoir traiter spécifiquement certains processus biologiques dérégulés et de pallier les problèmes de résistances liés à l'inhibition pan-BET (214-217), *via* la conception de « candidats médicaments ». Dans ce contexte, j'ai au cours de ma thèse identifié par HTRF, validé par ITC et TSA et caractérisé par cristallographie, des inhibiteurs sélectifs des BDs de la famille BET.

Conclusions générales

La première partie de ce manuscrit concerne l'**évaluation de la chimiothèque 2P2I_{3D}** sur plusieurs complexes protéine-protéine et protéine-peptide, dont le **complexe BRD4(1)-peptide lysines acétylées**. Cette étude, qui a fait l'objet d'une publication en 2016 dans le journal ACS Chemical Biology (218) a permis de valider cette chimiothèque composée de 1664 molécules focalisées sur les interactions PPIs. Sabine Milhas a pu identifier et valider des inhibiteurs pour l'ensemble des complexes évalués, dont sept molécules ciblant les BDs de la famille BET. Pour ma part, j'ai pu caractériser les complexes BD-ligand pour cinq d'entre elles. Parmi ces sept inhibiteurs de BDs, cinq présentaient des profils de sélectivité très classique, de type **pan-BET**. Une seule d'entre elles semblait particulièrement intéressante, à savoir le composé **4N15** qui avait une affinité élevée pour le domaine BRD4(1), proche des critères d'une sonde épigénétique (page 85). Nous avons initié un programme de recherche d'optimisation pour cette molécule, cependant nous n'avons pas encore pu identifier d'analogues plus affins. Nous n'avons pas perdu de vue notre objectif d'améliorer son affinité. *À contrario*, les deux autres (des sept) inhibiteurs présentaient des profils de sélectivité très inattendus, **5A14** étant **sélective du domaine BRD4(1)** et **3J7** étant quant à elle **sélective des domaines BD1 de la famille BET (BET-BD1)**. Bien qu'en terme d'affinité ces molécules étaient très loin des critères d'une sonde épigénétique, elles étaient très intéressantes de par leur profil de sélectivité. En effet, aucune molécule sélective envers un seul BD de la famille BET ou contre les domaines BD1 n'avait été publiée. Seule la molécule RVX208 était décrite comme sélective des domaines BD2 de la famille BET (204). C'est pour cette raison que nous avons concentré tous nos efforts sur ces deux molécules afin d'améliorer leur affinité par SAR et d'obtenir des **sondes épigénétiques sélectives**.

La deuxième partie de ce mémoire s'intéresse à la **compréhension de l'origine de la sélectivité de la molécule 5A14**. Dans ce contexte, nous avons initié un programme de reconstruction de **5A14** à partir de son cœur « xanthine » (plus de 130 molécules synthétisées et évaluées). Ce travail, qui a fait l'objet d'une publication en 2016 dans « journal of Medical Chemistry » (211) et d'un dépôt de brevet (EP 15 307 163.4), nous a permis de déterminer que le cœur « xanthine » était responsable de l'affinité *via* l'établissement de liaisons hydrogènes directe avec l'asparagine N140 et médiée par une molécule d'eau enfouie dans la cavité K_{ac} avec la tyrosine Y97. De plus, la condensation du groupe « triazolo » conduisait à un profil BET-BD1 sélectif *via* l'établissement d'une liaison hydrogène avec le résidu conservé d'acide aspartique D88 ; et que la condensation du groupe « pyrimidinyle » était quant à elle à l'origine de la sélectivité pour BRD4(1) *via* l'établissement de contacts de type van der Waals avec le résidu glutamine Q85. En effet, cette glutamine est remplacée par un résidu plus encombrant pour

BRDT(1) et les domaines BD2 (arginine et lysine respectivement), ce qui peut expliquer en partie la sélectivité observée. Nous avons également proposé un rôle de la dynamique de la boucle ZA pour expliquer ce profil de sélectivité. En effet, la molécule **5A14** induit le déplacement des résidus leucine L94, proline P95 et acide aspartique D96 ce qui conduirait à la sélectivité BRD4(1).

La troisième partie de ce manuscrit concerne **l'optimisation de la molécule 5A14**. Bien que les premières modifications chimiques apportées aux molécules **5A14** et **3J7** (plus de 200 molécules synthétisées et évaluées) n'aient pas permis d'améliorer leur affinité, nous avons néanmoins poursuivi leur optimisation (plus de 70 autres molécules synthétisées et évaluées) dans le but de développer des sondes épigénétiques.

L'une de nos stratégies fut de combiner les informations expérimentales (IC_{50} , K_D , EC_{50}) et structurales obtenus précédemment, à une approche de modélisation moléculaire originale utilisant l'outil S4MPLE. Cet outil de criblage virtuel, a permis entre autre de sélectionner une trentaine de building-blocks (chlorure de sulfonyl) présentant une diversité chimique à greffer sur le cœur *N*-benzylamine-xanthine de l'intermédiaire de synthèse. Parmi, ces building-blocks, se trouvait le groupement oxo-benzodiazépine que nous avons déjà identifié pour la molécule BET-BD1 sélective **3J7** (au moins un log par rapport aux domaines BD2). Par ailleurs, au cours du programme de SAR que nous avons initié afin de comprendre l'origine de la sélectivité de la molécule **3J7**, nous avons constaté que le groupement oxo-benzodiazépine était probablement à l'origine de la sélectivité pour les domaines BD1. C'est pourquoi le premier composé synthétisé fut une molécule hybride entre **5A14** et **3J7**, baptisée **5292**.

Notre travail fut récompensé quand nous avons évalué la molécule **5292** et que nous avons constaté que c'était une **sonde épigénétique pan-BET inhibitrice**. En effet, l'ajout du cycle benzodiazépinone a permis d'améliorer l'affinité d'un facteur de presque dix et trente fois par rapport à respectivement **5A14** et **3J7**. Par ailleurs, cette molécule inhibe l'ensemble des BDs de la famille BET (pan-BET) avec une préférence (au moins un demi-log) pour les domaines BD1 et n'a aucun effet sur d'autre famille de bromodomaine comme le domaine ATAD2A. Au vu de ces premiers résultats, nous avons condensé le groupe triazolo-pyrimidine responsable de la sélectivité de **5A14** avec la molécule **5292**, pour voir si on pouvait déplacer la sélectivité observée vers le domaine BRD4(1) ou du moins vers les domaines BD1 de la famille BET. Bien que cet analogue (**5316**) montre une préférence nette d'un facteur cinquante pour le domaine BD1 de la protéine BRD4 par rapport à son domaine BD2, elle touche également les autres BDs de la famille BET avec cette-fois une préférence d'un demi-log pour BRD4(1).

À ce stade, nous avons synthétisé les autres analogues sélectionnés *via* la méthode S4MPLE afin de voir, si nous pouvions identifier d'autres sondes épigénétiques. Nous avons identifié la molécule **5386**, comme **sonde épigénétique sélective des domaines BD1 des protéines BRD4 et BRD3** avec un facteur supérieur à trente par rapport aux domaines BD2 de ces protéines. La comparaison du mode de liaison de plusieurs analogues de **5292** et plus particulièrement les contacts de van der Waals, a permis de constater que les molécules pan-BET étaient plutôt centrées au niveau du « WPF channel » et établissaient autant de contacts avec les résidus de l'hélice α que le motif « WPF », alors que **5386** établit 90% de ses contacts (16 interactions) avec le « WPF ».

Une autre de nos stratégies d'optimisation de **5A14**, fut de concevoir des molécules **5A14** bivalentes. Ces molécules seraient des homo- ou hétéro-dimères de **5A14**, reliés entre eux par un espaceur PEG_n de longueur n variable. Cette idée provenait de l'observation de la structure cristallographique du complexe BRD4(1)-**5A14** : deux molécules **5A14** se faisant face en tête-bêche. Par ailleurs, deux articles (230, 231) portant sur le développement de molécules JQ1 bivalentes avait permis d'améliorer l'affinité d'un facteur cent. Ces résultats nous ont confortés dans le développement de cette stratégie. Le travail préliminaire que nous avons initié nous a permis d'identifier la molécule **PEG₁-5A14** comme **sonde épigénétique BET-BD1 sélective**. En effet, cet homodimère, **PEG₁-5A14** a montré une affinité (facteur dix) et une meilleure cytotoxicité (facteur trois) par rapport à **5A14**.

Le tableau ci-après (**Tableau 18**) reprend l'ensemble des données d'IC₅₀, de K_D, de ΔT_m et d'EC₅₀ obtenues pour les molécules 5292, 5386 et PEG1-5A14.

Tableau 18 : Tableau récapitulatif pour les trois sondes épigénétiques identifiées.

		5292	5386	PEG₁-5A14
HTRF IC₅₀ (nM)	BRD4(1)	283,4 ± 0,1	379	491,0 ± 1,1
	BRD3(1)	456,6 ± 0,1	837	814,2 ± 1,1
	BRD2(1)	336 ± 0,1	2263	1247 ± 0,9
	BRDT(1)	335,8 ± 0,1	> 50000	> 50000
	BRD4(2)	2779 ± 0,3	> 50000	> 50000
	BRD3(2)	3473 ± 0,2	> 50000	> 50000
	BRD2(2)	1605 ± 0,1	> 50000	> 50000
	ATAD2A	NC	> 50000	NC
ITC K_p (nM)	BRD4(1)	128	641	ND
TSA ΔTm (°C)	BRD4(1)	+ 4,5	+ 5,5	ND
	BRD3(1)	+ 3	+ 4,0	ND
	BD1	0 < X < + 2	0 < X < + 2	ND
	BD2	0 < X < + 2	0 < X < + 2	ND
	Autres BDs	0 < X < + 2	0 < X < + 2	ND
Cellular EC50 (μM)	Jurkat	> 50	10,2	14, 0 ± 0,1
	Molm14	> 50	9,2	15,8 ± 0,1

Perspectives

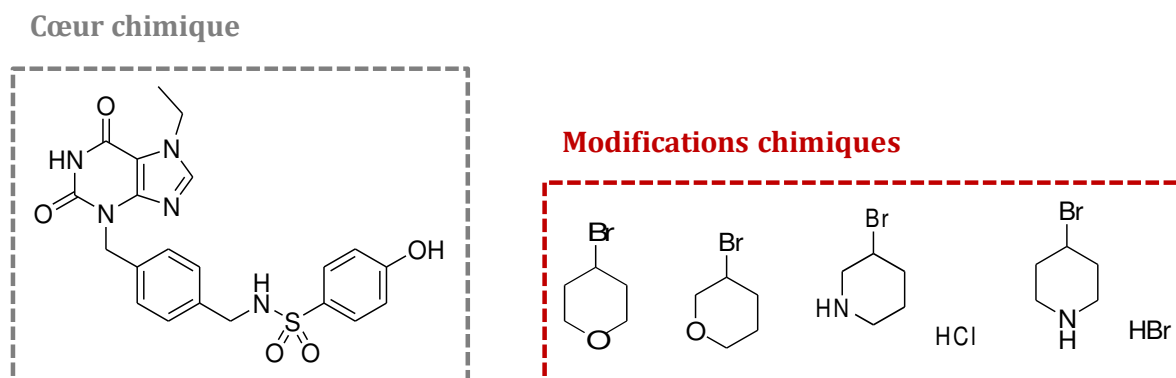
L'ensemble des résultats obtenus au cours de ma thèse, nous a permis d'identifier trois sondes épigénétiques : la molécule 5292 qui est pan-BET inhibitrice, la molécule 5386 qui est spécifique des domaines BRD4(1) et BRD3(1) et la molécule PEG₁-5A14 qui est quant à elle BET-BD1 sélective.

À présent, nos objectifs sont de poursuivre le travail d'optimisation initié pour les molécules bivalentes PEG_n-5A14, d'améliorer la solubilité des molécules 5292 et 5386 et enfin d'initier une étude cellulaire plus approfondie afin de déterminer l'impact biologique de ces molécules et d'élucider le rôle des BD1 et des domaines BRD4(1) et BRD3(1).

Lors de l'évaluation de nos molécules, nous avons pu constater que l'affinité (IC₅₀) autour d'une centaine de nanomolaire de celles-ci, n'était pas corrélée avec l'effet cytotoxique (EC₅₀) autour de la dizaine de micromolaire. Nous suspectons que la solubilité de nos molécules soit en cause, c'est pourquoi nous allons travailler en premier leur **solubilité**, afin de pouvoir envisager par la suite un essai préclinique.

À l'heure actuelle, la modélisation moléculaire réalisé par les chémo-informaticiens, a montré que les positions para ou ortho du dernier phényle pointent vers l'extérieur du site des BD1. Ceci nous indique, que nous pourrions attacher un groupement chimique au niveau de ces deux positions en vue d'améliorer la solubilité. Dans un premier temps, nous avons sélectionné cinq « building blocks » commerciaux présentés dans la **Figure 45**.

Figure 45 : Amélioration de la solubilité de nos molécules.



Légende : Le cœur chimique est encadré en rouge et les cinq building-blocks sont encadrés en rouge.

Lorsque les chimistes auront synthétisé ces molécules, nous pourrons vérifier par HTRF que l'affinité et la sélectivité de celles-ci sont conservées et par cristallographie que le mode de fixation au domaine BRD4(1) est identique aux molécules initiales. Enfin, nous verrons si l'efficacité en modèle cellulaire (cytotoxicité sur des lignées cancéreuses et expression de l'oncogène c-Myc) est améliorée. En fonction de ces résultats nous pourrons envisager d'autres types de modifications tels que les acides carboxyliques et les isostères (232) qui sont des groupements chimiques présentant des critères physicochimiques plus proches de ceux correspondant à ceux d'un « candidat médicament » (page 85). D'autres alternatives sont possibles, telles que l'utilisation du complexe d'inclusion β -cyclodextrine que j'ai déjà testé, ou encore de tensioactifs pour former des micelles, ou encore d'encapsuler nos composés dans des nanoparticules (233, 234).

Il se pourrait également que ces molécules ne soient pas cytotoxiques (on ne touche qu'une seule ou deux protéines de la famille BET et non pas toutes les protéines) et par conséquent non utilisable en monothérapie en cancérologie. Elles auraient par contre un impact important sur d'autres processus biologiques puisque non cytotoxiques et sélectives. Nous envisagerons alors de collaborer avec d'autres centres de recherche pour explorer d'autres pathologies associées aux protéines à bromodomains, telles que les maladies cardiovasculaires, du système nerveux central, ou inflammatoires.

En parallèle, nous pourrons d'ores et déjà initier une étude cellulaire plus approfondie pour nos trois molécules sondes en attendant d'obtenir des molécules plus solubles. Dans ce contexte, Kendall Carrasco, (étudiant en première année de thèse) développera ces aspects pour la partie biochimie structurale, en collaboration avec Eddy Pasquier (chercheur de notre équipe) et Estelle Duprez (CRCM), pour la partie cellulaire, et la plateforme TrGET du CRCM pour la partie concernant les modèles cellulaires et animaux. Nous réaliserons des études par **chémoprotéomique, génomique et transcriptomique** principalement, associées à des études biophysiques *in cellulo* (**CHIP-seq, FRAP et nano-BRET**) et à des études plus classiques de biologie cellulaire (**cytotoxicité, expression c-Myc, apoptose, caspases**). *Ces études sont essentielles pour décider si les efforts fournis dans le développement d'inhibiteurs sélectifs d'un seul BD ou d'une sous-famille doivent être continués ou non dans le futur.*

En **collaboration avec la plateforme de protéomique du CRCM**, nous analyserons le chémo-protéome (molécules couplées à des billes de biotine, synthèses en cours) en présence de nos molécules et en comparaison avec la molécule JQ1 pan-BET inhibitrice.

Les expériences de **FRAP** et de nano- **Bioluminescence Resonance Energy Transfer** (nano-**BRET**) seront réalisées en **collaboration avec les Drs. Stefan Knapp, Susanne Müller-Knapp et Catherine Rogers** (SGC de Frankfurt et Oxford). Ces techniques permettent toutes les deux d'étudier les interactions protéine-protéine *in cellulo* et de mesurer l'inhibition de l'interaction d'un BD avec la chromatine induite par une molécule. Le FRAP (235) est basé sur la redistribution de la fluorescence après photoblanchiment, tandis que nano-BRET (236) est basé sur le transfert d'énergie et l'émission de fluorescence.

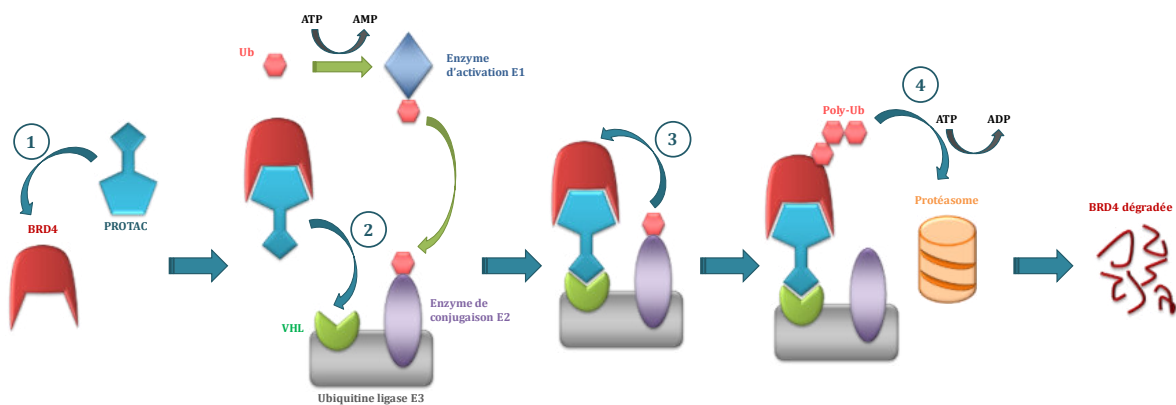
Nous évaluons déjà aussi **dans l'équipe** (Dr. Eddy Pasquier) l'impact biologique de nos molécules sur l'inhibition sélective des bromodomaines au niveau de la chromatine, par analyses **génomiques et transcriptomiques : transcriptome comparatif en présence de l'inhibiteur pan-BET JQ1, expérience par immunoprécipitation de chromatine (ChIP) ou à l'aide de puces « ChIP-seq »** (237). Le profilage épigénétique sera déterminé afin d'évaluer les effets autour ou dans l'environnement de la chromatine. En particulier la cartographie avec les marqueurs chromatinien (H3K4_{me1}, H3K4_{me3}, H3K27_{ac}) permettra de marquer les pertes ou gains des promoteurs actifs lors du traitement avec les inhibiteurs. Les expériences de ChIP-seq se feront dans un premier temps sur des lignées cellulaires établies. Dans un second temps, pour les composés montrant des changements sélectifs au niveau de la chromatine, la validation sera également effectuée sur des échantillons de patients. Les résultats seront intégrés aux données de RNA-Seq et de transcriptomiques obtenus à partir des mêmes échantillons de patients.

Nous envisageons de compléter les études cellulaire que j'ai initiées, afin de déterminer l'efficacité biologique de nos molécules sur des lignées cancéreuses (exemples : évaluation de la cytotoxicité, détermination de l'impact sur l'expression de l'oncogène c-Myc, étude de l'apoptose et du niveau caspases). Actuellement, Camille Montersino (plateforme TrGET) a déjà réalisé l'étude de l'apoptose et le niveau de caspases pour l'une de ces molécules. À terme, nous espérons pouvoir poursuivre l'étude en modèle animal **en collaboration avec la plateforme d'essai préclinique TrGET du CRCM**, afin d'évaluer l'efficacité biologique de nos molécule dans un système plus complexe.

Enfin, nous envisagerons d'utiliser la technologie « Proteolysis Targeted Chimeras » plus communément appelée « **PROTACs** » pour étudier la dégradation des protéines de la famille BET en présence de nos molécules. Cette méthode a été développée il y a une quinzaine d'années (238) et testée avec succès sur multitude de protéines. Par exemple les facteurs de transcription (239-241), les sérines/thréonines (242) ou tyrosines kinases (243), les protéines cytosoliques (244) et les « readers » épigénétiques et notamment les bromodomains (245-247).

Le PROTAC est une molécule hétéro-bifonctionnelle comprenant deux ligands reliés entre eux par un espaceur. Le premier ligand se fixe à la protéine d'intérêt et le second à la protéine ubiquitine ligase E3. De cette manière la protéine cible va pouvoir être poly-ubiquitinée par l'ubiquitine ligase E3, puis dégradée par le protéasome (**Figure 46**).

Figure 46 : Dégradation par le protéasome de BRD4 via l'utilisation du PROTAC.



Légende : Le PROTAC est bifonctionnel. Il comprend un ligand capable de se fixer à la protéine d'intérêt (Étape 1), qui est lié par un espaceur à un second ligand capable de se lier à la protéine ubiquitine ligase E3 via le VHL (Étape 2). Dans un premier temps, l'ubiquitine (Ub) est activée par l'enzyme E1, puis elle est transférée à l'enzyme de conjugaison E2 qui va ubiquitinyler la protéine cible (étape 3) grâce à l'ubiquitine ligase E3. La ligase E3 va permettre la poly-ubiquitination de la protéine cible ce qui va induire sa dégradation par le protéasome (étape 4).

Nous grefferons sur le core xanthine de nos molécules à un phtalamide (246), reconnu par l'E3 ubiquitine ligase qui va ainsi ubiquitinyler la protéine séquestrée par notre inhibiteur et la diriger vers la voie de dégradation. Le niveau d'expression des différentes protéines BET (BRD2, BRD3, BRD4 et BRDT) en présence ou non de nos molécules sera comparé par western-blot. Cette méthodologie nous donnera accès à une lecture de notre sélectivité au niveau protéique.

*Chapitre 4 : Principes &
Protocoles Expérimentaux*

Ce chapitre a pour objectif de présenter les divers techniques utilisées et de détailler les différents protocoles expérimentaux employés au cours de ma thèse afin d'identifier et de caractériser des inhibiteurs de Bromodomains. J'ai décidé de détailler les protocoles expérimentaux pour assurer un suivi. La synthèse des inhibiteurs ayant été réalisée par les chimistes de l'équipe, plus particulièrement par les Drs. Sébastien Combes et Jean-Michel Brunel ainsi qu'une étudiante en thèse Yulia Voitovich, je ne détaillerai pas les protocoles de synthèse et de purification des molécules. Je m'intéresserai donc dans cette partie d'une part à la production et la purification des protéines recombinantes d'intérêts et d'autre part à la mise en place des divers tests permettant l'identification par la méthode d'HTRF, la validation par les techniques biophysiques d'ITC et de TSA, ainsi que par des études cellulaires (viabilité cellulaire et expression de l'oncogène c-Myc) et la caractérisation structurale de complexe protéine-ligand (cristallographie). L'aspect optimisation des « touches » qui utilise des outils de modélisation moléculaire et de docking de complexes protéine-ligand ne sera pas développé dans cette partie, puisque ce travail a été réalisé par les chémo-informaticiens de l'équipe, les Drs. Philippe Roche, Laurent Hoffer et Christophe Muller.

Partie 1 : Biologie Moléculaire :
Clonage & Transformation
bactérienne

I. Plasmides utilisés

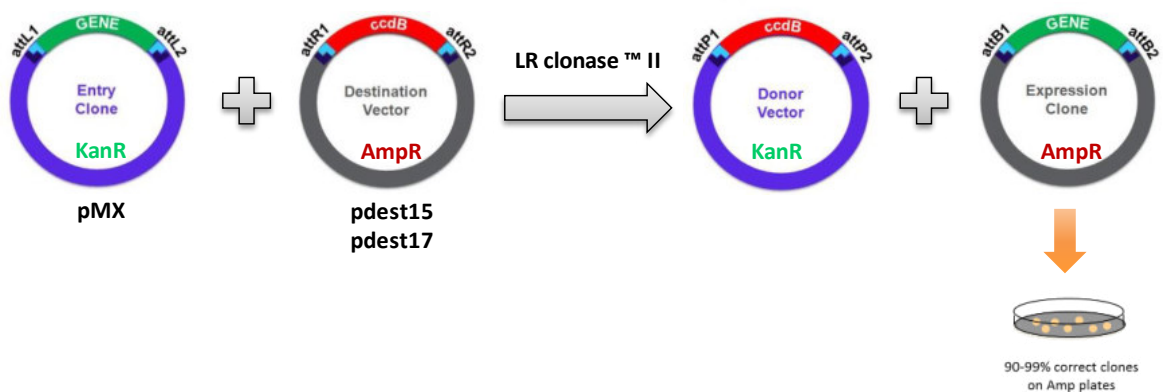
Deux plasmides de type pNIC28BSA4 nous ont été fournis par Stefan Knapp (SGC ; Université d'Oxford), l'un codant la protéine 6-His-TEV-BRD4(1) et l'autre la protéine 6-His-TEV-BRD3(2). Les autres protéines recombinantes (GST-TEV-Bromodomaine, 6-His-TEV-Bromodomaine, GST-6-His-TEV-Bromodomaine) ont été obtenues au sein de l'équipe par moi-même pour la protéine BRD4(1) et par Agnès Amouric pour toutes les autres protéines, par clonage de l'ADN correspondant à la protéine d'intérêt dans les plasmides pDEST15, pDEST17 (clonages Gateway) et pET42b (clonage par Assemblage Gibson). La liste de ces constructions est résumée dans l'**Annexe 1**.

Pour mimer les ADNc de départ, des gènes synthétiques (GeneArt ou Thermo Fisher Scientific) ont été commandés. Un site *attL*, a été ajouté à chaque extrémité du gène, ainsi qu'un site de clivage spécifique de la protéase Tobacco Etch Virus (TEV) en 5' du gène. Ils sont portés par des plasmides de type pMX, avec un gène de résistance à la kanamycine.

II. Clonage Gateway

Les plasmides de destination (pDEST15, pDEST17) contiennent un gène suicide (*ccdB*, pour une sélection négative), borné par les séquences *attR1* et *attR2* qui permettent la double recombinaison homologue (ou LR réaction), avec les séquences *attR1* et *attR2*, flanquant le gène codant la protéine d'intérêt du plasmide transport ou « entry clone » dans le cas du vecteur pMX. Ces derniers possèdent un gène de résistance à la kanamycine, alors que le plasmide de destination « destination vector » possède un gène de résistance à l'ampicilline, permettant de sélectionner uniquement les clones bactériens ayant intégrés le plasmide d'expression codant pour le gène d'intérêt flanqué des deux sites *attB*. Le principe du clonage Gateway est détaillé dans la .

Figure 47 : Principe du clonage de type Gateway.



Les plasmides pDEST15 et pDEST17 permettent d'insérer une étiquette à l'extrémité N-terminale de la protéine recombinante pour favoriser sa purification. Il s'agit de la Glutathionne S-Transférase (GST, ~220 acides aminés) et de la poly-histidine (6-His), respectivement. Ces deux vecteurs ne possèdent pas de site de clivage.

Les LR réactions, ont été réalisées selon les instructions du fournisseur (Invitrogen), en mélangeant 1 µL de plasmide pMX contenant le gène synthétique (100 ng/µL), 6 µL de tampon d'extraction (TE), 2 µL d'enzyme LR clonase II et 1 µL de vecteur de destination (150 ng/µL). Les mélanges ont été incubés à température ambiante pendant 4 heures ou à 25°C pendant 14 heures. Puis 2 µL du mélange ont été introduits dans des cellules chimio-compétentes *E. coli* DH5α (transformation bactérienne).

III. Clonage par Assemblage Gibson

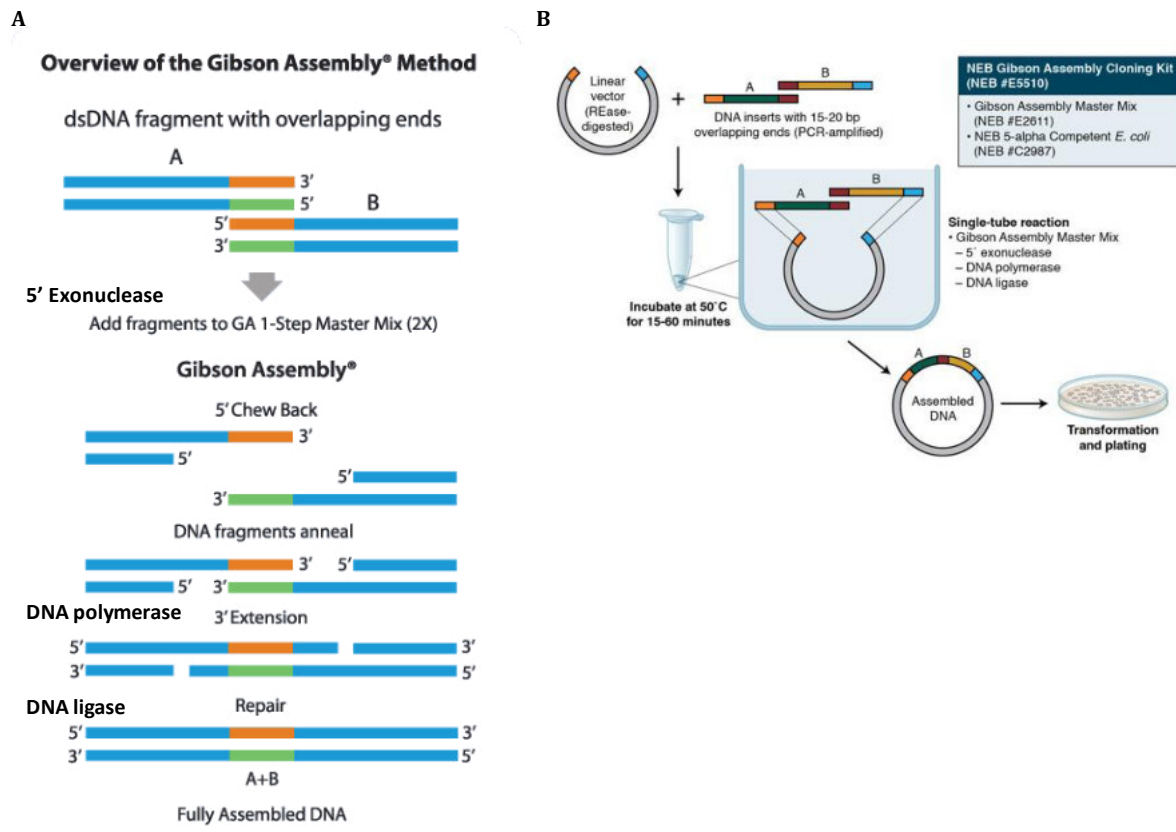
Parmi les techniques innovantes de clonage, la méthode d'Assemblage Gibson (248), permet tout comme la **SLIC** (Sequence and Ligation Independent Cloning), de s'affranchir notamment des enzymes de restrictions et de la séquence dans la construction du plasmide et consiste en l'assemblage de fragments d'ADN chevauchants.

Les fragments nécessaires au clonage (inserts et vecteurs) sont produits par amplification PCR en utilisant l'ADN polymérase Q5 (New England Biolabs) et assemblés au sein d'un mix contenant une exonucléase T5 permettant de créer des extrémités 3' simple brin pour faciliter l'appariement avec un fragment dont la séquence est complémentaire, une ADN polymérase et une ADN ligase pour sceller les ADN assemblés. Le principe du clonage assemblage Gibson est détaillé dans la **Figure 48-A**.

Dans notre cas c'est au moyen du kit NEBuilder ® HiFi DNA assembly Cloning (New England Biolabs) et du plasmide pET42b que les constructions ont été réalisées. Le plasmide pET42b possède un gène de résistance à la kanamycine, ainsi qu'un site multiple de clonage et permet de fusionner différentes étiquettes dont la GST- à l'extrémité N-terminale et la poly-histidine (6-His) à l'extrémité C-terminale de la protéine recombinante. Afin de cliver ces étiquettes, un site de clivage spécifique à la protéase Tobacco Etch Virus (TEV) a été ajouté en 5' du gène, ainsi qu'un site de clivage spécifique à la thrombine en 3'.

Les réactions d'assemblage ont été réalisées selon les instructions du fournisseur (New England Biolabs), en mélangeant 0,5 µL de produit de PCR (gène d'intérêt à 30 ng/µL), 1 µL de vecteur de destination linéarisé par amplification PCR (100 ng/µL) et 10 µL de master mix. Les mélanges ont été incubés à 50°C pendant 40 minutes (**Figure 48-B**). Puis 2 µL du mélange ont été introduits dans des cellules compétentes *E. coli* NEB5α (transformation bactérienne).

Figure 48 : Principe du clonage de type Assemblage Gibson.



IV. Transformation des bactéries

Tous les vecteurs d'expression utilisés dans cette étude possèdent le promoteur du bactériophage T7 qui permet de contrôler l'expression de gènes hétérologues chez *Escherichia coli*. C'est la souche hôte qui possède le gène codant l'ARN T7 polymérase sous le contrôle du promoteur *lacUV5* inducible par l'Isopropyl β -D-1-thiogalactopyranoside (IPTG). Ainsi, après ajout d'IPTG dans le milieu de culture, l'ARN polymérase T7 est produite en grande quantité pour permettre la transcription du gène d'intérêt cloné en aval du promoteur T7.

Les vecteurs ont été utilisés pour transformer des cellules compétentes DH5 α ou NEB5 α afin d'amplifier le gène, pour pouvoir extraire l'ADN en vue de vérifier le séquençage *via* la société GATC Biotech. Les constructions vérifiées par séquençage ont ensuite été introduites dans des cellules compétentes BL21Star, BL21pLys, C41 ou Rosetta en fonction de la protéine d'intérêt à produire.

La transformation des bactéries compétentes est réalisée par choc thermique. Pour cela les bactéries sont mises au contact du plasmide et maintenues 20 minutes dans la glace, puis 45 secondes à 42°C et enfin 3 minutes dans la glace. Les mélanges sont ensuite repris dans du milieu SOC et incubés à 37°C sous agitation 130 rpm, pendant 1 heure. Les cultures bactériennes (150 µL pour les DH5α ou NEB5α et 50 µL pour les autres cellules compétentes) sont ensuite étalées sur boîtes de pétri contenant du milieu LB-Agar additionné de l'antibiotique approprié et incubées à 37°C, sur la nuit. Deux clones par boîte sont piqués et mis en culture dans 5 mL de milieu LB, à 37°C sous agitation à 200 rpm, sur la nuit. Puis l'ADN est extrait et purifié à l'aide du kit « Plasmid DNA purification » selon les instructions du fournisseur (Macherey-Nagel), avant d'être envoyé au séquençage. Un stock glycérolé des clones sélectionnés est conservé à -80°C en milieu LB.

*Partie 2 : Production & Purification
de protéines recombinantes*

I. Production de protéines recombinantes

Les pré-cultures bactériennes consistent à mettre en culture une colonie de chaque souche transformée avec le plasmide contenant le gène codant la protéine d'intérêt (stock glycérolé). Ces pré-cultures sont réalisées en fiole Erlenmeyer de 500 mL contenant 100 mL de milieu Luria Broth (**LB**) à 37°C, sous agitation à 190 rpm et sur la nuit. Un ou plusieurs antibiotiques sont ajoutés au milieu de culture en fonction du ou des gènes de résistance présent(s) sur le plasmide. Ces pré-cultures permettent d'obtenir un stock dense de bactéries en croissance pour inoculer les cultures bactériennes en gros volumes pour les surproductions protéiques.

Les cultures sont réalisées en fiole Erlenmeyer de 2 L à ailettes (pour une meilleure oxygénation de la culture) et à 37°C, sous agitation à 240 rpm pendant 2 à 3 heures. Pour cela, 20 mL de pré-culture sont inoculées dans 1 L de milieu LB, additionné d'un ou plusieurs antibiotique(s) en fonction du ou des gène(s) de résistance présent(s) sur le plasmide. En général 4 L de culture de LB sont réalisés à chaque fois, afin d'avoir une grosse quantité de protéine, par exemple 40 mg de protéine BRD4(1) ont été purifiées pour 4 L de culture. La croissance bactérienne est suivie par mesure de la densité optique (D.O) à 600 nm. Lorsque la D.O atteint environ 0,7 la température est baissée à 20°C. La production de la protéine recombinante est induite par l'ajout d'IPTG (Isopropyl β -D-1-thiogalactopyranoside) à une concentration de 0,1 mM, lorsque la D.O à 600 nm atteint 1-1,2. La culture est ensuite prolongée 24 heures à 20°C.

