

Université Victor Segalen Bordeaux 2

Année 2010

Thèse n°1778

THESE

Pour le

DOCTORAT DE L'UNIVERSITE BORDEAUX 2

Mention : Sciences, Technologie, Santé

Option : Microbiologie

Présentée et soutenue publiquement

Le 13 décembre 2010

Par

Muriel FAURE-DELLA CORTE

Née le 8 août 1974 à Périgueux

**Cytomégalo­virus Humain :
Variabilité, Recombinaison et Protéine pUL40**

Membres du jury

Monsieur le Professeur Michael KANN
Madame le Professeur Sophie ALAIN
Monsieur le Professeur Henri AGUT
Madame le Docteur Thérèse ASTIER-GIN
Monsieur le Professeur Hervé FLEURY
Madame le Professeur Marie-Edith LAFON

Président
Rapporteur
Rapporteur
Membre
Membre invité
Directeur de thèse

Remerciements

Je remercie Monsieur le Professeur Michael Kann d'avoir accepté d'être le Président de ce jury et de m'accepter parmi les nouveaux membres de son équipe.

Je tiens à remercier particulièrement Madame le Professeur Sophie Alain et Monsieur le Professeur Henri Agut pour avoir accepté de juger ce travail en tant que rapporteurs après avoir suivi les travaux de notre équipe sur le cytomégalo virus humain depuis 2004.

Je remercie vivement Madame le Docteur Thérèse Astier-Gin, mon tuteur au sein de l'école doctorale « Sciences de la Vie et de la Santé » de Bordeaux II, pour ses conseils et son écoute tout au long de ce travail.

J'exprime ma gratitude à Monsieur le Professeur Hervé Fleury qui m'a permis de mener à bien ce doctorat en parallèle avec mes activités de technicienne dans l'EA2968.

Je remercie tout particulièrement Madame Marie-Edith Lafon pour m'avoir dirigée tout au long de ce travail de thèse. Merci pour ta bonne humeur, pour ton soutien sans faille et pour tous ces moments partagés autour de plaquettes de chocolat.

Je remercie tous les membres de l'équipe EA2968, à Isabelle, Evelyne, Valérie D, Jacques, Valérie R, Paty, Sandrine, Panxicha et Laurent pour leur soutien, leur bonne humeur et leurs conseils.

Je remercie tout spécialement Jacqueline Plissoneau et à Michel Castroviejo pour m'avoir initiée à la purification de protéines, leur aide précieuse concernant les manip de la deuxième partie et leur bonne humeur.

Je remercie mes parents qui ont toujours cru en moi et pour leur soutien quelles que soient les circonstances de la vie.

Enfin, j'adresse un remerciement tout spécial à mes deux amours Erik et Pierre. Merci pour votre soutien inébranlable au quotidien et pour m'accepter avec mon caractère particulier. Je vous aime.

Résumé

Le cytomégalo virus humain (CMV) appartient à la sous-famille des Betaherpès virus. L'homme est son seul réservoir et il infecte 40 à 90% de la population mondiale. Ce virus, rarement dangereux pour le sujet immunocompétent, représente une réelle menace chez le patient immunodéprimé comme le transplanté d'organe. Après la primo-infection, le CMV diffuse dans tout l'organisme et alterne des phases de latence et de réactivation. Il consacre 20% de son génome à diverses stratégies d'échappement immunitaire.

Les objectifs de notre travail sont de décrire dans une première partie

1- la variabilité de six gènes, codant quatre protéines immunomodulatrices (UL18, UL40, UL111a et US3) et deux protéines immunodominantes (IE1 et pp65), dans des souches de CMV directement séquencées à partir de sang total ou du LBA de patients infectés et immunodéprimés,

2- le mécanisme conduisant à cette variabilité

3- la répartition des souches virales dans le sang total et le LBA (compartimentation).

La deuxième partie est consacrée à l'expression de la protéine immunomodulatrice pUL40.

Nos travaux ont permis de confirmer l'existence d'un polymorphisme notable, supérieur à ce qui est déjà décrit. Des conséquences fonctionnelles pourraient en découler dans les protéines correspondantes. L'étude *in silico* de la variabilité du CMV met en lumière le rôle clé de la recombinaison comme mécanisme évolutif majeur. Nous n'avons pas mis en évidence de compartimentation des souches virales dans les échantillons analysés.

Nos résultats préliminaires indiquent la faisabilité de la sur-expression de la protéine pUL40, permettant ultérieurement de mieux comprendre ses fonctions.

Mots clés : CMV, variabilité, recombinaison, souches cliniques de CMV, compartimentation, outils bioinformatiques, expression pUL40

Ce travail a été réalisé au sein de l'EA2968, Université Victor Segalen Bordeaux 2

146 rue Léo Saignat, 33076 Bordeaux Cédex.

Titre en anglais

Human Cytomegalovirus : Variability, Recombination and Protein pUL40

Abstract

Human Cytomegalovirus (CMV) belongs to the Betaherpesviruses sub-family. Human beings are its sole reservoir and it infects 40 to 90% of the world population. This virus is rarely dangerous for the immunocompetent host, but represents a problematic opportunistic agent in immunocompromised patients, such as transplant recipients. After primary infection, CMV spreads widely in the organism and alternates latency and reactivation phases. CMV dedicates 20% of its genome to various immune escape strategies.

Our aims were to describe, in a first part,

1- the variability of six CMV genes, which encode 4 immunomodulatory proteins (UL18, UL40, UL111a and US3) and two immunodominant proteins (IE1 and pp65), after direct amplification and sequencing from whole blood and bronchoalveolar lavages (BAL) taken from infected immune compromised patients,

2- the mechanism responsible for this variability

3- the distribution of CMV strains in whole blood and BAL (compartmentalization).

In a second part, we attempted pUL40 expression.

Our results confirmed the existence of a noticeable polymorphism, superior to what was previously described. Functional consequences could arise for the corresponding proteins. Our *in silico* study of CMV variation indicated a key role for recombination as a major evolutionary mechanism. We have observed little CMV compartmentalization among the investigated samples.

Our preliminary results indicated pUL40 may be over-expressed, which could allow a better understanding of its functions in a near future.

Keywords : CMV, variability, recombination, CMV clinical strains, compartmentalization, sequence analysis, pUL40 expression

Table des matières

Résumé	4
Titre en anglais	5
Abstract	6
Table des matières	7
Abbréviations.....	9
Introduction	11
Introduction	12
I. Biologie générale des <i>Herpesviridae</i>	12
II. Généralités sur le CMV	13
A. Historique	13
B. Epidémiologie	14
C. Pathogénèse	16
III. Structure du virus	18
IV. Le cycle de réplication viral	23
A. L'attachement et la pénétration du virus	23
B. La synthèse des protéines virales et réplication de l'ADN	25
1. La phase très précoce IE (Immediate Early, de H0 à H2) :	25
2. La phase précoce E (Early, de H12 à H24)	25
3. La phase tardive L (Late)	27
C. La libération du virus	27
V. Réponse immunitaire	27
A. Réponse immunitaire innée	27
B. Réponse immunitaire adaptative	28
C. Schéma récapitulatif de la réponse immunitaire anti-CMV	29
VI. Infection par le CMV chez le patient transplanté	31
VII. Echappement immunitaire	31
A. Gènes immunomodulateurs	31
B. Gènes immunodominants	37
C. Variabilité du CMV	39
Première Partie.....	43
Variabilité génomique de souches de CMV isolées chez des patients transplantés	44
VIII. Introduction	44
IX. Objectifs	45
X. Méthodes	45
A. Extraction de l'ADN	45
B. Amplification des gènes viraux par PCR	46
C. Purification des produits PCR	48
D. Réaction de séquence	48
E. Clonage.....	48
1. Préparation des bactéries compétentes	48
2. Ligation	48
3. Transformation	49

4. Sélection des bactéries ayant intégré le plasmide recombinant (système blanc/bleu).....	49
5. Repiquage.....	50
6. Stockage.....	50
F. F. Extraction plasmidique.....	51
F. F. Extraction plasmidique.....	52
G. Digestion plasmidique.....	52
H. Analyse phylogénique.....	52
1. Clustal W 1.83.....	52
2. SplitsTree.....	53
I. Méthode de bootscan.....	54
J. Analyse des épitopes immunodominants.....	54
XI. Patients.....	54
XII. Résultats.....	56
A. Valorisation des résultats.....	56
B. Variabilité globale.....	58
1. Les quatre gènes immunomodulateurs.....	58
2. Les deux gènes immunodominants.....	65
C. Compartimentation.....	70
D. Recombinaison.....	75
1. Arbres phylogéniques et analyse de la recombinaison avec le logiciel SplitsTree..	77
2. Analyse par bootscan.....	81
XIII. Discussion.....	85
Deuxième Partie.....	92
Expression de la protéine pUL40.....	93
XIV. pUL40.....	93
XV. Objectifs.....	95
XVI. Matériel.....	95
XVII. Méthodes.....	98
A. Amplification.....	98
B. Purification.....	99
C. Description du système hétérologue.....	99
D. Clonage.....	100
E. Expression de la protéine.....	102
F. Extraction.....	103
G. Purification de la protéine par chromatographie d'affinité.....	103
H. Séparation des protéines par SDS-PAGE.....	104
1. Coloration des gels protéiques au bleu de Coomassie.....	104
2. Coloration au nitrate d'argent.....	104
3. Immunodétection des protéines séparées sur gel de polyacrylamide : Western blot.....	105
XVIII. Résultats et discussion.....	106
Conclusion.....	112
Conclusion Générale.....	113
Bibliographie.....	114
Annexes 1 et 2.....	122

Abbreviations

A	Alanine
aa	Acide aminé
Ag	Antigénémie
ARNm	ARN messenger
CMH	Complexe Majeur d'Histocompatibilité
CMV	Cytomégalovirus
C	Cystéine
D	Selon le contexte D pour Donneur ou D pour Aspartate
DO	Densité optique
E	Selon le contexte E pour Early ou phase précoce, ou E pour Glutamate
EBV	Epstein-Barr virus
EGF-R	Epithelial Growth Factor Receptor
F	Phénylalanine
G	Glycine
gp	Glycoprotéine
H	Histidine
HHV-5	Herpes virus humain 5
HHV-6	Herpes virus humain 6
HHV-7	Herpes virus humain 7
HHV-8	Herpes virus humain 8
HLA	Human Leucocyte Antigen
HSPG	Heparane Sulfate Proteoglycap
HSV-1 et -2	Herpes simplex virus types 1 et 2
I	Isoleucine
IE	Immediate Early ou phase très précoce
IL	Interleukine
ILT	Ig-Like Transcripts
IRL	Long internal repeat
IRS	Short internal repeat
K	Lysine
kDa	KiloDalton
kpb	Kilopaire de bases
L	Selon le contexte, L pour late ou phase tardive, ou L pour Leucine
LBA	Liquide de Lavage Broncho Alvéolaire
LIR	Leukocyte Ig-like Receptor
M	Méthionine
MIEP	Major Immediate Early Promotor
moi	Multiplicity of infection
N	Asparagine
NF-κB	Nuclear Factor κB
NK	Cellules Natural Killer
NKG2	Famille des récepteur des NK
ORF	Cadre ouvert de lecture
P	Proline
PCR	Réaction de Polymérisation en Chaîne
PDI	Protein disulfide isomérase

pb	Paire de bases
pp	Phosphoprotéine
Q	Glutamine
R	Selon le contexte, R pour receveur, ou R pour Arginine
RE	Réticulum endoplasmique
S	Sérine
SDS	Sodium dodécylsulfate
T	Thréonine
TAP	Transporter associated with antigen processing
TLR	Toll Like Receptor
TRL	Terminal long repeat
TRS	Terminal short repeat
UL	Unit Long
Us	Unit Short
V	Valine
VIH	Virus de l'Immunodéficience Humaine
vIL-10	Interleukine 10 virale
VZV	Varicelle-zona virus
W	Tryptophane
WB	Sang total
Y	Tyrosine

Introduction

Introduction

I. Biologie générale des *Herpesviridae*

Dans la nature, les herpesvirus infectent les espèces vertébrées et non vertébrées. Parmi la centaine d'herpesvirus au moins partiellement caractérisés, seulement huit infectent l'homme : herpes simplex virus types 1 et 2 (HSV-1 et-2), varicelle-zona virus (VZV ou HHV-3 pour herpesvirus 3 humain), Epstein-Barr virus (EBV ou HHV-4), cytomegalovirus (CMV ou HHV-5), herpesvirus 6 humain (HHV-6, variants A et B), herpesvirus 7 humain (HHV-7), et herpesvirus 8 humain (HHV-8). Un virus simien, nommé virus B, peut infecter occasionnellement les humains.

Les herpesvirus infectant l'homme possèdent quatre propriétés biologiques importantes. Ils codent tous des enzymes particulières, impliquées dans la biosynthèse des acides nucléiques viraux ; ces enzymes sont structurellement différentes et fournissent des sites spécifiques pour l'inhibition par des antiviraux. La synthèse et l'assemblage de l'ADN viral sont initiés dans le noyau, tout comme l'assemblage de la capsid. Le relargage de nouveaux virions à partir de la cellule infectée s'accompagne de la mort de la cellule. Enfin, tous les herpesvirus peuvent établir une infection latente à l'intérieur de tissus spécifiques, caractéristiques de chaque virus, reflétant le tropisme tissulaire de chaque membre de cette famille.

Les herpesvirus ont un arrangement de séquences dans leur génome leur permettant d'être divisés en six groupes classés de A à F. Les herpesvirus infectant l'homme appartiennent aux groupes C, D et E. Dans le groupe C, où sont classés le virus EBV et le virus du sarcome de Kaposi, les séquences répétées terminales divisent le génome en plusieurs domaines bien délimités. Les génomes du groupe D, incluant le virus varicelle-zona, possèdent des séquences terminales répétées avec une orientation inversée. Les génomes de groupe E, comme celui des herpes simplex virus et du cytomegalovirus, sont flanqués à chaque extrémité de séquences répétées internes et terminales et les séquences internes ont une orientation inversée.

Indépendamment de ces groupes, les herpesvirus sont divisés en sous-familles déterminant la parenté évolutive résumant les propriétés propres à chaque membre.

- **Alpha herpesvirus ou *Alphaherpesvirinae*** : Ils sont caractérisés par un cycle reproductif extrêmement court (quelques heures), une destruction rapide de la cellule infectée et une capacité à se répliquer dans une large variété de tissus de l'hôte. Ils

établissent leur latence dans les ganglions sensoriels des nerfs. Cette sous-famille comprend les herpes simplex virus 1 et 2 et le virus varicelle-zona.

- **Beta herpesvirus ou *Betaherpesvirinae*** : Leur cycle de multiplication est long (jours), avec une infection progressant lentement dans les cellules. Une de leurs caractéristiques est leur capacité à former de grandes cellules, spécifique d'une infection par le cytomégalovirus humain. Ils établissent leur latence dans les glandes sécrétrices, les cellules endothéliales et les reins. Contrairement au CMV et au HHV-6 qui appartiennent à la même sous-famille des *Betaherpesvirinae*, le HHV-7 n'a été spécifiquement rattaché avec certitude à aucune maladie bien que ces trois virus soient suspectés d'interagir ensemble pour induire des maladies chez les sujets immunodéprimés.
- **Gamma herpesvirus ou *Gammaherpesvirinae*** : Les gamma herpesvirus ont un nombre très limité de cibles cellulaires. Ils se répliquent dans les cellules lymphoblastoïdiques *in vitro* et peuvent causer des infections lytiques dans certaines cellules. La latence des virus a été démontrée dans les tissus lymphoïdes. Le virus EBV et le virus du sarcome de Kaposi sont des membres de cette sous-famille.

II. Généralités sur le CMV

A. Historique

En 1904, Ribbert, Jesionek et Kiolemenoglou décrivent pour la première fois la présence de grandes cellules à inclusion intranucléaire dans le foie, les poumons, les reins et dans les parotides de fœtus et d'enfants morts-nés (Jesionek and Kiolemenoglou, 1904; Ribbert, 1904). Ils donnent le nom de « maladie des inclusions cytomégaliqes ». Le rapprochement de ces lésions avec les cellules de lésions cutanées de la varicelle suggère l'origine virale de la maladie dans les années 20. En 1932, il est démontré par Farber que 12% des enfants morts de pathologies variées sans point commun apparent possèdent des inclusions cytomégaliqes dans leur glande salivaire (Farber and Wolbach, 1932). Le virus associé est désigné sous le nom de « virus des glandes salivaires ». Dans les années 50, le virus est isolé pour la première fois (Craig *et al.*, 1957; Weller and Hanshaw, 1962; Wyatt, Saxton, and *et al.*, 1950). En 1956, la répliqation de virus responsable de la « maladie des inclusions cytomégaliqes » est obtenue par Smith dans des fibroblastes humains cultivés *in vitro* (Smith, 1956). Cette même année, la souche AD169 est isolée par Rowe dans le tissu

adénoïdien d'un enfant apparemment sain (Rowe *et al.*, 1956). Avec la souche Towne isolée dans l'urine d'un enfant malade (Plotkin *et al.*, 1975), c'est la souche de laboratoire la plus utilisée. Le terme « cytomégalovirus » est donné en 1960 par Weller *et al* en raison de la morphologie des cellules infectées (Weller, Hanshaw, and Scott, 1960).

B. Epidémiologie

L'homme est le seul réservoir connu du CMV humain. Aucun animal de laboratoire n'est sensible à ce virus. Il appartient à la sous-famille des *Betaherpesvirinae*. Le CMV est un virus endémique, présent dans le monde entier. Il n'existe pas d'influence saisonnière, mais la séroprévalence varie en fonction de plusieurs facteurs :

- Elle augmente avec l'âge.
- Elle est plus importante dans les régions en voie de développement (90% en Afrique et en Asie, 40 à 80% dans les régions industrialisées). Elle diminue avec l'augmentation du niveau socio-économique et sanitaire (Cannon, Schmid, and Hyde, 2010). En France, en 2002, environ 50% des jeunes adultes sont séropositifs vis-à-vis du CMV (Imbert-Marcille, 2001).

On observe deux pics de contamination pour les infections communautaires : un au cours des premières années de vie (infection materno-fœtale et crèches et cela concerne 10 à 25% des enfants) et le deuxième au début de la vie sexuelle active.

Chez le sujet infecté, le virus est présent dans toutes les sécrétions corporelles : salive, sang, urine, sperme, larmes, lait maternel, sécrétions vaginales. Le CMV étant un virus enveloppé, il ne résiste pas dans le milieu extérieur et est transmis par contact étroit avec ces liquides biologiques, par voie transplacentaire, par voie sanguine.

Chez le sujet immunocompétent, l'infection à CMV est le plus souvent asymptomatique. La primo-infection peut donner lieu à une mononucléose similaire à celle provoquée par le virus EBV : syndrome pseudo-grippal, lymphocytose atypique, pharyngite, anomalies de la fonction hépatique (Wreghitt *et al.*, 2003). Les résultats sérologiques permettent de distinguer ces deux étiologies. Dans de rares cas, des complications telles une pneumopathie, une myocardite, une méningite aseptique ou un syndrome de Guillain-Barré peuvent survenir (Gandhi and Khanna, 2004).

L'infection congénitale à CMV est courante et se produit pour 0,5 à 2% des naissances. Environ 10% des enfants infectés *in utero* vont développer des symptômes sévères, comme une hépatosplénomégalie, une rétinite, un rash pétéchial ou purpurique ou

une atteinte du système nerveux central (calcifications intracrâniennes, ...). Les enfants asymptomatiques à la naissance peuvent développer des syndromes ultérieurement, le plus fréquent étant une surdité. La maladie congénitale symptomatique est plus fréquente quand la femme enceinte développe une primo-infection à CMV (30 à 50% de risque de transmission au fœtus) plutôt qu'une réactivation et/ou une réinfection (0,5 à 2% de risque de transmission au fœtus). La transmission de la mère à l'enfant se fait par plusieurs chemins incluant le placenta (CMV congénital), le lait maternel, les sécrétions cervicales et la salive. Un enfant peut transmettre l'infection à sa mère ou à une femme enceinte par la salive ou les urines. De plus, l'excrétion du CMV par les enfants infectés représente un risque particulièrement chez la femme enceinte séronégative qui risque de développer une primo-infection durant sa grossesse. Cependant, une étude récente, portant sur 5173 femmes enceintes (Vauloup-Fellous *et al.*, 2009), indiquait que des informations simples sur les mesures d'hygiène données à une large population de femmes enceintes au début de leur grossesse réduisaient significativement l'incidence d'une infection congénitale durant la grossesse et le nombre de fœti infectés. Les mesures préconisées étaient un lavage fréquent des mains surtout après exposition à la salive ou à l'urine d'un enfant et la limitation des contacts rapprochés avec de jeunes enfants (baiser sur la bouche, dormir avec, partage de la cuillère ou des boissons).

Une autre population à risque est celle des patients immunodéprimés et en particulier celle des transplantés pour lesquels primo-infection et réactivation sont source de morbidité et de mortalité. Ces patients vont développer une infection symptomatique ou asymptomatique. Les patients séronégatifs ayant reçu un organe de patients séropositifs pour le CMV peuvent développer une primo-infection à CMV. La réactivation de l'infection à CMV peut aussi être problématique et l'ampleur de l'immunosuppression est un déterminant majeur de la sévérité de la maladie.

En 2004, Paya *et al* ont établi des définitions pour les termes « infection à CMV », « maladie invasive à CMV » et « syndrome CMV » (Paya *et al.*, 2004). Le patient asymptomatique développe une **infection à CMV** lorsque la présence du virus est prouvée dans le sang circulant (virémie, Ag pp65, ADNémie). Lorsque l'infection est symptomatique, il s'agit d'une **maladie à CMV** : soit un **syndrome CMV**, soit une **maladie invasive à CMV**. Le syndrome CMV associe fièvre $\geq 38^{\circ}\text{C}$ et l'un des signes suivants : malaise, leucopénie $< 3500/\mu\text{L}$, lymphocytose atypique $\geq 5\%$, thrombocytopénie, élévation des transaminases > 2 fois la normale (sauf chez les transplantés hépatiques). La maladie est invasive lorsqu'il y a dysfonction d'organe (excluant le rejet aigu de l'organe greffé) avec

preuve de la présence virale dans cet organe (dans une biopsie, un liquide de lavage broncho-alvéolaire (LBA) par exemple).

Actuellement, l'infection à CMV représente la virose opportuniste la plus fréquente après transplantation d'organe et, depuis la diffusion large de la vaccination contre la rubéole, la plus fréquente des infections virales congénitales dans les pays industrialisés (Rowshani *et al.*, 2005).

C. Pathogenèse

Après contamination, le virus va disséminer dans tout l'organisme grâce à la distribution ubiquitaire de ses cibles : cellules endothéliales, épithéliales, fibroblastiques (Alain and Mazon, 2003). La transmission du virus se fait de cellule à cellule par contiguïté. Les épithélia forment l'interface entre les organes infectés et leur environnement et interviennent dans la pénétration et l'excrétion du virus. Elles peuvent se détacher, circuler, obstruer la lumière des capillaires et induire l'infection d'un organe (Alain and Mazon, 2003). Elles transmettent le virus par contact aux polynucléaires et aux monocytes circulants. L'infection de ces derniers est abortive, mais ils deviennent permissifs après leur différenciation en macrophages. Les fibroblastes, cellules des tissus conjonctifs, représentent une cible majeure du virus.

L'infection des monocytes-macrophages et des cellules dendritiques joue un rôle important dans l'évolution de la maladie. Ces deux types de cellules présentatrices d'antigène dans la réponse immunitaire représentent également des sites de latence.

Le tropisme pour les cellules souches myélomonocytaires et les cellules stromales de la moelle osseuse entraîne des perturbations de l'hématopoïèse.

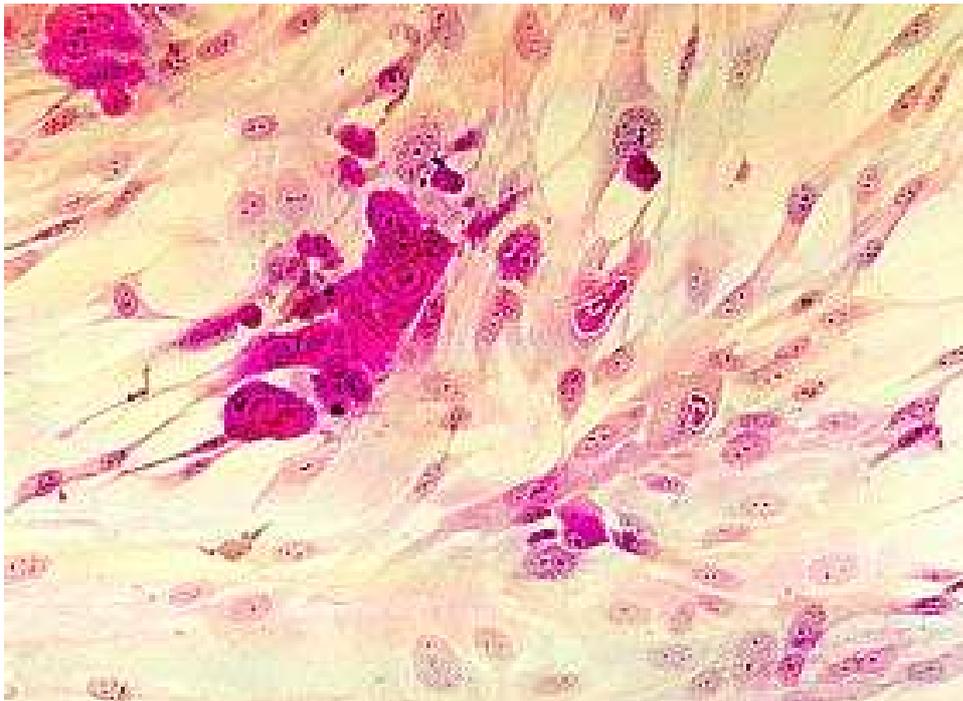
Le virus peut également infecter les cellules musculaires lisses et provoquer des ulcérations du tractus gastro-intestinal ou des perforations. Ce tropisme pourrait également intervenir dans le processus d'accélération de l'athérosclérose (Plachter, Sinzger, and Jahn, 1996; Sinzger and Jahn, 1996; Sinzger *et al.*, 1996).

L'atteinte des cellules du système nerveux central et des neurones périphériques est à l'origine de manifestations neurologiques (Luo, Schwartz, and Fortunato, 2008).

La transmission materno-fœtale s'explique par l'infection des cellules du trophoblaste et du syncytiotrophoblaste.

La multiplication du virus dans les organes est lente. Il induit des cellules géantes caractéristiques avec une inclusion intranucléaire associée à une inclusion intracytoplasmique (Figure 1).

Figure 1 : Effet cytopathique : foyer ovalaire sur fibroblaste, cellules augmentées de volume et réfringentes, aspect en banc de poisson. Délai 8-20 jours. umr5558-mq1.univ-lyon1.fr



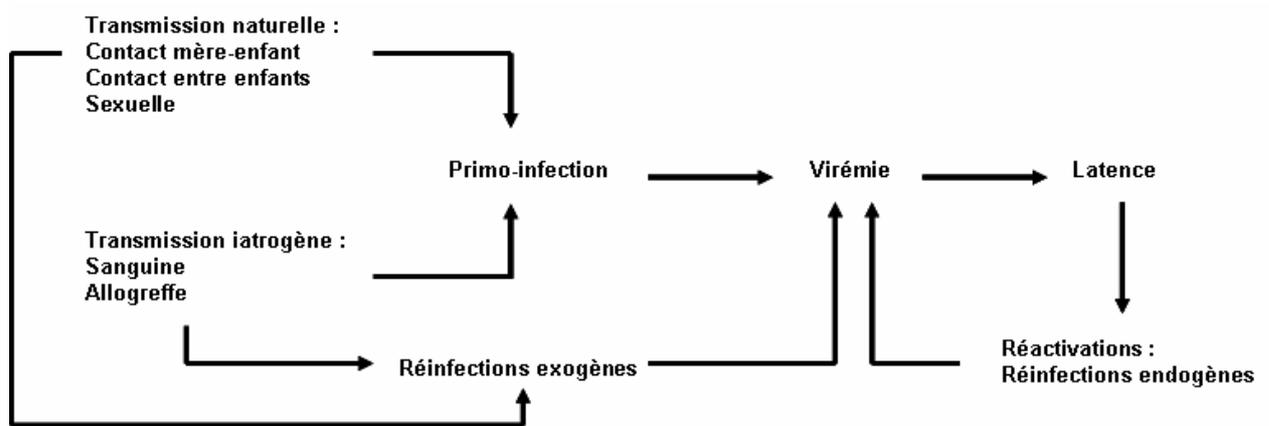
Au cours d'une transplantation, les organes du donneur peuvent être une source de contamination par le cytomégalovirus pour le receveur, et les leucocytes du sang périphérique peuvent être impliqués lors d'une transmission du cytomégalovirus durant une transfusion sanguine, cette transmission est prévenue par la déleucocytation des dons de sang.

Après la primo-infection, le CMV, comme tous les herpesvirus, persiste dans l'organisme et alterne des épisodes de latence et de réactivation chez son hôte. Les sites de latence sont les cellules endothéliales, les monocytes circulants, les progéniteurs médullaires. Le virus est présent à l'état latent dans de très nombreux tissus au niveau des cellules épithéliales, des fibroblastes, des cellules du muscle lisse. Les mécanismes de la latence et de la réactivation sont encore mal connus. Lors de la phase de latence, l'ADN viral subit une réplication et une transcription limitées à un petit nombre de gènes viraux appelés gènes « latents ». Ainsi, le virus peut persister dans la cellule (et donc dans l'hôte) indéfiniment. Tandis que la primo-infection est souvent accompagnée d'une période limitée de maladie clinique, la latence à long terme est asymptomatique. Les virus latents peuvent se réactiver et

entrer dans un cycle répliatif à n'importe quel moment. Cette réactivation est un phénomène biologique bien connu mais il reste à en comprendre les mécanismes génétiques et biochimiques. Elle peut être asymptomatique ou entraîner une maladie.

La figure 2 suivante illustre la physiopathologie (simplifiée) de l'infection à CMV :

Figure 2 : Physiopathologie de l'infection à CMV (d'après Alain et Mazon en 2003 (Alain and Mazon, 2003))

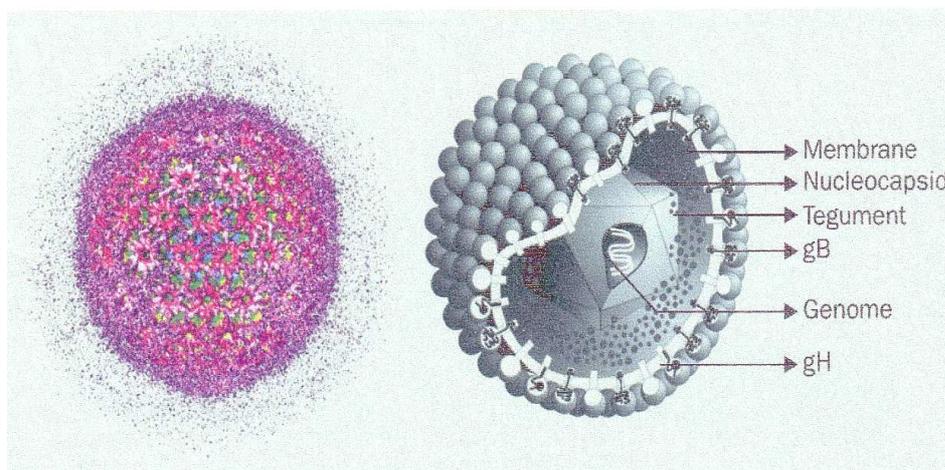


III. Structure du virus

L'appartenance à la famille des Herpesviridae est fondée sur la structure du virion (Figure 3).

Figure 3 : Structure du CMV (<http://www.bigrafix.del>)

Représentation du virion CMV à partir d'une photographie en microscopie électronique à balayage et structure schématique du CMV et des β herpesviridae

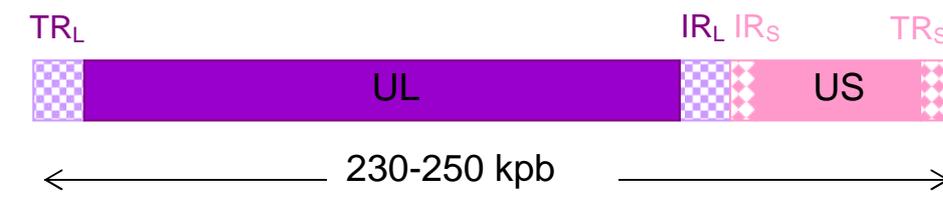


gB : glycoprotéine B ; gH : glycoprotéine H

Le virion (Figure 3) de 150 à 200 nm de diamètre est constitué de différents éléments :

- Le génome est le plus long et le plus complexe de tous les Herpesviridae avec un ADN bicaténaire linéaire de 230 à 250 kpb enroulé autour d'un noyau protéique (core). Il se circularise dans le noyau de la cellule infectée lorsqu'il est libéré de la capsid. Ce génome code potentiellement plus de 200 protéines (Davison *et al.*, 2003; Dolan *et al.*, 2004). Il est organisé en deux segments uniques : d'une part un segment unique long, U_L (Unit Long), correspondant à 85 % du génome et d'autre part, un segment unique court, U_S (Unit Short). Chacun de ces segments est flanqué à ses extrémités de séquences répétées et inversées terminales, les TRL (terminal long repeat) et les TRS (terminal short repeat). A l'intersection de ces deux segments se trouvent deux séquences internes, les IRL (long internal repeat) et les IRS (short internal repeat) (Figure 4).

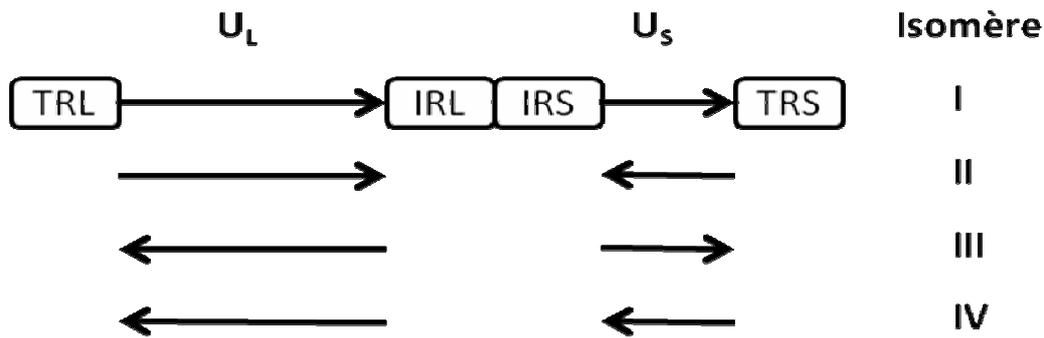
Figure 4 : Structure du génome du CMV



UL : long unit ; US : short unit; IRL: long internal repeat; IRS: short internal repeat; TRL: terminal long repeat, TRS: terminal short repeat

Quatre isoformes différentes du génome viral peuvent apparaître au cours de la réplication virale selon l'orientation des segments U_L et U_S (figure 5). La nomenclature des gènes et des protéines associe la désignation de la région codante (UL, US, TRL/IRL, TRS/IRS) et le cadre ouvert de lecture de façon séquentielle. La protéine porte l'indication de sa nature : pp pour phosphoprotéine, gp pour glycoprotéine.

Figure 5 : Les quatre isomères du génome viral du CMV



Dans la section UL, sept groupes de gènes conservés (I à VII) sont présents et l'origine lytique de réplication *oriLyt* se situe à proximité du groupe II.

La souche AD169 est la souche de laboratoire la plus étudiée. Elle a été entièrement séquencée et compte 230 288 paires de base (NC_001347). D'autres souches ont été entièrement séquencées comme Merlin (NC_006273), Towne (AC_146851) ou Toledo (AC_146905) (Chee *et al.*, 1990; Davison *et al.*, 2003; Dolan *et al.*, 2004; Dunn *et al.*, 2003).

Parmi les 200 cadres ouverts de lecture (ORF) identifiés, 45 codent des protéines essentielles pour la réplication virale dans des fibroblastes *in vitro*, même si leur rôle n'est pas identifié pour tous. Les autres ORF codent des protéines dites « non essentielles » pour la culture cellulaire *in vitro* et leur fonction demeure pour la plupart inconnue (Chee *et al.*, 1990; Dolan *et al.*, 2004; Dunn *et al.*, 2003).

- La capside composée de 162 capsomères présente une structure icosaédrique et mesure environ 100 nanomètres de diamètre. Elle est constituée d'au moins sept protéines (Mocarski and Courcelle, 2001), dont la Major Capsid Protein (MCP, pUL86, 150 kDa) constituant les hexons et les pentons (les capsomères). L'espace entre eux est occupé par la minor Capsid Protein (mCP, pUL85) et par la minor Capsid Binding Protein (mC-BP, pUL46, 34 kDa).
- Le tégument ou matrice, est situé entre la capside et l'enveloppe et maintient leur cohésion. Il est constitué d'au moins 39 protéines (Gibson, 1996; Kalejta, 2008). Les plus importantes sont pp65 (UL83), pp150 (UL32), pp28 (UL99), pp71 (UL82), pUL47 et pUL48. En général, elles sont phosphorylées (souvent désignées avec le préfixe pp signifiant phosphoprotéine) et fortement immunogènes. Les protéines pp65

et pp150 semblent jouer un rôle important dans la régulation des gènes viraux et dans le contrôle du métabolisme cellulaire au cours de la réplication virale. La pp65 représente le constituant majeur des protéines du tégment et de la particule virale. Nous en reparlerons au chapitre gènes immunodominants.

La pp150 représente presque 1/10^{ème} de la masse protéique du virion (Varnum *et al.*, 2004).

L'expression du promoteur très précoce (major immediate early promoteur, MIEP) est activée par la pp71.

La pp28 est essentielle pour la production de virions infectieux en jouant un rôle dans l'assemblage des particules de l'enveloppe (Britt *et al.*, 2004; Jones and Lee, 2004; Seo and Britt, 2006; Seo and Britt, 2007).

Parmi les autres protéines du tégment, le cycle cellulaire est modifié par ppUL69 alors que pTRS1 et pIRS1 transactivent l'expression des gènes viraux. La terminase, qui convertit les concatémères d'ADN en monomère lors de l'encapsidation virale, est constituée de pUL89 (petite sous-unité) et pUL56 (grande sous-unité) (Thoma *et al.*, 2006).

Cependant, la fonction de la plupart des protéines du tégment demeure inconnue (Hayashi, Blankenship, and Shenk, 2000; Kalejta, 2008a; Kalejta, 2008b).

- L'enveloppe dérive des membranes internes cytoplasmiques et nucléaires. Elle confère au virion une sensibilité aux solvants des lipides, aux pH bas et à la chaleur. Elle est constituée de polyamines, de lipides et de glycoprotéines (le CMV code potentiellement plus de 60 glycoprotéines présentables sur l'enveloppe virale), dont les plus étudiées sont les glycoprotéines B, H, M/N, L et 48, respectivement notées gB, gH, gM/N, gL et gp48. Elles sont ancrées dans la membrane phospholipidique et émergent sous forme de spicules à la surface du virion. Elles permettent la fixation du virus aux récepteurs des cellules hôtes au cours de la réplication virale. Ces glycoprotéines constituent les principaux antigènes contre lesquels l'hôte est capable de répondre.

La glycoprotéine gB est codée par UL55. Associée en dimère, elle compose le complexe protéique gCI. Elle constitue le composant majeur de l'enveloppe et est essentielle pour la réplication. Elle semble indispensable dans le phénomène d'adhésion cellulaire (Britt and Mach, 1996), est présente sur toutes les membranes de la cellule infectée et est incorporée dans le virion lors du processus d'enveloppement

(Sanchez *et al.*, 2000). Elle intervient également dans la transmission cellulaire et la fusion des cellules adjacentes. Elle est très immunogène. Elle constitue une cible majeure pour les anticorps neutralisants et les cellules CD8⁺ et CD4⁺ (Pass, 2004).

Le complexe protéique gCIII est composé des glycoprotéines gH (UL75), gL (UL115) et gO (UL74). Elles facilitent le transport vers la surface cellulaire, influencent le tropisme cellulaire (Wang *et al.*, 1998) et sont essentielles pour l'entrée du virus dans la cellule.

Le complexe gCII est composé des glycoprotéines gM (UL100) et gN (UL73). La gM serait plus abondante que la gB (10% des protéines totales virales) (Varnum *et al.*, 2004). Ces protéines ne sont pas essentielles pour la réplication virale *in vitro*.

- Les protéines de l'hôte : dans le virion plus de 70 protéines humaines ont été identifiées. Il s'agit de protéines structurelles cellulaires (α - et β - actine, tubuline, α -actinine...), des protéines cellulaires (impliquées dans la traduction, incluant des facteurs d'initiation et d'élongation), d'enzymes et de chaperonnes. D'autres protéines sont impliquées dans le trafic vers le réticulum endoplasmique et vers le Golgi, suggérant un rôle dans l'enveloppement du virion et sa progression cellulaire (Varnum *et al.*, 2004). La présence d'une enzyme cellulaire serait nécessaire à l'activité kinase de la pp65 (Roby and Gibson, 1986). Ces protéines cellulaires font partie du virion infectieux et seraient localisées dans le tégument (Kalejta, 2008).
- Deux classes d'ARN sont présentes dans le virion :
 - L'initiation de la réplication est permise grâce à la présence de 2 ARN hybridés avec l'origine de réplication *oriLyt* associés avec UL84 (Colletti *et al.*, 2007; Prichard *et al.*, 1998).
 - Dans le tégument, des ARNm (ARN messenger) ont été identifiés : transcrits très précoces (UL106 à 109), précoces (TRL/IRL2 et 5, TRL/IRL7 et 13) et tardifs (UL21.5). Ces transcrits sont fortement exprimés tardivement au cours de l'assemblage du virion. Deux autres ARN viraux (UL65 et UL123) et deux ARN cellulaires (celui codant la glyceraldéhyde-3-phosphate déshydrogénase (GAPDH) et U1A) sont également trouvés dans le virion (Greijer, Dekkers, and Middeldorp, 2000). Ces ARN sont libérés dans les cellules nouvellement infectées dès l'entrée du virus et permettent une expression des gènes viraux en absence de la

transcription du génome viral (Bresnahan and Shenk, 2000; Terhune, Schroer, and Shenk, 2004).

Dans la cellule infectée, de nombreuses protéines codées par le génome viral mais non assemblées dans la particule, sont présentes. Parmi celles-ci, se trouvent diverses enzymes impliquées dans la réplication virale et des protéines participant à l'échappement immunitaire.

IV. Le cycle de réplication viral

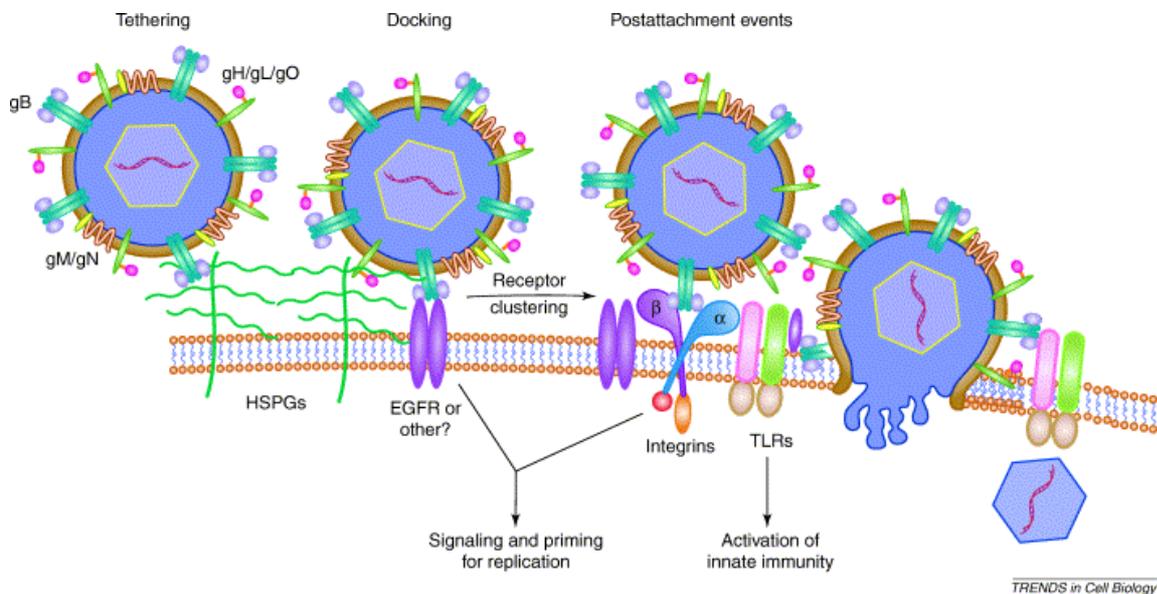
De très nombreuses cellules permettent la réplication du CMV dans l'organisme. Pour la culture cellulaire, on utilise principalement les fibroblastes humains car ce sont les seuls à permettre une production élevée de virions. Ce sont donc des cellules de choix pour l'étude de la multiplication virale et le diagnostic. Le cycle répliatif du CMV en culture cellulaire est lent (48h à plusieurs semaines pour les isolats cliniques) par rapport au temps de doublement de la charge virale circulante chez un patient immunodéprimé (0,5 à 2 jours) (Pass, 2004).

Dans les cellules infectées, trois types de particules virales sont trouvées : les virions complets infectieux, les particules virales non infectieuses (sans ADN) et les corps denses (principalement des protéines du tégument dont la pp65 représente 60% des protéines, enveloppées d'une membrane cytoplasmique).

A. L'attachement et la pénétration du virus

La fixation et la pénétration du virus dans la cellule hôte sont un processus très rapide. Elle implique la participation d'un grand nombre de récepteurs cellulaires encore mal connus (Evers, Wang, and Huang, 2004). En 2004 Teresa Compton a proposé un mécanisme (Compton, 2004) illustré dans la figure 6. Les glycoprotéines gM/gN et gB interagiraient avec les molécules d'Heparane Sulfate (HSPG). Ces interactions seraient suivies de l'attachement du virus à des récepteurs, d'un recrutement de co-récepteurs puis de la fusion de l'enveloppe virale et de la membrane cytoplasmique, avec libération de la capsid dans le noyau.

Figure 6: Mécanisme général d'attachement et de pénétration du CMV d'après Teresa Compton, 2004



Il n'existe pas de récepteur spécifique pour le CMV. Les glycoprotéines gB et gH se lieraient de manière spécifique au récepteur membranaire EGF-R (epithelial growth factor) (Compton, Nowlin, and Cooper, 1993; Wang *et al.*, 2005; Wang *et al.*, 2003), et de manière aspécifique à l'annexine II ou à la $\beta 2$ microglobuline. Elles seraient également reconnues par les TLR2 (Toll Like Receptor 2) présents sur les monocytes (Isaacson, Feire, and Compton, 2007), entraînant la production de cytokines pro-inflammatoires comme IL-8 (Boehme, Guerrero, and Compton, 2006; Compton *et al.*, 2003), ou le recrutement de co-récepteurs comme les intégrines $\alpha\beta 3$ dont l'interaction avec EGR-F serait essentielle pour la transmission du signal (Wang *et al.*, 2005). Cependant, pour certains auteurs le récepteur EGF-R est essentiel (Bentz and Yurochko, 2008) et pour d'autres non (Isaacson, Feire, and Compton, 2007).

Cette fixation permet une activation très précoce du métabolisme cellulaire dont les facteurs de transcription comme NF- κ B, et le rapprochement du virus de la membrane phospholipidique. Le virus pénètre par endocytose. Ceci aboutit à la libération de la capsidie et des protéines du tégument dans le cytoplasme cellulaire. La capsidie virale est transportée dans le noyau où l'ADN est libéré, circularisé et l'expression des gènes très précoces commence. La protéine du tégument pp65 est immédiatement dirigée vers le noyau et se fixe à la matrice nucléaire pour interférer avec le métabolisme cellulaire : on a ainsi une réorganisation du noyau, avec condensation de la chromatine, ce qui prépare la réplication virale (figure 7A). Le

cycle de réplication viral (figure 7) s'effectue exclusivement sur des cellules en phase S de croissance avec une grande spécificité d'hôte. Il a été étudié *in vitro* dans des fibroblastes embryonnaires.

B. La synthèse des protéines virales et réplication de l'ADN

La protéine du tégment pp71, en activant l'expression du promoteur majeur très précoce, le MIEP, induit l'expression en cascade des gènes viraux en 3 phases successives : très précoce α (ou IE pour Immediate Early), précoce β (ou E pour Early) et tardive γ (ou L pour Late) (Pass, 2004). Chaque phase contrôle la suivante. La transcription des gènes viraux est assurée par l'ARN polymérase II cellulaire.

1. La phase très précoce IE (Immediate Early, de H0 à H2) :

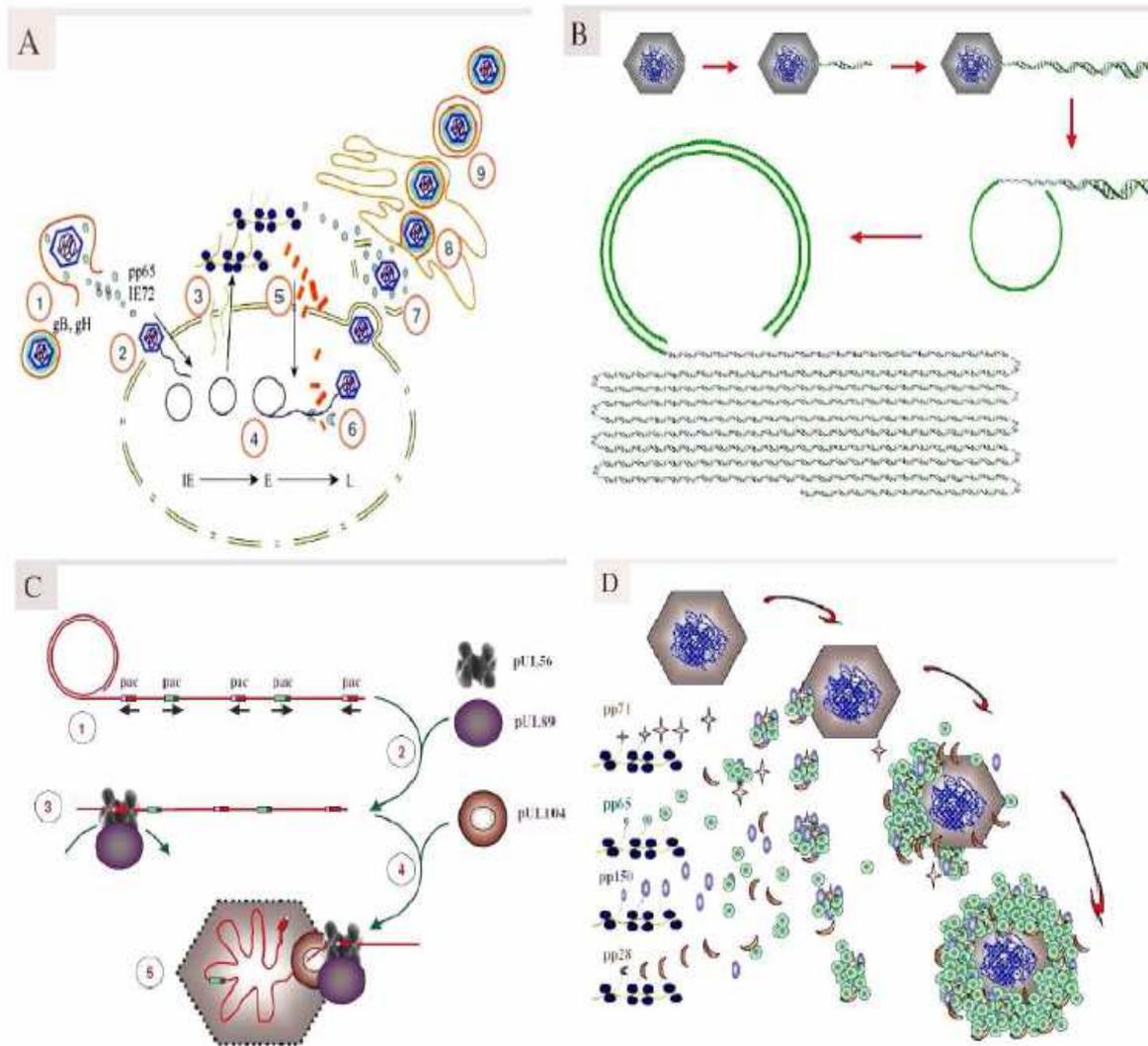
Elle correspond à l'expression des gènes très précoces du virus (gènes IE1 (UL123 codant IE72) et IE2 (UL122 codant IE86), 80% des gènes IE) sous contrôle d'un promoteur majeur très puissant le MIEP. Les protéines très précoces IE86, IE72 et IE55 sont des transactivateurs positifs et négatifs des gènes viraux et cellulaires et régulent leur propre expression. Leur expression permet le détournement du métabolisme cellulaire au profit de la réplication virale, l'inhibition de l'ADN cellulaire et le déclenchement de la phase précoce.