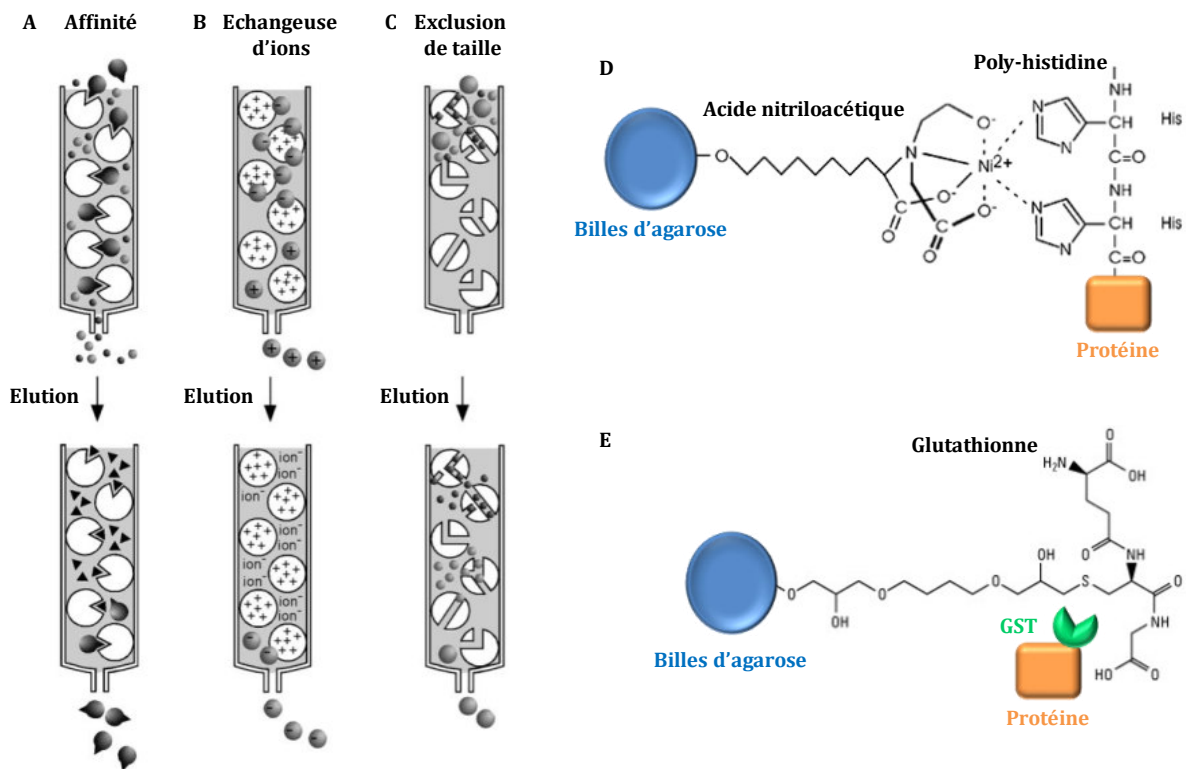
Les bactéries sont ensuite récupérées par centrifugation à 4°C, pendant 20 minutes, à une vitesse de 6000 rpm dans des pots de 1 L. Le surnageant est éliminé et les culots sont collectés dans des falcons de 50 mL. Les parois des pots sont lavés avec du surnageant (milieu de culture) afin de récupérer la totalité des bactéries. Les falcons sont ensuite centrifugés à 4°C, pendant 20 minutes, à une vitesse de 6000 rpm, le surnageant est de nouveau éliminé et les culots bactériens sont pesés et stockés à -80°C.

II. Purification de protéines recombinantes

1. Description des différents types de chromatographies utilisées

Les quatre principales chromatographies sont la chromatographie d'affinité, la chromatographie échangeuse d'ions (anions ou cations), la chromatographie d'interaction hydrophobe et la chromatographie d'exclusion de taille. Dans la plupart des cas, la purification d'une protéine recombinante nécessite plusieurs étapes de purification afin d'obtenir une pureté suffisante pour les applications en aval. Au cours de ma thèse, j'ai purifié des protéines recombinantes à l'aide des chromatographies d'affinité, de dessalage et d'exclusion de taille, c'est pourquoi je vais brièvement vous présenter leurs principes. La **Figure 49** présente différents types de chromatographies.

Figure 49 : Différents types de chromatographies utilisées.



Légende : **A-** La chromatographie d'affinité consiste à fixer des protéines sur une matrice via une étiquette de purification et d'éluer la protéine d'intérêt par l'ajout d'une molécule compétitrice ou par un gradient de pH (pH <6). **B-** La chromatographie échangeuse d'ions consiste à purifier les protéines en fonction de leur charge nette. Les protéines chargées positivement sont purifiées sur des chromatographies échangeuse d'anions (chargées négativement), tandis que les protéines chargées négativement sont purifiées sur des échangeuses de cations (chargées positivement). **C-** La chromatographie d'exclusion de taille permet de séparer les protéines en fonction de leurs rayons hydrodynamiques (poids moléculaire et forme). Les grosses protéines de tailles supérieures au diamètre des pores de la matrice sont éluées en premier et les petites en dernier. **D-** Les protéines fusionnées avec une étiquette poly-histidine sont capables de se fixer de manière covalente aux ions chélaté Ni²⁺ immobilisés sur une matrice inerte composée de billes d'agarose. **E-** Les protéines fusionnées avec une étiquette Glutathionne S-transférase sont capables de se fixer de manière covalente à la glutathionne immobilisée sur une matrice inerte composée de billes d'agarose.

La chromatographie d'affinité :

La chromatographie d'affinité présentée en **Figure 49-A**, repose sur la capacité d'une protéine à se fixer de manière spécifique et réversible à un ligand lié sur une matrice inerte composée d'agarose réticulé ou de polyacrylamide. La protéine peut se fixer au ligand de manière covalente soit directement ou soit *via* l'intermédiaire d'une étiquette de purification placée en position N- ou C-terminale de la protéine. Ce type de chromatographie est très robuste et elle est généralement utilisée dans les premières étapes de purification. Après fixation de l'éluât protéique sur la colonne, celle-ci est lavée afin d'éliminer les contaminants fixés de manière non spécifiques. La protéine d'intérêt est ensuite éluée par l'ajout d'un tampon contenant une molécule compétitrice qui est capable de se lier à la matrice et ainsi de décrocher la protéine de celle-ci. Il existe différents type de chromatographie d'affinité selon le type de protéine à purifier, soit des protéines étiquetées (poly-histidine, GST, MBP, FLAG...), soit des anticorps. Au cours de ma thèse j'ai purifié des protéines possédant des étiquettes de purification de type Glutathionne S-transférase (GST) et poly-histidine (6-His). Les protéines fusionnées avec une étiquette GST sont fixées sur une matrice sur laquelle est immobilisée de la Glutathionne (GSH) et sont éluées par l'ajout d'une solution contenant de 10 à 20 mM de GSH réduit. Les protéines fusionnées avec une étiquette 6-His sont quant à elles fixées sur une matrice sur laquelle sont immobilisés des ions chélatés Ni^{2+} ou Cu^{2+} , (résines nickel et cobalt) et sont éluées par ajout d'imidazole 250 à 500 mM, ou par un gradient de pH (en effet, à pH < 6,0 les résidus histidines sont réduits, ainsi la protéine se décroche de la matrice). Ces deux types de matrice sont détaillés dans la **Figure 49-D & E**.

La chromatographie dite de « dessalage » :

À l'issue de la ou des première(s) étape(s) de purification, il est parfois préférable de se débarrasser des petites molécules telles que la glutathionne, l'imidazole ou le sel, avant de passer à l'étape suivante, pour cela une chromatographie dite de « dessalage » est utilisée. Contrairement aux autres chromatographies, les protéines ne sont pas fixées à une phase stationnaire. Elles sont séparées en fonction de la vitesse à laquelle elles vont passer au travers des pores de la phase stationnaire inerte. La matrice de la colonne de dessalage a une limite d'exclusion (taille des pores) très basse permettant de laisser passer uniquement les petites molécules (sels, sucres, acides aminés...). Les protéines (macromolécules) sont donc exclues de la résine et éluées dans le volume mort de la colonne, tandis que les petites molécules sont éluées plus tardivement. La chromatographie de dessalage peut également être utilisée pour réaliser un échange rapide de tampon. L'échange de tampon sera en effet plus rapide et plus sûr qu'une dialyse.

La chromatographie d'exclusion de taille :

La chromatographie d'exclusion de taille présentée en **Figure 49-C**), permet de séparer les protéines en fonction de leur rayon hydrodynamique, qui est lui-même déterminé à la fois par la taille et la forme des protéines. Tout comme la chromatographie « de dessalage », les protéines sont séparées en fonction de la vitesse à laquelle elles vont traverser la phase stationnaire inerte. Une colonne d'exclusion de taille a également une limite de séparation des protéines, ainsi les grosses protéines ayant une taille supérieure au diamètre des pores de la colonne sont exclues et sont éluées au niveau du volume mort de la colonne, tandis que les moyennes et petites molécules sont éluées plus tardivement en fonction de leur taille. Les protéines de taille moyenne étant éluées avant celle de petite taille. Une proportionnalité existe entre le rayon de giration des protéines et leur volume d'éluion en sortie de colonne. Il existe des colonnes avec des tailles de pores différentes afin de purifier des protéines de petites tailles comme la Superdex 75, pour les protéines de 3 kDa à 70 kDa, ou de grandes tailles comme la Superdex 200, pour les protéines de 10 kDa à 600 kDa. Ce type de chromatographie est utilisé dans les étapes finales de purification du fait qu'elle permet de différencier les différentes espèces protéiques (différentes espèces oligomériques, protéines dépliées, protéines tronquées). Cette étape est notamment indispensable pour les expériences de cristallographie, où la protéine à cristalliser doit être très pure et monodisperse. Il existe différents type de phases stationnaires, celles composées de dérivés du dextran telles que les colonnes GE de type Sephadex ou Sephacryl et celles constituées d'agarose comme les colonnes GE de type Sepharose, Superose et Superdex. Le choix du type de colonne dépendra du poids moléculaire de la protéine et des conditions de tampon dans lequel la purification sera réalisée. Les colonnes de type Superdex restent le plus souvent le meilleur choix, puisqu'elle est compatible avec la plupart des solvants et offre la résolution la plus élevée par rapport aux autres phases stationnaires.

2. Lyse bactérienne

Les protéines à bromodomaines ne sont pas excrétées, il faut donc les extraire de la bactérie par lyse de la membrane plasmique. Ainsi, les culots bactériens sont décongelés et re-suspendues dans du tampon de lyse. Les bactéries sont ensuite lysées dans un bac de glace par sonication (5 cycles de 2 minutes, entrecoupés de 5 minutes de pause, à 80% d'impulsion et à 80% de puissance) sur un sonificateur – homogénéiseur à ultrasons (Ultrasonic Homogenizer, BioLogics). Le lysat bactérien est ensuite centrifugé pendant 1h30 à 4°C à une vitesse de 13000 rpm. L'extrait protéique brut contenu dans le surnageant est récupéré, passé sur filtre 0,8 µM, puis conservé à 4°C jusqu'à l'étape de purification.

Le tampon de lyse diffère selon le type de protéine bromodomaine et de l'étiquette (6-His- ou GST-), cependant il est toujours additionné de : Lysozyme (1 mg/mL), d'anti-protéases (1 pastille pour 50 mL, Complete – Roche), de DNase I (10 µg/mL, Invitrogen), de Glycérol (5%), d'EDTA (2 mM) et de DTT (2 mM). Le lysozyme sert à hydrolyser le peptidoglycane des parois cellulaires bactériennes et d'améliorer l'extraction des protéines (ou ADN/ARN). La DNase permet d'hydrolyser l'ADN génomique. Le glycérol est utilisé pour stabiliser la protéine. L'EDTA permet de chélater les ions et bloquer l'action d'éventuelles protéases sur la protéine d'intérêt, tandis que le DTT permet de réduire les ponts disulfures.

3. Liste des tampons de lyse et purification

La liste des tampons utilisés pour les différentes étapes de purifications figurent dans l'**Annexe 2**. La purification des protéines a été réalisée en batch ou sur des colonnes pré-paquées en utilisant un système **FPLC** (Fast Protein Liquid Chromatography) automatisé (ÄKTA Prime ou Pure, GE Healthcare).

4. Purification des GST-protéines

a. Chromatographie d'affinité

Pour la purification de 1 L culture nous utilisons 2 mL de résine GST (Glutathionne Sepharose™ 4B, GE Healthcare). La résine est déposée dans une colonne en plastique, puis rincée avec 20 mL d'eau milliQ, avant d'être équilibrée avec 1 mL de tampon de lyse. L'extrait protéique brut filtré est incubé avec la résine pendant 2 heures à 4°C sur une roue. La fraction non retenue est éluée et la résine est lavée avec 60 mL de tampon de lavage. La résine est ensuite incubée avec 5 mL de tampon d'éluion (2,5 fois le volume de résine) pendant 15 minutes à 4°C sous agitation (roue), puis la protéine est éluée. La résine est de nouveau incubée avec 5 mL de tampon d'éluion afin de décrocher toute la protéine fixée sur la résine. Plusieurs aliquotes sont récupérées pour analyse sur gel SDS-PAGE (Mini-PROTEAN TGX 4-20%, 10 well, Bio Rad) : fraction non retenue, fraction issue du lavage, protéine éluée et la résine.

La protéine est ensuite concentrée à l'aide d'un concentrateur par centrifugation de type Millipore Amicon Ultra 15® avec un seuil de séparation de 30 kDa et d'une contenance de 15 mL. La membrane de filtration du concentrateur est au préalable rincée avec le tampon de la protéine. La protéine est centrifugée à 4°C, à une vitesse de 4000 g, jusqu'à un volume maximal de 7,5 mL.

b. Chromatographie de dessalage

Afin de séparer la protéine d'intérêt (> 30 kDa) de la glutathionne (307 Da) une colonne de dessalage est utilisée. Cette étape de purification est réalisée sur un système FPLC (ÅKTA Prime ou Pure, GE Healthcare). Nous utilisons une colonne HiPrep 26/10 (53 mL, GE Healthcare) préalablement équilibrée avec 2,5 volumes de colonne de tampon. L'éluât protéique concentré est injecté automatiquement sur la colonne *via* une boucle d'injection de 15 mL. Les fractions correspondant au pic d'absorbance à 280 nm contiennent la protéine en cours de purification. Ces fractions sont rassemblées et une aliquote est gardée pour une analyse sur un gel SDS-PAGE (Mini-PROTEAN TGX 4-20%, 10 well, Bio Rad).

c. Chromatographie d'exclusion de taille

Cette étape de purification est également réalisée sur un système FPLC (ÅKTA Prime ou Pure, GE Healthcare). Nous utilisons une colonne Superdex 75 Hiload 16/600 (124 mL, GE Healthcare) avec des billes de sépharose qui possèdent une limite de séparation de 3 kDa à 70 kDa. La colonne est préalablement équilibrée avec 2,5 volumes de colonne de tampon. L'éluât protéique récupéré en sortie de chromatographie de dessalage est concentré jusqu'à un volume maximal de 5 mL à l'aide d'un Amicon Ultra 15 ® avec un seuil de séparation de 30 kDa et d'une contenance de 15 mL. Cette solution est injectée automatiquement sur la colonne *via* une boucle de 10 mL et est éluée par le tampon d'éluion. Les différentes fractions du pic de la protéine sont déposées sur un gel SDS-PAGE (Mini-PROTEAN TGX 4-20%, 10 well, Bio Rad). Seules les fractions les plus pures sont rassemblées et concentrées jusqu'à la concentration désirée (autour de 30 µM). La protéine est aliquotée (aliquotes de 5 µL et de 50 µL) et congelée dans un bain carboglace – éthanol ou en azote liquide avant d'être stockée à -80°C.

d. Gel SDS-PAGE

La pureté de la protéine est vérifiée à chaque étape de purification sur un gel SDS-PAGE (Mini-PROTEAN TGX 4-20%, 10 well, Bio Rad). La protéine est dénaturée par ajout de bleu de charge (1:1) et chauffage pendant 10 min à 90°C, avant d'être déposée sur gel (10 µL ou 1 mg/mL). Un marqueur de poids moléculaire (PageRuler™ Prestained Protein Ladder, Thermo Fisher Scientific) est également déposé sur gel (10 µL). Le gel migre tout d'abord 10 minutes à 140 V, puis 25 minutes à 200 V. Le gel SDS-PAGE est ensuite coloré et décoloré simultanément en utilisant le système Thermo pierce C2 fast blotter. Pour cela, un feuillet de pads (Mini Gels Pads, Thermo) est imbibé de solution de coloration et un autre d'une solution de décoloration (solutions power stain et destain, Thermo), puis le gel est positionné en sandwich entre ces deux pads (solution de coloration en dessous et de décoloration au-dessus) et les bulles d'air sont

chassées à l'aide d'un rouleau. Le sandwich ainsi formé est ensuite soumis à un courant électrique permettant la migration du colorant et du décolorant simultanément en sens opposé afin de colorer les zones du gel contenant de la protéine et décolorer les zones du gel dépourvues de protéines en une étape unique de 6 minutes pour un gel ou de 11 minutes pour deux gels.

5. Purification des 6-His-protéines

a. Chromatographie d'affinité

Cette étape de purification est réalisée sur un système FPLC (ÄKTA Prime ou Pure, GE Healthcare). Nous utilisons une colonne d'affinité His-Trap FF crude (5 mL, GE Healthcare), avec des billes de sépharose sur lesquelles sont greffées des ions Ni²⁺. La colonne est préalablement équilibrée avec 2,5 volumes de colonne de tampon, avant que l'extrait protéique brut filtré ne soit chargé automatiquement sur celle-ci *via* la pompe de chargement du système FPLC. Les contaminants faiblement fixés sur la colonne sont éliminés à l'aide d'un lavage avec 60 mL d'imidazole à 5 mM. La protéine est ensuite éluée par un gradient de 5 à 500 mM d'imidazole sur 40 mL. Les fractions contenant la protéine d'intérêt, définies après une analyse sur un gel SDS-PAGE (Mini-PROTEAN TGX 4-20%, 10 well, Bio Rad), sont rassemblées.

b. Chromatographie de dessalage

Afin de séparer la protéine d'intérêt (> 10 kDa) de l'imidazole (68 Da) une colonne de dessalage est utilisée. Cette étape de purification est également réalisée sur un système FPLC (ÄKTA Prime ou Pure, GE Healthcare). Nous utilisons une colonne HiPrep 26/10 (53 mL, GE Healthcare). La colonne est préalablement équilibrée avec 2,5 volumes de colonne de tampon. La protéine est concentrée jusqu'à un volume maximal de 7,5 mL sur Amicon Ultra 15® avec un seuil de séparation de 10 kDa, avant d'être injectée automatiquement sur la colonne *via* une boucle d'injection de 15 mL. Les fractions du pic de la protéine sont rassemblées et une aliquote est gardée pour un analyse sur un gel SDS-PAGE (Mini-PROTEAN TGX 4-20%, 10 well, Bio Rad).

c. Clivage de l'étiquette

Afin de cliver l'étiquette de la protéine, nous ajoutons à la protéine, 2 à 4 mg de protéase TEV (Tobacco Etch Virus) selon la quantité de protéine à cliver, 2 mM de DTT et 10% de glycérol. Le tout est incubé sur la nuit à 4°C sur une roue. À noter que la protéine TEV possède une étiquette poly-histidine et qu'elle est produite et purifiée au sein de l'équipe.

d. Chromatographie d'affinité

La TEV ayant tendance à précipiter, l'extrait protéique est préalablement centrifugé 5 minutes à 4°C à une vitesse de 4000 g afin de culotter le précipité. Ensuite, la protéine d'intérêt clivée est séparée du reste (protéine non clivée, étiquette et TEV) à l'aide d'une chromatographie d'affinité. Nous utilisons de la résine Ni²⁺ en batch (His-pure-NiNTA, Thermo) ou pré-paquée sur colonne de 5 mL (His-Trap FF crude, GE Healthcare). Dans le cas de l'utilisation de résine en batch, nous incubons 2 mL de résine pour 60 mg de protéine, pendant 30 minutes à 4°C sur une roue. La fraction non retenue correspondant à la protéine clivée est collectée. Puis l'étiquette, la TEV ainsi qu'éventuellement la protéine qui n'aurait pas été clivée sont éluées avec du tampon d'éluion. Dans le cas de l'utilisation de la colonne His-Trap FF crude, la purification est réalisée sur un système FPLC (ÄKTA Prime ou Pure, GE Healthcare). Pour cela, la colonne est préalablement équilibrée avec 2,5 volumes de colonne de tampon complété de 5 mM d'imidazole pour éviter les interactions non spécifiques avec le support. L'éluât protéique filtré est chargé automatiquement sur la colonne *via* la pompe de chargement, puis la colonne est lavée avec 60 mL de tampon de lavage et enfin une étape à 100% de tampon d'éluion (500 mM d'imidazole) est réalisée sur 40 mL. La protéine clivée sort dans la fraction non retenue, tandis que l'étiquette, la protéase TEV et éventuellement la protéine non clivée sont éluées à 500 mM d'imidazole. Les fractions contenant la protéine d'intérêt (clivée et non clivée), définies après une analyse du gel SDS-PAGE (Mini-PROTEAN TGX 4-20%, 10 well, Bio Rad), sont rassemblées respectivement.

e. Chromatographie d'exclusion

Cette étape de purification est réalisée sur un système FPLC (ÄKTA Prime ou Pure, GE Healthcare). Nous utilisons une colonne Superdex 75 Hiload 16/600, préalablement équilibrée avec 2,5 volumes de colonne de tampon. L'éluât protéique récupéré en sortie de chromatographie d'affinité est concentré jusqu'à un volume maximum de 5 mL à l'aide d'un Amicon Ultra 15® de 10 kDa. Cette solution est injectée automatiquement sur la colonne *via* une boucle d'injection de 10 mL et éluée par le tampon d'éluion. Les différentes fractions du pic de la protéine sont déposées sur un gel SDS-PAGE (Mini-PROTEAN TGX 4-20%, 10 well, Bio Rad). Seules les fractions les plus pures sont rassemblées et concentrées jusqu'à la concentration désirée (environ 25 mg/mL pour la cristallographie ou 500 µM pour l'ITC et le TSA). La protéine est aliquotée (aliquotes de 100 µL) et congelée dans un bain carboglace – éthanol ou en azote liquide avant d'être stockée à -80°C.

III. Quantification de protéines recombinantes

La quantification des protéines peut être déterminée par la mesure de l'absorbance à 280 nm ou par l'utilisation de méthodes colorimétriques telles que le BCA et le Bradford. Ci-dessous sont détaillés les principes et les protocoles utilisés pour chacune ces trois méthodes. Les avantages et inconvénients de chaque méthode sont quant à eux détaillés dans le **Tableau 19**.

1. Absorbance à 280 nm

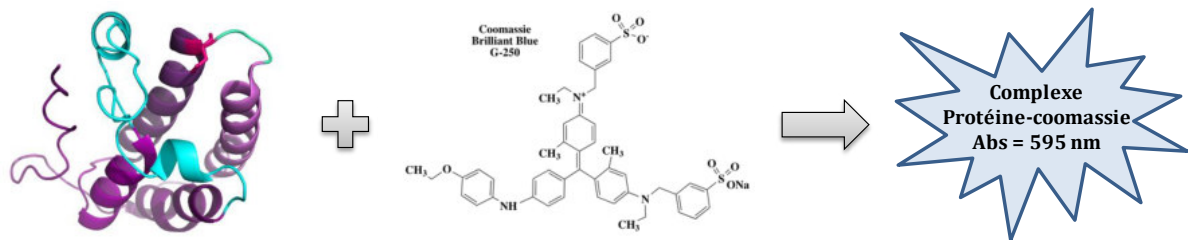
La méthode la plus rapide pour déterminer la concentration d'une protéine est d'utiliser un spectromètre et de mesurer l'absorbance à 280 nm. Cette méthode utilise la loi de Beer-Lambert qui est $A = \epsilon \times b \times c$; avec **A** l'absorbance, ϵ le coefficient d'extinction en L/mol-cm, **b** la longueur du trajet en cm et **c** la concentration en molaire (M). À la longueur d'onde de 280 nm, les acides aminés aromatiques, tyrosine et tryptophane absorbent. La phénylalanine et les ponts disulfures contribuent également à cette absorption, bien que faiblement.

L'utilisation d'un spectromètre microvolumes à fibre optique comme le nanodrop 1000 de Thermo Fisher Scientific permet de mesurer des petits volumes de 1 à 2 μ l d'échantillon sur des trajets optiques très courts ce qui permet de limiter les pertes de matériel. Ce type de spectromètre a également l'avantage de permettre la mesure précise d'échantillons concentrés (plage dynamique de concentration plus large que les spectromètres traditionnels) ce qui permet de s'affranchir des étapes de dilution des échantillons. Pour mesurer la concentration de la protéine, il est nécessaire de soustraire l'absorbance du tampon. Pour cela l'absorbance du tampon de la protéine (2 μ L) est mesurée, puis celle de la protéine (2 μ L) est mesurée.

2. Bradford

La méthode de Bradford est basée sur la formation d'un complexe entre le bleu de Coomassie G-250 et certains acides aminés de la protéine, à savoir les résidus arginine, tryptophane, tyrosine, histidine et phénylalanine. Sous cette forme complexée (forme anionique bleue) la protéine est capable d'absorber à 595 nm. Ainsi, la concentration de la protéine peut être estimée par mesure de l'absorbance à une longueur d'onde proche de 595 nm à l'aide d'un spectromètre et par comparaison avec une protéine standard telle que l'albumine bovine sérique ou BSA. Le principe de cette méthode est détaillé dans la figure ci-dessous (**Figure 50**).

Figure 50 : Principe de la quantification par la méthode de Bradford.

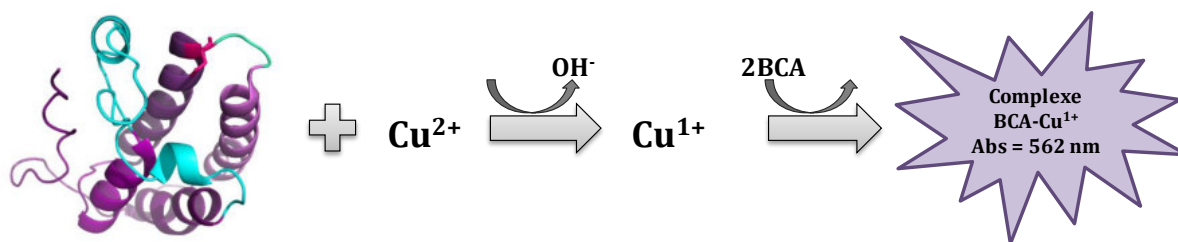


Le Kit Coomassie Plus [™] (Thermo Fisher Scientific) utilisé comprend, 950 mL de Coomassie G-250 dye, methanol phosphoric acid, solubilisé dans de l'eau et 10 * 1 mL de BSA à 2 mg/mL dans une solution 0,9% saline et 0,05% sodium azide. Une gamme avec le standard BSA est réalisée et comprend huit concentrations, allant de 0 à 2000 µg/mL. La protéine d'intérêt est quant à elle diluée quatre fois (1/5^e, 1/10^e, 1/20^e, 1/40^e). Pour chaque concentration, 10 µL de protéine ou de BSA (Kit Coomassie Plus [™] de Thermofisher) sont mélangées à 300 µL de bleu de Coomassie Plus (Kit Coomassie Plus [™]). Un triplicata est réalisé pour chaque point. La plaque est agitée énergétiquement pendant 30 secondes, puis incubée 10 minutes à 20°C, avant d'être lue à 595 nm sur un lecteur Sunrise [™] (TECAN).

3. BCA

La méthode à l'acide bicinchonique ou BCA, est basée sur la réaction de Biuret qui consiste en la conversion du Cu²⁺ en Cu¹⁺ sous des conditions alcalines. Cette réaction est influencée par les acides aminés suivants : cystéines, cystine, tyrosine et tryptophane, ainsi que par la liaison peptidique. Lors de cette conversion, la couleur de l'échantillon protéique passe du bleu au violet, permettant ainsi à la protéine d'absorber à 562 nm. La quantité de Cu²⁺ réduit est proportionnelle à la concentration en protéine et l'absorbance est donc directement proportionnelle à la quantité de protéine. Ainsi, la concentration de la protéine peut être estimée par mesure de l'absorbance à une longueur d'onde proche de 562 nm à l'aide d'un spectromètre et par comparaison avec une protéine standard telle que la BSA. Le principe de cette méthode est détaillé dans la figure ci-dessous (**Figure 51**).

Figure 51 : Principe de la quantification par la méthode BCA.



Le kit Pierce [™] BCA (Thermo Fisher Scientific) utilisé comprend, 1000 mL de réactif A (sodium carbonate, sodium bicarbonate, bicinchoninic acid, sodium tartrate dans 0,1 M de sodium hydroxyde); 25 mL de réactif B (4% cupric sulfate) et 10 * 1 mL de BSA à 2 mg/mL dans une solution 0,9% saline et 0,05% sodium azide. Tout comme le Bradford, le BCA nécessite d'établir la même gamme de concentration avec le standard BSA et de diluer de la même manière la protéine d'intérêt. Cette méthode nécessite de mélanger deux réactifs A et B ratio 50:1 (Kit Pierce [™] BCA de Thermo Fisher Scientific). Pour chaque concentration, 25 µL de protéine ou de BSA sont mélangées à 200 µL du mélange des réactifs A et B. Un triplicata est réalisé pour chaque point. La plaque est agitée énergétiquement pendant 30 secondes, puis incubée 2 heures à 20°C avant d'être lu à 595 nm sur le lecteur Sunrise [™] (TECAN).

Tableau 19 : Comparatif de trois méthodes de quantifications.

Méthodes de quantification			
Absorbance à 280nm			
Range concentration en protéine ex 0,1 à 100 mg/mL de BSA	Avantages Nécessite que 2 µL de protéine Rapide Aucune préparation nécessaire Aucun réactif nécessaire Pas besoin de courbe de standard	Inconvénients Incompatible avec les détergents et les agents dénaturants Erreur possible pour les protéines non caractérisées ou les mélanges de protéines Interférence possible avec les acides nucléiques mélangés avec la protéine	
Bradford			
Range concentration en protéine ex 0,1 à 8 mg/mL de BSA (50:1) range linéaire 0,1- 1mg/mL. ex 15 à 100 µg/mL de BSA (1:1)	Avantages Facile et rapide pour tout type de protéines	Inconvénients Incompatible avec les détergents Nécessite > 100 µL de protéine Nécessite une préparation préalable Utilise du Coomassie Utilise une gamme étalon de BSA Les tensioactifs peuvent conduire à la précipitation des réactifs Les résultats présentent deux fois plus de variation par rapport au BCA	
BCA			
Range concentration en protéine ex 0,2 à 8 mg/mL de BSA (20:1) réactif: protéine ex 0,01 à 20 mg/mL de BSA (1:1)	Avantages Compatible avec les tensioactifs jusque 5%	Inconvénients Incompatible avec les détergents Nécessite > 100 µL de protéine Nécessite une préparation préalable Utilise deux réactifs A et B Utilise une gamme étalon de BSA Nécessite une incubation 2 h à 20°C ou 30 min à 37°C Interférences possible avec les chélateurs, les agents réducteurs et certains tampons	

Partie 3 : Équipements Scientifiques

I. Plateforme DOSynth

Depuis un peu plus d'un an, l'équipe a fait l'acquisition d'un robot de synthèse chimique semi-automatisée « ChemSpeed » qui a été mis en place au sein de la plateforme DOSynth (**Figure 52**) dirigée par un membre de l'équipe. Grâce à ce robot plus de 200 molécules ont déjà été synthétisées pour le projet Bromodomaine.

Figure 52 : Plateforme robotisée DOSynth.



Légende : La plateforme DOSynth est équipée d'un robot de synthèse chimique semi-automatisée « Chemspeed » qui comprend huit réacteurs permettant la synthèse à froid ou à haute température, d'une balance de précision, de plaques pour disposer des vials, de quatre seringues d'injection de volume différent, d'un bloc de filtration et d'une pompe à vide.

II. Plateforme IPC_{DD}

Depuis un peu plus d'un an, l'Institut Paoli Calmette a créé une plateforme nommée **IPC_{DD}** (*IPC drug discovery*), dont les directeurs scientifiques sont le Drs. Yves Collette et Xavier Morelli. Cette plateforme robotique « Access Labcyte » (**Figure 53**) est articulée autour d'un distributeur nanovolumes à énergie acoustique « Echo® 550 » de la société Labcyte et d'un distributeur microvolumes « Multidrop Combi » (Thermo Fisher Scientific). Elle dispose également, d'une centrifugeuse à microplaques « Veolcity 11 » (Agilent Technologies), d'une scelleuse de plaque thermique sous argon « PlateLoc » (Agilent Technologies), d'un lecteur d'absorbance et de fluorescence équipé d'un laser de type « PHERAstar FS » (BMG Labtech). L'enceinte sécurisée, est équipée d'un bras robotique « Access » et de positions de stockage et de gestion des couvercles (Labcyte).

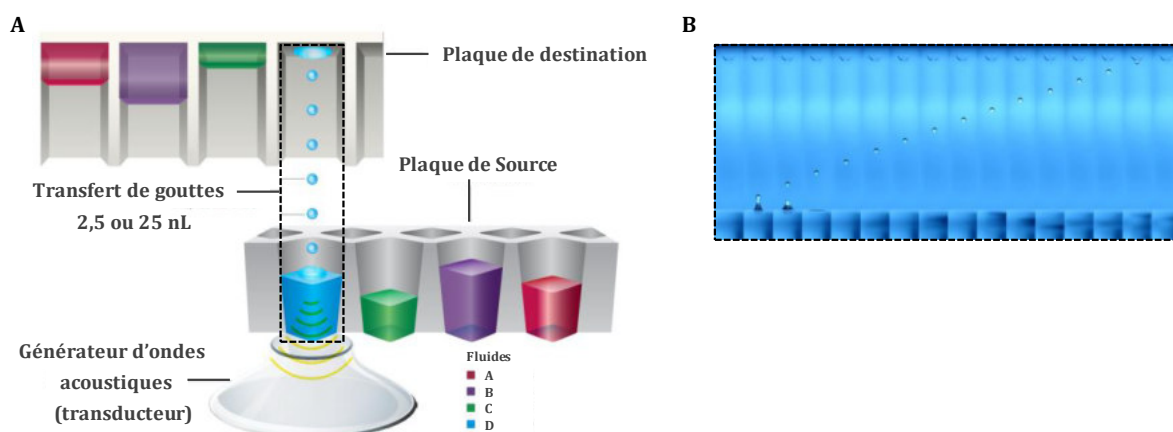
Figure 53 : Plateforme robotisée IPC_{DD}.



***Légende :** La plateforme IPC_{DD} est équipée d'un robot automatisé « Echo® » de chez Labcyte basé sur le principe d'onde acoustique pour dispenser des petits volumes (2,5 nL), d'un « Multidrop Combi » permettant de distribuer des plus gros volumes, d'un bras robotisé, d'une centrifugeuse à microplaque, d'une scelleuse de plaque sous argon et d'un lecteur de fluorescence PHERAstar FS. Cette plateforme peut aussi bien être utilisée pour des expériences d'HTRF, de TSA, de viabilité cellulaire (cytotoxicité) et de cristallogénèse.*

La technologie « Echo® Acoustic Liquid Handling » de ce robot utilise l'énergie sonore ou onde acoustique, comme illustré en **Figure 54**, pour dispenser automatiquement et avec une très grande précision des nanovolumes et sans aucun contact. Grâce à la technologie « Dynamic Fluid Analysis™ de Labcyte », le système est capable en quelques millisecondes de déterminer à la fois la composition du liquide (aqueux, DMSO, glycérol), ainsi que la hauteur du liquide et la puissance nécessaire pour éjecter un volume précis de liquide d'une microplaque source, sous forme de gouttelettes de 2,5 nL dans le puits d'une microplaque de destination inversée. Le transfert de plus gros volume est réalisé en transférant plusieurs centaines de gouttelettes par seconde. Cette technologie est sans contact avec le liquide transféré, de ce fait il n'y pas de risque de contamination croisée ni de risque d'absorption de l'échantillon sur des seringues, cônes ou capillaires, comme dans le cas des robots pipeteurs traditionnels.

Figure 54 : Principe d'onde acoustique du robot Echo® de Labcyte.



Légende : Figure issue de Labcyte. A- Principe de l'onde acoustique. Le transducteur situé en dessous de la plaque source génère une onde acoustique permettant la formation de gouttelettes de 2,5 nL à 25 nL. Ces gouttelettes vont être transférées grâce à cette onde, jusqu'à une plaque destination placée juste au-dessus de la plaque source. Selon le type de fluide (aqueux, glycérol ou DMSO) le transducteur va générer une onde d'intensité plus ou moins importante B- visualisation du déplacement d'une goutte d'un puits de la plaque source vers un puits de la plaque de destination.

Nous avons appliqué cette technologie à la fois pour l'identification d'inhibiteurs *via* la méthode d'HTRF, leur validation *via* les méthodes de TSA et de viabilité cellulaire, ainsi que leur caractérisation structurale *via* la méthode de cristallographie aux rayons X. Pour la réalisation de ces différents tests, trois modules différents de l'Echo® sont utilisés : **audit**, **dose-response** et **reformat**.

Le module "audit" sert à contrôler le volume des inhibiteurs contenu dans les plaques sources 384 puits (384PP, Labcyte) préalablement étiquetées avec un code à barre que le robot est capable de lire. Pour cela, l'onde acoustique de l'Echo® est calibrée pour le DMSO (calibration DMSO2). Ce programme permet aussi, indirectement de vérifier le niveau d'hydratation d'un composé ce qui donne une indication sur sa concentration. En effet au cours du temps les molécules s'hydratent, le volume augmente ce qui entraîne la diminution du pourcentage en DMSO et par conséquent la diminution de la concentration en inhibiteur.

Le module "dose-response" permet quant à lui de programmer les transferts permettant de réaliser des gammes de concentration en molécules « inhibitrices ». Pour cela, l'onde acoustique de l'Echo® est également calibrée pour du DMSO. Par exemple, lorsque nous réalisons pour nos molécules, une gamme de dix concentrations étendue allant de 50 μ M à 13 nM à partir d'une solution stock à 10 mM (100% DMSO), il est nécessaire d'effectuer quatre dilutions intermédiaires dans 100% DMSO. Un certain volume de la molécule à 10 mM est transféré de la plaque source 384 puits (384PP, Labcyte) vers une plaque intermédiaire

384 puits (384LDV, Labcyte), puis le volume est complété avec du DMSO à l'aide du « Multidrop Combi », afin d'obtenir la concentration intermédiaire désirée. La plaque source et la plaque intermédiaire sont ensuite utilisées pour réaliser la gamme de concentrations finales. Un certain volume est transféré de ces deux plaques, vers une plaque destination (Ex : 384 Small Volume™ HiBase Polystyrene Microplates; Greiner bio-one ou 384 RT-PCR ; Bio Rad) et le volume est complétée à 100 nL avec du DMSO à l'aide du robot, afin d'obtenir la concentration finale désirée et une concentration finale de 0,5% en DMSO, *via* l'utilisation d'une plaque Fixe 384 puits (384PP) préalablement remplie de DMSO. À l'issue de ce programme la plaque source est filmée automatiquement sous argon avec un film d'aluminium (Peelable Aluminum Seal, Agilent Technologies).

Le module "reformat" permet de distribuer les réactifs. Pour cela, l'onde acoustique de l'Echo® est calibrée pour du tampon aqueux dans le cas de l'HTRF et du TSA ou pour du glycérol dans le cas de la cristallographie (calibration AQ-BP2 ou Glycérol). Les réactifs sont transférés d'une plaque source 384PP vers une plaque de destination (Ex : 384 Small Volume™ HiBase Polystyrene Microplates; Greiner bio-one ou 384 RT-PCR ; Bio Rad ou CrystalQuick™X ; Greiner bio-one) puis le volume est éventuellement complété avec du tampon à l'aide du « Multidrop Combi ».

Partie 4 : Chimiothèques de composés

I. Chimiothèque de composés commerciaux

L'équipe des Drs. Xavier Morelli et Yves Collettes a développé une base de données 2P2I_{DB} (220-222), répertoriant des inhibiteurs d'interactions protéine-protéine (PPIs) à l'interface (inhibiteurs orthostériques) pour lesquels sont disponibles dans la PDB les structures tridimensionnelles des complexes protéine-ligand et protéine-petite molécule. Les inhibiteurs se liant de manière covalente ne sont pas intégrés à cette base de données. Les complexes au sein de la 2P2I_{DB} sont classifiés en trois catégories : les complexes protéine/peptides présentant un ou deux segment à la surface (classe I), les complexes protéine/protéine présentant au moins six segments à la surface (classe II) et les complexes bromodomaines/lysine acétylées (classe III). Actuellement (juillet 2017) cette base de données comprend 31 complexes protéine-protéine ou protéine-peptide, dont 11 pour les bromodomaines, 558 complexes protéine-ligand, dont 322 pour les bromodomaines et 494 ligands seuls. Celle-ci est actualisée semi automatiquement tous les mois par Marie-Jeanne Basse et les nouveaux inhibiteurs sont validés manuellement avant d'être ajoutés à la base de données. Les molécules présentes dans la 2P2I_{DB} répondent à la « règle des quatre » spécifique des inhibiteurs de PPI (249). Ainsi, les molécules ont une masse moléculaire supérieure à 400 Da, le nombre de cycles, ainsi que le nombre de liaisons hydrogènes accepteurs et que le « AlogP » sont supérieurs à 4. Le coefficient LogP (LogP = Log (Coct/Ceau), permet de prédire le caractère hydrophobe des molécules et plus il est élevé et positif, plus celle-ci sera lipophile.

L'équipe a également développée plusieurs outils afin de créer la chimiothèque 2P2I_{3D} (219), focalisée sur les interactions protéine/protéine, constituée de 1664 composés. Au total 8,3 millions de molécules disponibles commercialement ont été sélectionnées et plusieurs filtres ont été appliqués. Le premier filtre a consisté à ne garder que les modulateurs d'interaction PPI, correspondant 143218 molécules. Le second a permis de sélectionner uniquement les petites molécules, au total 51476 molécules, présentant des structures privilégiées identifiées comme structures de bases pour le développement de médicaments. Un filtre de diversité moléculaire a ensuite été appliqué, permettant de ne garder que 8217 molécules différentes. Le dernier filtre est basé sur l'indice de saturation des liaisons carbone a été utilisé, $Fsp3 \geq 0,4$, celui a permis d'obtenir la 2P2I_{3D} constituée de 1664 molécules. L'ensemble de ces molécules sont stockés à -80°C, en plaques 96 puits (Thermo Scientific™ Nunc™ 96-Well Polystyrene Conical Bottom) à une concentration de 10 mM dans 100% DMSO. La qualité de toutes ces molécules est assurée par les fournisseurs comme étant purs à plus de 90%. L'ensemble des molécules de la 2P2I_{3D} ont été évaluées sur la protéine BRD4(1), afin d'identifier des inhibiteurs sélectifs de bromodomaines, ainsi que sur d'autres complexes protéine-protéine (218).

II. Composés synthétisés au sein de l'équipe

Suite au criblage de la chimiothèque 2P2I_{3D} nous avons pu identifier et valider des inhibiteurs sélectifs de bromodomains. Ainsi nous avons initié avec les chimistes et les chimio-informaticiens de l'équipe des programmes d'optimisation par SAR (relation structure-fonction) pour les meilleures touches identifiées, qui présentaient des profils de sélectivité originaux à savoir les molécules 4N15, 3J7, et 5A14. À ce jour, près de 500 molécules ont été synthétisées, dont 200 à l'aide du robot de synthèse ChemSpeed et purifiées principalement par le Dr. Sébastien Combes et une étudiante en thèse Yulia Voitovich, ainsi que par le Dr. Jean-Michel Brunel. Les détails concernant les voies de synthèses et les purifications de l'ensemble de ces molécules ne sera pas exposés dans cette partie puisqu'ils feront l'objet du mémoire de thèse de Yulia.

*Partie 4 : Identification de « touches »
par la technique d'Homogeneous Time
Resolved Fluorescence (HTRF)*

Il existe de nombreuses méthodes in vitro permettant d'étudier les interactions de type protéine-protéine. Dans le cadre du criblage de petites molécules il est nécessaire d'utiliser des méthodes robustes et sensibles pouvant être miniaturisées et robotisées telles que le l'ELISA, la polarisation de fluorescence, l'AlphaScreen (ou, l'AlphaLISA) ou encore l'HTRF qui utilisent la fluorescence... Parmi ces méthodes, l'HTRF présente de nombreux avantages et nous avons sélectionné cette technologie pour la mettre en place dans l'équipe sur la plateforme de criblage et avons initié une collaboration étroite avec la société Cisbio qui la développe et la distribue.

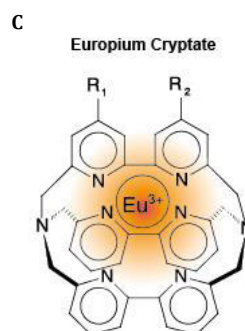
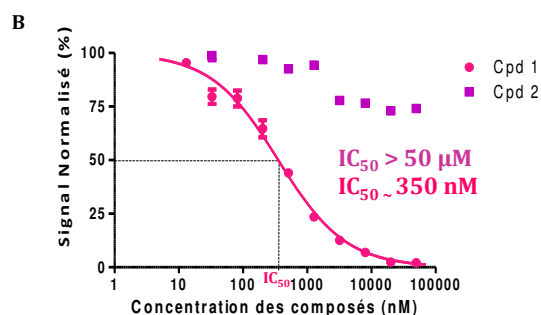
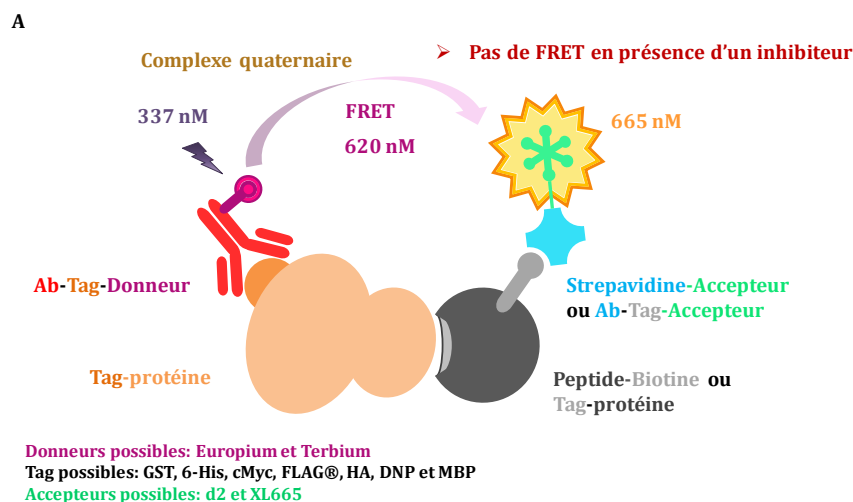
I. Principe de l'Homogeneous Time Resolved Fluorescence

À l'origine l'HTRF était une technique d'immuno-essais, aujourd'hui c'est une technique permettant de réaliser aussi bien des tests biochimiques basiques que des tests cellulaires plus complexes. Cette méthode permet entre autre d'identifier des inhibiteurs d'interaction de types protéine-protéine ou protéine-peptide. Cette technique qui est développée par la société Cisbio, est basée sur le principe du **FRET** (**F**luorescence **R**esonance **E**nergy **T**ransfer) (250). Le phénomène de FRET a lieu entre deux fluorophores, un donneur et un accepteur, ayant des longueurs d'onde d'excitation différentes. Le donneur est excité à sa longueur d'onde d'excitation, il va alors réémettre à une longueur d'onde d'émission correspondant à la longueur d'onde d'excitation de l'accepteur. Si les deux fluorophores sont suffisamment proches (distance inférieure au rayon de Förster) l'énergie émise par le donneur va pouvoir être transférée à l'accepteur et celui-ci va à son tour réémettre à sa propre longueur d'onde d'émission. Le rayon de Förster R_0 qui représente la distance entre le donneur et l'accepteur pour laquelle l'efficacité de FRET est de 50% est généralement compris entre 55 et 65 Å pour les couples donneurs/accepteurs que nous utilisons à savoir Streptavidine/Europium.

Le principe de cette méthode est détaillé dans la **Figure 55-A**. Les donneurs que nous utilisons sont excités à une longueur d'onde de 337 nm et sont capables de réémettre à la longueur de 620 nm. Les accepteurs que nous utilisons sont quant à eux excités à une longueur de 620 nm et sont capables de réémettre à la longueur d'onde de 665 nm. Dans le cadre du projet « Bromodomaine », nous étudions l'interaction protéine-peptide (Bromodomaine-Histone lysine acétylée). Ainsi, lorsqu'une molécule se fixe à la protéine empêchant alors son interaction avec le peptide, la distance entre les deux fluorophores deviennent trop importante ce qui induit une perte de fluorescence La concentration où il y a 50% d'inhibition ou IC_{50} peut alors être déterminée au niveau du point d'inflexion de la courbe d'inhibition, comme indiqué dans la **Figure 55-B**.

La société Cisbio commercialise plusieurs réactifs dirigés contre différents tag protéiques (GST, 6-His, cMyc, FLAG®, HA, DNP et MBP), ainsi que deux types de fluorophores donneurs, l'Europium cryptates (Eu³⁺+cryptate, **Figure 55-C**) et le Terbium cryptates (Lumina®-Tb). Ces deux fluorophores sont formés d'une « cage » de macrocycles à l'intérieur de laquelle se trouve l'ion de terre rare à savoir l'Europium ou le Terbium. Cette architecture permet en effet l'accumulation de l'énergie au sein de la cage, avant d'être transférée à l'ion de terre rare. L'ion va à son tour être capable de libérer l'énergie sur une longue durée contrairement aux fluorophores classiques (1 à 2 ms). Cette propriété, confère à la technique d'HTRF la particularité de pouvoir mesurer la fluorescence en « temps résolue », c'est-à-dire qu'elle n'est pas mesurée immédiatement après l'excitation du fluorophore. Cette spécificité est notamment intéressante pour éliminer la fluorescence après excitation, provenant soit du plastique des plaques, soit de l'un des constituant du tampon utilisé ou encore lorsqu'un composé fluorescent est testé. En effet, une grande majorité de cette fluorescence pouvant interférer avec le signal d'HTRF va être éliminé du fait que sa durée d'émission est très courte par rapport au donneur. La société Cisbio commercialise également deux types de fluorophores accepteurs : la d2 (1 kDa) et la XL665 (105 kDa) qui possèdent les mêmes caractéristiques d'excitation et d'émission.

Figure 55 : Homogeneous Time Resolved Fluorescence.



Légende : A- Principe de l'HTRF : Un Ab-anti-Tag-Donneur reconnaît une Tag-Protéine, tandis qu'un Streptavidine-Accepteur reconnaît un Peptide-Biotine ou un Ab-anti-Tag-Accepteur reconnaît une Tag-protéine. Lorsque la Tag-protéine interagit avec le peptide ou l'autre protéine, les deux fluorophores couplés au donneur et à l'accepteur sont assez proches pour qu'il y ait transfert de fluorescence. Ainsi lorsque le donneur est excité à la longueur d'onde de 337 nm, celui-ci va émettre à la longueur d'onde d'excitation du second fluorophore de 620 nm et ainsi le fluorophore va à son tour réémettre à 665 nm. La société Cisbio commercialise deux types de donneurs (Europium et Terbium), deux types d'accepteurs (d2 et XL665), ainsi que des réactifs dirigés contre différents tag protéiques (GST, 6-His, cMyc, FLAG®, HA, DNP et MBP). B- Lorsqu'une molécule interfère avec l'interaction protéine-protéine ou protéine-peptide, cela empêche le transfert de fluorescence entre donneur et accepteur. Ainsi en présence d'une molécule inhibitrice, la fluorescence diminue (courbe bleue), le point d'inflexion de la courbe correspond à l' IC_{50} (concentration où il y a 50% d'inhibition). C- L'ion de terre rare Europium se trouve à l'intérieur d'une « cage » de macrocycles, au sein de laquelle l'énergie s'accumule avant d'être transférée à l'ion qui va à son tour transférer son énergie sur une longue période.

II. Avantages et inconvénients de l'HTRF

Les avantages : C'est une méthode qui permet de travailler en milieu hétérogène et ne nécessite aucune fixation sur un support solide, ni lavage ou pré-incubation particulière contrairement à l'ELISA. La quantité de protéine nécessaire peut-être réduite grâce à la miniaturisation en plaque 384 ou 1536 puits. Par ailleurs, la société Cisbio fournit une variété de réactifs dirigés contre divers « tags ». Enfin, le principal avantage est tout de même la possibilité de travailler en temps résolu et ainsi de pouvoir éliminer une partie des interférences.

Les inconvénients : Bien qu'elle ne consomme peu de protéine, celles-ci sont onéreuses, de même que les réactifs (donneur, accepteur). Tout comme le TSA qui utilise la fluorescence et malgré la mesure en temps résolu, des interférences sont possibles avec certaines molécules et notamment avec des molécules auto-fluorescentes. L'utilisation d'anticorps (donneur et accepteur) peut également modifier la dynamique d'interaction entre la protéine et le ligand.

III. Réactifs utilisés pour l'HTRF

Les protéines à Bromodomaines que nous avons utilisé pour les tests d'HTRF sont toutes étiquetées GST- en position N-terminale de la protéine. Certaines protéines ont été produites et purifiées au sein de l'équipe, d'autres ont été achetées auprès de la société Cisbio (fournisseur BPS Bioscience). Seul le domaine BRDT(2) n'est pas commercialisé par Cisbio. À ce jour nous avons déjà produit et purifié huit Bromodomaines à savoir BRD4(1), BRD4(2), BRD3(1), BRD3(2), BRD2(1), BRD2(2), ATAD2A et CREBBP et nous essayons de produire et purifier les deux bromodomaines BD1 et BD2 de la protéine BRDT. Pour rappel, les protéines à Bromodomaines sont capables de reconnaître les lysines acétylées des histones. Ainsi, le peptide biotinylé dérivé de l'histone H3 acétylé sur les lysines en positions 5, 8, 12 et 16 (H3K5_{ac}K8_{ac}K12_{ac}K16_{ac}) que nous avons utilisé a été acheté auprès de la société Genic Bio Synthetic. Nous achetons également auprès de Cisbio le donneur « GST-Europium cryptate » et les accepteurs « Strepta-d2 » et « Strepta-XL665 ». Le donneur est couplé à un anticorps qui reconnaît l'étiquette GST- des protéines, tandis que les accepteurs sont couplés à de la streptavidine qui reconnaît le groupement biotine fusionné au peptide. Les molécules inhibitrices sont pour la plupart synthétisées au sein de l'équipe, d'autres sont commerciales et constituent notre chimiothèque 2P2I_{3D}. L'ensemble de nos molécules sont conditionnées à 10 mM dans 100% DMSO dans des plaques 384 puits compatibles avec la technologie d'onde acoustique (Plaque 384PP Polypropylène, Labcyte). Le tampon utilisé pour diluer l'ensemble des réactifs est composé de 50 mM Hepes à pH 7,5, 400 mM de KF et de 0,1 % de BSA.

Les concentrations des différentes protéines, du peptide, ainsi que des réactifs d'HTRF (donneur et accepteur) utilisés sont détaillées dans le **Tableau 20**. Ces conditions ont été optimisées afin d'avoir une valeur du ratio signal/bruit suffisante ($S/B > 6$) afin de s'assurer d'une bonne qualité des données obtenues. À noter que le ratio peptide : accepteur est toujours de 8:1. Enfin pour les inhibiteurs, une gamme de concentrations (8, 10 ou 12 points) est réalisée à partir d'un stock à 10 mM dans 100% DMSO. Celle-ci est généralement comprise entre 50 μ M et 10 nM, toutefois elle est ajustée en fonction de l' IC_{50} observé pour les molécules (par exemple 28 μ M à 1 nM pour les molécules ayant un IC_{50} de l'ordre de la centaine de nM).

Tableau 20 : Concentrations finales pour les tests d'HTRF.

Master Mix					Peptide		Tampon	DMSO finale	
GST-BDs	(nM)	Donneur	(nM)	Accepteur	(nM)	Peptide			(nM)
BRD4(1)	5	MAb Anti GST-Keu	0.5	Streptavidin d2	1,88	H4 KAc 5/8/12/16	15	Tampon A	0,50%
BRD3(1)									
BRD2(1)									
BRDT(1)									
BRD4(2)				Streptavidin XL665	18,8		150		
BRD3(2)									
BRD2(2)									
ATAD2A									

Tampon A : 50 mM Hepes, pH7,5, 400 mM KF, BSA 0,1%

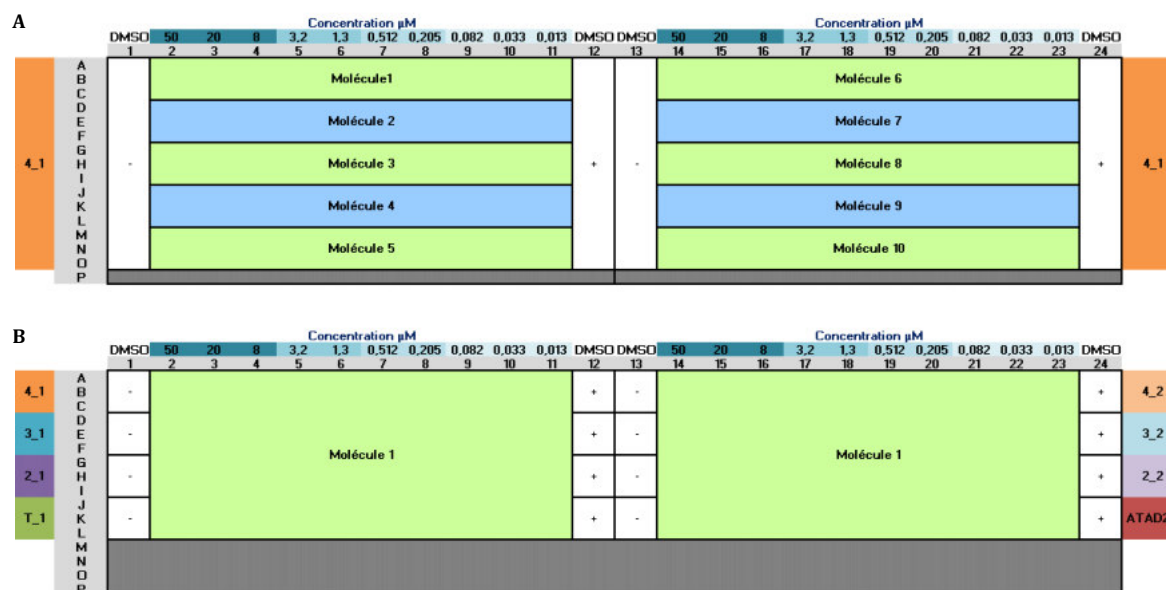
Peptide (Genic Bio Synthetic): SGRGK(Ac)CGK(Ac)GLGK(Ac)GGAK(Ac)RHRKVG

IV. Protocole d'HTRF (manuel)

Les tests d'HTRF sont réalisés en triplicata. Deux exemples de plan de plaques pour l'HTRF classique et en sélectivité sont détaillés dans la **Figure 56**, respectivement. Des contrôles avec du DMSO (contrôles négatifs sans peptide et positifs avec peptide) sont ajoutés, en plus de la gamme de dix concentrations en inhibiteurs. Les contrôles négatifs permettent de mesurer le signal minimal (bruit de fond) proche du signal qu'on obtient lorsqu'on a 100% d'inhibition, tandis que les contrôles positifs permettent de mesurer le signal maximal proche du signal obtenu lorsqu'il n'y a aucune inhibition. Le rapport signal/bruit (S/B) est déterminé en calculant le rapport [(665 nm/620 nm du contrôle positif) / (665 nm/620 nm du contrôle négatif)]. Afin de s'assurer d'avoir des données de bonne qualité, nous avons défini un seuil minimal de 6 pour le ratio S/B.

Le volume final de l'essai est de 20 μ L. L'inhibiteur ou le DMSO dans le cas des contrôles (2 μ L) sont déposés en premier dans des plaques 384 puits de type Small Volume™ HiBase Polystyrene Microplates (Greiner bio-one). La plaque est centrifugée 1 minute à 1500 rpm. Le Master-mix composé de la protéine, du donneur et de l'accepteur est ensuite déposé (14 μ L) dilué dans le tampon d'HTRF, puis la plaque est centrifugée 1 minute à 1500 rpm. Enfin le peptide (dilué dans le tampon d'HTRF) ou du tampon dans le cas du contrôle négatif sont déposés (4 μ L), puis la plaque est centrifugée 2 minutes à 1500 rpm.

Figure 56 : Plans de plaques HTRF manuel.



Légende : A- Plan de plaque utilisé pour l'évaluation de dix molécules en triplicata sur un seul BD par exemple BRD4(BD1). Une gamme de dix concentrations en molécule allant de 50 μM à 13 nM est réalisée (0,5% DMSO finale). Des contrôles DMSO, négatifs (sans peptide en colonne 1 et 13) et positifs (avec peptide en colonne 12 et 24) sont également réalisés. B- Plan de plaque pour l'évaluation de la sélectivité d'une seule molécule en triplicata sur huit BDs, à savoir BRD4(1), BRD3(1), BRD2(1), BRDT(1), BRD4(2), BRD3(2), BRD2(2), ATAD2A. Une gamme de dix concentrations en molécule allant de 50 μM à 13 nM, ainsi que des contrôles DMSO, négatifs (colonnes 1 et 13) et positifs (colonnes 12 et 24) pour chacune des protéines sont également réalisés.

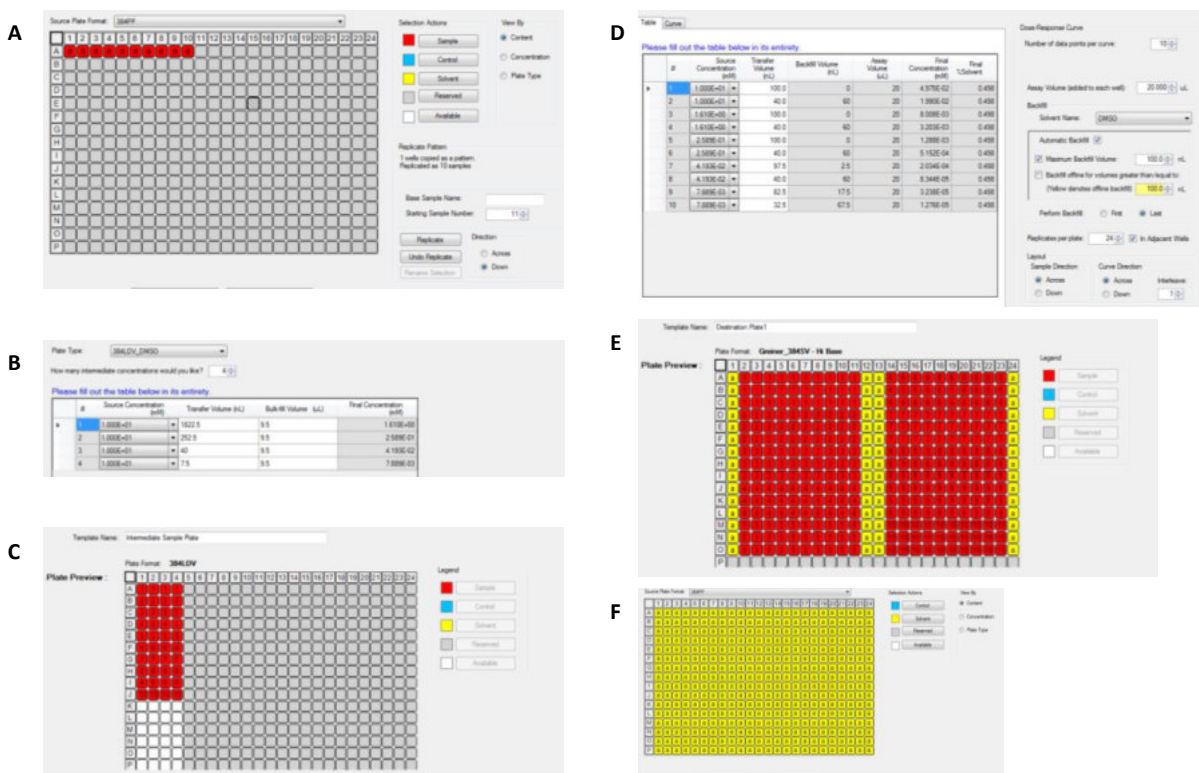
V. Automatisation des essais d'HTRF

Dès la mise en place de la **plateforme IPC_{DD}**, les tests HTRF ont été réalisés en routine à l'aide du robot « Echo® ».

Le module "**dose-response**" de l'Echo® est utilisé pour programmer le pipetage permettant de réaliser une gamme de dix concentrations étendue allant de 50 μM à 13 nM, avec la calibration DMSO2. Cette gamme peut être ajustée en fonction de l'IC₅₀ observée pour ces molécules. Deux exemples d'expériences HTRF sont présentés en **Figure 57** pour un test HTRF classique sur un seul BD et en **Figure 58** pour un test HTRF sur huit BDs (sélectivité). Les molécules sont stockées dans des plaques sources 384 puits (384PP, Labcyte) à 10 mM dans 100% de DMSO. Pour la réalisation de cette gamme de concentration, quatre dilutions intermédiaire sont réalisées : 1,6 mM ; 258,9 μM ; 41,9 μM et 7,8 μM . Un certain volume est transféré à l'aide de l'Echo® des plaques sources vers les plaques intermédiaires 384 puits (384LDV, Labcyte), puis le volume est complété avec du DMSO à l'aide du « Multidrop Combi ». Le robot réalise ensuite la gamme de concentration finale : 50 μM ; 20 μM ; 8 μM ; 1,3 μM ; 512 nM ; 205 nM ; 82 nM ; 33 nM et 13 nM. Un certain volume est transféré à l'aide de l'Echo®

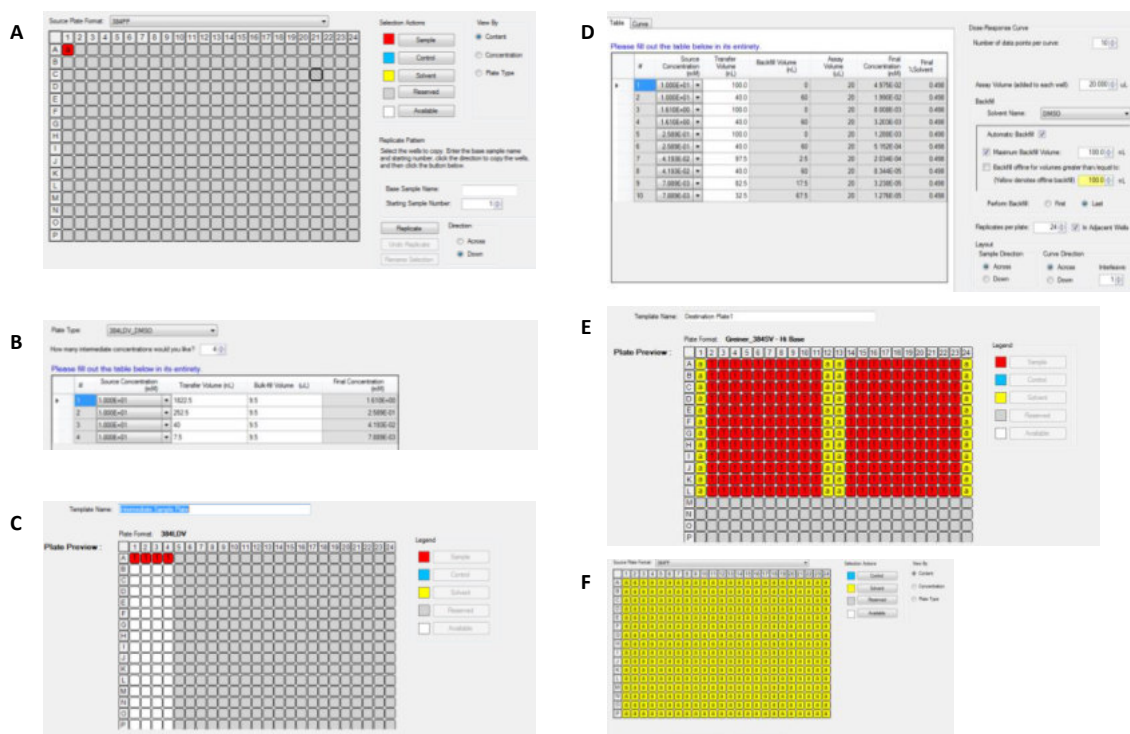
des plaques sources et intermédiaires vers les plaques de destination 384 puits de type Small Volume™ HiBase Polystyrene Microplates (Greiner bio-one), puis le volume est complété jusqu'à un volume final de 100 nL (pour une concentration final en DMSO de 0,5%) à l'aide de l'Echo® avec du DMSO, en utilisant une plaque fixe 384 puits (384PP, Labcyte) remplie de DMSO. Pour les contrôles en présence de DMSO 100 nL sont dispensés à partir de la plaque Fixe. La plaque de destination est ensuite centrifugée 1 minute à 1500 rpm, puis filmées avec un film optique (Greiner bio-one).

Figure 57 : Protocole « dose response » pour l'HTRF classique.



Légende : Impressions écran du protocole de transfert programmé via le module "dose-response" à l'Echo® pour la distribution des inhibiteurs et des contrôles DMSO lors d'un test HTRF classique. **A-** Plan pour la plaque Source 384-PP contenant les dix molécules à 10 mM dans 100% DMSO. **B-** Exemple de dilutions intermédiaires réalisées pour une gamme de concentration allant de 50 µM à 13 nM. Quatre dilutions intermédiaires sont réalisées : 1,6 mM ; 258 µM ; 41,9 µM et 7,9 µM. Les volumes de molécule à 10 mM, ainsi que le volume « bulkfill » à compléter avec du DMSO sont indiqués. **C-** Plan pour la plaque Intermédiaire 384-LDV où sont réalisées les quatre dilutions intermédiaires pour les dix molécules. **D-** Exemple d'IC₅₀ réalisée pour une gamme de concentration allant de 50 µM à 13 nM : 50 µM ; 20 µM ; 8 µM ; 3,2 µM ; 1,3 µM ; 512 nM ; 203 nM ; 83 nM ; 32 nM et 13 nM ; à 0,5% DMSO finale. Les volumes de molécules stocks ainsi que le volume « backfill » à compléter avec du DMSO à l'aide du « Multidrop Combi » sont indiqués. **E-** Plan pour la plaque Destination 384-HiBase, où sont indiqués les contrôles DMSO (en jaune, colonnes 1, 12, 13 et 24) et les dix composés distribués en ligne et en triplicata (en rouge). **F-** Plan pour la plaque Fixed 384-PP. Cette plaque est remplie de DMSO et elle permet d'une part de faire le « backfill » jusqu'à 100 nL pour la réalisation des concentrations finales et d'autre part de remplir les puits avec 100 nL de DMSO pour les contrôles négatifs et positifs (sans molécule).