2. La phase précoce E (Early, de H12 à H24)

Elle correspond à l'expression des gènes E, qui permettent la production des protéines impliquées dans la réplication de l'ADN viral, notamment l'ADN polymérase (UL54) et sa protéine accessoire de processivité (UL44), une hélicase et une DNase. La synthèse de l'ADN viral se fait le modèle du cercle roulant, à partir de l'origine de réplication *oriLyt* (figure 7B) pour laquelle UL84 et le cluster UL112-113 sont requis.

Le génome viral commence sa réplication

Figure 7 : Cycle de réplication du CMV ; détails des étapes de la synthèse d'ADN à la tégumentation (Alain and Mazon, 2001).



A : Représentation simplifiée du cycle viral : le cycle viral se déroule en 8 grandes étapes : (1) Fusion du virion et entrée de la capsidie par endocytose, (2) : translocation nucléaire de l'ADN viral et circularisation de l'ADN ; (3) : phase IE et E ; (4) : Réplication de l'ADN selon le modèle du cercle roulant et phase L ; (5 et 6) : Maturation et encapsidation de l'ADN viral, détaillé en C ; (7) : tégumentation : enveloppement des virions, (détaillé en D), transit intra cellulaire et acquisition des enveloppes ; (8) : sortie des virions néoformés par exocytose

B : Réplication de l'ADN selon le modèle du cercle roulant. Après libération et translocation dans le noyau, l'ADN viral est circularisé et répliqué en une multitude de copies assemblées bout à bout formant une longue molécule d'ADN double brin, nommée concatémère. Chaque unité de génome est flanquée de séquences répétées et inversées, riches en bases AT, les séquences pac, qui sont des signaux de clivage et d'encapsidation.

C : Encapsidation de l'ADN viral. (12) : la reconnaissance de séquences pac, par UL56 permet le premier clivage de l'ADN et la fixation du complexe UL56/ADN à pUL89. Puis le complexe UL56/UL89 se fixe à la protéine pUL104 de la capsidie, qui transfère l'ADN dans la capsidie. Le complexe UL56/UL89 réalise le second clivage de l'ADN alors libéré dans la capsidie puis se dissocie pour permettre une nouvelle encapsidation

D : Tégumentation des capsides : Les protéines du tégument : pp28, pp71, pp150 et pp65 forment des amas qui s'agrègent autour des capsides pour former le tégument : l'ensemble sera recouvert d'une enveloppe d'origine cellulaire lors de l'entrée des capsides dans le réticulum endoplasmique puis extériorisé par exocytose.

3. La phase tardive L (Late)

Durant cette phase, ce sont les gènes L qui sont transcrits en protéines de structure du virion. La réplication de l'ADN viral se prolonge durant la phase tardive, puis a lieu sa maturation, son encapsidation dans les capsides néoformées (figure 7C) puis la tégumentation des virions. Les protéines du tégument sont acquises dans le cytoplasme, l'enveloppe définitive et les protéines virales d'enveloppe sont acquises au niveau de l'ergastoplasme, et /ou l'appareil de Golgi, où a lieu la glycosylation de la membrane virale.

Le virus peut être observé sous plusieurs formes dans la cellule :

- Le virion complet
- Les corps denses, résultant d'une production excessive de protéines virales, constitués uniquement de capsides et de tégument.
- Les particules enveloppées non infectieuses (sans ADN).

C. La libération du virus

Les particules enveloppées sortent de la cellule par fusion des vésicules golgiennes avec la membrane plasmique, 120 heures après l'infection en culture cellulaire (figure 7D) (Alain and Mazon, 2001; Pass, 2004). La libération des virions matures s'accompagne de la mort de la cellule hôte. Dans certaines cellules, le virus peut se maintenir à l'état latent.

V. Réponse immunitaire

A. Réponse immunitaire innée

La réponse immunitaire dirigée contre l'infection à CMV est complexe. Elle fait intervenir en premier lieu **l'immunité innée**, avec les cellules natural killer (NK), les monocytes/macrophages, les cellules dendritiques et l'activité antivirale des interférons alpha et bêta. Elle constitue une première barrière à l'infection virale.

Les cellules NK possèdent de nombreux récepteurs dont certains sont sollicités au cours de l'infection à CMV :

- des récepteurs inhibiteurs, avec par exemple :
 - les Leucocytes Ig-like receptors (LIR). LIR-1 reconnaît la protéine UL18 du CMV

- les lectines CD94/NKG2A et NKG2B. Elles reconnaissent la molécule non classique HLA-E qui présente des peptides dérivés des séquences leaders d'autres molécules HLA, mais qui peut aussi présenter un nonapeptide codé par le gène UL40 du CMV.

- des récepteurs activateurs dont :

- les lectines CD94/NKG2C, E et H
- les NCR
- NKG2D

Les monocytes/macrophages produisent des cytokines et des chimiokines, ils permettent la dissémination virale (les monocytes hébergent le génome viral latent et la réplication peut avoir lieu dans les macrophages).

Les cellules dendritiques possèdent des récepteurs du CMV (molécules DC-SIGN, DC-Specific ICAM Grabbing Non-integrin) et des Toll Like Receptors (TLR). Après capture via leur récepteur d'un antigène viral, les cellules dendritiques matures vont présenter les antigènes aux lymphocytes T par les molécules CMH-I ou II sans être infectées (Guermontprez *et al.*, 2002), initiant la réponse innée et adaptative.

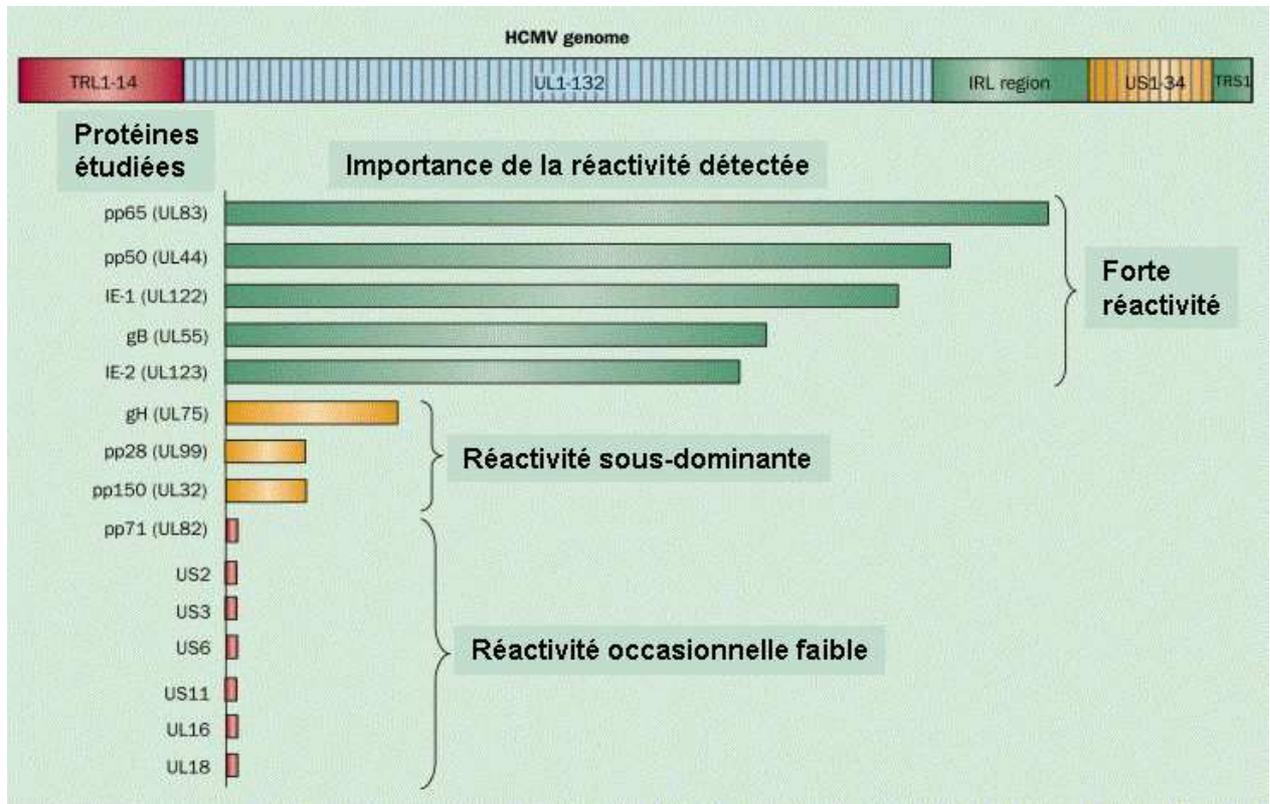
B. Réponse immunitaire adaptative

L'immunité adaptative médiée par les CD4+ entraîne l'apparition d'INF γ dans la circulation sanguine, suivie par la production d'anticorps spécifiques et l'apparition des lymphocytes T CD8+.

La réponse cellulaire spécifique est dirigée essentiellement contre les glycoprotéines de surface gB et gH, des protéines très précoces comme IE1, IE2 et pUL69 et contre la protéine du tégument pp65 (Kondo *et al.*, 2004; Le Roy and Davignon, 2005; Prod'homme *et al.*, 2003a; Prod'homme *et al.*, 2003b; Provenzano *et al.*, 2003; Retiere *et al.*, 2000). Les acteurs de cette réponse sont les lymphocytes T CD4+ et cytotoxiques CD8+ (figure 8).

La réponse humorale est dirigée contre les protéines du tégument, des protéines non structurales (pp52, IE1, IE2) et surtout contre les glycoprotéines d'enveloppe gB et gH. Les anticorps interviendraient lors de l'attachement, de la fusion avec les membranes plasmiques et de la pénétration du virus en empêchant la liaison du virion à la cellule (Burton, Saphire, and Parren, 2001; Gicklhorn *et al.*, 2003).

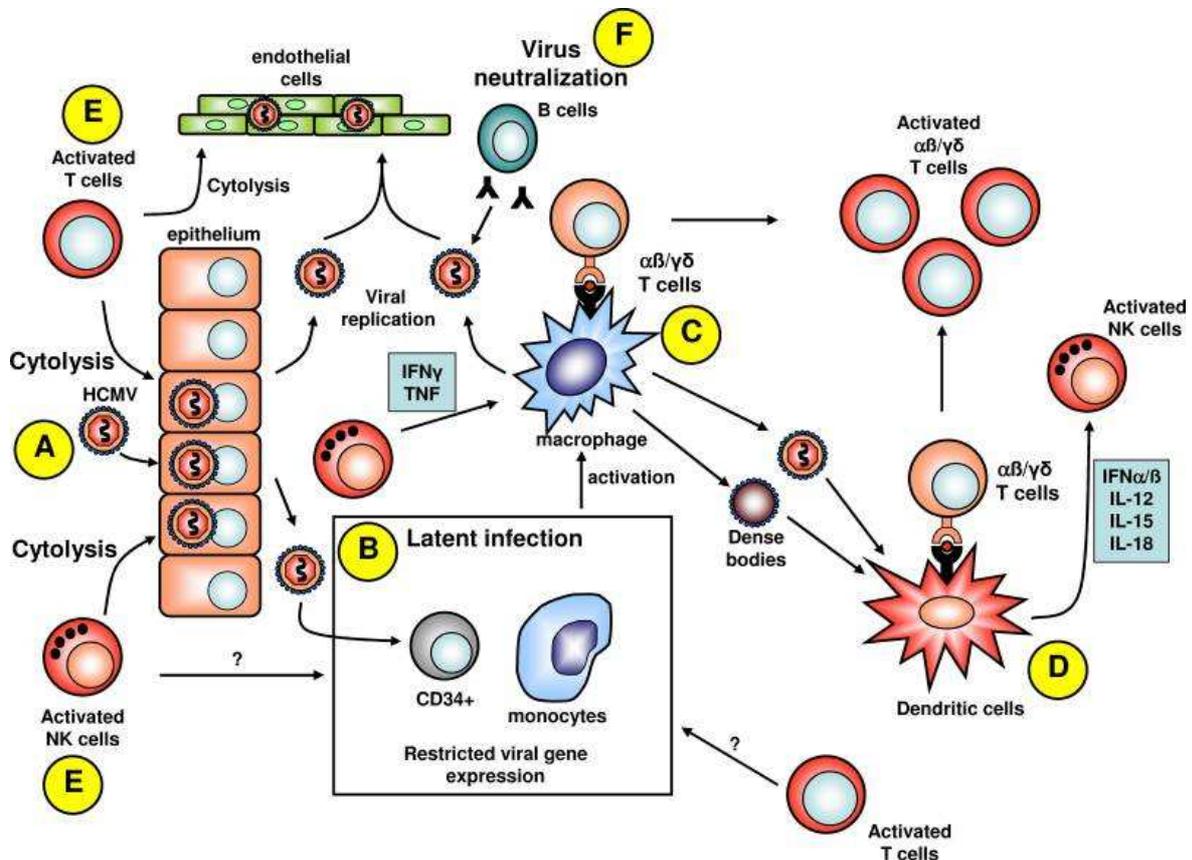
Figure 8 : Représentation schématique des antigènes viraux reconnus par la réponse immunitaire médiée par les cellules T CD4 et CD8 d'après Gandhi et Khanna (Gandhi and Khanna, 2004)



C. Schéma récapitulatif de la réponse immunitaire anti-CMV

En 2009, Crough et Khanna ont proposé le schéma suivant (figure 9), récapitulant le contrôle du CMV par l'immunité innée et adaptative.

Figure 9 : Contrôle du CMV par l'immunité innée et adaptative (Crough and Khanna, 2009). L'infection primaire par le CMV chez les individus en bonne santé commence typiquement dans les épithéliums des muqueuses (A), après quoi le virus diffuse dans les cellules myéloïdes incluant les monocytes et les cellules CD34+, où il établira sa latence (B). Une expression restreinte des gènes viraux est observée dans ces cellules infectées de manière latente ce qui limite leur reconnaissance par les cellules effectrices. La différenciation de ces monocytes infectés en macrophage ou cellule dendritique peut initier l'infection productive (C). Les particules virales ou les corps denses peuvent être présentés par les cellules dendritiques pouvant alors stimuler les cellules T spécifique de l'antigène (D). De plus, les cellules dendritiques activées via les TLR peuvent aussi sécréter un panel de cytokines activant les cellules effectrices de la réponse innée (D). Les macrophages infectés peuvent aussi activer directement les lymphocytes T spécifique d'antigène (C). Ces lymphocytes T activés (CD8, CD4 et gd) et les cellules NK peuvent directement lyser les cellules infectées par cytolysse ou bloquer la réplication virale au travers de la sécrétion de cytokines telles que l'IFN- γ ou encore du TNF (E). Une autre réponse importante est médiée par les lymphocytes B activés qui vont contrôler les virions libres dans le milieu extracellulaire par le biais de la production d'anticorps neutralisants (F). Une autre réponse importante est médiée par les lymphocytes B activés qui vont contrôler les virions libres dans le milieu extracellulaire par le biais de la production d'anticorps neutralisants (F).



VI. Infection par le CMV chez le patient transplanté

Un organe transplanté peut contenir du virus à l'état latent. Ainsi, un patient transplanté peut présenter une primo-infection à CMV (receveur séronégatif pour le CMV : R- ; donneur séropositif : D+), une réactivation endogène (R+) ou exogène (réactivation ou surinfection à partir du virus transmis par le donneur). Plusieurs souches virales peuvent donc circuler chez un même patient, à l'occasion de surinfections ou co-infections (D+ / R+). Ceci a été démontré par l'analyse de divers gènes viraux chez des patients suivis à distance de la greffe (Aquino and Figueiredo, 2000; Meyer-Konig *et al.*, 1998).

De plus, la réplication virale peut être compartimentée, pouvant conduire à une évolution différente des populations virales dans plusieurs sites de l'organisme, comme le sang périphérique et le poumon ou le sang périphérique et le cerveau. La compartimentation pourrait jouer un rôle essentiel lors de l'adaptation du virus à son environnement. Il a été montré l'existence de variants résistants aux anti-viraux dans le sang total d'un patient alors que le liquide céphalo-rachidien hébergeait des variants sensibles aux traitements malgré une charge virale élevée (Miller *et al.*, 2006).

VII. Echappement immunitaire

La réponse immunitaire complexe dirigée contre le CMV fait appel en premier lieu à l'immunité innée, par l'intermédiaire des macrophages, des cellules Natural Killer (NK) et dendritiques. Les réponses cellulaires spécifiques reposent sur les lymphocytes T CD8+ et CD4+. Des épitopes sont décrits sur de nombreuses protéines du CMV (Britt and Mach, 1996).

Le CMV est avant tout remarquable par sa capacité à déjouer le système immunitaire, en partie par la limitation de l'expression protéique virale pendant la phase de latence, par les nombreux gènes viraux codant des protéines immuno-modulatrices, et par une certaine variabilité génomique. Pour assurer une persistance à long terme, le CMV a élaboré de nombreuses stratégies de modulation et d'évasion immunitaire, impliquant pour cela plus de 20% de ses gènes.

A. Gènes immunomodulateurs

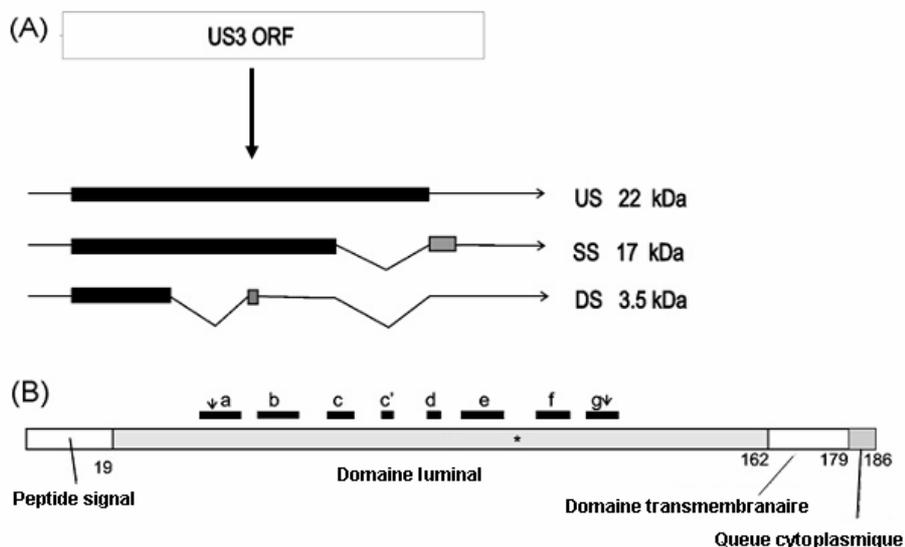
Les protéines virales interférant avec le système immunitaire interviennent à différents niveaux (Mocarski, 2002) :

- Altération de la présentation des peptides viraux aux lymphocytes T (séquestration, clivage ou inhibition du transport des protéines des complexes majeurs d'histocompatibilité (CMH) I et II).
- Interférence avec la fonction des cellules NK (par l'intermédiaire de leurs récepteurs) et le comportement des leucocytes (par exemple pseudo-récepteurs de chimiokines).
- Modification du cycle cellulaire et inhibition de l'apoptose.

Notre travail s'intéresse plus particulièrement à quatre de gènes codant des protéines immunomodulatrices :

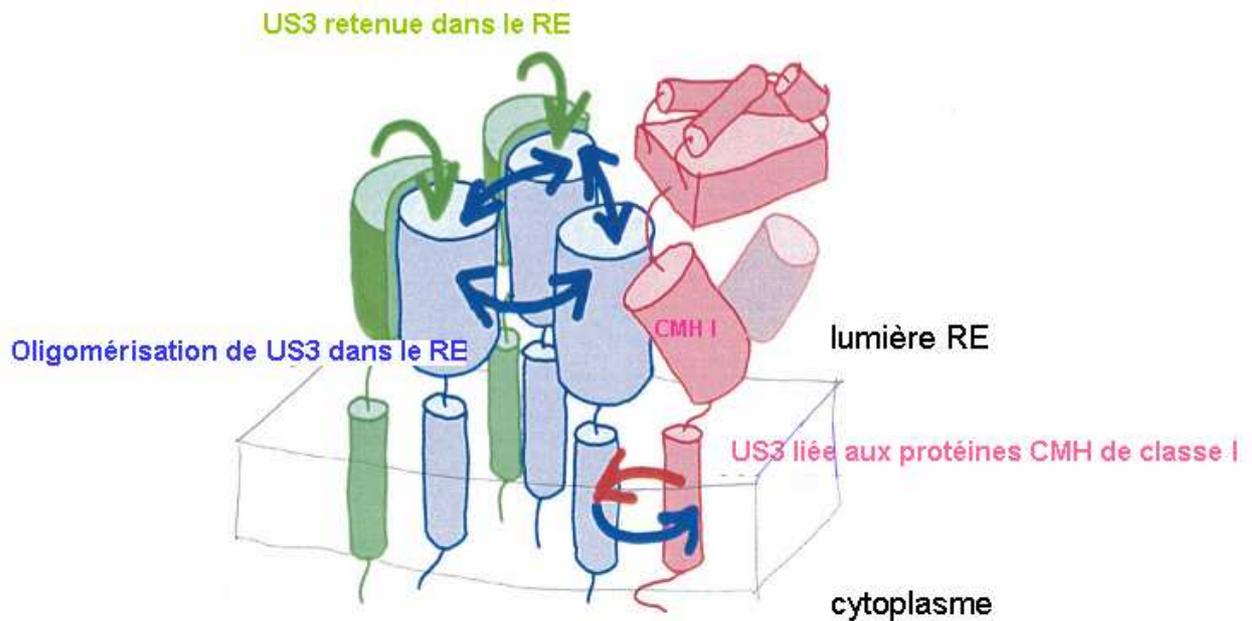
1. **US3** est un gène très précoce, impliqué dans l'évasion immunitaire. Son expression entraîne la rétention des molécules du CMH-I dans le réticulum endoplasmique, interférant avec la présentation de l'antigène à la surface des cellules exprimant US3 (Liu, Winkler, and Biegelke, 2009). L'expression est contrôlée par des éléments régulateurs activateurs ou inhibiteurs de la transcription associés à un réseau complexe de protéines. Les transcrits sont détectables une heure après l'infection, avec un pic à trois heures suivi d'une chute brutale cinq heures post infection. Son transcrit primaire peut subir un épissage alternatif conduisant à la génération de trois transcrits : un ARN non épissé de 22 kDa, le premier à apparaître après le début de l'infection ; un ARN épissé une fois codant une protéine de 17 kDa ; un ARN doublement épissé générant une protéine de 3,5 kDa dont nous n'avons pas encore la preuve biochimique de son existence contrairement aux deux premières. La protéine US3-22kDa est localisée dans le réticulum endoplasmique. Elle est composée d'un peptide signal, d'un domaine luminal, d'une région transmembranaire de 20 acides aminés suivie d'une courte queue cytoplasmique (figure10).

Figure 10 : Organisation des transcrits du gène US3 (Liu, Winkler, and Biegelke, 2009)



La structure du domaine luminal contient sept feuillets β repliés avec un pont disulfure conservé entre les feuillets A et G. (figure 11).

Figure 11 : Représentation schématique de la protéine gpUS3 (Misaghi *et al.*, 2004).

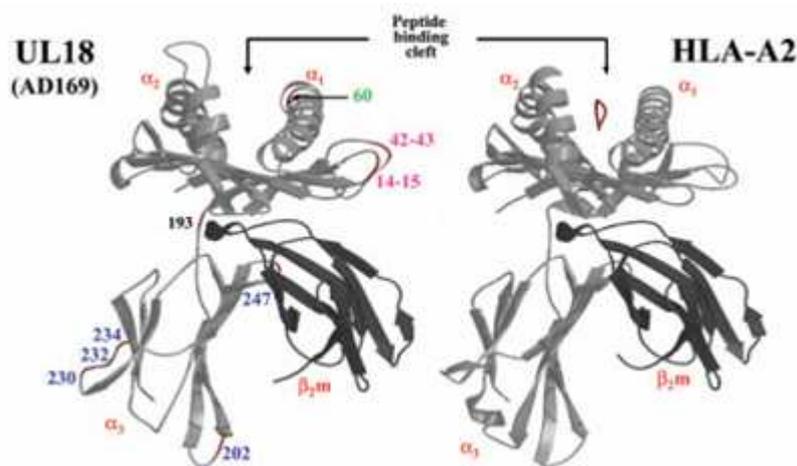


Cette protéine s'oligomérisise *in vivo* et *in vitro* et se lie à la chaîne lourde des molécules du CMH-I portant le peptide viral chargé et les retient dans le réticulum endoplasmique, retardant ainsi la présentation de l'antigène viral à la surface des cellules infectées. Elle se lie aussi à TAP (Transporter associated with antigen processing), une tapasine (TAP binding protein) et à des PDI (protein disulfide isomerase), membres du complexe chargeant les peptides viraux au CMH-I. L'interaction entre pUS3-22kDa et la tapasine interfère avec le chargement et l'assemblage du CMH-I entraînant un retard de maturation. La fonction des PDI est de réguler l'état d'oxydation des molécules du CMH-I, facilitant la sélection des peptides optimaux à la formation du complexe CMH-I. L'expression de pUS3-22kDa a pour conséquence une baisse du taux des PDI et une augmentation de molécules du CMH-I réduites interférant ainsi avec le chargement du peptide. La protéine pUS3-17kDa possède 134 acides aminés en commun avec celle de 22 kDa sans sa région transmembranaire. Elle est localisée dans le réticulum endoplasmique, les vésicules

cytoplasmiques et l'appareil de Golgi, suggérant une prise en charge par la voie de sécrétion. Elle ne se lie pas aux molécules du CMH-I. Dans les cellules exprimant à la fois pUS3-22kDa et pUS3-17kDa, davantage de molécules du CMH-I arrivent à maturation en comparaison avec les cellules n'exprimant que pUS3-22kDa. La protéine pUS3-17kDa se lie aussi à la tapasine entraînant une compétition entre les deux protéines : pUS3-22kDa bloque l'interaction de la tapasine avec le système TAP ; cette interaction est restaurée après addition de pUS3-17kDa. La compétition entre les deux protéines suggère que le gène US3 code un autorégulateur de sa fonction dans l'échappement immunitaire. De plus, sa séquence très conservée aussi bien dans les souches de laboratoire que dans les souches cliniques indiquerait un rôle important dans l'infection.

2. **UL18** code une protéine agissant à la fois comme un homologue du CMH-I qui interfère avec les monocytes, les lymphocytes B, et les cellules NK et comme un leurre des cellules NK. Cette protéine est composée d'une queue cytoplasmique, d'une région transmembranaire et de trois domaines extracellulaires (α_1 , α_2 et α_3) avec approximativement 25% d'homologie avec les molécules classiques du HLA-I (human leukocyte antigen). Elle s'associe avec la β_2 -microglobuline, et la stabilité du complexe dépend de la présence d'un peptide signal (figure 12).

Figure 12 : Modèle moléculaire de l'homologue du CMH-I, gpUL18 (Cerboni *et al.*, 2006) : Comparaison des domaines extra-cellulaires de l'homologue codé par AD169 à gauche avec ceux de HLA-A2, à droite (Peptide binding cleft : site de liaison du peptide).

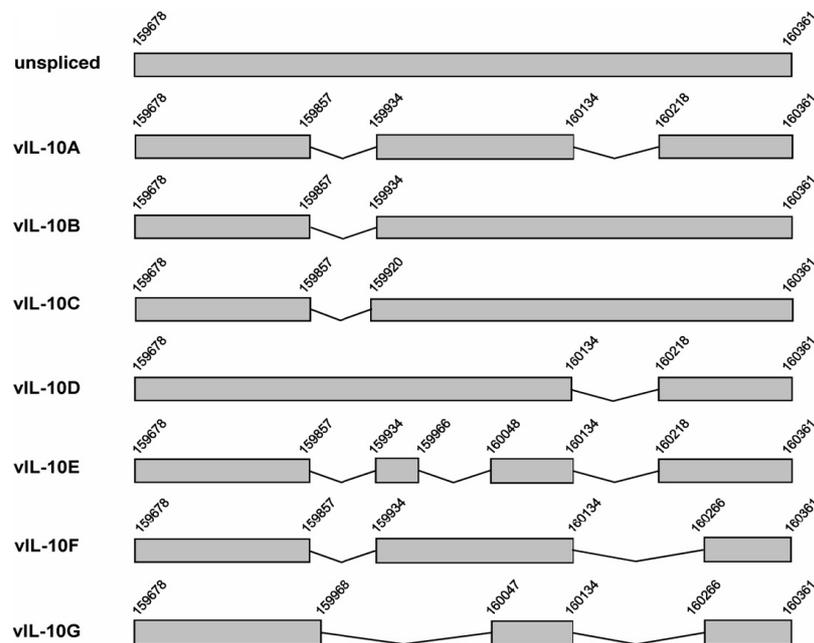


UL18 agit en se liant spécifiquement au récepteur inhibiteur CD85j/LIR-1/ILT2 (LIR-1 : Ig-like Leukocyte Inhibitory Receptor), une molécule transmembranaire exprimée par différentes cellules immunitaires, comme les monocytes, les cellules dendritiques, les lymphocytes B aussi bien qu'un sous-ensemble de cellules NK et de lymphocytes T. LIR-1 reconnaît les molécules du CMH de classe I mais son affinité pour pUL18 est 1000 fois supérieure. Ce contact étroit est influencé d'une part par l'expression à la surface cellulaire de pUL18 (deux motifs (YKVR et KVRS) présents dans la queue cytoplasmique, motifs semblent être responsables de la rétention cytoplasmique et la ré-internalisation de pUL18) et d'autre part par des mutations localisées dans les domaines extracellulaires $\alpha 1$ (K42 et A43) et $\alpha 3$ (Q202). De plus, quatre cystéines (C211, C240, C255 et C267) sont essentielles pour lier pUL18 et la $\beta 2$ -microglobuline, grâce à des ponts disulfures. Cette interaction est nécessaire pour lier pUL18 à son récepteur. (Cerboni *et al.*, 2006; Maffei *et al.*, 2008; Wagner, Ljunggren, and Achour, 2008).

3. La protéine codée par *UL111a* est un homologue viral de l'interleukine 10 (vIL-10) avec laquelle elle possède 27% d'identité (Kotenko *et al.*, 2000). La délétion de l'ORF UL111a n'a pas de conséquence sur la réplication virale *in vitro* (Chang *et al.*, 2004). La protéine vIL-10 présente différentes propriétés immunosuppressives : inhibition de la prolifération des cellules mononucléées sanguines et de la sécrétion de cytokines pro-inflammatoires, diminution du CMH-I et II à la surface des monocytes, interaction avec la maturation et la migration des cellules dendritiques (Chang *et al.*, 2004; Spencer *et al.*, 2002). Le gène UL111a possède deux introns et l'ARN pré-messager peut subir un épissage alternatif. L'épissage des deux introns entraîne la production d'une protéine (vIL-10A, figure 13) de 175 aa sécrétée durant la phase lytique de l'infection (Jenkins, Abendroth, and Slobedman, 2004; Kotenko *et al.*, 2000). Dans des cellules infectées en phase de latence, un transcrit plus court que vIL-10A a été mis en évidence et nommé LAcmvIL-10 (vIL-10B, figure 13). Depuis, il a été trouvé en phase d'infection productive de fibroblastes par des souches de laboratoire (Slobedman *et al.*, 2002). Un virus dépourvu de l'ORF UL111a est capable d'établir, maintenir et réactiver une phase de latence dans des cellules infectées. Cependant, les niveaux de CMH-II augmentent significativement à la surface de ces cellules, entraînant une reconnaissance plus efficace des cellules infectées par les cellules T CD4+. La protéine supprimerait la reconnaissance des cellules infectées en phase de

latence par les CD4⁺ constituant ainsi une stratégie d'évasion immunitaire en augmentant la capacité du CMV à persister à l'état latent dans l'hôte (Cheung *et al.*, 2009). Cinq autres isoformes ont été identifiées par Lin *et al* en 2008 (figure 13). Les différentes isoformes de gpUL111a sont identiques à l'extrémité N-terminale, pour les 60 premiers aa. Toutes peuvent établir un complexe avec l'IL-10 humaine, suggérant que le CMV a développé un mécanisme sophistiqué pour interférer sur la voie de signalisation de l'IL-10 (Cheung *et al.*, 2009; Jenkins *et al.*, 2008a; Jenkins *et al.*, 2008b; Lin *et al.*, 2008; Spencer *et al.*, 2008).

Figure 13 : Représentation schématique des transcrits vIL-10 (Lin *et al.*, 2008)



Les chiffres indiqués correspondent à la séquence ADN du génome de la souche AD169. L'ARN total a été purifié à partir de cellules MRC5 infectées. Après transcription inverse, l'ADN complémentaire a été amplifié, cloné et séquencé.

4. **UL40** contribue à protéger les cellules infectées par le CMV de la lyse médiée par les cellules NK. L'expression à la surface des molécules HLA-E dépend du chargement d'un peptide dérivé de séquences leaders présentes dans d'autres molécules du HLA de classe I. HLA-E chargé se lie essentiellement au récepteur inhibiteur CD94/NKG2A et inhibe la lyse cellulaire médiée par les NK. La séquence signal de pUL40 contient une séquence de 9 acides aminés (VMAPRTLIL) strictement homologue de celle du peptide chargé par HLA-E et dérivant des molécules HLA-Cw.

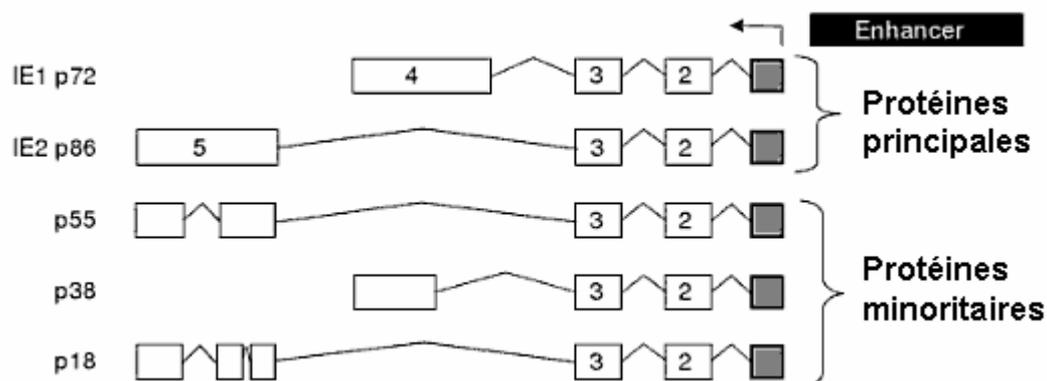
Cette interaction entre HLA-E et le nonapeptide UL40 augmente l'expression à la surface de HLA-E, permettant l'inhibition des cellules NK et de la lyse des cellules infectées. Ceci est d'autant plus important que les cellules NK devraient être activées par la diminution de l'expression des molécules MHC-I induite par le CMV. HLA-E, chargé du nonapeptide, peut aussi se lier à un récepteur activateur, CD94/NKG2C, avec une moindre affinité (Wilkinson *et al.*, 2008). Nous parlerons plus longuement d'UL40 dans la seconde partie de notre travail.

B. Gènes immunodominants

Nous avons choisi d'étudier deux autres gènes, les gènes UL123 et UL83, qui codent respectivement une protéine très précoce IE1 et une protéine du tégument, la pp65. IE1 et IE2 sont largement considérés comme les régulateurs centraux d'une cascade d'expression de gènes viraux et cellulaires pendant la phase lytique de l'infection à CMV. De plus, IE1 et la phosphoprotéine pp65 sont les cibles majeures de la réponse cellulaire immune. Des épitopes immunodominants ont été décrits pour de nombreux types HLA (Slezak *et al.*, 2007).

- La protéine **IE1** : Les protéines IE1 et IE2 sont le produit de l'épissage alternatif du gène UL122-123. IE1 et IE2 représentent les protéines très précoces principales. Elles ont en commun les exons 2 et 3 mais diffèrent pour le plus grand, le 4 pour IE1, le 5 pour IE2 (figure 14). Ces protéines agissent en synergie pour activer l'expression des protéines virales précoces mais de façon antagoniste pour autoréguler le promoteur majeur MIEP (Cherrington, Khoury, and Mocarski, 1991; Cherrington and Mocarski, 1989; Liu, Hermiston, and Stinski, 1991; Pizzorno and Hayward, 1990; Stenberg *et al.*, 1990).

Figure 14 : Les protéines de l'épissage alternatif du transcrit primaire IE (Meier and Stinski, 2006)



La fonction des protéines minoritaires demeurent inconnues (Shirakata *et al.*, 2002). Pour prévenir l'apoptose, des protéines incluant pUL38, IE1, IE2 et l'ARN viral beta2.7 (produit de l'ORF TRL4) se lient au complexe respiratoire mitochondrial I, maintenant la production d'ATP et prévenant la mort cellulaire induite par la mitochondrie (McCormick, 2008; Reeves *et al.*, 2007; Zhao *et al.*, 2010).

IE1 et IE2 sont essentielles pour le déroulement du cycle viral. IE n'est pas essentielle dans l'infection productive mais est très importante pour le succès de la réplication. En 2001, Mocarski *et al* montrent qu'une délétion d'IE1 dans le génome viral a pour conséquence des perturbations dans la réplication à de faibles moi (multiplicity of infection) (Mocarski *et al.*, 1996). Leurs résultats indiquent également que des protéines virales transactivatrices, comme les protéines du tégument, peuvent compenser l'absence d'IE1 à de fortes moi. En 2001, la délétion d'IE2 entraîne l'incapacité à exprimer des gènes lytiques très précoces aboutissant à l'empêchement de la réplication virale (Marchini, Liu, and Zhu, 2001). Ces protéines sont localisées dans le noyau de la cellule infectée et leur fonction principale est la régulation de la transcription des gènes viraux précoces et tardifs. Pour cela, elles interagissent avec les histones déacétylases HDAC-1, -2 et -3, régulant ainsi l'acétylation des histones (Murphy *et al.*, 2002; Reeves *et al.*, 2005; Woodhall *et al.*, 2006) et l'assemblage des nucléosomes de l'ADN viral (Nitzsche, Paulus, and Nevels, 2008). Ces paramètres sont essentiels pour contrôler la transcription des gènes viraux dans le cycle lytique et la latence.

La protéine IE1 va stimuler de nombreux promoteurs viraux et cellulaires en interagissant avec des protéines cellulaires telles que CTF-1, SP1, E2F1 à 5, TAF_{II}130/TAF4, p107, HDAC-2, PML et Daxx (Meier, J.L., and M.F. Stinski. 2006)(Nevels, Paulus, and Shenk, 2004) mais aussi, grâce à une activité kinase, avec des facteurs de transcription comme Rb, E2F et c-jun (Pajovic *et al.*, 1997). Elle peut induire l'expression des sous-unités de NF- κ B qui est considéré comme le facteur de transcription nécessaire à l'activation du MIEP (Hooks *et al.*, 2006). En 2009, Khan *et al* démontrent que durant les phases précoces de l'infection dans les macrophages, le virus utilise les sous-unités p50/p65 alors que pendant la phase tardive de l'infection il utilisera les sous-unités p100/Bcl-3 (Khan *et al.*, 2009).

En 2006 l'équipe de Paulus démontre qu'IE1 fonctionne apparemment comme un antagoniste du signal de transduction de l'interféron de type I (Paulus, Krauss, and Nevels, 2006).

La SUMOylation d'IE1, c'est-à-dire la liaison covalente avec la protéine SUMO-1 (Small Ubiquitin-like Modifier) permet à IE1 de disloquer les domaines ND10 cellulaires pour empêcher leur rôle répressif sur la réplication virale (Muller and Dejean, 1999; Woodhall *et al.*, 2006). Elle facilite également l'expression d'IE2 et promeut la réplication virale (Nevels, Brune, and Shenk, 2004).

L'expression d'IE1 peut être inhibée par IE2 qui agit comme un transactivateur de promoteurs cellulaires et viraux (comme IE1, Meier and Stonski, 2006) mais aussi comme un trans-répresseur du MIEP en se fixant sur la séquence CRS (*cis*-repressor signal) située entre la TATA box et le site de départ du MIEP (Pizzorno and Hayward, 1990).

- La protéine **pp65** (pour revue Kalejta 2008(Kalejta, 2008)) : la phosphoprotéine pp65 est le constituant majeur de la particule virale et est codée par le gène UL83. Elle est dirigée vers le noyau des cellules infectées au tout début de la phase lytique de l'infection (Revello *et al.*, 1992). Elle est la cible majeure des lymphocytes T cytotoxiques (Grefte *et al.*, 1992; Jahn *et al.*, 1987; Kern *et al.*, 2002; Wills *et al.*, 1996). La persistance de cette protéine virale malgré la forte pression immune suggère qu'elle a un rôle très important durant le cycle viral. Étonnamment, le gène UL83 n'est pas essentiel pour la réplication dans des cultures de fibroblastes mais il est essentiel pour la formation des corps denses (Schmolke *et al.*, 1995). La protéine, qui posséderait une activité kinase, phosphoryle les protéines IE inhibant leur présentation par le CMH-I (Gilbert *et al.*, 1996). Elle entraîne l'accumulation des molécules HLA de classe II dans le lysosome (Odeberg *et al.*, 2003). Elle inhibe également les NK en interagissant avec le récepteur activateur NKp30 (Arnon *et al.*, 2005). La réponse interféron est également atténuée par la pp65 (Abate, Watanabe, and Mocarski, 2004; Browne and Shenk, 2003) comme par IE1 (Paulus, Krauss, and Nevels, 2006).

C. Variabilité du CMV

Des facteurs virologiques et environnementaux sont impliqués dans l'émergence et la ré-émergence de maladies virales. Les virus n'occupent pas une niche écologique unique et permanente. En raison de leur capacité intrinsèque aux modifications génétiques, et à leur haut niveau d'évolutivité, ils montrent un potentiel pour parasiter différentes espèces. Mutation, recombinaison, réassortiment de segments génomiques et combinaison de ces

événements moléculaires produisent des populations virales complexes et phylogénétiquement variées qui constituent la matière première sur laquelle la sélection agit. La majorité des maladies virales émergentes touchant l'homme a une origine animale. Des facteurs sociologiques et environnementaux induisent des changements environnementaux divers à l'occasion desquels les sous-populations virales ont de nombreuses occasions d'être sélectionnées à l'intérieur des populations virales hétérogènes, particulièrement dans le cas des virus à ARN. De cette façon, de nouveaux virus contaminant l'homme, les animaux ou les plantes ont émergé périodiquement, et continueront d'apparaître. Un des mécanismes identifié pour cette émergence virale est l'importance de la variabilité génétique des virus. Cette variabilité pourrait jouer un rôle dans l'échappement immunitaire notamment grâce à la présence de quasiespèces virales (Domingo, 2010; Domingo et al., 2006; Domingo et al., 1978; Domingo and Wain-Hobson, 2009; Luring and Andino, 2010). Domingo a défini la notion de quasiespèces virales comme étant la distribution dynamique de génomes soumis à une variation génétique, à la concurrence et à la sélection, et agissant comme une unité de sélection (Domingo *et al.*, 1978).

Une variabilité génétique notable est observée entre les souches de CMV de laboratoire (Towne, Merlin, AD169, Toledo). Des gènes non essentiels sont entièrement délétés dans certaines souches sélectionnées par culture cellulaire prolongée. Il existe moins d'informations concernant les souches cliniques, cultivées brièvement ou non cultivées. Plusieurs gènes viraux ont été étudiés, en particulier *UL55* codant la glycoprotéine B, aboutissant à une classification du CMV en 4 génotypes gB (Pignatelli *et al.*, 2004; Rasmussen *et al.*, 2002; Zwegberg Wirgart *et al.*, 1998).

La variabilité du CMV serait en partie due à des phénomènes de recombinaison, déjà montrés *in vitro* mais aussi *in vivo* et facilités par les infections multiples et simultanées chez un même patient (Haberland, Meyer-Konig, and Hufert, 1999). Ainsi, la présence de plusieurs variants viraux chez un patient greffé rénal a été étudiée par les premiers travaux sur le CMV du laboratoire de Virologie EA2968. L'analyse des quatre gènes immunomodulateurs cités précédemment chez plusieurs patients montre que certaines souches virales sont identiques pour un gène ou deux, mais différent pour les autres, ce qui rend chacune d'elle unique et est compatible avec l'hypothèse de la recombinaison (Garrigue et al., 2007b).

Pour les virus à génome ARN (virus de l'immunodéficience humaine, virus de l'hépatite C), la notion de quasi-espèce a été retenue pour décrire la famille de variants viraux qui circulent simultanément chez un individu infecté. L'existence de ces quasi-espèces pourrait jouer un rôle clé dans l'évolution de l'infection, en permettant par exemple au virus de

s'adapter à la réponse immunitaire de l'hôte, de modifier son tropisme cellulaire, ou en facilitant l'émergence de virus résistants aux médicaments antiviraux. Ce concept est peu admis à l'heure actuelle pour les virus à génome ADN, sauf pour le virus de l'hépatite B (qui possède une transcriptase inverse). Mais les résultats de notre équipe suggèrent qu'une évolution par quasi-espèce pourrait également exister pour le CMV. A notre connaissance, notre équipe a été la première à « oser » le terme de quasi-espèces pour qualifier le CMV. Depuis il a été utilisé par Kowalik *et al* (3rd Congenital CMV Conference, septembre 2010, Paris) pour décrire l'existence du CMV sous forme de quasiespèces chez des enfants infectés congénitalement.

Notre travail comporte deux parties principales, l'une portant sur l'analyse de la variabilité génomique du CMV, et l'autre, plus préliminaire, sur l'étude de la protéine pUL40.

Le travail présenté dans la première partie a fait l'objet de deux publications principales dans lesquelles je suis premier auteur ou coauteur :

1. UL40 human cytomegalovirus variability evolution patterns over time in renal transplant recipients. Garrigue I*, **Faure-Della Corte M***, Magnin N, Recordon-Pinson P, Couzi L, Lebrette ME, Schrive MH, Roncin L, Taupin JL, Déchanet-Merville J, Fleury H, Lafon ME. *Transplantation*. **2008** Sep 27;86(6):826-35 (annexe 1)

*Les deux premiers auteurs ont contribué également à ce travail

2. Variability and recombination of clinical human cytomegalovirus strains from transplantation recipients. **Faure-Della Corte M**, Samot J, Garrigue I, Magnin N, Reigadas S, Couzi L, Dromer C, Velly JF, Déchanet-Merville J, Fleury HJ, Lafon ME. *J Clin Virol*. **2010** Feb;47(2):161-9 (annexe 2)

Il a également été présenté durant des congrès :

- Human Cytomegalovirus (HCMV) UL40 variability in renal transplant recipients. ME. Lafon, **M. Faure-Della Corte**, N. Magnin, L. Couzi, J. Dechanet-Merville, P. Merville, H. Fleury, I. Garrigue. *Annual Meeting of the European Society for Clinical Virology (ESCV) 2009*, Amsterdam (NL).
- Human Cytomegalovirus (HCMV) compartmentalization in pulmonary transplant recipients. E. Lafon, **M. Faure-Della Corte**, J. Samot, N. Magnin, C. Dromer, E. Rosier, JF. Velly, Merville, H. Fleury, I. Garrigue. *Annual Meeting of the European Society for Clinical Virology (ESCV) 2009*, Amsterdam (NL) (Communication Orale).

- Variability and recombination of clinical human cytomegalovirus strains in transplantation recipients. **M. Faure-Della Corte**, J. Samot, I. Garrigue, N. Magnin, S. Reigadas, L. Couzi, JF. Velly, J. Déchanet-Merville, H. Fleury and ME. Lafon. *12th International CMV and Betaherpes virus Workshop 2009*, Boston (USA).

J'ai par ailleurs participé à d'autres études qui ont donné lieu à trois publications supplémentaires dont deux en cours de soumission et à une communication orale :

1. Is there an association between positive cytomegalovirus (CMV) plasma DNAemia and the risk of cancer among HIV-infected patients? A case-control study within the ANRS CO3 Aquitaine Cohort. G. Marcel, MA. Vandenhende, M. Bruyand, S. Geffard, ME. Lafon, **M. Faure-Della Corte**, E. Lazaro, M. Dupon, H. Fleury, F. Bonnet, R. Thiébaud and I. Garrigue, for the Groupe d'Epidémiologie Clinique du SIDA en Aquitaine (GECSA). *4th European Congress of Virology 2010*. Villa Erba Cernobbio, Como (Italie).
2. Is there an association between positive CMV plasma DNAemia and the risk of cancer among HIV-infected patients? A case-control study within the ANRS CO3 Aquitaine Cohort. M Bruyand, MA Vandenhende, G Marcel, S Geffard, ME Lafon, **M Faure-Della Corte**, E Lazaro, M Dupon, F Dabis, H Fleury, F Bonnet, R Thiébaud and I Garrigue; for the Groupe d'Epidémiologie Clinique du SIDA en Aquitaine (GECSA). **soumis** à *J. Clin. Virol.*
3. Hepatitis B virus genotypes : a retrospective survey in southwestern France, 1999-2004. P. Trimoulet, M. Boutonnet, M. Winnock, **M. Faure**, M.A. Loko, V. de Ledinghen, P.H. Bernard, L. Castera, J. Foucher, M. Dupon, J.M. Ragnaud, M.E. Lafon, P. Couzigou, F. Dabis, H. Fleury. *Gastroentérologie Clinique et Biologique*. **2007** Dec;31(12):1088-94
4. HCV protease variability and anti-HCV protease inhibitors resistance in HIV-HCV coinfecting patients. P. Trimoulet, C. Belzunce, **M. Faure**, L. Wittkop, S. Reigadas, M. Dupon, J.M. Ragnaud, H. Fleury, D. Neau **soumis** à *HIV Medecine*.

Première Partie

Variabilité génomique de souches de CMV isolées chez des patients transplantés

VIII. Introduction

Chez les patients immunodéprimés, comme les transplantés pulmonaires, le Cytomégalo virus humain ou CMV est encore responsable des plus fréquentes infections virales opportunistes.

Nous avons choisi quatre gènes codant des protéines immunomodulatrices, UL18, UL40, UL111a et US3 et deux gènes immunodominants UL83 codant pp65 et UL123 codant IE1. Les quatre gènes immunomodulateurs ont été sélectionnés car peu de séquences de souches cliniques sans étape préalable de culture cellulaire étaient disponibles quand nous avons débuté notre étude.

Le CMV est un virus à ADN décrit comme génétiquement très stable. Toutefois, des mutations (y compris de résistance) ont été observées. *In vitro*, la recombinaison virale a lieu quand deux souches de CMV co-infectent la même cellule et interagissent durant la réplication pour générer une descendance dont les génomes sont composés de segments génétiques obtenus à partir des souches parentales. Des recombinaisons homologues ont été décrites pour le HSV et l'EBV. Pendant la réplication de l'ADN, la polymérase peut sauter d'un brin à l'autre, comme cela a été montré pour le VIH : des séquences peuvent être échangées entre deux variants. Durant l'infection à CMV, la topoisomérase II cellulaire, présente dans le tégument viral, est activée. Elle joue un rôle essentiel dans la réplication de l'ADN, en favorisant l'apparition de cassures dans la double hélice d'ADN entraînant alors des interactions entre les différentes souches présentes dans une même cellule. *In vivo*, un mécanisme de recombinaison peut être présumé pour des populations virales de patients infectées par le CMV comme les personnes séropositives pour le VIH, les transplantés ou les nouveaux nés.

Différentes populations virales peuvent co-exister et évoluer de façon variable. C'est le cas lors des infections compartimentées. Verbraak *et al* en 1998 ont montré que la distribution des variants (sous-types gB) peut être différente entre compartiments sanguin et ophtalmique chez le patient séropositif pour le VIH et développant une infection à CMV (Verbraak *et al.*, 1998). Tarrago *et al* l'ont démontré en 2003 en comparant le compartiment sanguin et le liquide céphalo-rachidien (Tarrago, Quereda, and Tenorio, 2003). Au cours

d'une transplantation d'organe, des co- ou sur-infections peuvent avoir lieu notamment pour les greffés D+/R+. Ils sont infectés par une population virale et reçoivent *via* le greffon une autre population virale à l'état latent. Ils peuvent ainsi héberger à la fois des souches sensibles aux molécules antivirales et des souches résistantes (Ducancelle *et al.*, 2004). Les charges virales dans les différents compartiments peuvent être différentes (Chemaly *et al.*, 2004; Miller *et al.*, 2006; Ruell *et al.*, 2007) et des mutations de résistance peuvent être détectées dans un site et pas dans un autre (Miller *et al.*, 2006).

IX. Objectifs

Les objectifs de notre étude sont de :

- décrire le polymorphisme naturel des gènes viraux *UL18*, *UL40*, *UL83*, *UL111a*, *UL123* et *US3* pour des souches de CMV isolées chez des patients transplantés pulmonaires ou rénaux présentant une infection à CMV;
- chez les patients transplantés pulmonaires, rechercher une compartimentation de l'infection virale en comparant différents échantillons biologiques à la même date, le sang et le liquide de lavage broncho alvéolaire (LBA) et d'évaluer l'évolution des variants viraux dans le temps (suivi longitudinal pour un patient et le gène *UL40*) ;
- apporter des arguments d'une évolution par recombinaison des souches virales pour des souches de CMV isolées chez des patients transplantés rénaux ou pulmonaires en comparant *in silico* les séquences des gènes viraux *UL18*, *UL40*, *UL111a*, *UL123* et *US3*.

X. Méthodes

A. Extraction de l'ADN

L'ADN est extrait à partir de 200µl de sang total ou de LBA en utilisant l'automate MagNA™ Pure LC (Roche Molecular Biochemicals) avec la trousse MagNA™ Pure LC Total Nucleic Acid Isolation (Roche Diagnostics) selon les recommandations du fournisseur. L'élution s'effectue dans 100µL.

B. Amplification des gènes viraux par PCR

En accord avec les alignements des séquences de CMV disponibles dans la base de données Genbank, des amorces spécifiques ont été conçues et utilisées pour l'amplification complète de cinq gènes (UL18, UL40, UL111a, UL123 et US3) et partielle pour le gène codant pp65 (UL83) selon les conditions définies dans le tableau 1 :

Tableau 1 : Séquences des amorces et conditions de PCR

		Amorces sens	Amorces anti-sens	Conditions de PCR
UL18	1 ^{ère} PCR	5'- GAAGATAGGAGGGGT CAAAACGCGG-3'	5'- AGAAACATGACCGTGTT GACGGGGG-3'	94°C 12 min 94°C 30 s - 60°C 30 s - 72°C 2 min : 50 cycles 72°C 7 min
	PCR nichée	5'- ATAGGAGGGGTCAAA ACGCGGACTG-3'	5'- AACAGCTCATAACAGGGC GTTGCTGG-3'	94°C 12 min 94°C 30 s - 60°C 30 s - 72°C 2 min : 40 cycles 72°C 7 min
UL40	1 ^{ère} PCR	5'- TCCTCCCTGGTACCCG ATAACAG-3'	5'- CGGGCCAGGACTTTTTA ATGGCC-3'	94°C 12 min 94°C 30 s - 59°C 30 s - 72°C 1min30: 50 cycles 72°C 7 min
	PCR nichée	5'- GGTAAGGGCACTCGTG AGGATGTGC-3'	5'- TCCGAACGCTCGTGAGC AACAGTCG-3'	94°C 12 min 94°C 30 s - 59°C 30 s - 72°C 1min30: 40 cycles 72°C 7 min
UL111a	1 ^{ère} PCR	5'- GCCATCAAGTGGTGAC GATAACCCG-3'	5'- ACACCCACAAACAACGT CACGACGC-3'	94°C 12 min 94°C 30 s - 59°C 30 s - 72°C 1min3 : 50 cycles 72°C 7 min
	PCR nichée	5'- AAAGGACCACCTACCT GGGACGCGC-3'	5'- GACACCGGAGCGGACT GCAAATCGC-3'	94°C 12 min 94°C 30 s - 59°C 30 s - 72°C 1min30: 40 cycles 72°C 7 min
US3	1 ^{ère} PCR	5'- TGAGCAGCCATAGCAC CAGAGTCC-3'	5'- CAGTCCACACGCTACTT CTCAGCG-3'	94°C 12 min 94°C 30 s - 60°C 30 s - 72°C 1min30: 50 cycles 72°C 7 min
	PCR nichée	5'- CCCGCGCCTGGGTGT TAGAGTCCG-3'	5'- GCACTGCTGCAGCCAGA CCGGAGCG-3'	94°C 12 min 94°C 30 s - 65°C 30 s - 72°C 1min30: 40 cycles 72°C 7 min
UL123	1 ^{ère} PCR	5'- AGGCGTGACACGTTTA TTGAGTAGG-3'	5'- TAACTCCYGTTGCGGTG CTGTTAAC-3'	94°C 12 min 94°C 30 s - 60°C 30 s - 72°C 2 min : 50 cycles 72°C 7 min
	PCR nichée	5'- AGTATAACAATAGTGAC GTGGGATCC-3'	5'- TAGCTGACAGACTAACA GACTGTTC-3'	94°C 12 min 94°C 30 s - 60°C 30 s - 72°C 2 min : 40 cycles 72°C 7 min
UL83	1 ^{ère} PCR	5'- AGCTTGCCCTGGATGC GATACTGGC-3'	5'- ATCAACGTGCACCACTA CCCGTCGG-3'	94°C 12 min 94°C 30 s - 61°C 30 s - 72°C 2 min : 50 cycles 72°C 7 min
	PCR nichée	5'- AGCTTGCCCTGGATGC GATACTGGC-3'	5'- TTGGTCGCCTGCCTATG TAAGGCGG-3'	94°C 12 min 94°C 30 s - 61°C 30 s - 72°C 2 min : 40 cycles 72°C 7 min

Les positions des amorces par rapport à la souche de référence AD169 (NC_001347) et les amorces nécessaires pour le séquençage sont présentées dans les tableaux 2 et 3 :

Tableau 2 : Longueur des gènes amplifiés et position des amorces par rapport à AD169

	UL18	UL40	UL83	UL111a	UL123	US3
Longueur (nt)	1107	666	1685	687	1759	561
Position	23639-24745	53215-53880	120283-121968	160609-161295	171937-173696	195061-195621
Par rapport à AD169						
PCR	23563-23588 /	53094-53119 /	120915-120940 /	160437-160462 /	171825-171850 /	194914-194939 /
Par rapport à AD169	24791-24815	53930-54952	122036-122061	161367-161392	173821-173846	195682-195706
PCR nichée	23567-23592 /	53178-53203/	120915-120940 /	160549-160574 /	171881-171906 /	194989-195014 /
Par rapport à AD169	24746-24770	53886-53910	122019-122044	161299-161324	173736-173661	195655-195679

Tableau 3 : Amorces supplémentaires nécessaires pour le séquençage

UL18is	5'-ATAACCTGACACTATGGACGGGCCC-3'
UL18ias	5'-TGTCATTCTGGTTTCGGACACCGCC-3'
IE1-c	5'-TGGCCAAAGTGTAGGCTRCAATAGC-3'
IE1-d	5'-AAGGTCTTTGCCAGTACATTCTGG-3'
IE1-e	5'-AGGCTGAGTTCTTGGTAAAGAACTC-3'
IE1-f	5'-GAGATGTGGATGGCTTGTATTAAGG-3'
IE1-g	5'-ACATACAYAAGTAAATCCRCGTCCC-3'
IE1-h	5'-AGTGACCGAGGATTGCAACGAGAAC-3'
T7	5'-TAATACGACTCACTATAGG-3'
SP6	5'-GATTTAGGTGACACTATA-3'

La première PCR s'effectue dans un mélange réactionnel comportant par échantillon : 1.25 U d'ADN polymérase (Ampli Taq Gold, Applied Biosystems), 5 µl de tampon 10X, 2,5 mM de MgCl₂, 200 µM de dNTP et 200 nM de chaque amorce pour un volume final de 50µL par tube.

Pour la PCR nichée, le mélange réactionnel comporte pour chaque tube : 2U d'ADN polymérase (Ampli Taq Gold, Applied Biosystems), 10 µl de tampon 10X, 2,5 mM de MgCl₂, 200 µM de dNTP et 200 nM de chaque amorce pour un volume final de 100µL.

Selon la charge virale CMV initiale détectée en routine par PCR en temps réel, 10 à 20 µL d'extrait d'ADN sont introduits dans le premier mélange réactionnel et 2 à 10 µL du produit de première PCR dans le second mélange réactionnel.

L'amplification des produits PCR est vérifiée sur un gel d'agarose 1,5%.

C. Purification des produits PCR

Les produits de PCR nichée sont purifiés à l'aide des colonnes de purification Microspin S400-HR (GE Healthcare) par centrifugation (Centrifugeuse Eppendorf 5417R) selon les recommandations du fournisseur. Les produits d'amplification purifiés sont conservés à une température de -20 °C avant séquençage.

D. Réaction de séquence

Les fragments de PCR ont été séquencés sur la totalité des brins sens et antisens comme précédemment décrit (Merel *et al.*, 2001) et une séquence consensus a été générée.

La réaction de séquence est réalisée à l'aide de la trousse CEQ DTCS Quick Start (Beckman Coulter) selon les recommandations du fournisseur.

Les amorces utilisées pour le séquençage sont celles présentées dans le tableau 2 pour la PCR nichée à l'exception du gène *UL18* qui, du fait de sa longueur (1107pb), a nécessité l'utilisation d'amorces supplémentaires internes sens (*UL18is*) et antisens (*UL18ias*), tout comme le gène *UL123* (1759 pb) où huit amorces ont été nécessaires pour séquencer le gène entier (Tableau 3). Les séquences ont été assemblées en utilisant le programme Staden Package (Staden, 1994).

Les produits de réaction de séquence sont purifiés sur Sephadex® dans des plaques du système MultiScreen® HTS (Millipore) selon les recommandations du fournisseur.

Le séquençage est réalisé avec le séquenceur capillaire CEQ 8000 de Beckman Coulter.

E. Clonage

Le clonage du gène *UL40* a été réalisé chez le patient 101Bx car les résultats du séquençage direct du gène *UL40* pour ce patient montrent une variabilité intéressante entre les temps 1 et 2.

1. Préparation des bactéries compétentes

La souche d'*Escherichia coli* utilisée pour le clonage est la souche DH5 α sensible à l'ampicilline. Les bactéries sont rendues compétentes par la technique du chlorure de calcium.

2. Ligation

Le vecteur utilisé est le plasmide pGEM-T Easy Vector System (Promega). Il possède la séquence ori, origine de répllication d'*E. Coli* et un gène de résistance à l'ampicilline.

Ce vecteur contient également le gène lacZ qui code la β -galactosidase. A l'intérieur du gène lacZ se trouve un multisite de clonage ou polylinker qui autorise l'insertion de la séquence d'ADN à amplifier.

Pour la ligation, le mélange réactionnel est détaillé dans le tableau 4.

Une ligation courte a été réalisée (1h à température ambiante).

Tableau 4 : Préparation des échantillons, témoin positif et négatif pour la ligation

	Echantillon	Témoin positif	Témoin négatif
Tampon 5 μ L	+	+	+
Vecteur 1 μ l	+	+	+
Produit PCR 36.6ng	+	-	-
Contrôle positif 2 μ L	-	+	-
T4 ligase 1 μ L	+	+	+
H2O qsp 10 μ L	+	+	+

Concentration du vecteur : 50ng/ μ L, concentration ligase 3U/ μ L

Rapport produit PCR/vecteur : 3/1

La quantité de produit PCR à ajouter a été déterminée selon l'équation recommandée par le fournisseur (Promega) : [(ng de vecteur x taille en kb de l'insert)/ taille en kb du vecteur] x ratio molaire insert / vecteur.

3. Transformation

Les bactéries DH5 α compétentes (200 μ L) sont incubées en présence du produit de ligation, et le mélange est placé dans la glace durant 20 minutes. Un choc thermique à 42°C (1 min 30) est effectué puis les bactéries sont immédiatement plongées dans la glace pendant 2 minutes. Elles sont ensuite mises en culture à 37°C pendant une heure dans du milieu LB contenant 20mM de glucose, sous agitation.

4. Sélection des bactéries ayant intégré le plasmide recombinant (système blanc/bleu)

Les bactéries transformées vont croître sur le milieu en présence de l'antibiotique (1^{ère} sélection). La séquence d'ADN à amplifier s'est insérée au niveau du site de multiclonage du

gène Lac Z aboutissant à son inactivation et l'absence de β -galactosidase fonctionnelle. En présence d'isopropylthio- β -D-galactoside (IPTG, inducteur de l'opéron lactose) et de X-gal (substrat de la β -galactosidase qui libère un chromophore bleu lors de son clivage), les bactéries ayant intégré le plasmide recombinant se présentent sous forme de colonies de couleur blanchâtre, celles ayant un plasmide sans insert sont de couleur bleue (2^{ème} sélection).

5. Repiquage

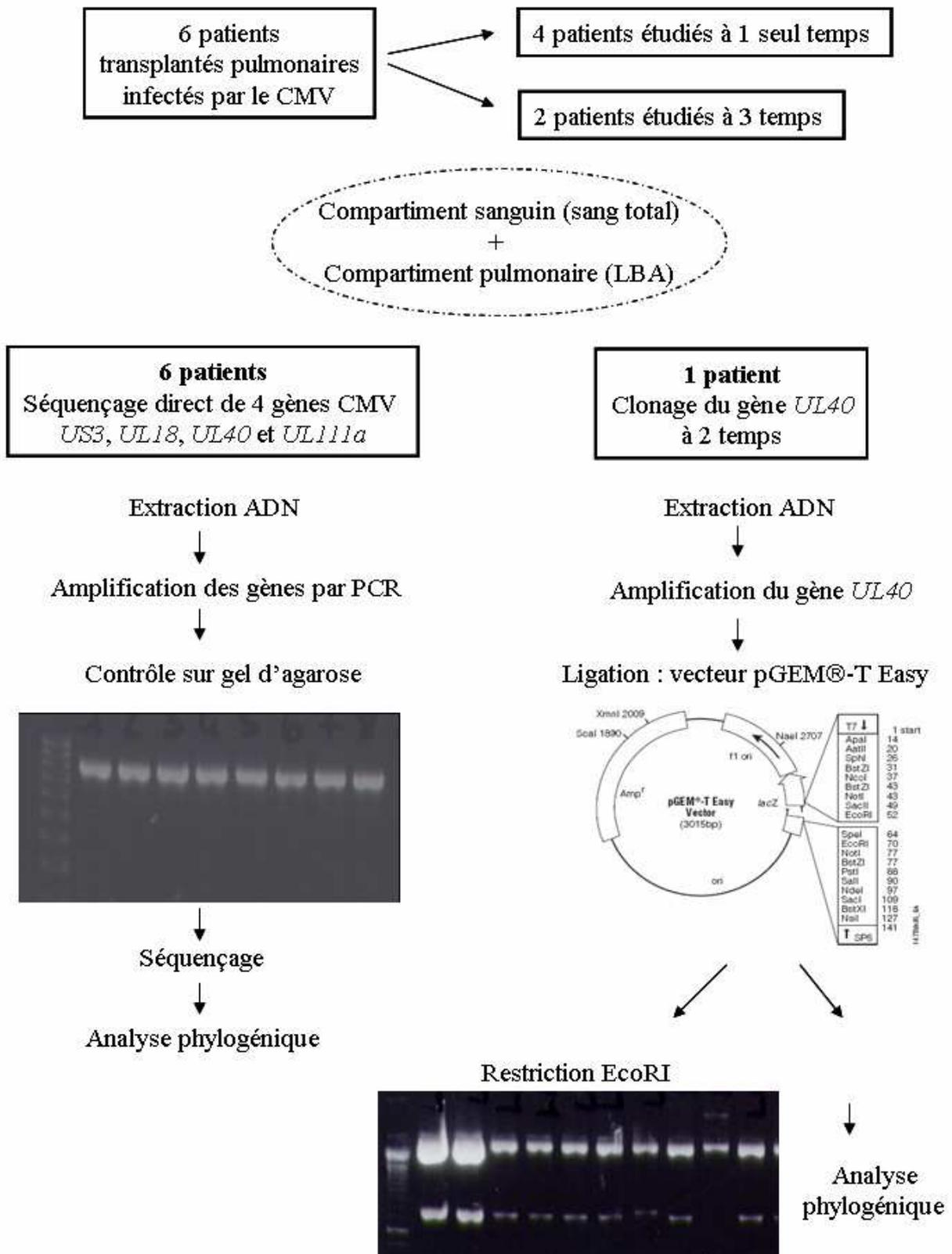
Les colonies blanches sélectionnées sont repiquées sur milieu LB + ampicilline en boîte de Petri et mises en culture en milieu liquide (2 ml de milieu LB + ampicilline) à 37°C, une nuit sous agitation.

6. Stockage

Avant extraction, les bactéries sont stockées à la fois sous la forme d'un culot de centrifugation et dans un mélange LB + glycérol, conservés tous deux à -80°C.

La figure 15 illustre la stratégie utilisée pour les transplantés pulmonaires.

Figure 15: Schéma récapitulatif des patients transplantés pulmonaires, matériels et méthodes



F. Extraction plasmidique

L'ADN plasmidique est extrait à partir du culot bactérien à l'aide de la trousse NucleoSpin Plasmid (Macherey-Nagel) selon les recommandations du fabricant. L'élution finale s'effectue dans un volume de 50µL.

G. Digestion plasmidique

Une digestion plasmidique est réalisée afin de vérifier la présence et la taille de l'insert (732 pb). Elle est obtenue par l'enzyme *EcoRI* (Promega) selon les recommandations du fabricant.

Les produits de digestion (20µL) sont déposés sur un gel d'agarose à 1.5%.

Les clones d'intérêt sont séquencés selon la méthode décrite plus haut et à l'aide des amorces T7 et SP6.

H. Analyse phylogénique

1. Clustal W 1.83

Clustal W 1.83 a été utilisé pour aligner les séquences nucléotidiques en utilisant des souches connues comme référence : AD169 (numéro d'accèsion Genbank NC_001347), Merlin (numéro d'accèsion Genbank NC_006273), Towne (numéro d'accèsion Genbank AY072777) et Toledo (numéro d'accèsion Genbank AH013698). Les arbres phylogénétiques traditionnels sont construits en utilisant la méthode du plus proche voisin (neighbour-joining).

L'arbre phylogénique issu de cet alignement permet une étude de la compartimentation pour chaque patient et de la variabilité inter-patients des souches de CMV pour les souches de CMV isolées chez les patients transplantés pulmonaires.

De même, un arbre phylogénique est construit avec les séquences issues du clonage du gène *UL40* du patient 101Bx dans les 2 compartiments et à 2 temps différents afin d'étudier la répartition des variants viraux. En particulier, l'analyse de l'arbre permet l'identification de variants viraux minoritaires. La fiabilité de l'arbre est évaluée grâce aux valeurs de bootstrap (méthode de ré-échantillonnage).

Les gènes *UL18*, *UL40*, *UL111a*, *UL123* et *US3* ont été analysés séparément aussi bien que concaténés, c'est-à-dire mis bout à bout, en respectant l'ordre dans lequel ils se présentent dans le génome viral. L'identité et l'homologie des séquences concaténées ou non ont été

calculées en utilisant le programme Needle inclus dans le programme EMBOSS (Rice, Longden, and Bleasby, 2000).

2. SplitsTree

Les liens d'évolution sont généralement représentés sous forme d'arbres phylogéniques traditionnels, basés sur un modèle d'évolution dominé par les mutations et les réunions d'espèces. Des modèles plus réalistes doivent prendre en compte la genèse des gènes, les événements de perte ou de duplication génique, l'hybridation, les transferts horizontaux des gènes et les recombinaisons. Un arbre phylogénique simple n'est souvent pas approprié pour représenter une histoire phylogénique ou différents signaux phylogéniques incompatibles. Il est nécessaire d'utiliser un graphique plus général, un réseau, pour illustrer ces données. Pour acquérir des informations concernant les liens d'évolution pouvant exister entre les souches cliniques isolées chez des patients transplantés, les séquences concaténées des cinq gènes de CMV ont été analysées en utilisant le programme SplitsTree (Huson and Bryant, 2006). Des souches recombinantes de HSV-1 et HSV-2 ont été détectées avec cette méthode (Norberg *et al.*, 2007). A notre connaissance, le programme SplitsTree est ici utilisé pour la première fois pour analyser des gènes de CMV. Ce programme construit des réseaux de recombinaison, illustrant les liens d'évolution existant entre différentes souches en présence d'événements de recombinaison, contrairement aux algorithmes traditionnels qui élaborent des arbres phylogéniques bifurquant. L'histoire évolutive des isolats est construite à partir des signaux conflictuels contenus dans les séquences alignées quand des événements de recombinaison ont participé à l'évolution des isolats inclus. Une recombinaison ne sera pas caractérisée par une branche mais par de nombreuses branches connectées aux souches parentales. Cela signifie que si un réseau complexe représente les relations existant entre les isolats inclus, il est probable que des événements de recombinaison ont participé à leur évolution. Un réseau réticulé est utilisé pour décrire une évolution réticulée. Un Phi-test (Bruen, Philippe, and Bryant, 2006) est inclus dans le programme SplitsTree, version 4, pour valider les signaux conflictuels détectés à cause d'un événement de recombinaison. Nous avons aussi étudié les séquences concaténées des cinq gènes du CMV pour simplifier l'alignement nucléotidique (4804 nucléotides en moyenne par séquence concaténée) et gagner une vue synthétique des séquences comparées.

I. Méthode de bootscan

Tous les candidats recombinants du CMV isolé dans le patient 101Bx ont été testés avec la technique de bootscan incluse dans le programme Simplot (Lole *et al.*, 1999).

Pour renforcer l'hypothèse de la présence de recombinaison dans les séquences concaténées de CMV, la méthode dite du bootscan incluse dans le programme Simplot (Lole *et al.*, 1999) a été appliquée aux candidats recombinants.

En résumé, le bootscanning relie à partir de leur alignement une séquence recombinante potentielle (séquence cible) à des séquences parentales potentielles. Après l'obtention d'un alignement optimal, il est « cassé » séquentiellement, en segments chevauchant de 20 nucléotides qui alimentent l'analyse phylogénétique du programme. Des arbres phylogéniques sont construits pour chaque segment. Finalement, la séquence inconnue est placée en fonction de la valeur de bootstrap de chaque segment le long du génome en tenant compte du fait que plus la valeur de bootstrap est élevée plus il y a de chance d'être proche de la séquence parentale. Schématiquement, le bootscan indique quelle séquence ressemble le plus à la séquence cible et dans quelle proportion (pourcentage visible sur l'axe des ordonnées), tout au long de la séquence nucléique. En cas d'événement de recombinaison, un message apparaît : « Détection d'un possible événement de recombinaison », qui peut être confirmé en utilisant l'onglet disponible « FindSites ».

J. Analyse des épitopes immunodominants

Pour étudier la variabilité des épitopes décrits pour la protéine immunodominante IE1, les logiciels BIMAS (http://thr.cit.nih.gov/cgi-bin/molbio/ken_parker_comboform) et SYFPEITHI (www.syfpeithi.de) ont été utilisés pour calculer le score d'affinité théorique de chaque séquence peptidique précédemment décrite en association avec le génotype HLA du patient infecté.

XI. Patients

Notre étude rétrospective a été réalisée sur des échantillons biologiques conservés à -80°C pour des souches isolées chez des patients transplantés.

Les patients, dénommés 1 à 17Bx (transplantés rénaux) et 101 à 106Bx (transplantés pulmonaires), et les prélèvements étudiés sont présentés dans le tableau ci-dessous. La charge virale (CV) a été mesurée par PCR en temps réel (laboratoire de virologie, CHU de

Bordeaux). Elle est exprimée en copies ADN/mL. Les patients ont été inclus car ils développaient une infection à CMV avec une PCR en temps réelle positive pour le CMV à la fois dans le sang et dans le liquide de lavage broncho-alvéolaire pour les patients transplantés pulmonaires. Pour dix patients, des prélèvements séquentiels étaient disponibles et espacés d'au moins six semaines (délai pour lequel nous considérons qu'il s'agit d'un épisode d'infection à CMV différent).

Un organe transplanté peut contenir du virus à l'état latent. Ainsi, un patient transplanté peut présenter une primo-infection à CMV (receveur séronégatif pour le CMV : R-, donneur séropositif : D+), une réactivation endogène (R+) ou exogène (réactivation ou surinfection à partir du virus transmis par le donneur). Plusieurs souches virales peuvent donc circuler chez un même patient, à l'occasion de surinfections ou co-infections (D+ / R+).

Parmi les 23 patients étudiés, 10 étaient D+/R+, un D-/R+ et 12 D+/R- (Tableau 5).

Tableau 5 : Patients et prélèvements (nature et charge virale (CV))

Patient	Code patient	Date examen	Nature prélèvement	CV	Statut
1Bx	1Bx0408	31/08/2004	sang total	1316	D+R+
	1Bx0411	22/11/2004	sang total	1070	
	1Bx0501	05/01/2005	sang total	60440	
2Bx	2Bx0410	05/10/2004	sang total	1391	D+R+
	2Bx0503	25/03/2005	sang total	1970	
3Bx	3Bx0412	08/12/2004	sang total	1033	D+R+
	3Bx0503	09/03/2005	sang total	819	
4Bx	4Bx0409	17/09/2004	sang total	+	D+R+
5Bx	5Bx0502	23/02/2005	sang total	13850	D+R+
6Bx	6Bx0409	22/09/2004	sang total	+	D+R+
7Bx	7Bx0506	08/06/2005	sang total	3101	D-R+
8Bx	8Bx0312	17/12/2003	sang total	+	D+R-
	8Bx0402	17/02/2004	sang total	+	
	8Bx0405	26/05/2004	sang total	+	
	8Bx0410	06/10/2004	sang total	34140	
	8Bx0501	05/01/2005	sang total	1006	
	8Bx0604	05/04/2006	sang total	3609	
	8Bx0610	02/10/2006	sang total	2616	
9Bx	9Bx0411	26/11/2004	sang total	64800	D+R-
	9Bx0501	06/01/2005	sang total	1076	
	9Bx0502	23/02/2005	sang total	794	
	9Bx0504	21/04/2005	sang total	+	
10Bx	10Bx0502	24/02/2005	sang total	4417	D+R-
	10Bx0504	07/04/2005	sang total	980	
	10Bx0508	11/08/2005	sang total	1373	
11Bx	11Bx0409	17/09/2004	sang total	+	D+R-
	11Bx0412	23/12/2004	sang total	116700	
12Bx	12Bx0501	26/01/2005	sang total	758000	D+R-
	12Bx0505	04/05/2005	sang total	1056	
13Bx	13Bx0408	25/08/2004	sang total	+	D+R-
14Bx	14Bx0507	27/07/2005	sang total	118400	D+R-

15Bx	15Bx0506	02/06/2005	sang total	5507	D+R-
16Bx	16Bx0503	21/03/2005	sang total	16260	D+R-
17Bx	17Bx0508	03/08/2005	sang total	183000	D+R-
101Bx	101Bx0406WB	17/06/2004	sang total	+ faible	D+R-
	101Bx0406BAL	21/06/2004	LBA	+ faible	
	101Bx0508BAL	23/08/2005	LBA	+	
	101Bx0509WB	12/09/2005	sang total	< 500	
	101Bx0705BAL	10/05/2007	LBA	+ faible	D-R+
102Bx	101Bx0705WB	24/05/2007	sang total	2006	
	102Bx0602WB	28/02/2006	sang total	689	D+R-
	102Bx0603BAL	17/03/2006	LBA	+	
	102Bx0605BAL	09/05/2006	LBA	+ faible	
	102Bx0605WB	09/05/2006	sang total	1375	
103Bx	102Bx0608BAL	11/08/2006	LBA	+	
	102Bx0608WB	21/08/2006	sang total	1705	
	103Bx0607BAL	31/07/2006	LBA	+	D+R+
104Bx	103Bx0607WB	31/07/2006	sang total	2449000	
	104Bx0604BAL	14/04/2006	LBA	+	D+R+
105Bx	104Bx0604WB	13/04/2006	sang total	2847000	
	105Bx0502BAL	24/02/2005	LBA	+	D+R+
106Bx	105BX0502WB	21/02/2005	sang total	123700	
	106Bx0411WB	04/11/2004	sang total	17680	D+R+
	106Bx0411BAL	03/11/2004	LBA	+	

Sang total : PCR qualitative (résultat + ou + faible) ou PCR quantitative (nombre de copies/mL).

LBA : PCR qualitative.

Informations apportées par le code patient. Exemple du 101Bx0406WB :

101 : rang du patient, Bx : Bordeaux, 04 : année de prélèvement, 06 : mois de prélèvement, WB : compartiment (WB pour sang total, BAL pour liquide de lavage broncho-alvéolaire).

XII. Résultats

A. Valorisation des résultats

Les 263 séquences obtenues sont disponibles dans Genbank avec les « access numbers » suivants indiqués dans le tableau 6.

Tableau 6 : Numéros d'accès des séquences obtenues

Patient	Code patient	UL18	UL40	UL83	UL111a	UL123	US3
1Bx	1Bx0408	DQ367970	DQ367999	HM142972	DQ367940	FJ491265	DQ368030
	1Bx0411		DQ368000		DQ367941	FJ491266	DQ368031
	1Bx0501	DQ367971	DQ368001	HM142973	DQ367942	FJ491267	DQ368032
2Bx	2Bx0410	DQ367972	DQ368002		DQ367943	FJ491268	DQ368033
	2Bx0503	DQ367973	DQ368003	HM142974	DQ367944	FJ491269	DQ368034
3Bx	3Bx0412	DQ367974	DQ368004	HM142975	DQ367945	FJ491270	DQ368035
	3Bx0503	DQ367975	DQ368005		DQ367946	FJ491271	DQ368036
4Bx	4Bx0409	DQ367976	DQ368006		DQ367947	FJ491272	DQ368037

5Bx	5Bx0502	DQ367977	DQ368007	HM142976	DQ367948	FJ491273	DQ368038
6Bx	6Bx0409		DQ368008		DQ367949	FJ491274	DQ368039
7Bx	7Bx0506	DQ367978	DQ368009	HM142977	DQ367950	FJ491275	DQ368040
8Bx	8Bx0312	DQ367979	DQ368010	HM142978	DQ367951	FJ491276	DQ368041
	8Bx0402	DQ367980	DQ368011	HM142979	DQ367952	FJ491277	DQ368042
	8Bx0405	DQ367981	DQ368012	HM142980	DQ367953	FJ491278	DQ368043
	8Bx0410	DQ367982	DQ368013	HM142981	DQ367954	FJ491279	DQ368044
	8Bx0501	DQ367983	DQ368014		DQ367955	FJ491280	DQ368045
	8Bx0604		EU048965		HM142995		HM142997
	8Bx0610		EU048996		HM142996		HM142998
9Bx	9Bx0411	DQ367984	DQ368015	HM142982	DQ367956	FJ491281	DQ368046
	9Bx0501	DQ367985	EU049061		DQ367957	FJ491282	DQ368047
	9Bx0502	DQ367986	DQ368017		DQ367958	FJ491283	DQ368048
	9Bx0504		EU049138				
10Bx	10Bx0502	DQ367987	DQ368018	HM142983	EF448364	FJ491284	DQ368049
	10Bx0504	DQ367988	DQ368019	HM142984	EF488365	FJ491285	DQ368050
	10Bx0508	DQ367989	DQ368020	HM142985	EF488366	FJ491286	DQ368051
11Bx	11Bx0409	DQ367990	DQ368021	HM142986	DQ367962	FJ491287	DQ368052
	11Bx0412	DQ367991	DQ368022	HM142987	DQ367963	FJ491288	DQ368053
12Bx	12Bx0501	DQ367992	DQ368023	HM142988	DQ367964	FJ491289	DQ368054
	12Bx0505	DQ367993	DQ368024	HM142989	DQ367965	FJ491290	DQ368055
13Bx	13Bx0408	DQ367994	DQ368025	HM142990	DQ367966	FJ491291	DQ368056
14Bx	14Bx0507	DQ367995	DQ368026	HM142991	EF535834	FJ491292	DQ368057
15Bx	15Bx0506	DQ367996	DQ368027	HM142992	DQ367967	FJ491293	DQ368058
16Bx	16Bx0503	DQ367997	DQ368028	HM142993	DQ367968	FJ491294	DQ368059
17Bx	17Bx0508	DQ367998	DQ368029	HM142994	DQ367969	FJ491295	DQ368060
101Bx	101Bx0406WB		FJ491325		FJ491344		
	101Bx0406BAL	FJ491307	FJ491326				
	101Bx0508BAL	FJ491308	FJ491327		FJ491345	FJ491296	FJ491361
	101Bx0509WB	FJ491309	FJ491328		FJ491346		FJ491362
	101Bx0705BAL	FJ491310	FJ491329		FJ491347		FJ491363
	101Bx0705WB	FJ491311	FJ491330		FJ491348		FJ491364
102Bx	102Bx0602WB	FJ491312	FJ491331		FJ491349		
	102Bx0603BAL	FJ491313	FJ491332		FJ491350	FJ491297	FJ491365
	102Bx0605BAL	FJ491314	FJ491333		FJ491351	FJ491298	FJ491366
	102Bx0605WB	FJ491315	FJ491334		FJ491352		
	102Bx0608BAL	FJ491316	FJ491335		FJ491353		FJ491367
	102Bx0608WB						
103Bx	103Bx0607BAL	FJ491317	FJ491336		FJ491354	FJ491299	FJ491368
	103Bx0607WB	FJ491318	FJ491337		FJ491355	FJ491300	FJ491369
104Bx	104Bx0604BAL	FJ491319	FJ491338		FJ491356	FJ491301	FJ491370
	104Bx0604WB	FJ491320	FJ491339		FJ491357	FJ291302	FJ491371
105Bx	105Bx0502BAL	FJ491321	FJ491340		FJ491358	FJ291303	FJ491372
	105BX0502WB	FJ491322	FJ491341			FJ291304	FJ491373
106Bx	106Bx0411WB	FJ491323	FJ491342		FJ491359	FJ291305	FJ491374
	106Bx0411BAL	FJ491324	FJ491343		FJ491360	FJ291306	FJ491375

B. Variabilité globale

1. Les quatre gènes immunomodulateurs

Cette étude rétrospective a été menée sur des souches cliniques de CMV isolées chez 23 patients transplantés avec ou non un suivi longitudinal dans le temps. Les résultats d'amplification des quatre gènes immunomodulateurs sont présentés dans le tableau 7.

Tableau 7 : Résultats d'amplification

Patient	Code patient	UL18	UL40	UL111a	US3	Patient	Code patient	UL18	UL40	UL111a	US3
1Bx	1Bx0408	+	+	+	+	101Bx	101Bx0406WB	-	+	+	-
	1Bx0411	-	+	+	+		101Bx0406BAL	+	+	-	-
	1Bx0501	+	+	+	+		101Bx0508BAL	+	+	+	+
2Bx	2Bx0410	+	+	+	+	101Bx0509WB	+	+	+	+	
	2Bx0503	+	+	+	+	101Bx0705BAL	+	+	+	+	
3Bx	3Bx0412	+	+	+	+	101Bx0705WB	+	+	+	+	
	3Bx0503	+	+	+	+	102Bx	102Bx0602WB	+	+	+	-
4Bx	4Bx0409	+	+	+	+	102Bx0603BAL	+	+	+	+	
	5Bx0502	+	+	+	+	102Bx0605BAL	+	+	+	+	
5Bx	5Bx0502	+	+	+	+	102Bx0605WB	+	+	+	-	
6Bx	6Bx0409	-	+	+	+	102Bx0608BAL	+	+	+	+	
7Bx	7Bx0506	+	+	+	+	102Bx0608WB	-	-	-	+	
8Bx	8Bx0312	+	+	+	+	103Bx	103Bx0607BAL	+	+	+	+
	8Bx0402	+	+	+	+	103Bx0607WB	+	+	+	+	
	8Bx0405	+	+	+	+	104Bx	104Bx0604BAL	+	+	+	+
	8Bx0410	+	+	+	+	104Bx0604WB	+	+	+	+	
	8Bx0501	+	+	+	+	105Bx	105Bx0502BAL	Partielle	+	+	+
9Bx	8Bx0604	-	+	+	+	105BX0502WB	Partielle	+	-	+	
	8Bx0610	-	+	+	+	106Bx	106Bx0411WB	+	+	+	+
	9Bx0411	+	+	+	+	106Bx0411BAL	+	+	+	+	
	9Bx0501	+	+	+	+						
	9Bx0502	+	+	+	+						
10Bx	9Bx0504	-	+	-	-						
	10Bx0502	+	+	+	+						
	10Bx0504	+	+	+	+						
11Bx	10Bx0508	+	+	+	+						
	11Bx0409	+	+	+	+						
12Bx	11Bx0412	+	+	+	+						
	12Bx0501	+	+	+	+						
13Bx	12Bx0505	+	+	+	+						
	13Bx0408	+	+	+	+						
14Bx	14Bx0507	+	+	+	+						
15Bx	15Bx0506	+	+	+	+						
16Bx	16Bx0503	+	+	+	+						
17Bx	17Bx0508	+	+	+	+						

+ : obtention d'une séquence

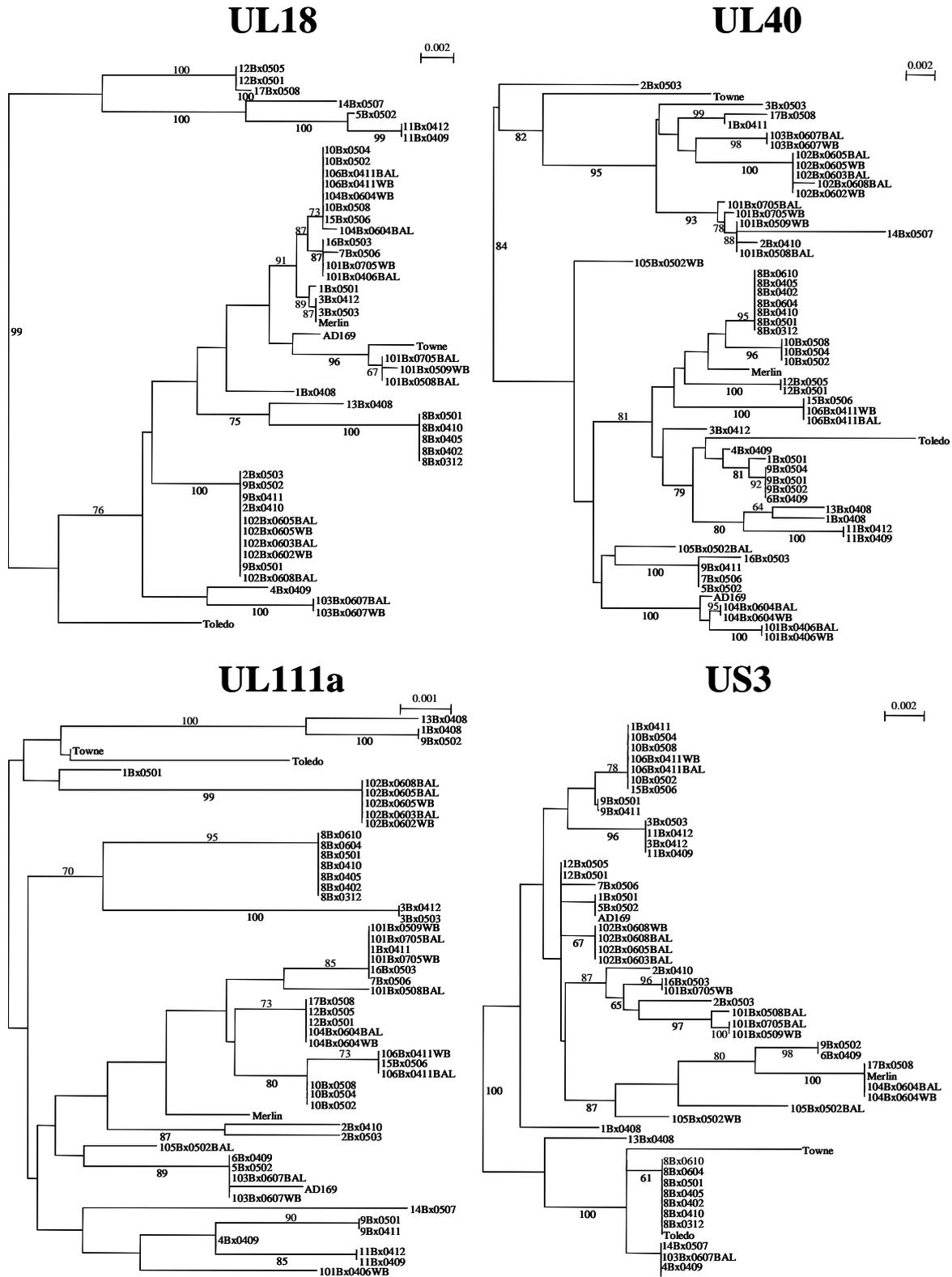
- : échec d'amplification ou de séquençage

Les arbres phylogéniques (figure 16) obtenus après séquençage direct pour chaque gène indiquent que le CMV de chaque patient a une combinaison de séquences différente pour les 4 gènes, ce qui signifie que chaque patient est infecté par des variants de CMV particuliers. L'analyse de l'ensemble des séquences montre qu'une séquence identique peut être trouvée

pour un gène donné chez différents patients comme nous l'avons déjà montré (Garrigue et al., 2008b). Les valeurs de bootstrap confirment la robustesse des arbres phylogéniques. De plus, la variabilité est modérée. Par exemple, pour le gène UL40 le pourcentage de variabilité le plus important au niveau acide aminé est de 12% entre 13Bx et 17Bx.

Dix patients (quatre greffés pulmonaires et six greffés rénaux) sont D+R+. Chez ces patients, pendant un épisode de réactivation du CMV, les souches virales présentes à la fois chez le donneur et chez le receveur peuvent se réactiver concomitamment. Il n'est pas surprenant de trouver des populations virales mixtes, se traduisant par la présence de double population dans la séquence nucléique.

Figure 16 : Arbres phylogéniques des quatre gènes immunomodulateurs



Certains commentaires peuvent être faits à propos de chacun des gènes viraux :

UL111a : Par comparaison avec la souche de laboratoire AD169 prise comme référence, dans le gène UL111a, des insertions de trois nucléotides (ACG en position 28-30), de six nucléotides (ACGACG, en position 28-33) ou de quinze nucléotides (ACGACGACGACGACG, en position 28-42) ont été trouvées chez deux, neuf et un isolats respectivement. De plus, une délétion de cinq nucléotides (TGGTC) est observée en position 29-33 dans l'isolat 10Bx pour chacun des trois temps disponibles pour ce patient et se traduit par un gène UL111a probablement non fonctionnel.

US3 : Par comparaison avec la souche de laboratoire AD169 prise comme référence, aucune délétion ou insertion n'a été trouvée dans la séquence du gène US3.

UL18 : Par comparaison avec la souche de laboratoire AD169 prise comme référence, pour le gène UL18, une délétion de trois nucléotides (ACA) est trouvée en position 40-42 pour six isolats.

Nous n'avons pas observé de modifications au niveau des motifs YRKV et KVRS localisés au niveau de la queue cytoplasmique, ni au niveau des quatre cystéines (C211, C240, C255 et C267) impliquées dans la liaison à la b2-microglobuline..

Sur 45 isolats, 29 avaient les mutations K42R et A43T et 27 la mutation Q202H.

Ces résultats sont illustrés dans la figure 17.

Figure 17 : Sélection des zones d'intérêt après alignements des séquences *ULI8* en acides aminés obtenues chez nos patients avec la séquence d'AD169.

	1	33	$\alpha 1$	193	$\alpha 3$	233	253	343								
AD169	MMTMWCLTLF	VLWMLRVVGM	[.] FTYHVNS	DK AS	SRANGTIS	[.] PVVKGGV	NDNRAEA	FCI	[.] PEDSEPO	ENB	[.] QGC	WALFCN	QNYFC	VTHG	[.] FFRKVR	S*
102Bx0608B	..T.....	..R.....	[.]	..R..T.	[.]	..H.....	[.]	[.]	[.]
9Bx0501	..T.....	..R.....	[.]	..R..T.	[.]	..H.....	[.]	[.]	[.]
102Bx0602W	..T.....	..R.....	[.]	..R..T.	[.]	..H.....	[.]	[.]	[.]
102Bx0603B	..T.....	..R.....	[.]	..R..T.	[.]	..H.....	[.]	[.]	[.]
102Bx0605W	..T.....	..R.....	[.]	..R..T.	[.]	..H.....	[.]	[.]	[.]
102Bx0605B	..T.....	..R.....	[.]	..R..T.	[.]	..H.....	[.]	[.]	[.]
2Bx0410	..T.....	..R.....	[.]	..R..T.	[.]	..H.....	[.]	[.]	[.]
9Bx0411	..T.....	..R.....	[.]	..R..T.	[.]	..H.....	[.]	[.]	[.]
9Bx0502	..T.....	..R.....	[.]	..R..T.	[.]	..H.....	[.]	[.]	[.]
2Bx0503	..T.....	..R.....	[.]	..R..T.	[.]	..H.....	[.]	[.]	[.]
101Bx0508B	..T.....	..R.....	[.]	..R..T.	[.]	..H.....	[.]	[.]	[.]
101Bx0509W	..T.....	..R.....	[.]	..R..T.	[.]	..H.....	[.]	[.]	[.]
101Bx0705B	..T.....	..R.....	[.]	..R..T.	[.]	..H.....	[.]	[.]	[.]
Towne	[.]	..R..T.	[.]	[.]	[.]	[.]
101Bx0406B	[.]	[.]	..D.....	[.]	[.]	[.]
101Bx0705W	[.]	[.]	..D.....	[.]	[.]	[.]
16Bx0503	[.]	[.]	..D.....	[.]	[.]	[.]
7Bx0506	[.]	[.]	..D.....	[.]	[.]	[.]
106Bx0411B	[.]	[.]	..D.....	[.]	[.]	[.]
15Bx0506	[.]	[.]	..D.....	[.]	[.]	[.]
10Bx0508	[.]	[.]	..D.....	[.]	[.]	[.]
104Bx0604W	[.]	[.]	..D.....	[.]	[.]	[.]
106Bx0411W	[.]	[.]	..D.....	[.]	[.]	[.]
10Bx0502	[.]	[.]	..D.....	[.]	[.]	[.]
10Bx0504	[.]	[.]	..D.....	[.]	[.]	[.]
104Bx0604B	[.]	[.]	..D.....	[.]	[.]	[.]
Merlin	[.]	[.]	..D.....	[.]	[.]	[.]
3Bx0503	[.]	[.]	..D.....	[.]	[.]	[.]
3Bx0412	[.]	[.]	..D.....	[.]	[.]	[.]
1Bx0501	[.]	[.]	..D.....	[.]	[.]	[.]
103Bx0607W	..T.....	..R.....	[.]	[.]	..H.....	[.]	..D.....	[.]	[.]
103Bx0607B	..T.....	..R.....	[.]	[.]	..H.....	[.]	..D.....	[.]	[.]
4Bx0409	..T.....	..R.....	[.]	..R..	[.]	..H.....	[.]	..D.....	[.]	[.]
8Bx0312	[.]	..R..T.	[.]	..H.....	[.]	..D.....	[.]	[.]
8Bx0402	[.]	..R..T.	[.]	..H.....	[.]	..D.....	[.]	[.]
8Bx0405	[.]	..R..T.	[.]	..H.....	[.]	..D.....	[.]	[.]
8Bx0410	[.]	..R..T.	[.]	..H.....	[.]	..D.....	[.]	[.]
8Bx0501	[.]	..R..T.	[.]	..H.....	[.]	..D.....	[.]	[.]
13Bx0408	[.]	..R..	[.]	..H.....	[.]	..D.....	[.]	[.]
1Bx0408	[.]	..R..	[.]	..H.....	[.]	..D.....	[.]	[.]
Toledo	..T.....	..R.....	[.]	[.]	..H.....	[.]	[.]	[.]
11Bx0409	..T.....	..R.....	[.]	..R..T.	[.]	..H.....	[.]	[.]	[.]
11Bx0412	..T.....	..R.....	[.]	..R..T.	[.]	..H.....	[.]	[.]	[.]
5Bx0502	..T.....	..R.....	[.]	..R..T.	[.]	..H.....	[.]	[.]	[.]
14Bx0507	..T.....	..R.....	[.]	..R..T.	[.]	..H.....	[.]	[.]	[.]
17Bx0508	..T.....	..R.....	[.]	..R..T.	[.]	..H.....	[.]	..D.....	[.]	[.]
12Bx0501	..T.....	..R.....	[.]	..R..T.	[.]	..H.....	[.]	..D.....	[.]	[.]
12Bx0505	..T.....	..R.....	[.]	..R..T.	[.]	..H.....	[.]	..D.....	[.]	[.]

En vert : mutation décrite par Cerboni comme étant la plus importante – En rouge : mutations conservatives décrites par Cerboni
 [.] indique une coupure dans la séquence protéique.

UL40 : Par comparaison avec la souche de laboratoire AD169 prise comme référence, aucune délétion ou insertion n'a été trouvée dans la séquence du gène UL40.

La traduction en acides aminés des 53 séquences nucléotidiques de *UL40* a permis de mettre en évidence des variations dans la séquence du nonapeptide (VMAPRTLIL pour AD169), présenté par HLA-E et permettant son expression à la surface cellulaire (tableau 8). Certaines de ces mutations n'avaient pas été décrites auparavant. Les sites d'ancrage du nonapeptide à HLA-E se situent en position 2 et 9.

Tableau 8 : Séquence du nonapeptide et nombre d'isolats ou de souche de référence associés:

Nombre d'isolats	Séquences du nonapeptide
27 + AD169 + Merlin + Towne	VMAPRTLIL
4	VMAPRTLL <u>L</u>
2 + Toledo	VMAPRTL <u>V</u> L
1	<u>A</u> MAPRTLIL
2	VM <u>G</u> PRTLIL
3	VMAPR <u>I</u> LIL
3	VMAPR <u>S</u> LLL
1	VMAP <u>W</u> SLLL
5	V <u>T</u> APRTLLL
1	VMAP <u>L</u> TLIL
1	<u>A/V</u> MA <u>L/P</u> RTLIL
1	VMAP <u>R/W</u> TLLL

Les acides aminés mutés sont en gras et soulignés.

Trois sites potentiels de N-glycosylation ont été suggérés dans cette protéine (acides aminés 89-91, 110-112 et 143-145) (Tomasec *et al.*, 2000). Pour les sites putatifs de N-glycosylation de pUL40, les régions 110-112 et 143-145 sont conservées pour toutes les souches. En position 89-91, sept isolats présentent la séquence NNT décrite chez la souche AD169, alors que onze isolats de greffés rénaux et 106Bx dans les deux compartiments étudiés ont la séquence SNT comme les souches Toledo, Merlin et Towne.

Figure 18 : Variabilité de la protéine pUL40

Strain	MNKFSNTRIG	FTCAV N APRT	LILLTVGLLGM	RIRSLCSPA	ETTIVTTAAVT	SAHGPLCLPLV	FQGWAYAVYH	QQDMALMTLD	VIYCCRQTSNN	TVVAFSHHPA	DNTLLIEVGN	NTRRHVDGIS	QQDFHRAHQ	DCPAQTVHVR	GVN S SAFGLT	HLQSCCLNEH	SQLSERVAYH	LKLRPATFGL	ETWAMYTVGI	LALGFSFSFY	SQIARSLGVL	PNDHHYALKK	A*		
AD169																									
101Bx0406W		I	L																						
101Bx0406B		I	L																						
104Bx0604W																									
104Bx0604B																									
5Bx0502																S							M		
7Bx0506																								M	
9Bx0411																								M	
16Bx0503	Y	A																						M	
8Bx0402																									
8Bx0312						S					G														
8Bx0501																S									
8Bx0410																									
8Bx0604																									
8Bx0610																									
8Bx0405																									
10Bx0502																									
10Bx0504																									
10Bx0508																									
12Bx0501																									
12Bx0505																									
Merlin																									
106Bx0411B																									
106Bx0411W																									
15Bx0506																									
11Bx0409																									
11Bx0412																									
1Bx0408																									
6Bx0409																									
9Bx0502																									
9Bx0501																									
9Bx0504																									
Toledo																									
101Bx0508B																									
2Bx0410																									
101Bx0509W																									
14Bx0507																									
101Bx0705B																									
101Bx0705W																									
102Bx0602W																									
102Bx0608B																									
102Bx0603B																									
102Bx0605W																									
102Bx0605B																									
103Bx0607W																									
103Bx0607B																									
1Bx0411																									
17Bx0508																									
3Bx0503																									
Towne																									
1Bx0501																									
105Bx0502W																									
105Bx0502B																									
2Bx0503																									
3Bx0412																									
13Bx0408																									
4Bx0409																									

La figure 18 illustre la comparaison de la séquence protéique pUL40 de la souche AD169 avec les séquences protéiques de tous nos isolats. Quand la séquence protéique est différente, l'acide aminé correspondant est indiqué sur la figure. Le nonapeptide est signalé en rouge sur la figure et les trois sites potentiels de N-glycosylation en vert.