Figure 58 : Protocole « dose response » pour l’HTRF en sélectivité.



Légende : Impressions écran du protocole de transfert programmé via le module "dose-response" à l’Echo® pour la distribution des inhibiteurs et des contrôles DMSO lors d’un test HTRF en sélectivité. De la même manière que dans la **Figure 57** : **A-** Plan pour la plaque Source 384-PP. **B-** Exemple de dilution intermédiaires réalisées. **C-** Plan pour la plaque Intermédiaire 384-LDV. **D-** Exemple d’IC50 réalisée pour une gamme de concentration allant de 50 μM à 13 nM. **E-** Plan pour la plaque Destination 384-HiBase. **F-** Plan pour la plaque Fixed 384-PP.

Le module "**reformat**" de l’Echo® est ensuite utilisé pour programmer la distribution du Master-Mix et du peptide ou du tampon dans le cas des contrôles négatifs sans peptide, avec la calibration AQ-BP2. Deux exemples d’expériences HTRF sont présentés en **Figure 59** pour un test HTRF classique sur un seul BD et en **Figure 60** pour un test HTRF sur huit BDs (sélectivité). La plaque source 384PP contenant les réactifs (Master-mix, peptide et le tampon d’HTRF) est préalablement centrifugée 2 minutes à 1500 rpm. Les réactifs (200 nL) sont transférés d’une plaque source 384PP vers une plaque de destination 384-HiBase et le « Multidrop Combi » complète le volume avec 19,5 μL de tampon. Des contrôles DMSO, positifs en présence de peptide et négatifs en absence de peptide, sont toujours testés en plus de la gamme de concentration en molécule. Au cours de ces étapes de transferts acoustiques prolongés dans un même puits, il arrive que l’adhésion du liquide aux parois de la plaque modifie la configuration du ménisque lorsque le volume diminue. Le ménisque est alors décalé et n’est plus perpendiculaire aux parois du puits. Ce genre de décalage du ménisque empêche les pipetages ultérieurs car l’Echo® ne parvient plus à mesurer le volume de liquide restant ni à calibrer

l'énergie acoustique à appliquer. Quand ce cas de figure se présente, il est nécessaire de centrifuger la plaque source 2 minutes à 1500 rpm et de reprogrammer les transferts non réalisés à partir des données d'exception générées par le robot. Lorsque les transferts sont terminés, la plaque de destination est centrifugée une dernière fois 2 minutes à 1500 rpm et filmée avec un film optique (Greiner bio-one).

Figure 59 : Protocole « reformat » pour l'HTRF classique.

A

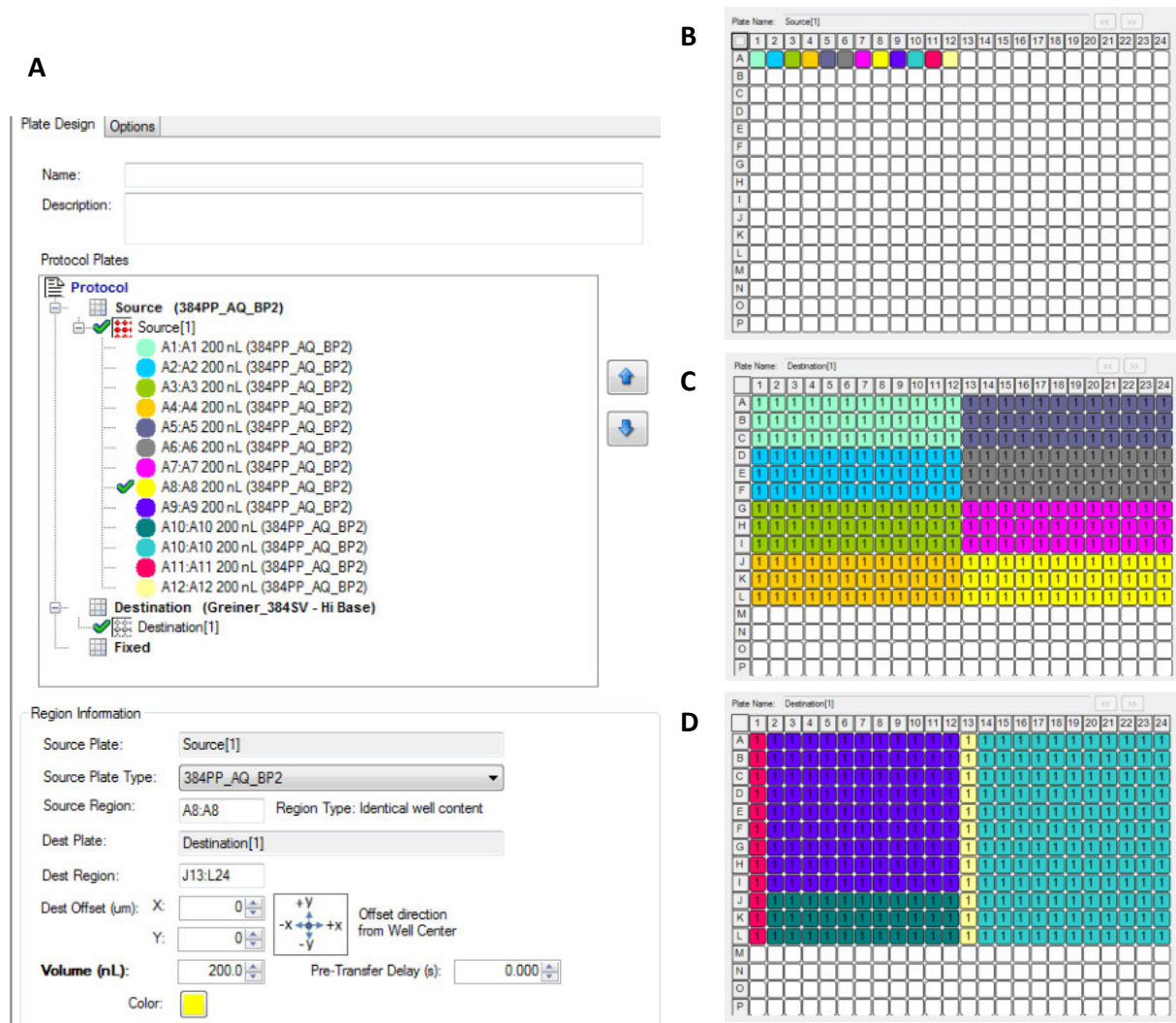
B

C

D

Légende : Impressions écran du protocole de transfert programmé via le module "reformat" à l'Echo® pour la distribution des réactifs lors d'un test HTRF classique (dose réponse de dix molécules à 10 concentrations en triplicata pour un seul BD). **A-** Ici figure la liste des volumes à transférer : 200 nL de Master-Mix (vert), 200 nL de peptide (jaune et bleu) ou de tampon (rose et violet). La calibration AQ-BP2 a été utilisée pour dispenser l'ensemble de ces réactifs. **B-** Plan pour la plaque source : le Master-Mix et le peptide sont déposés dans trois puits adjacents respectivement et le tampon dans le dernier puits. **C-** Premier plan pour la plaque destination : le Master-Mix est dispensé sur les lignes A à O et les colonnes 1 à 24, ce qui correspond à dix molécules testées en ligne en triplicata (Ex : lignes A à C et colonnes 1 à 12). **D-** Second plan pour la plaque destination : le peptide est dispensé sur les colonnes 2 à 12 et 14 à 24 et les lignes A à O. Le peptide est dispensé pour les dix concentrations en inhibiteur ainsi que pour le contrôle positif DMSO ; le tampon (contrôle négatif sans peptide) est dispensé dans les colonnes 1 et 13 sur les lignes A à O.

Figure 60 : Protocole « reformat » pour l'HTRF en sélectivité.



Légende : Impressions écran du protocole de transfert programmé via le module "reformat" à l'Echo® pour la réalisation d'un test de sélectivité d'une molécule pour huit bromodomains. **A-** Ici figure la liste des volumes à transférer pour chaque réactif : 200 nL de Master-Mix (un par protéine), 200 nL de deux concentrations possibles de Peptide (15 nM et 150 nM) ou de tampon. La calibration AQ-BP2 a été utilisée pour dispenser l'ensemble de ces réactifs. **B-** Plan pour la plaque source : les huit BDs sont déposés chacune dans un puits, les deux concentrations utilisées pour le peptide (15 nM et 150 nM) sont également déposés respectivement dans deux puits et enfin le tampon est déposé dans le dernier puits. **C-** Premier plan pour la plaque destination : les Master-Mix pour les huit protéines sont distribuées en ligne (exemple A1 à A12) et en triplicata (exemple A, B, C). **D-** Second plan pour la plaque destination : le peptide à 15 nM est dispensé pour trois des protéines (lignes A à I, colonnes 2 à 12) et celui à 150 nM est dispensé pour les cinq autres protéines (lignes J à L, colonnes 2 à 12 et lignes A à L, colonnes 14 à 24). Le peptide est dispensé pour les dix concentrations en inhibiteur ainsi que pour le contrôle positif DMSO ; le tampon (contrôle négatif sans peptide) est quant à lui dispensé dans les colonnes 1 et 13 sur les lignes A à L.

Les avantages de l'automatisation des tests d'HTRF sont d'une part la reproductibilité des répliquâtes (la barre d'erreur est quasiment imperceptible), d'autre part la rapidité de la distribution dans une plaque (1h30/plaque au robot versus 3h à la main). Par ailleurs, la plateforme permet de programmer automatiquement la lecture des plaques HTRF par le lecteur PHERAstar FS (BMG Labtech).

Le principal inconvénient est que le volume minimum permettant le transfert à l'écho dans les plaques sources 384PP est de 20 µL. Ceci consomme beaucoup plus de réactif lorsqu'on fait une expérience de sélectivité ou une plaque non complète. Dans le cadre de criblages d'un nombre élevé de molécules, le Master-Mix peut être distribué directement en même temps que le tampon à l'aide du « Multidrop Combi » et ainsi éviter ce type de problème.

VI. Analyse des résultats

Les plaques (HTRF classique ou sélectivité) sont incubées à 4°C pendant 3 heures pour les protéines BRD4(1), BRD3(1) et BRD2(1) ou 6 heures pour les protéines BRDT(1), BRD4(2), BRD3(2), BRD2(2) et ATAD2A. Ces conditions ont été optimisées afin d'avoir une valeur du ratio signal/bruit suffisante ($S/B > 6$) afin de s'assurer d'une bonne qualité des données obtenues. La fluorescence est ensuite mesurée à l'aide du PHERAstar FS (BMG Labtech), avec un filtre d'excitation à 337 nm et une mesure de la longueur d'onde à 620 et 665 nm, un retard d'intégration de 60 µs et un temps d'intégration de 400 µs.

Les résultats sont analysés en calculant le rapport du signal (665 nm/620 nm) $\times 10^4$. Les ratios (665/620) mesurés sont normalisés et soumis à un ajustement algébrique à l'aide du logiciel Prism (GraphPad, version 5.03) en utilisant l'équation suivante $Y = 100 / (1 + ((X/IC_{50})^n))$, dans laquelle **X** correspond à la concentration en inhibiteur, **Y** la valeur de fluorescence normalisée et **n** la pente de la courbe à la valeur d' IC_{50} (ou coefficient de Hill, celui-ci est contraint à une valeur supérieure à 0,2 au cours de l'ajustement). Le facteur **Z'** exprime en déviation standard, l'écart par rapport à la valeur moyenne. Il est calculé en utilisant l'équation : $Z' = 1 - [3(SD \text{ maximale}) + (3 (SD \text{ minimale}) / [(moyenne \text{ du signal maximal}) - (moyenne \text{ du signal minimal})]$; le SD correspondant à la déviation standard.

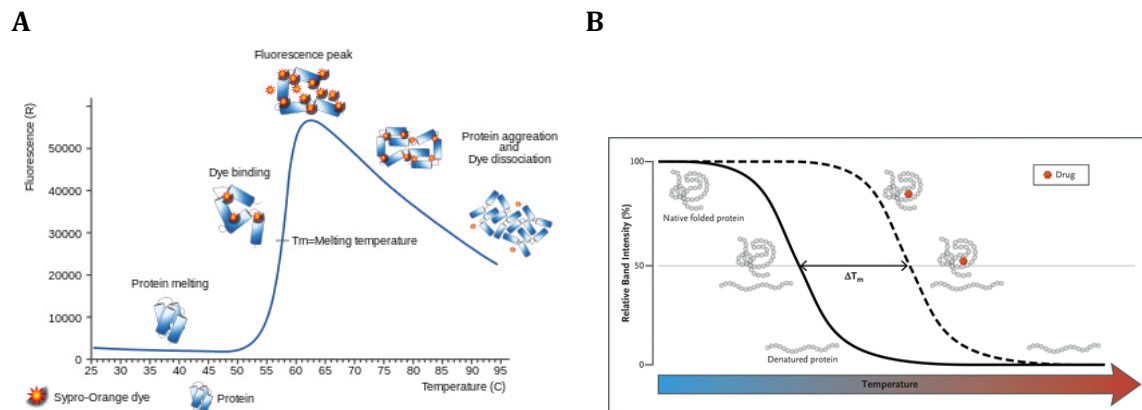
*Partie 6 : Validation orthogonale de
« touches » par la technique de Thermal
Shift Assay (TSA)*

Dans l'équipe nous utilisons deux méthodes afin de valider les « touches » identifiées par la méthode d'HTRF, à savoir le TSA qui permet de déterminer l'effet stabilisant ou déstabilisant d'une molécule sur une protéine et l'ITC qui permet de mesurer les échanges de chaleur produits lors de l'interaction. L'équipe a également accès à d'autres techniques de biophysique permettent de valider des interactions de type protéine - petite molécule, comme par exemples les méthodes d'interférométrie : Surface Plasmon Resonance (SPR) ou Biolayer Interferometry, le Microscale Thermophoresis. Dans le cadre de ma thèse, j'ai utilisé les méthodes de TSA et d'ITC.

I. Principe du Thermal Shift Assay

La technique de TSA, aussi appelée **Differential Scanning Fluorimetry**, est basée sur la mesure de la température de dénaturation d'une protéine au cours d'un gradient de température allant de 20 à 90°C (251, 252). Le TSA repose sur l'utilisation d'un fluorophore, généralement le SYPRO® Orange qui se lie de manière non-spécifique aux régions hydrophobes de la protéine. La longueur d'onde d'excitation est comprise entre 450 et 490 nm et la longueur d'onde d'émission est comprise entre 560 et 580 nm (filtres d'émission et d'excitation FRET). En milieu aqueux, le signal du fluorophore est « quenché », mais dès qu'il se fixe à la protéine, son signal ne l'est plus. En effet, lors de la dénaturation de la protéine, sous l'action de la chaleur, celle-ci se déplie et expose ses régions hydrophobes au fluorophore. Le fluorophore se fixe alors à la protéine, ce qui induit la délocalisation des électrons du fluorophore entraînant ainsi l'augmentation de la fluorescence. Dans le cas du SYPRO® orange, le signal peut être augmenté d'un ratio de plus de 500 fois, ce qui permet de détecter les changements d'état de la protéine (253). Le maximum de fluorescence est atteint lorsque la protéine a exposée toutes ses régions hydrophobes. Dès lors que la protéine est complètement dénaturée, celle-ci s'agrège ce qui induit la dissociation du fluorophore de la protéine et par conséquent la diminution de la fluorescence. La stabilité d'une protéine est déterminée par sa température de fusion (T_m), elle correspond au point d'inflexion de la courbe de dénaturation (**Figure 61-A**). Cette méthode permet donc d'évaluer la stabilité d'une protéine en fonction de son tampon de conditionnement ou encore en présence de ligand. En effet, en présence de tampon ou ligand ayant pour effet de stabiliser la protéine, la dénaturation de celle-ci est retardée et son T_m est augmenté (**Figure 61-B**).

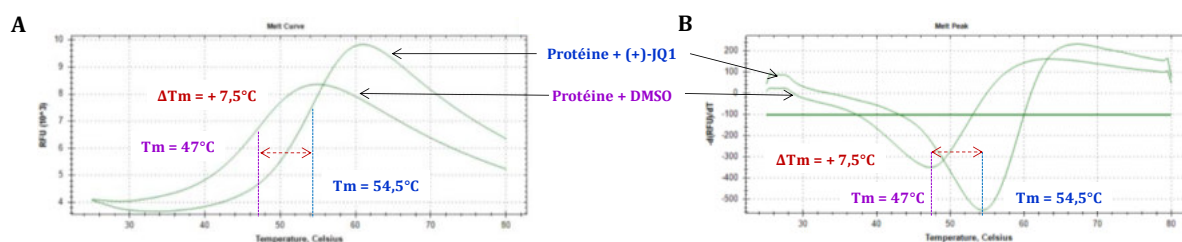
Figure 61 : Principe du Thermal Shift Assay.



Légende : **A-** Lorsque la protéine commence à se dénaturer, elle expose ses régions hydrophobes au fluorophore, c'est ainsi que le fluorophore se fixe à la protéine, ce qui induit l'augmentation progressive de la fluorescence jusqu'à exposition totale des régions hydrophobes de la protéine. Puis la protéine s'agrège et le fluorophore se dissocie de la protéine, ce qui induit une diminution progressive de la fluorescence. Le T_m correspond au point d'inflexion de la courbe de dénaturation. **B-** Lorsqu'on incube une protéine en présence d'un inhibiteur, celui-ci peut avoir soit un effet stabilisant ou alors déstabilisant. Dans cette figure, la molécule qui se fixe à la protéine a un effet stabilisateur qui protège la protéine de la dénaturation. Ainsi la dénaturation de la protéine est retardée et son T_m est augmenté. On peut alors calculer le ΔT_m entre le T_m de la protéine seule (en présence de DMSO) et celui de la protéine en présence de l'inhibiteur (dans du DMSO). Les figures sont issues des sites suivant : https://de.wikipedia.org/wiki/Thermal_Shift_Assay <http://www.nejm.org/doi/full/10.1056/NEJMcibr1308868?rss=searchAndBrowse&#>.

Prenons l'exemple de l'inhibiteur très connu de bromodomains, (+)-JQ1 qui inhibent l'ensemble des membres de la famille BET (**Figure 62**). Cette molécule est capable de se fixer aux protéines de la famille BET et d'établir des contacts avec celles-ci, de type liaisons hydrogènes, ce qui aura pour effet de stabiliser la protéine et de retarder la dénaturation de la protéine. La protéine BRD4(1) est mise en présence de DMSO (contrôle) et de la molécule JQ1. Les T_m de la protéine BRD4(1) mesurées sont de 47°C et de 54,5°C, respectivement en présence de DMSO et de (+)-JQ1. Le ΔT_m calculé entre les deux est positif (stabilisation de la protéine) et il est de + 7,5°C. Inversement, lorsque le tampon ou le ligand déstabilisent la protéine, celle-ci est plus rapidement dénaturée, son T_m est diminué et on observe un ΔT_m négatif. Contrairement à l'ITC, le TSA ne permet pas de mesurer un K_D , cependant une corrélation entre le T_m et le K_D est observée pour différents analogues d'une molécule possédant un « cœur chimique » commun lorsqu'on compare les résultats obtenus avec d'autres techniques (254).

Figure 62 : Exemple de résultats obtenus.



Légende : **A-** Courbe de fluorescence (RFU) en fonction de la température pour la protéine en présence de DMSO ou de la molécule (+)-JQ1. La température de dénaturation (T_m) observée est de 47°C en présence de DMSO (violet) et de 54,5°C avec (+)-JQ1 (bleu). Le ΔT_m calculé entre les deux (rouge) est de + 7,5°C. **B-** Courbe dérivée $-d(RFU)/dT$. Seuls les pics négatifs en dessous de la barre sont regardés. Les T_m sont indiqués en violet en présence de DMSO et en bleu pour (+)-JQ1 et le ΔT_m entre les deux est indiqué en rouge.

II. Avantages et inconvénients du TSA

Les avantages : Les longueurs d'onde d'excitation et d'émission du SYPRO® Orange sont compatibles avec les systèmes qPCR retrouvés couramment dans les laboratoires de recherche, c'est donc une technique très facile à mettre en place. Cette méthode ne consomme pas beaucoup de protéine (bien que plus que l'HTRF) et elle est peu onéreuse. Par ailleurs, une corrélation est possible avec le K_D mesuré par une autre technique comme par exemple l'ITC (252, 255).

Les inconvénients : Tout comme l'HTRF qui utilise la fluorescence, des interférences sont possibles avec des molécules auto-fluorescentes. Certaines protéines (20 à 30% des protéines) sont incompatibles avec cette méthode car elles ne produisent pas de signal. C'est une méthode qualitative qui ne donne pas d'information quant à l'interaction protéine-ligand, contrairement aux méthodes quantitatives telles que l'ITC ou l'HTRF.

III. Protocole de TSA (manuel)

La protéine est testée en présence de DMSO (0,5% final) afin de déterminer son T_m , ainsi qu'en présence de composé (0,5% DMSO final) en vue de déterminer le ΔT_m (positif ou négatif) induit par la présence de la molécule. Le composé est également testé seul (la protéine étant remplacée par du tampon) pour vérifier l'absence d'auto-fluorescence et d'interférence quelconque avec le signal. Chaque point est réalisé en triplicata.

Tous les réactifs sont distribués dans des plaques 384 puits de RT-PCR (Hard-Shell® PCR plates, 384-well, thin-wall ; Bio-Rad) et dans la glace, dans l'ordre qui suit : le fluorophore (Protein Thermal Shift™ Dye, ThermoFisher Scientific, excitation 580 +/- 10 nM et émission 623 +/- 14 nM) en premier au 1/1000^e (2,5 μ L), puis la protéine à une concentration de 2 à 5 μ M ou

le tampon (5 μ L) et enfin la molécule à une concentration comprise entre 10 et 50 μ M ou le DMSO (12,5 μ L); pour un volume final de 20 μ L. Les concentrations en protéine et ligand sont optimisées en fonction de l'intensité du signal observé. Le tampon utilisé est composé de 50 mM d'Hepes à pH 7,5 et 200 mM de NaCl.

La plaque est centrifugée entre chaque ajout de réactif à 4°C pendant 1 minute à une vitesse de 1000 rpm. Elle est ensuite scellée par un film optique (Microseal® 'B' seal, Bio-Rad) et placée sur un appareil de type CFX384 RTQPCR (Bio-Rad), pour être est tout d'abord chauffée à 25°C pendant 1 minute, puis la température est augmentée de 25°C à 70°C par palier de 0,5°C toutes les 25 secondes. Les signaux de fluorescence sont suivis en utilisant le filtre FRET.

IV. Protocole de TSA au robot

Un exemple de protocole utilisé au robot est détaillé dans la **Figure 63**. Dans cet exemple, quinze molécules, ainsi qu'un contrôle DMSO ont été testés en triplicata sur huit protéines différentes. Lorsque les expériences de TSA sont réalisées au robot « Echo® », 100 nL de molécules (concentration finale comprise entre 10 et 50 μ M ; 0,5% DMSO finale), 300 nL de protéine à 133 μ M (concentration finale 2 μ M), 200 nL de SYPRO® 100X (concentration finale 1X) sont dispensés à l'aide de l'Echo® et le volume est complété à 20 μ L final avec du tampon à l'aide du « Multidrop Combi ». La plaque est centrifugée à 4°C pendant 1 minute à une vitesse de 1000 rpm, avant d'être scellée par un film optique (Microseal® 'B' seal, Bio-Rad) et placée sur un appareil de type CFX384 RTQPCR (Bio-Rad), pour être est tout d'abord chauffée à 25°C pendant 1 minute, puis la température est augmentée de 25°C à 70°C par palier de 0,5°C toutes les 25 secondes. Les signaux de fluorescence sont suivis en utilisant le filtre FRET.

Figure 63 : Protocole « reformat » pour le TSA.

A

B

C

D

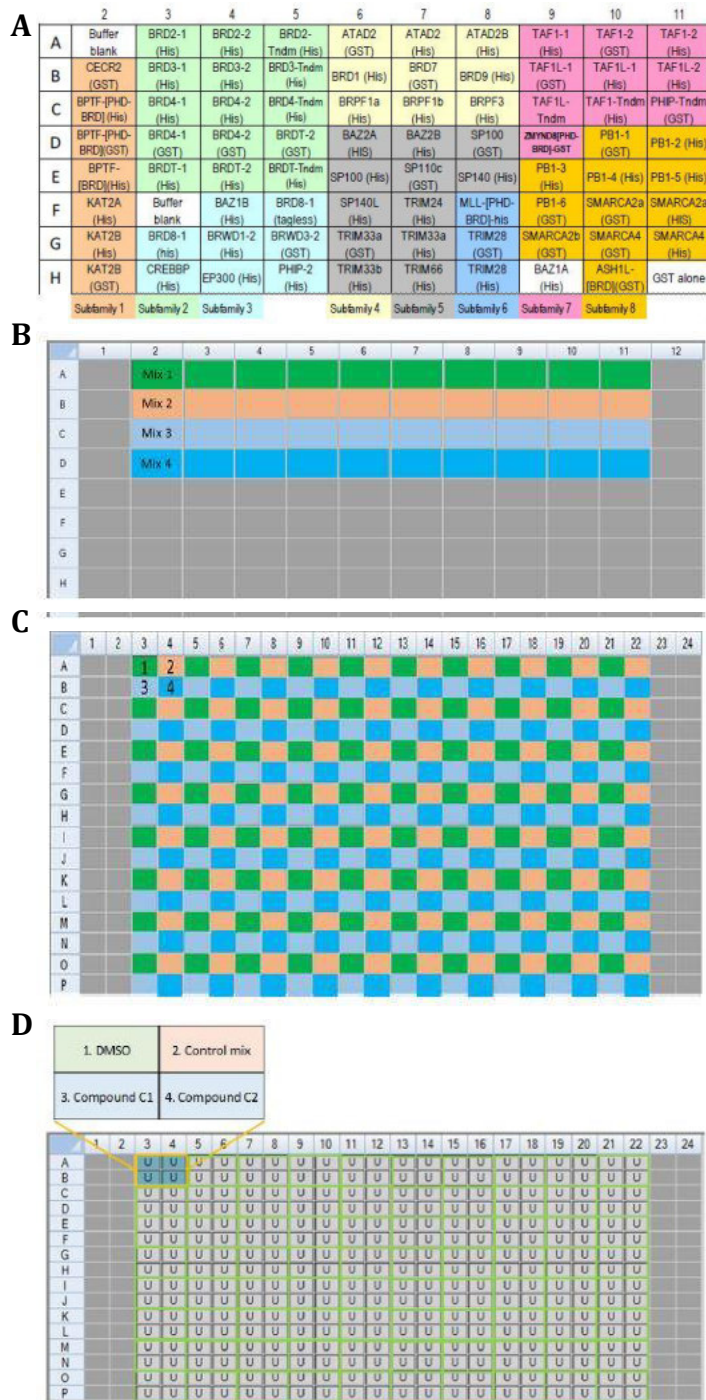
Légende : Comme pour les expériences d'HTRF, les inhibiteurs ou le DMSO (100 nL ; 0,5% DMSO finale) sont dispensés à l'aide de l'Echo® en ligne (24 puits, correspondant à 8 protéines en triplicata). Impressions écran du protocole de transfert des réactifs programmé via le module "reformat" à l'Echo® pour la réalisation d'un test de TSA sur huit protéines bromodomaines. **A-** Cet écran liste les pipetages et volumes programmés pour chaque réactif : 300 nL de protéine, puis 200 nL de SYPRO® Orange. La calibration AQ-BP2 a été utilisée pour dispenser l'ensemble de ces réactifs. **B-** Plan pour la plaque source : les huit protéines sont déposés dans un puits chacune et le SYPRO® est déposé dans deux puits. **C-** Premier plan pour la plaque destination : les protéines sont dispensées en colonne et en triplicata (exemple A1, A2 et A3), et sur 16 lignes (pour les quinze composés et le contrôle DMSO). **D-** Second plan pour la plaque destination : le SYPRO® est dispensé sur l'ensemble de la plaque (384 puits).

V. Protocole de TSA (Bromoscan)

Nous utilisons le kit commercial « BromoMelt™ (Reaction Biology) pour la réalisation du TSA sur la quasi-totalité des bromodomains (excepté 6 BDs non présent dans le kit : BRD8(2), BRDW3(2), PHIP(2), BRDW1(2), PB1, SP110A. Ce kit comprend une plaque source 96 puits où sont conditionnées les protéines concentrées à 10X (**Figure 64-A**), ainsi que des tubes contenant du 5X reaction buffer, 5000X SYPRO® Orange, DMSO 100% et un mix contrôle composé de cinq inhibiteurs (JQ1/CBP112/IBET151/GKS2801/PFI3) à 400X dans 100% de DMSO. Ce kit étant très onéreux, seules les meilleures molécules validées par ITC et TSA sur l'une des protéines d'intérêts, sont testées sur l'ensemble des BDs disponibles dans ce kit.

Tout le protocole qui suit est réalisé dans la glace. Dans un premier temps, 3 mL de « 1X protein dilution buffer » est préparé par dilution du stock 2,5X dans de l'H₂O milliQ. Les protéines 10X sont ensuite diluées à 5X par ajout de 10 µL de « 1X protein dilution buffer » dans les colonnes 2 à 11 de la plaque source et le mélange est homogénéisé à la pipette. La plaque est centrifugée à 4°C pendant 1 minute à une vitesse de 1000 rpm. Ensuite, 3 µL de chaque protéine sont déposés dans quatre puits adjacents (**Figure 64-C**), dans une plaque 384 puits de RT-PCR, puis celle-ci est centrifugée à 4°C pendant 1 minute à une vitesse de 1000 rpm. Quatre tubes contenant 240 µL de 5X reaction buffer, 955 µL d'H₂O milliQ, 1,5 µL de 5000X SYPRO® Orange et 3,75 µL de (1) le DMSO, (2) le mix contrôle, (3) le composé à tester à une concentration initiale de 400X pour une concentration finale de 10 ou 25 µM, (4) le composé est testé à la même concentration qu'en (3) ou à une autre concentration. Puis 118 µL de chaque tube sont déposés dans 10 puits (en ligne) sur une plaque 96 puits fournie par reaction biology. (**Figure 64-B**). Enfin 12 µL de chaque puits sont déposés dans la plaque de destination (**Figure 64-D**). Le mélange est homogénéisé à la pipette. La plaque de RT-PCR est une dernière fois centrifugée à 4°C pendant 1 minute à une vitesse de 1000 rpm, avant d'être scellée par un film optique (Bio Rad). La plaque est placée sur un appareil de type CFX384 RTQPCR (Bio Rad), elle est tout d'abord chauffée à 25°C pendant 10 secondes, puis la température est augmentée de 25°C à 90°C par palier de 0,5°C toutes les 30 secondes. Les signaux de fluorescence sont suivis en utilisant le filtre FRET.

Figure 64 : Protocole TSA Bromoscan.



Légende : **A-** Plaque 96 puits contenant les protéines 10X ; **B-** Plan de plaque (96 puits) où sont dispensés les quatre mix préparés ; **C-** Premier plan de plaque RT-PCR (384puits) où sont répartis les protéines ainsi que les quatre mix préparés ; **D-** Second plan de plaque RT-PCR (384puits) où sont répartis les quatre mix préparés.

*Partie 7 : Validation orthogonal de
« touches » par la technique d' Isothermal
Titration Calorimetry (ITC)*

I. Principe de l'Isothermal Titration Calorimetry

La technique d'ITC (256), est basée sur la mesure directe de la chaleur libérée ou absorbée au cours de l'interaction entre deux biomolécules. Cette méthode permet de mesurer en solution, les paramètres thermodynamiques de l'interaction protéine-ligand, à savoir la constante de liaison (K_D de 10^2 à 10^8 M^{-1}), la stœchiométrie (n), l'enthalpie (ΔH). L'enthalpie libre (ΔG) et l'entropie (ΔS) de l'interaction sont calculées à l'aide des équations suivantes : $\Delta G = \Delta H - T \Delta S = -RT \ln K_A = -RT \ln (1/K_D)$, où T correspond à la température en Kelvin et R à la constante des gaz parfait = $8,314 \text{ J mol}^{-1} \text{ K}^{-1}$.

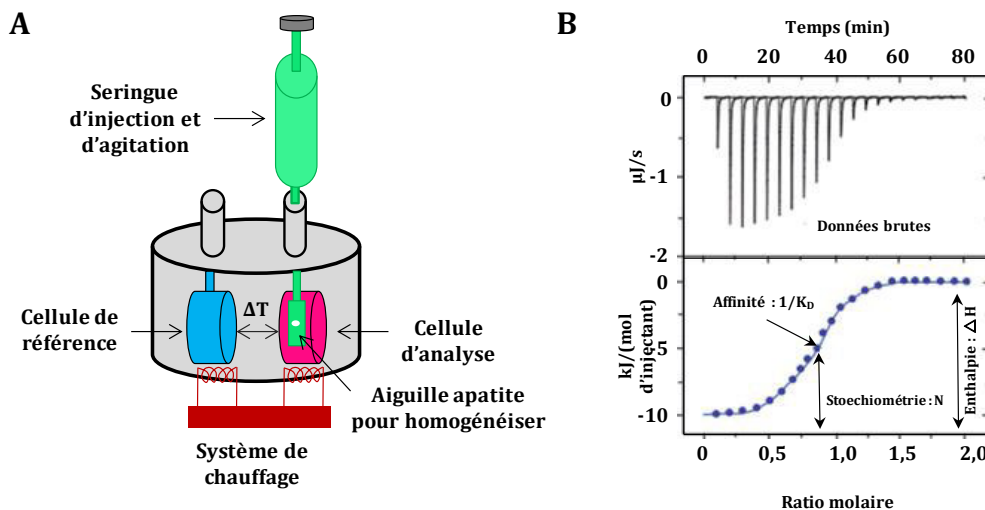
Cette méthode est plus particulièrement utilisée pour la recherche et le développement de médicaments (257, 258) car elle permet de quantifier l'affinité de liaison, mesurer les paramètres thermodynamiques, déterminer la spécificité et la stœchiométrie d'une liaison, d'identifier des molécules « hit » à optimiser, de valider des valeurs d' IC_{50} et EC_{50} obtenues par d'autres techniques au cours de l'optimisation de molécules « hit to lead », de mesurer des paramètres cinétiques d'une enzyme.

Le microcalorimètre est composé de deux cellules, la cellule de référence qui contient de l'eau et la cellule d'analyse qui contient l'échantillon (protéine et ligand), d'une seringue d'injection et d'un système de chauffage. (Figure 65-A) Au cours de la mesure, le microcalorimètre doit maintenir une température rigoureusement identique dans ces deux cellules. Ce système est capable de détecter des différences infimes de température entre ces deux cellules au cours de la formation de liaisons entre la protéine et le ligand (protéine, petites molécule, acides nucléiques, lipides...) et de compenser cette différence de température *via* son système de chauffage afin de ramener les deux cellules à la même température.

Au début de l'expérience la cellule de référence et la cellule d'analyse sont placées à la même température, le plus souvent 15°C , 25°C ou 35°C . En général, le ligand est chargé dans une seringue placée dans un dispositif d'injection automatique extrêmement précis, puis la seringue est insérée dans la cellule d'analyse contenant la protéine. Le ligand va être injecté progressivement par le biais d'une série de petites injections dans la solution protéique. Au cours de ma thèse, j'ai réalisé des expériences de titrations dites « inverses », puisque le ligand est placé dans la cellule d'analyse tandis que la protéine est chargée dans la seringue et injectée dans la solution contenant le ligand. Si le ligand se lie à la protéine, des variations de chaleur de l'ordre de quelques millièmes de degré Celsius vont pouvoir être détectées et mesurées. La quantité de chaleur mesurée est directement proportionnelle au nombre de liaisons établies entre la protéine et le ligand.

La **Figure 65-B**, présente un exemple de résultat obtenu. Dans cet exemple, la réaction est exothermique, c'est-à-dire qu'elle libère de la chaleur. Ainsi la cellule d'analyse devient plus chaude que la cellule de référence. Le pic du signal est alors orienté vers le bas et lorsque la température des deux cellules redevient identique, le signal revient à sa valeur initiale (le zéro). À chaque injection, le microcalorimètre compense la petite variation de température détectée jusqu'au retour du système à l'équilibre. Au fur et à mesure des injections et de l'augmentation de la concentration en ligand, il y a de moins en moins de site de fixation disponible sur la protéine et on observe progressivement une diminution de la chaleur dégagée au cours des différentes injections. En fin d'expérience, lorsque tous les sites de liaison sont saturés par la présence du ligand on observe uniquement la chaleur de dilution du ligand dans le tampon. L'aire sous chaque pic de chaleur mesurée est ensuite intégrée et tracée en fonction du rapport ligand/protéine et le thermogramme résultant prend l'aspect d'une courbe sigmoïde qui permet de mesurer l'enthalpie de liaison ΔH (chaleur mesurée en début d'expérience), sa stœchiométrie qui correspond au rapport molaire au point d'inflexion de la courbe, ainsi que la constante de dissociation K_D de l'interaction correspondant à la pente de la courbe mesurée en ce point. Les valeurs d'énergie libre de Gibbs (ΔG) et d'entropie (ΔS) sont déduites des valeurs d'enthalpie et de K_D .

Figure 65 : Principe de l'ITC.



Légende : A- Le système de titration est composé d'une cellule de référence remplie d'eau, d'une cellule d'analyse dans laquelle se trouve la protéine (ou l'inhibiteur dans le cas d'une titration inverse), d'une seringue d'injection possédant une aiguille aplatie permettant l'homogénéisation de la solution et chargée avec l'inhibiteur (ou la protéine dans le cas de la titration inverse). B- Exemple de résultat de titration obtenue pour une réaction exothermique. En haut la courbe des chaleurs mesurées lors de l'injection du ligand dans la cellule au cours du temps, en bas l'isotherme résultant de l'intégration de l'aire sous la courbe en fonction du ratio molaire ligand/protéine. La stœchiométrie de la réaction correspond au ratio molaire au niveau du point d'inflexion de la courbe de titration. L'affinité, plus précisément $1/K_D$ correspond à la pente de la courbe au point d'inflexion. L'enthalpie correspond à la différence d'énergie entre le début et la fin de la titration.

II. Avantages et inconvénients de l'ITC

Les avantages : Contrairement à l'HTRF et le TSA, elle n'utilise pas la fluorescence, ainsi il n'y a pas de problèmes d'interférence avec des molécules auto-fluorescentes. Par ailleurs, c'est une méthode en solution, qui ne nécessite aucun substrat comme par exemple dans le cas de l'HTRF, ni de marquage ou d'immobilisation protéique comme dans le cas des méthodes d'interférométrie et de SPR. Le principal avantage de cette méthode est qu'elle permet de quantifier l'interaction protéine-ligand et qu'elle donne accès aux paramètres thermodynamiques de l'interaction en une seule expérience.

Les inconvénients : Cette méthode nécessite une grande quantité de protéine (ou de ligand en mode titration inverse). L'appareil est très sensible aux faibles variations de concentrations entre les solutions contenues dans la cellule et dans la seringue et il est nécessaire de s'assurer d'avoir exactement le même tampon pour la protéine et le ligand. Ce paramètre peut poser problème dans le cas de la mesure des interactions avec les petites molécules pour lesquelles le DMSO utilisé pour solubiliser les petites molécules peut générer de grandes quantités de chaleur de dilution lors des injections. Le temps de réalisation d'une expérience est assez long, en général 1 heure par ligand, de ce fait le criblage de molécules par cette méthode n'est pas possible.

III. Protocole d'ITC (titration inverse)

La protéine est passée au préalable sur une petite colonne de dessalage de 2,5 mL de type PD10 (GE Healthcare) avec du tampon 10 mM Hepes pH 7,5 et 150 mM NaCl. L'inhibiteur est dilué avec ce même tampon. La titration « inverse » est réalisée à l'aide d'un ITC200 (GE Healthcare, Piscataway, NJ). La protéine est chargée dans la seringue (43 μ L) tandis que la molécule est placée dans la cellule d'échantillon (300 μ L). La concentration en protéine est fixée entre 10 à 15 fois celle de la molécule qui est elle-même comprise entre 10 et 40 μ M. L'expérience est réalisée à une température fixe de 15, 25 ou 35°C. En général, on effectue entre 13 et 17 injections de protéine dans la cellule contenant la molécule et la première injection est fixée à 0,2 μ L pour éliminer les bulles d'air présentes éventuellement dans la seringue avant de débiter la titration. Les données brutes sont mises à l'échelle, pour cela le zéro peut être ajusté en soustrayant un contrôle négatif (protéine injectée dans la cellule contenant le tampon) ou alors il peut être ajusté avec la valeur de saturation lorsque cela est possible.

Les données d'ITC brutes sont ensuite intégrées et ajustées avec le logiciel Origin 9.1 et le plugin MicroCal Origin en suivant un modèle à un site et un ajustement des moindres carrés. Enfin, les valeurs ΔG et ΔS ont été calculées à partir des valeurs ajustées ΔH et K_A , en utilisant les équations suivantes : $\Delta G = - R \times T \times \ln(K_A)$ et $\Delta G = \Delta H - T \times \Delta S$.

*Partie 8 : Viabilité cellulaire &
Expression de c-Myc*

Dans le cadre de notre étude cellulaire, différentes lignées cellulaires leucémiques humaines ont été choisies. À savoir, la lignée cellulaire Jurkat, qui sont des cellules de leucémie lymphoblastique T aiguë et la lignée cellulaire Molm14, qui sont des cellules de leucémie myéloïde aiguë. La lignée de cancer du sein HeLa FTR a également été utilisée comme contrôle positif pour les tests HTRF c-Myc.

I. Culture des lignées cellulaires

Une ampoule contenant 10 millions de cellules Jurkat (ATCC® TIB-152) ou Molm14 (obtenues auprès du Dr. J.E Sarry) ou HeLa FTR (T-Rex™ HeLa Cell Line, ThermoFisher Scientific) supplémentée de 10% DMSO est décongelée dans un bain marie à 37°C jusqu'à ce qu'il ne reste qu'un petit glaçon. Le contenu de l'ampoule est transféré dans un falcon 15 mL contenant 10 mL de milieu RPMI-1640 pour les cellules Jurkat, de milieu MEM α pour les cellules Molm14 et de milieu DMEM pour les cellules HeLa, tous trois additionnés de 10% Sérum de Veau Fœtal déplémenté d'immunoglobuline IgG (FCS Δ), pour une concentration finale de 1 million de cellules / mL. Le falcon est centrifugé à 4°C pendant 5 minutes à une vitesse de 1400 rpm et le surnageant est éliminé. Le culot cellulaire est repris avec 10 mL de milieu RPMI-1640 ou de MEM α ou DMEM additionnés de 10 % FCS Δ . Les cellules sont ensuite mises en culture à 37°C et 5% de CO₂ dans un flacon Corning de 25 cm².

Les cellules Jurkat et Mol14 sont en suspension, ainsi tous les deux jours, elles sont coupées à 0,3 million de cellules / mL et remises en culture. Les cellules HeLa FTR sont quant à elles, adhérentes, ainsi tous les deux jours elles sont observées et dès que la confluence atteint 80% elles sont trypsinisées. Pour cela le surnageant est aspiré, la flasque est rincée avec 5 mL de PBS afin d'éliminer le FCS, puisque l'activité de la trypsine est inhibée par le FCS. Le PBS est éliminé, puis 2 mL de trypsine 0,05% sont déposés sur les cellules et les flasque est mise à l'incubateur à 37°C pendant 1 à 2 minutes jusqu'à ce que les cellules se décroche. La réaction de trypsinisation est stoppée par l'ajout de 8 mL de milieu DMEM additionné de 10% FCS. Elles sont ensuite coupées au 1/10^e et remises en culture. Toutes les cultures cellulaires sont entretenues au maximum un mois, au-delà elles dérivent.

Pour le comptage de cellules, 10 μ L de cellules sont mélangées à 10 μ L de bleu trypan et sont déposés sur une cellule Kova. La formule utilisée pour compter le nombre de cellules est la suivante : (nombre de cellules comptées / nombre de carreaux comptés) * 90 (ratio cellule de Kova) * 1000 (pour passer de μ L à mL) * facteur de dilution.

II. Viabilité cellulaire ou cytotoxicité (manuel)

Les cellules sont préalablement visualisées au microscope inversé (Primovert, Zeiss), puis comptées à l'aide du compteur Luna™*fl* de la société mokascience. Pour cela, 10 µL de cellules ont été mélangées avec 10 µL de bleu trypan. Les cellules HeLa sont ensuite diluées afin d'obtenir une concentration finale 5000 cellules/ puits/ 90 µL, sont distribuées dans des plaques 96 puits de type Corning Costar ® 3610 (White with Clear Flat Bottom) et sont incubées au préalable 24h à 37°C à et 5% de CO₂, afin qu'elles adhèrent au fond des puits. Les cellules Jurkat et Molm14 sont quant à elles diluées afin d'obtenir une concentration finale de 10000 cellules/ puits/ 90 µL et aucune pré-incubation n'est nécessaire, puisqu'elles sont en suspension.

Une gamme de concentration en inhibiteur (8 points) est préparée par dilution en cascade du composé dans des plaques 96 puits (Cellstar, cell culture plate, V-bottom, Greiner Bio-one). Deux contrôles sont également réalisés, avec la molécule JQ1 (un pan BET-inhibiteur connu) et du DMSO. Chaque point est réalisé en triplicata. Dans les premières expériences, les composés ont été dilués en cascade directement dans du milieu RPMI-1640 free sans FCS et en DMSO constant (10% DMSO final). La gamme de concentration allant de 100 µM à 3,1 nM. Suite à l'observation de précipités pour les doses les plus élevées (100, 50, 25 µM) pour certaines de nos molécules, entraînant des dilutions en cascades faussées, les dilutions ont par la suite été réalisées en deux étapes. Tout d'abord en cascade en 100% DMSO, puis chaque dilution est transférée dans du milieu RPMI-1640 sans FCS (10% DMSO final). La gamme de concentration étendue allant de 50 µM à 1,3 nM. Enfin, certaines molécules très peu solubles ont également été diluées en cascade en deux étapes. La première étape en 100% DMSO, et la seconde étape consistant en une dilution au 1/100^e en 10% hydroxypropyl-β-cyclodextrine ou HPCD (1% en DMSO final). La gamme de concentration allant de 50 µM à 1,3 nM.

Dans une plaque 96 puits de type Corning Costar ® 3610 et 3903, (White with Clear Flat Bottom), 90 µL de cellules sont incubées à 37°C et 5% de CO₂ pendant 72 heures, en présence de 10 µL d'inhibiteurs, de la molécule JQ1 ou de DMSO (contrôles). Au bout de 72 heures, la plaque de cellules est laissée à TA pendant 30 minutes et le tampon CellTiter-Glo est décongelé également à TA. Ensuite, 50 µL de CellTiter-Glo (1/2 du volume du puits) sont ajoutés aux cellules. La plaque est mélangée 2 minutes à l'aide d'un shaker orbital permettant la lyse cellulaire, puis elle est laissée 10 minutes à TA. La luminescence engendrée par l'action de la luciférase est mesurée à l'aide du Centro luminometer LB960 (Berthold). Le signal lumineux est proportionnel au niveau d'ATP et la quantité d'ATP présent dans le milieu reflète le nombre de cellules viables (métaboliquement actives). Les courbes de dose-réponse sont générées et la dose efficace à 50% (EC₅₀) est calculée en utilisant une régression non-linéaire à l'aide du logiciel

Graph Pad Prism (version 5.03). L'équation utilisée est la suivante : $Y=100/(1+10^{((\text{LogEC}_{50}-X)*n)})$, dans laquelle **X** correspond à la concentration en inhibiteur, **Y** la valeur de fluorescence normalisée et **n** la pente de la courbe à la valeur d' EC_{50} (ou coefficient de Hill).

III. Viabilité cellulaire ou cytotoxicité (au robot)

Pour les tests de viabilité réalisés au robot « Echo® », les cellules sont diluées afin d'obtenir une concentration finale de 2500 cellules/ puits/ 25 μL . La gamme de concentration en inhibiteur allant de 100 μM à 3,1 nM est réalisée à l'aide du module « **dose response** » de l'Echo® avec une concentration finale de 1% en DMSO. La molécule JQ1 (pan-BET inhibiteur), ainsi que le DMSO seul sont également utilisés comme contrôles. Dans une plaque 96 puits de type Corning Costar ® 3610 et 3903, (White with Clear Flat Bottom), 250 nL de molécules ou de DMSO (1% final DMSO) sont transférés à l'Echo® et 25 μL de cellules sont dispensés au « Multidrop Combi ». Les plaques sont ensuite incubées à 37°C et 5% de CO_2 pendant 72 heures.

Au bout de 72 heures, la plaque de cellules est laissée à TA pendant 30 minutes et le tampon CellTiter-Glo est également décongelé à TA. Ensuite 12,5 μL de CellTiter-Glo (1/2 du volume du puits) sont ajoutés aux cellules. Les cellules sont lysées 2 minutes par agitation sur le « Multidrop Combi », puis la plaque est laissée 10 minutes à TA. La luminescence est mesurée à l'aide du type « PHERAstar FS » (BMG Labtech) et les courbes de dose-réponse sont générées et la dose efficace à 50% (EC_{50}) est calculée en utilisant une régression non-linéaire à l'aide du logiciel Graph Pad Prism (version 5.03). L'équation utilisée est la suivante : $Y=100/(1+10^{((\text{LogEC}_{50}-X)*n)})$, dans laquelle **X** correspond à la concentration en inhibiteur, **Y** la valeur de fluorescence normalisée et **n** la pente de la courbe à la valeur d' EC_{50} (ou coefficient de Hill).

IV. Expression de c-Myc (western-blot)

Une partie des cellules Jurkat ne sont pas traitées (contrôle) et les autres sont traitées 24 heures avec un inhibiteur, ainsi que la molécule JQ1 et du DMSO (contrôles). Les cellules non traitées et traitées sont ensuite lysées dans du tampon de lyse RIPA additionné d'anti-protéase 100X, de PMSF 1 mM et de Triton 1%. Pour cela, le culot cellulaire est lavé dans 1 mL de PBS, puis centrifugé dans un eppendorf pendant 5 minutes à une vitesse de 1500 rpm. Un million de cellules sont lysées avec 50 μL de tampon de lyse pendant 20 minutes à 4°C sur une roue. Les cellules sont ensuite centrifugées 20 minutes à 4°C à une vitesse de 1300 rpm. Le surnageant est récupéré dans un nouvel eppendorf. La quantité de protéine est déterminée par la méthode du Bradford.

Le Bradford est préalablement dilué au 1/5^e et une gamme étalon de BSA est préparée, allant de 0 à 15 µg/mL en ajoutant 2 µL de tampon de lyse afin d'avoir la même concentration en tampon que pour la protéine. 2 µL d'échantillons sont mélangés à 1 mL de Bradford dans des cuvettes, en veillant à bien homogénéiser. Le tout est incubé 5 minutes à TA avant lecture à 595 nm sur un spectromètre.

L'échantillon est préalablement dénaturé par ajout de bleu de charge et chauffage 5 minutes à 95°C avant d'être déposé sur gel d'acrylamide SDS-PAGE 10%, (75 µg ou 15 µL). Un marqueur de poids moléculaire est également déposé sur le gel. Le gel migre tout d'abord 15 minutes à 90 V, puis 1h30 à 130 V. Les protéines du gel sont ensuite transférées sur une membrane de nitrocellulose (Hybond C-extra membranes de 45 micron, GE Healthcare). Le transfert s'effectue avec un bloc de glace pendant 1h30 à 70 V. La membrane est saturée avec 5% (wt/vol) de lait en poudre dans du tampon TBST pendant 1 heure à TA (ou sur la nuit à 4°C). La membrane est ensuite lavée 10 minutes avec du tampon TBST avant d'être incubée avec l'anticorps primaire anti-myc (clone 9E10, sc40 Santa Cruz) au 1/500^e dans du TBST avec 0,5% (wt/vol) de lait en poudre sur la nuit à 4°C. La membrane est de nouveau lavée trois fois avec du TBST, puis incubée avec l'anticorps secondaire conjugué HRP anti-mouse (polyclonal goat anti mouse P0447, Pierce) au 1/20000^e dans du TBST pendant 1 heure à TA. Enfin, la membrane est lavée trois fois 10 minutes dans du TBST. Les bandes immuno-réactives sont révélées à l'aide des réactifs de détections SuperSignal™ West Pico Extended Duration Substrate (Pierce). Pour cela, les solutions du kit Stable Peroxide et Luminol/Enhancer sont mélangés à volume équivalent (5 mL par membrane) et incubés 5 minutes sous agitation à l'abri de la lumière, avant révélation. La quantification des bandes est effectuée à l'aide du logiciel ImageJ.

V. Expression de c-Myc (HTRF)

Les expériences d'expression de c-Myc ont été réalisées à l'aide du kit Human c-Myc Cell Based Assay fourni par la société Cisbio. Ce kit comprend un tampon de lyse, un lysat cellulaire (contrôle positif) et deux anticorps anti c-Myc couplés à des fluorophores, l'un est donneur (c-Myc cryptate) et l'autre accepteur (c-Myc d2). Par ailleurs, ce test HTRF a déjà été monté par la société Cisbio, sur la lignée du cancer du sein HeLa en absence et en présence de la molécule JQ1. J'ai donc réalisé les tests d'expression de c-Myc sur les lignées Jurkat, Molm14 et HeLa comme contrôle positif. Pour la lignée HeLa les conditions étaient connues (40000 cellules par puits/ 90 µL de milieu), tandis que pour les lignées Jurkat et Molm14, j'ai dû optimiser les conditions et tester plusieurs concentrations (10.000, 50.000, 100.000, 200.000 et 400.000 cellules par puits/ 90 µL de milieu) ont été testés afin d'identifier la condition donnant le meilleur signal d'HTRF. Le rapport signal/bruit pour la lignée HeLa était semblable à celui du contrôle positif du kit, celui

des Jurkat était inférieur mais suffisant pour pouvoir interpréter les résultats, en revanche celui des Molm14 était trop faible. L'expression de c-Myc et son inhibition de manière dose dépendante en présence JQ1 (contrôle positif) a pu être déterminée sur les lignées HeLa et Jurkat, mais pas sur les Molm14.

Dans une plaque 96 puits de type Corning Costar ® 3610, (White with Clear Flat Bottom), les cellules HeLa (40000 cellules/puits/90µL) sont incubées au préalable 24 heures à 37°C et 5% de CO₂, afin qu'elles adhèrent au fond des puits. Le lendemain 10 µL d'inhibiteurs, de la molécule JQ1 ou de DMSO (contrôles) sont ajoutés, puis la plaque est incubée de nouveau 24 heures à 37°C et 5% de CO₂. Les cellules Jurkat et Molm14 (90µL/puits) sont quant à elles déposés directement sans pré-incubation 24h, dans des plaques 96 puits de type Corning Costar® 3610, puis 10 µL d'inhibiteurs, de la molécule (+)-JQ1 ou de DMSO (contrôles) sont ajoutés et les plaques sont incubées 24 heures à 37°C et 5% de CO₂.

Au bout de 24 heures, les plaques de cellules sont laissées à TA quelques minutes. Pour les cellules HeLa (adhérentes) le surnageant est aspirées, tandis que pour les cellules Jurkat et Molm14 (en suspension) les cellules sont transférées dans des tubes eppendorf de 1,5 mL, les puits sont rincés avec du PBS afin de récupérer toutes les cellules, puis le tout est centrifugés pendant 5 min à 1400 rpm à 20°C et enfin le surnageant est aspiré. Ensuite 50 µL de tampon de lyse sont ajoutés et les cellules sont mélangées 45 minutes à TA l'aide d'un shaker orbital permettant la lyse cellulaire. Dans un plaque 384 Small Volume™ HiBase Polystyrene Microplates, 10 µL de Master mix contenant les deux anticorps à volume égal (5 µL de c-Myc Cryptate + 5 µL de c-Myc d2) sont mélangés à 10 µL de lysat cellulaire (non traitées et traitées avec l'inhibiteur ou JQ1). Des contrôles supplémentaires sont ajoutés, d'une part les cellules non traitées (10 µL) en présence uniquement de l'anticorps donneur (5 µL de c-Myc Cryptate + 5 µL de tampon) et d'autre part un contrôle négatif avec du tampon de lyse (10 µL) et du Master mix (10 µL).

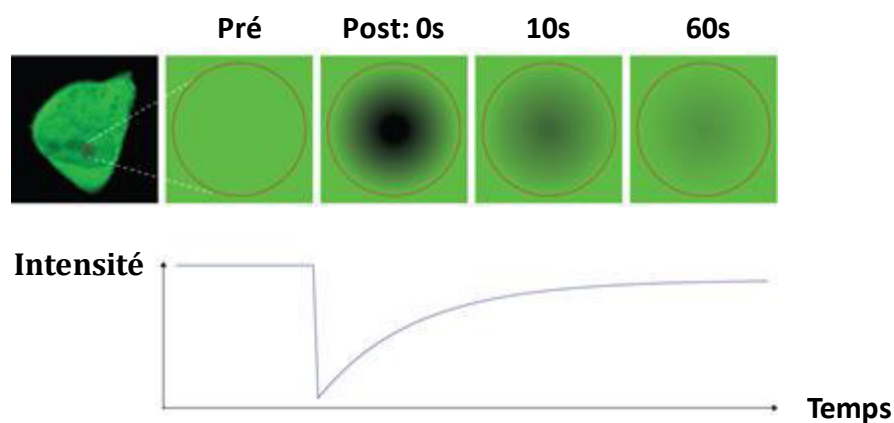
Les courbes de dose-réponse sont générées et la dose efficace à 50% (EC₅₀) est calculée en utilisant une régression non-linéaire à l'aide du logiciel Graph Pad Prism (version 5.03). L'équation utilisée est la suivante : $Y=100/(1+10^{((\text{LogIC50}-X)*n)})$, dans laquelle **X** correspond à la concentration en inhibiteur, **Y** la valeur de fluorescence normalisée et **n** la pente de la courbe à la valeur d'IC50 (ou coefficient de Hill).

*Partie 9: Validation Cellulaire :
Fluorescence Recovery After
Photobleaching (FRAP)*

I. Principe du Fluorescence Recovery After Photobleaching

La technique de FRAP, est basée sur la redistribution de la fluorescence après photoblanchiment. (Figure 66) qui sera influencée par la vitesse de diffusion des molécules. La capacité d'une population cellulaire homogène, à émettre de la fluorescence peut être inhibée de manière réversible à l'aide d'un flash lumineux (laser) et dans une zone délimitée. Deux populations très distinctes se forment alors, l'une fluorescente et l'autre pas. Une redistribution des deux populations est observée et enregistrée jusqu'à homogénéisation des populations. Le temps de régénération (demi-période de rétablissement de la fluorescence) correspond à la vitesse de déplacement de la protéine.

Figure 66: Principe du FRAP.



Légende : La figure est issue du site : <https://cam.facilities.northwestern.edu/588-2/fluorescence-recovery-after-photobleaching/>. Un flash laser est envoyé sur la cellule à $t=0$ sec, une tache noire apparaît dans la zone où le laser a tapé la cellule. Cette tache disparaît au cours du temps grâce à la redistribution de la fluorescence.

II. Protocole du FRAP

Les expériences de FRAP ont été réalisées essentiellement comme décrit dans l'article de (259) et en collaboration avec **Stefan Knapp**, **Susanne Müller** et **Catherine Rogers** au sein du **SGC** à d'Oxford. Brièvement, les lignées cellulaires humaines d'ostéosarcome (U₂OS, ATCCR HTB-96™) ont été transfectées (Fugene HD; Roche) avec une construction de mammifère sur-exprimant la protéine BRD4 entière fusionnée avec la GFP en position N-terminale. Les cellules U₂OS sont maintenues en culture dans du milieu DMEM supplémenté de 10% de Sérum de Bœuf Fœtal (FBS) à 37°C et 5% de CO₂.

Le FRAP et le système d'imagerie est composé d'un système Zeiss LSM 710 à balayage laser, couplé à un microscope « inversé » Zeiss Axio Observer.Z1 possédant un objectif à immersion 40 x (Zeiss) de très grande ouverture (N. A. 1,3). L'échantillon cellulaire est placé dans une chambre d'incubation où la température et l'humidité sont maintenues. Les expériences de fluorescence FRAP et GFP ont été réalisées avec un laser à argon (488 nm) et un détecteur PMT (500 – 550 nm). Un premier balayage laser est effectué, puis une région d'intérêt correspondant à 50% du noyau positif à la GFP est arbitrairement sélectionnée pour le photoblanchiment. Une série de mesures est ensuite effectuée, afin d'enregistrer le retour de la fluorescence GFP en utilisant 1% de la puissance du laser. Les images et les données de fluorescence ont été exportées du logiciel de contrôle du microscope ZEN 2009 vers un fichier Excel pour déterminer la demi-période moyenne pour la récupération complète de l'équilibre de la fluorescence, et ce pour 10 à 20 cellules par point de traitement (avec inhibiteur). En général l'inhibiteur est ajouté 1 heure avant le début des enregistrements et 24 heures après la transfection.

*Partie 10 : Caractérisation de
complexes protéine-ligand par
cristallographie aux rayons X*

Différentes techniques permettent de caractériser les interactions au sein de complexes protéine-ligand, telles que la RMN, la microscopie électronique et la cristallographie aux rayons X. Au cours de ma thèse, j'ai cristallisé les protéines suivantes : BRD4(1), BRD4(2), BRD3(2) et BRD2(2), seules ou en complexe avec des ligands et j'ai résolues les structures cristallographiques.

I. Définitions

La cristallogenèse est un processus au cours duquel un soluté va passer d'un état liquide désordonné à un état solide ordonné, formant un *cristal*.

Un cristal est un solide ordonné, au sein duquel l'ensemble de ses constituants cristallisés (molécules, atomes et ions) forment un motif. Plusieurs motifs sont empilés de manière régulière et périodique à l'identique, formant une « *maille* » qui est répétée dans les trois dimensions de l'espace. Selon la symétrie du cristal une maille peut être constituée d'un ou plusieurs motif(s). Dans le cas de cristaux de protéines, certaines molécules de solvant, ainsi que les régions flexibles des protéines n'obéissent pas à cette périodicité. Le cristal est constitué de 30 à 70% de solvant (eau, tampon, ions...). La cohésion entre les protéines au sein d'une même maille, ainsi qu'entre les différentes mailles du cristal sont permises grâce à l'établissement d'un réseau de liaisons hydrogènes, d'interactions de van der Waals et de ponts salins au sein du cristal. Cet arrangement périodique confère au cristal la propriété de diffracter les rayons X.

II. Principe de la cristallogenèse

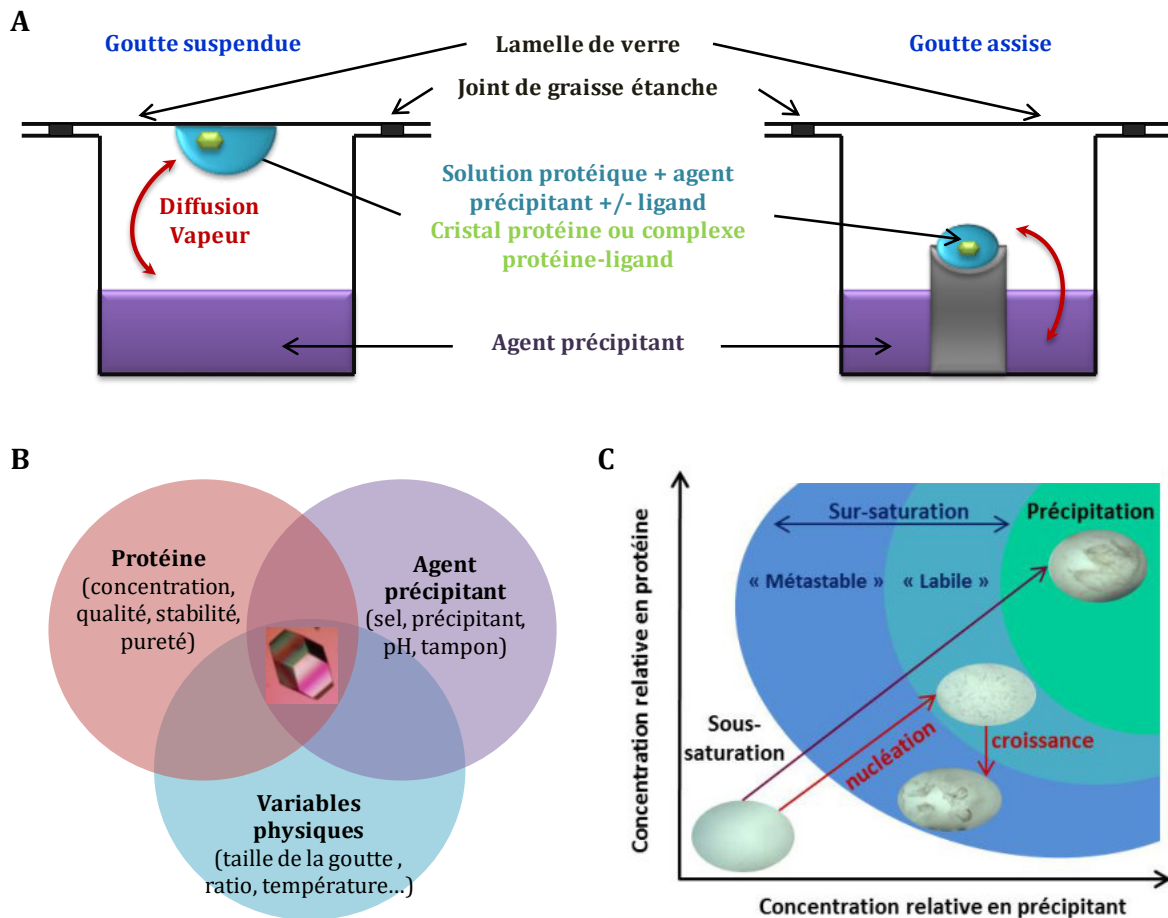
La cristallogenèse est basée sur le principe de diffusion d'espèces volatiles sous forme gazeuse entre une goutte (petit volume allant de 200 nL à 10 µL) contenant la protéine, une solution d'agent précipitant et éventuellement un ligand et un réservoir (grand volume allant de 70 µL à 1 mL) contenant une solution du même agent précipitant, dans une enceinte fermée hermétiquement (**Figure 67-A**). Les agents précipitants les plus couramment utilisés sont les sels (par exemple le sulfate d'ammonium), les polymères (par exemple le polyéthylène glycol 400 à 20.000) et les solvants organiques (par exemple l'éthanol). Ils ont pour rôle de diminuer la solubilité de la protéine dans la goutte et de favoriser la nucléation. En effet, l'agent précipitant étant majoritaire dans le réservoir, le transfert de molécules d'eau va se faire de la goutte vers le réservoir jusqu'à l'équilibre. Ainsi, la concentration de tous les constituants de la goutte va augmenter, entraînant la protéine dans un état de sursaturation propice à la formation cristalline. La formation de cristaux est donc dépendante de la cinétique d'échange entre le réservoir et la goutte, qui est elle-même dépendante de la température, la nature et les

concentrations initiales des constituants de la goutte et de la surface d'échange de la goutte « assise » ou « suspendue » (**Figure 67-B**).

Deux critères sont déterminants pour la réussite et la reproductibilité de la cristallisation. La protéine doit être la plus **pure** possible **chimiquement**, c'est-à-dire l'absence de contaminants dans la solution protéique et **conformationnellement**, c'est-à-dire **monodisperse** et **sans agrégats** (**Figure 67-B**). La monodispersité de la solution de protéine peut être évaluée par diffusion dynamique de la lumière (**DLS** pour **Dynamic Light Scattering**) ou chromatographie d'exclusion de taille analytique seule ou couplée à des analyses de la masse (par exemple SEC-MALLS).

La formation de cristaux suit le diagramme de phase de la protéine (**Figure 67-C**). La protéine est initialement soluble à l'état d'équilibre thermodynamique dans la zone de sous-saturation. Au cours des échanges d'espèces volatiles entre la goutte et le réservoir, la protéine se concentre jusqu'à atteindre sa limite de solubilité. La protéine passe d'abord par une zone de sursaturation dite « labile » qui est propice à la formation de germes cristallins (nucléation) et à leur croissance. La vitesse de nucléation est d'autant plus rapide qu'on se trouve loin dans la zone « labile ». Plus on se trouve loin dans cette zone, plus il y aura formation de cristaux de petites tailles et imparfaits qui ne seront pas exploitables en termes de diffraction des rayons X. Au-delà de la zone de sursaturation, il y aura précipitation. La croissance cristalline est permise grâce à la fixation de molécules disponibles en solution et cette étape dure tout le temps que la protéine en solution est disponible (zone de sursaturation). La croissance cristalline se poursuit ensuite dans la zone de sursaturation dite « métastable », puis elle s'arrête lorsque l'état d'équilibre est atteint en sortie de zone de sursaturation. Dès lors des défauts structuraux ou des impuretés sur les faces du cristal peuvent apparaître.

Figure 67 : Cristallogenèse.



Légende : A- Principe de cristallogenèse en goutte suspendue (gauche) ou assise (droite). Le réservoir est rempli d'une solution d'agent précipitant. La goutte contient la protéine, l'agent précipitant et éventuellement un ligand. Un échange sous forme de diffusion de vapeur se fait entre le réservoir et la goutte, ce qui favorise la formation de cristaux. B- Critères déterminant la cristallisation. La cristallisation est dépendante de la protéine, de l'agent précipitant et de variables physiques. C- Diagramme de phase d'une protéine en fonction de sa propre concentration et celle en agent précipitant. Au cours de sa concentration, la protéine passe d'un état soluble (sous-saturation), à un état de sursaturation. La zone de sursaturation « labile » est favorable à la formation de germes cristallins et à la croissance cristalline. La zone de sursaturation « métastable » est quant à elle uniquement favorable à la croissance cristalline. Les zones de sous-saturation et précipitation sont défavorables à la cristallisation. En effet, dans la zone de sous-saturation la concentration en protéine ou agent précipitant n'est pas assez importante pour favoriser la nucléation, à l'inverse dans la zone de précipitation, la concentration en protéine ou agent précipitant est trop élevée.

III. Criblage de conditions de cristallisation

Lorsqu'on souhaite cristalliser une protéine, la première étape consiste à chercher dans la PDB s'il existe une ou plusieurs structure(s) connue(s) pour celle-ci et de reproduire la ou les condition(s) de cristallisation décrites(s). Néanmoins, il arrive parfois que les conditions décrites dans la PDB ou dans les articles soient incomplètes, difficilement reproductibles, voir même non reproductibles. C'est pour ces raisons qu'un criblage de conditions de cristallisation est parfois nécessaire lorsqu'on cristallise pour la première fois une protéine.

Ayant eu des difficultés à reproduire les conditions de cristallisation publiées pour la protéine BRD3(2), j'ai réalisé des criblages pour la protéine seule et en complexe avec certains ligands. Ceux-ci ont été réalisés au sein de la plateforme de biologie structurale (plateforme labélisée FRISBI) du laboratoire **AFMB** (Architecture et Fonction des Macromolécules) de Marseille. Cette plateforme dispose de robots (Tecan Genesis, LabX et Mosquito® Crystal, TTPLabtech) nécessaires à la réalisation des criblages de conditions et d'un large panel de kits de conditions provenant de divers fournisseurs.