2. Les deux gènes immunodominants

a) UL123 codant IE1

Cette étude rétrospective a été réalisée pour tous nos patients greffés rénaux et pulmonaires. Sur 55 échantillons, 42 séquences ont pu être obtenues. Le détail des résultats est présenté dans le tableau suivant 9 :

Tableau 9 : Résultats d'amplification du gène UL123

Patient	Code patient	UL123	Patient	Code patient	UL123	
1Bx	1Bx0408	+	101Bx	101Bx0406WB	-	
	1Bx0411	+		101Bx0406BAL	-	
	1Bx0501	+		101Bx0508BAL	+	
2Bx	2Bx0410	+	101Bx0509WB	-		
	2Bx0503	+	101Bx0705BAL	-		
3Bx	3Bx0412	+	101Bx0705WB	-		
	3Bx0503	+	102Bx	102Bx0602WB	-	
4Bx	4Bx0409	+		102Bx0603BAL	+	
	5Bx	5Bx0502	+	102Bx0605BAL	+	
6Bx		6Bx0409	+	102Bx0605WB	-	
	7Bx	7Bx0506	+	102Bx0608BAL	-	
8Bx		8Bx0312	+	102Bx0608WB	-	
	8Bx0402	+	103Bx	103Bx0607BAL	+	
	8Bx0405	+		103Bx0607WB	+	
	8Bx0410	+	104Bx	104Bx0604BAL	+	
	8Bx0501	+		104Bx0604WB	+	
9Bx	8Bx0604	-	105Bx	105Bx0502BAL	+	
	8Bx0610	-		105BX0502WB	+	
	9Bx	9Bx0411	+	106Bx	106Bx0411WB	+
		9Bx0501	+		106Bx0411BAL	+
	10Bx	9Bx0502	+			
9Bx0504		-				
10Bx		10Bx0502	+			
		10Bx0504	+			
11Bx	10Bx0508	+				
	11Bx0409	+				
12Bx	11Bx0412	+				
	12Bx0501	+				
13Bx	12Bx0505	+				
	13Bx0408	+				
14Bx	14Bx0507	+				
15Bx	15Bx0506	+				
16Bx	16Bx0503	+				
17Bx	17Bx0508	+				

+ : obtention d'une séquence

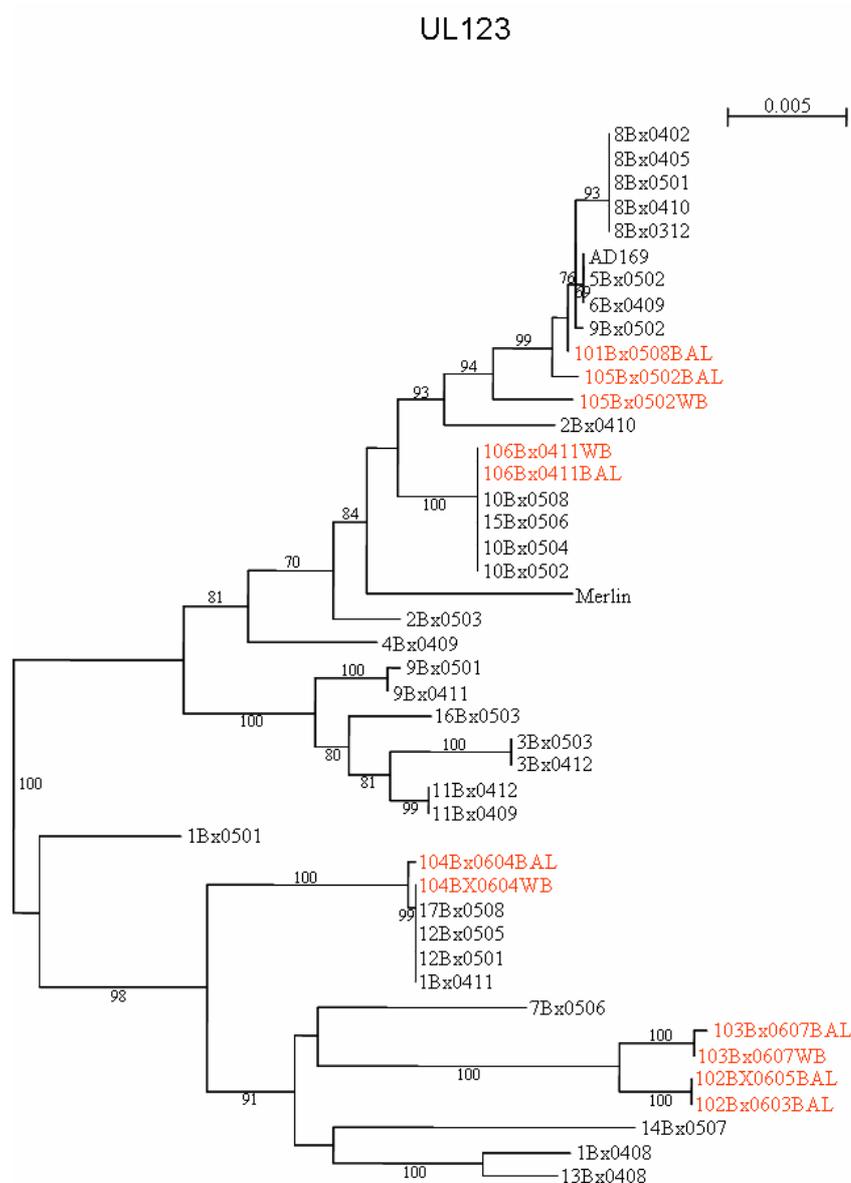
- : échec d'amplification ou de séquençage

Dans le gène UL123, une délétion de neuf nucléotides (ACAGGAGGA) est présente dans l'exon 4 de la souche 16Bx0503 ; de plus, l'intron 2 a une longueur variable (114 nucléotides pour AD169, et de 111 à 116 nucléotides dans nos isolats cliniques).

La charge virale du CMV dans les prélèvements pour lesquels nous n'avons pas obtenu d'amplification est faible. Ceci pourrait expliquer cet échec.

L'arbre phylogénique (figure 19) obtenu après séquençage direct montre une variabilité non négligeable. Cette variabilité peut être nulle (par exemple pour 10Bx, 15Bx et 106Bx) ou aller jusqu'à 6.2% de la séquence nucléique totale (par exemple pour 13Bx et 2Bx0503).

Figure 19 : Arbre phylogénique du gène UL123



Après traduction en acides aminés de la séquence codante, en comparaison avec la protéine IE1 de la souche AD169, des mutations dans les épitopes de IE1 ont été détectées pour de nombreux isolats et sont résumées dans le tableau 10 :

Tableau 10 : Mutations des épitopes d'IE1

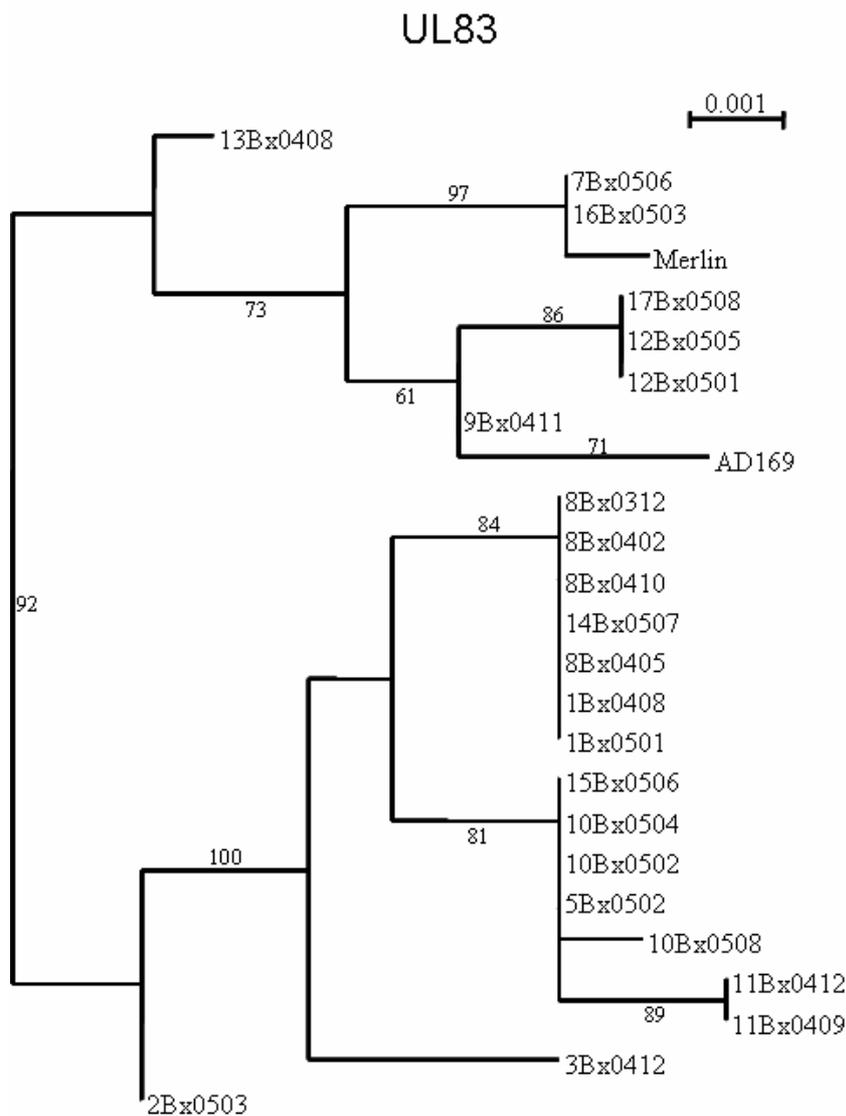
Position de l'épitope	Séquence sauvage (AD169)	HLA	Séquence mutée	Nombre d'isolats
87-95	KQIKVRVDM	C*06	KQIKIRVDM	38
			KQIK(I/V)RVDM	2
198-207	DELKRKMIYM	B*08	DELKRKMMYM	7
			DELRRKMMYM	4
			EELRRKMMYM	23
			DEL(K/R)RKMMYM	5
			DEL(K/R)RKM(I/M)YM	1
199-207	ELKRKMIYM	B18	ELKRKMMYM	2
			ELRRKMMYM	7
			EL(K/R)RKMMYM	4
			EL(K/R)RKM(I/M)YM	28
279-287	CVETMCNEY	B18	CVETMCSEY	1
			CVETMC(N/S)EY	26
				14
297-304	TMYGGISL	A*0201	TMYGSISL	2
			GMYGAISL	16
			TMYG(S/P)ISL	7
			TMYG(S/T/R/P)ISL	14
			(T/S/A/G)MYG(P/T)ISL	2
			(T/S/A/G)MYG(G/A)ISL	1
				1
305-313	LSEFCRVLC	C*07	LTEFCRVLC	25
			L(S/T)EFCRVLC	14
				2
308-316	FCRVLCYV	C*07	FCRVLCYI	23
			FCRVLCY(V/I)	14
309-317	CRCVLCCYVL	B*0702	CRCVLCCYIL	5
			CRCVLCCY(V/I)L	23
				14
312-320	LCCYVLEET	C*07	LCCYILEET	5
			LCCY(V/I)LEET	23
315-323	YILEETSVM	A02(*0201)	YILEETSVL	7
			YVLEETSVM	6
			YVLEETSVL	16
			YILEETSV(L/M)	7
			Y(I/V)LEETSVM	1
			Y(I/V)LEETSVL	3
			Y(I/V)LEETSV(M/L/I)	1
316-324	ILEETSVML	A*0201	ILEETSVLL	1
			VLEETSVML	7
			VLEETSVLL	16
			ILEETSV(L/M)L	7
			(I/V)LEETSVML	1
			(I/V)LEETSVLL	3
			(I/V)LEETSV(M/L/I)(L/M)	1
				1
319-327	ETSVMLAKR	A*68	ETSVLLAKR	26
			ETSV(M/L)LAKR	14
			ETSV(M/L/I)LAKR	1
				1
354-363	YILGADPLRV	A02(*0201)		42
379-387	DEEDAIAAY	B18	DEEEAIVAY	20
			DEENAIVAY	18
			DEEEAI(A/V)AY	1
			DEE(D/E)AIVAY	1
			DEE(D/E)AIAAY	1
				1
				1

L'impact théorique de cette variabilité sur la capacité de l'épitope à se lier aux molécules du CMH a été étudié avec les logiciels BIMAS et SYFPEITHI mais nous n'avons pas trouvé de modification du score d'affinité.

b) UL83 codant pp65

Cette étude a été réalisée avec les souches isolées chez les transplantés rénaux. Nous avons obtenu une séquence partielle correspondant à la partie 5' du gène UL83 (les 1125 premiers nucléotides) pour 23 prélèvements sur 34. Au niveau nucléique, nous observons jusqu'à 1.5% de variabilité entre deux souches (11Bx0412 et 17Bx0508, par exemple). La figure 20 illustre l'arbre phylogénique obtenu pour ces séquences partielles.

Figure 20 : Arbre phylogénique du gène UL83



Au niveau protéique, dans la partie N-terminale de la séquence disponible, la variabilité est quasiment nulle (avec un maximum de 0,5% de divergence entre les souches 10Bx0508 et 11Bx0412 par exemple). Les épitopes préalablement décrits dans la souche AD169 pour la protéine pp65 (Kern *et al.*, 2000; Trivedi *et al.*, 2005) ne sont mutés pour aucune de nos souches cliniques. La figure 21 illustre la variabilité de la protéine pp65.

Figure 21 : Variabilité de la protéine pp65 (375 premiers acides aminés).

AD169	MESRGRRCPE	MISVLGPISG	HVLKAVFSRG	DTPVLPHETR	LLQTGIHVRV	SQPSLLILVSQ	YTPDSTPCHR	GDNLQVQHT	YFTGSEVENV	SVNVHNPTGR
1Bx0501
14Bx0507
8Bx0410
1Bx0408
8Bx0402
8Bx0312
8Bx0405
11Bx0409
11Bx0412
5Bx0502
10Bx0508
10Bx0502
10Bx0504
15Bx0506
3Bx0412
2Bx0503
9Bx0411
12Bx0501
12Bx0505
17Bx0508
16Bx0503
7Bx0506
13Bx0408
AD169	SICPSQEPMS	IYVYALPLKM	LNIPISNVHH	YPSAAERKHH	HLPVADAVIH	ASGKQMWQAR	LTVSGLAWTR	QQNQWKEPDV	YYTSAFVFPT	KDVALRHVVC
1Bx0501
14Bx0507
8Bx0410
1Bx0408
8Bx0402
8Bx0312
8Bx0405
11Bx0409
11Bx0412
5Bx0502
10Bx0508
10Bx0502
10Bx0504
15Bx0506
3Bx0412
2Bx0503
9Bx0411
12Bx0501
12Bx0505
17Bx0508
16Bx0503
7Bx0506
13Bx0408
AD169	AHELVCSMEN	TRATKMQVIG	DQYVKVYLES	FCEDVPSGKL	FMHVTLGSDV	EEDLTMTRNP	QPFMRPHERN	GFTVLCPKNM	IIKPKGISHI	MLDVAFTSHE
1Bx0501
14Bx0507
8Bx0410
1Bx0408
8Bx0402
8Bx0312
8Bx0405
11Bx0409
11Bx0412
5Bx0502
10Bx0508
10Bx0502
10Bx0504
15Bx0506
3Bx0412
2Bx0503
9Bx0411
12Bx0501
12Bx0505
17Bx0508
16Bx0503
7Bx0506
13Bx0408
AD169	HFGLLCPKSI	PGLSISGNLL	MNGQIFLEV	QAIRETVELR	QYDPVAALFF	FDIDLLLRG	PQYSEHPTFT	SQYRI		
1Bx0501		
14Bx0507		
8Bx0410		
1Bx0408		
8Bx0402		
8Bx0312		
8Bx0405		
11Bx0409		
11Bx0412		
5Bx0502		
10Bx0508		
10Bx0502		
10Bx0504		
15Bx0506		
3Bx0412		
2Bx0503		
9Bx0411		
12Bx0501		
12Bx0505		
17Bx0508		
16Bx0503		
7Bx0506		
13Bx0408		

Quand un point est indiqué, cela signifie que l'acide aminé correspond à celui de la souche AD169. Quand la séquence protéique diffère, l'acide aminé correspondant est indiqué sur la figure.

C. Compartimentation

Six patients transplantés pulmonaires (trois hommes, trois femmes) ont été inclus dans cette étude (patients 101Bx à 106Bx). Quatre patients sont donneur positif / receveur positif (D+/R+) et deux D+/R-. Pour ces deux derniers patients des prélèvements séquentiels dans les deux compartiments étaient disponibles et espacés d'au moins six semaines (délai pour lequel nous considérons qu'il s'agit d'un épisode d'infection à CMV différent).

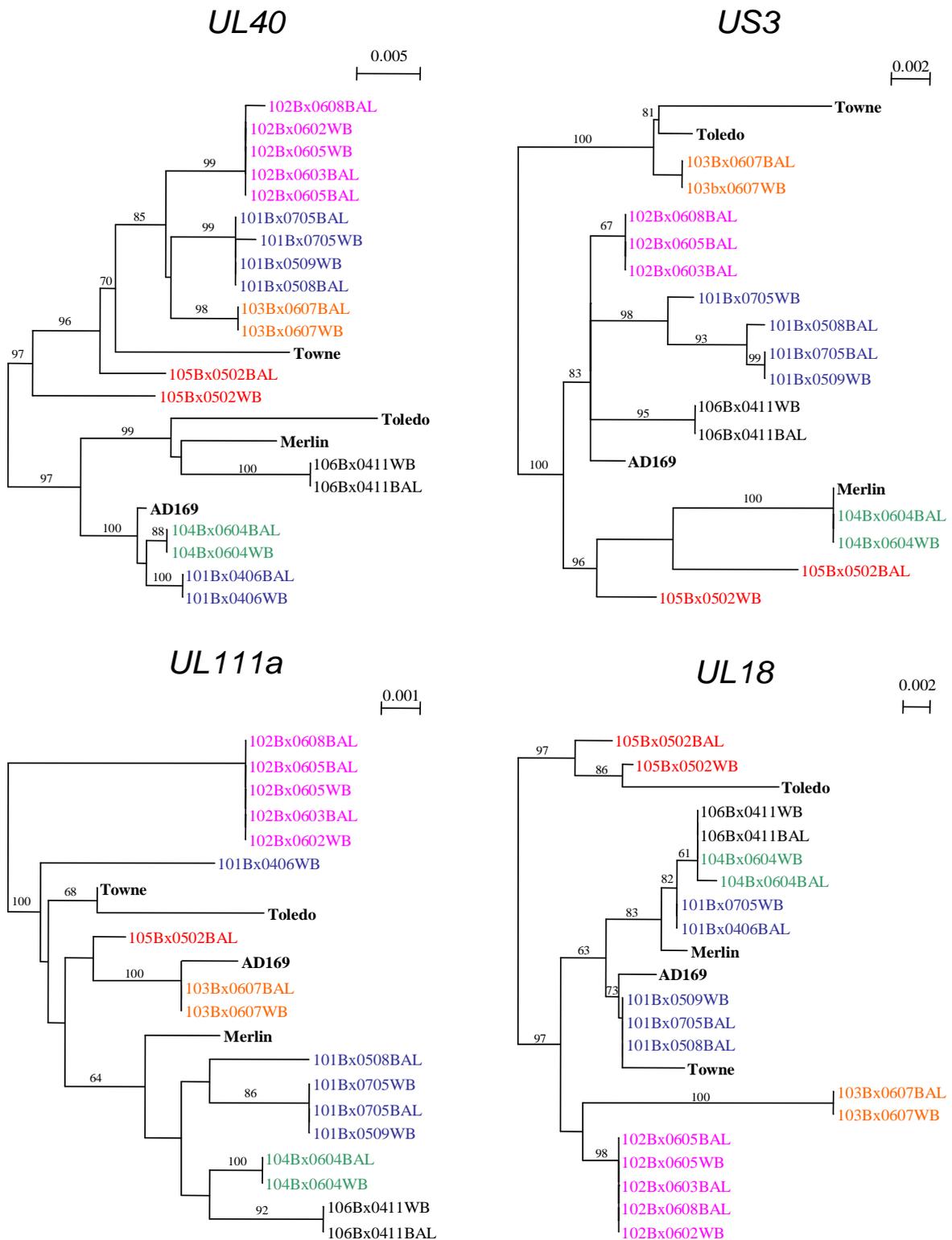
L'ADN du CMV a été amplifié et séquencé à partir de 10 échantillons de sang total et 10 liquides de lavage broncho-alvéolaire.

Les séquences des gènes *UL40*, *US3*, *UL18* et *UL111a* ont été obtenues pour les compartiments sang et poumon pour les six patients et pour deux, à trois temps différents. Les prélèvements pour lesquels une séquence directe a été obtenue sont résumés dans le tableau 11 et les arbres phylogéniques dans la figure 22.

Tableau 11 : Séquences directes disponibles (+) ou non (-)

Code patient	UL40	US3	UL111a	UL18
101Bx D+R- puis D-R+				
101Bx0406WB	+	-	+	-
101Bx0406BAL	+	-	-	+
101Bx0508BAL	+	+	+	+
101Bx0509WB	+	+	+	+
101Bx0705BAL	+	+	+	+
101Bx0705WB	+	+	+	+
102Bx D+R-				
102Bx0602WB	+	-	+	+
102Bx0603BAL	+	+	+	+
102Bx0605BAL	+	+	+	+
102Bx0605WB	+	-	+	+
102Bx0608BAL	+	+	+	+
102Bx0608WB	-	-	-	-
103Bx D+R+				
103Bx0607BAL	+	+	+	+
103Bx0607WB	+	+	+	+
104Bx D+R+				
104Bx0604BAL	+	+	+	+
104Bx0604WB	+	+	+	+
105Bx D+R+				
105Bx0502BAL	+	+	+	+
105BX0502WB	+	+	-	+
106Bx D+R+				
106Bx0411WB	+	+	+	+
106Bx0411BAL	+	+	+	+

Figure 22 : Arbres phylogéniques obtenus avec les séquences directes pour les 4 gènes du CMV.



en gras : séquences des souches choisies comme références.

bleu 101Bx, mauve 102Bx, orange 103Bx, vert 104Bx, rouge 105Bx et noir 106Bx.

Dans certains cas, les séquences des quatre gènes sont identiques dans les deux compartiments (103Bx et 106Bx).

Dans d'autres, trois gènes sont identiques dans les deux compartiments étudiés (104Bx et 102Bx). Pour 104Bx, en comparant la séquence du gène UL18 (1107 pb) dans le sang et le liquide de lavage broncho-alvéolaire, nous avons identifié trois nucléotides différents. Pour 102Bx, pour lequel nous disposons de prélèvements séquentiels, la variabilité observée concerne le gène UL40 dans le compartiment du LBA au troisième temps ; dans le sang total, nous n'avons pas pu amplifier le gène US3 quelque soit le temps, ni aucun gène pour le troisième temps.

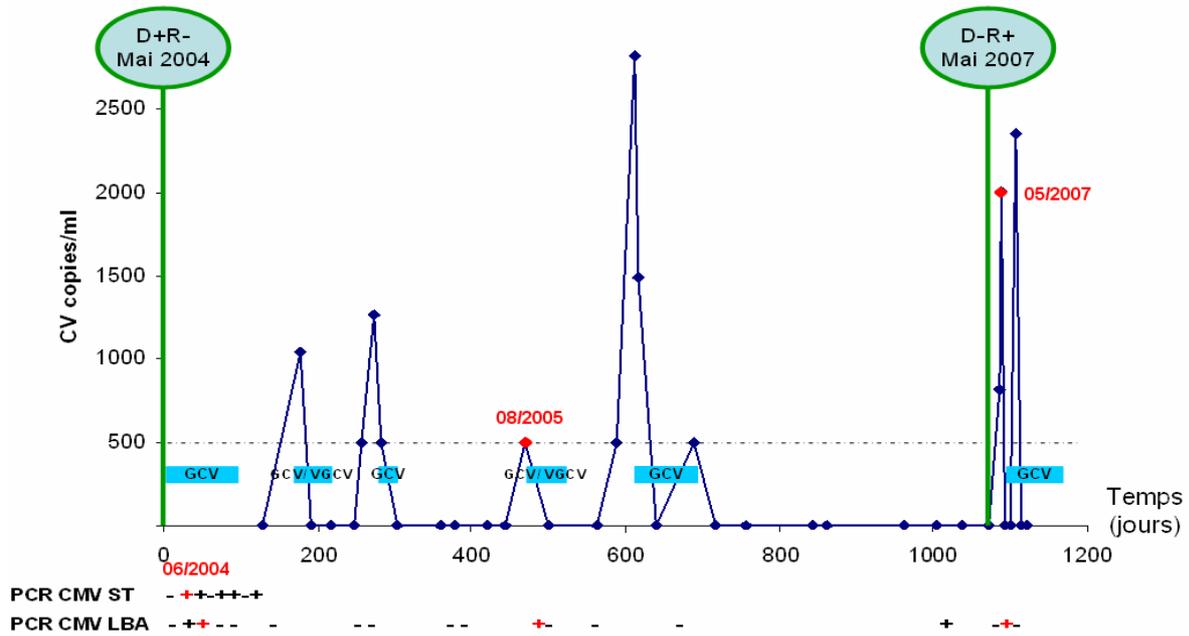
Pour 105Bx (D+/R+), trois gènes sont différents en comparant les deux compartiments (la séquence du gène UL111a n'a pas pu être amplifiée dans le compartiment sang total). Des doubles populations sont observées pour chaque séquence directe.

Le patient 101Bx a été transplanté deux fois. Les séquences directes changent entre le premier épisode infectieux (juin 2004), le deuxième (septembre 2005) et le troisième (mai 2007). Suite à cette observation, nous avons sélectionné les souches isolées chez le patient 101Bx pour cloner les produits de PCR du gène UL40 avant séquençage.

Analyse de la variabilité de UL40 pour le patient 101Bx (séquençage après clonage)

Ce patient a bénéficié de deux transplantations bi-pulmonaires : la première fin mai 2004 (patient D+R-), la seconde trois ans plus tard (patient devenu D-R+). Le premier épisode d'infection à CMV, probablement transmis par le greffon, est survenu quelques jours après la première transplantation (figure 23).

Figure 23 : Histoire de l'infection à CMV du patient 101Bx après ses deux transplantations.



Sang total : PCR qualitative (PCR CMV ST, 1^{ère} ligne sous le graphe) puis PCR quantitative (courbe bleue)

LBA : PCR qualitative uniquement (PCR CMV LBA, 2^{ème} ligne sous le graphe)

Traitement : GCV = Ganciclovir ; GCV / VGCV = Ganciclovir poursuivi par cure de valganciclovir (la durée du traitement est matérialisée par la longueur de la bande bleue)

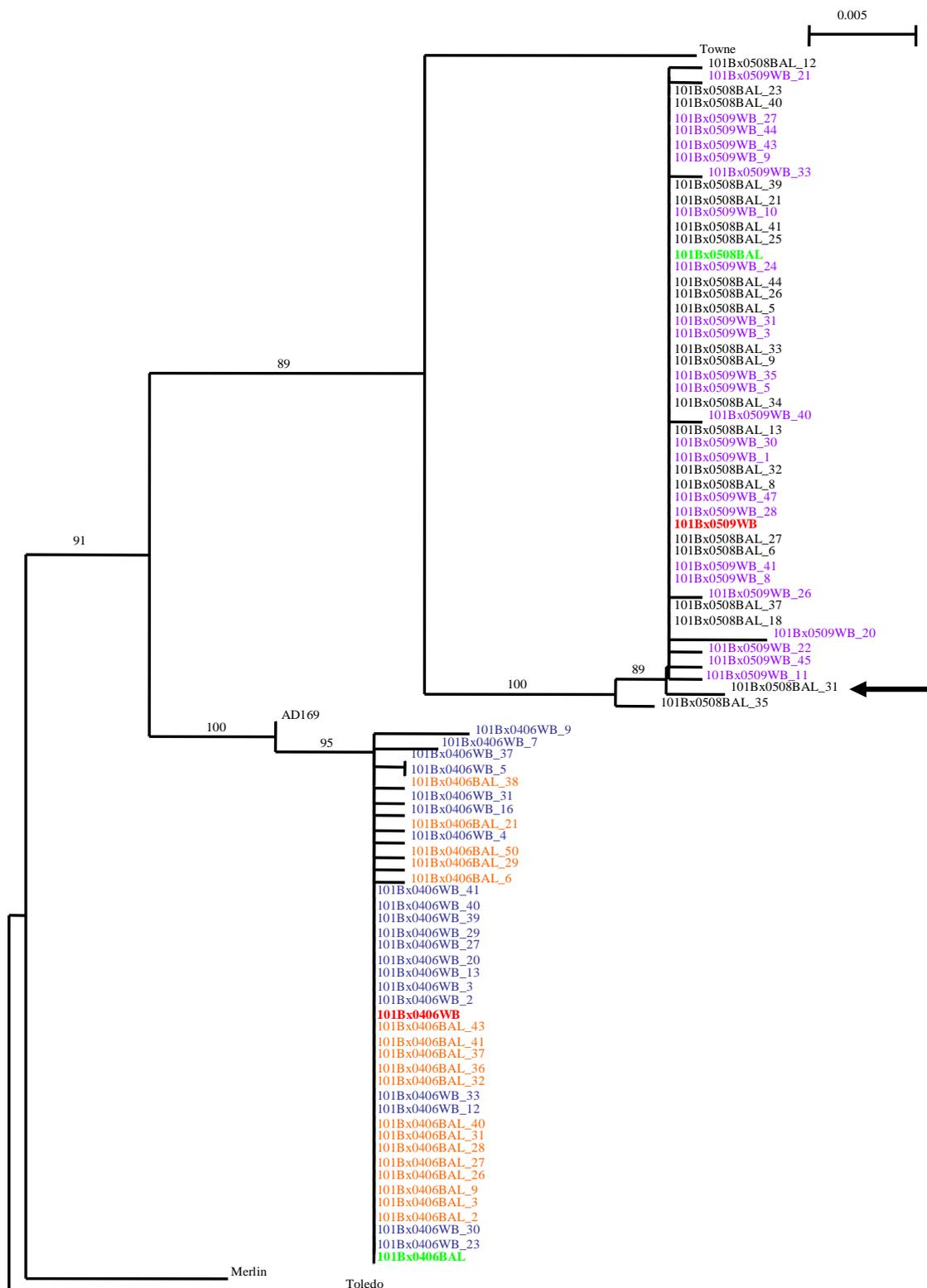
Transplantation bi-pulmonaire

+ ● Prélèvement utilisé dans notre étude avec la date correspondante

Les échantillons positifs dont la CV < 500 copies/ml sont placés sur la ligne horizontale qui représente le seuil de linéarité de la quantification.

L'arbre phylogénique issu du séquençage des clones *UL40* (figure 24) dans les 2 compartiments du patient 101Bx conforte les résultats du séquençage direct : les clones des temps 1 et 2 forment des « clusters » séparés. Ceci est confirmé par des valeurs élevées de bootstrap. On n'observe pas de phénomène de compartimentation.

Figure 24 : Arbre phylogénique des clones *UL40* de 101Bx des temps 1 et 2



Séquence directe temps 1 et 2 : vert compartiment pulmonaire, rouge compartiment sanguin.

Séquence clones Temps 1 : bleu compartiment sanguin, orange compartiment pulmonaire.

Séquence clones Temps 2 : mauve compartiment sanguin, noir compartiment pulmonaire.

Les valeurs de bootstrap sont indiquées au-dessus des branches. Le clone 101Bx0508BAL_31 est indiqué par une flèche.

Le clone 31 du LBA temps 2 présente des mutations retrouvées dans les séquences directes des 2 premiers temps.

Ceci peut être expliqué par un événement de recombinaison comme le suggère l'analyse par Bootscan (figure 25). La séquence du clone BAL 31 ressemble à la séquence directe 101Bx0406BAL (courbe bleue) dans sa région 5' et à la séquence directe 101Bx0508BAL (courbe noire) dans sa région 3'. Le clone BAL 31 pourrait être issu de la recombinaison de ces 2 souches parentales.

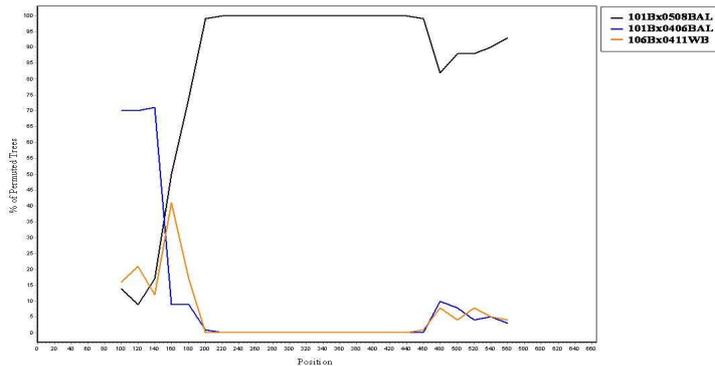


Figure 25 :

Bootscan réalisé en comparant le clone *UL40* BAL 31 temps 2 avec les séquences directes de *UL40* compartiment poumon temps 1 (bleu) et 2 (noir) pour 101Bx et la séquence directe d'*UL40* compartiment sanguin pour 106Bx (comme témoin, orange).

D. Recombinaison

Vingt et un patients (13 hommes, 8 femmes ; âge médian de 55 (44 ; 56) ans au moment de la transplantation) ont été inclus dans cette étude car ils développaient une infection à CMV. Seize patients avaient reçu un rein (patients 1Bx à 17Bx) et cinq un poumon (patients 101Bx à 106Bx). Huit patients avaient un statut sérologique de donneur positif / receveur positif (D+/R+), un patient était D-/R+ et 12 D+/R-. Les souches de CMV ont été isolées dans le sang de 19 patients (1Bx à 17Bx et 103Bx à 106Bx) et dans le liquide de lavage broncho-alvéolaire pour deux patients (101Bx et 102Bx).

Fig. 26A

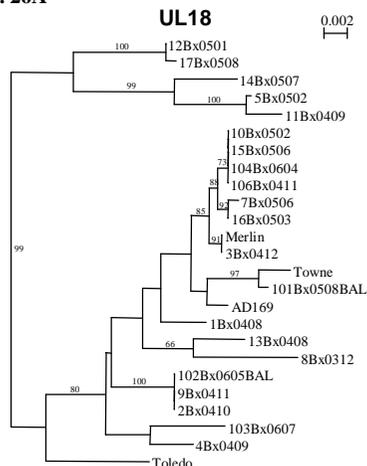


Fig. 26D

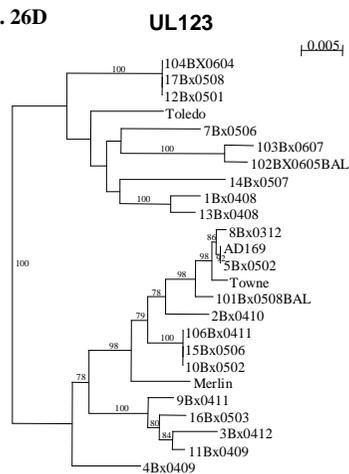


Fig. 26B

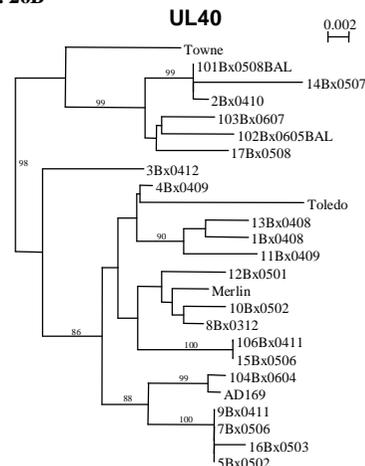


Fig. 26E

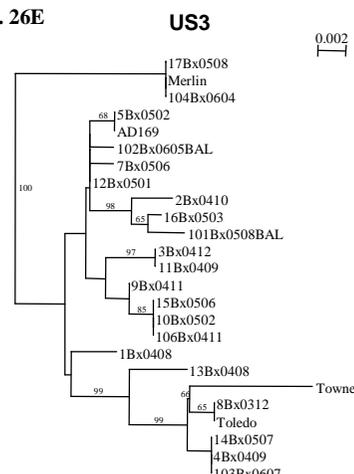


Fig. 26C

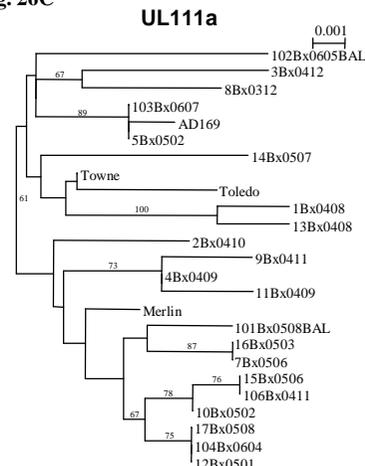


Fig. 26F

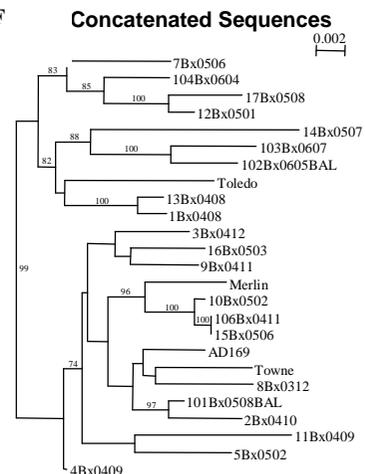


Figure 26 : Représentations phylogéniques de la variabilité nucléique du CMV pour les cinq gènes étudiés et pour la séquence concaténée. Les souches sont ainsi identifiées : - numéro du patient (1 à 17 pour les transplantés rénaux ou 101 à 106 pour les greffés pulmonaires)
 -Bx (Bordeaux, France)
 -année (03, 04,0 5 ou 06)
 -mois du prélèvement (01 à 12).

Les séquences sont comparées avec des séquences homologues disponibles dans Genbank: AD169 (numéro d'accès NC_001347), Merlin (NC_006273), Towne (AY072777), et Toledo (AH013698).

FIG. 26A, 26B, 26C, 26D et 26E illustrent les séquences des gènes UL18, UL40, UL111a, UL123 et US3.

FIG. 26F illustre la variabilité nucléique des séquences concaténées (4800 nucléotides en moyenne).

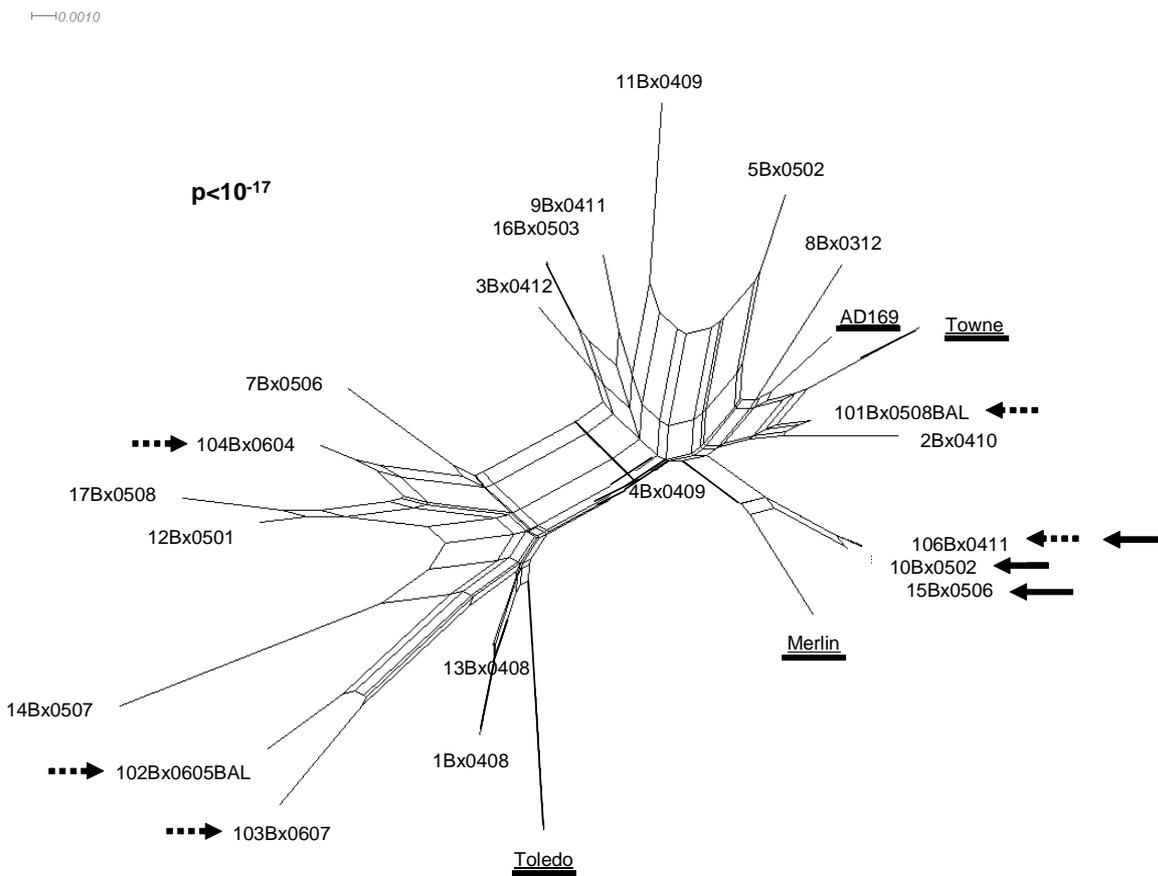
1. Arbres phylogéniques et analyse de la recombinaison avec le logiciel SplitsTree

Des arbres phylogéniques sont construits pour chaque gène et pour les séquences concaténées (figure 26). Les valeurs de bootstrap confirment la robustesse de ce calcul.

En considérant les séquences des cinq gènes, tous les isolats cliniques présentent un profil génétique unique et chaque patient est infecté par un variant CMV particulier. Néanmoins, pour un gène donné comme UL18, des séquences homologues peuvent être identiques chez différents isolats (10Bx0502, 15Bx0506, 104Bx0604 et 106Bx0411 ou 2Bx0410, 9Bx0411 et 102Bx0605BAL, par exemple). Dans l'arbre phylogénique issu des séquences concaténées, une distribution en deux clusters dominants est observée. L'identité nucléique est de 95.4% pour les deux souches les plus éloignées (14Bx et 4Bx) et de 99.9% pour les deux plus proches (106Bx et 15Bx). Ces deux dernières souches clustérisent ensemble quelque soit le gène avec des séquences identiques pour les gènes UL18, UL40, UL123 et US3. Pour le gène UL111a, il existe une insertion de 6 nucléotides dans la souche 106Bx comparée à celle de 15Bx.

Cette approche phylogénique traditionnelle n'est pas suffisante pour suggérer une recombinaison. Pour chaque gène et pour les séquences concaténées, nous avons construit des réseaux de recombinaison en utilisant le logiciel SplitsTree et en incluant tous les isolats cliniques et quatre souches de référence (Figures 27A à I).

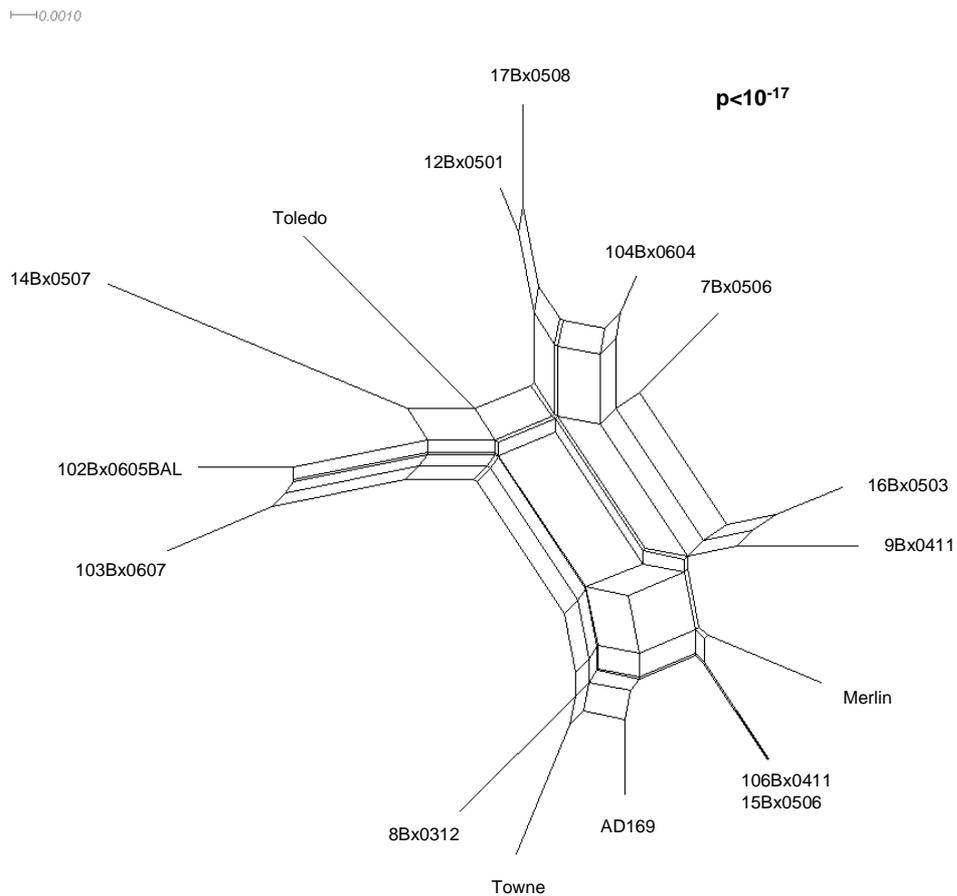
Figure 27A : Séquences concaténées



La figure 27A illustre l'analyse par SplitsTree des séquences concaténées des cinq gènes. Les séquences des virus infectant les transplantés pulmonaires sont indiquées par une flèche en pointillé et les souches de référence sont soulignées. Les souches présentées dans la figure 27A après analyse par bootscan, que nous décrivons plus loin, sont indiquées avec des flèches pleines. Le réseau présente un profil réticulé renforçant l'hypothèse de souches évoluant par recombinaison. Le Phi-test, calculé à partir de l'alignement des séquences, détecte 297 sites informatifs (297/4804 nucléotides) et met en évidence des événements de recombinaison de façon statistiquement significative ($p < 10^{-17}$).

La présence de doubles populations dans 9 séquences concaténées aurait pu introduire un biais dans l'analyse. C'est pourquoi nous l'avons renouvelée avec les 12 séquences dépourvues de doubles populations et les quatre souches de référence. Les patients correspondants sont 3 D+/R+, un D-/R+ et 8 D+/R- (Figure 27B).

Figure 27B : Séquences concaténées sans double population



Comme précédemment, le Phi-test est statistiquement significatif ($p < 10^{-17}$) et le profil réticulé de l'analyse est en faveur d'une évolution du CMV par recombinaison.

Nous avons ensuite analysé avec SplitsTree les alignements obtenus pour chaque gène. Tous les réseaux présentent un profil réticulé, renforçant l'hypothèse de souches évoluant par recombinaison, avec à chaque fois un phi-test statistiquement significatif ($p < 10^{-17}$).

Les réseaux les plus complexes sont ceux du gène UL40 (666 pb) et du gène UL123 (1762 pb en moyenne), reflétant un haut niveau de recombinaison durant l'évolution du CMV. Ces deux gènes sont aussi ceux qui présentent le plus grand nombre de sites variables (12% et 11,8% respectivement). Le gène US3 est le moins complexe.

Dans nos travaux précédents, nous avons cloné les produits de PCR du gène UL40 pour des souches infectant des patients, inclus dans la présente étude, pour lesquels nous disposions de prélèvements séquentiels (Garrigue et al., 2008b)). L'analyse par bootscan et l'arbre phylogénétique étaient en faveur d'événements de recombinaison pour un patient (1Bx) et l'arbre phylogénétique était en faveur d'une évolution par quasi-espèces pour un autre (9Bx). Nous avons analysé ces séquences clonées (22 pour 9Bx et 14 pour 1Bx) avec le programme SplitsTree (Figure 27H et I).

Figure 27H: Clones 1Bx UL40

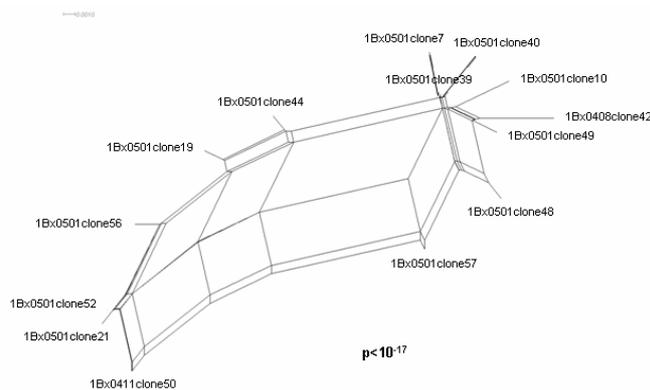
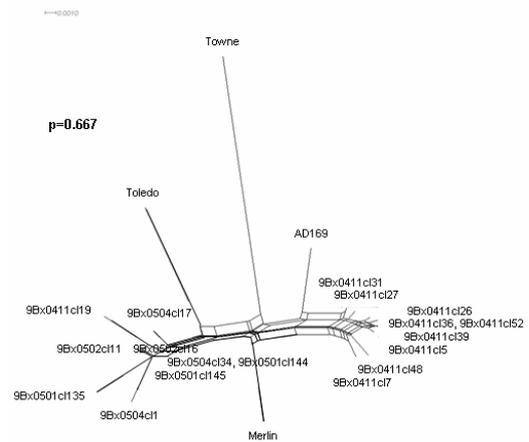


Figure 27I: Clones 9Bx UL40



Nos précédentes hypothèses sont confortées avec, pour le patient 1Bx, un Phi-test statistiquement significatif ($p < 10^{-17}$) en faveur d'une évolution par recombinaison, alors que pour le patient 9Bx, le Phi-test n'est pas statistiquement significatif ($p = 0,667$) et donc en défaveur d'une évolution par recombinaison. Des résultats obtenus avec des clones du gène UL40 peuvent ainsi préfigurer ceux obtenus avec les séquences concaténées.

2. Analyse par bootscan

Les séquences concaténées ont été analysées par la méthode du bootscan. Elle indique les séquences qui se rapprochent le plus de la séquence analysée, et dans quelle proportion (pourcentage indiqué au niveau de l'axe des ordonnées), tout le long de la séquence

nucléique. La règle au bas du graphique indique la position et la longueur des cinq gènes étudiés. Bien que ces séquences ne soient pas clonées, elles ne contiennent pas de double population. Ainsi, nos résultats ont peu de chance d'être la conséquence de co-infections. L'illustration de ces résultats est présentée dans les figures 28A à C :

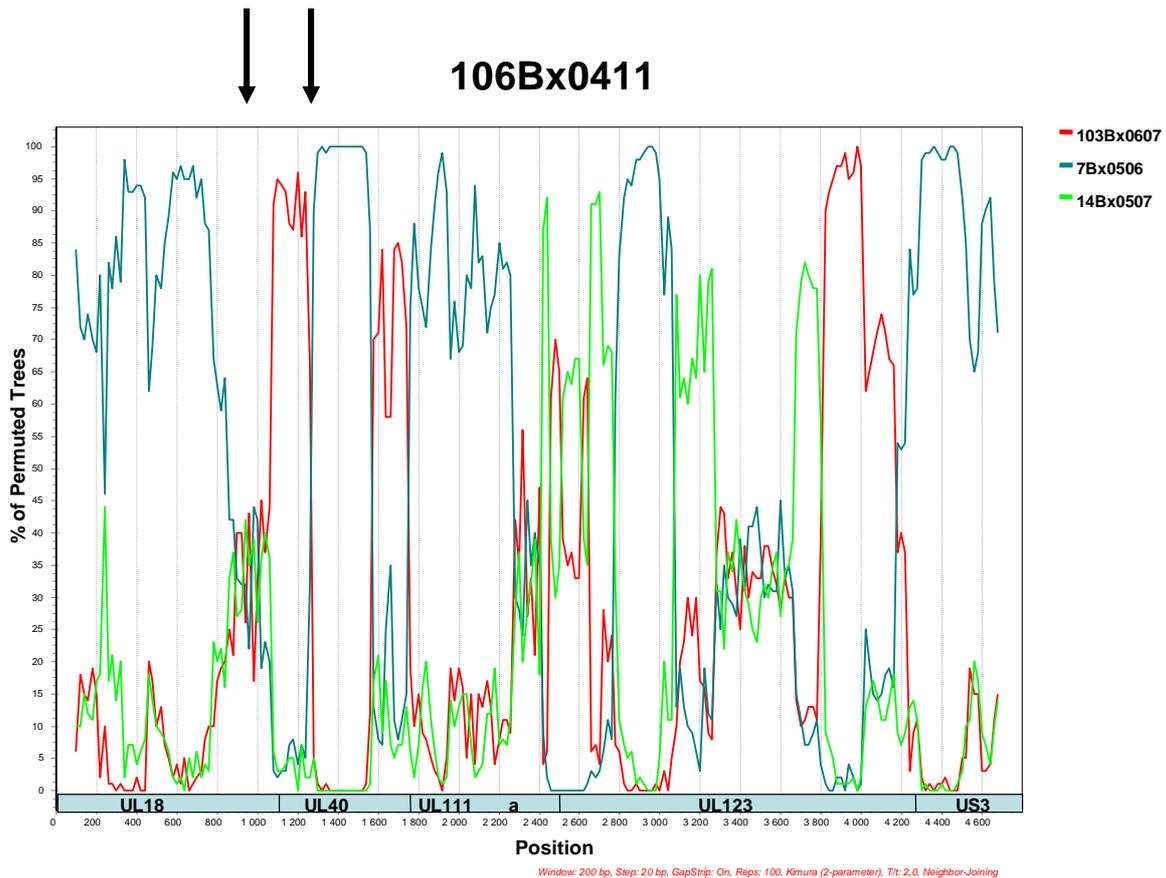


Figure 28A : la séquence 106Bx0411 (D+/R+) est comparée avec celles de 103Bx0607 (D+/R+), 7Bx0506 (D-/R+) et 14Bx0507 (D+/R-). Les 5' et le 3' termini du gène UL40 sont similaires entre 106Bx0411 et 103Bx0607; la séquence centrale du gène UL40 est similaire entre 106Bx0411 et 7Bx0506. Le bootscan considère que ce profil est hautement suggestif d'événements de recombinaison intra-génique au sein du gène UL40. Des sites de recombinaison vraisemblables sont indiqués par des flèches en haut de la figure.