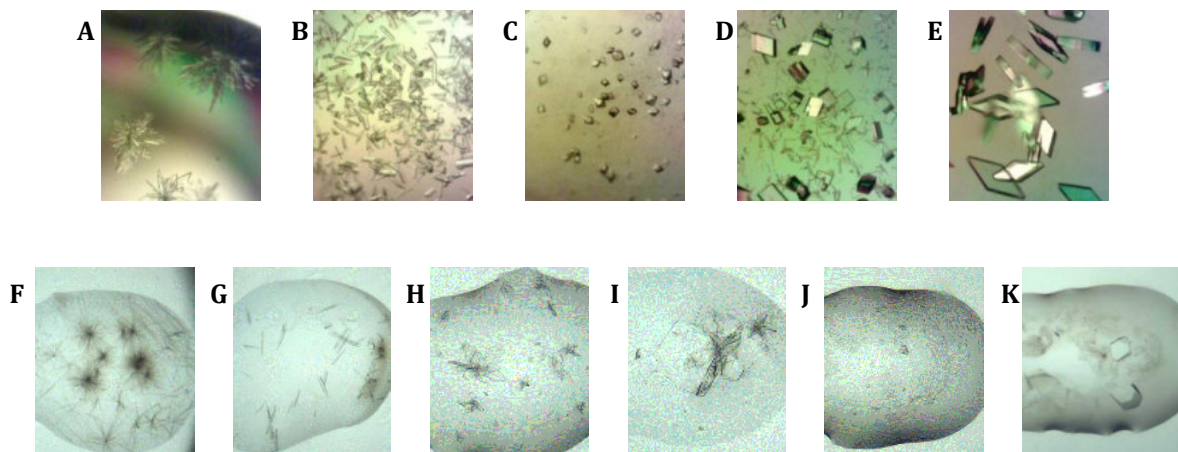
La protéine BRD3(2) a été concentrée à 20 mg/mL puis incubée avec l'inhibiteur à une concentration finale de 2 mM (10% DMSO). La protéine a ensuite été centrifugée afin d'éliminer les précipités et la concentration finale en présence de ligand a été mesurée à 8 mg/mL. Les criblages ont été réalisés dans des plaques de cristallogénèse 96 puits SwisSCI à trois cupules. Le réservoir a été préalablement rempli à l'aide du Tecan Genesis (LabX), avec 70 µL de solution de précipitant issus de divers kits. Les kits utilisés ont été : STURA, Wizard, JCSG+ et PACT (Fournisseurs : Molecular Dimension, Qiagen et Hampton). Le mélange protéine-ligand, ainsi que l'agent précipitant ont ensuite été distribués à l'aide d'un robot Mosquito® Crystal (TTPLabtech) en faisant varier le rapport de volumes protéine/réservoir dans les gouttes, à savoir 300:100, 200:100 et 100:100 nL. Les plaques ont ensuite été incubées à 20°C dans un incubateur de type Rock Imager (Formulatrix) et la cristallisation a été suivie par collecte automatisée d'images des gouttes pour analyse de ces clichés *via* un serveur de visualisation HTXLab CRIMS (Crystallization Information Management System) développé par l'EMBL de Grenoble.

IV. Optimisation des conditions de cristallisation

La limite de diffraction d'un cristal (résolution) a un impact majeur sur la qualité des données de diffraction permettant de résoudre la structure tridimensionnelle de celui-ci, c'est pourquoi il est important de déterminer la meilleure condition favorable à la croissance cristalline de cristaux de bonnes qualités (cristaux uniques ne formant pas de clusters) et ayant une taille suffisante (supérieure au faisceau de rayons X) et de bonnes propriétés de diffraction aux rayons X. Il existe des lignes microfoc (par exemple les lignes ESRF ID23-2 et ID30A-3) pour collecter les microcristaux, seulement nous n'y avons accès que quelques fois par an. Au cours de ma thèse, j'ai été amené à optimiser les conditions de cristallisation pour les protéines BRD4(1), BRD3(2) et BRD2(2).

Ainsi, lorsqu'on souhaite optimiser une condition de cristallisation, notamment après les premières touches obtenues par une approche de criblage de conditions, on peut faire varier de nombreux paramètres tels que la concentration en protéine et/ou en agent précipitant, la température, la nature et la concentration du tampon, le pH, la force ionique, la nature de l'agent précipitant et la présence d'additifs (260, 261). La **Figure 68**, présente différents types de cristaux obtenus après optimisation des conditions de cristallisation.

Figure 68 : Optimisation des conditions de cristallisation.



Légende : Exemples de cristaux obtenus après optimisation des conditions de cristallisation pour la protéine BRD2(2) photos du haut et pour BRD3(2) photos du bas. **A-** cluster de cristaux, **B-** nombreux cristaux, **C-** cristaux de très petite taille < 50 μM , **D-** cristaux de plus grande taille, **E-** cristaux de taille idéale. **F-** oursins, **G-** très fines baguettes, **H-** petit cluster de plaquettes, **I-** cluster de grosses plaquettes très fines, **J-** nombreux micro-cristaux, **K-** cristaux de taille et de forme optimale.

V. Cristallisation des complexes protéine/ligand

Afin de développer des inhibiteurs de bromodomains, il est crucial de caractériser au niveau moléculaire le mode de reconnaissance des inhibiteurs identifiés par HTRF et validés par ITC et TSA. Pour cela, il est nécessaire de résoudre les structures cristallographiques aux rayons X des complexes protéine-ligand. L'obtention de ces complexes peut se faire *via* deux méthodes à savoir la co-cristallisation et le trempage.

1. La co-cristallisation

Le principe de la co-cristallisation d'un complexe protéine–ligand est très similaire à celui de la cristallisation d'une protéine seule. Par contre, la présence de ligand peut drastiquement changer les conditions de cristallisation, c'est pourquoi il est parfois nécessaire d'optimiser les conditions préalablement identifiées. Ainsi, la cristallisation consiste à mélanger la protéine et l'agent précipitant dans une même goutte, tandis que la co-cristallisation consiste à rajouter

à ce mélange le ligand en concentration suffisante pour assurer la saturation du site actif de toutes les protéines en cours de cristallisation.

Les protéines de la famille BET sont sensibles au DMSO et plus particulièrement les BD2. De plus, il a été montré que le DMSO peut se fixer dans la poche de reconnaissance des lysines acétylées et inhiber l'interaction avec le peptide à de trop fortes concentrations. Il est donc primordial que la concentration en DMSO dans la goutte ne soit pas trop élevée et j'ai donc défini un seuil maximal en DMSO de 10% pour les BD1 et de 5% pour les BD2, dans mes tests de cristallogenèse. Classiquement 0,1 à 0,2 μL d'inhibiteur à 10, 25 ou 50 mM en DMSO (100%) ont été déposés sur une lamelle, puis 1 μL de protéine à la concentration optimale et enfin 1 μL d'agent précipitant ont été ajoutés avant homogénéisation à la pipette et mise en équilibre.

Plus de 80% des structures que j'ai résolues (sur plus de 200 structures) étaient la protéine sous forme libre ou en présence de DMSO dans la poche et non pas en complexe avec les inhibiteurs. C'est pourquoi, j'ai été amené à faire sécher préalablement une goutte d'inhibiteur afin d'évaporer le DMSO selon une méthodologie proposée par **Gelin *et al***, (262). L'inhibiteur a été ensuite resolubilisé avec la protéine et l'agent précipitant. À noter que si le ligand ne se resolubilise pas, les PEGs de faible poids moléculaire comme par exemple le PEG400 peuvent également être utilisés pour faciliter la resolubilisation.

2. Le trempage

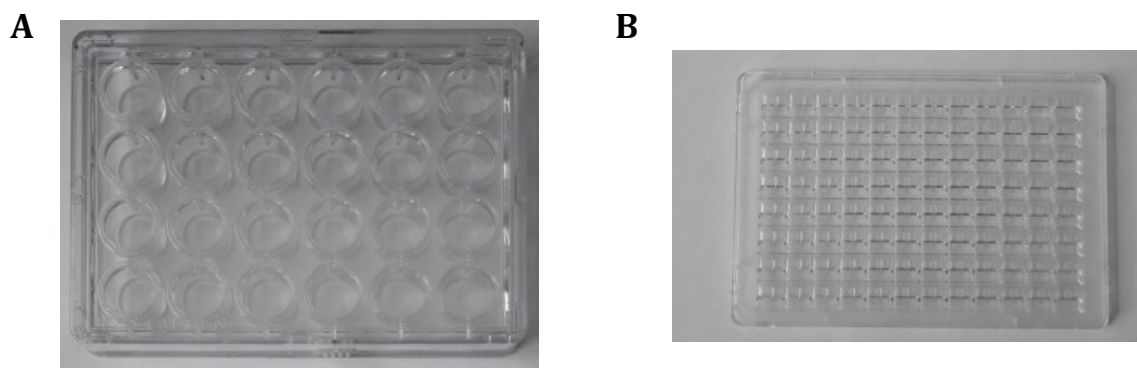
La méthode de trempage utilise quant à elle les propriétés intrinsèques des cristaux de protéines. En effet, les cristaux de protéines sont constitués d'environ 30 à 70% de solvant, en volume. La présence de solvant autorise la protéine à avoir une certaine flexibilité et une dynamique de ses chaînes latérales à l'intérieur du cristal, de cette manière la protéine peut garder son activité biologique. La technique de trempage exploite la présence de ces nombreux canaux de solvant au sein du cristal. Cette méthode nécessite de cristalliser dans un premier temps la protéine seule (sous forme libre), puis de récolter manuellement (sous un microscope) un cristal à l'aide d'une boucle (dual thickness microLoops™ SPINE, Jena Bioscience) afin de le tremper dans une solution contenant une petite molécule inhibitrice. La molécule va pouvoir diffuser à l'intérieur du cristal *via* les canaux de solvant, jusqu'à atteindre le site actif de la protéine. L'état des cristaux est surveillé tout au long du trempage à l'aide d'un microscope (LEICA MZ16 FA, Lieca Microsystems). La concentration du ligand, le nombre et la durée du trempage est fonction de l'affinité du ligand et de la résistance des cristaux à la solution de trempage et à la concentration en DMSO. Ces paramètres sont souvent à optimiser à chaque nouvelle molécule.

3. Co-cristallisation pour une collecte in-situ

Pour la co-cristallisation et les collectes *in-situ*, nous avons collaboré avec le **CBS** (Centre de Biochimie Structurale) de Montpellier et plus particulièrement avec le **Pr. Jean-François Guichou** et le **Dr. Gilles Labesse** (262).

Dans le cadre de cette collaboration, nous avons mis en place des tests de cristallisation en utilisant le robot distributeur « Echo® ». Pour cela nous avons utilisé des plaques de cristallisation de type Crystal Quick™ X (Greiner BioOne) (96 puits avec deux cupules par puits, Jena Bioscience). Cette fois-ci, la co-cristallisation se fait en goutte assise. L'inhibiteur à 10% DMSO final (100 nL complété jusqu'à 1 µL avec du DMSO, afin de recouvrir la surface du puit) est déposé dans un premier temps dans chaque cupule, puis séché dans un incubateur une semaine à 37°C. Lorsque le DMSO est complètement évaporé, la protéine puis l'agent précipitant sont dispensés à l'aide de l'Echo® dans la cupule pour un volume final de 1 µL. Les rapports protéine : solution d'agent précipitant testés sont les suivants, 1:1, 1:2, 2:1, 1:3, 3:1. Le réservoir est, quant à lui, rempli manuellement avec 35 µL d'agent précipitant. La plaque de cristallisation est ensuite scellée par un film (Greiner bio-one) et incubée à la température désirée jusqu'à l'obtention de cristaux. La présente **Figure 69** les deux types de plaque de cristallisation utilisés.

Figure 69 : plaques de cristallisation.



Légende : **A-** Plaque Crystalgen pour les expériences en goutte suspendue. **B-** Plaque Crystal Quick™ X (Greiner BioOne) pour les expériences in situ en gouttes assises.

VI. Avantages et inconvénients de la cristallographie

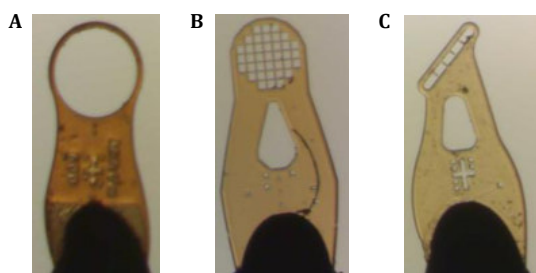
Les avantages : Cette méthode donne accès aux données structurales très précises de l'interaction protéine-ligand.

Les inconvénients : Cette méthode nécessite une grande quantité de protéine pure et monodisperse. La surface d'interaction protéine-ligand doit être accessible (canaux de solvant suffisants pour permettre la diffusion de ligand). La cristallisation est un processus complexe et parfois difficile à mettre au point et de manière reproductible. Par ailleurs la présence de ligand (en co-cristallisation) modifie parfois drastiquement les conditions de cristallisation.

VII. Cryoprotection des cristaux

Les cristaux sont visualisés régulièrement sous un microscope (LEICA MZ16 FA, Lieca Microsystems) et ils sont récoltés à l'aide de boucles présentées dans la **Figure 70** (dual thickness microLoops™ SPINE ou micromeshes™ SPINE pour les gouttes contenant beaucoup de petits cristaux, Jena Bioscience) dès lors qu'ils ont atteint une taille suffisante permettant l'enregistrement des jeux de données de diffraction aux rayons X. Les boucles utilisées sont de tailles variables allant de 50 mm à 600 mm, elles doivent être toujours légèrement plus grandes que la taille du cristal. Nous disposons également de boucle de type « peigne » (microLoopsE™ SPINE, Jena Bioscience), afin de récolter les cristaux en forme de fines baguettes.

Figure 70 : Différents types de boucles utilisées.



Légende : A- Boucle dual thickness microLoops™ SPINE. B- Boucle micromeshes™ SPINE. C- Boucle microLoopsE™ SPINE.

Les cristaux sont tout d'abord cryoprotégés par trempage bref dans la solution de cristallisation additionnée de l'inhibiteur dans le cas des complexes protéine-ligand et d'un agent cryoprotectant qui est dans mon cas soit 10% (v/v) de glycérol ou 15 % (v/v) d'éthylène glycol. Ils sont ensuite rapidement vitrifiés dans de l'azote liquide, ou préalablement refroidis sous flux d'azote gazeux 100K avant d'être conservés dans l'azote liquide. Les données de diffraction aux rayons X des cristaux peuvent ensuite être enregistrées.

VIII. Collecte de jeux de diffractions aux rayons X

1. Collecte classique sous flux d'azote

Un faisceau de rayons X monochromatique est envoyé sur le cristal et les rayons diffractés par celui-ci sont enregistrés à l'aide d'un détecteur, sous forme de clichés de diffraction rayons X présentant des taches d'intensité variable. Une rotation du cristal est effectuée par incrément de 0,1 ou 0,2° selon les cristaux et l'enregistrement se fait pour une rotation complète du cristal (360°) ou parfois sur une rotation plus petite notamment lorsque le cristal présente de nombreuses symétries internes. Les clichés de diffractions sont enregistrés tout au long de la rotation du cristal. La limite de résolution du cristal peut être estimée à partir de ces clichés de diffractions. Plus les taches seront éloignées de la position du faisceau transmis (origine du réseau réciproque), plus la résolution du cristal sera élevée et plus la structure sera détaillée. L'enregistrement des jeux de diffraction aux rayons X des cristaux se fait le plus souvent au synchrotron ESRF de Grenoble ou SOLEIL de Paris. Pour quelques-uns de mes cristaux, les données de diffraction aux rayons X ont été enregistrées sur la plateforme de biologie structurale de l'AFMB en utilisant un générateur de rayons X de type microfocus (Bruker AXS MICROSTAR) équipé de miroirs Osmic et d'un détecteur Mar345 (MarXperts). À titre d'exemple, pour collecter un jeu de données complet pour un cristal appartenant au groupe d'espace $P2_12_12_1$, il faudra 15 heures sur le système anode tournante (MAR345) disponible à l'AFMB, alors qu'il ne faudra que 2 minutes sur une ligne de lumière au synchrotron. Par ailleurs, certaines lignes de lumière des synchrotrons donnent accès à l'enregistrement de données de très haute résolutions (autours de 1 Å) et ce qui permet de pouvoir analyser des détails atomiques très précis de nos interactions protéine/ligand/molécules d'eau.

2. Collecte *In situ* à température ambiante

Contrairement aux collectes de données de diffraction aux rayons X classiques se faisant sous flux d'azote après congélation du cristal, les collectes *in-situ* se font à température ambiante directement dans les plaques de cristallisation et sans cryoprotection. Les cristaux n'étant pas refroidis à 100 K, ils sont dégradés par le faisceau de rayons X. De plus le plastique des plaques diffuse sous l'action du faisceau de rayons X ce qui diminue le rapport signal sur bruit. Par ailleurs, les cristaux étant collectés directement en plaque, l'angle de rotation maximal pour les collectes est de 80° à cause de la présence du réservoir et des puits adjacents. Pour ces raisons, il est nécessaire d'enregistrer les données de diffraction de plusieurs cristaux (même condition de cristallisation) afin d'avoir un jeu de données avec une bonne complétude pour pouvoir résoudre la structure cristallographique. Les collectes *in situ* ont été enregistrées au synchrotron ESRF de Grenoble sur la ligne de lumière BM30 (**FIP** pour **F**rench **b**eamline for **I**nteraction of

Proteins). Les données ont été enregistrées entre les angles de rotation -35° et $+35^\circ$ par rapport à la position verticale de la plaque et pour trois cristaux pour chaque condition de cristallisation.

IX. Traitement des données de diffraction aux rayons X

Lorsqu'un cristal est soumis à un faisceau de rayons X, celui-ci diffracte les rayons. En effet les rayons X induisent le déplacement du nuage électronique des atomes du cristal ce qui provoque une réémission de l'onde électromagnétique, c'est le phénomène de diffusion Rayleigh. L'organisation périodique des molécules, atomes et ions au sein d'un cristal conduit à l'apparition d'interférences constructives (les ondes s'additionnent) dans certaines directions de l'espace et destructives (les ondes s'annulent) dans toutes les autres directions ; c'est le phénomène de diffraction des rayons X. On peut retrouver les directions de ces différentes ondes diffractées grâce à la loi de Bragg. Les ondes diffractées sont caractérisées par une amplitude et une phase. La diffraction suit une fonction mathématique, la « transformée de Fourier ». La transformée de Fourier d'un réseau cristallin conduit à un réseau réciproque dont les dimensions sont inversées. La transformée de Fourier inverse des informations obtenues à partir des clichés de diffraction va donc permettre de retrouver la densité électronique dans le cristal et ainsi la structure. Pour cela il est nécessaire de connaître l'amplitude qui est mesurée à partir des clichés de diffraction et la phase qui est déterminée de manière indirecte.

Pour chaque cristal, une série d'image de diffraction aux rayons X (en moyenne 2000 à 3000 images) est enregistrée. Afin de résoudre la structure cristallographique du cristal, il est nécessaire d'extraire les paramètres d'amplitude et de phase. Le traitement des jeux de données (images) consiste à définir l'intensité et la position de l'ensemble des taches de diffraction RX. Cette phase est divisée en trois étapes successives : **l'indexation**, **l'intégration** et la **mise à l'échelle** de chacune des réflexions. **L'indexation** sert à identifier les taches de diffraction présentes sur l'ensemble des clichés, à déterminer leur coordonnées, et à leurs attribuer les indices de Miller (h, k, l) du nœud du réseau réciproque auquel elles correspondent. Cette étape permet d'émettre l'hypothèse du groupe d'espace et les paramètres de maille (a, b, c et α , β , γ)

Tableau 21. L'intégration sert à déterminer l'intensité de chaque tache de diffraction, pour cela une fraction d'images successives est analysée. Les tâches intenses considérées comme fiables sont utilisées pour estimer l'intensité des tâches plus faibles. **La mise à l'échelle** permet de moyenniser l'intensité de toutes les taches observées pour une même réflexion, ou ses équivalentes cristallographiques. Du fait qu'au cours de l'enregistrement, l'intensité du faisceau de rayons X peut varier et que le cristal peut subir des dommages liés à l'irradiation, les tâches de diffraction n'ont pas toute la même intensité. Il est nécessaire de corriger l'intensité des réflexions pour ces différents paramètres. Après la mise à l'échelle, il est nécessaire de regarder

la valeur du R_{sym} ou R_{merge} , qui correspond au pourcentage moyen de déviation entre les réflexions équivalentes. Le R_{sym} est un indicateur du bon choix du groupe de Laue, il permet de valider ou invalider l'hypothèse de groupe d'espace faite lors de l'indexation. Lorsque le R_{sym} est faible (2 à 15%), cela indique que la symétrie imposée est correcte, tandis que lorsqu'il est élevé (40-50 %), cela signifie que le groupe d'espace choisi peut ne pas être le bon. À l'issue de ces trois étapes, différentes statistiques sont générées, à savoir : le R_{sym} , le nombre total de réflexions, le nombre unique de réflexions, la multiplicité qui correspond au rapport entre le nombre de réflexions totales et uniques, la complétude correspondant à la fraction de données réellement mesurées par rapport à celles théoriquement mesurables et le rapport signal sur bruit ($I/\sigma I$). Ces étapes de traitement des jeux de données ont été réalisées à l'aide du logiciel XDS (version : 1 novembre 2016).

Tableau 21 : système cristallin.

axe de symétrie présent	vecteur parallèle à l'axe (convention)	contraintes sur la maille	nom du système cristallin
aucun		aucune	triclinique
1 axe de rotation de 180° (axe 2)	\vec{b}	$\alpha = \gamma = 90^\circ$	monoclinique
3 axes 2	\vec{a}, \vec{b} et \vec{c}	$\alpha = \beta = \gamma = 90^\circ$	orthorhombique
1 axe de rotation de 90° (axe 4)	\vec{c}	$\alpha = \beta = \gamma = 90^\circ$ $a = b$	quadratique (tétragonal en anglais)
1 axe de rotation de 120° (axe 3)	\vec{c}	$\alpha = \beta = 90^\circ, \gamma = 120^\circ$ $a = b$	trigonal
1 axe de rotation de 60° (axe 6)	\vec{c}	$\alpha = \beta = 90^\circ, \gamma = 120^\circ$ $a = b$	hexagonal
1 axe 4 et des axes 3	axe 4 le long de \vec{c} , axe 3 le long de $\vec{a} + \vec{b} + \vec{c}$ (la diagonale du cube)	$\alpha = \beta = \gamma = 90^\circ$ $a = b = c$	cubique

X. Résolution et affinement des structures cristallographiques

Comme expliqué précédemment, ce sont les paramètres d'amplitude et de phase des ondes diffractées par le cristal qui permettent de réaliser la transformée de Fourier inverse pour pouvoir déterminer la densité électronique. Le paramètre de phase est malheureusement perdu au cours de l'enregistrement des données de diffractions, de ce fait il est nécessaire de le déterminer. Pour cela, le remplacement isomorphe utilisant des atomes lourds peut être utilisé, mais c'est une méthode complexe qui peut être remplacée dans certains cas par une méthode simple et rapide appelée remplacement moléculaire lorsque la structure d'un modèle proche de notre protéine existe. Le phasage par remplacement moléculaire consiste à utiliser la structure d'une protéine homologue ou de la même protéine disponible dans la PDB qui sera utilisée pour calculer les phases à partir de nos données de diffraction. Pour cela nous récupérons les coordonnées de chaque atome de la protéine homologue dans le fichier PDB du modèle choisi pour le remplacement moléculaire et nous éliminons du fichier toutes molécules d'eau et ligand.

Un « cristal virtuel » va ensuite être généré avec les mêmes paramètres de maille que pour notre cristal. Les molécules du modèle vont ensuite être placées comme celles de notre cristal par une série de rotations et puis de translations. À partir des coordonnées spatiales (x, y, z) des atomes présents dans le « cristal virtuel » généré, il est possible de calculer les ondes diffractées à savoir les amplitudes et les phases. Les phases ainsi calculées vont pouvoir être appliquées aux amplitudes mesurées expérimentalement pour notre cristal. Il est alors possible de calculer la carte de densité électronique pour notre cristal. Dans la mesure du possible, la protéine du modèle doit être identique ou très similaire (identité de séquence élevée) à la protéine cristallisée, le modèle doit être si possible une forme libre de la protéine car la présence de ligand peut induire des changements structuraux au sein de la protéine. Trois critères permettent de juger de la qualité du remplacement moléculaire. Le coefficient de corrélation entre les amplitudes mesurées expérimentalement pour notre cristal (F_{obs}) et celles calculées pour le cristal virtuel (F_{calc}) est compris entre 0 et 1 et il doit être le plus élevé possible. Le facteur de désaccord (R) entre les F_{obs} et F_{calc} est également compris entre 0 et 1 et il doit être le plus petit possible. Le dernier critère reste l'observation de la carte de densité électronique qui doit être en accord avec le modèle choisi.

Dans le cadre du projet bromodomaine il est possible de déterminer simplement la phase par remplacement moléculaire puisque des structures sont disponibles dans la PDB pour quasiment tous les modules bromodomains de la famille BET. Les modèles utilisés (PDB ID : 2OSS, 2OUO, 2NXB) ont préalablement été nettoyés avant utilisation, c'est-à-dire que les molécules d'eau, le ligand éventuel et les conformations alternatives des chaînes latérales ont été supprimées du fichier PDB. Le module Phaser MR de la suite CCP4 (version : CCP4 7.0.0.4.2 et CCP4I2 4.8.6) a été utilisé pour le remplacement moléculaire.

La structure doit être ensuite affinée, afin d'optimiser l'accord entre les amplitudes calculées à partir du modèle et celles mesurées expérimentalement pour le cristal. La reconstruction se fait à partir de deux types de carte de densité électronique à savoir la carte de densité $2mF_{obs}-DF_{calc}$ qui permet de visualiser la densité électronique de la protéine cristallisée et la carte de Fourier différence $mF_{obs}-DF_{calc}$, qui permet de visualiser les différences entre le cristal et le modèle, à savoir les électrons manquants et en excès. À cette étape, les chaînes latérales des résidus d'acides aminés de la protéine sont repositionnées correctement dans la densité électronique et les molécules d'eau visibles sont ajoutées. Les modules COOT (version : 0.8.8.EL) et Refmac5 (version : CCP4 7.0.0.4.2 et CCP4I2 4.8.6) de la suite CCP4, ont été utilisés pour l'affinement des structures cristallographiques. Le module COOT permet de corriger manuellement le positionnement des résidus d'acides aminés, de placer les molécules d'eau

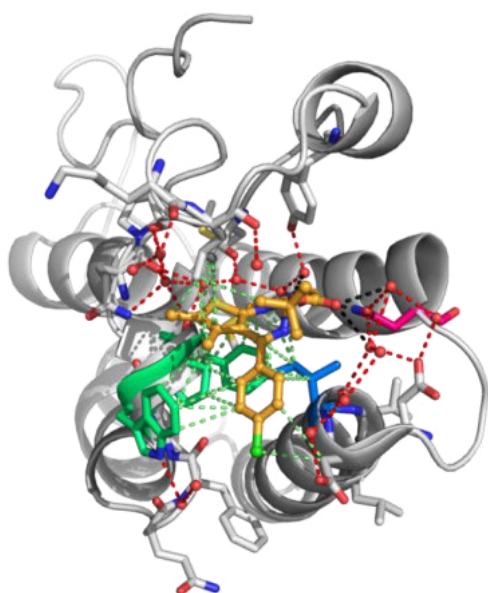
visibles et de placer éventuellement le ligand, le tout dans la densité électronique. Le module Refmac5 permet quant à lui de réaliser à proprement parler l'affinement de la structure.

Après plusieurs cycles d'affinement, le ligand est construit et positionné dans sa densité électronique à l'aide du logiciel COOT et plus particulièrement avec les modules « ligand builder » et « find ligand ». Depuis peu, à partir de l'interface du logiciel COOT il est possible de générer automatiquement une structure tridimensionnelle d'une molécule chimique à partir d'un dessin de sa structure 2D grâce au module PRODRG (version : CCP4 7.0.0.4.2). Ce module génère une structure de la molécule directement utilisable dans les cycles d'affinement CCP4 puisqu'il génère également un fichier de topologie au format cif qui décrit les contraintes propres au ligand (liaisons atomiques, angles de valence, multiplicité des liaisons, chiralité, planéité...).

XI. Analyse des contacts et génération de figures

Une fois que l'affinement de la structure est terminé, nous pouvons visualiser et analyser les interactions entre la protéine, le ligand et les molécules d'eau à l'aide du logiciel Pymol (version : 1.8.0.3). La **Figure 71** présente un exemple d'analyse des contacts entre la protéine BRD4(1) et la molécule inhibitrice JQ1 (PDB : 3MXF).

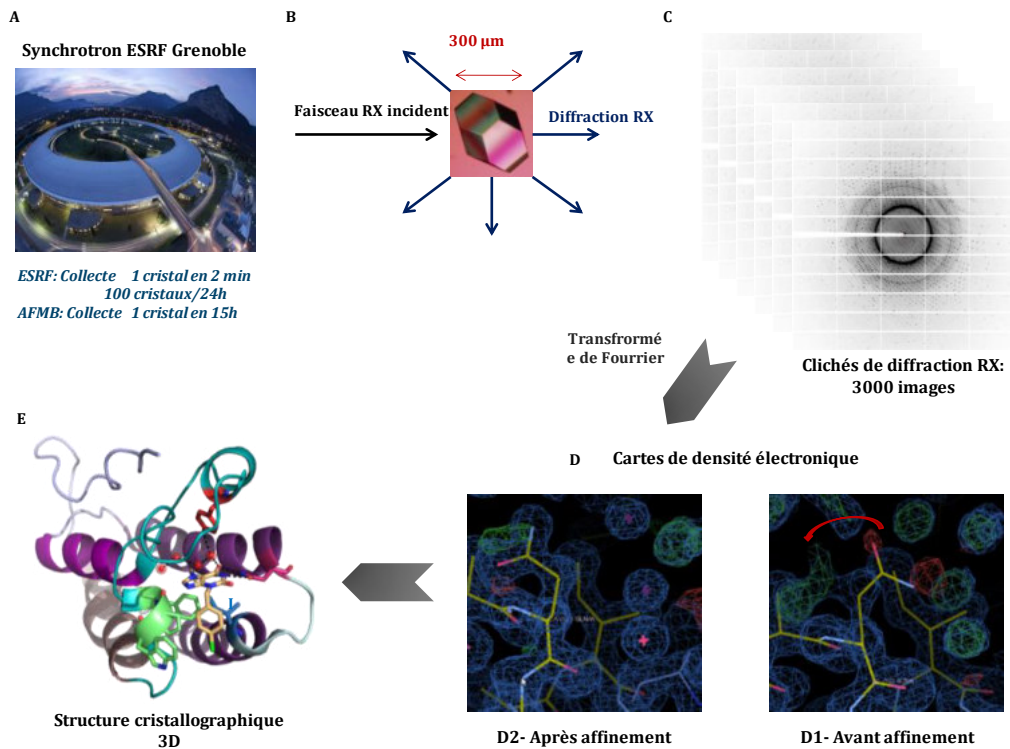
Figure 71 : Exemple d'analyse des contacts BD-inhibiteur.



Légende : Analyse des contacts entre BRD4(1) et son inhibiteur JQ1. Structure du complexe BRD4(1)-JQ1 (PDB : 3MXF). La protéine BRD4(1) est représentée en « cartoon » gris. Les résidus clefs sont représentés en « balls and sticks », en rose l'asparagine N140, en vert le « WPF », en bleu le « gatekeeper » I146 et en gris les autres résidus importants. Les molécules d'eau sont représentées par des sphères rouges. Le ligand JQ1, est quant à lui représenté en « balls and sticks » orange. Les liaisons hydrogènes (rouge et noir, respectivement) et les interactions de type der Walls (vert) sont représentées par des lignes en pointillés.

La **Figure 72** est un récapitulatif des étapes permettant de résoudre la structure cristallographique aux rayons X d'un cristal de protéine.

Figure 72 : De la cristallogénèse à la structure 3D d'une protéine.



Légende : **A-** Synchrotron ESRF de Grenoble où ont été collectés les jeux de diffraction aux rayons X. **B-** Un faisceau de rayon X est envoyé sur le cristal et les faisceaux diffractés résultant sont enregistrés sous forme de clichés de diffraction RX. **C-** En moyenne pour un jeu de données complet, environ 3000 images sont collectées en 2 minutes au synchrotron. **D-** Après avoir subi différents traitements (indexation, intégration et mise à l'échelle par le logiciel XDS) et après avoir réalisé le phasage dans notre cas par remplacement moléculaire en utilisant la suite CCP4, le jeu de données est converti en une carte de densité électronique. La carte 2mFobs-DFcalc permet de visualiser la densité électronique de la protéine cristallisée, tandis que la mFobs-DFcalc qui permet de visualiser les différences entre le cristal et le modèle avec en vert les électrons manquants et en rouge les électrons en excès. **D1-** Montre une glutamine mal positionnée (présence de rouge et de vert avant affinement) et des molécules d'eau (sphères vertes) non affinées. **D2-** Montre le même résidu glutamine correctement positionné dans la densité, ainsi que les molécules d'eau affinées. **E-** Après avoir construit le ligand à l'aide du logiciel COOT et réalisé différents cycles d'affinement avec la suite CCP4, la carte de densité électronique est convertie en un fichier PDB qui est utilisé pour générer la figure de la structure 3D du complexe protéine-ligand à l'aide du logiciel Pymol.

XII. Dépôt des structures dans la PDB

Le dépôt d'une structure dans la PDB est divisé en cinq étapes : (1) la vérification de la séquence protéique, (2) la vérification du ligand, (3) la préparation du fichier pour la déposition, (4) la validation de la structure et (5) le dépôt de la structure. Différents outils listés sur le site de la PDB sont disponibles pour faciliter la réalisation de ces différentes étapes, un tutoriel est également disponible sur le site <https://www.wwpdb.org/deposition/tutorial>. Les structures sont déposées sur le site <https://deposit-1.wwpdb.org/>.