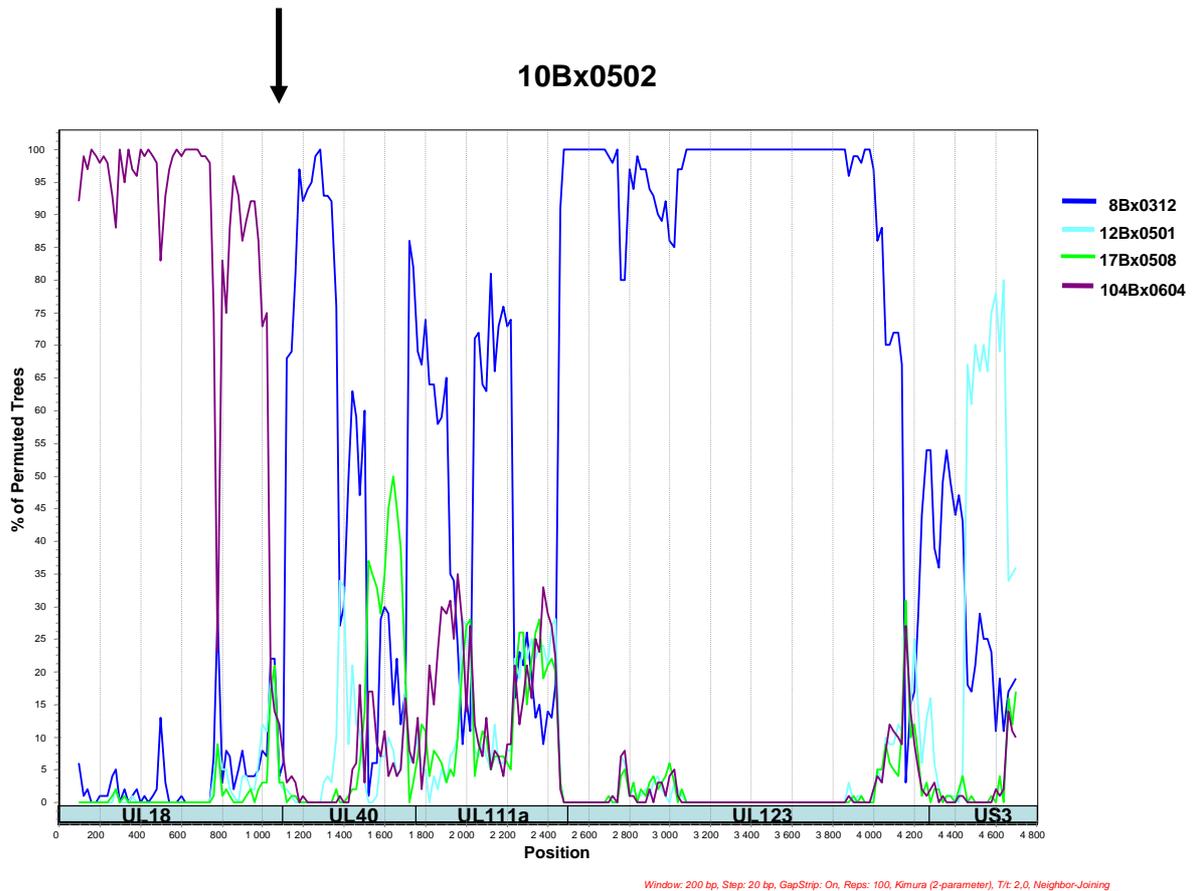


Figure 28B : la séquence 10Bx0502 (D+/R-) est comparée avec celles de 8Bx312 (D+/R-), 12Bx0501 (D+/R-), 17Bx0508 (D+/R-) et 104Bx0604 (D+/R+). La séquence du gène UL18 est similaire entre 10Bx0502 et 104Bx0604. Les séquences des gènes UL40, UL111a, UL123 et US3 sont proches entre 10Bx0502 et 8Bx0312. Le bootscan considère que ce profil est en faveur d'événements de recombinaison qui ont eu lieu dans la séquence génomique entre les gènes UL18 et UL40. La zone de recombinaison est indiquée par une flèche en haut de la figure.

15Bx0506

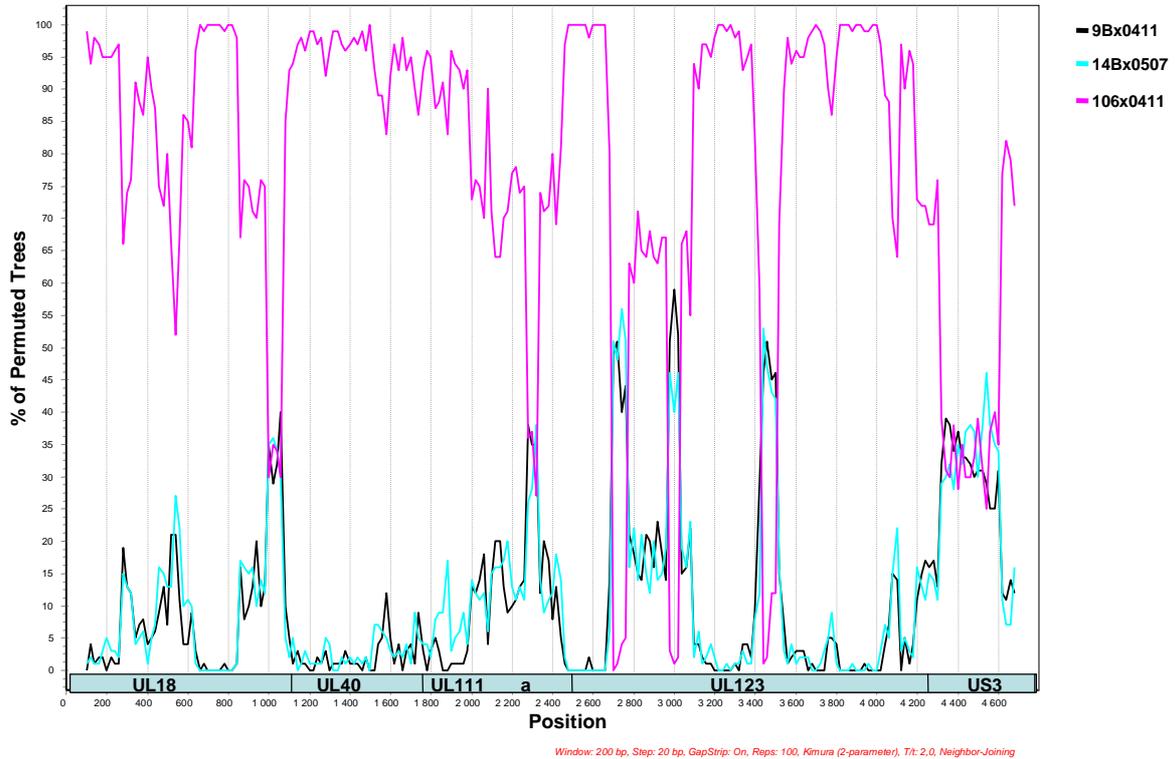


Figure 28C : la séquence 15Bx0506 (D+/R-) est comparée avec celles de 9Bx0411 (D+/R-), 14Bx0411 (D+/R-) et 106Bx0411 (D+/R+). 15Bx0506 est similaire tout au long de la séquence concaténée à 106Bx0411 et le bootscan n'indique pas d'événement de recombinaison.

Nous n'avons observé au cours de nos différentes analyses aucune zone préférentielle de recombinaison.

XIII. Discussion

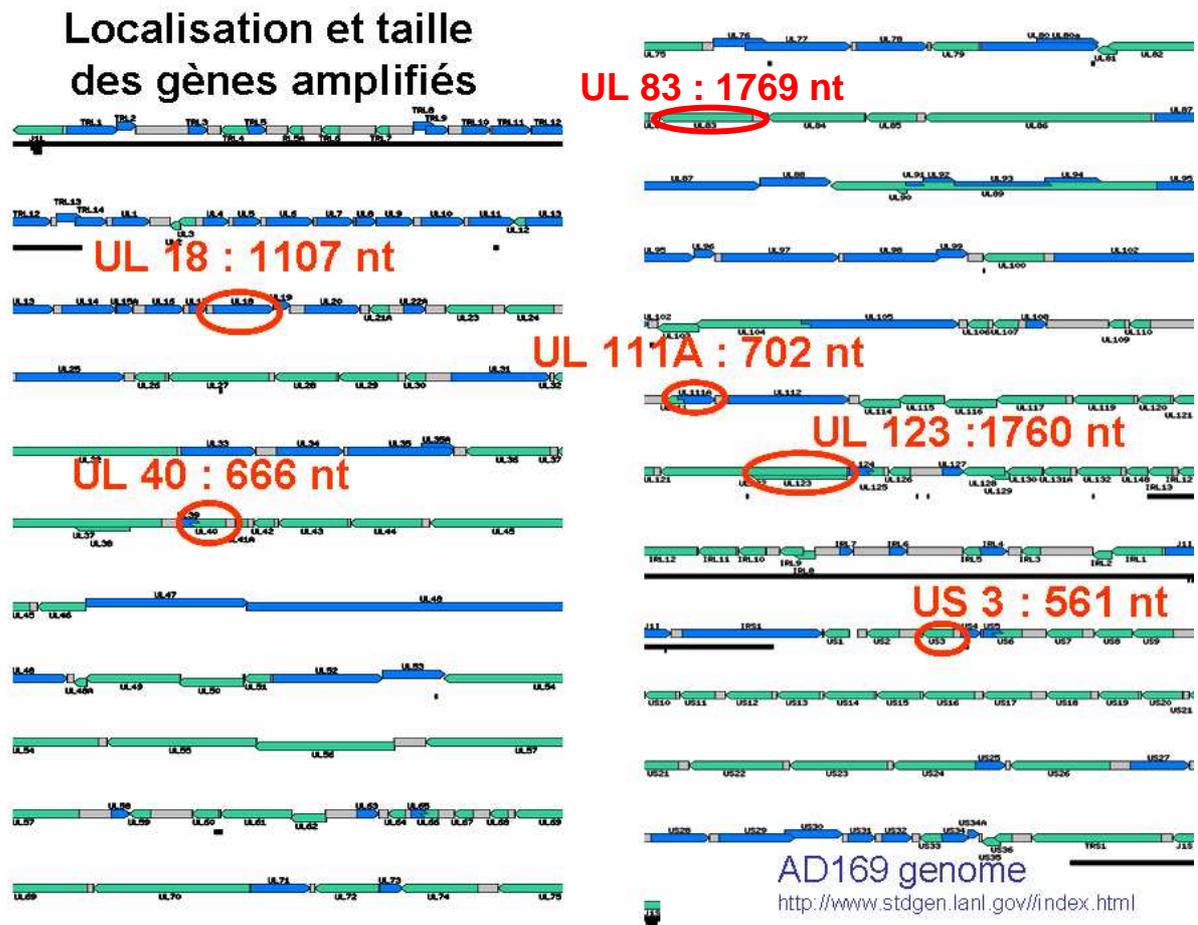
De nombreux paramètres sont susceptibles d'influencer l'évolution d'une infection à CMV : complexité des défenses immunitaires spécifiques et non spécifiques, traitements immunosuppresseurs, variabilité et virulence des souches, traitements anti-viraux ...

D'autres auteurs et nous-mêmes avons montré un polymorphisme significatif pour des souches cliniques de CMV pour lequel différents mécanismes peuvent être impliqués : polymorphisme naturel, recombinaison ou évolution par quasi-espèces (Baldanti *et al.*, 2003; Dolan *et al.*, 2004; Garrigue *et al.*, 2007; Garrigue *et al.*, 2008). Dans nos travaux précédents, nous avons mis en évidence une variabilité intra-génique compatible avec un événement de recombinaison pour le gène UL40 cloné à partir de prélèvements effectués chez un patient infecté et suivi sur du long terme (Garrigue *et al.*, 2008b). Des recombinaisons homologues avaient été montrées pour des souches de laboratoire de CMV et pour des souches cliniques mais toujours après une étape de culture (Arav-Boger, Zong, and Foster, 2005; Baldanti *et al.*, 2003; Chou, 1989; Dolan *et al.*, 2004; Haberland, Meyer-Konig, and Hufert, 1999; Rasmussen, Geissler, and Winters, 2003). Dans ce travail nous avons examiné *ex vivo* des acides nucléiques viraux extraits directement de l'échantillon biologique du patient et nous sommes focalisés sur l'évolution du CMV par recombinaison en comparant des souches cliniques infectant différents patients.

Choix des gènes viraux

Les six gènes sélectionnés UL18, UL40, UL111a, UL123 et US3 sont répartis tout le long du génome du CMV et ne sont pas contigus, comme indiqué dans la figure 29.

Figure 29 : Sélection des gènes viraux étudiés



Notre choix de gène peut être contesté puisque le CMV possède des gènes plus variables (par exemple le gène UL55 codant la glycoprotéine B). Cependant, les mutations sont suffisantes pour que l'analyse par SimPlot conclue à l'existence d'événements de recombinaison.

Pression de sélection ?

Les séquences alignées de nos isolats cliniques de CMV contiennent de nombreux signaux phylogénétiquement conflictuels (297 sites informatifs sur 4800 nucléotides), beaucoup pouvant résulter de recombinaison homologue. La pression de sélection peut aussi conduire à l'accumulation de tels signaux au niveau de sites spécifiques ou d'épitopes. Cependant, notre analyse par bootscan détecte de tels signaux tout au long des cinq gènes ce qui est en défaveur de la pression de sélection.

Artefacts d'amplification ?

Ces signaux peuvent aussi être le résultat d'une substitution nucléotidique introduite au hasard par la Taq polymérase durant la PCR (Zong *et al.*, 2007). Ceci est hautement improbable avec nos conditions expérimentales et l'enzyme AmpliTaq Gold d'Applied Biosystems. Premièrement, le taux d'erreur de cette enzyme est de 5.10^{-6} nucléotides incorporés par cycle (environ une erreur tous les 200000 nucléotides) et l'erreur doit être introduite très tôt dans la réaction d'amplification. Deuxièmement, la séquence consensus générée à partir des brins en 5' et en 3' montre très peu de différences. Les deux brins sont séquencés indépendamment : la présence de la même erreur due à la polymérase, à la même position et sur les deux brins est hautement improbables. Les différences observées sont levées manuellement en comparant les deux brins et en tenant compte de la qualité des électrophorogrammes obtenus pour chaque brin. Enfin, troisièmement, ceci est incompatible avec le haut pourcentage de sites variables dans chaque gène.

Événements de recombinaison

Donc, nos résultats suggèrent fortement l'existence d'événements antérieurs à l'origine des 21 souches de CMV que nous avons étudiées. L'analyse par SplitsTree des séquences concaténées présente un profil réticulé suggestif d'une recombinaison génomique. Ceci est confirmé par l'étude bootscan correspondante, qui est en faveur d'événements de recombinaison inter- et intra-génique. Quand le Phi-test est calculé, une probabilité hautement significative pour une recombinaison est trouvée. Malheureusement, ni le bootscan, ni SplitsTree n'indiquent de localisation précise des sites de recombinaison.

Nos résultats sont fortement influencés par les séquences des gènes UL40 et UL123, à cause de leur variabilité (respectivement 12% et 11,8%) et de la grande longueur du gène UL123. Les deux clusters observés dans l'analyse phylogéniques des séquences concaténées ne sont pas expliqués par la nature des organes transplantés (rein ou poumon), un donneur commun ou une hospitalisation concomitante. Les souches de référence utilisées, incluant Merlin et Toledo qui n'ont pas subi de nombreux passages en culture cellulaire *in vitro*, sont distribuées dans les deux clusters.

Conséquences fonctionnelles du polymorphisme

La présente étude confirme un polymorphisme significatif dans des souches de CMV par séquençage direct à partir de sang ou de liquide de lavage broncho-alvéolaire de transplantés.

Des variations peuvent induire des conséquences fonctionnelles pour les protéines correspondantes ce qui souligne leur importance potentielle dans l'interférence immunitaire durant l'infection clinique.

Dans cette étude, nous avons observé des mutations dans le nonapeptide de pUL40 qui n'avaient pas encore été décrites. Sa variabilité est assez importante si on considère le nombre limité de patients étudiés. Cette variabilité touche de façon préférentielle le huitième acide aminé. Des études ont montré que ces peptides mutés demeuraient fonctionnels (Cerboni *et al.*, 2001; Tomasec *et al.*, 2000). De plus, dans 102Bx0605BAL, le nonapeptide présente la mutation M2T qui est un des deux sites d'ancrage du nonapeptide à HLA-E ; des mutations au niveau de ces deux positions n'avaient pas encore été décrites pour des souches de CMV. Cependant, ce nonapeptide muté correspond au peptide de la séquence leader de HLA-B aussi présenté par HLA-E et qui est correctement exprimé à la surface cellulaire mais qui n'est pas reconnu par les lymphocytes T NK cytotoxiques spécifiques de HLA-E. A notre connaissance, il n'existe pas de logiciel disponible actuellement pour évaluer la capacité théorique de liaison des épitopes UL40 à HLA-E.

L'équipe CMV du laboratoire de virologie EA2968 a d'ailleurs identifié d'autres mutations dans ce nonapeptide chez des patients greffés rénaux. Enfin, deux cas particuliers sont à souligner :

- pour le patient 101Bx, les deux mutations du nonapeptide observées au temps 1 disparaissent pour faire place à partir du temps 2 à la séquence décrite pour AD169. Cette séquence pourrait donc peut être présenter pour le virus un avantage ;
- le patient 105Bx présente une double population dans le LBA ce qui pourrait apparaître comme un préalable à la sélection d'un des variants viraux, le mieux adapté à l'environnement de l'hôte. Malheureusement, nous ne disposons pas de prélèvements ultérieurs pour conforter cette hypothèse.

Mis à part le nonapeptide, il y a peu de connaissances concernant la structure et les fonctions de la protéine pUL40. Les trois sites potentiels de N-glycosylation ont été identifiés dans le domaine de 144 acides aminés par des logiciels bioinformatiques à partir de la séquence de la souche de laboratoire AD169. En position 89-91, la moitié de nos patients ne présentent pas de séquence signature de site de N-glycosylation, comme les souches de référence Toledo, Merlin et Towne. Ce n'est probablement pas un site de N-glycosylation. Enfin, pour nos souches cliniques, aucune mutation n'a été observée dans la partie transmembranaire putative de pUL40 identifiée chez AD169.

Une variabilité nucléique importante a été observée pour le gène UL123 codant IE1, une protéine immunodominante cible des réponses T-lymphocytaires durant la réactivation du CMV et durant sa latence. Des mutations ont été observées dans treize des quatorze épitopes étudiés pour de nombreuses souches cliniques. Cependant, ces variations ne semblent pas avoir de conséquences fonctionnelles car la capacité théorique de liaison de ces épitopes aux molécules CMH du patient, évaluée avec les logiciels BIMAS et SYFPHEITHI, demeure inchangée. Toutefois, une étude *in silico* ne remplace pas une étude immunologique en culture cellulaire de la capacité de reconnaissance des peptides mutés.

L'étude de l'ensemble du gène UL83 codant pp65 n'a pas été possible. Nous n'avons jamais pu amplifier le gène en entier. Ceci peut être dû à un problème d'amorce en 3' du gène mais nous avons testé d'autres amorces qui n'ont pas non plus fonctionné. Cela pourrait être également dû à la présence d'une structure secondaire compliquée mais la première étape de notre PCR est une dénaturation à 95°C qui doit dérouler l'ADN. Le faible nombre de séquences disponibles dans Genbank en 3' du gène UL83 ne nous permet pas d'avoir d'explication satisfaisante pour comprendre et résoudre cet échec. Néanmoins, nous avons la séquence partielle de pp65 pour 1125 pb sur 1769. Quand nous traduisons les acides nucléiques en acides aminés et que nous les comparons avec la séquence protéique pp65 de la souche AD169, nous n'observons pas de mutation dans les régions identifiées comme étant les cibles des réponses immunitaires.

La structure de la protéine pUL18 a été beaucoup étudiée à cause de son homologie avec les molécules du CMH de classe I. Pour nos isolats, nous observons une stricte conservation des deux motifs localisés dans la queue cytoplasmique de pUL18 impliqués dans son expression à la surface et des quatre cystéines nécessaire à sa liaison avec la b2-microglobuline. L'absence de variabilité des ces motifs dans la littérature et chez nos patients infectés est peut être un argument pour penser qu'ils doivent être conservés pour permettre l'expression de la protéine à la surface. 14 isolats présentent les mutations K42R et A43T et 13 isolats la mutation Q202H qui peuvent moduler l'interaction de la protéine avec son récepteur LIR-1 (Leukocyte Inhibitor Receptor-1) (Cerboni *et al.*, 2006; Wagner *et al.*, 2007). La liaison entre pUL18 et LIR-1 dépend essentiellement de résidus spécifiques situés dans les domaines extracellulaires $\alpha 1$ et $\alpha 3$ de pUL18 : K42/A43 et Q202. L'effet de ces mutations sur l'affinité entre pUL18 et son récepteur est encore débattu. Pour Cerboni *et al.*, 2006, elles augmenteraient la capacité de liaison à LIR-1, tandis que pour Wagner *et al.*, 2007, elles réduiraient de manière très significative la liaison à LIR-1. Dans le domaine $\alpha 3$ de pUL18, pour les souches isolées chez les patients 102Bx et 103Bx, la mutation Q202H a été décrite.

En accord avec Cerboni *et al*, 2006, H202 remplirait mieux l'espace entre pUL18 et LIR-1 et augmenterait leur capacité de liaison. Même si les conséquences de ces mutations ne sont pas encore bien éclaircies, on peut supposer qu'elles visent à accroître l'affinité de UL18 pour son récepteur (chez le patient 101Bx par exemple, elles n'apparaissent qu'à partir du deuxième temps).

Infections multiples et compartimentation

La connaissance du statut sérologique CMV initial du patient se révèle capitale lors d'une greffe d'organe solide. Les infections multiples par le CMV sont bien connues (Puchhammer-Stockl *et al.*, 2006). La réinfection par la souche d'un donneur positif a été montrée comme plus fréquente que la réactivation de la souche endogène chez des greffés rénaux (Grundy *et al.*, 1988). Ceci fournit une explication pertinente à la présence probable d'une seule population virale pour les patients D+R+ exception faite de 105Bx. Il serait intéressant de caractériser la souche CMV du donneur mais ceci nécessite l'accès à un liquide biologique ou à un tissu et la mise en œuvre d'une méthode de réactivation *in vitro* du virus.

Le cas du patient 105Bx est intéressant. On peut évoquer différents mécanismes aboutissant à la présence concomitante de plusieurs variants viraux pour chaque gène :

- surinfection (multiplication simultanée de la souche exogène du greffon et de la souche endogène de l'hôte)
- recombinaison, favorisée par l'action du CMV sur la topoisomérase II cellulaire (Benson and Huang, 1990)
- mutations ponctuelles avec persistance de la population virale initiale (peu probable car observable pour tous les gènes).

C'est le seul patient dans notre étude pour lequel on pourrait parler d'une éventuelle compartimentation pouvant mener à terme à la sélection du variant viral le plus adapté à l'environnement (le virus a intérêt à établir une infection persistante tolérable par le patient). Ce patient n'a pas présenté d'autres épisodes d'infection à CMV, il n'a donc pas été possible de suivre l'évolution de ces souches virales différentes dans le temps.

Chez le patient 102Bx, l'impossibilité d'obtenir les séquences directes des 4 gènes pour le troisième temps dans le compartiment sanguin, est intéressante. Comme il est peu probable qu'il existe une variabilité au niveau des zones d'hybridation des amorces pour les 4 gènes, nous avons recherché la présence d'inhibiteurs de PCR. Nous avons tenté en vain des

amplifications des produits de 1^{ère} PCR dilués au 1/5, 1/10, 1/100. La réalisation d'une PCR d'un gène de ménage n'a pas encore été effectuée.

Nous avons choisi les virus du patient 101Bx pour une analyse plus détaillée avec clonage avant le séquençage, en raison du changement de population virale observé entre les temps 1 et 2. En prenant en compte le fait que la deuxième transplantation bi-pulmonaire a eu lieu juste avant le troisième temps étudié, la question du moteur de cette évolution virale se pose alors : adaptation de la souche du greffon à l'hôte (mutation), surinfection du patient à l'occasion d'une nouvelle contamination (recombinaison). Nos résultats ne permettent pas pour l'instant de trancher cette question. La différence importante observée entre les virus des temps 1 et 2 est plutôt en faveur d'une surinfection par une nouvelle souche de CMV. Ce patient a été transfusé à plusieurs reprises entre mai 2004 et mai 2007 ; néanmoins une contamination transfusionnelle est peu probable en raison de la dé-leucocytation systématique des dons. Une surinfection par contact avec une personne de son entourage est plus vraisemblable. La complexité d'une population de variants viraux reflète le nombre de variants et leurs proportions relatives. Parmi les différents clones de *UL40*, le clone 31 (LBA temps 2) est intéressant à analyser. Il représente une population minoritaire du LBA et peut témoigner d'un phénomène de recombinaison. En effet, il cumule l'essentiel des mutations des temps 1 et 2. L'analyse du bootscan et de l'alignement des séquences nucléotidiques de ce clone 31 et des clones dominants des temps 1 et 2 (donnée non montrée) situerait l'événement de recombinaison entre les nucléotides 65 et 274. Cette zone correspond sur la protéine aux acides aminés situés entre le nonapeptide se liant à HLA-E et le premier des 3 sites potentiels de N-glycosylation présents sur *UL40*. Cette région pourrait représenter une zone préférentielle de recombinaison, favorisant une adaptation du virus à son environnement.

Deuxième Partie

Expression de la protéine pUL40

XIV. pUL40

Le CMV a élaboré de nombreuses stratégies d'immunomodulation sous-tendues par 20% de ses gènes. Certains d'entre eux ont été très étudiés à la fois structurellement et fonctionnellement. Pour d'autres cependant, seules quelques données structurales ou des fonctions hypothétiques sont disponibles car les protéines qu'ils codent sont dites non-essentiels *in vitro*.

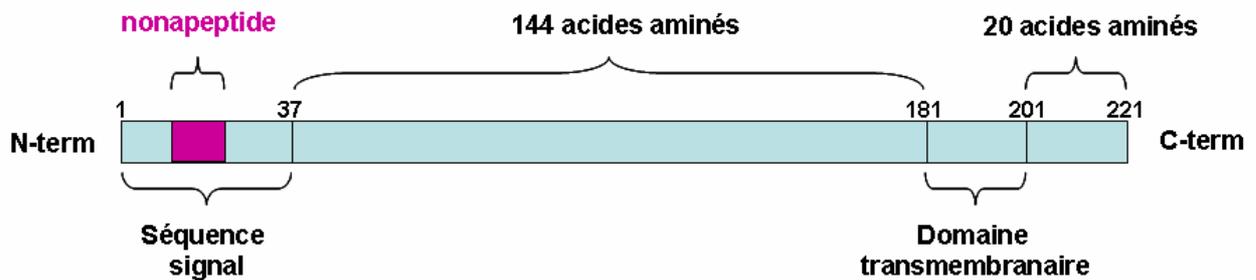
Les cellules NK constituent une première ligne de défense immunitaire efficace contre le CMV et représentent une cible majeure dans les stratégies d'évasion immunitaire. Le CMV code de nombreuses protéines impliquées dans la modulation des fonctions des cellules NK (UL16, UL18, UL40, UL83, UL141 et UL142 et un microARN (miR-UL112)) (Wilkinson *et al.*, 2008). Le gène UL40 contribue à la protection des cellules infectées par le CMV contre la lyse cellulaire médiée par les NK (Carboni *et al.*, 2001; Tomasec *et al.*, 2000; Ulbrecht *et al.*, 2000; Wang *et al.*, 2002).

Physiologiquement, l'expression à la surface des molécules HLA-E dépend du chargement d'un peptide dérivé des séquences leader des molécules classiques du CMH-I ou du HLA-G, ce chargement étant dépendant du système TAP (Transporter associated with Antigen Processing). HLA-E chargé se lie essentiellement au récepteur inhibiteur CD94/NKG2A localisé à la surface des cellules NK et inhibe la lyse cellulaire médiée par les NK (Lee *et al.*, 1998).

Durant l'infection à CMV, la protéine virale pUS6 bloque le transport des peptides liant HLA-E dans le réticulum endoplasmique, entraînant un déficit de molécules HLA-E à la surface cellulaire. Ceci devrait aboutir à la lyse de la cellule infectée. Le nonapeptide dérivé de pUL40 se lie à HLA-E, indépendamment du système TAP, entraînant sa surexpression à la surface et permettant ainsi indirectement le maintien de la réplication virale.

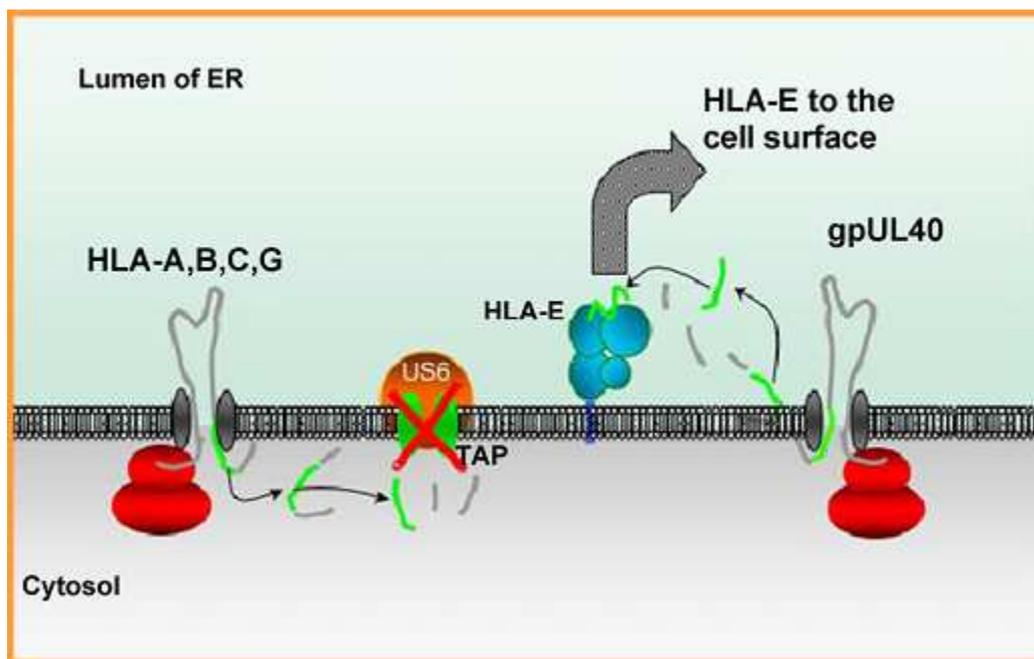
Le cadre ouvert de lecture UL40 code une protéine putative de 221 acides aminés (aa) composée d'un peptide signal de 37 aa en N-terminal, d'un domaine central de 144 aa, d'un domaine transmembranaire putatif de 20 aa et d'une queue de 20 aa en C-terminal (Figure 30). Le peptide signal de pUL40 contient une séquence de 9 aa (VMAPRTLIL) strictement homologue de celle du peptide leader dérivé des molécules du HLA-Cw et liant HLA-E (Braud *et al.*, 1998; Ulbrecht *et al.*, 2000). Trois sites potentiels de N-glycosylation ont été décrits en position 89-91 (NNT), 110-112 (NNT) et 143-145 (NES)(Ulbrecht *et al.*, 2000).

Figure 30 : Représentation schématique de la protéine pUL40



Pour rappel, l'interaction entre HLA-E et le nonapeptide UL40 augmente l'expression à la surface de HLA-E, permettant l'inhibition des cellules NK et de la lyse des cellules infectées (figure 31). Ceci est d'autant plus important que les cellules NK devraient être activées par la diminution de l'expression des molécules MHC-I induite par le CMV.

Figure 31 : Régulation positive de HLA-E par la gpUL40 (Wilkinson *et al.*, 2008)



HLA-E, chargé du nonapeptide, peut aussi se lier à un récepteur activateur, CD94/NKG2C, avec une moindre affinité (Wilkinson *et al.*, 2008).

Les logiciels bioinformatiques disponibles donnent des informations d'interprétation délicate quand la séquence de pUL40 est analysée. Cependant, la protéine semble globulaire avec une hydrophobicité moyenne (ProtParam, ExPASy Tools). Il n'y a d'homologie de séquence avec aucune des protéines disponibles dans Genbank.

La fonction du nonapeptide a déjà été largement décrite (Cerboni *et al.*, 2001; Tomasec *et al.*, 2000; Ulbrecht *et al.*, 2000; Wang *et al.*, 2002). Nous ne connaissons rien des 184 acides aminés restants.

XV. Objectifs

Les objectifs du projet sont de mieux connaître la localisation cellulaire et la fonction de la protéine pUL40.

Notre but est d'exprimer la protéine pUL40 dans le système adéquat pour :

- obtenir des anticorps polyclonaux dirigés contre pUL40 chez le lapin
- faire une analyse biochimique de pUL40 par Western blot
- étudier la localisation intracellulaire de pUL40 dans des cultures cellulaires primaires (fibroblastes, cellules endothéliales) au niveau photonique et électronique afin d'identifier son trajet intracellulaire, de suggérer des fonctions et d'identifier d'éventuels partenaires cellulaires.

A terme, la cristallisation de la protéine permettrait de rechercher une homologie structurale avec une protéine dont la fonction est déjà identifiée.

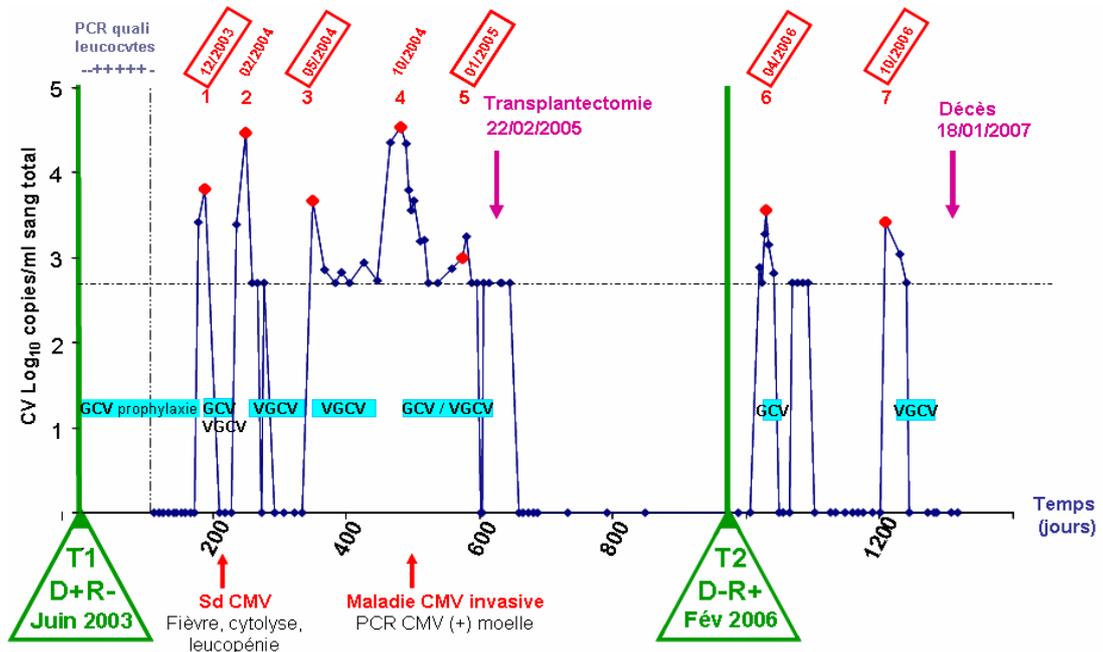
XVI. Matériel

La souche virale choisie pour l'étude de la protéine pUL40 a été isolée chez un patient présentant une histoire virologique compliquée et pour lequel nous disposons d'un suivi sur 42 mois.

Le patient 8Bx: présentation clinique et virologique

Il s'agit d'un patient de 56 ans porteur d'une néphropathie primitive d'étiologie inconnue, responsable d'une insuffisance rénale chronique et nécessitant la mise sous dialyse durant 1 an. Il reçoit une première greffe rénale en juin 2003. Le patient est D+/R-. Pour aider le malade à contenir l'infection à CMV, le traitement immunosuppresseur est arrêté en janvier 2005. La charge virale devient indétectable mais le patient perd son greffon en février 2005 et est remis sous dialyse. Il bénéficie d'une deuxième greffe rénale en février 2006. Il est alors D-/R+. Malheureusement, le patient décède des complications d'une aspergillose le 18 janvier 2007. Son histoire virologique est résumée dans la figure 32 et nous le nommerons 8Bx.

Figure 32 : Schéma récapitulatif de l'histoire virologique du patient 8Bx, après 2 transplantations rénales (T1 et T2).



Séquençage direct + clonage 01/2005

Traitement anti-viral

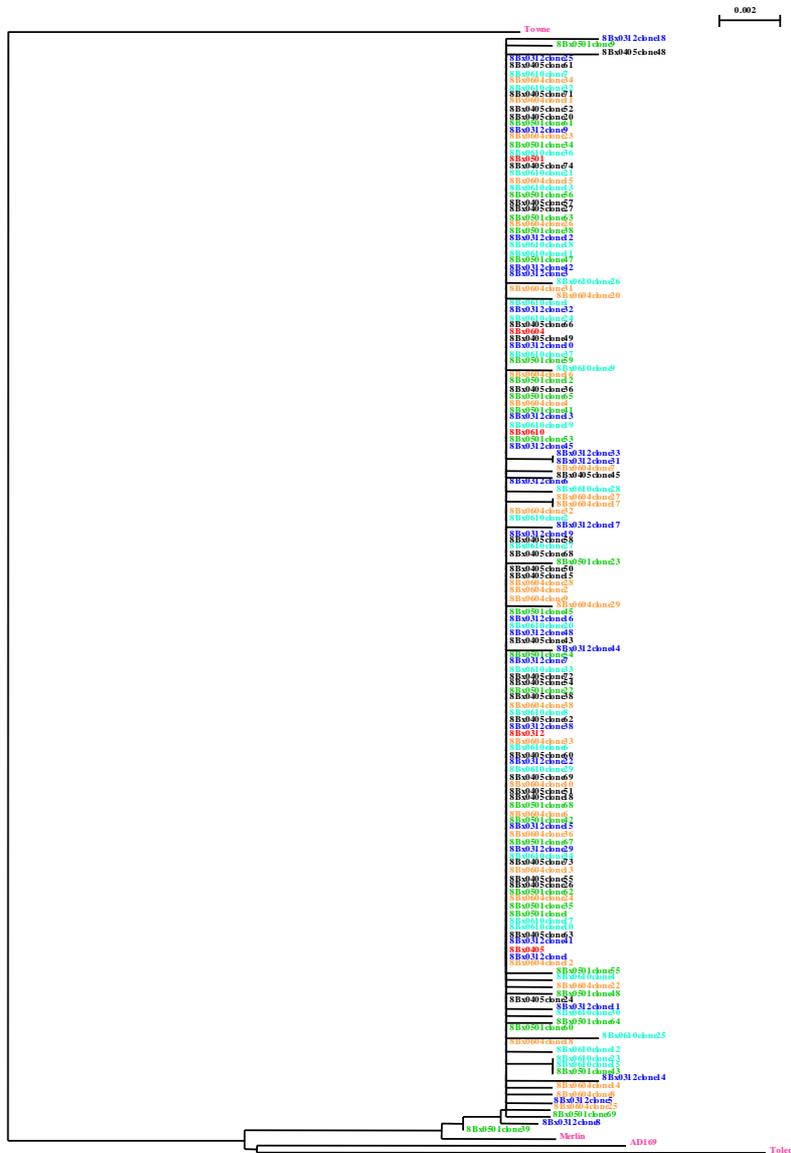
Séquençage direct 10/2004

Nous avons effectué une étude longitudinale de ce patient. Pour cela, nous avons sélectionné sept épisodes infectieux. Pour chacun nous avons amplifié et séquencé le gène UL40 et pour cinq, nous avons également réalisé un clonage avant séquençage afin d'étudier les quasi-espèces présentes chez ce patient (Figure 33). Le tableau 12 indique les prélèvements sélectionnés pour notre étude.

Tableau 12 : Prélèvements du patient 8Bx

Patient	Code patient	UL40	Clonage
8Bx	8Bx0312	Oui	Oui
	8Bx0402	Oui	Non
	8Bx0405	Oui	Oui
	8Bx0410	Oui	Non
	8Bx0501	Oui	Oui
	8Bx0604	Oui	Oui
	8Bx0610	Oui	Oui

Figure 33 : Représentation phylogénétique des variations nucléotidiques du gène viral UL40 chez le patient 8Bx.



Séquences bleu foncées : 29 clones pour décembre 2003

Séquences noires : 30 clones pour mai 2004

Séquences vertes : 29 clones pour janvier 2005

Séquences oranges : 30 clones pour avril 2006

Séquences bleues pâles : 30 clones pour octobre 2006

Séquences rouges : 5 séquences directes à chacun des 5 temps

Séquences roses : séquences de références

Au cours des 42 mois de suivi, la séquence du gène UL40 ne change pas. Ces travaux ont été publiés dans *Transplantation* en 2008 par Isabelle Garrigue, Muriel Faure-Della Corte *et al.* Nous avons choisi cette souche virale pour la stabilité de sa séquence sur une longue période et parce que nous disposions par conséquent d'un matériel abondant. Ces séquences présentent une homologie de 99.4% avec Merlin. Le reste de l'étude a été réalisé avec le prélèvement 8Bx0410.

XVII. Méthodes

A. Amplification

Des amorces spécifiques ont été conçues, selon les recommandations préconisées pour le clonage dans le vecteur pET44 Ek/LIC (Novagen), et utilisées pour l'amplification du gène UL40 selon les conditions définies dans le tableau 13 :

Tableau 13 : Conditions de PCR

		Amorces sens	Amorces anti-sens	Conditions de PCR
pUL40	1 ^{ère} PCR	5'- GACGACGACAAGATGAA CAAATTCAGCAAC-3'	5'- GAGGAGAAGCCCGGTAGC CTTCTTCAAGGC-3'	95°C 12 min 95°C 30 s - 55°C 30 s - 72°C 1 min : 35 cycles 72°C 7 min
	PCR nichée	5'- GACGACGACAAGATGTCT CCTGCCGAGACG-3'	5'- GAGGAGAAGCCCGGTAGC CTTCTTCAAGGC-3'	95°C 12 min 95°C 30 s - 55°C 30 s - 72°C 1 min : 35 cycles 72°C 7 min

La 1^{ère} PCR s'effectue dans un mélange réactionnel comportant: 1,25 U d'ADN polymérase (Ampli Taq Gold, Applied Biosystems), 5 µL de tampon 10X, 2,5 mM de MgCl₂, 200 µM de dNTP et 200 nM de chaque amorce pour un volume final de 50 µL par tube.

Pour la PCR nichée, le mélange réactionnel comporte : 2 U d'ADN polymérase (Ampli Taq Gold, Applied Biosystems), 10 µL de tampon 10X, 2,5 mM de MgCl₂, 200 µM de dNTP et 200 nM de chaque amorce pour un volume final de 100 µL.

Dix µL d'extrait d'ADN sont introduits dans le premier mélange réactionnel et 2 µL du produit de première PCR dans le second mélange réactionnel.

L'amplification des produits PCR est vérifiée sur un gel d'agarose 1,5%.

B. Purification

Les produits de PCR nichée sont purifiés à l'aide de la trousse QIAquick Gel Extraction Kit (QIAGEN) par centrifugation (Centrifugeuse Eppendorf 5417R) selon les recommandations du fournisseur. Les produits d'amplification purifiés sont conservés à une température de -20°C avant séquençage avec les amorces utilisées pour la PCR nichée et selon la méthode décrite dans la première partie.

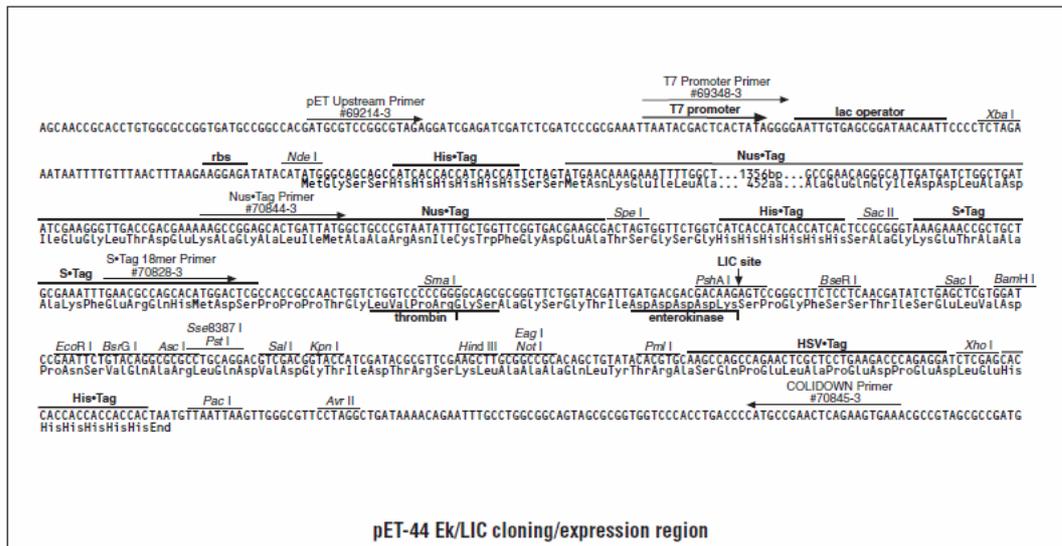
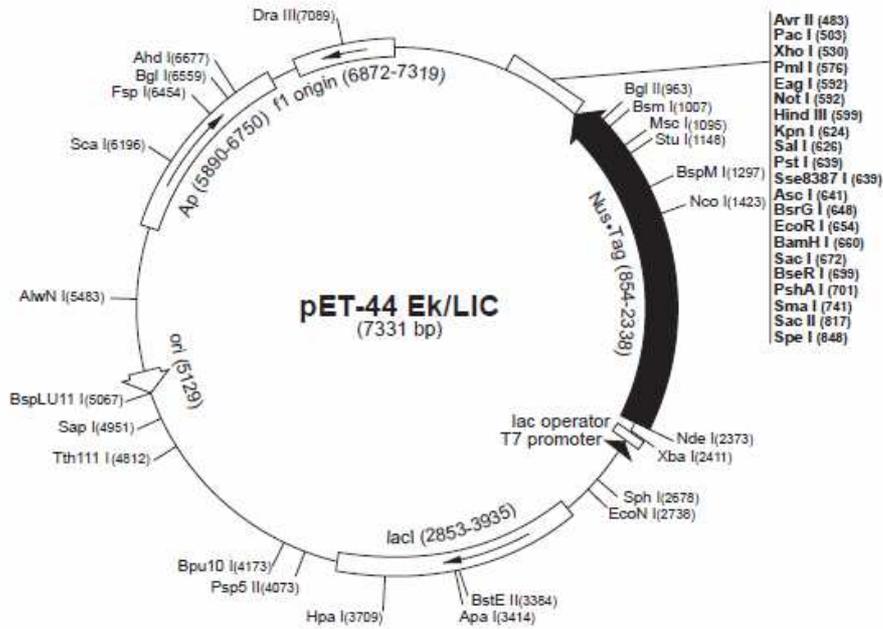
C. Description du système hétérologue

La souche d'*E. coli* utilisée est la souche Rosetta-gami 2 ((DE3)pLysS, Novagen). Plusieurs acides aminés sont codés par plus d'un codon, et chaque organisme possède son propre biais dans l'utilisation des 61 codons disponibles. Dans chaque cellule, la population d'ARNt (ARN de transfert) reflète la disponibilité en codons. Dans une souche classique d'*E. coli*, de nombreux codons sont sur-représentés et d'autres sous-représentés. Quand l'ARNm (ARN messager) d'un système hétérologue est surexprimé chez *E. coli*, les différences dans l'utilisation des codons peuvent influencer sur la traduction en raison de la demande d'un ou plusieurs ARNt qui peuvent être rares ou absents. Cette insuffisance peut entraîner des erreurs dans la traduction d'un ARNm en protéine comme un décalage du cadre de lecture, une terminaison précoce, une erreur dans l'incorporation des acides aminés. Bien que la présence d'un petit nombre de codons rares n'ait souvent pas une forte influence sur l'expression de la protéine cible, cette dernière peut être inhibée quand un gène code un cluster de plusieurs codons rares chez *E. coli*. L'excès d'utilisation de codons rares par le gène cible a été mis en cause dans des niveaux faibles d'expression protéique (Sorensen *et al*, 1989 ; Zhang *et al*, 1991) aussi bien que dans des protéines tronquées. Les souches Rosetta-gami 2 pallient ce manque grâce à la présence d'un plasmide compatible avec les vecteurs pET et codant les ARNt complémentaires des codons AUA, AGG, AGA, CUA, CCC, GGA et CGG. Ce plasmide porte également les gènes codant le lysosyme T7 et le represseur lac. De plus, le milieu non réducteur de ces souches favorise la formation des ponts disulfures dans le cytoplasme quand la protéine hétérologue est exprimée. Les Rosetta-gami 2 portent également une copie du gène de l'ARN polymérase du bactériophage T7 placée sous la dépendance du promoteur *lacUV5*. L'expression du lysosyme T7 supprime l'expression basale de l'ARN polymérase T7 en absence d'induction. Ces souches sont capables de produire les protéines des gènes cibles clonés dans les vecteurs pET par induction avec l'IPTG.

D. Clonage

Le vecteur d'expression utilisé pour l'expression hétérologue de la protéine pUL40 est le pET-44 Ek/LIC (Novagen) (Figure 34).

Figure 34 : Carte du pET-44 Ek/LIC



Une PCR a été réalisée pour amplifier la partie du gène correspondant aux acides aminés 38 à 221 (sans la séquence signal). Pour la ligation 0,2 pmol d'insert est nécessaire. Avant ligation, le produit PCR ou l'insert contrôle sera traité à la T4 DNA polymérase selon les recommandations du fournisseur (Tableau 14). Ce traitement génère des extrémités

complémentaires aux extrémités situées sur le vecteur. La figure 35 illustre la stratégie employée.

Figure 35 : Stratégie de clonage Ek/LIC

Ek/LIC Strategy

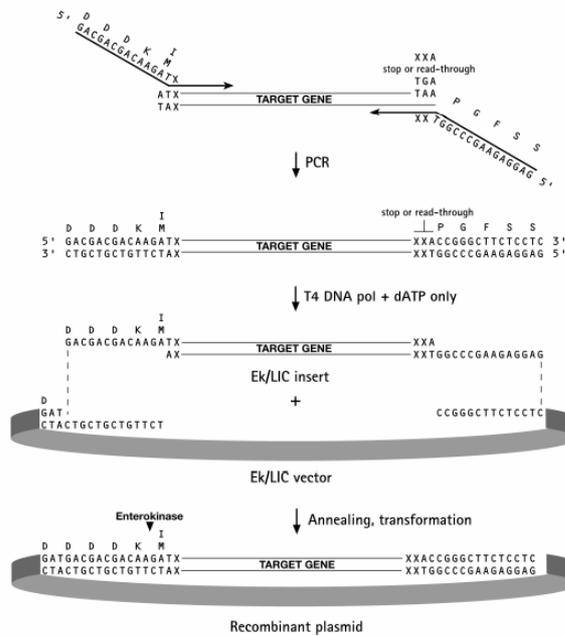


Tableau 14 : Traitement de l'insert à la T4 DNA polymérase

Tampon T4 DNA Polymerase 10X	2 µl	2 µl
dATP 25mM	2 µl	2 µl
DTT 100mM	1 µl	1 µl
T4 DNA Polymerase 2,5U/µl	0,4 µl	0,4 µl
H ₂ O nuclease free qsp 20µl	9,7 µl	10,6 µl
Produit PCR (0,2 pmol)	4,9 µl	-
Insert contrôle (Témoin +, ou T+)	-	4 µl

Après une incubation de 30 minutes à 20 °C, l'enzyme est inactivée par chauffage à 75 °C pendant 20 minutes. L'insert traité est ensuite mis en contact avec le vecteur comme indiqué dans le tableau 15.

Tableau 15 : Ligation

Vecteur	1 µl	1 µl	1 µl
Insert traité	2 µl	2 µl	-
	Produit PCR	T+	T-

Incubation 5 minutes à 22 °C, puis

EDTA	1 µl	1 µl	1 µl
------	------	------	------

Le mélange insert-vecteur est incubé 5 minutes à 22°C.

Vingt µl de cellules compétentes (*E. coli*, Rosetta-gami 2, Novagen) sont transformées avec 1 µl du mélange réactionnel précédent suivant les recommandations du fournisseur. Ces cellules sont étalées sur boîte LB avec de l'ampicilline et incubées la nuit à 37 °C. Seules les bactéries avec un plasmide peuvent pousser en présence de l'antibiotique. Elles sont repiquées dans 2 ml liquide de LB en présence d'ampicilline puis placées sous agitation toute la nuit à 37 °C. Vingt µl de milieuensemencé sont stockés dans 200 µl de milieu LB plus glycérol 33% à -80 °C. Le reste est centrifugé et l'ADN plasmidique est extrait à partir du culot bactérien à l'aide de la trousse NucleoSpin Plasmid (Macherey-Nagel) selon les recommandations du fabricant. L'élution finale s'effectue dans un volume de 50µL. La présence de l'insert est vérifiée par séquençage selon la méthode décrite dans la première partie en utilisant les amorces encadrant le site d'insertion recommandées par le fournisseur (Tableau 16).

Tableau 16 : Séquence des amorces recommandées par le fournisseur

pET Upstream Primer	5'-GATGCGTCCGGCGTAGA-3'
COLIDOWN Primer	5'-TTTCACTTCTGAGTTCGGCATG-3'

E. Expression de la protéine

Les expériences suivantes ont été réalisées en parallèle avec un clone bactérien contenant l'insert correspondant à la protéine que nous souhaitons exprimer et avec un clone possédant un plasmide sans insert et servant de témoin négatif.

Pour induire l'expression d'une protéine, la DO de la culture bactérienne doit se situer entre 0,5 et 0,8. correspondant à la phase exponentielle de croissance.

Cinq litres de milieu LB liquide avec ampicilline sontensemencés. La culture est incubée à 37°C sous agitation jusqu'à ce que la DO de la culture atteigne 0,5. L'expression de la protéine est alors induite avec de l'IPTG pour une concentration finale de 1 mM. La culture est laissée toute la nuit à 16 °C. Le culot bactérien obtenu par centrifugation est ensuite conservé à -80°C jusqu'à l'extraction.

F. Extraction

Le culot bactérien est décongelé dans la glace puis repris dans 20 ml d'un tampon extraction (tampon natif) contenant du Tris pH7,5 50 mM, du NaCl 500 mM, du glycérol 10%, β -mercaptoethanol 1 mM, auxquels sont ajoutés 10 mg de lysozyme pour hydrolyser la paroi bactérienne et un cocktail inhibiteur de protéases (EDTA free). Ce mélange est incubé 30 minutes à 4°C, soumis aux ultrasons quelques secondes pour réduire sa viscosité puis centrifugé 30 minutes à 100000 g.

Le culot C1 et le surnageant S1 sont séparés et conservés. Le surnageant S1 constitue la fraction soluble soumise à la chromatographie d'affinité.

Le culot C1 contient les membranes et autres débris cellulaires et constitue la fraction non soluble. La protéine cible peut se trouver dans cette fraction. Le culot est resuspendu dans 20 ml du tampon natif additionné d'urée à 8 M (tampon dénaturant), incubé 30 minutes à 4°C puis il est soumis à centrifugation 15 minutes à 100000 g. Le surnageant S2 constitue l'extrait protéique qui sera soumis à la chromatographie d'affinité.

G. Purification de la protéine par chromatographie d'affinité

Le surnageant S1 constitue la fraction soluble. Pour concentrer et ne conserver que les protéines d'intérêt, il est soumis à la chromatographie d'affinité. Les ions Ni^{2+} vont chélater les histidines accessibles contenues dans les protéines présentes dans le surnageant. Plus la protéine est riche en histidine, plus la force de liaison avec les ions Ni^{2+} sera importante. La protéine recombinante possède une extrémité de 6 résidus histidine qui permet de la purifier grâce à son affinité pour les ions nickel. Elle sera retenue sur la colonne, les autres dépourvues d'histidine étant éliminées. Le surnageant est chargé sur la colonne puis la colonne est lavée avec un tampon contenant du Tris/HCl pH 7,5 10 mM, du NaCl 500 mM, du Triton X100 0,001% et du β -mercaptoethanol 1 mM. Les protéines sont éluées avec le même tampon contenant 50% d'imidazole. L'imidazole ayant une plus grande affinité pour les ions Ni^{2+} que l'histidine, il va entrer en compétition avec les protéines retenues et entraîner leur élution. En suivant la courbe d'élution des protéines, la fraction contenant notre protéine peut être récupérée.

H. Séparation des protéines par SDS-PAGE

Les protéines sont séparées selon leur masse moléculaire sur un gel discontinu de polyacrylamide en conditions dénaturantes en SDS (Sodium dodécylsulfate). Le gel de polyacrylamide en plaque mince (épaisseur 1,5 mm) est composé d'un gel de séparation contenant 10 à 15 % de polyacrylamide (acrylamide : bisacrylamide 30/08 ; tampon de résolution (Tris/HCl 3 M pH 8,8) ; SDS 10% ; ammonium persulfate 10% ; TEMED) selon la taille de la protéine, au sommet duquel est coulé un gel de concentration de 1 cm contenant 4,5 % d'acrylamide (acrylamide : bisacrylamide 30/08 ; tampon de concentration (Tris/HCl 0,5 M pH 6,8) ; SDS 10% ; ammonium persulfate 10% ; TEMED). Les échantillons protéiques sont déposés après addition d'un volume de tampon de charge (Tris/HCl pH 6,8 50 mM, SDS 2%, glycérol 10%, Bleu de bromophénol, 10% de β -mercaptoethanol extemporanément). Le gel SDS-PAGE est placé dans un tampon d'électrophorèse (Tris 0,25 M ; glycine 1,92 M ; SDS 1%). L'électrophorèse est réalisée à voltage constant (200V) pendant 1 heure 30 minutes.

En fin de migration, les gels sont colorés au bleu de Coomassie ou au nitrate d'argent ou les protéines sont transférées sur une membrane afin de réaliser un Western blot.

1. Coloration des gels protéiques au bleu de Coomassie

Cette coloration permet de mettre en évidence la présence des protéines. Après électrophorèse, le gel de polyacrylamide est incubé au moins une heure et sous agitation dans la solution de coloration contenant de l'eau, du méthanol, de l'acide acétique et du bleu de Coomassie R-250 dans les proportions 5/5/1/0,1%. Il est ensuite décoloré par plusieurs lavages dans un mélange eau/méthanol/acide acétique, 5/5/1. Seules les protéines restent colorées.

A partir de ce gel, la protéine d'intérêt peut être élue puis utilisée pour produire des anticorps. Pour la production d'anticorps chez le lapin, les fabricants recommandent d'avoir au minimum 500 μ g de protéines/ml. Les protéines sont dosées selon la méthode de Bradford.

2. Coloration au nitrate d'argent

La spectrométrie de masse permet de vérifier la séquence d'une protéine. A partir d'un gel SDS-Page, les protéines sont colorées au nitrate d'argent avec la trousse ProteoSilverTM Plus Silver Stain Kit (Sigma-Aldrich) selon les recommandations du fournisseur.

La bande d'intérêt est découpée et envoyée au service de spectrométrie de masse.

3. Immunodétection des protéines séparées sur gel de polyacrylamide : Western blot

La protéine d'intérêt est détectée *via* sa queue histidine par des anticorps anti-tag histidine.

A partir du gel de polyacrylamide, les protéines sont transférées sur une membrane de nitrocellulose par électrophorèse horizontale grâce à un champ électrique. De l'anode vers la cathode, le dispositif est composé de :

- 2 papiers Whatman incubés dans un tampon anode I (Tris Base 0,3 M pH 10,4 ; méthanol),
- 1 papier Whatman incubé dans un tampon anode II (Tris Base 25 mM pH 10,4 ; méthanol),
- la membrane incubée dans le tampon anode II après un passage rapide dans du méthanol puis de l'eau,
- le gel SDS-Page incubé 15 minutes dans le tampon cathode (Tris Base 25 mM pH9,4 ; méthanol, glycine 40 mM),
- 3 papiers Whatman incubés dans le tampon cathode.

Le transfert se fait pendant 30 minutes grâce à un courant de 2,5 mA/cm².

La membrane est ensuite « saturée » toute la nuit à 4°C dans un tampon WB INC 1X (PBS 1X, lait écrémé 5%, BSA 3%, Tween 20 0,05%). Ce tampon permet la diminution des réactions non spécifiques. Il est récupéré et dilué au 1/5^{ème}.

Les étapes suivantes sont réalisées sous agitation à température ambiante. La membrane est incubée une heure dans le tampon dilué en présence d'un anticorps anti-tag histidine primaire, puis elle est lavée cinq fois durant 5 minutes dans du PBS 1X - Tween 0,05%. Elle est ensuite incubée une heure dans le tampon dilué en présence d'un anticorps secondaire couplé à une peroxydase. La membrane est lavée cinq fois durant 5 minutes dans du PBS 1X - Tween 0,05%. Les anticorps sont révélés avec le kit ECL (Pierce) en suivant les recommandations du fournisseur. La membrane est séchée puis enveloppée dans du papier cellophane. La membrane est autoradiographiée avec des temps compris entre une seconde et une minute.

XVIII. Résultats et discussion

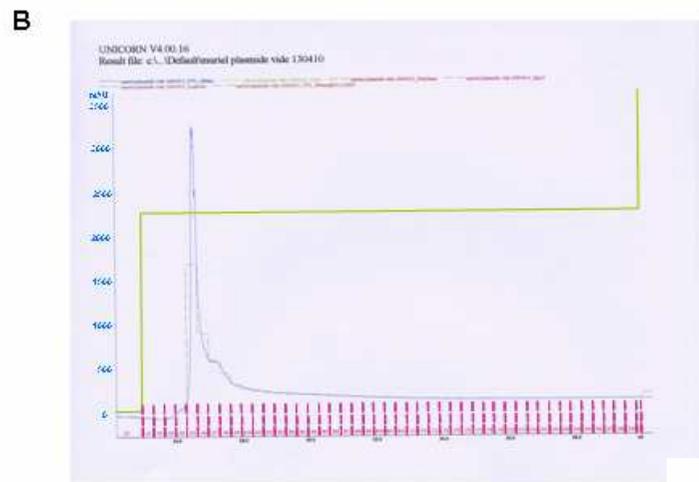
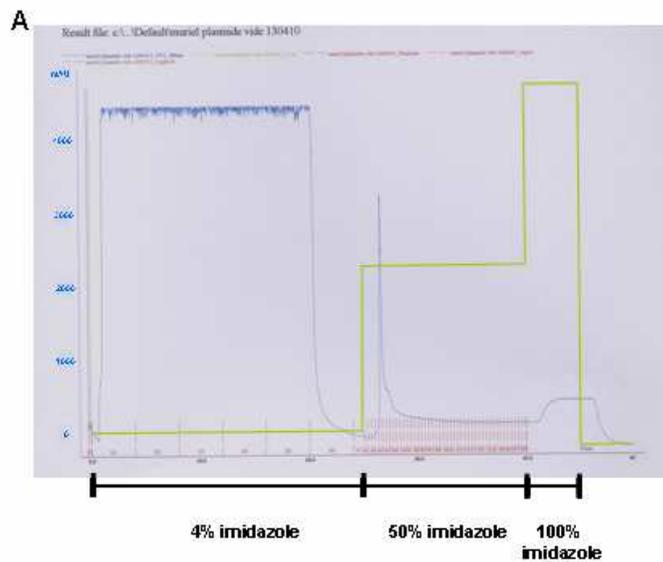
Nos résultats concernant la production de la protéine pUL40 nous ont permis d'aborder de nouvelles techniques, mais demeurent encore préliminaires.

Nous avons obtenu l'expression de pUL40 tronquée des 37 aa de son extrémité N-terminale, notre objectif étant l'obtention d'anticorps dirigés contre des déterminants distincts du peptide signal.

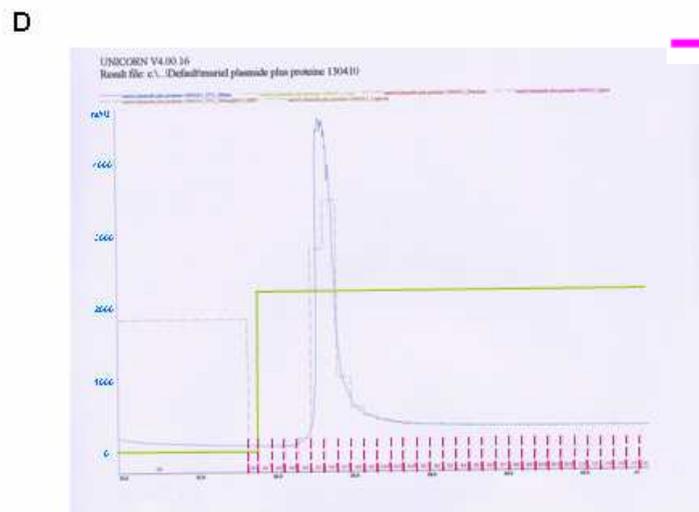
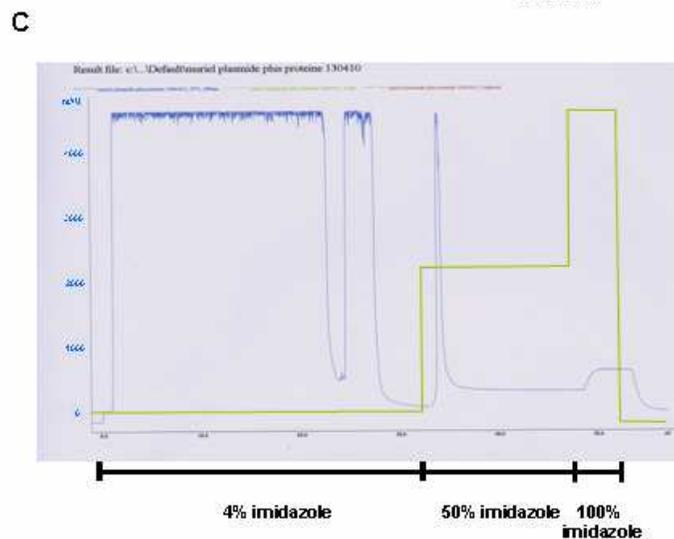
Après extraction des protéines dans un tampon non dénaturant, nous disposons d'une fraction soluble (surnageant S1) et d'un culot C1 obtenu par centrifugation susceptible de contenir des protéines associées aux membranes.

La figure 36 illustre l'élution des protéines contenues dans le surnageant S1. Les profils A et B correspondent au surnageant du plasmide vide et les profils C et D à celui du plasmide avec insert. Les diagrammes A et C montrent l'élution des protéines d'abord dans un tampon contenant 4% d'imidazole. Les protéines dépourvues d'histidine, ou dont l'histidine n'est pas accessible aux ions Ni⁺, sont éliminées. Un tampon contenant 50% d'imidazole permet d'éluer les protéines retenues par les ions nickel (diagrammes B et D). La comparaison du pic d'élution des protéines (profils B et D) montre que celui du plasmide avec insert est à la fois plus large, symétrique et quantitativement plus important, sous-entendant une production plus importante de protéines pour les souches contenant le plasmide avec insert. Les fractions d'élution contenant les protéines ont été récupérées.

Figure 36A à D : Elution des protéines présentes dans le surnageant S1

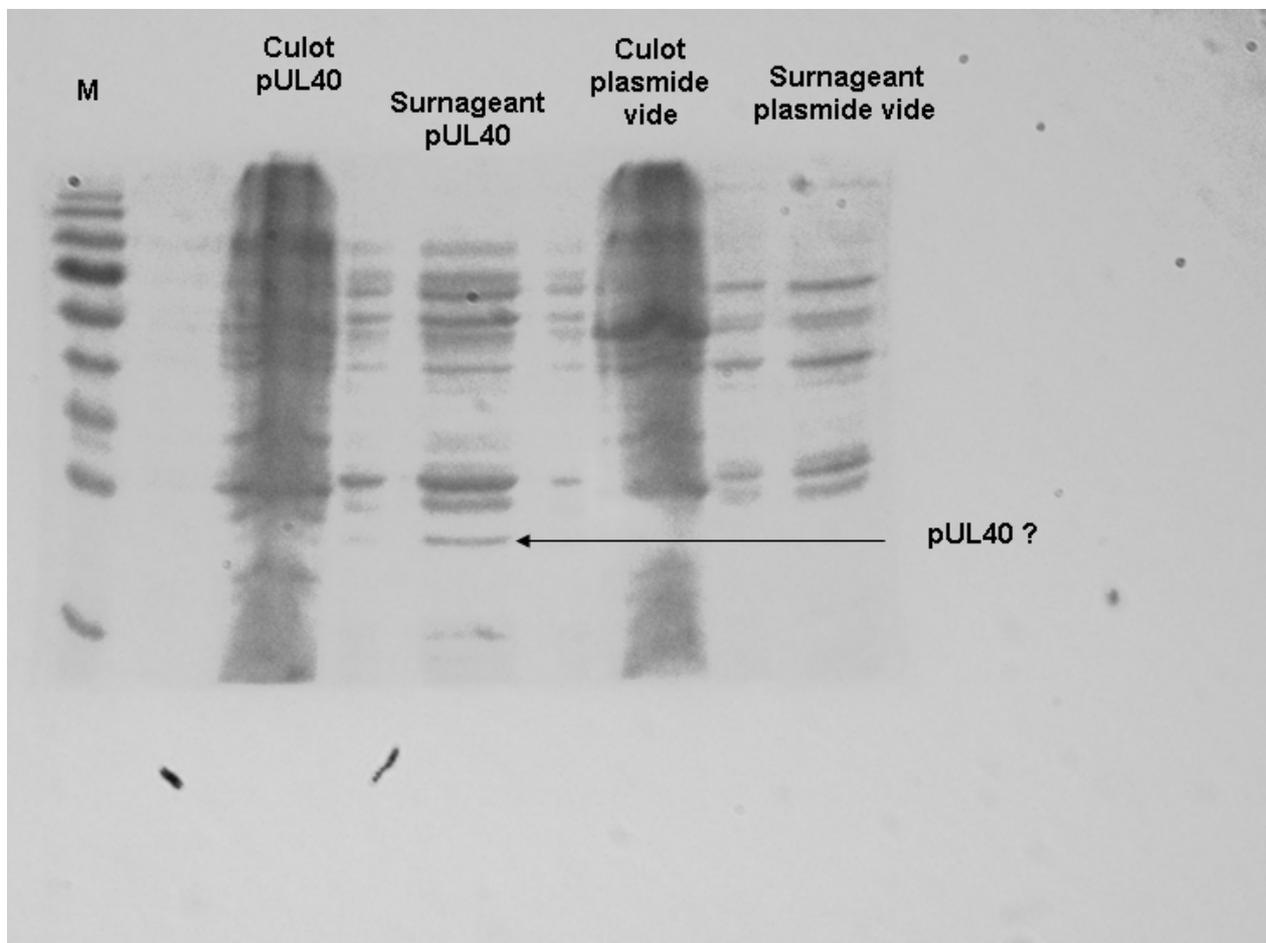


— Concentration d'imidazole
— Profil d'éluion des protéines
— Fractions d'éluion



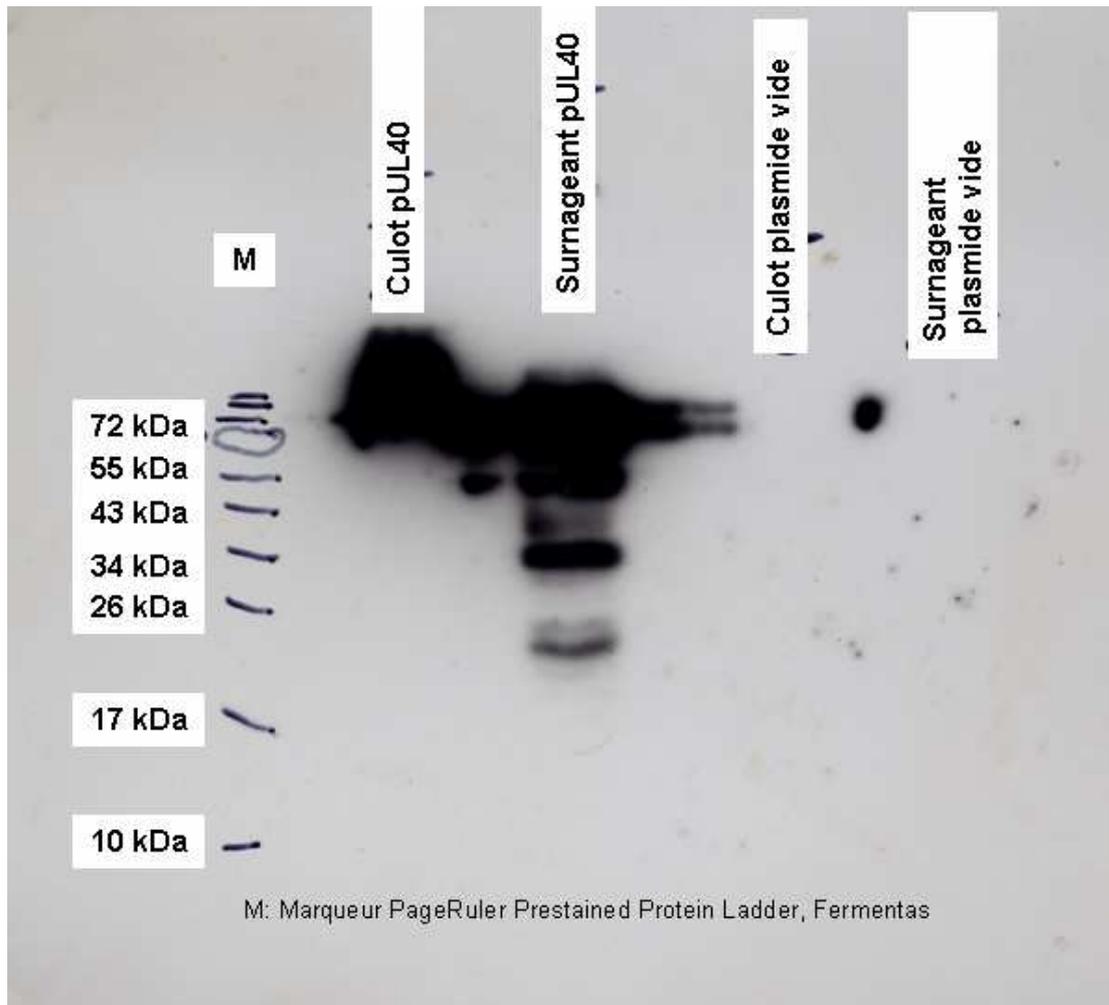
La séparation des protéines est obtenue par électrophorèse en polyacrylamide en condition dénaturante SDS-Page. Le gel, présenté en figure 37, est coloré au bleu de Coomassie. Les colonnes 2 et 3 correspondent aux protéines exprimées dans des bactéries contenant le plasmide avec insert (Culot C1 et Surnageant pUL40 S1) et les colonnes 4 et 5 correspondent aux protéines exprimées dans des bactéries contenant le plasmide sans insert (Culot C1 et Surnageant plasmide vide S1). Une plus grande quantité de protéines est produite dans le surnageant pUL40 et une bande aux alentours de 20kDa est observée. Elle pourrait correspondre à notre protéine d'intérêt.

Figure 37 : Gel de polyacrylamide coloré au bleu de Coomassie



Un Western-blot (Figure 38) est réalisé à l'aide d'anticorps anti-tag histidine, pour vérifier de façon relativement spécifique la présence de la protéine cible. En parallèle, pour affiner ce résultat, une coloration au nitrate d'argent (Figure 40), plus sensible et plus simple pour l'éluion des protéines du gel, est effectuée.

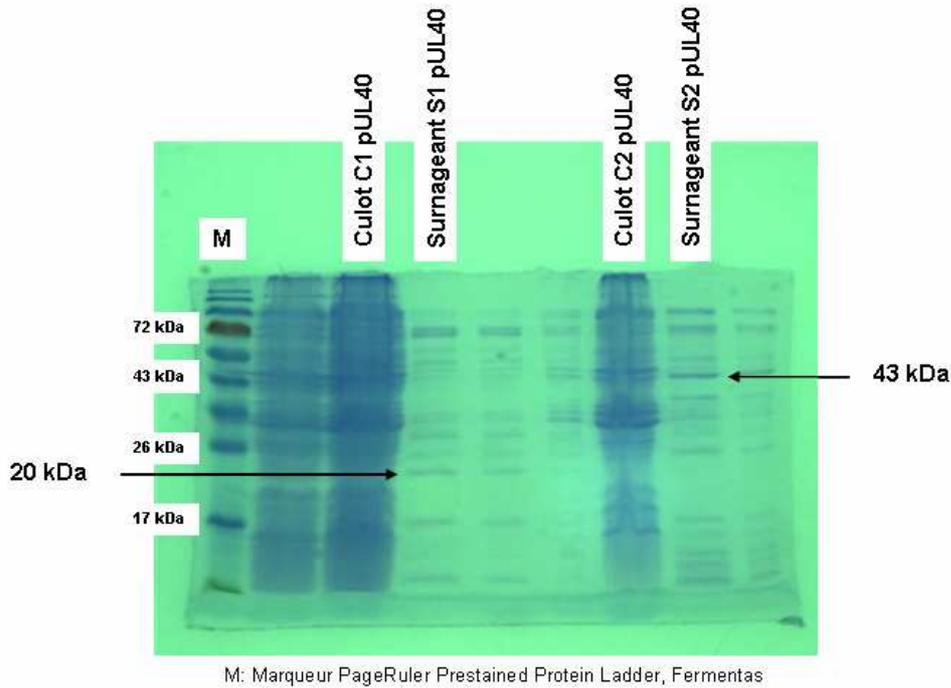
Figure 38 : Western-blot utilisant des anticorps anti-tag histidine.



Aucune protéine n'est marquée dans le culot et le surnageant du plasmide vide. Plusieurs bandes sont révélées dans le surnageant pUL40 y compris celle que nous avons identifiée sur le gel coloré au bleu de Coomassie. Le culot pUL40 est fortement marqué aux anticorps anti-histidine, suggérant que la majorité de notre protéine d'intérêt serait associée à des membranes.

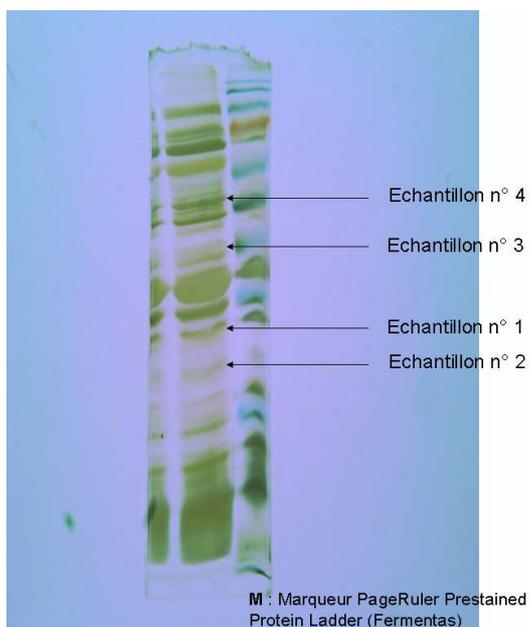
Afin de vérifier cette hypothèse, plausible en raison du domaine transmembranaire putatif identifié par les outils bioinformatiques dans la portion C-terminale de pUL40, nous avons soumis le culot C1 à l'action d'un tampon fortement dénaturant (Urée 8M), avant centrifugation et sélection sur colonne de nickel. Nous espérons ainsi dissocier les protéines (S2) des fractions membranaires (C2). La figure 39 illustre cette expérience. Nous observons effectivement un enrichissement en protéines solubles, notamment à la taille de 43 kDa, mais avec la disparition de la fraction protéique de 20 kDa. Cette dernière fraction pourrait correspondre à des monomères pUL40 solubles.

Figure 39 : Gel de polyacrylamide coloré au bleu de Coomassie



Pour vérifier la présence de la protéine pUL40, nous avons choisi pour la spectrométrie de masse quatre échantillons différents correspondant aux quatre bandes identifiées sur le gel de polyacrylamide coloré au nitrate d'argent (Figure40).

Figure 40 : Gel de polyacrylamide coloré au nitrate d'argent. L'électrophorèse correspond à la fraction S1 décrite précédemment.



La spectrométrie de masse permet de déterminer la séquence des protéines dans les échantillons. Après élution des protéines du gel, les échantillons sont soumis à l'action de la trypsine. Cette dernière clive la protéine chaque fois qu'elle rencontre un résidu arginine (R) ou lysine (K). Les peptides obtenus sont introduits dans le spectromètre de masse, chargés d'un H⁺ puis chaque liaison peptidique est clivée. Chaque peptide puis chaque acide aminé sont identifiés par leur masse. Les séquences obtenues sont envoyées dans des banques de données peptidiques où elles seront identifiées. La protéine pUL40 que nous attendions a un poids théorique de 20 kDa (échantillon n°1). De façon surprenante, elle a été identifiée dans l'échantillon n°4, d'un poids d'environ 43 kDa. Bien que nous soyons en conditions dénaturantes, nous ne pouvons exclure l'hypothèse d'un dimère pUL40-pUL40 ou pUL40 avec une autre protéine. En effet, dans l'échantillon n°4 a été également identifiée une protéine bactérienne (facteur d'élongation) qui pourrait représenter une contamination ou un partenaire cellulaire.

Cependant, la quantité de protéine pUL40 est insuffisante pour l'injection au lapin en vue de l'obtention d'anticorps. De façon intéressante, après traitement du culot C1 avec un tampon dénaturant à l'urée, nous réussissons à dissocier des membranes une protéine de 43 kDa (Figure 39) qui pourrait, elle, être en quantité suffisante. Ces premiers résultats nous encouragent donc à poursuivre ce travail.

Conclusion

Conclusion Générale

Notre travail de séquençage et analyse phylogénétique de six gènes du cytomégalo virus humain, codant quatre protéines immunomodulatrices (UL18, UL40, UL111a et US3) et deux protéines immunodominantes (IE1 et pp65), a permis de confirmer l'existence de nombreuses mutations de polymorphisme, dans des souches de CMV directement séquencées à partir de sang total ou du LBA de patients infectés et immunodéprimés. Cette variabilité est plus importante que l'estimation obtenue par comparaison de séquences de souches cliniques cultivées et de souches de laboratoire (Dolan et al., 2004). Des conséquences fonctionnelles pourraient en découler dans les protéines correspondantes, soulignant leur importance dans l'évasion immunitaire. La recombinaison entre souches a été prouvée in vitro (Haberland, Meyer-Konig, and Hufert, 1999) mais ceci est la première étude, à l'aide de logiciels d'analyse bioinformatiques, de séquences de souches cliniques sans étape de culture cellulaire. La complexité du profil de recombinaison qui interviendrait à travers de nombreux événements intra- et inter-géniques a été mise en lumière par l'absence de sélection in vitro avant analyse, fournissant ainsi de nouvelles données en faveur du rôle clé de la recombinaison comme mécanisme évolutif majeur du CMV et suggérant qu'elle doit représenter une stratégie virale cruciale pour contourner le contrôle immunitaire et établir une infection persistante.

Au cours des derniers mois de ce doctorat, nous avons tenté d'aborder des aspects plus fondamentaux de l'infection par le cytomégalo virus humain, en débutant l'étude de la protéine immunomodulatrice pUL40, la plus variable parmi celles mentionnées ci-dessus, et dont la fonction demeure pour l'essentiel encore inconnue. Nos résultats bien qu'encore préliminaires suggèrent que pUL40 peut être surexprimée en quantité suffisante pour espérer l'obtention d'anticorps polyclonaux par injection au lapin. L'étude du parcours intracellulaire (« trafficking ») de pUL40 est envisageable au sein de la future UMR Microbiologie Fondamentale et Pathogénicité. L'identification de partenaires cellulaires potentiels (structures membranaires, protéines cellulaires) permettrait de mieux appréhender le rôle de cette protéine dans la physiopathologie complexe de l'infection à CMV.

Bibliographie

- Akter, P., Cunningham, C., McSharry, B. P., Dolan, A., Addison, C., Dargan, D. J., Hassan-Walker, A. F., Emery, V. C., Griffiths, P. D., Wilkinson, G. W., and Davison, A. J. (2003). Two novel spliced genes in human cytomegalovirus. *J Gen Virol* **84**(Pt 5), 1117-22.
- Alain, S., and Mazon, M. C. (2001). "Infections à cytomégalo­virus." EMC 8-052-C-10.
- Alain, S., and Mazon, M. C. (2003). *Traité de virologie médicale. eds Estem Chapitre 12*, 195-211.
- Aquino, V. H., and Figueiredo, L. T. (2000). High prevalence of renal transplant recipients infected with more than one cytomegalovirus glycoprotein B genotype. *J Med Virol* **61**(1), 138-42.
- Arav-Boger, R., Foster, C. B., Zong, J. C., and Pass, R. F. (2006). Human Cytomegalovirus-Encoded alpha -Chemokines Exhibit High Sequence Variability in Congenitally Infected Newborns. *J Infect Dis* **193**(6), 788-91.
- Arav-Boger, R., Zong, J. C., and Foster, C. B. (2005). Loss of linkage disequilibrium and accelerated protein divergence in duplicated cytomegalovirus chemokine genes. *Virus Genes* **31**(1), 65-72.
- Baldanti, F., Revello, M. G., Percivalle, E., Labo, N., and Gerna, G. (2003). Genomes of the endothelial cell-tropic variant and the parental Toledo strain of human cytomegalovirus are highly divergent. *J Med Virol* **69**(1), 76-81.
- Benson, J. D., and Huang, E. S. (1990). Human cytomegalovirus induces expression of cellular topoisomerase II. *J Virol* **64**(1), 9-15.
- Borrego, F., Ulbrecht, M., Weiss, E. H., Coligan, J. E., and Brooks, A. G. (1998). Recognition of human histocompatibility leukocyte antigen (HLA)-E complexed with HLA class I signal sequence-derived peptides by CD94/NKG2 confers protection from natural killer cell-mediated lysis. *J Exp Med* **187**(5), 813-8.
- Braud, V. M., Allan, D. S., O'Callaghan, C. A., Soderstrom, K., D'Andrea, A., Ogg, G. S., Lazetic, S., Young, N. T., Bell, J. I., Phillips, J. H., Lanier, L. L., and McMichael, A. J. (1998). HLA-E binds to natural killer cell receptors CD94/NKG2A, B and C. *Nature* **391**(6669), 795-9.
- Braud, V. M., Tomasec, P., and Wilkinson, G. W. (2002). Viral evasion of natural killer cells during human cytomegalovirus infection. *Curr Top Microbiol Immunol* **269**, 117-29.
- Britt, W. J., and Mach, M. (1996). Human cytomegalovirus glycoproteins. *Intervirology* **39**(5-6), 401-12.
- Bruen, T. C., Philippe, H., and Bryant, D. (2006). A simple and robust statistical test for detecting the presence of recombination. *Genetics* **172**(4), 2665-81.
- Burton, D. R., Saphire, E. O., and Parren, P. W. (2001). A model for neutralization of viruses based on antibody coating of the virion surface. *Curr Top Microbiol Immunol* **260**, 109-43.
- Cannon, M. J., Schmid, D. S., and Hyde, T. B. (2010). Review of cytomegalovirus seroprevalence and demographic characteristics associated with infection. *Rev Med Virol* **20**(4), 202-13.
- Cerboni, C., Mousavi-Jazi, M., Wakiguchi, H., Carbone, E., Karre, K., and Soderstrom, K. (2001). Synergistic effect of IFN-gamma and human cytomegalovirus protein UL40 in the HLA-E-dependent protection from NK cell-mediated cytotoxicity. *Eur J Immunol* **31**(10), 2926-35.

- Chang, W. L., Baumgarth, N., Yu, D., and Barry, P. A. (2004). Human cytomegalovirus-encoded interleukin-10 homolog inhibits maturation of dendritic cells and alters their functionality. *J Virol* **78**(16), 8720-31.
- Chee, M. S., Bankier, A. T., Beck, S., Bohni, R., Brown, C. M., Cerny, R., Horsnell, T., Hutchison, C. A., 3rd, Kouzarides, T., Martignetti, J. A., and et al. (1990). Analysis of the protein-coding content of the sequence of human cytomegalovirus strain AD169. *Curr Top Microbiol Immunol* **154**, 125-69.
- Chou, S. W. (1989). Reactivation and recombination of multiple cytomegalovirus strains from individual organ donors. *J Infect Dis* **160**(1), 11-5.
- Compton, T. (2004). Receptors and immune sensors: the complex entry path of human cytomegalovirus. *Trends Cell Biol* **14**(1), 5-8.
- Crough, T., and Khanna, R. (2009). Immunobiology of human cytomegalovirus: from bench to bedside. *Clin Microbiol Rev* **22**(1), 76-98, Table of Contents.
- Davison, A. J., Dolan, A., Akter, P., Addison, C., Dargan, D. J., Alcendor, D. J., McGeoch, D. J., and Hayward, G. S. (2003). The human cytomegalovirus genome revisited: comparison with the chimpanzee cytomegalovirus genome. *J Gen Virol* **84**(Pt 1), 17-28.
- Dechanet, J., Merville, P., Lim, A., Retiere, C., Pitard, V., Lafarge, X., Michelson, S., Meric, C., Hallet, M. M., Kourilsky, P., Potaux, L., Bonneville, M., and Moreau, J. F. (1999). Implication of gammadelta T cells in the human immune response to cytomegalovirus. *J Clin Invest* **103**(10), 1437-49.
- Dolan, A., Cunningham, C., Hector, R. D., Hassan-Walker, A. F., Lee, L., Addison, C., Dargan, D. J., McGeoch, D. J., Gatherer, D., Emery, V. C., Griffiths, P. D., Sinzger, C., McSharry, B. P., Wilkinson, G. W., and Davison, A. J. (2004). Genetic content of wild-type human cytomegalovirus. *J Gen Virol* **85**(Pt 5), 1301-12.
- Domingo, E., Sabo, D., Taniguchi, T., and Weissmann, C. (1978). Nucleotide sequence heterogeneity of an RNA phage population. *Cell* **13**(4), 735-44.
- Ducancelle, A., Belloc, S., Alain, S., Scieux, C., Malphettes, M., Petit, F., Brouet, J. C., Sanson Le Pors, M. J., and Mazon, M. C. (2004). Comparison of sequential cytomegalovirus isolates in a patient with lymphoma and failing antiviral therapy. *J Clin Virol* **29**(4), 241-7.
- Dunn, W., Chou, C., Li, H., Hai, R., Patterson, D., Stolc, V., Zhu, H., and Liu, F. (2003). Functional profiling of a human cytomegalovirus genome. *Proc Natl Acad Sci U S A* **100**(24), 14223-8.
- Evers, D. L., Wang, X., and Huang, E. S. (2004). Cellular stress and signal transduction responses to human cytomegalovirus infection. *Microbes Infect* **6**(12), 1084-93.
- Farber, S., and Wolbach, S. B. (1932). Intranuclear and Cytoplasmic Inclusions ("Protozoan-Like Bodies") in the Salivary Glands and Other Organs of Infants. *Am J Pathol* **8**(2), 123-136 3.
- Gandhi, M. K., and Khanna, R. (2004). Human cytomegalovirus: clinical aspects, immune regulation, and emerging treatments. *Lancet Infect Dis* **4**(12), 725-38.
- Garrigue, I., Boucher, S., Couzi, L., Caumont, A., Dromer, C., Neau-Cransac, M., Tabrizi, R., Schrive, M. H., Fleury, H., and Lafon, M. E. (2006). Whole blood real-time quantitative PCR for cytomegalovirus infection follow-up in transplant recipients. *J Clin Virol* **36**(1), 72-5.
- Garrigue, I., Corte, M. F., Magnin, N., Couzi, L., Capdepon, S., Rio, C., Merville, P., Dechanet-Merville, J., Fleury, H., and Lafon, M. E. (2007a). Variability of UL18, UL40, UL111a and US3 immunomodulatory genes among human cytomegalovirus clinical isolates from renal transplant recipients. *J Clin Virol* **40**(2), 120-8.

- Garrigue, I., Doussau, A., Asselineau, J., Bricout, H., Couzi, L., Rio, C., Merville, P., Fleury, H., Lafon, M. E., and Thiebaut, R. (2008a). Prediction of cytomegalovirus (CMV) plasma load from evaluation of CMV whole-blood load in samples from renal transplant recipients. *J Clin Microbiol* **46**(2), 493-8.
- Garrigue, I., Faure-Della Corte, M., Magnin, N., Couzi, L., Capdepon, S., Rio, C., Merville, P., Dechanet-Merville, J., Fleury, H., and Lafon, M. E. (2007b). Variability of UL18, UL40, UL111a and US3 immunomodulatory genes among human cytomegalovirus clinical isolates from renal transplant recipients. *J Clin Virol* **40**(2), 120-8.
- Garrigue, I., Faure-Della Corte, M., Magnin, N., Recordon-Pinson, P., Couzi, L., Lebrette, M. E., Schrive, M. H., Roncin, L., Taupin, J. L., Dechanet-Merville, J., Fleury, H., and Lafon, M. E. (2008b). UL40 human cytomegalovirus variability evolution patterns over time in renal transplant recipients. *Transplantation* **86**(6), 826-35.
- Geraghty, D. E., Stockscheider, M., Ishitani, A., and Hansen, J. A. (1992). Polymorphism at the HLA-E locus predates most HLA-A and -B polymorphism. *Hum Immunol* **33**(3), 174-84.
- Gicklhorn, D., Eickmann, M., Meyer, G., Ohlin, M., and Radsak, K. (2003). Differential effects of glycoprotein B epitope-specific antibodies on human cytomegalovirus-induced cell-cell fusion. *J Gen Virol* **84**(Pt 7), 1859-62.
- Gilbert, M. J., Riddell, S. R., Plachter, B., and Greenberg, P. D. (1996). Cytomegalovirus selectively blocks antigen processing and presentation of its immediate-early gene product. *Nature* **383**(6602), 720-2.
- Greijer, A. E., Dekkers, C. A., and Middeldorp, J. M. (2000). Human cytomegalovirus virions differentially incorporate viral and host cell RNA during the assembly process. *J Virol* **74**(19), 9078-82.
- Grundy, J. E., Lui, S. F., Super, M., Berry, N. J., Sweny, P., Fernando, O. N., Moorhead, J., and Griffiths, P. D. (1988). Symptomatic cytomegalovirus infection in seropositive kidney recipients: reinfection with donor virus rather than reactivation of recipient virus. *Lancet* **2**(8603), 132-5.
- Guermontprez, P., Valladeau, J., Zitvogel, L., Thery, C., and Amigorena, S. (2002). Antigen presentation and T cell stimulation by dendritic cells. *Annu Rev Immunol* **20**, 621-67.
- Haberland, M., Meyer-Konig, U., and Hufert, F. T. (1999). Variation within the glycoprotein B gene of human cytomegalovirus is due to homologous recombination. *J Gen Virol* **80** (Pt 6), 1495-500.
- Humar, A., Kumar, D., Gilbert, C., and Boivin, G. (2003). Cytomegalovirus (CMV) glycoprotein B genotypes and response to antiviral therapy, in solid-organ-transplant recipients with CMV disease. *J Infect Dis* **188**(4), 581-4.
- Huson, D. H., and Bryant, D. (2006). Application of phylogenetic networks in evolutionary studies. *Mol Biol Evol* **23**(2), 254-67.
- Imbert-Marcille, B. M. (2001). Histoire naturelle des infections à cytomégalo-virus. *Médecine Thérapeutique* **7**(8), 577-584.
- Isaacson, M. K., Feire, A. L., and Compton, T. (2007). Epidermal growth factor receptor is not required for human cytomegalovirus entry or signaling. *J Virol* **81**(12), 6241-7.
- Jesionek, A., and Kiolemenoglou, B. (1904). Ueber einen Befund von protozoenartigen Gebilden in den Organen eines hereditär-luetischen Foetus. *Muenchner Med Wochenschr* **51**:1905–1907.
- Kalejta, R. F. (2008). Tegument proteins of human cytomegalovirus. *Microbiol Mol Biol Rev* **72**(2), 249-65, table of contents.
- Kern, F., Faulhaber, N., Frommel, C., Khatamzas, E., Prosch, S., Schonemann, C., Kretzschmar, I., Volkmer-Engert, R., Volk, H. D., and Reinke, P. (2000). Analysis of

- CD8 T cell reactivity to cytomegalovirus using protein-spanning pools of overlapping pentadecapeptides. *Eur J Immunol* **30**(6), 1676-82.
- Kondo, E., Akatsuka, Y., Kuzushima, K., Tsujimura, K., Asakura, S., Tajima, K., Kagami, Y., Kodera, Y., Tanimoto, M., Morishima, Y., and Takahashi, T. (2004). Identification of novel CTL epitopes of CMV-pp65 presented by a variety of HLA alleles. *Blood* **103**(2), 630-8.
- Lafarge, X., Merville, P., Cazin, M. C., Berge, F., Potaux, L., Moreau, J. F., and Dechanet-Merville, J. (2001). Cytomegalovirus infection in transplant recipients resolves when circulating gammadelta T lymphocytes expand, suggesting a protective antiviral role. *J Infect Dis* **184**(5), 533-41.
- Le Roy, E., and Davignon, J. L. (2005). Human cytomegalovirus-specific CD4(+) T-cell clones recognize cross-reactive peptides from the immediate early 1 protein. *Viral Immunol* **18**(2), 391-6.
- Lee, N., Goodlett, D. R., Ishitani, A., Marquardt, H., and Geraghty, D. E. (1998). HLA-E surface expression depends on binding of TAP-dependent peptides derived from certain HLA class I signal sequences. *J Immunol* **160**(10), 4951-60.
- Legrand-Abravanel, F., Claudinon, J., Nicot, F., Dubois, M., Chapuy-Regaud, S., Sandres-Saune, K., Pasquier, C., and Izopet, J. (2007). New natural intergenotypic (2/5) recombinant of hepatitis C virus. *J Virol* **81**(8), 4357-62.
- Lin, Y. L., Chang, P. C., Wang, Y., and Li, M. (2008). Identification of novel viral interleukin-10 isoforms of human cytomegalovirus AD169. *Virus Res* **131**(2), 213-23.
- Liu, Z., Winkler, M., and Biegalka, B. (2009). Human cytomegalovirus: host immune modulation by the viral US3 gene. *Int J Biochem Cell Biol* **41**(3), 503-6.
- Lole, K. S., Bollinger, R. C., Paranjape, R. S., Gadkari, D., Kulkarni, S. S., Novak, N. G., Ingersoll, R., Sheppard, H. W., and Ray, S. C. (1999). Full-length human immunodeficiency virus type 1 genomes from subtype C-infected seroconverters in India, with evidence of intersubtype recombination. *J Virol* **73**(1), 152-60.
- Marchini, A., Liu, H., and Zhu, H. (2001). Human cytomegalovirus with IE-2 (UL122) deleted fails to express early lytic genes. *J Virol* **75**(4), 1870-8.
- Matte, C., Lacaille, J., Zijenah, L., Ward, B., and Roger, M. (2000). HLA-G and HLA-E polymorphisms in an indigenous African population. The ZVITAMBO Study Group. *Hum Immunol* **61**(11), 1150-6.
- McCormick, A. L. (2008). Control of apoptosis by human cytomegalovirus. *Curr Top Microbiol Immunol* **325**, 281-95.
- Meier, J. L., and Stinski, M. F. (2006). Major Immediate early enhancer and its gene products. *Cytomegaloviruses, molecular biology and immunology*. **Chapter 8:151-166**.
- Merel, P., Pellegrin, I., Garrigue, I., Caumont, A., Schrive, M. H., Birac, V., Bonot, P., and Fleury, H. (2001). Comparison of capillary electrophoresis sequencing with the new CEQ 2000 DNA Analysis System to conventional gel based systems for HIV drug resistance analysis. *J Virol Methods* **98**(1), 9-16.
- Millo, E., Pietra, G., Armirotti, A., Vacca, P., Mingari, M. C., Moretta, L., and Damonte, G. (2007). Purification and HPLC-MS analysis of a naturally processed HCMV-derived peptide isolated from the HEK-293T/HLA-E+/UL40+ cell transfectants and presented at the cell surface in the context of HLA-E. *J Immunol Methods* **322**(1-2), 128-36.
- Misaghi, S., Sun, Z. Y., Stern, P., Gaudet, R., Wagner, G., and Ploegh, H. (2004). Structural and functional analysis of human cytomegalovirus US3 protein. *J Virol* **78**(1), 413-23.
- Mocarski, E. S., and Courcelle, C. T. (2001). "Cytomegaloviruses and their replication in Fields Virology. Knipe DM and Howley PM, editors-in-chief." Philadelphia, PA, USA:Lippincott Williams & Wilkins ed.

- Mocarski, E. S., Jr. (2002). Immunomodulation by cytomegaloviruses: manipulative strategies beyond evasion. *Trends Microbiol* **10**(7), 332-9.
- Murphy, E., Yu, D., Grimwood, J., Schmutz, J., Dickson, M., Jarvis, M. A., Hahn, G., Nelson, J. A., Myers, R. M., and Shenk, T. E. (2003). Coding potential of laboratory and clinical strains of human cytomegalovirus. *Proc Natl Acad Sci U S A* **100**(25), 14976-81.
- Nattermann, J., Nischalke, H. D., Hofmeister, V., Ahlenstiel, G., Zimmermann, H., Leifeld, L., Weiss, E. H., Sauerbruch, T., and Spengler, U. (2005a). The HLA-A2 restricted T cell epitope HCV core 35-44 stabilizes HLA-E expression and inhibits cytotoxicity mediated by natural killer cells. *Am J Pathol* **166**(2), 443-53.
- Nattermann, J., Nischalke, H. D., Hofmeister, V., Kupfer, B., Ahlenstiel, G., Feldmann, G., Rockstroh, J., Weiss, E. H., Sauerbruch, T., and Spengler, U. (2005b). HIV-1 infection leads to increased HLA-E expression resulting in impaired function of natural killer cells. *Antivir Ther* **10**(1), 95-107.
- Nevels, M., Brune, W., and Shenk, T. (2004). SUMOylation of the human cytomegalovirus 72-kilodalton IE1 protein facilitates expression of the 86-kilodalton IE2 protein and promotes viral replication. *J Virol* **78**(14), 7803-12.
- Nevels, M., Paulus, C., and Shenk, T. (2004). Human cytomegalovirus immediate-early 1 protein facilitates viral replication by antagonizing histone deacetylation. *Proc Natl Acad Sci U S A* **101**(49), 17234-9.
- Nitzsche, A., Paulus, C., and Nevels, M. (2008). Temporal dynamics of cytomegalovirus chromatin assembly in productively infected human cells. *J Virol* **82**(22), 11167-80.
- Norberg, P., Kasubi, M. J., Haarr, L., Bergstrom, T., and Liljeqvist, J. A. (2007). Divergence and recombination of clinical herpes simplex virus type 2 isolates. *J Virol* **81**(23), 13158-67.
- Pass, R. F. (2004). A key role for adolescents in the epidemiology of cytomegalovirus and genital herpes infections. *Clin Infect Dis* **39**(10), 1439-40.
- Paya, C., Humar, A., Dominguez, E., Washburn, K., Blumberg, E., Alexander, B., Freeman, R., Heaton, N., and Pescovitz, M. D. (2004). Efficacy and safety of valganciclovir vs. oral ganciclovir for prevention of cytomegalovirus disease in solid organ transplant recipients. *Am J Transplant* **4**(4), 611-20.
- Pignatelli, S., Dal Monte, P., Rossini, G., and Landini, M. P. (2004). Genetic polymorphisms among human cytomegalovirus (HCMV) wild-type strains. *Rev Med Virol* **14**(6), 383-410.
- Plotkin, S. A., Furukawa, T., Zygraich, N., and Huygelen, C. (1975). Candidate cytomegalovirus strain for human vaccination. *Infect Immun* **12**(3), 521-7.
- Prod'homme, V., Retiere, C., Imbert-Marcille, B. M., Bonneville, M., and Hallet, M. M. (2003a). Modulation of HLA-A*0201-restricted T cell responses by natural polymorphism in the IE1(315-324) epitope of human cytomegalovirus. *J Immunol* **170**(4), 2030-6.
- Prod'homme, V., Retiere, C., Valtcheva, R., Bonneville, M., and Hallet, M. M. (2003b). Cross-reactivity of HLA-B*1801-restricted T-lymphocyte clones with target cells expressing variants of the human cytomegalovirus 72kDa-IE1 protein. *J Virol* **77**(12), 7139-42.
- Provenzano, M., Lim, J. B., Mocellin, S., Monsurro, V., Bettinotti, M., Marincola, F. M., and Stroncek, D. F. (2003). The matrix protein pp65(341-350): a peptide that induces ex vivo stimulation and in vitro expansion of CMV-specific CD8+ T cells in subjects bearing either HLA-A*2402 or A*0101 allele. *Transfusion* **43**(11), 1567-74.