Liste des figures

Figure 1 : Compaction de l'ADN.	35
Figure 2 : Méthylation de l'ADN.	37
Figure 3 : Modifications post-traductionnelles des protéines histones.	38
Figure 4 : Régulateurs épigénétiques : « writers », « erasers » et « readers ».	39
Figure 5 : Régulation de transcription.	45
Figure 6 : Facteurs génétiques et épigénétiques conduisant au cancer.	49
Figure 7 : Caractéristiques d'une cellule cancéreuse et inhibiteurs connus.	50
Figure 8 : Arbre phylogénétique des BDs du protéome humains.	61
Figure 9 : Étude de "druggability" des BDs.	62
Figure 10 : Organisation des BCPs de la famille BET.	66
Figure 11 : Évolution du nombre de structures de BDs dans la PDB.	72
Figure 12 : Architecture des BDs et mode de liaison avec les histones acétylées.	73
Figure 13 : BDs canoniques et non-canoniques.	74
Figure 14 : Représentation de la charge de surface des différents BDs.	76
Figure 15 : Architecture et résidus clefs des BDs de la famille BET.	77
Figure 16 : Comparaison des BDs au sein de la sous-famille BD1.	80
Figure 17 : Comparaison des BDs au sein de la sous-famille BD2.	81
Figure 18 : Comparaison des BD1 et BD2 au sein de la famille BET.	83
Figure 19 : Différents "cœurs chimiques" retrouvés parmi les inhibiteurs de BETs.	86
Figure 20 : Mode de liaison de différents inhibiteurs de BETs.	87
Figure 21 : Exploration des résidus autour de la poche Kac des BETs.	102
Figure 22 : Validation orthogonale par ITC des « touches » identifiées.	113
Figure 23 : Détermination des profils de sélectivité de 2F20 et 3G21.	115
Figure 24 : Études cellulaires pour les composés 2F20 et 3G21.	116
Figure 25 : Structures de BRD4(1) en complexe avec les composés 2F20 et 3G21.	117
Figure 26 : Identification et validation orthogonale de la molécule 5A14.	127
Figure 27 : Structure cristallographique du complexe BRD4(1)-5A14.	128
Figure 28 : Comparaison du mode de fixation de 5A14 avec l'inhibiteur de BAZ2B.	129
Figure 29 : Fragmentation de la molécule 5A14.	130
Figure 30 : Profils de sélectivité des molécules 1915 et 1916.	131
Figure 31 : Structures de BRD4(1)-1915 et BRD4(1)-1916 vs BRD4(1)-5A14.	132
Figure 32 : Profils de sélectivité des molécules 1913 et 1928.	134
Figure 33 : Structure de BRD4(1)-1928 et comparaison avec BRD4(1)-5A14.	135
Figure 34 : Comparaison de la forme libre et complexée de BRD4(1).	136
Figure 35 : Méthode S4MPLE pour l'optimisation de 5A14.	145
Figure 36 : Structure chimique de 5292, un hybride entre 3J7 & 1916.	147
Figure 37 : Identification, Validations orthogonale et cellulaire de 5292.	148

Figure 38 : Comparaison des modes de liaisons des molécules 5292 et 5157.	150
Figure 39 : Comparaison de mode de liaison des molécules 5292, 5318 et 5A14.	153
Figure 40 : Validations orthogonales par ITC et TSA de 5386 et 5392.	156
Figure 41 : Comparaison des modes de liaison de 5292, 5386, 5387, 5392 et 5393.....	158
Figure 42 : Molécule JQ1 bivalente.	159
Figure 43 : Molécules 5A14 bivalentes.	160
Figure 44 : Évaluation des molécules 5A14-PEGn.....	161
Figure 45 : Amélioration de la solubilité de nos molécules.....	170
Figure 46 : Dégradation par le protéasome de BRD4 <i>via</i> l'utilisation du PROTAC.	173
Figure 47 : Principe du clonage de type Gateway.	182
Figure 48 : Principe du clonage de type Assemblage Gibson.	183
Figure 49 : Différents types de chromatographies utilisées.	188
Figure 50 : Principe de la quantification par la méthode de Bradford.	196
Figure 51 : Principe de la quantification par la méthode BCA.	197
Figure 52 : Plateforme robotisée DOSynth.	201
Figure 53 : Plateforme robotisée IPC _{DD}	202
Figure 54 : Principe d'onde acoustique du robot Echo® de Labcyte.....	203
Figure 55 : Homogeneous Time Resolved Fluorescence.	213
Figure 56 : Plans de plaques HTRF manuel.	216
Figure 57 : Protocole « dose response » pour l'HTRF classique.	217
Figure 58 : Protocole « dose response » pour l'HTRF en sélectivité.	218
Figure 59 : Protocole « reformat » pour l'HTRF classique.	219
Figure 60 : Protocole « reformat » pour l'HTRF en sélectivité.	220
Figure 61 : Principe du Thermal Shift Assay.	226
Figure 62 : Exemple de résultats obtenus.	227
Figure 63 : Protocole « reformat » pour le TSA.	229
Figure 64 : Protocole TSA Bromoscan.	231
Figure 65 : Principe de l'ITC.	236
Figure 66: Principe du FRAP.	249
Figure 67 : Cristallogenèse.	255
Figure 68 : Optimisation des conditions de cristallisation.	257
Figure 69 : plaques de cristallisation.....	259
Figure 70 : Différents types de boucles utilisées.....	260
Figure 71 : Exemple d'analyse des contacts BD-inhibiteur.	265
Figure 72 : De la cristallogenèse à la structure 3D d'une protéine.	266

Liste des tableaux

Tableau 1 : Régulateurs épigénétiques.....	40
Tableau 2 : Expression tissulaire des différents BCPs.....	58
Tableau 3 : Fonction biologique et localisation subcellulaire des BCPs.....	59
Tableau 4 : Maladies associées aux BCPs.....	60
Tableau 5 : Liste des PPI BDs-peptides histones acétylés connues pour les BETs.....	68
Tableau 6 : Identité de séquence au sein de la famille BET.....	70
Tableau 7 : Alignement du motif « WPF » et du « gatekeeper ».....	78
Tableau 8 : Inhibiteurs de BETs en essais précliniques.....	96
Tableau 9 : Inhibiteurs de BETs en essais cliniques.....	97
Tableau 10 : Comparaisons des résidus autour de la poche Kac des BETs.....	103
Tableau 11 : Criblage de la chimiothèque 2P2I3D.....	112
Tableau 12 : Récapitulatifs des résultats pour les sept molécules validées.....	114
Tableau 13 : Évaluation des analogues « xanthine » par HTRF et ITC.....	131
Tableau 14 : Évaluation des analogues « triazolo-pyrimidine » par HTRF et ITC.....	133
Tableau 15 : Valeurs d'IC ₅₀ mesurées par HTRF pour les analogues de 5292.....	151
Tableau 16 : Profils de sélectivités des analogues de 5318 et 5316.....	152
Tableau 17 : Évaluation HTRF & viabilité cellulaire des neuf dérivés sulfonamides.....	155
Tableau 18 : Tableau récapitulatif pour les trois sondes épigénétiques identifiées.....	169
Tableau 19 : Comparatif de trois méthodes de quantifications.....	197
Tableau 20 : Concentrations finales pour les tests d'HTRF.....	215
Tableau 21 : système cristallin.....	263

Annexes

Annexe 1 : Plasmides utilisés pour l'expression de protéines recombinantes

Clonage		
Gènes "Bromodomaine"	Plasmide pour l'HTRF	Plasmide pour ITC, TSA, Cristallographie
BRD4(BD1)	GST-TEV-pDEST15	His-TEV-pNIC BSA4 (S. Knapp)
BRD4(BD2)	GST-TEV-pDEST15	GST-TEV-pet42
BRD3(BD1)	GST-TEV-pet42-Thr-His	-
BRD3(BD2)	-	His-TEV-pNIC BSA4 (S. Knapp)
BRD2(BD1)	GST-TEV-pet42-Thr-His	-
BRD2(BD2)	GST-TEV-pet42	GST-TEV-pet42
BRDT(BD1)	-	-
BRDT(BD2)	GST-TEV-pet42-Thr-His	-
ATAD2A	GST-TEV-pet42-Thr-His	HIS-TEV-pDEST17
CREBBP	GST-TEV-pDEST15	GST-TEV-pDEST15

Annexe 2 : Purification de protéines recombinantes

Lyse bactérienne		
Tampon de "Base"	GST-protéines	6-His-protéines
1 mg/mL de Lysozyme	"Base" additionnée de	"Base" additionnée de
Une pastille de Complete pour 50 mL	20 mM Tris pH 8,0	50 mM Hepes pH 7,5
10 µg/mL de DNase	150mM NaCl	500mM NaCl
5% de Glycérol		5 mM Imidazole
2 mM d'EDTA		
2 mM de DTT		

Purification des Bromodomains		
Purification des GST-protéines		
Lavage GST	Elution GST	Dessalage & Gel Filtration
20 mM Tris pH 8,0	20 mM Tris pH 8,0	50 mM Hepes pH 7,5
500 mM NaCl	150 mM NaCl	150 mM NaCl
	10 mM Glutathione	
Purification des 6-His-protéines		
Lavage His	Elution His	Dessalage & Gel Filtration
50 mM Hepes pH 7,5	50 mM Hepes pH 7,5	10 mM Hepes pH 7,5
500mM NaCl	500mM NaCl	150 mM NaCl
5 mM Imidazole	500 mM Imidazole	

Annexe 3 : Conditions de cristallisation

BRD4(1)		
BRD4(1) 25 mg/mL 200 mM AmSO4 100 mM Tris pH 8.5 25 % PEG 3350 100 mM NaCl	BRD4(1) 20 mg/mL 200 mM AmSO4 100 mM Tris pH 8.5 19 % PEG 3350 100 mM NaCl	BRD4(1) 12 mg/mL 300 mM NaNO3 100 mM NaCl 22 % PEG 3350 10 % EG
BRD4(1) 12 mg/mL 300 mM NaNO3 16 % PEG 3350 5 % EG	BRD4(1) 12 mg/mL 300 mM NaNO3 18 % PEG 3350	BRD4(1) 12 mg/mL 300 mM Na Formate 100 mM NaCl 22 % PEG 3350 10 % EG

BRD4(2)		
BRD4(2) 20/15/12 mg/mL 30% PEG 1000 100 mM MIB pH 7,0	BRD4(2) 20/15/12 mg/mL 2,5 M (NH4)2SO4 100 mM tris pH 8,5	BRD4(2) 16/12/18 mg/mL 50% PEG 300 100 mM Imidazole
BRD4(2) 20/15/12 mg/mL 20% PEG 6000 100 mM Citrate pH 5,6	BRD4(2) 20/15/12 mg/mL 20% Peg 3350 10% EG 200 mM Na - K Tartrate	BRD4(2) 20/15/12 mg/mL 20% PEG3350 200 mM Na Formate 100 mM Bis-Tris propane

BRD3(2)		
BRD3(2) 20 mg/mL 200 mM imidazole malate pH 22.5 % PEG 10K	BRD3(2) 12 mg/mL 200 mM sulfate d'ammonium 100 mM BTP pH 5.5 25 % PEG 3350	BRD3(2) 20 mg/mL 200 mM sodium chloride 100 mM HEPES pH 7.0 20 % PEG 6K

BRD2(2)		
BRD2(2) 20/15/10/5 mg/ml 20% PEG 6000 100 mM Citrate pH 5,6	BRD2(2) 20/15/10/5 mg/ml 20% PEG 3350 10% EG 200 mM Na K Tartrate	BRD2(2) 20/15/10/5 mg/mL 20% PEG 3350 200 mM Na Formate 100 mM Bis-Tris Propane
BRD2(2) 20/15/10/5 mg/ml 25% PEG 3350 200 mM KCN thiocyanate	BRD2(2) 20/15/10/5 mg/ml 2,6 M Na Malonate	BRD2(2) 20/15/10/5 mg/mL 33% PEG 3350 83 mM Na Acetate 167 mM AmSO4

Annexe 4 : Tables des statistiques pour les structures cristallographiques

A	Criblage 2P2I3D		
	2F20	3G21	5A14
	Resolution range	39.22 - 1.7 (1.761 - 1.7)	33.89 - 1.901 (1.969 - 1.901)
Space group	P 21 21 21	P 21 21 21	P 21 21 21
Unit cell	37.263 44.236 78.436 90 90 90	41.723 48.509 58.094 90 90 90	69.686 83.612 107.391 90 90 90
Unique reflections	14321 (1372)	9663 (918)	31316 (2976)
Completeness (%)	96.55 (96.48)	99.30 (96.73)	97.85 (94.87)
Wilson B-factor	10.02	33.29	22.62
Reflections used in refinement	14313 (1371)	9663 (918)	31315 (2977)
Reflections used for R-free	715 (68)	484 (46)	1565 (148)
R-work	0.2523 (0.2786)	0.1429 (0.1876)	0.1836 (0.2080)
R-free	0.3282 (0.3959)	0.2121 (0.3501)	0.2309 (0.2806)
Number of non-hydrogen atoms	1133	1115	4355
macromolecules	1062	1043	3952
ligands	34	35	140
solvent	37	37	263
Protein residues	127	125	473
RMS(bonds)	0.028	0.028	0.019
RMS(angles)	2.33	2.31	1.83
Ramachandran favored (%)	98.40	98.37	97.82
Ramachandran allowed (%)	1.60	1.63	2.18
Ramachandran outliers (%)	0.00	0.00	0.00
Rotamer outliers (%)	1.67	0.85	1.81
Clashscore	3.25	0.94	2.24
Average B-factor	12.91	42.51	25.08
macromolecules	12.32	42.22	25.12
ligands	29.62	52.31	22.06
solvent	14.48	41.40	26.02

B	SAR 5A14		
	1915	1916	1928
	Resolution range	38.67 - 1.3 (1.346 - 1.3)	38.83 - 1.6 (1.657 - 1.6)
Space group	P 21 21 21	P 21 21 21	P 21 21 21
Unit cell	34.437 47.059 77.347 90 90 90	35.935 46.545 77.658 90 90 90	37.353 44.069 78.272 90 90 90
Unique reflections	30646 (2309)	17747 (1740)	60685 (5874)
Completeness (%)	96.68 (74.01)	99.51 (99.15)	99.52 (98.10)
Wilson B-factor	8.42	19.66	8.63
Reflections used in refinement	30645 (2310)	17741 (1740)	60684 (5875)
Reflections used for R-free	1533 (116)	888 (87)	3035 (294)
R-work	0.1726 (0.1967)	0.1889 (0.3296)	0.1778 (0.2580)
R-free	0.1954 (0.2248)	0.2265 (0.3500)	0.1990 (0.2605)
Number of non-hydrogen atoms	1282	1204	1312
macromolecules	1079	1071	1115
ligands	17	29	33
solvent	186	104	164
Protein residues	127	127	127
RMS(bonds)	0.028	0.019	0.030
RMS(angles)	2.41	1.86	2.40
Ramachandran favored (%)	99.20	96.80	99.20
Ramachandran allowed (%)	0.80	3.20	0.80
Ramachandran outliers (%)	0.00	0.00	0.00
Rotamer outliers (%)	0.82	2.48	0.79
Clashscore	3.23	1.38	10.54
Average B-factor	11.20	27.62	11.75
macromolecules	9.72	26.86	10.51
ligands	6.94	25.66	10.82
solvent	20.15	36.02	20.35

Légende : Table générée avec le logiciel Phenix (version : 1.11.1-2575)

Annexe 4 : Tables des statistiques pour les structures cristallographiques (suite)

C	SAR 5292		
	5292	5157	5318
Resolution range	36.87 - 0.95 (0.984 - 0.95)	35.41 - 0.911 (0.9438 - 0.911)	35,52 - 1,092 (1,131 - 1,092)
Space group	P 21 21 21	P 21 21 21	P 21 21 21
Unit cell	44.284 47.498 58.468 90 90 90	44.524 47.509 58.425 90 90 90	44,415 46,924 59,147 90 90 90
Unique reflections	78063 (7498)	84919 (6408)	51906 (5111)
Completeness (%)	99.61 (96.26)	95.17 (71.78)	99,75 (98,44)
Wilson B-factor	9.83	7.79	11,75
Reflections used in refinement	78000 (7444)	84709 (6297)	51811 (5034)
Reflections used for R-free	3904 (375)	4235 (314)	2590 (251)
R-work	0.2343 (0.3879)	0.2998 (0.5230)	0,1963 (0,5373)
R-free	0.2516 (0.3751)	0.3249 (0.5008)	0,2274 (0,5819)
Number of non-hydrogen atoms	1160	1185	1212
macromolecules	1062	1062	1062
ligands	4	35	58
solvent	94	88	92
Protein residues	127	127	127
RMS(bonds)	0.026	0.021	0,899
RMS(angles)	2.36	2.26	11,97
Ramachandran favored (%)	98.40	100.00	99,2
Ramachandran allowed (%)	1.60	0.00	0,8
Ramachandran outliers (%)	0.00	0.00	0
Rotamer outliers (%)	2.50	3.36	1,67
Clashscore	5.17	6.50	2,28
Average B-factor	13.66	12.13	14,95
macromolecules	13.22	11.57	14,72
ligands	18.92	10.25	12,32
solvent	18.40	19.71	19,33

D	SAR 5292		
	5386	5387	5393
Resolution range	35.12 - 1.401 (1.451 - 1.401)	35.21 - 1.193 (1.236 - 1.193)	32.44 - 1.194 (1.236 - 1.194)
Space group	P 21 21 21	P 21 21 21	P 21 21 21
Unit cell	44.137 47.255 57.981 90 90 90	44.3287 47.7411 57.9355 90 90 90	44.3149 47.6315 57.5301 90 90 90
Unique reflections	23509 (2030)	29841 (496)	30919 (522)
Completeness (%)	95.98 (84.72)	74.20 (12.30)	77.20 (12.77)
Wilson B-factor	9.82	8.52	7.82
Reflections used in refinement	23508 (2030)	29535 (482)	30425 (495)
Reflections used for R-free	1176 (102)	1506 (27)	1443 (30)
R-work	0.1687 (0.2090)	0.1812 (0.2860)	0.1772 (0.3156)
R-free	0.2035 (0.2392)	0.1973 (0.2192)	0.1946 (0.2306)
Number of non-hydrogen atoms	1241	1241	1255
macromolecules	1062	1062	1062
ligands	42	37	41
solvent	137	142	152
Protein residues	127	127	127
RMS(bonds)	0.024	0.027	0,028
RMS(angles)	2.18	2.30	2.42
Ramachandran favored (%)	100.00	99.20	100.00
Ramachandran allowed (%)	0.00	0.80	0.00
Ramachandran outliers (%)	0.00	0.00	0.00
Rotamer outliers (%)	0.00	1.67	0.00
Clashscore	0.92	2.78	1.84
Average B-factor	12.04	12.03	10,73
macromolecules	11.35	11.13	9,70
ligands	8.50	7.80	8,06
solvent	18.50	19.86	18,63

Légende : Table générée avec le logiciel Phenix (version : 1.11.1-2575)

Références bibliographiques

1. E. Bianconi *et al.*, An estimation of the number of cells in the human body. *Ann Hum Biol* **40**, 463-471 (2013).
2. S. Choudhuri, From Waddington's epigenetic landscape to small noncoding RNA: some important milestones in the history of epigenetics research. *Toxicol Mech Methods* **21**, 252-274 (2011).
3. A. D. Goldberg, C. D. Allis, E. Bernstein, Epigenetics: a landscape takes shape. *Cell* **128**, 635-638 (2007).
4. S. L. Berger, T. Kouzarides, R. Shiekhattar, A. Shilatifard, An operational definition of epigenetics. *Genes Dev* **23**, 781-783 (2009).
5. J. D. WATSON, F. H. CRICK, The structure of DNA. *Cold Spring Harb Symp Quant Biol* **18**, 123-131 (1953).
6. J. J. Champoux, DNA topoisomerases: structure, function, and mechanism. *Annu Rev Biochem* **70**, 369-413 (2001).
7. C. Alabert, A. Groth, Chromatin replication and epigenome maintenance. *Nat Rev Mol Cell Biol* **13**, 153-167 (2012).
8. K. Luger, A. W. Mäder, R. K. Richmond, D. F. Sargent, T. J. Richmond, Crystal structure of the nucleosome core particle at 2.8 Å resolution. *Nature* **389**, 251-260 (1997).
9. T. J. Richmond, J. T. Finch, B. Rushton, D. Rhodes, A. Klug, Structure of the nucleosome core particle at 7 Å resolution. *Nature* **311**, 532-537 (1984).
10. R. D. Kornberg, Chromatin structure: a repeating unit of histones and DNA. *Science* **184**, 868-871 (1974).
11. R. D. Kornberg, J. O. Thomas, Chromatin structure; oligomers of the histones. *Science* **184**, 865-868 (1974).
12. R. D. Kornberg, Structure of chromatin. *Annu Rev Biochem* **46**, 931-954 (1977).
13. S. Khochbin, Histone H1 diversity: bridging regulatory signals to linker histone function. *Gene* **271**, 1-12 (2001).
14. M. Vignali, J. L. Workman, Location and function of linker histones. *Nat Struct Biol* **5**, 1025-1028 (1998).
15. S. Khorasanizadeh, The nucleosome: from genomic organization to genomic regulation. *Cell* **116**, 259-272 (2004).
16. G. Felsenfeld, M. Groudine, Controlling the double helix. *Nature* **421**, 448-453 (2003).

17. V. Marx, Epigenetics: Reading the second genomic code. *Nature* **491**, 143-147 (2012).
18. A. S. Quina, M. Buschbeck, L. Di Croce, Chromatin structure and epigenetics. *Biochem Pharmacol* **72**, 1563-1569 (2006).
19. L. D. Moore, T. Le, G. Fan, DNA methylation and its basic function. *Neuropsychopharmacology* **38**, 23-38 (2013).
20. T. H. Bestor, The DNA methyltransferases of mammals. *Hum Mol Genet* **9**, 2395-2402 (2000).
21. M. Okano, D. W. Bell, D. A. Haber, E. Li, DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. *Cell* **99**, 247-257 (1999).
22. W. Reik, W. Dean, DNA methylation and mammalian epigenetics. *Electrophoresis* **22**, 2838-2843 (2001).
23. A. M. Deaton, A. Bird, CpG islands and the regulation of transcription. *Genes Dev* **25**, 1010-1022 (2011).
24. M. Weber, D. Schübeler, Genomic patterns of DNA methylation: targets and function of an epigenetic mark. *Curr Opin Cell Biol* **19**, 273-280 (2007).
25. M. Hirst, M. A. Marra, Epigenetics and human disease. *Int J Biochem Cell Biol* **41**, 136-146 (2009).
26. M. A. Dawson, T. Kouzarides, Cancer epigenetics: from mechanism to therapy. *Cell* **150**, 12-27 (2012).
27. P. V. Hornbeck *et al.*, PhosphoSitePlus, 2014: mutations, PTMs and recalibrations. *Nucleic Acids Res* **43**, D512-520 (2015).
28. T. Bonaldi, A. Imhof, J. T. Regula, A combination of different mass spectroscopic techniques for the analysis of dynamic changes of histone modifications. *Proteomics* **4**, 1382-1396 (2004).
29. T. Bonaldi, J. T. Regula, A. Imhof, The use of mass spectrometry for the analysis of histone modifications. *Methods Enzymol* **377**, 111-130 (2004).
30. K. R. Karch, S. Sidoli, B. A. Garcia, Identification and Quantification of Histone PTMs Using High-Resolution Mass Spectrometry. *Methods Enzymol* **574**, 3-29 (2016).
31. T. Kouzarides, Chromatin modifications and their function. *Cell* **128**, 693-705 (2007).
32. M. A. Christophorou *et al.*, Citrullination regulates pluripotency and histone H1 binding to chromatin. *Nature* **507**, 104-108 (2014).
33. B. M. Turner, Reading signals on the nucleosome with a new nomenclature for modified histones. *Nat Struct Mol Biol* **12**, 110-112 (2005).
34. C. D. Allis *et al.*, New nomenclature for chromatin-modifying enzymes. *Cell* **131**, 633-636 (2007).

35. C. Sawan, Z. Herceg, Histone modifications and cancer. *Adv Genet* **70**, 57-85 (2010).
36. B. D. Strahl, C. D. Allis, The language of covalent histone modifications. *Nature* **403**, 41-45 (2000).
37. N. Lacoste, J. Côté, [The epigenetic code of histones]. *Med Sci (Paris)* **19**, 955-959 (2003).
38. T. Jenuwein, C. D. Allis, Translating the histone code. *Science* **293**, 1074-1080 (2001).
39. R. Schneider, R. Grosschedl, Dynamics and interplay of nuclear architecture, genome organization, and gene expression. *Genes Dev* **21**, 3027-3043 (2007).
40. B. E. Bernstein, A. Meissner, E. S. Lander, The mammalian epigenome. *Cell* **128**, 669-681 (2007).
41. C. H. Arrowsmith, C. Bountra, P. V. Fish, K. Lee, M. Schapira, Epigenetic protein families: a new frontier for drug discovery. *Nat Rev Drug Discov* **11**, 384-400 (2012).
42. V. G. ALLFREY, R. FAULKNER, A. E. MIRSKY, ACETYLATION AND METHYLATION OF HISTONES AND THEIR POSSIBLE ROLE IN THE REGULATION OF RNA SYNTHESIS. *Proc Natl Acad Sci U S A* **51**, 786-794 (1964).
43. T. Kouzarides, Acetylation: a regulatory modification to rival phosphorylation? *EMBO J* **19**, 1176-1179 (2000).
44. B. M. Turner, Histone acetylation and an epigenetic code. *Bioessays* **22**, 836-845 (2000).
45. F. Gong, K. M. Miller, Mammalian DNA repair: HATs and HDACs make their mark through histone acetylation. *Mutat Res* **750**, 23-30 (2013).
46. A. Ganesan, L. Nolan, S. J. Crabb, G. Packham, Epigenetic therapy: histone acetylation, DNA methylation and anti-cancer drug discovery. *Curr Cancer Drug Targets* **9**, 963-981 (2009).
47. A. Csordas, On the biological role of histone acetylation. *Biochem J* **265**, 23-38 (1990).
48. G. Zhang, S. G. Smith, M. M. Zhou, Discovery of Chemical Inhibitors of Human Bromodomains. *Chem Rev* **115**, 11625-11668 (2015).
49. V. G. Allfrey, A. E. Mirsky, Structural Modifications of Histones and their Possible Role in the Regulation of RNA Synthesis. *Science* **144**, 559 (1964).
50. K. Marushige, Activation of chromatin by acetylation of histone side chains. *Proc Natl Acad Sci U S A* **73**, 3937-3941 (1976).
51. S. Peleg *et al.*, Altered histone acetylation is associated with age-dependent memory impairment in mice. *Science* **328**, 753-756 (2010).
52. M. Shogren-Knaak *et al.*, Histone H4-K16 acetylation controls chromatin structure and protein interactions. *Science* **311**, 844-847 (2006).

53. R. Murr *et al.*, Histone acetylation by Trrap-Tip60 modulates loading of repair proteins and repair of DNA double-strand breaks. *Nat Cell Biol* **8**, 91-99 (2006).
54. Y. Wang *et al.*, Histone H3 lysine 14 acetylation is required for activation of a DNA damage checkpoint in fission yeast. *J Biol Chem* **287**, 4386-4393 (2012).
55. M. D. Shahbazian, M. Grunstein, Functions of site-specific histone acetylation and deacetylation. *Annu Rev Biochem* **76**, 75-100 (2007).
56. S. L. Schreiber, B. E. Bernstein, Signaling network model of chromatin. *Cell* **111**, 771-778 (2002).
57. P. Filippakopoulos *et al.*, Selective inhibition of BET bromodomains. *Nature* **468**, 1067-1073 (2010).
58. A. H. Peters, D. Schübeler, Methylation of histones: playing memory with DNA. *Curr Opin Cell Biol* **17**, 230-238 (2005).
59. R. J. Klose, Y. Zhang, Regulation of histone methylation by demethylimination and demethylation. *Nat Rev Mol Cell Biol* **8**, 307-318 (2007).
60. U. M. Bauer, S. Daujat, S. J. Nielsen, K. Nightingale, T. Kouzarides, Methylation at arginine 17 of histone H3 is linked to gene activation. *EMBO Rep* **3**, 39-44 (2002).
61. R. Cao *et al.*, Role of histone H3 lysine 27 methylation in Polycomb-group silencing. *Science* **298**, 1039-1043 (2002).
62. A. Di Lorenzo, M. T. Bedford, Histone arginine methylation. *FEBS Lett* **585**, 2024-2031 (2011).
63. V. Migliori *et al.*, Symmetric dimethylation of H3R2 is a newly identified histone mark that supports euchromatin maintenance. *Nat Struct Mol Biol* **19**, 136-144 (2012).
64. R. Marmorstein, R. C. Trievel, Histone modifying enzymes: structures, mechanisms, and specificities. *Biochim Biophys Acta* **1789**, 58-68 (2009).
65. K. K. Li, C. Luo, D. Wang, H. Jiang, Y. G. Zheng, Chemical and biochemical approaches in the study of histone methylation and demethylation. *Med Res Rev* **32**, 815-867 (2012).
66. E. L. Greer, Y. Shi, Histone methylation: a dynamic mark in health, disease and inheritance. *Nat Rev Genet* **13**, 343-357 (2012).
67. M. Yun, J. Wu, J. L. Workman, B. Li, Readers of histone modifications. *Cell Res* **21**, 564-578 (2011).
68. R. Loury, P. Sassone-Corsi, Histone phosphorylation: how to proceed. *Methods* **31**, 40-48 (2003).
69. C. Prigent, S. Dimitrov, Phosphorylation of serine 10 in histone H3, what for? *J Cell Sci* **116**, 3677-3685 (2003).
70. D. Rossetto, N. Avvakumov, J. Côté, Histone phosphorylation: a chromatin modification involved in diverse nuclear events. *Epigenetics* **7**, 1098-1108 (2012).

71. S. H. Lecker, A. L. Goldberg, W. E. Mitch, Protein degradation by the ubiquitin-proteasome pathway in normal and disease states. *J Am Soc Nephrol* **17**, 1807-1819 (2006).
72. J. Cao, Q. Yan, Histone ubiquitination and deubiquitination in transcription, DNA damage response, and cancer. *Front Oncol* **2**, 26 (2012).
73. H. He, N. Lehming, Global effects of histone modifications. *Brief Funct Genomic Proteomic* **2**, 234-243 (2003).
74. A. Shilatifard, Chromatin modifications by methylation and ubiquitination: implications in the regulation of gene expression. *Annu Rev Biochem* **75**, 243-269 (2006).
75. J. Gräff, I. M. Mansuy, Epigenetic codes in cognition and behaviour. *Behav Brain Res* **192**, 70-87 (2008).
76. M. Tan *et al.*, Identification of 67 histone marks and histone lysine crotonylation as a new type of histone modification. *Cell* **146**, 1016-1028 (2011).
77. K. Kobza *et al.*, K4, K9 and K18 in human histone H3 are targets for biotinylation by biotinidase. *FEBS J* **272**, 4249-4259 (2005).
78. Y. Wang *et al.*, Histone hypercitrullination mediates chromatin decondensation and neutrophil extracellular trap formation. *J Cell Biol* **184**, 205-213 (2009).
79. A. Dhall *et al.*, Sumoylated human histone H4 prevents chromatin compaction by inhibiting long-range internucleosomal interactions. *J Biol Chem* **289**, 33827-33837 (2014).
80. A. C. Mirabella, B. M. Foster, T. Bartke, Chromatin deregulation in disease. *Chromosoma* **125**, 75-93 (2016).
81. K. Jones-Mason, I. E. Allen, N. Bush, S. Hamilton, Epigenetic marks as the link between environment and development: examination of the associations between attachment, socioeconomic status, and methylation of the SLC6A4 gene. *Brain Behav* **6**, e00480 (2016).
82. S. B. Baylin, P. A. Jones, A decade of exploring the cancer epigenome - biological and translational implications. *Nat Rev Cancer* **11**, 726-734 (2011).
83. G. V. Denis, Bromodomain coactivators in cancer, obesity, type 2 diabetes, and inflammation. *Discov Med* **10**, 489-499 (2010).
84. P. Anand *et al.*, Cancer is a preventable disease that requires major lifestyle changes. *Pharm Res* **25**, 2097-2116 (2008).
85. S. Virani, J. A. Colacino, J. H. Kim, L. S. Rozek, Cancer epigenetics: a brief review. *ILAR J* **53**, 359-369 (2012).
86. D. Hanahan, R. A. Weinberg, The hallmarks of cancer. *Cell* **100**, 57-70 (2000).
87. D. Hanahan, R. A. Weinberg, Hallmarks of cancer: the next generation. *Cell* **144**, 646-674 (2011).

88. G. Bergers, L. E. Benjamin, Tumorigenesis and the angiogenic switch. *Nat Rev Cancer* **3**, 401-410 (2003).
89. C. Holohan, S. Van Schaeybroeck, D. B. Longley, P. G. Johnston, Cancer drug resistance: an evolving paradigm. *Nat Rev Cancer* **13**, 714-726 (2013).
90. G. Housman *et al.*, Drug resistance in cancer: an overview. *Cancers (Basel)* **6**, 1769-1792 (2014).
91. S. Wee *et al.*, Targeting epigenetic regulators for cancer therapy. *Ann N Y Acad Sci* **1309**, 30-36 (2014).
92. P. W. Manley *et al.*, Imatinib: a selective tyrosine kinase inhibitor. *Eur J Cancer* **38 Suppl 5**, S19-27 (2002).
93. J. C. Hernández-Boluda, F. Cervantes, Imatinib mesylate (Gleevec, Glivec): a new therapy for chronic myeloid leukemia and other malignancies. *Drugs Today (Barc)* **38**, 601-613 (2002).
94. A. Lau, K. Seiter, Second-line therapy for patients with chronic myeloid leukemia resistant to first-line imatinib. *Clin Lymphoma Myeloma Leuk* **14**, 186-196 (2014).
95. J. K. Christman, 5-Azacytidine and 5-aza-2'-deoxycytidine as inhibitors of DNA methylation: mechanistic studies and their implications for cancer therapy. *Oncogene* **21**, 5483-5495 (2002).
96. K. Raj, G. J. Mufti, Azacytidine (Vidaza(R)) in the treatment of myelodysplastic syndromes. *Ther Clin Risk Manag* **2**, 377-388 (2006).
97. A. K. Bubna, Vorinostat-An Overview. *Indian J Dermatol* **60**, 419 (2015).
98. M. Duvic, J. Vu, Vorinostat in cutaneous T-cell lymphoma. *Drugs Today (Barc)* **43**, 585-599 (2007).
99. S. R. Haynes *et al.*, The bromodomain: a conserved sequence found in human, Drosophila and yeast proteins. *Nucleic Acids Res* **20**, 2603 (1992).
100. S. Hunter *et al.*, InterPro: the integrative protein signature database. *Nucleic Acids Res* **37**, D211-215 (2009).
101. T. Fujisawa, P. Filippakopoulos, Functions of bromodomain-containing proteins and their roles in homeostasis and cancer. *Nat Rev Mol Cell Biol* **18**, 246-262 (2017).
102. P. Filippakopoulos, S. Knapp, The bromodomain interaction module. *FEBS Lett* **586**, 2692-2704 (2012).
103. P. Filippakopoulos *et al.*, Histone recognition and large-scale structural analysis of the human bromodomain family. *Cell* **149**, 214-231 (2012).
104. S. Muller, P. Filippakopoulos, S. Knapp, Bromodomains as therapeutic targets. *Expert Rev Mol Med* **13**, e29 (2011).
105. D. Boehm *et al.*, BET bromodomain-targeting compounds reactivate HIV from latency via a Tat-independent mechanism. *Cell Cycle* **12**, 452-462 (2013).

106. D. Boehm, R. J. Conrad, M. Ott, Bromodomain proteins in HIV infection. *Viruses* **5**, 1571-1586 (2013).
107. U. Egner, R. C. Hillig, A structural biology view of target drugability. *Expert Opin Drug Discov* **3**, 391-401 (2008).
108. X. U. Dai *et al.*, Predicting the druggability of protein-protein interactions based on sequence and structure features of active pockets. *Curr Pharm Des* **21**, 3051-3061 (2015).
109. L. R. Vidler, N. Brown, S. Knapp, S. Hoelder, Druggability analysis and structural classification of bromodomain acetyl-lysine binding sites. *J Med Chem* **55**, 7346-7359 (2012).
110. O. Fedorov *et al.*, Selective targeting of the BRG/PB1 bromodomains impairs embryonic and trophoblast stem cell maintenance. *Sci Adv* **1**, e1500723 (2015).
111. B. Vangamudi *et al.*, The SMARCA2/4 ATPase Domain Surpasses the Bromodomain as a Drug Target in SWI/SNF-Mutant Cancers: Insights from cDNA Rescue and PFI-3 Inhibitor Studies. *Cancer Res* **75**, 3865-3878 (2015).
112. P. Filippakopoulos, S. Knapp, Targeting bromodomains: epigenetic readers of lysine acetylation. *Nat Rev Drug Discov* **13**, 337-356 (2014).
113. G. A. Josling, S. A. Selvarajah, M. Petter, M. F. Duffy, The role of bromodomain proteins in regulating gene expression. *Genes (Basel)* **3**, 320-343 (2012).
114. M. H. Jones, M. Numata, M. Shimane, Identification and characterization of BRDT: A testis-specific gene related to the bromodomain genes RING3 and Drosophila fsh. *Genomics* **45**, 529-534 (1997).
115. M. S. Kim *et al.*, A draft map of the human proteome. *Nature* **509**, 575-581 (2014).
116. C. Pivot-Pajot *et al.*, Acetylation-dependent chromatin reorganization by BRDT, a testis-specific bromodomain-containing protein. *Mol Cell Biol* **23**, 5354-5365 (2003).
117. S. Dhar, A. Thota, M. R. Rao, Insights into role of bromodomain, testis-specific (Brdt) in acetylated histone H4-dependent chromatin remodeling in mammalian spermiogenesis. *J Biol Chem* **287**, 6387-6405 (2012).
118. J. Govin *et al.*, Histone acetylation-mediated chromatin compaction during mouse spermatogenesis. *Ernst Schering Res Found Workshop*, 155-172 (2006).
119. E. Shang, H. D. Nickerson, D. Wen, X. Wang, D. J. Wolgemuth, The first bromodomain of Brdt, a testis-specific member of the BET sub-family of double-bromodomain-containing proteins, is essential for male germ cell differentiation. *Development* **134**, 3507-3515 (2007).
120. M. M. Matzuk *et al.*, Small-molecule inhibition of BRDT for male contraception. *Cell* **150**, 673-684 (2012).
121. B. D. Berkovits, D. J. Wolgemuth, The role of the double bromodomain-containing BET genes during mammalian spermatogenesis. *Curr Top Dev Biol* **102**, 293-326 (2013).

122. B. D. Berkovits, L. Wang, P. Guarnieri, D. J. Wolgemuth, The testis-specific double bromodomain-containing protein BRDT forms a complex with multiple spliceosome components and is required for mRNA splicing and 3'-UTR truncation in round spermatids. *Nucleic Acids Res* **40**, 7162-7175 (2012).
123. G. LeRoy, B. Rickards, S. J. Flint, The double bromodomain proteins Brd2 and Brd3 couple histone acetylation to transcription. *Mol Cell* **30**, 51-60 (2008).
124. G. V. Denis *et al.*, Identification of transcription complexes that contain the double bromodomain protein Brd2 and chromatin remodeling machines. *J Proteome Res* **5**, 502-511 (2006).
125. G. V. Denis, C. Vaziri, N. Guo, D. V. Faller, RING3 kinase transactivates promoters of cell cycle regulatory genes through E2F. *Cell Growth Differ* **11**, 417-424 (2000).
126. J. Peng *et al.*, Brd2 is a TBP-associated protein and recruits TBP into E2F-1 transcriptional complex in response to serum stimulation. *Mol Cell Biochem* **294**, 45-54 (2007).
127. R. Gamsjaeger *et al.*, Structural basis and specificity of acetylated transcription factor GATA1 recognition by BET family bromodomain protein Brd3. *Mol Cell Biol* **31**, 2632-2640 (2011).
128. S. Rahman *et al.*, The Brd4 extraterminal domain confers transcription activation independent of pTEFb by recruiting multiple proteins, including NSD3. *Mol Cell Biol* **31**, 2641-2652 (2011).
129. D. Houzelstein *et al.*, Growth and early postimplantation defects in mice deficient for the bromodomain-containing protein Brd4. *Mol Cell Biol* **22**, 3794-3802 (2002).
130. E. Shang, X. Wang, D. Wen, D. A. Greenberg, D. J. Wolgemuth, Double bromodomain-containing gene Brd2 is essential for embryonic development in mouse. *Dev Dyn* **238**, 908-917 (2009).
131. T. Kanno *et al.*, Selective recognition of acetylated histones by bromodomain proteins visualized in living cells. *Mol Cell* **13**, 33-43 (2004).
132. B. N. Devaiah *et al.*, BRD4 is an atypical kinase that phosphorylates serine2 of the RNA polymerase II carboxy-terminal domain. *Proc Natl Acad Sci U S A* **109**, 6927-6932 (2012).
133. M. J. Scanlan *et al.*, Expression of cancer-testis antigens in lung cancer: definition of bromodomain testis-specific gene (BRDT) as a new CT gene, CT9. *Cancer Lett* **150**, 155-164 (2000).
134. F. Wang *et al.*, Brd2 disruption in mice causes severe obesity without Type 2 diabetes. *Biochem J* **425**, 71-83 (2009).
135. A. C. Belkina, B. S. Nikolajczyk, G. V. Denis, BET protein function is required for inflammation: Brd2 genetic disruption and BET inhibitor JQ1 impair mouse macrophage inflammatory responses. *J Immunol* **190**, 3670-3678 (2013).
136. A. C. Belkina, G. V. Denis, BET domain co-regulators in obesity, inflammation and cancer. *Nat Rev Cancer* **12**, 465-477 (2012).
137. D. K. Pal *et al.*, BRD2 (RING3) is a probable major susceptibility gene for common juvenile myoclonic epilepsy. *Am J Hum Genet* **73**, 261-270 (2003).

138. E. N. Yavuz *et al.*, Bromodomain-containing protein 2 gene in photosensitive epilepsy. *Seizure* **21**, 646-648 (2012).
139. R. J. Greenwald *et al.*, E mu-BRD2 transgenic mice develop B-cell lymphoma and leukemia. *Blood* **103**, 1475-1484 (2004).
140. M. E. Lenburg, A. Sinha, D. V. Faller, G. V. Denis, Tumor-specific and proliferation-specific gene expression typifies murine transgenic B cell lymphomagenesis. *J Biol Chem* **282**, 4803-4811 (2007).
141. C. French, NUT midline carcinoma. *Nat Rev Cancer* **14**, 149-150 (2014).
142. C. A. French, The importance of diagnosing NUT midline carcinoma. *Head Neck Pathol* **7**, 11-16 (2013).
143. C. A. French, Pathogenesis of NUT midline carcinoma. *Annu Rev Pathol* **7**, 247-265 (2012).
144. C. A. French, NUT midline carcinoma. *Cancer Genet Cytogenet* **203**, 16-20 (2010).
145. C. A. French *et al.*, BRD-NUT oncoproteins: a family of closely related nuclear proteins that block epithelial differentiation and maintain the growth of carcinoma cells. *Oncogene* **27**, 2237-2242 (2008).
146. J. Zuber *et al.*, RNAi screen identifies Brd4 as a therapeutic target in acute myeloid leukaemia. *Nature* **478**, 524-528 (2011).
147. M. A. Dawson *et al.*, Inhibition of BET recruitment to chromatin as an effective treatment for MLL-fusion leukaemia. *Nature* **478**, 529-533 (2011).
148. J. E. Delmore *et al.*, BET bromodomain inhibition as a therapeutic strategy to target c-Myc. *Cell* **146**, 904-917 (2011).
149. G. V. Denis, B. S. Nikolajczyk, G. R. Schnitzler, An emerging role for bromodomain-containing proteins in chromatin regulation and transcriptional control of adipogenesis. *FEBS Lett* **584**, 3260-3268 (2010).
150. S. Y. Wu, C. M. Chiang, The double bromodomain-containing chromatin adaptor Brd4 and transcriptional regulation. *J Biol Chem* **282**, 13141-13145 (2007).
151. C. M. Chiang, Nonequivalent response to bromodomain-targeting BET inhibitors in oligodendrocyte cell fate decision. *Chem Biol* **21**, 804-806 (2014).
152. J. Shi, C. R. Vakoc, The mechanisms behind the therapeutic activity of BET bromodomain inhibition. *Mol Cell* **54**, 728-736 (2014).
153. S. R. Floyd *et al.*, The bromodomain protein Brd4 insulates chromatin from DNA damage signalling. *Nature* **498**, 246-250 (2013).
154. J. M. Lamonica *et al.*, Bromodomain protein Brd3 associates with acetylated GATA1 to promote its chromatin occupancy at erythroid target genes. *Proc Natl Acad Sci U S A* **108**, E159-168 (2011).

155. Y. Nakamura *et al.*, Crystal structure of the human BRD2 bromodomain: insights into dimerization and recognition of acetylated histone H4. *J Biol Chem* **282**, 4193-4201 (2007).
156. N. Borcherding *et al.*, Paracrine WNT5A Signaling Inhibits Expansion of Tumor-Initiating Cells. *Cancer Res* **75**, 1972-1982 (2015).
157. H. Fukazawa, A. Masumi, The conserved 12-amino acid stretch in the inter-bromodomain region of BET family proteins functions as a nuclear localization signal. *Biol Pharm Bull* **35**, 2064-2068 (2012).
158. P. Garcia-Gutierrez, M. Mundi, M. Garcia-Dominguez, Association of bromodomain BET proteins with chromatin requires dimerization through the conserved motif B. *J Cell Sci* **125**, 3671-3680 (2012).
159. D. Gallenkamp, K. A. Gelato, B. Haendler, H. Weinmann, Bromodomains and their pharmacological inhibitors. *ChemMedChem* **9**, 438-464 (2014).
160. Q. Zhang *et al.*, Structural Mechanism of Transcriptional Regulator NSD3 Recognition by the ET Domain of BRD4. *Structure* **24**, 1201-1208 (2016).
161. J. You *et al.*, Kaposi's sarcoma-associated herpesvirus latency-associated nuclear antigen interacts with bromodomain protein Brd4 on host mitotic chromosomes. *J Virol* **80**, 8909-8919 (2006).
162. B. Florence, D. V. Faller, You bet-cha: a novel family of transcriptional regulators. *Front Biosci* **6**, D1008-1018 (2001).
163. Z. Yang *et al.*, Recruitment of P-TEFb for stimulation of transcriptional elongation by the bromodomain protein Brd4. *Mol Cell* **19**, 535-545 (2005).
164. E. A. Abbate, C. Voitenleitner, M. R. Botchan, Structure of the papillomavirus DNA-tethering complex E2:Brd4 and a peptide that ablates HPV chromosomal association. *Mol Cell* **24**, 877-889 (2006).
165. J. A. Smith *et al.*, Genome-wide siRNA screen identifies SMCX, EP400, and Brd4 as E2-dependent regulators of human papillomavirus oncogene expression. *Proc Natl Acad Sci U S A* **107**, 3752-3757 (2010).
166. S. Y. Wu *et al.*, Brd4 links chromatin targeting to HPV transcriptional silencing. *Genes Dev* **20**, 2383-2396 (2006).
167. C. Choudhary, B. T. Weinert, Y. Nishida, E. Verdin, M. Mann, The growing landscape of lysine acetylation links metabolism and cell signalling. *Nat Rev Mol Cell Biol* **15**, 536-550 (2014).
168. E. Verdin, M. Ott, 50 years of protein acetylation: from gene regulation to epigenetics, metabolism and beyond. *Nat Rev Mol Cell Biol* **16**, 258-264 (2015).
169. C. Dhalluin *et al.*, Structure and ligand of a histone acetyltransferase bromodomain. *Nature* **399**, 491-496 (1999).
170. S. Mujtaba *et al.*, Structural basis of lysine-acetylated HIV-1 Tat recognition by PCAF bromodomain. *Mol Cell* **9**, 575-586 (2002).

171. R. H. Jacobson, A. G. Ladurner, D. S. King, R. Tjian, Structure and function of a human TAFII250 double bromodomain module. *Science* **288**, 1422-1425 (2000).
172. D. J. Owen *et al.*, The structural basis for the recognition of acetylated histone H4 by the bromodomain of histone acetyltransferase gcn5p. *EMBO J* **19**, 6141-6149 (2000).
173. L. Zeng *et al.*, Selective small molecules blocking HIV-1 Tat and coactivator PCAF association. *J Am Chem Soc* **127**, 2376-2377 (2005).
174. F. A. Romero *et al.*, Disrupting Acetyl-Lysine Recognition: Progress in the Development of Bromodomain Inhibitors. *J Med Chem* **59**, 1271-1298 (2016).
175. J. W. Tamkun *et al.*, brahma: a regulator of Drosophila homeotic genes structurally related to the yeast transcriptional activator SNF2/SWI2. *Cell* **68**, 561-572 (1992).
176. F. Jeanmougin, J. M. Wurtz, B. Le Douarin, P. Chambon, R. Losson, The bromodomain revisited. *Trends Biochem Sci* **22**, 151-153 (1997).
177. R. Sanchez, M. M. Zhou, The role of human bromodomains in chromatin biology and gene transcription. *Curr Opin Drug Discov Devel* **12**, 659-665 (2009).
178. J. Meslamani, S. G. Smith, R. Sanchez, M. M. Zhou, Structural features and inhibitors of bromodomains. *Drug Discov Today Technol* **19**, 3-15 (2016).
179. M. Brand *et al.*, Correction to Small Molecule Inhibitors of Bromodomain-Acetyl-lysine Interactions. *ACS Chem Biol*, (2016).
180. M. Brand *et al.*, Small molecule inhibitors of bromodomain-acetyl-lysine interactions. *ACS Chem Biol* **10**, 22-39 (2015).
181. A. J. Barr *et al.*, Large-scale structural analysis of the classical human protein tyrosine phosphatome. *Cell* **136**, 352-363 (2009).
182. A. Unzue *et al.*, The "Gatekeeper" Residue Influences the Mode of Binding of Acetyl Indoles to Bromodomains. *J Med Chem* **59**, 3087-3097 (2016).
183. C. Galdeano, A. Ciulli, Selectivity on-target of bromodomain chemical probes by structure-guided medicinal chemistry and chemical biology. *Future Med Chem*, (2016).
184. C. H. Arrowsmith *et al.*, The promise and peril of chemical probes. *Nat Chem Biol* **11**, 536-541 (2015).
185. C. H. Arrowsmith *et al.*, Corrigendum: The promise and peril of chemical probes. *Nat Chem Biol* **11**, 887 (2015).
186. M. Pérez-Salvia, M. Esteller, Bromodomain inhibitors and cancer therapy: From structures to applications. *Epigenetics* **12**, 323-339 (2017).
187. T. Ito *et al.*, Real-time imaging of histone H4K12-specific acetylation determines the modes of action of histone deacetylase and bromodomain inhibitors. *Chem Biol* **18**, 495-507 (2011).

188. P. Ciceri *et al.*, Dual kinase-bromodomain inhibitors for rationally designed polypharmacology. *Nat Chem Biol* **10**, 305-312 (2014).
189. M. P. Martin, S. H. Olesen, G. I. Georg, E. Schönbrunn, Cyclin-dependent kinase inhibitor dinaciclib interacts with the acetyl-lysine recognition site of bromodomains. *ACS Chem Biol* **8**, 2360-2365 (2013).
190. S. W. Ember *et al.*, Acetyl-lysine binding site of bromodomain-containing protein 4 (BRD4) interacts with diverse kinase inhibitors. *ACS Chem Biol* **9**, 1160-1171 (2014).
191. A. Dittmann *et al.*, The commonly used PI3-kinase probe LY294002 is an inhibitor of BET bromodomains. *ACS Chem Biol* **9**, 495-502 (2014).
192. S. Venkataraman *et al.*, Inhibition of BRD4 attenuates tumor cell self-renewal and suppresses stem cell signaling in MYC driven medulloblastoma. *Oncotarget* **5**, 2355-2371 (2014).
193. X. Ran *et al.*, Structure-Based Design of γ -Carboline Analogues as Potent and Specific BET Bromodomain Inhibitors. *J Med Chem* **58**, 4927-4939 (2015).
194. M. Boi *et al.*, The BET Bromodomain Inhibitor OTX015 Affects Pathogenetic Pathways in Preclinical B-cell Tumor Models and Synergizes with Targeted Drugs. *Clin Cancer Res* **21**, 1628-1638 (2015).
195. M. M. Coudé *et al.*, BET inhibitor OTX015 targets BRD2 and BRD4 and decreases c-MYC in acute leukemia cells. *Oncotarget* **6**, 17698-17712 (2015).
196. A. Chaidos *et al.*, Potent antimyeloma activity of the novel bromodomain inhibitors I-BET151 and I-BET762. *Blood* **123**, 697-705 (2014).
197. E. Nicodeme *et al.*, Suppression of inflammation by a synthetic histone mimic. *Nature* **468**, 1119-1123 (2010).
198. P. Filippakopoulos *et al.*, Benzodiazepines and benzotriazepines as protein interaction inhibitors targeting bromodomains of the BET family. *Bioorg Med Chem* **20**, 1878-1886 (2012).
199. B. K. Albrecht *et al.*, Identification of a Benzoisoxazoloazepine Inhibitor (CPI-0610) of the Bromodomain and Extra-Terminal (BET) Family as a Candidate for Human Clinical Trials. *J Med Chem* **59**, 1330-1339 (2016).
200. J. Seal *et al.*, Identification of a novel series of BET family bromodomain inhibitors: binding mode and profile of I-BET151 (GSK1210151A). *Bioorg Med Chem Lett* **22**, 2968-2972 (2012).
201. A. Wyce *et al.*, BET inhibition silences expression of MYCN and BCL2 and induces cytotoxicity in neuroblastoma tumor models. *PLoS One* **8**, e72967 (2013).
202. S. Picaud *et al.*, PFI-1, a highly selective protein interaction inhibitor, targeting BET Bromodomains. *Cancer Res* **73**, 3336-3346 (2013).
203. P. V. Fish *et al.*, Identification of a chemical probe for bromo and extra C-terminal bromodomain inhibition through optimization of a fragment-derived hit. *J Med Chem* **55**, 9831-9837 (2012).

204. S. Picaud *et al.*, RVX-208, an inhibitor of BET transcriptional regulators with selectivity for the second bromodomain. *Proc Natl Acad Sci U S A* **110**, 19754-19759 (2013).
205. K. G. McLure *et al.*, RVX-208, an inducer of ApoA-I in humans, is a BET bromodomain antagonist. *PLoS One* **8**, e83190 (2013).
206. D. Bailey *et al.*, RVX-208: a small molecule that increases apolipoprotein A-I and high-density lipoprotein cholesterol in vitro and in vivo. *J Am Coll Cardiol* **55**, 2580-2589 (2010).
207. O. A. Kharenko *et al.*, RVX-297- a novel BD2 selective inhibitor of BET bromodomains. *Biochem Biophys Res Commun* **477**, 62-67 (2016).
208. X. Lucas *et al.*, 4-Acyl pyrroles: mimicking acetylated lysines in histone code reading. *Angew Chem Int Ed Engl* **52**, 14055-14059 (2013).
209. G. Zhang *et al.*, Structure-guided design of potent diazobenzene inhibitors for the BET bromodomains. *J Med Chem* **56**, 9251-9264 (2013).
210. M. Gacias *et al.*, Selective chemical modulation of gene transcription favors oligodendrocyte lineage progression. *Chem Biol* **21**, 841-854 (2014).
211. B. Raux *et al.*, Exploring Selective Inhibition of the First Bromodomain of the Human Bromodomain and Extra-terminal Domain (BET) Proteins. *J Med Chem*, (2016).
212. A. Chaidos, V. Caputo, A. Karadimitris, Inhibition of bromodomain and extra-terminal proteins (BET) as a potential therapeutic approach in haematological malignancies: emerging preclinical and clinical evidence. *Ther Adv Hematol* **6**, 128-141 (2015).
213. J. E. Bolden *et al.*, Inducible in vivo silencing of Brd4 identifies potential toxicities of sustained BET protein inhibition. *Cell Rep* **8**, 1919-1929 (2014).
214. C. Y. Fong *et al.*, BET inhibitor resistance emerges from leukaemia stem cells. *Nature* **525**, 538-542 (2015).
215. S. Shu *et al.*, Response and resistance to BET bromodomain inhibitors in triple-negative breast cancer. *Nature* **529**, 413-417 (2016).
216. A. M. Kurimchak *et al.*, Resistance to BET Bromodomain Inhibitors Is Mediated by Kinome Reprogramming in Ovarian Cancer. *Cell Rep* **16**, 1273-1286 (2016).
217. J. Settleman, Cancer: Bet on drug resistance. *Nature*, (2016).
218. S. Milhas *et al.*, Protein-Protein Interaction Inhibition (2P2I)-Oriented Chemical Library Accelerates Hit Discovery. *ACS Chem Biol* **11**, 2140-2148 (2016).
219. V. Hamon *et al.*, 2P2I HUNTER: a tool for filtering orthosteric protein-protein interaction modulators via a dedicated support vector machine. *J R Soc Interface* **11**, 20130860 (2014).
220. R. Bourgeas, M. J. Basse, X. Morelli, P. Roche, Atomic analysis of protein-protein interfaces with known inhibitors: the 2P2I database. *PLoS One* **5**, e9598 (2010).

221. M. J. Basse *et al.*, 2P2Idb: a structural database dedicated to orthosteric modulation of protein-protein interactions. *Nucleic Acids Res* **41**, D824-827 (2013).
222. M. J. Basse, S. Betzi, X. Morelli, P. Roche, 2P2Idb v2: update of a structural database dedicated to orthosteric modulation of protein-protein interactions. *Database (Oxford)* **2016**, (2016).
223. L. Drouin *et al.*, Structure enabled design of BAZ2-ICR, a chemical probe targeting the bromodomains of BAZ2A and BAZ2B. *J Med Chem* **58**, 2553-2559 (2015).
224. J. Bennett *et al.*, Discovery of a Chemical Tool Inhibitor Targeting the Bromodomains of TRIM24 and BRPF. *J Med Chem* **59**, 1642-1647 (2016).
225. W. S. Palmer *et al.*, Structure-Guided Design of IACS-9571, a Selective High-Affinity Dual TRIM24-BRPF1 Bromodomain Inhibitor. *J Med Chem* **59**, 1440-1454 (2016).
226. C. Abad-Zapatero, Ligand efficiency indices for effective drug discovery. *Expert Opin Drug Discov* **2**, 469-488 (2007).
227. M. Kuang *et al.*, Binding Kinetics versus Affinities in BRD4 Inhibition. *J Chem Inf Model* **55**, 1926-1935 (2015).
228. L. Hoffer, C. Chira, G. Marcou, A. Varnek, D. Horvath, S4MPLE--Sampler for Multiple Protein-Ligand Entities: Methodology and Rigid-Site Docking Benchmarking. *Molecules* **20**, 8997-9028 (2015).
229. L. Hoffer, D. Horvath, S4MPLE--sampler for multiple protein-ligand entities: simultaneous docking of several entities. *J Chem Inf Model* **53**, 88-102 (2013).
230. M. Tanaka *et al.*, Design and characterization of bivalent BET inhibitors. *Nat Chem Biol* **12**, 1089-1096 (2016).
231. M. J. Waring *et al.*, Potent and selective bivalent inhibitors of BET bromodomains. *Nat Chem Biol* **12**, 1097-1104 (2016).
232. C. Ballatore, D. M. Huryn, A. B. Smith, Carboxylic acid (bio)isosteres in drug design. *ChemMedChem* **8**, 385-395 (2013).
233. A. C. Groo *et al.*, Comparison of 2 strategies to enhance pyridoclox solubility: Nanoemulsion delivery system versus salt synthesis. *Eur J Pharm Sci* **97**, 218-226 (2017).
234. K. T. Savjani, A. K. Gajjar, J. K. Savjani, Drug solubility: importance and enhancement techniques. *ISRN Pharm* **2012**, 195727 (2012).
235. M. Philpott *et al.*, Assessing cellular efficacy of bromodomain inhibitors using fluorescence recovery after photobleaching. *Epigenetics Chromatin* **7**, 14 (2014).
236. T. Machleidt *et al.*, NanoBRET--A Novel BRET Platform for the Analysis of Protein-Protein Interactions. *ACS Chem Biol* **10**, 1797-1804 (2015).
237. R. Rodriguez, K. M. Miller, Unravelling the genomic targets of small molecules using high-throughput sequencing. *Nat Rev Genet* **15**, 783-796 (2014).

238. K. M. Sakamoto *et al.*, Protacs: chimeric molecules that target proteins to the Skp1-Cullin-F box complex for ubiquitination and degradation. *Proc Natl Acad Sci U S A* **98**, 8554-8559 (2001).
239. K. M. Sakamoto *et al.*, Development of Protacs to target cancer-promoting proteins for ubiquitination and degradation. *Mol Cell Proteomics* **2**, 1350-1358 (2003).
240. D. Puppala, H. Lee, K. B. Kim, H. I. Swanson, Development of an aryl hydrocarbon receptor antagonist using the proteolysis-targeting chimeric molecules approach: a potential tool for chemoprevention. *Mol Pharmacol* **73**, 1064-1071 (2008).
241. K. Okuhira *et al.*, Development of hybrid small molecules that induce degradation of estrogen receptor-alpha and necrotic cell death in breast cancer cells. *Cancer Sci* **104**, 1492-1498 (2013).
242. D. P. Bondeson *et al.*, Catalytic in vivo protein knockdown by small-molecule PROTACs. *Nat Chem Biol* **11**, 611-617 (2015).
243. A. C. Lai *et al.*, Modular PROTAC Design for the Degradation of Oncogenic BCR-ABL. *Angew Chem Int Ed Engl* **55**, 807-810 (2016).
244. D. L. Buckley *et al.*, HaloPROTACS: Use of Small Molecule PROTACs to Induce Degradation of HaloTag Fusion Proteins. *ACS Chem Biol* **10**, 1831-1837 (2015).
245. J. Lu *et al.*, Hijacking the E3 Ubiquitin Ligase Cereblon to Efficiently Target BRD4. *Chem Biol* **22**, 755-763 (2015).
246. G. E. Winter *et al.*, DRUG DEVELOPMENT. Phthalimide conjugation as a strategy for in vivo target protein degradation. *Science* **348**, 1376-1381 (2015).
247. M. Zengerle, K. H. Chan, A. Ciulli, Selective Small Molecule Induced Degradation of the BET Bromodomain Protein BRD4. *ACS Chem Biol* **10**, 1770-1777 (2015).
248. D. G. Gibson *et al.*, Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat Methods* **6**, 343-345 (2009).
249. X. Morelli, R. Bourgeas, P. Roche, Chemical and structural lessons from recent successes in protein-protein interaction inhibition (2P2I). *Curr Opin Chem Biol* **15**, 475-481 (2011).
250. F. Degorce *et al.*, HTRF: A technology tailored for drug discovery - a review of theoretical aspects and recent applications. *Curr Chem Genomics* **3**, 22-32 (2009).
251. F. H. Niesen, H. Berglund, M. Vedadi, The use of differential scanning fluorimetry to detect ligand interactions that promote protein stability. *Nat Protoc* **2**, 2212-2221 (2007).
252. C. J. Layton, H. W. Hellinga, Quantitation of protein-protein interactions by thermal stability shift analysis. *Protein Sci* **20**, 1439-1450 (2011).
253. F. Vollrath, N. Hawkins, D. Porter, C. Holland, M. Boulet-Audet, Differential Scanning Fluorimetry provides high throughput data on silk protein transitions. *Sci Rep* **4**, 5625 (2014).

254. O. Fedorov, F. H. Niesen, S. Knapp, Kinase inhibitor selectivity profiling using differential scanning fluorimetry. *Methods Mol Biol* **795**, 109-118 (2012).
255. A. Schön, R. K. Brown, B. M. Hutchins, E. Freire, Ligand binding analysis and screening by chemical denaturation shift. *Anal Biochem* **443**, 52-57 (2013).
256. M. W. Freyer, E. A. Lewis, Isothermal titration calorimetry: experimental design, data analysis, and probing macromolecule/ligand binding and kinetic interactions. *Methods Cell Biol* **84**, 79-113 (2008).
257. W. H. Ward, G. A. Holdgate, Isothermal titration calorimetry in drug discovery. *Prog Med Chem* **38**, 309-376 (2001).
258. N. C. Garbett, J. B. Chaires, Thermodynamic studies for drug design and screening. *Expert Opin Drug Discov* **7**, 299-314 (2012).
259. A. Gaulton *et al.*, ChEMBL: a large-scale bioactivity database for drug discovery. *Nucleic Acids Res* **40**, D1100-1107 (2012).
260. H. Sedgwick, J. E. Cameron, W. C. Poon, S. U. Egelhaaf, Protein phase behavior and crystallization: effect of glycerol. *J Chem Phys* **127**, 125102 (2007).
261. B. Heras, J. L. Martin, Post-crystallization treatments for improving diffraction quality of protein crystals. *Acta Crystallogr D Biol Crystallogr* **61**, 1173-1180 (2005).
262. M. Gelin *et al.*, Combining 'dry' co-crystallization and in situ diffraction to facilitate ligand screening by X-ray crystallography. *Acta Crystallogr D Biol Crystallogr* **71**, 1777-1787 (2015).

Summary

Development of protein-protein interaction inhibitors targeting bromodomain-containing proteins: implications in epigenetics and cancer development.

Bromodomain-containing proteins (BCPs) play a crucial role in the regulation of gene transcription, cell signaling and many other biological processes. Some epigenetic dysregulations can induce overexpression of these proteins, leading to the development of pathologies, such as inflammatory, metabolism and cardiovascular diseases, and more particularly cancers. BCPs consist of one or more **BromoDomain(s) module(s) (BDs)** involved in the recognition of one epigenetic mark (acetylation) on lysines of the histone tails, as well as in the recruitment of transcription factors. The forty-two BCPs of the human proteome are classified into eight families, including the "BET" which comprises four members BRD4, BRD3, BRD2 and BRDT. These four proteins are composed of a tandem of two BDs named BRDx(1) and BRDx(2) each belonging to the BD1 or the BD2 subfamily. The architecture of the central cavity of the BDs, qualified as "druggable", allows the emergence of these proteins as new promising **epigenetic targets** for the development of potent and specific inhibitors and drug candidates. To date, about twenty clinical trials targeting different types of cancer, including leukemia, have been initiated for "pan-BET" molecules that target all the members of this family. However, "pan-BET" inhibition is clinically problematic because it impacts many transcriptional pathways and causes the appearance of resistant cells.

The current challenge is to develop more selective inhibitors, for example towards the BET-BD1 subfamily or the BET-BD2 subfamily or ideally towards one single BD inside the BET family. The development of such "**selective epigenetic probes**" targeting BET family BDs should allow **deciphering their role and mechanism of action in various biological processes**. Identifying "**drug candidates**" should lead to **new targeted therapies** especially against cancers, in order to **overcome the resistances related to the use of pan-BET molecules**. In this context, our team focused on the search for selective inhibitors of the BET family by **screening our "2P2I_{3D}" library**, consisting of 1664 molecules with protein-protein interactions inhibition properties (PPI). The screening led to the selection of **seven molecules** further validated *in vitro* by various biophysical methods (HTRF, ITC, TSA, X-ray crystallography) and *in cellulo* on several cancer cell lines. Among these molecules, five exhibited a very classical "**pan-BET**" selectivity profile and two others showed original selectivity profiles that drew our attention. Indeed, the first is selective to the BD1 subdomain of the BETs family while the second is selective only toward the first subdomain of BRD4 protein.

My thesis project focused on optimization through a structure / function relationship program of the two identified selective molecules, as well as the best identified "pan-BET" molecule **to improve their affinity while maintaining their observed selectivity profile**. By combining the information obtained from various series of synthesized analogs (IC₅₀, K_D, EC₅₀ and crystallographic structures) with molecular modeling, we have successfully developed three epigenetic probe-like molecules: one pan-BET inhibitor; one compound with a selective profile towards the BD1 subdomains; and the last towards the BD1 subdomains of BRD4 and BRD3. We are now initiating a more comprehensive study on these compounds combining mainly chemoproteomics, genomics, transcriptomics, *in vivo* biophysical approaches (CHIP-seq, FRAP and nano-BRET) with more classical cellular studies (cytotoxicity, c-Myc expression, apoptosis...). These data are essential to define whether our efforts to develop selective inhibitors of a single BD or subfamily should be continued in the future.

Key Words:

(1) *protein-protein interaction*, (2) *inhibitors*,
(3) *bromodomains*, (4) *epigenetics*, (5) *cancer*,
(6) *structural biology*.

Brigitt RAUX

Résumé

Développement d'inhibiteurs d'interaction protéine-protéine ciblant les protéines à bromodomains : implication en épigénétique et dans le développement de cancers.

Les protéines à bromodomains (BCPs) jouent un rôle essentiel dans la régulation de la transcription de gènes, la signalisation cellulaire et bien d'autres processus biologiques. Certaines dérégulations épigénétiques peuvent conduire à une surexpression de ces protéines, entraînant le développement de diverses pathologies, telles que les maladies inflammatoires, du métabolisme, cardiovasculaires et plus particulièrement les cancers... Les BCPs sont constituées d'un ou plusieurs module(s) **BromoDomaine(s)** (BDs), qui sont impliqués dans la reconnaissance des marques épigénétiques (acétylation) au niveau des lysines des queues des histones, ainsi que dans le recrutement de facteurs de transcription. Les quarante-deux BCPs du protéome humain sont classifiées en huit familles, dont la famille « BET » qui comprend quatre membres BRD4, BRD3, BRD2 et BRDT. Ces quatre protéines sont constituées de deux BDs en tandem, nommés BRDx(1) et BRDx(2), formant deux sous-familles, BD1 et BD2. L'architecture de la cavité centrale des BDs, qualifiée de « druggable », a permis l'émergence de ces protéines en tant que **nouvelles cibles épigénétiques prometteuses** pour le développement d'inhibiteurs affins et spécifiques et de « **candidat médicaments** ». À ce jour, une vingtaine d'essais cliniques ciblant divers types de cancers, incluant les leucémies, ont été initiés pour des molécules « pan-BET », inhibant l'ensemble des membres de cette famille. Cependant, l'inhibition « pan-BET » est problématique au niveau clinique puisqu'elle impacte de nombreuses voies de transcription et engendre l'apparition de cellules résistantes.

Le challenge actuel est de développer des inhibiteurs plus sélectifs, par exemple envers la sous-famille BET-BD1 ou la sous-famille BET-BD2 ou plus idéalement envers un seul BD de la famille BET. Le développement de « **sondes épigénétiques sélectives** » ciblant des BDs de la famille BET, devrait permettre de **décrypter leur rôle et leur mécanisme d'action dans les divers processus biologiques**. L'identification de « **candidat médicament** » devrait aboutir à de nouvelles thérapies plus ciblées notamment contre les cancers, afin de **pallier les résistances liées à l'utilisation de molécules pan-BET** inhibitrices. Dans ce contexte, notre équipe s'est concentrée à la recherche d'inhibiteurs sélectifs de la famille BET **en criblant notre chimiothèque**, « **2P2I3D** », constituées de 1664 molécules focalisées sur les interactions protéine-protéine (PPI). Le criblage de celle-ci a permis de sélectionner **sept molécules** qui ont été validées par la suite in vitro par diverses techniques biophysiques (HTRF, ITC, TSA, et cristallographie aux rayons X) et in cellulo sur plusieurs lignées cancéreuses. Parmi ces molécules, **cinq** d'entre elles présentaient un profil de sélectivité très classique de type « **pan-BET** » et les **deux** autres montraient une **sélectivité très originale**. La première présentait une sélectivité dirigée contre les **sous-domaines BD1 des protéines BET**, la seconde dirigée contre le **domaine BRD4(1)**.

Mon projet de thèse s'est focalisé sur l'optimisation par une étude de relation structure/fonction de nos deux molécules sélectives, ainsi que de notre meilleure molécule « pan-BET » **inhibitrice afin d'améliorer leur affinité tout en tentant de conserver leur profil de sélectivité observé**. En combinant les informations obtenues pour les diverses séries d'analogues synthétisés (IC₅₀, K_D, EC₅₀ et structures cristallographiques) à de la modélisation moléculaire, nous avons réussi à développer **trois pseudo-sondes épigénétiques : une pan-BET, une sélective des sous-domaines BD1 des protéines BET et la dernière sélective des domaines BD1 de BRD4 et BRD3**. Nous avons débuté une étude cellulaire plus approfondie combinant de la chimoprotéomique, de la génomique et de la transcriptomique, associées à des approches de biophysique in cellulo (CHIP-seq, FRAP et nano-BRET) et à des études plus classiques de biologie cellulaire (cytotoxicité, expression c-Myc, apoptose...). Ces études sont essentielles pour décider si les efforts fournis dans le développement d'inhibiteurs sélectifs d'un seul BD ou d'une sous-famille doivent être continués ou non dans le futur.

Mots clefs :

(1) interactions protéine-protéine, (2) inhibiteurs, (3) bromodomains, (4) épigénétique, (5) cancer, (6) biologie structurale.

Brigitt RAUX