- Puchhammer-Stockl, E., Gorzer, I., Zoufaly, A., Jaksch, P., Bauer, C. C., Klepetko, W., and Popow-Kraupp, T. (2006). Emergence of multiple cytomegalovirus strains in blood and lung of lung transplant recipients. *Transplantation* **81**(2), 187-94.
- Rasmussen, L., Geissler, A., Cowan, C., Chase, A., and Winters, M. (2002). The genes encoding the gCIII complex of human cytomegalovirus exist in highly diverse combinations in clinical isolates. *J Virol* **76**(21), 10841-8.
- Rasmussen, L., Geissler, A., and Winters, M. (2003). Inter- and intragenic variations complicate the molecular epidemiology of human cytomegalovirus. *J Infect Dis* **187**(5), 809-19.
- Reeves, M. B., Davies, A. A., McSharry, B. P., Wilkinson, G. W., and Sinclair, J. H. (2007). Complex I binding by a virally encoded RNA regulates mitochondria-induced cell death. *Science* **316**(5829), 1345-8.
- Retiere, C., Prod'homme, V., Imbert-Marcille, B. M., Bonneville, M., Vie, H., and Hallet, M. M. (2000). Generation of cytomegalovirus-specific human T-lymphocyte clones by using autologous B-lymphoblastoid cells with stable expression of pp65 or IE1 proteins: a tool to study the fine specificity of the antiviral response. *J Virol* **74**(9), 3948-52.
- Ribbert (1904). Ueber protozoenartige Zellen in der Niere eines syphilitischen Neugeborenen und in der Parotis von Kindern. *Zbl All Pathol* **15**:945-948.
- Rice, P., Longden, I., and Bleasby, A. (2000). EMBOS: the European Molecular Biology Open Software Suite. *Trends Genet* **16**(6), 276-7.
- Roby, C., and Gibson, W. (1986). Characterization of phosphoproteins and protein kinase activity of virions, noninfectious enveloped particles, and dense bodies of human cytomegalovirus. *J Virol* **59**(3), 714-27.
- Rowe, W. P., Hartley, J. W., Waterman, S., Turner, H. C., and Huebner, R. J. (1956). Cytopathogenic agent resembling human salivary gland virus recovered from tissue cultures of human adenoids. *Proc Soc Exp Biol Med* **92**(2), 418-24.
- Rowshani, A. T., Bemelman, F. J., van Leeuwen, E. M., van Lier, R. A., and ten Berge, I. J. (2005). Clinical and immunologic aspects of cytomegalovirus infection in solid organ transplant recipients. *Transplantation* **79**(4), 381-6.
- Sanchez, V., Greis, K. D., Sztul, E., and Britt, W. J. (2000). Accumulation of virion tegument and envelope proteins in a stable cytoplasmic compartment during human cytomegalovirus replication: characterization of a potential site of virus assembly. *J Virol* **74**(2), 975-86.
- Schmolke, S., Kern, H. F., Drescher, P., Jahn, G., and Plachter, B. (1995). The dominant phosphoprotein pp65 (UL83) of human cytomegalovirus is dispensable for growth in cell culture. *J Virol* **69**(10), 5959-68.
- Slezak, S. L., Bettinotti, M., Selleri, S., Adams, S., Marincola, F. M., and Stroncek, D. F. (2007). CMV pp65 and IE-1 T cell epitopes recognized by healthy subjects. *J Transl Med* **5**, 17.
- Slobedman, B., Mocarski, E. S., Arvin, A. M., Mellins, E. D., and Abendroth, A. (2002). Latent cytomegalovirus down-regulates major histocompatibility complex class II expression on myeloid progenitors. *Blood* **100**(8), 2867-73.
- Smith, M. G. (1956). Propagation in tissue cultures of a cytopathogenic virus from human salivary gland virus (SGV) disease. *Proc Soc Exp Biol Med* **92**(2), 424-30.
- Staden, R. (1994). Staden: introduction. *Methods Mol Biol* **25**, 9-26.
- Stanton, R., Westmoreland, D., Fox, J. D., Davison, A. J., and Wilkinson, G. W. (2005). Stability of human cytomegalovirus genotypes in persistently infected renal transplant recipients. *J Med Virol* **75**(1), 42-6.

- Strong, R. K., Holmes, M. A., Li, P., Braun, L., Lee, N., and Geraghty, D. E. (2003). HLA-E allelic variants. Correlating differential expression, peptide affinities, crystal structures, and thermal stabilities. *J Biol Chem* **278**(7), 5082-90.
- Sylwester, A. W., Mitchell, B. L., Edgar, J. B., Taormina, C., Pelte, C., Ruchti, F., Sleath, P. R., Grabstein, K. H., Hosken, N. A., Kern, F., Nelson, J. A., and Picker, L. J. (2005). Broadly targeted human cytomegalovirus-specific CD4+ and CD8+ T cells dominate the memory compartments of exposed subjects. *J Exp Med* **202**(5), 673-85.
- Tamouza, R., Rocha, V., Busson, M., Fortier, C., El Sherbini, S. M., Esperou, H., Filion, A., Socie, G., Dulphy, N., Krishnamoorthy, R., Toubert, A., Gluckman, E., and Charron, D. (2005). Association of HLA-E polymorphism with severe bacterial infection and early transplant-related mortality in matched unrelated bone marrow transplantation. *Transplantation* **80**(1), 140-4.
- Tarrago, D., Quereda, C., and Tenorio, A. (2003). Different cytomegalovirus glycoprotein B genotype distribution in serum and cerebrospinal fluid specimens determined by a novel multiplex nested PCR. *J Clin Microbiol* **41**(7), 2872-7.
- Thoma, C., Borst, E., Messerle, M., Rieger, M., Hwang, J. S., and Bogner, E. (2006). Identification of the interaction domain of the small terminase subunit pUL89 with the large subunit pUL56 of human cytomegalovirus. *Biochemistry* **45**(29), 8855-63.
- Tomasec, P., Braud, V. M., Rickards, C., Powell, M. B., McSharry, B. P., Gadola, S., Cerundolo, V., Borysiewicz, L. K., McMichael, A. J., and Wilkinson, G. W. (2000). Surface expression of HLA-E, an inhibitor of natural killer cells, enhanced by human cytomegalovirus gpUL40. *Science* **287**(5455), 1031.
- Trivedi, D., Williams, R. Y., O'Reilly, R. J., and Koehne, G. (2005). Generation of CMV-specific T lymphocytes using protein-spanning pools of pp65-derived overlapping pentadecapeptides for adoptive immunotherapy. *Blood* **105**(7), 2793-801.
- Ulbrecht, M., Martinozzi, S., Grzeschik, M., Hengel, H., Ellwart, J. W., Pla, M., and Weiss, E. H. (2000). Cutting edge: the human cytomegalovirus UL40 gene product contains a ligand for HLA-E and prevents NK cell-mediated lysis. *J Immunol* **164**(10), 5019-22.
- Varnum, S. M., Streblow, D. N., Monroe, M. E., Smith, P., Auberry, K. J., Pasa-Tolic, L., Wang, D., Camp, D. G., 2nd, Rodland, K., Wiley, S., Britt, W., Shenk, T., Smith, R. D., and Nelson, J. A. (2004). Identification of proteins in human cytomegalovirus (HCMV) particles: the HCMV proteome. *J Virol* **78**(20), 10960-6.
- Vauloup-Fellous, C., Picone, O., Cordier, A. G., Parent-du-Chatelet, I., Senat, M. V., Frydman, R., and Grangeot-Keros, L. (2009). Does hygiene counseling have an impact on the rate of CMV primary infection during pregnancy? Results of a 3-year prospective study in a French hospital. *J Clin Virol* **46 Suppl 4**, S49-53.
- Verbraak, F. D., Bruinenberg, M., van den Horn, G. J., Meenken, C., van der Lelij, A., Hoyng, C. B., Kijlstra, A., and Peek, R. (1998). Cytomegalovirus (CMV) strain differences between the eye and blood in AIDS patients with CMV retinitis. *Aids* **12**(7), 713-8.
- Vossen, M. T., Westerhout, E. M., Soderberg-Naucler, C., and Wiertz, E. J. (2002). Viral immune evasion: a masterpiece of evolution. *Immunogenetics* **54**(8), 527-42.
- Wang, E. C., McSharry, B., Retiere, C., Tomasec, P., Williams, S., Borysiewicz, L. K., Braud, V. M., and Wilkinson, G. W. (2002). UL40-mediated NK evasion during productive infection with human cytomegalovirus. *Proc Natl Acad Sci U S A* **99**(11), 7570-5.
- Wang, X., Huang, D. Y., Huang, S. M., and Huang, E. S. (2005). Integrin alphavbeta3 is a coreceptor for human cytomegalovirus. *Nat Med* **11**(5), 515-21.

- Wang, X., Kenyon, W. J., Li, Q., Mullberg, J., and Hutt-Fletcher, L. M. (1998). Epstein-Barr virus uses different complexes of glycoproteins gH and gL to infect B lymphocytes and epithelial cells. *J Virol* **72**(7), 5552-8.
- Weller, T. H., Hanshaw, J. B., and Scott, D. E. (1960). Serologic differentiation of viruses responsible for cytomegalic inclusion disease. *Virology* **12**, 130-2.
- Wilkinson, G. W., Tomasec, P., Stanton, R. J., Armstrong, M., Prod'homme, V., Aichelcer, R., McSharry, B. P., Rickards, C. R., Cochrane, D., Llewellyn-Lacey, S., Wang, E. C., Griffin, C. A., and Davison, A. J. (2007). Modulation of natural killer cells by human cytomegalovirus. *J Clin Virol*.
- Wilkinson, G. W., Tomasec, P., Stanton, R. J., Armstrong, M., Prod'homme, V., Aichelcer, R., McSharry, B. P., Rickards, C. R., Cochrane, D., Llewellyn-Lacey, S., Wang, E. C., Griffin, C. A., and Davison, A. J. (2008). Modulation of natural killer cells by human cytomegalovirus. *J Clin Virol* **41**(3), 206-12.
- Wolinsky, S. M., Korber, B. T., Neumann, A. U., Daniels, M., Kunstman, K. J., Whetsell, A. J., Furtado, M. R., Cao, Y., Ho, D. D., and Safrit, J. T. (1996). Adaptive evolution of human immunodeficiency virus-type 1 during the natural course of infection. *Science* **272**(5261), 537-42.
- Wreghitt, T. G., Teare, E. L., Sule, O., Devi, R., and Rice, P. (2003). Cytomegalovirus infection in immunocompetent patients. *Clin Infect Dis* **37**(12), 1603-6.
- Zhao, J., Sinclair, J., Houghton, J., Bolton, E., Bradley, A., and Lever, A. (2010). Cytomegalovirus beta2.7 RNA transcript protects endothelial cells against apoptosis during ischemia/reperfusion injury. *J Heart Lung Transplant* **29**(3), 342-5.
- Zong, J. C., Arav-Boger, R., Alcendor, D. J., and Hayward, G. S. (2007). Reflections on the interpretation of heterogeneity and strain differences based on very limited PCR sequence data from Kaposi's sarcoma-associated herpesvirus genomes. *J Clin Virol* **40**(1), 1-8.
- Zweyberg Wirgart, B., Brytting, M., Linde, A., Wahren, B., and Grillner, L. (1998). Sequence variation within three important cytomegalovirus gene regions in isolates from four different patient populations. *J Clin Microbiol* **36**(12), 3662-9.

Annexes 1 et 2

Annexe 1 : UL40 human cytomegalovirus variability evolution patterns over time in renal transplant recipients. Garrigue I*, **Faure-Della Corte M***, Magnin N, Recordon-Pinson P, Couzi L, Lebrette ME, Schrive MH, Roncin L, Taupin JL, Déchanet-Merville J, Fleury H, Lafon ME. *Transplantation*. **2008** Sep 27;86(6):826-35

Annexe 2 : Variability and recombination of clinical human cytomegalovirus strains from transplantation recipients. **Faure-Della Corte M**, Samot J, Garrigue I, Magnin N, Reigadas S, Couzi L, Dromer C, Velly JF, Déchanet-Merville J, Fleury HJ, Lafon ME. *J Clin Virol*. **2010** Feb;47(2):161-9

Annexe 1

UL40 HCMV variability evolution patterns over time in renal transplant recipients.

**Isabelle Garrigue^{1,2*}, Muriel Faure-Della Corte^{1*}, Noël Magnin¹,
Patricia Recordon-Pinson², Lionel Couzi³, Marie-Elise Lebrette², Marie-Hélène
Schrive¹, Loïc Roncin^{1,2}, Jean-Luc Taupin⁴, Julie Déchanet-Merville⁴,
Hervé Fleury^{1,2}, Marie-Edith Lafon^{1,2}.**

* I. Garrigue and M. Faure-Della Corte contributed equally to this study.

Correspondent footnote

Isabelle Garrigue, Laboratoire de Virologie EA2968, Université Victor Segalen Bordeaux 2,
146 rue Léo Saignat, 33 076 BORDEAUX Cedex, FRANCE.

Tel 33 5 57 57 13 63. Fax : 33 5 56 99 11 40. Mail isabelle.garrigue@chu-bordeaux.fr

Key words: HCMV UL40, cloning, renal transplantation, recombination, quasispecies.

Running title : HCMV UL40 variability.

Abstract word count : 223

Manuscript word count: 2608

This manuscript contains 3 Tables (Tables 1, 2 and 3)

3 Figures of which 2 color Figures (Figures 2 and 3).

¹Université Bordeaux 2, Laboratoire de Virologie EA2968 and IFR66, F-33 076 Bordeaux, France ;

²Centre Hospitalier Universitaire de Bordeaux, Laboratoire de Virologie, F-33 076 Bordeaux, France ;

³Centre Hospitalier Universitaire de Bordeaux, Unité de Transplantation Rénale, Département de Néphrologie, F-33 076 Bordeaux, France ;

⁴Université Bordeaux 2, Centre National de la Recherche Scientifique and IFR 66, Laboratoire d'Immunologie UMR5164 CIRID, F-33 076 Bordeaux, France.

Conflict of interest

The authors have no commercial or other association that might pose a conflict of interest.

Financial support : no

Meeting

This work was partly presented at the Third European Congress of Virology, Nurnberg, Germany in September 2007. Abstract VPI005.

ABBREVIATIONS

D/R, donor/recipient

GCV, ganciclovir

HCMV, Human cytomegalovirus

HCV, Hepatitis C Virus

HIV, Human Immunodeficiency Virus

NK cells, natural killer cells

VGCV, valganciclovir

ABSTRACT

Background. Human cytomegalovirus (HCMV) is the most common opportunistic pathogen infecting immunocompromised patients after transplantation. Although its immunomodulatory capacities and genomic variability participate in immune system evasion, they are poorly studied in clinical strains without culture amplification. One of HCMV immunomodulatory genes, UL40, confers HCMV infected cells protection from natural killer mediated lysis, through its encoded nonapeptide presented in the context of HLA-E.

Methods. In three renal transplant recipients with different HCMV serostatus, we aimed to evidence the co-evolution of mixtures of HCMV variants over time with sequencing and cloning of HCMV UL40 gene.

Results. Six months after renal transplantation in Patient 1Bx, D+/R+, UL40 phylogenetic and bootscan analysis strongly suggested the emergence of a recombination between two previous viral strains. In Patient 8Bx, initially D+/R-, distribution of variants in five samples over 43 months was notably stable, with no visible emerging variants despite two renal engraftments and extended episodes of active infection. In patient 9Bx, also D+/R-, phylogenetic tree of viral variants revealed in the first sample a minor clone, confirmed by a specific PCR, related to the three subsequent dominant clones.

Conclusions. In three HCMV-infected renal transplant recipients, we have evidenced different viral evolutive polymorphisms including point mutations, recombination and occasionally suggesting the intervention of several HCMV strains or a quasispecies-like distribution. This variability could contribute to viral adaptability in pathogenesis.

INTRODUCTION

Human cytomegalovirus (HCMV) is still responsible for the most frequent opportunistic infection after transplantation. The complex immune response to HCMV infection involves in the first place innate immunity, via macrophages, natural killer (NK) and dendritic cells, and type I interferons. Specific cellular responses depend on T lymphocytes (Sylwester et al., 2005). CD4-positive and CD8-positive T $\alpha\beta$ lymphocytic responses are mostly directed against surface glycoproteins gB and gH (UL55 and UL75), immediate early proteins IE1 and IE2 (UL123, UL122) and the tegument protein pp65 (UL83) (Sylwester et al., 2005). Interestingly, the elaboration of a persisting T $\gamma\delta$ lymphocytic response, involving exclusively T $\gamma\delta$ V $\delta 2^{\text{neg}}$ cells and acting with and between innate and adaptive components of host immunity, is highly correlated with HCMV infection control in immunocompromised patients (Dechanet et al., 1999). Finally, humoral responses directed against tegument (pp65) and envelope (gB and gH) proteins elicit the synthesis of neutralising antibodies.

During the co-evolution with its host, HCMV has built numerous immune evasion and modulation strategies, involving up to 20% of its genes. Some immuno-modulatory proteins alter viral antigen presentation, by interfering with or mimicking the expression of host major histocompatibility complex (MHC or HLA for Human Leukocyte Antigen) proteins (i.e. HCMV US2-, US3-, US6-, US11- and UL18-encoded proteins). Others interfere with NK cell function (such as UL16- or UL40-encoded products) or leukocyte behaviour (such as UL111a-encoded viral IL10). Other viral products modify cell cycle or inhibit apoptosis (Mocarski, 2002; Reeves et al., 2007; Vossen et al., 2002; Wilkinson et al., 2007).

HCMV proteins also interact with HLA-E. While the down-regulation of classical HLA I molecules is able to activate NK cells and T $\gamma\delta$ lymphocytes through NK receptors, an UL40-encoded nonapeptide is presented in the context of HLA-E which is then up-regulated at the cell surface (Tomasec et al., 2000). This leads to an inhibition of NK cells function after the commitment of the inhibitory receptors CD94/NKG2A-B, preventing NK mediated lysis of the HCMV infected cells (Braud et al., 1998; Braud, Tomasec, and Wilkinson, 2002; Tomasec et al., 2000; Ulbrecht et al., 2000; Wang et al., 2002). Little information is available on the corresponding amino acids sequence in clinical strains (Carboni et al., 2001).

Viral genomic variability also participates in HCMV evasion. Most available HCMV sequences arise from laboratory or cultured clinical strains (Dolan et al., 2004; Murphy et al., 2003), which bear gene deletions in dispensable genes, point mutations and underwent homologous recombination events (Dolan et al., 2004; Rasmussen, Geissler, and Winters,

2003). Data from clinical samples are sparse and to our knowledge, there is no UL40 sequence from clinical strains devoid of culture passages in GenBank. In our previous work (Garrigue et al., 2007a), using direct sequencing of 4 viral genes (US3, UL18, UL40 and UL111a) in blood samples of renal transplant patients developing an active HCMV infection, we showed a noticeable viral polymorphism. Moreover, viral evolution over time seemed to be linked to the pregraft donor/recipient (D/R) serological status and to the existence of co-infection. Briefly, each patient harbored a unique strain (i. e. a unique combination of the 4 genes), but a single gene could be demonstrated among various patients, suggesting recombination events. Three evolutive patterns could be distinguished: in D+/R- recipients, either complete gene stability or significant variability; in D+/R+ patients, multiple gene variations, possibly linked to super- or co-infections.

We aimed to pursue investigations on HCMV variability with UL40 cloning in sequential samples of 3 patients, each representative of one of the above-mentioned patterns. The goal of the present work was to evidence the co-evolution of mixtures of HCMV variants, and to refine its analysis by studying recombination events and variants dynamics.

PATIENTS, MATERIALS AND METHODS

Patients (Table 1)

Among 17 patients who developed an HCMV infection, three patients were selected according to the preliminary results of direct sequencing of four HCMV genes (US3, UL18, UL40 and UL111a) in sequential samples (Garrigue et al., 2007b): (i) a D+/R+ patient (patient 1Bx) with significant sequences variability observed over time for the four viral genes, (ii) a D+/R- patient (patient 8Bx) with identical viral sequences all along his prolonged follow-up, and (iii) finally, a D+/R- patient (patient 9Bx), with variable UL40, UL111a and US3 viral sequences.

During HCMV infection, the D+/R+ patient did not show any HCMV-related symptom; after several positive HCMV PCR in whole blood and plasma, he received valganciclovir for 5 weeks between the first and second studied samples and no other antiviral treatment between the second and third samples.

The 2 HCMV seronegative recipients were preventively treated with oral ganciclovir during 3 months post-transplantation. The long HCMV story of patient 8Bx is presented in Figure 1: after a first renal transplantation in June 2003, this patient developed an active HCMV infection, with 2 symptomatic episodes. Viral loads were persistently positive and high from time to time, despite repeated ganciclovir and valganciclovir therapies. UL97 genotyping did not show any ganciclovir-related mutation. The immunosuppressive treatment was finally interrupted, and the patient lost his graft. After a second renal graft in February 2006 (D-/R+ status) a more efficient control of HCMV infection was obtained probably due to both antiviral therapy and partial immune improvement. This patient finally died from another opportunistic infection, a cardiopulmonary aspergillosis.

The third patient, 9Bx, presented one symptomatic episode with fever, leucopenia and cytolysis at the end of November, 2004. He received a single and effective 3 weeks treatment with ganciclovir.

Materials and methods

UL40 PCR, cloning and sequencing

UL40 cloning was performed in 3, 5 and 4 samples from patients 1Bx, 8Bx and 9Bx, respectively. For a single patient, samples were collected at least 6 weeks apart. The date of each sample is indicated in the sample identification number (See Table 1 and Figure 2).

DNA extraction

DNA was extracted from 200 µl of whole blood using the MagNA™ Pure instrument (Roche Molecular Biochemicals) with the MagNA™ Pure LC Total Nucleic Acid Isolation Kit (Roche Diagnostics) and 100µl buffer for elution.

Polymerase chain reactions

Primer pairs (MWG) were designed according to alignments of HCMV sequences available in the Genbank database, to amplify UL40 (666 nucleotides) as described elsewhere (Garrigue et al., 2007b). Moreover, specific primers to 9Bx0411clone19, 9Bx0501, 9Bx0502 and 9Bx0504 were secondarily designed to confirm the presence in the original blood samples of UL40 variants identified thanks to cloning.

Standard reaction mixtures were used (AmpliTaq Gold, Applied Biosystems, dNTP, Promega). Oligonucleotides and cycle conditions are indicated in Table 2.

Cloning and sequencing

After DNA purification (S400HR, Amersham Pharmacia Biotech), cloning experiments were performed for UL40 PCR products using a bacterial plasmid (pGEMT Easy Vector, Promega). After transformation, overnight culture of competent bacteria (DH5-α) and plasmid extraction (Plasmid DNA purification, Macherey-Nagel), 30 clones per sample were sequenced using T7 and SP6 primers (MWG) on both strands as previously described (Merel et al., 2001) (using Beckman Coulter Instrument, CEQ 8000 with DTCS Quickstart and CEQ 8000 Investigator), and a consensus sequence was generated. Clustal W1.83 was used for alignment of derived nucleotide sequences with known reference strains: AD169 (access number NC_001347), Merlin (NC_006273), Towne (AY072777) and Toledo (AH013698). Phylogenetic trees were inferred using the neighbor-joining method from matrix distances calculated after gapstripping of alignments, with a Kimura two-parameters algorithm. Quasispecies complexity was calculated using Shannon's normalized entropy (S_n): S_n values range from 0 (a unique variant) to 1 (totally distinct variants) (Wolinsky et al., 1996).

Bootscan method

All recombinant candidates from patient 1Bx samples were tested with the bootscan method included in the Simplot program (Legrand-Abravanel et al., 2007; Lole et al., 1999; Norberg et al., 2007). Briefly, bootscanning relies on the alignment of a suspected recombinant sequence with a set of potential parental reference sequences. After optimal alignment, the alignment is broken into sequential, overlapping segments (or windows) which are fed to a program for phylogenetic analysis. Bootstrapped phylogenetic trees are built for each segment and finally, the bootstrap value for placing the unknown with each of the reference sequences

is tabulated and plotted along the genome. The assumption is that the unknown will always reach high bootstrap values with its parental subtypes in windows.

RESULTS

Sequences from cloning were deposited in GenBank with accession numbers DQ367999-DQ368001, EF488288-EF488363, EF523615-EF523629 for patient 1Bx, DQ368010-DQ368014, EU048877-EU049026 for patient 8Bx and DQ368015, DQ368017, EU049027- EU049168 for patient 9Bx.

Phylogenetic trees of HCMV nucleotidic variation in UL40 for the 3 patients are presented in Figure 2. Altogether, 182 polymorphic positions were observed in the 666 bp fragment. Viral strains infecting Patient 1Bx, D+/R+, (Figure 2A) showed an evolutive distribution within less than 5 months. In the limit of the 91 analyzed clones, the tree strongly suggested the emergence of the third group of variants from the first two viral populations. The proportion of the dominant clone, lower in January 2005 (10/29) than in the first two samples (16/30 and 21/32, respectively) is evidenced with the variation of S_n (shown in the Figure); third direct sequencing resulted in a mixed population, leading to an isolated direct sequence in the tree. Moreover, bootscan analysis of UL40 gene was in favor of a recombination of the first two direct sequences for 1Bx0501 (Figure 3A) as well as for clonal sequences (Figure 3B, 1Bx0501clone 57, surrounded in Figure 2A). On the other hand, 1Bx0501clone49 and 1Bx0501clone52 seemed to be close to 1Bx0408 and 1Bx0411, respectively (Figures 3C and 3D). Unfortunately, there was no available sample to follow this evolution.

Distribution of viral variants in 5 consecutive samples of Patient 8Bx, D+/R- at the time of his first transplantation, was notably stable with no visible emerging variants despite a long follow-up, 2 renal engraftments and a total of 148 analyzed clones (Figure 2B).

Phylogenetic tree of viral variants in patient 9Bx who was D+/R-, presented differently (Figure 2C). Among the variants identified in the first sample, a single clone (9Bx0411clone19, surrounded in blue) was closer to the clones from the 3 later samples. More clones from January 2005 were analyzed (data not shown) and they were all close to the dominant clone of this date. To invalidate any doubt about this clone 19, we performed a PCR with specific primers designed after alignment of 9Bx0411clone19, 9Bx0501, 9Bx0502 and 9Bx0504. This PCR and subsequent sequencing (data not shown) confirmed the presence of this minor variant in the original blood sample.

We studied the variability of the UL40-encoded nonapeptide presented by HLA-E, which could let suspect differences in the interaction between infected and NK cells, depending on the viral variant. As shown in Table 3, a few mutations were indeed observed in 2 of the 3 patients.

DISCUSSION

CMV has evolved multiple functions to bypass and hijack the immune system, thanks to viral proteins (Mocarski, 2002), RNA (Reeves et al., 2007) and genomic variability. Unlike others (Arav-Boger et al., 2006; Cerboni et al., 2001; Norberg et al., 2007; Rasmussen, Geissler, and Winters, 2003), we studied genomic viral polymorphism from clinical samples without any culture step to avoid *in vitro* strains selection and evolution, detectable as early as after the first passage (Akter et al., 2003). Moreover, since PCR only allows the detection of major variants, we have also looked for this polymorphism after cloning of PCR products of HCMV UL40, the most variable gene in our previous work (up to 10% of divergence at the amino acidic level) (Garrigue et al., 2007b).

UL40 may play an important role in the escape from host defenses, thanks to the presentation of its nonapeptide, leading to the up-regulation of HLA-E, a mechanism also shared by HIV and HCV (Nattermann et al., 2005a; Nattermann et al., 2005b). The two known HLA-E alleles differ by a single amino acid substitution in the $\alpha 2$ heavy chain domain (Geraghty et al., 1992; Matte et al., 2000) and by cell surface expression leading to a possible functional difference (Strong et al., 2003; Tamouza et al., 2005). We compared the nonapeptide polymorphism in our strains with the literature (Table 3). Cerboni et al (Cerboni et al., 2001) described substitutions in clinical isolates with no effect on HLA-E regulation. These were also found in Toledo (Tomasec et al., 2000) and in 34 clones of patient 1Bx. Interestingly, we observed no mutation at position 2 (Met), suggested to be functionally crucial (Borrego et al., 1998). In patients 1Bx and 8Bx clones, we found 3 new substitutions (positions 1, 4 and 6) which potential effects deserve further consideration. Methods are being developed to identify peptides presented in the context of the non classical HLA class I molecule HLA-E and investigate their functionality (Braud et al., 1998; Millo et al., 2007). To our knowledge, there is no available software to evaluate the theoretical binding capacity of the UL40 epitope to HLA-E. However, besides the nonapeptide variability, other factors interfere during the course of HCMV infection in order to prevent NK attack of infected cells (Cerboni et al., 2001).

HCMV DNA variability has been described, in particular in laboratory strains (Baldanti et al., 2003; Dolan et al., 2004). In transplanted patients, different variants may be observed because of the possible reactivation of recipient and/or donor strains (Aquino and Figueiredo, 2000; Arav-Boger et al., 2006; Ducancelle et al., 2004; Humar et al., 2003; Stanton et al., 2005); reinfection with donor virus was shown to be more frequent than

reactivation of endogenous strain in symptomatic kidney recipients (Grundy et al., 1988). Unfortunately, the variant(s) infecting donor and recipient before transplantation are not available for characterization.

Homologous recombination participates in the evolution of *Herpesviridae* and was shown both *in vivo* and *in vitro* for HCMV (Arav-Boger, Zong, and Foster, 2005; Baldanti et al., 2003; Chou, 1989; Dolan et al., 2004; Haberland, Meyer-Konig, and Hufert, 1999; Rasmussen, Geissler, and Winters, 2003), favored by HCMV-induced expression of cellular topoisomerase II (Benson and Huang, 1990). Indeed, the 3 UL40 mixtures identified for patient 1Bx, suggested that the third viral population was possibly born from the first two (Fig. 2A). One can suspect that this D+/R+ patient presented successively exogenous and endogenous reactivations (whatever the first), or also a super-infection, associated with a recombination phenomenon. His favorable evolution contrasts with previous data suggesting that the presence of mixed CMV-strain infections may be disadvantageous (Puchhammer-Stockl et al., 2006).

Patient 9Bx, D+/R-, showed the highest CMV loads (data not shown) reaching 287,000 copies/ml, favoring mutations. However, CMV super-infection is possible even though he did not receive any blood transfusion after engraftment. Variability can also emerge from strains evolution as suggested by the similarity between a single early clone and the subsequent variants. The term of “quasispecies” could be suitable in spite of the replication characteristics of a DNA virus. The existence of quasispecies could be another key in the evolution of infection, helping the virus to evade and modify its tropism, or leading to the emergence of antiviral therapy resistance.

Zong *et al.* (Zong et al., 2007) argued that sequencing of cloned PCR products yields a complex mixture of randomly re-assorted variants (chimeras and randomly introduced cloned PCR product errors). This is highly improbable with our standardized experimental conditions and the AmpliTaq Gold enzyme (Applied Biosystems). Firstly, the amplified fragment was 732 pb long and the mean error rate of the enzyme is 5.10^{-6} /incorporated nucleotide/cycle (i. e. an error every 200,000 nt). Secondly, consensus sequences generated from two strands in both directions showed very few discrepancies. Thirdly, the existence of a crucial minor clone (9Bx0411clone19) was ascertained by PCR performed from extracted DNA with specific primers, ruling out the introduction of an artificial variant by the sequencing reaction after cloning. Thus, we consider that we observed a true polymorphism, independent of PCR artefacts.

In patient 8Bx (D+/R-), despite a 43 months-long follow-up, 2 renal engraftments and extended episodes of active infection, UL40 variability was very low (Fig. 2B). This patient died from a fungal opportunistic infection. Interestingly, at the time of the engraftment, he was presenting a T lymphopenia (including CD4+ and CD8+ lymphocytes), which persisted in spite of a non depleting induction therapy. The anti-HCMV T $\gamma\delta$ V δ 2^{neg} cells population was stable and did not show any progression (data not shown). He seemed to display ineffective immune responses, possibly confronted to a HCMV strain more easily adapted to the host environment. On the opposite, patient 9Bx, D+/R- with higher viral variability and no preliminary lymphopenia, elaborated a significant response, with T $\gamma\delta$ V δ 2^{neg} cells expanding from 0.52% of T $\gamma\delta$ cells to 25.7%. He succeeded in controlling HCMV infection, in accordance with previous observations (Lafarge et al., 2001). Viral strain adaptation along with host environment represents a large field to investigate.

In this study, we have evidenced different viral evolutive patterns including point mutations, recombination and occasionally suggesting the intervention of several CMV strains or a quasispecies-like distribution, allowing adaptability in pathogenesis. Of course, no relationship can be established at that point between UL40 sequences and CMV infection outcome: beyond the study of mutations functional consequences, several other factors complicate results interpretation, such as immunosuppressive regimen, the occurrence of graft rejection or other infections.

Acknowledgments

We are indebted to Anne Caumont-Sarcos, Benoît Pinson, Dominique Bégu, Michel Ventura, and Vincent Parissi for their counselling and helpful discussions, all the Virology department technicians for their skilful participation in routine CMV quantification, Sophie Alain for anti-viral resistance testing and Catherine Rio from the nursing staff of the Department of Renal Transplantation.

REFERENCES

1. Sylwester AW, Mitchell BL, Edgar JB, et al. Broadly targeted human cytomegalovirus-specific CD4⁺ and CD8⁺ T cells dominate the memory compartments of exposed subjects. *J Exp Med* 2005; 202 (5): 673.
2. Dechanet J, Merville P, Lim A, et al. Implication of gammadelta T cells in the human immune response to cytomegalovirus. *J Clin Invest* 1999; 103 (10): 1437.
3. Mocarski ES, Jr. Immunomodulation by cytomegaloviruses: manipulative strategies beyond evasion. *Trends Microbiol* 2002; 10 (7): 332.
4. Vossen MT, Westerhout EM, Soderberg-Naucler C, Wiertz EJ. Viral immune evasion: a masterpiece of evolution. *Immunogenetics* 2002; 54 (8): 527.
5. Wilkinson GW, Tomasec P, Stanton RJ, et al. Modulation of natural killer cells by human cytomegalovirus. *J Clin Virol* 2007.
6. Reeves MB, Davies AA, McSharry BP, Wilkinson GW, Sinclair JH. Complex I binding by a virally encoded RNA regulates mitochondria-induced cell death. *Science* 2007; 316 (5829): 1345.
7. Tomasec P, Braud VM, Rickards C, et al. Surface expression of HLA-E, an inhibitor of natural killer cells, enhanced by human cytomegalovirus gpUL40. *Science* 2000; 287 (5455): 1031.
8. Ulbrecht M, Martinozzi S, Grzeschik M, et al. Cutting edge: the human cytomegalovirus UL40 gene product contains a ligand for HLA-E and prevents NK cell-mediated lysis. *J Immunol* 2000; 164 (10): 5019.
9. Braud VM, Allan DS, O'Callaghan CA, et al. HLA-E binds to natural killer cell receptors CD94/NKG2A, B and C. *Nature* 1998; 391 (6669): 795.
10. Wang EC, McSharry B, Retiere C, et al. UL40-mediated NK evasion during productive infection with human cytomegalovirus. *Proc Natl Acad Sci U S A* 2002; 99 (11): 7570.
11. Braud VM, Tomasec P, Wilkinson GW. Viral evasion of natural killer cells during human cytomegalovirus infection. *Curr Top Microbiol Immunol* 2002; 269: 117.
12. Cerboni C, Mousavi-Jazi M, Wakiguchi H, Carbone E, Karre K, Soderstrom K. Synergistic effect of IFN-gamma and human cytomegalovirus protein UL40 in the HLA-E-dependent protection from NK cell-mediated cytotoxicity. *Eur J Immunol* 2001; 31 (10): 2926.
13. Dolan A, Cunningham C, Hector RD, et al. Genetic content of wild-type human cytomegalovirus. *J Gen Virol* 2004; 85 (Pt 5): 1301.
14. Murphy E, Yu D, Grimwood J, et al. Coding potential of laboratory and clinical strains of human cytomegalovirus. *Proc Natl Acad Sci U S A* 2003; 100 (25): 14976.
15. Rasmussen L, Geissler A, Winters M. Inter- and intragenic variations complicate the molecular epidemiology of human cytomegalovirus. *J Infect Dis* 2003; 187 (5): 809.
16. Garrigue I, Corte MF, Magnin N, et al. Variability of UL18, UL40, UL111a and US3 immunomodulatory genes among human cytomegalovirus clinical isolates from renal transplant recipients. *J Clin Virol* 2007; 40 (2): 120.
17. Merel P, Pellegrin I, Garrigue I, et al. Comparison of capillary electrophoresis sequencing with the new CEQ 2000 DNA Analysis System to conventional gel based systems for HIV drug resistance analysis. *J Virol Methods* 2001; 98 (1): 9.
18. Wolinsky SM, Korber BT, Neumann AU, et al. Adaptive evolution of human immunodeficiency virus-type 1 during the natural course of infection. *Science* 1996; 272 (5261): 537.

19. Lole KS, Bollinger RC, Paranjape RS, et al. Full-length human immunodeficiency virus type 1 genomes from subtype C-infected seroconverters in India, with evidence of intersubtype recombination. *J Virol* 1999; 73 (1): 152.
20. Norberg P, Kasubi MJ, Haarr L, Bergstrom T, Liljeqvist JA. Divergence and recombination of clinical herpes simplex virus type 2 isolates. *J Virol* 2007; 81 (23): 13158.
21. Legrand-Abravanel F, Claudinon J, Nicot F, et al. New natural intergenotypic (2/5) recombinant of hepatitis C virus. *J Virol* 2007; 81 (8): 4357.
22. Arav-Boger R, Foster CB, Zong JC, Pass RF. Human Cytomegalovirus-Encoded alpha-Chemokines Exhibit High Sequence Variability in Congenitally Infected Newborns. *J Infect Dis* 2006; 193 (6): 788.
23. Akter P, Cunningham C, McSharry BP, et al. Two novel spliced genes in human cytomegalovirus. *J Gen Virol* 2003; 84 (Pt 5): 1117.
24. Nattermann J, Nischalke HD, Hofmeister V, et al. HIV-1 infection leads to increased HLA-E expression resulting in impaired function of natural killer cells. *Antivir Ther* 2005; 10 (1): 95.
25. Nattermann J, Nischalke HD, Hofmeister V, et al. The HLA-A2 restricted T cell epitope HCV core 35-44 stabilizes HLA-E expression and inhibits cytolysis mediated by natural killer cells. *Am J Pathol* 2005; 166 (2): 443.
26. Geraghty DE, Stockschleider M, Ishitani A, Hansen JA. Polymorphism at the HLA-E locus predates most HLA-A and -B polymorphism. *Hum Immunol* 1992; 33 (3): 174.
27. Matte C, Lacaille J, Zijenah L, Ward B, Roger M. HLA-G and HLA-E polymorphisms in an indigenous African population. The ZVITAMBO Study Group. *Hum Immunol* 2000; 61 (11): 1150.
28. Strong RK, Holmes MA, Li P, Braun L, Lee N, Geraghty DE. HLA-E allelic variants. Correlating differential expression, peptide affinities, crystal structures, and thermal stabilities. *J Biol Chem* 2003; 278 (7): 5082.
29. Tamouza R, Rocha V, Busson M, et al. Association of HLA-E polymorphism with severe bacterial infection and early transplant-related mortality in matched unrelated bone marrow transplantation. *Transplantation* 2005; 80 (1): 140.
30. Borrego F, Ulbrecht M, Weiss EH, Coligan JE, Brooks AG. Recognition of human histocompatibility leukocyte antigen (HLA)-E complexed with HLA class I signal sequence-derived peptides by CD94/NKG2 confers protection from natural killer cell-mediated lysis. *J Exp Med* 1998; 187 (5): 813.
31. Millo E, Pietra G, Armirotti A, et al. Purification and HPLC-MS analysis of a naturally processed HCMV-derived peptide isolated from the HEK-293T/HLA-E+/U140+ cell transfectants and presented at the cell surface in the context of HLA-E. *J Immunol Methods* 2007; 322 (1-2): 128.
32. Baldanti F, Revello MG, Percivalle E, Labo N, Gerna G. Genomes of the endothelial cell-tropic variant and the parental Toledo strain of human cytomegalovirus are highly divergent. *J Med Virol* 2003; 69 (1): 76.
33. Stanton R, Westmoreland D, Fox JD, Davison AJ, Wilkinson GW. Stability of human cytomegalovirus genotypes in persistently infected renal transplant recipients. *J Med Virol* 2005; 75 (1): 42.
34. Aquino VH, Figueiredo LT. High prevalence of renal transplant recipients infected with more than one cytomegalovirus glycoprotein B genotype. *J Med Virol* 2000; 61 (1): 138.
35. Humar A, Kumar D, Gilbert C, Boivin G. Cytomegalovirus (CMV) glycoprotein B genotypes and response to antiviral therapy, in solid-organ-transplant recipients with CMV disease. *J Infect Dis* 2003; 188 (4): 581.

36. Ducancelle A, Belloc S, Alain S, et al. Comparison of sequential cytomegalovirus isolates in a patient with lymphoma and failing antiviral therapy. *J Clin Virol* 2004; 29 (4): 241.
37. Grundy JE, Lui SF, Super M, et al. Symptomatic cytomegalovirus infection in seropositive kidney recipients: reinfection with donor virus rather than reactivation of recipient virus. *Lancet* 1988; 2 (8603): 132.
38. Arav-Boger R, Zong JC, Foster CB. Loss of linkage disequilibrium and accelerated protein divergence in duplicated cytomegalovirus chemokine genes. *Virus Genes* 2005; 31 (1): 65.
39. Chou SW. Reactivation and recombination of multiple cytomegalovirus strains from individual organ donors. *J Infect Dis* 1989; 160 (1): 11.
40. Haberland M, Meyer-Konig U, Hufert FT. Variation within the glycoprotein B gene of human cytomegalovirus is due to homologous recombination. *J Gen Virol* 1999; 80 (Pt 6): 1495.
41. Benson JD, Huang ES. Human cytomegalovirus induces expression of cellular topoisomerase II. *J Virol* 1990; 64 (1): 9.
42. Puchhammer-Stockl E, Gorzer I, Zoufaly A, et al. Emergence of multiple cytomegalovirus strains in blood and lung of lung transplant recipients. *Transplantation* 2006; 81 (2): 187.
43. Zong JC, Arav-Boger R, Alcendor DJ, Hayward GS. Reflections on the interpretation of heterogeneity and strain differences based on very limited PCR sequence data from Kaposi's sarcoma-associated herpesvirus genomes. *J Clin Virol* 2007; 40 (1): 1.
44. Lafarge X, Merville P, Cazin MC, et al. Cytomegalovirus infection in transplant recipients resolves when circulating gammadelta T lymphocytes expand, suggesting a protective antiviral role. *J Infect Dis* 2001; 184 (5): 533.
45. Garrigue I, Boucher S, Couzi L, et al. Whole blood real-time quantitative PCR for cytomegalovirus infection follow-up in transplant recipients. *J Clin Virol* 2006; 36 (1): 72.
46. Garrigue I, Doussau A, Asselineau J, et al. Prediction of cytomegalovirus (CMV) plasma load from evaluation of CMV whole-blood load in samples from renal transplant recipients. *J Clin Microbiol* 2008; 46 (2): 493.

FIGURE LEGENDS

TABLE 1

Characteristics of the 3 patients' samples.

TABLE 2 :

Amplification of HCMV UL40: sequences of sense and anti-sense primers, and cycle conditions. The confirmation of the presence of 9Bx0411clone19 in the sample of patient 9Bx in November 2004 was performed thanks to a semi-nested PCR with 2 other anti-sense primers (UL40 9Bx) specific to 9Bx0411clone19 and to direct sequences of 9Bx0501, 9Bx0502 and 9Bx0504.

TABLE 3

UL40 nonapeptide variability. Comparison of AD169 with published data and sequences from patients 1Bx and 8Bx. No viral sequence from patient 9Bx was found different from VMAPRTLIL.

FIGURE 1

Evolution of Patient 8Bx after two renal transplantations, T1 (first transplantation) and T2 (second transplantation). GCV, ganciclovir; VGCV, valganciclovir; proph, prophylaxy; VL, viral load; MMF, mycophenolate mofetil; Cs, cyclosporine; Tac, tacrolimus.

HCMV loads were evaluated in 2003 with qualitative PCR in leukocytes (results “-“ and “+” before the vertical dotted line) and afterwards with whole blood real-time quantitative PCR (Garrigue et al., 2006; Garrigue et al., 2008a) (blue graph following the vertical dotted line). Whole blood viral loads are quantitated from 2.7 log₁₀ copies/ml (500 copies/ml, horizontal dotted line). Below this value and for comprehension purposes, weak positive results have been reported on the horizontal dotted line.

FIGURE 2

Phylogenetic representations of HCMV nucleotidic variation in UL40 for the 3 graft recipients. Samples were identified as follows : patient's number (1, 8 or 9)-Bx (Bordeaux, France)-year (2003 to 2006) and month of collection. For each sample, results of direct sequencing are presented in red and results of sequencing after cloning are presented with the number of each clone. The sequences were compared to homologous sequences available in GenBank database: AD169 (access number NC_001347), Merlin (NC_006273), Towne

(AY072777), Toledo (AH013698). **FIG. 2A**, Patient 1Bx. The 3 clones of 1Bx0501 presented in Fig. 3 for Bootscan analysis are ringed in orange. **FIG. 2B**, Patient 8Bx. **FIG. 3C**, Patient 9Bx. Clone 19 of 9Bx0411, presented in PCR results in Fig. 4 is ringed in blue.

Normalized Shannon entropy (S_n) for nucleotidic sequences is indicated for each quasispecies.

FIGURE 3

Results from the Bootscan analysis to evidence UL40 recombinant viral candidates or strains filiation in patient 1Bx' samples.

FIG. 3A: 1Bx0501 compared to 1Bx0408, 1Bx0411 and 1Bx0501clone57.

FIG. 3B: 1Bx0501clone57 compared to 1Bx0408, 1Bx0411 and 1Bx0501.

FIG. 3C: 1Bx0501clone49 compared to 1Bx0408, 1Bx0411 and 1Bx0501.

FIG. 3D: 1Bx0501clone52 compared to 1Bx0408, 1Bx0411 and 1Bx0501.

TABLE 1

HCMV	patients	renal	delay ^a	samples	copies/mL ^b	sample
D/R serostatus		transplant				identity ^c
D+R+	1	2004/07/12	50	1	1826	1Bx0408
			133	2	1070	1Bx0411
			177	3	60440	1Bx0501
D+R-	8	2003/06/11	189	1	6456	8Bx0312
			251	2	29670	8Bx0402
			350	3	4743	8Bx0405
			483	4	34140	8Bx0410
			574	5	1006	8Bx0501
		1029	6	3609	8Bx0604	
		1209	7	2616	8Bx0610	
D+R-	9	2004/07/24	125	1	64800	9Bx0411
			166	2	1076	9Bx0501
			214	3	794	9Bx0502
			271	4	975	9Bx0504

^a delay (days) between (first) renal transplantation and sample.

^b whole blood viral load (Garrigue et al., 2006; Garrigue et al., 2008a).

^c samples were identified as follows : patient number (1, 8 or 9)-Bx (Bordeaux, France)-year (03, 04, 05 or 06) and month of collection.

TABLE 2

		sense primer	anti-sense primer	cycle conditions
UL40	1 st PCR	5'-TCCTCCCTGGTACCCGATAACAG-3'	5'-CGGGCCAGGACTTTTAAATGGCC-3'	94°C 12 min 94°C 30 s - 59°C 30 s - 72°C 1min30: 50 cycles 72°C 7 min
	nested PCR	5'-GGTAAGGGCACTCGTGAGGATGTGC-3'	5'-TCCGAACGCTCGTGAGCAACAGTCG-3'	94°C 12 min 94°C 30 s - 59°C 30 s - 72°C 1min30: 40 cycles 72°C 7 min
UL40 9Bx	1 st PCR	5'-TCCTCCCTGGTACCCGATAACAG-3'	5'-ACGGTAACAACCGCCGGCGCGATG-3'	94°C 12 min 94°C 30 s - 60°C 30 s - 72°C 1min30: 50 cycles 72°C 7 min
	nested PCR	5'-GGTAAGGGCACTCGTGAGGATGTGC-3'	5'-ACGGTAACAACCGCCGGCGCGATG-3'	94°C 12 min 94°C 30 s - 60°C 30 s - 72°C 1min30: 40 cycles 72°C 7 min

TABLE 3

	AD169	V	M	A	P	R	T	L	I	L	n° of clones
Tomasec <i>et al</i> (2000)	Toledo								V		
Cerboni <i>et al</i> (2001)	109B* 3C*			T					V L		
	1Bx0411 clones						S		L		32/32
	1Bx0501 clones						S		L		2/29
	1Bx0501 clones				L						26/29
	8Bx0604 clones	A									1/30

* clinical isolates after propagation on human lung fibroblasts

FIGURE 1

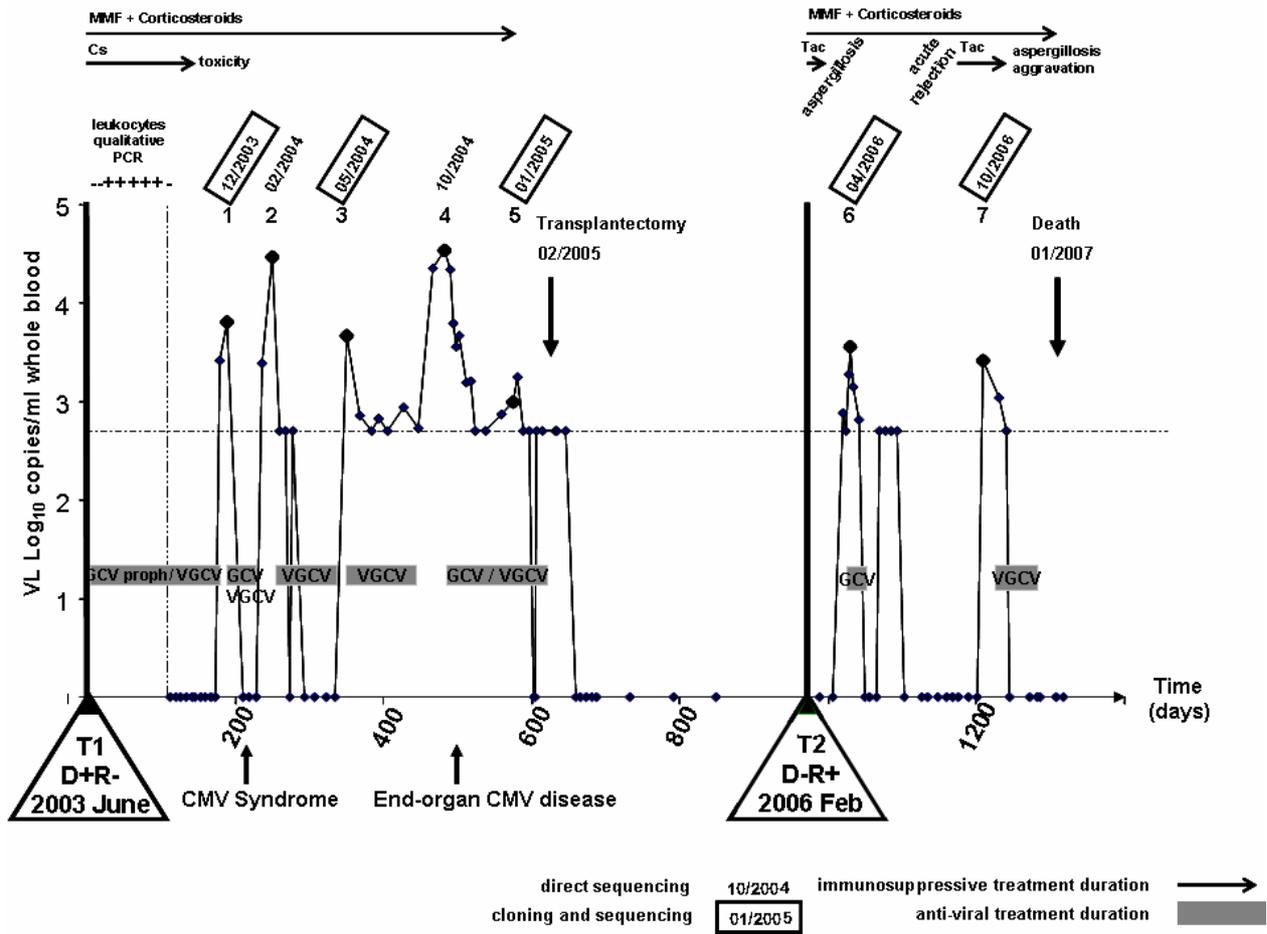


FIGURE 2

FIG 2A
Patient 1Bx HCMV UL40

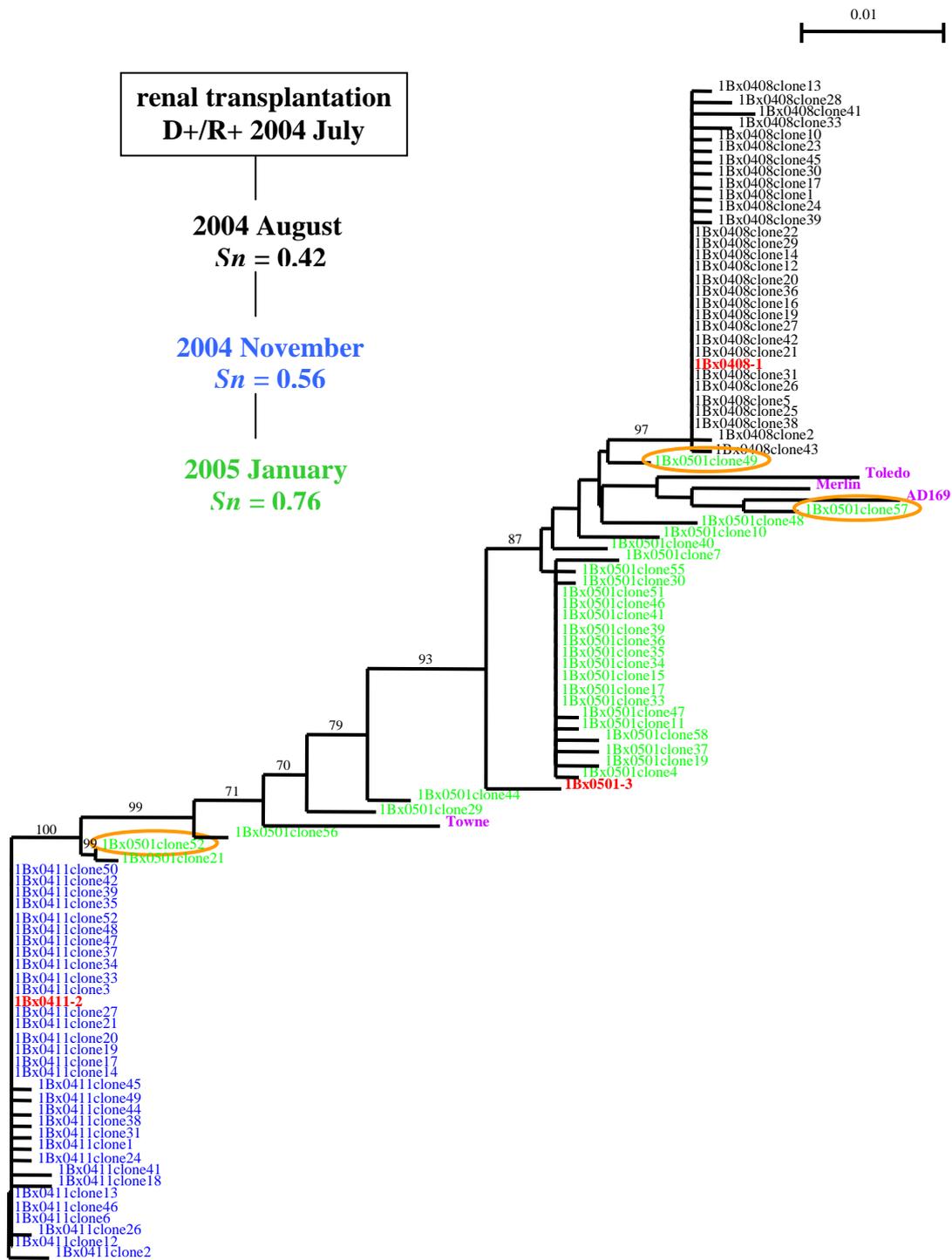


FIG 2B
Patient 8Bx HCMV UL40

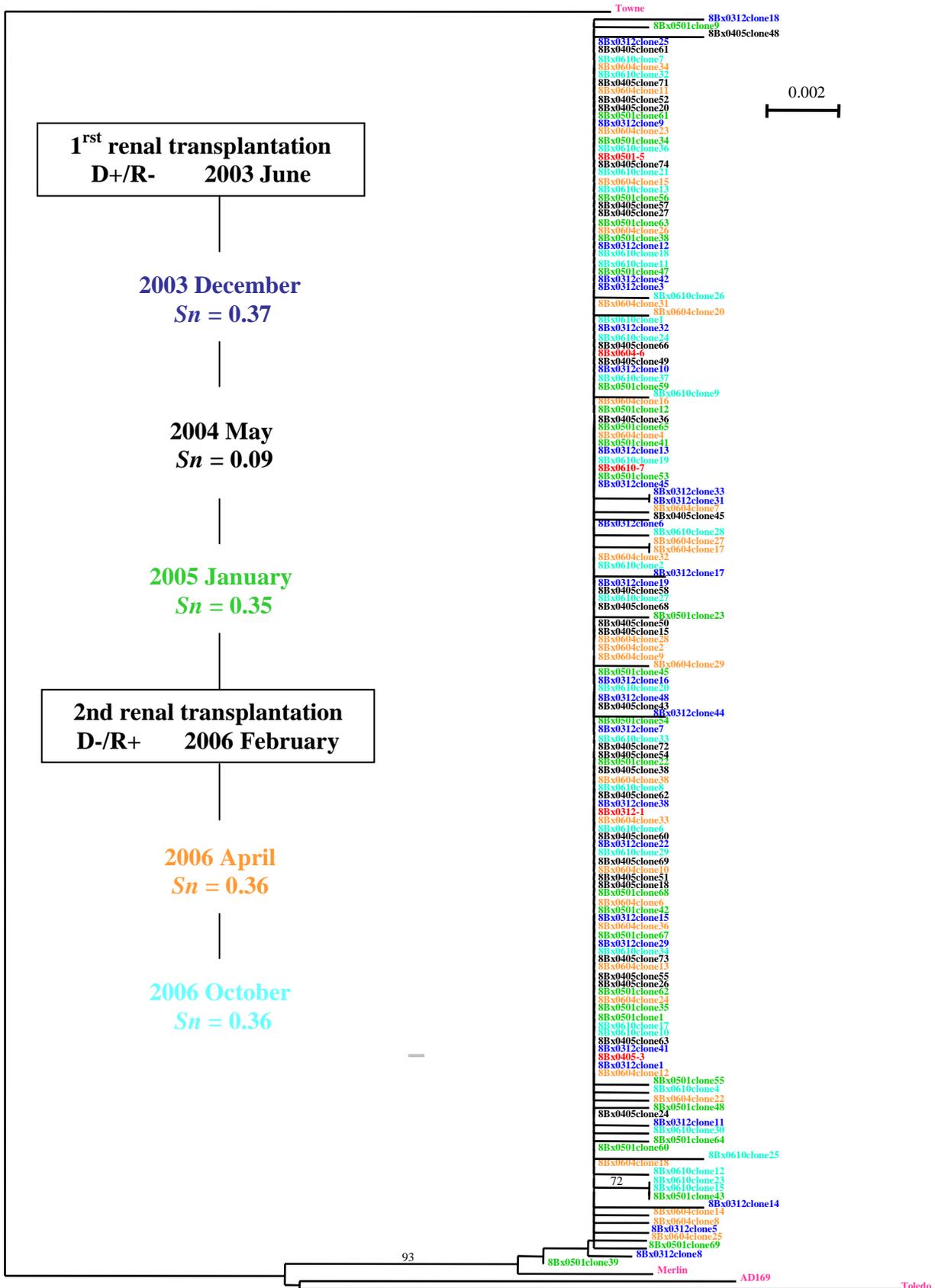


FIG 2C
Patient 9BX HCMV UL40

0.002

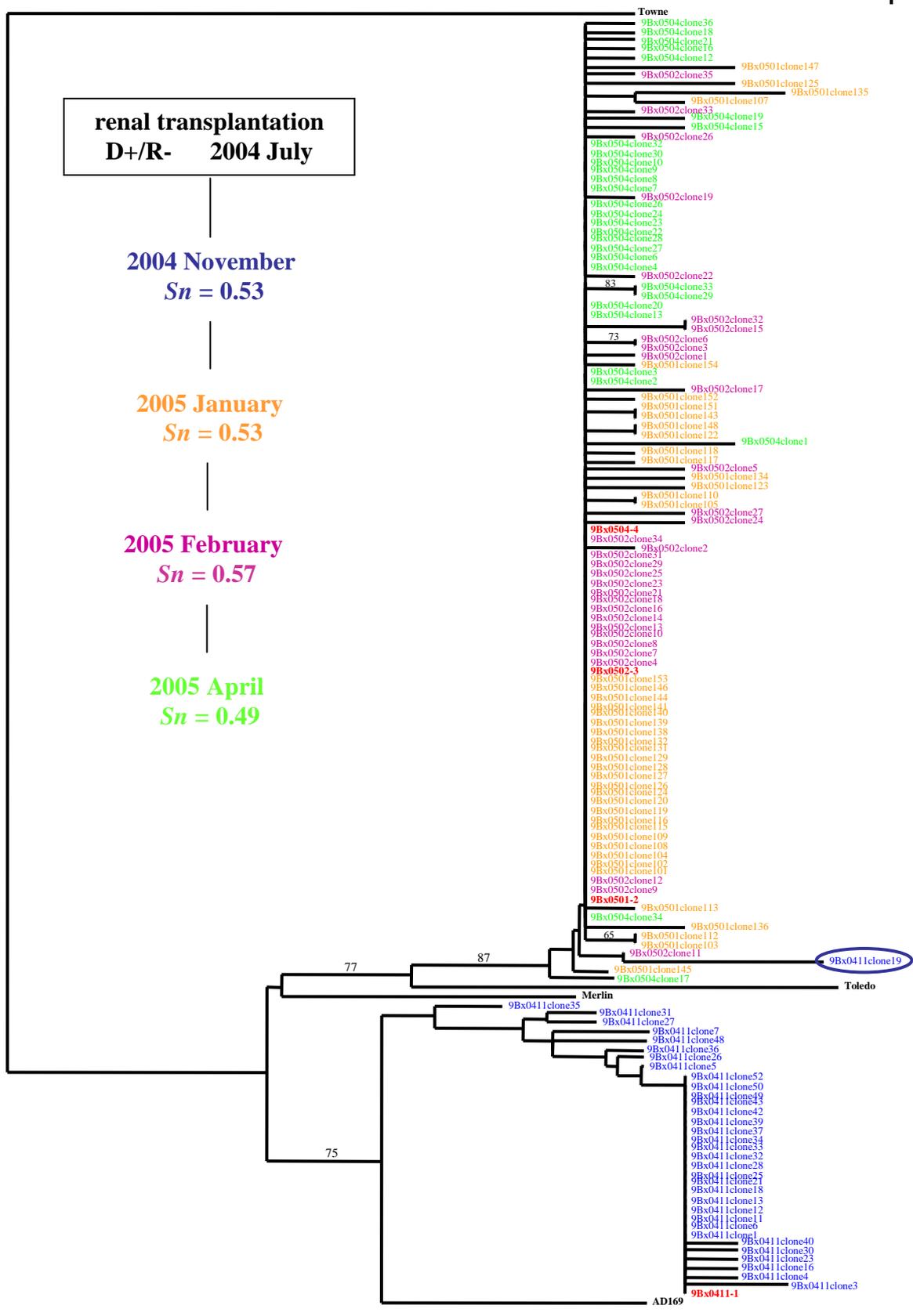
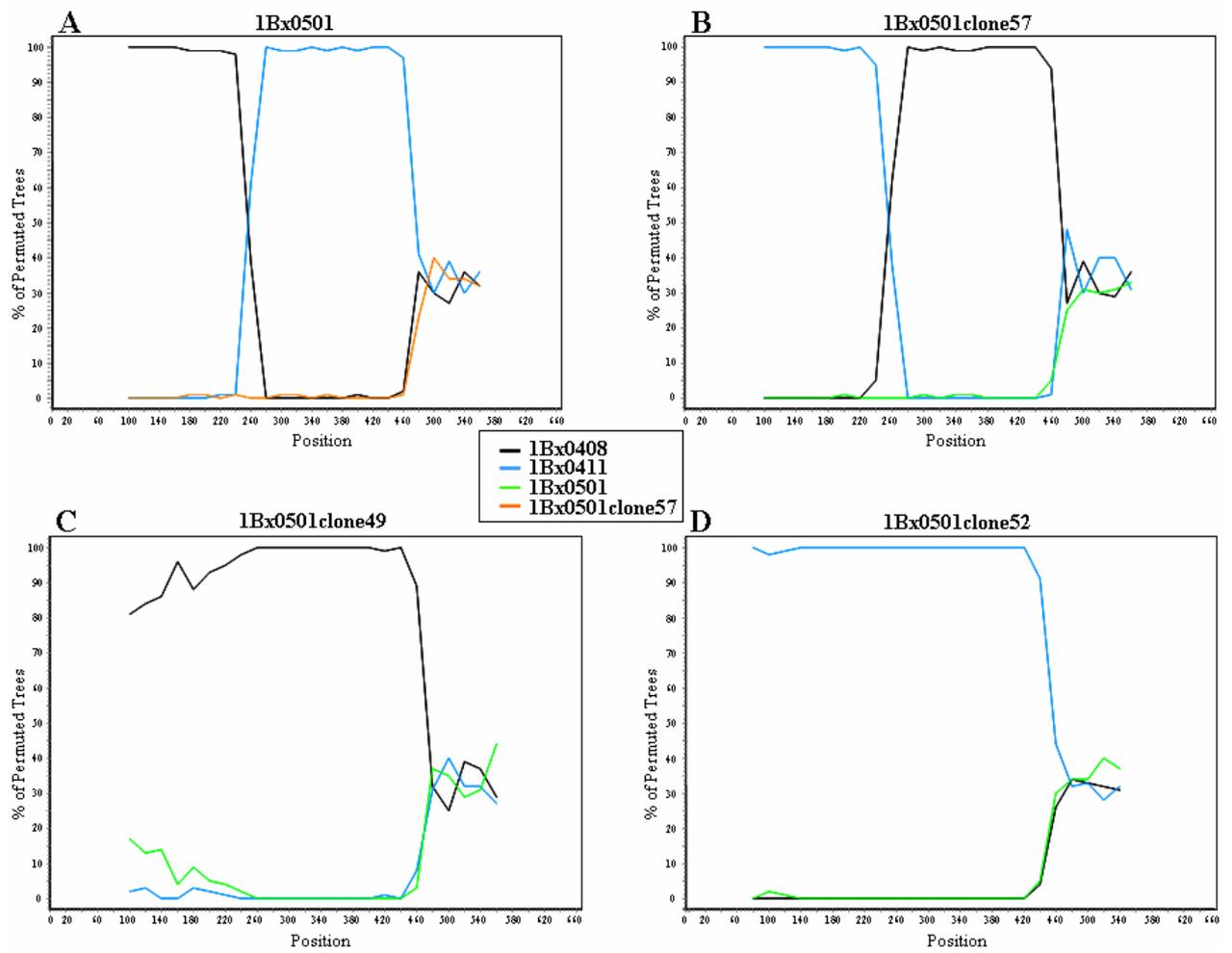


FIGURE 3



Annexe 2

Variability and recombination of clinical human cytomegalovirus strains from transplantation recipients

Muriel Faure-Della Corte^{a*}, Johan Samot^a, Isabelle Garrigue^{a,b}, Noël Magnin^a, Sandrine Reigadas^{a,b}, Lionel Couzi^c, Claire Dromer^d, Jean-François Velly^d, Julie Déchanet-Merville^e, Hervé J.A. Fleury^{a,b} and Marie-Edith Lafon^{a,b}.

Running title: HCMV variability and recombination.

Abstract word count: 146 words

Manuscript word count: 2584 words

Financial support : Centre Hospitalier Universitaire de Bordeaux and Institut Fédératif de Recherche 66 (IFR 66).

Abbreviations : BAL, bronchoalveolar lavage ; D, donor ; HCMV, human cytomegalovirus ; HSV, herpes simplex virus ; LIR-1, leukocyte inhibitory receptor-1 ; MHC, major histocompatibility complex ; NK, natural killer ; R, recipient ; WB, whole blood.

Abstract

Background: Human cytomegalovirus (HCMV) is the first cause of viral infection in immunocompromised transplanted patients.

Objectives: Here, five HCMV genes were studied to investigate the existence of recombination events in clinical strains *ex vivo*.

Study design: Sequencing and phylogenetic analysis were conducted on 21 strains from 16 renal and 5 lung transplant recipients.

Results: Nucleotidic polymorphism ranged from 6.6% (US3) to 12% (UL40), with a significant proportion of missense mutations (39% to 69%), some of which could have a functional impact. Analysis of the concatenated sequence (4804 nucleotides for each strain) evidenced two clusters of sequences presenting a reticulate topology suggestive of recombination events (SplitsTree). Phi-Test pointed numerous phylogenetically conflicting signals indicating a high statistical probability of recombination. The subsequent bootscan analysis was consistent with these data.

Conclusions: These results reinforce the prominent role of recombination in HCMV evolutionary history and adaptation to its host.

Key words: HCMV; variability; recombination; HCMV clinical strains; sequence analysis.

Background

Human cytomegalovirus (HCMV) is the largest virus in the *Herpesviridae* family, with a double-stranded DNA genome of 235 kbp encoding 164-167 predicted open reading frames (ORFs).^(Davison et al., 2003; Dolan et al., 2004) Like all herpesviruses, HCMV undergoes latency and reactivation episodes in the host. In the immunosuppressed patients, such as lung or renal graft recipients, HCMV is still responsible for the most frequent viral opportunistic infection.^(Rowshani et al., 2005)

To ascertain this long-lasting persistence in the infected persons, HCMV has evolved numerous immune evasion and modulation strategies, involving up to 20% of its genes, acting through three main mechanisms^(Mocarski, 2002): (1) mimicking/interference with major histocompatibility complex (MHC) class I- and class II-molecules (such as US3 and UL18); (2) modulation of natural killer (NK) function (e.g. UL40) and leukocyte behaviour (NK receptors agonists, pseudo-chemokines, -chemokine receptors, -cytokines (e.g. UL111a)); (3) modification of cell cycle and apoptosis events. Viral genomic variability, also depending on different mechanisms, could participate significantly in HCMV evasion.

In vitro viral recombination occurs when two HCMV strains co-infect the same cell and interact during replication to generate progeny which genomes consist of genetic segments obtained from both parental strains.^(Haberland, Meyer-Konig, and Hufert, 1999) Homologous recombinations were described for HSV and EBV (Lingen, Hengerer, and Falke, 1997; Walling et al., 1994), also belonging to the *Herpesviridae* family. During DNA replication, the polymerase may jump to another strand, as during HIV replication (Hu and Temin, 1990) or sequences may be exchanged between two variants. During HCMV infection, cellular topoisomerase II, also present in viral tegument, was shown to be activated.^(Benson and Huang, 1990) It plays an essential part in DNA replication, favours the appearance of breaks in double-stranded DNA and thus interactions between different strains replicating in a single cell. *In vivo*, a recombination mechanism was also alleged in different HCMV-infected patients populations such as HIV-infected persons, transplant recipients and infected newborns.^(Arav-Boger, Zong, and Foster, 2005; Chou, 1989; Ducancelle et al., 2004; Rasmussen et al., 2002)

Objectives

The aim of the present study was to bring data in favour of the occurrence of recombination events *ex vivo* among HCMV-uncultivated clinical strains.

In our previous work,^(Garrigue et al., 2007b) using direct sequencing of four immunomodulatory viral genes (US3, UL18, UL40 and UL111a) in blood samples of renal transplant patients developing an active HCMV infection, we showed a noticeable viral polymorphism. Briefly, each patient harboured a unique strain (i.e. a unique combination of the four genes), but an identical sequence for a given gene could be demonstrated among various patients. We subsequently followed the evolution of UL40 gene, by direct sequencing and sequencing after cloning, in patients with multiple HCMV reactivations (intra-patient studies).^(Garrigue et al., 2007b; Garrigue et al., 2008b) The present work is based on the transversal analysis of five HCMV genes (including UL123, which encodes regulatory protein IE1) in twenty-one patients. Our results indicate that recombination events do participate in the observed inter-patient variability.

Study design

1. Clinical HCMV strains

Twenty-one HCMV-infected patients (13 men, 8 women; median age (IQR (interquartile range)) of 55 (44; 56) years) were included. Sixteen underwent renal transplantation (1Bx to 17Bx) and five, lung transplantation (101Bx to 106Bx). Eight patients were donor-positive/recipient-positive (D+R+), 1 D-R+ and 12 D+R-. HCMV strains were collected from whole blood (WB) in 19 patients and from bronchoalveolar lavage (BAL) in two (101Bx and 102Bx).

2. DNA extraction

DNA was extracted from 200 µl WB or 200 µl BAL using the MagNATM Pure LC Total Nucleic Acid Isolation Kit (Roche Diagnostics).

3. PCR amplification and sequencing

Four HCMV genes were amplified prior to sequencing. For UL18, UL40, UL111a, US3, specific primers pairs were used as previously described^(Garrigue et al., 2007b). Additional primers (Eurofins, Germany) were designed to amplify and to sequence UL123, according to alignments of HCMV sequences available in the GenBank database (Table 1). The fragments were sequenced on both strands as previously described^(Merel et al., 2001) and a consensus sequence was generated.

4. Sequence analysis

The five genes were analysed separately as well as concatenated, according to their order in the viral genome. Identity and similarity values were calculated with the Needle program from the EMBOSS suite of software.^(Rice, Longden, and Bleasby, 2000)

Clustal W 1.83 was used for alignment of derived nucleotidic sequences with known strains as references: AD169 (access number NC_001347), Merlin (NC_006273), Towne (AY072777) and Toledo (AH013698). Traditional phylogenetic trees were inferred using the neighbour-joining method from matrix distances calculated after gapstripping of alignments, with a Kimura two-parameter algorithm.

Evolutionary relationships are usually represented using phylogenetic trees, based on a model of evolution dominated by mutations and speciation events. More realistic models must also account for other events, such as loss and duplication, or recombination. To be

represented, such data need a network graph, provided by the SplitsTree program.^(Huson and Bryant, 2006) The evolutionary history of the isolates is constructed from conflicting phylogenetic signals contained in sequence alignment when recombination events have participated in the evolution of the included isolates. A recombination will not be characterized by one but several branches connected to the parental strains. That is, if a complex network represents the relationships between isolates, it is likely that several recombination events have participated in their evolutionary history. A reticulate network is used to describe this reticulate evolution. Most importantly, the Phi-test^(Bruen, Philippe, and Bryant, 2006) is calculated from the SplitsTree program version 4 to validate whether detected conflicting signals are due to recombination. Recombinants have been detected with this method for HSV-1 and HSV-2.^(Norberg et al., 2007) To our knowledge, SplitsTree program was operated here to analyse HCMV genes for the first time. We studied the concatenated sequences of the five HCMV genes to simplify the nucleotidic alignments (4804 nucleotides) and gain a synthetic view of the compared sequences.

To reinforce the presence of recombination in the concatenated sequences of HCMV, the bootscan method included in the SimPlot program^(Lole et al., 1999) was applied to the recombinant candidates. Briefly, bootscanning relies on the alignment of a suspected recombinant sequence (target sequence) with a set of potential parental reference sequences. After optimal alignment is obtained, it is broken into sequential, overlapping segments (or windows) of 20 nucleotides which are fed into a phylogenetic analysis program. Bootstrapped phylogenetic trees are built for each segment. Finally, the bootstrap value for placing the unknown sequence with each of the reference ones is tabulated and plotted along the genome. The assumption is that the unknown will always reach high bootstrap values with its parental subtypes in windows. Schematically, bootscan indicates which sequence resembles mostly the target sequence, and in which proportion (percentage shown in the y axis), all along the nucleotidic sequence. In case of recombination, the software displays a message indicating: “Evidence of possible recombination event detected”, which can be confirmed by using the available “FindSites” tool.

5. Nucleotide sequence accession numbers

The sequences were deposited in Genbank with the following accession numbers: FJ491265, FJ491268, FJ491270, FJ491272, FJ491273, FJ491275, FJ491276, FJ491281, FJ491284, FJ491287, FJ491289, FJ491291 to FJ491296, FJ491298, FJ491299, FJ491301, FJ491306, FJ491308, FJ491315, FJ491317, FJ491319, FJ491324, FJ491327, FJ491334,

FJ491336, FJ491338, FJ491343, FJ491345, FJ491352, FJ491354, FJ491356, FJ491360, FJ491361, FJ491366, FJ491368, FJ491370, FJ491375, DQ367940, DQ367943, DQ367945, DQ367947 to DQ367951, DQ367956, DQ367962, DQ367964, DQ367966 to DQ367970, DQ367972, DQ367974, DQ367976 to DQ367979, DQ367984, DQ367987, DQ367990, DQ367992, DQ367994 to DQ367999, DQ368002, DQ368004, DQ368006 to DQ368010, DQ368015, DQ368018, DQ368021, DQ368023, DQ368025 to DQ368030, DQ368033, DQ368035, DQ368037 to DQ368041, DQ368046, DQ368049, DQ368052, DQ368054, DQ368056 to DQ368060, EF488364 and EF535834.

Results

1. Sequence variability of HCMV isolates

Sequence polymorphism was observed in the twenty-one HCMV clinical strains for the five selected genes (Table 2). Missense or silent variable sites were found throughout the five genes, affecting from 6.6% (US3) to 12% (UL40) of each gene sequence. Table 2 also indicates the maximum percentage of strain to strain variability for each gene.

2. Phylogenetic trees and SplitsTree recombination analysis

Phylogenetic trees were constructed for each of the five genes and for the concatenated sequences (Fig. 1). Considering the five HCMV sequences, all clinical isolates presented a unique genetic pattern and each patient appeared to be infected with a particular HCMV variant. The identity was 95.4% between the two most distant isolates (14Bx0507 and 4Bx0409) and 99.9% between the closest (106Bx0411 and 15Bx0506). The latter clustered whatever the gene, with identical UL18, UL40, UL123 and US3 genes. For UL111a gene, there was an insertion of six nucleotides in 106Bx0411 compared with 15Bx0506.

For each gene and for the concatenated sequences were constructed recombination networks including all clinical isolates and the four reference strains (Fig. 2A). All networks presented a reticulate topology, the more complex being found for UL40 (666 bp) and UL123 (1762 bp) genes (Fig. 2B and 2C). These results were confirmed when analysis was restricted to the 12 concatenated sequences presenting no double population (Fig. 2D). Moreover, the results obtained with concatenated sequences could be prefigured by those obtained with single UL40 gene cloned sequences: for example in patient 1Bx (Fig. 2E) the phi test did find statistically significant evidence for recombination with $p < 10^{-17}$ and in patient 9Bx (Fig. 2F), the phi test did not find statistically significant evidence for recombination with $p = 0.667$.

Phi-test was calculated from the alignment containing all the isolates and reference sequences : 297 informative sites were detected (297/4804) and evidence for recombination was statistically significant ($p < 0.05$).

3. Bootscan analysis

As shown in Fig. 3, intra- (Fig. 3A) or inter- (Fig. 3B) genic recombination events were suggested by the bootscan analysis, which compares the similarity of various genes all along the concatenated sequences. For example, 106Bx0411 seems to result from recombination events between 103Bx0607 and 7Bx0506, and 10Bx0502 from 104Bx0604 and 8Bx0312. On

the opposite, Simplot software could detect no evidence of recombination between other sequences (example given Fig. 3C).

Discussion

We and others showed a noticeable polymorphism in HCMV clinical strains, for which different mechanisms may be involved: natural polymorphism, recombination or even quasispecies-like evolution.^(Baldanti et al., 2003; Dolan et al., 2004; Garrigue et al., 2007b; Garrigue et al., 2008b) In our previous work, we evidenced an intragenic variability compatible with a recombination event in UL40 gene in one infected patient followed over-time^(Garrigue et al., 2008b). Homologous recombination was shown in HCMV laboratory strains and also in clinical strains, but always after a culture step.^(Arav-Boger, Zong, and Foster, 2005; Baldanti et al., 2003; Chou, 1989; Dolan et al., 2004; Haberland, Meyer-Konig, and Hufert, 1999; Rasmussen, Geissler, and Winters, 2003) In the present study, we investigated *ex vivo* uncultivated virus-infected samples and focused on HCMV evolution by recombination, by comparing clinical strains infecting different patients.

The five selected unlinked genes, UL18, UL40, UL111a, UL123 and US3, were evenly distributed within the genome. This choice could be contested, since HCMV genome comprises more variable sequences (e.g. UL55 encoding glycoprotein B). However, there were enough mutations to allow SimPlot analysis to conclude in the existence of recombination events.

The sequence alignment of our clinical HCMV strains contained several phylogenetically conflicting signals, most of which could result from homologous recombination. Selection pressure also leads to the accumulation of conflicting signals at specific sites or epitopes. However, our bootscan analysis detected such signals all along the five genes, which argued against selection pressure.

Conflicting phylogenetic signals could also have resulted from random introduction of nucleotide substitutions by *Taq* polymerase during the PCR.^(Zong et al., 2007) This is highly improbable with our standardized experimental conditions and the AmpliTaq Gold enzyme (Applied Biosystems). Firstly, the mean error rate of the enzyme is $5 \cdot 10^{-6}$ /incorporated nucleotide/cycle (i.e. an error every 200,000 nt) and the error must be introduced early in the amplification reaction. Secondly, consensus sequences generated from two strands in both directions showed very few discrepancies. Thirdly, this explanation is unlikely to account for the high percentage of variable sites in each gene.

The two clusters observed after phylogenetic analysis of the concatenated sequences (Fig. 2A) were not explained by the transplanted organ (kidney or lung), a common donor or a concomitant hospitalization.

The present study confirmed a significant polymorphism within HCMV strains directly sequenced from blood or BAL in transplant recipients. Unfortunately, we did not obtain enough sequences from BAL samples to address the question of HCMV compartmentalization and preferential site of recombination. Nucleotidic variations could induce functional consequences for the corresponding proteins, which underlines their potential importance in immune interference during clinical infection. For example, in pUL18, fourteen isolates presented K42R and A43T mutations, and thirteen the Q202H mutation, which could modulate its interaction with Leukocyte Inhibitory Receptor-1 (LIR-1). (Cerboni et al., 2006) *In vitro* functional tests are needed to evaluate the impact of HCMV protein variability.

Although recombination in HCMV strains has been proven *in vitro*,^(Haberland, Meyer-Konig, and Hufert, 1999) to our knowledge, this is the first study on clinical strains using such software. Principally, this work provided additional data arguing in favour of the key role of recombination as a major evolution mechanism for HCMV. The absence of *in vitro* selection before sequence analysis highlighted the complexity of the recombination pattern, which involved multiple both inter- and intra-genic events. To circumvent immune control and enable HCMV persistence, we suggest recombination may represent a crucial viral strategy.

Conflicts of interest

No conflicts of interest have been identified.

Acknowledgments

We thank Jean-Luc Taupin, Valérie Jauvin and Patricia Recordon-Pinson for so many helpful discussions. We are indebted to the technicians of the Virology Laboratory in Bordeaux University Hospital.

References

- Akter, P., Cunningham, C., McSharry, B. P., Dolan, A., Addison, C., Dargan, D. J., Hassan-Walker, A. F., Emery, V. C., Griffiths, P. D., Wilkinson, G. W., and Davison, A. J. (2003). Two novel spliced genes in human cytomegalovirus. *J Gen Virol* **84**(Pt 5), 1117-22.
- Alain, S., and Mazon, M. C. (2001). "Infections à cytomégalo virus." EMC 8-052-C-10.
- Alain, S., and Mazon, M. C. (2003). *Traité de virologie médicale. eds Estem Chapitre 12*, 195-211.
- Aquino, V. H., and Figueiredo, L. T. (2000). High prevalence of renal transplant recipients infected with more than one cytomegalovirus glycoprotein B genotype. *J Med Virol* **61**(1), 138-42.
- Arav-Boger, R., Foster, C. B., Zong, J. C., and Pass, R. F. (2006). Human Cytomegalovirus-Encoded alpha -Chemokines Exhibit High Sequence Variability in Congenitally Infected Newborns. *J Infect Dis* **193**(6), 788-91.
- Arav-Boger, R., Zong, J. C., and Foster, C. B. (2005). Loss of linkage disequilibrium and accelerated protein divergence in duplicated cytomegalovirus chemokine genes. *Virus Genes* **31**(1), 65-72.
- Baldanti, F., Revello, M. G., Percivalle, E., Labo, N., and Gerna, G. (2003). Genomes of the endothelial cell-tropic variant and the parental Toledo strain of human cytomegalovirus are highly divergent. *J Med Virol* **69**(1), 76-81.
- Benson, J. D., and Huang, E. S. (1990). Human cytomegalovirus induces expression of cellular topoisomerase II. *J Virol* **64**(1), 9-15.
- Borrego, F., Ulbrecht, M., Weiss, E. H., Coligan, J. E., and Brooks, A. G. (1998). Recognition of human histocompatibility leukocyte antigen (HLA)-E complexed with HLA class I signal sequence-derived peptides by CD94/NKG2 confers protection from natural killer cell-mediated lysis. *J Exp Med* **187**(5), 813-8.
- Braud, V. M., Allan, D. S., O'Callaghan, C. A., Soderstrom, K., D'Andrea, A., Ogg, G. S., Lazetic, S., Young, N. T., Bell, J. I., Phillips, J. H., Lanier, L. L., and McMichael, A. J. (1998). HLA-E binds to natural killer cell receptors CD94/NKG2A, B and C. *Nature* **391**(6669), 795-9.
- Braud, V. M., Tomasec, P., and Wilkinson, G. W. (2002). Viral evasion of natural killer cells during human cytomegalovirus infection. *Curr Top Microbiol Immunol* **269**, 117-29.
- Britt, W. J., and Mach, M. (1996). Human cytomegalovirus glycoproteins. *Intervirolgy* **39**(5-6), 401-12.
- Bruen, T. C., Philippe, H., and Bryant, D. (2006). A simple and robust statistical test for detecting the presence of recombination. *Genetics* **172**(4), 2665-81.
- Burton, D. R., Saphire, E. O., and Parren, P. W. (2001). A model for neutralization of viruses based on antibody coating of the virion surface. *Curr Top Microbiol Immunol* **260**, 109-43.
- Cannon, M. J., Schmid, D. S., and Hyde, T. B. (2010). Review of cytomegalovirus seroprevalence and demographic characteristics associated with infection. *Rev Med Virol* **20**(4), 202-13.
- Cerboni, C., Achour, A., Warnmark, A., Mousavi-Jazi, M., Sandalova, T., Hsu, M. L., Cosman, D., Karre, K., and Carbone, E. (2006). Spontaneous mutations in the human CMV HLA class I homologue UL18 affect its binding to the inhibitory receptor LIR-1/ILT2/CD85j. *Eur J Immunol* **36**(3), 732-41.
- Cerboni, C., Mousavi-Jazi, M., Wakiguchi, H., Carbone, E., Karre, K., and Soderstrom, K. (2001). Synergistic effect of IFN-gamma and human cytomegalovirus protein UL40 in

- the HLA-E-dependent protection from NK cell-mediated cytotoxicity. *Eur J Immunol* **31**(10), 2926-35.
- Chang, W. L., Baumgarth, N., Yu, D., and Barry, P. A. (2004). Human cytomegalovirus-encoded interleukin-10 homolog inhibits maturation of dendritic cells and alters their functionality. *J Virol* **78**(16), 8720-31.
- Chee, M. S., Bankier, A. T., Beck, S., Bohni, R., Brown, C. M., Cerny, R., Horsnell, T., Hutchison, C. A., 3rd, Kouzarides, T., Martignetti, J. A., and et al. (1990). Analysis of the protein-coding content of the sequence of human cytomegalovirus strain AD169. *Curr Top Microbiol Immunol* **154**, 125-69.
- Chou, S. W. (1989). Reactivation and recombination of multiple cytomegalovirus strains from individual organ donors. *J Infect Dis* **160**(1), 11-5.
- Compton, T. (2004). Receptors and immune sensors: the complex entry path of human cytomegalovirus. *Trends Cell Biol* **14**(1), 5-8.
- Crough, T., and Khanna, R. (2009). Immunobiology of human cytomegalovirus: from bench to bedside. *Clin Microbiol Rev* **22**(1), 76-98, Table of Contents.
- Davison, A. J., Dolan, A., Akter, P., Addison, C., Dargan, D. J., Alcendor, D. J., McGeoch, D. J., and Hayward, G. S. (2003). The human cytomegalovirus genome revisited: comparison with the chimpanzee cytomegalovirus genome. *J Gen Virol* **84**(Pt 1), 17-28.
- Dechanet, J., Merville, P., Lim, A., Retiere, C., Pitard, V., Lafarge, X., Michelson, S., Meric, C., Hallet, M. M., Kourilsky, P., Potaux, L., Bonneville, M., and Moreau, J. F. (1999). Implication of gammadelta T cells in the human immune response to cytomegalovirus. *J Clin Invest* **103**(10), 1437-49.
- Dolan, A., Cunningham, C., Hector, R. D., Hassan-Walker, A. F., Lee, L., Addison, C., Dargan, D. J., McGeoch, D. J., Gatherer, D., Emery, V. C., Griffiths, P. D., Sinzger, C., McSharry, B. P., Wilkinson, G. W., and Davison, A. J. (2004). Genetic content of wild-type human cytomegalovirus. *J Gen Virol* **85**(Pt 5), 1301-12.
- Domingo, E., Sabo, D., Taniguchi, T., and Weissmann, C. (1978). Nucleotide sequence heterogeneity of an RNA phage population. *Cell* **13**(4), 735-44.
- Ducancelle, A., Belloc, S., Alain, S., Scieux, C., Malphettes, M., Petit, F., Brouet, J. C., Sanson Le Pors, M. J., and Mazon, M. C. (2004). Comparison of sequential cytomegalovirus isolates in a patient with lymphoma and failing antiviral therapy. *J Clin Virol* **29**(4), 241-7.
- Dunn, W., Chou, C., Li, H., Hai, R., Patterson, D., Stolc, V., Zhu, H., and Liu, F. (2003). Functional profiling of a human cytomegalovirus genome. *Proc Natl Acad Sci U S A* **100**(24), 14223-8.
- Evers, D. L., Wang, X., and Huang, E. S. (2004). Cellular stress and signal transduction responses to human cytomegalovirus infection. *Microbes Infect* **6**(12), 1084-93.
- Farber, S., and Wolbach, S. B. (1932). Intranuclear and Cytoplasmic Inclusions ("Protozoan-Like Bodies") in the Salivary Glands and Other Organs of Infants. *Am J Pathol* **8**(2), 123-136 3.
- Gandhi, M. K., and Khanna, R. (2004). Human cytomegalovirus: clinical aspects, immune regulation, and emerging treatments. *Lancet Infect Dis* **4**(12), 725-38.
- Garrigue, I., Boucher, S., Couzi, L., Caumont, A., Dromer, C., Neau-Cransac, M., Tabrizi, R., Schrive, M. H., Fleury, H., and Lafon, M. E. (2006). Whole blood real-time quantitative PCR for cytomegalovirus infection follow-up in transplant recipients. *J Clin Virol* **36**(1), 72-5.
- Garrigue, I., Corte, M. F., Magnin, N., Couzi, L., Capdepon, S., Rio, C., Merville, P., Dechanet-Merville, J., Fleury, H., and Lafon, M. E. (2007a). Variability of UL18,

- UL40, UL111a and US3 immunomodulatory genes among human cytomegalovirus clinical isolates from renal transplant recipients. *J Clin Virol* **40**(2), 120-8.
- Garrigue, I., Doussau, A., Asselineau, J., Bricout, H., Couzi, L., Rio, C., Merville, P., Fleury, H., Lafon, M. E., and Thiebaut, R. (2008a). Prediction of cytomegalovirus (CMV) plasma load from evaluation of CMV whole-blood load in samples from renal transplant recipients. *J Clin Microbiol* **46**(2), 493-8.
- Garrigue, I., Faure-Della Corte, M., Magnin, N., Couzi, L., Capdepon, S., Rio, C., Merville, P., Dechanet-Merville, J., Fleury, H., and Lafon, M. E. (2007b). Variability of UL18, UL40, UL111a and US3 immunomodulatory genes among human cytomegalovirus clinical isolates from renal transplant recipients. *J Clin Virol* **40**(2), 120-8.
- Garrigue, I., Faure-Della Corte, M., Magnin, N., Recordon-Pinson, P., Couzi, L., Lebrette, M. E., Schrive, M. H., Roncin, L., Taupin, J. L., Dechanet-Merville, J., Fleury, H., and Lafon, M. E. (2008b). UL40 human cytomegalovirus variability evolution patterns over time in renal transplant recipients. *Transplantation* **86**(6), 826-35.
- Geraghty, D. E., Stockschleider, M., Ishitani, A., and Hansen, J. A. (1992). Polymorphism at the HLA-E locus predates most HLA-A and -B polymorphism. *Hum Immunol* **33**(3), 174-84.
- Gicklhorn, D., Eickmann, M., Meyer, G., Ohlin, M., and Radsak, K. (2003). Differential effects of glycoprotein B epitope-specific antibodies on human cytomegalovirus-induced cell-cell fusion. *J Gen Virol* **84**(Pt 7), 1859-62.
- Gilbert, M. J., Riddell, S. R., Plachter, B., and Greenberg, P. D. (1996). Cytomegalovirus selectively blocks antigen processing and presentation of its immediate-early gene product. *Nature* **383**(6602), 720-2.
- Greijer, A. E., Dekkers, C. A., and Middeldorp, J. M. (2000). Human cytomegalovirus virions differentially incorporate viral and host cell RNA during the assembly process. *J Virol* **74**(19), 9078-82.
- Grundy, J. E., Lui, S. F., Super, M., Berry, N. J., Sweny, P., Fernando, O. N., Moorhead, J., and Griffiths, P. D. (1988). Symptomatic cytomegalovirus infection in seropositive kidney recipients: reinfection with donor virus rather than reactivation of recipient virus. *Lancet* **2**(8603), 132-5.
- Guermonez, P., Valladeau, J., Zitvogel, L., Thery, C., and Amigorena, S. (2002). Antigen presentation and T cell stimulation by dendritic cells. *Annu Rev Immunol* **20**, 621-67.
- Haberland, M., Meyer-Konig, U., and Hufert, F. T. (1999). Variation within the glycoprotein B gene of human cytomegalovirus is due to homologous recombination. *J Gen Virol* **80** (Pt 6), 1495-500.
- Hu, W. S., and Temin, H. M. (1990). Genetic consequences of packaging two RNA genomes in one retroviral particle: pseudodiploidy and high rate of genetic recombination. *Proc Natl Acad Sci U S A* **87**(4), 1556-60.
- Humar, A., Kumar, D., Gilbert, C., and Boivin, G. (2003). Cytomegalovirus (CMV) glycoprotein B genotypes and response to antiviral therapy, in solid-organ-transplant recipients with CMV disease. *J Infect Dis* **188**(4), 581-4.
- Huson, D. H., and Bryant, D. (2006). Application of phylogenetic networks in evolutionary studies. *Mol Biol Evol* **23**(2), 254-67.
- Imbert-Marcille, B. M. (2001). Histoire naturelle des infections à cytomégalo-virus. *Médecine Thérapeutique* **7**(8), 577-584.
- Isaacson, M. K., Feire, A. L., and Compton, T. (2007). Epidermal growth factor receptor is not required for human cytomegalovirus entry or signaling. *J Virol* **81**(12), 6241-7.
- Jesionek, A., and Kiolemenoglou, B. (1904). Ueber einen Befund von protozoenartigen Gebilden in den Organen eines hereditär-luetischen Foetus. *Muenchner Med Wochenschr* **51**:1905–1907.

- Kalejta, R. F. (2008). Tegument proteins of human cytomegalovirus. *Microbiol Mol Biol Rev* **72**(2), 249-65, table of contents.
- Kern, F., Faulhaber, N., Frommel, C., Khatamzas, E., Prosch, S., Schonemann, C., Kretzschmar, I., Volkmer-Engert, R., Volk, H. D., and Reinke, P. (2000). Analysis of CD8 T cell reactivity to cytomegalovirus using protein-spanning pools of overlapping pentadecapeptides. *Eur J Immunol* **30**(6), 1676-82.
- Kondo, E., Akatsuka, Y., Kuzushima, K., Tsujimura, K., Asakura, S., Tajima, K., Kagami, Y., Kodera, Y., Tanimoto, M., Morishima, Y., and Takahashi, T. (2004). Identification of novel CTL epitopes of CMV-pp65 presented by a variety of HLA alleles. *Blood* **103**(2), 630-8.
- Lafarge, X., Merville, P., Cazin, M. C., Berge, F., Potaux, L., Moreau, J. F., and Dechanet-Merville, J. (2001). Cytomegalovirus infection in transplant recipients resolves when circulating gammadelta T lymphocytes expand, suggesting a protective antiviral role. *J Infect Dis* **184**(5), 533-41.
- Le Roy, E., and Davignon, J. L. (2005). Human cytomegalovirus-specific CD4(+) T-cell clones recognize cross-reactive peptides from the immediate early 1 protein. *Viral Immunol* **18**(2), 391-6.
- Lee, N., Goodlett, D. R., Ishitani, A., Marquardt, H., and Geraghty, D. E. (1998). HLA-E surface expression depends on binding of TAP-dependent peptides derived from certain HLA class I signal sequences. *J Immunol* **160**(10), 4951-60.
- Legrand-Abravanel, F., Claudinon, J., Nicot, F., Dubois, M., Chapuy-Regaud, S., Sandres-Saune, K., Pasquier, C., and Izopet, J. (2007). New natural intergenotypic (2/5) recombinant of hepatitis C virus. *J Virol* **81**(8), 4357-62.
- Lin, Y. L., Chang, P. C., Wang, Y., and Li, M. (2008). Identification of novel viral interleukin-10 isoforms of human cytomegalovirus AD169. *Virus Res* **131**(2), 213-23.
- Lingen, M., Hengerer, F., and Falke, D. (1997). Mixed vaginal infections of Balb/c mice with low virulent herpes simplex type 1 strains result in restoration of virulence properties: vaginitis/vulvitis and neuroinvasiveness. *Med Microbiol Immunol* **185**(4), 217-22.
- Liu, Z., Winkler, M., and Biegelke, B. (2009). Human cytomegalovirus: host immune modulation by the viral US3 gene. *Int J Biochem Cell Biol* **41**(3), 503-6.
- Lole, K. S., Bollinger, R. C., Paranjape, R. S., Gadkari, D., Kulkarni, S. S., Novak, N. G., Ingersoll, R., Sheppard, H. W., and Ray, S. C. (1999). Full-length human immunodeficiency virus type 1 genomes from subtype C-infected seroconverters in India, with evidence of intersubtype recombination. *J Virol* **73**(1), 152-60.
- Marchini, A., Liu, H., and Zhu, H. (2001). Human cytomegalovirus with IE-2 (UL122) deleted fails to express early lytic genes. *J Virol* **75**(4), 1870-8.
- Matte, C., Lacaille, J., Zijenah, L., Ward, B., and Roger, M. (2000). HLA-G and HLA-E polymorphisms in an indigenous African population. The ZVITAMBO Study Group. *Hum Immunol* **61**(11), 1150-6.
- McCormick, A. L. (2008). Control of apoptosis by human cytomegalovirus. *Curr Top Microbiol Immunol* **325**, 281-95.
- Meier, J. L., and Stinski, M. F. (2006). Major Immediate early enhancer and its gene products. *Cytomegaloviruses, molecular biology and immunology*. **Chapter 8:151-166**.
- Merel, P., Pellegrin, I., Garrigue, I., Caumont, A., Schrive, M. H., Birac, V., Bonot, P., and Fleury, H. (2001). Comparison of capillary electrophoresis sequencing with the new CEQ 2000 DNA Analysis System to conventional gel based systems for HIV drug resistance analysis. *J Virol Methods* **98**(1), 9-16.
- Millo, E., Pietra, G., Armirotti, A., Vacca, P., Mingari, M. C., Moretta, L., and Damonte, G. (2007). Purification and HPLC-MS analysis of a naturally processed HCMV-derived

- peptide isolated from the HEK-293T/HLA-E+/UI40+ cell transfectants and presented at the cell surface in the context of HLA-E. *J Immunol Methods* **322**(1-2), 128-36.
- Misaghi, S., Sun, Z. Y., Stern, P., Gaudet, R., Wagner, G., and Ploegh, H. (2004). Structural and functional analysis of human cytomegalovirus US3 protein. *J Virol* **78**(1), 413-23.
- Mocarski, E. S., and Courcelle, C. T. (2001). "Cytomegaloviruses and their replication in Fields Virology. Knipe DM and Howley PM, editors-in-chief." Philadelphia, PA, USA:Lippincott Williams & Wilkins ed.
- Mocarski, E. S., Jr. (2002). Immunomodulation by cytomegaloviruses: manipulative strategies beyond evasion. *Trends Microbiol* **10**(7), 332-9.
- Murphy, E., Yu, D., Grimwood, J., Schmutz, J., Dickson, M., Jarvis, M. A., Hahn, G., Nelson, J. A., Myers, R. M., and Shenk, T. E. (2003). Coding potential of laboratory and clinical strains of human cytomegalovirus. *Proc Natl Acad Sci U S A* **100**(25), 14976-81.
- Nattermann, J., Nischalke, H. D., Hofmeister, V., Ahlenstiel, G., Zimmermann, H., Leifeld, L., Weiss, E. H., Sauerbruch, T., and Spengler, U. (2005a). The HLA-A2 restricted T cell epitope HCV core 35-44 stabilizes HLA-E expression and inhibits cytotoxicity mediated by natural killer cells. *Am J Pathol* **166**(2), 443-53.
- Nattermann, J., Nischalke, H. D., Hofmeister, V., Kupfer, B., Ahlenstiel, G., Feldmann, G., Rockstroh, J., Weiss, E. H., Sauerbruch, T., and Spengler, U. (2005b). HIV-1 infection leads to increased HLA-E expression resulting in impaired function of natural killer cells. *Antivir Ther* **10**(1), 95-107.
- Nevels, M., Brune, W., and Shenk, T. (2004). SUMOylation of the human cytomegalovirus 72-kilodalton IE1 protein facilitates expression of the 86-kilodalton IE2 protein and promotes viral replication. *J Virol* **78**(14), 7803-12.
- Nevels, M., Paulus, C., and Shenk, T. (2004). Human cytomegalovirus immediate-early 1 protein facilitates viral replication by antagonizing histone deacetylation. *Proc Natl Acad Sci U S A* **101**(49), 17234-9.
- Nitzsche, A., Paulus, C., and Nevels, M. (2008). Temporal dynamics of cytomegalovirus chromatin assembly in productively infected human cells. *J Virol* **82**(22), 11167-80.
- Norberg, P., Kasubi, M. J., Haarr, L., Bergstrom, T., and Liljeqvist, J. A. (2007). Divergence and recombination of clinical herpes simplex virus type 2 isolates. *J Virol* **81**(23), 13158-67.
- Pass, R. F. (2004). A key role for adolescents in the epidemiology of cytomegalovirus and genital herpes infections. *Clin Infect Dis* **39**(10), 1439-40.
- Paya, C., Humar, A., Dominguez, E., Washburn, K., Blumberg, E., Alexander, B., Freeman, R., Heaton, N., and Pescovitz, M. D. (2004). Efficacy and safety of valganciclovir vs. oral ganciclovir for prevention of cytomegalovirus disease in solid organ transplant recipients. *Am J Transplant* **4**(4), 611-20.
- Pignatelli, S., Dal Monte, P., Rossini, G., and Landini, M. P. (2004). Genetic polymorphisms among human cytomegalovirus (HCMV) wild-type strains. *Rev Med Virol* **14**(6), 383-410.
- Plotkin, S. A., Furukawa, T., Zygraich, N., and Huygelen, C. (1975). Candidate cytomegalovirus strain for human vaccination. *Infect Immun* **12**(3), 521-7.
- Prod'homme, V., Retiere, C., Imbert-Marcille, B. M., Bonneville, M., and Hallet, M. M. (2003a). Modulation of HLA-A*0201-restricted T cell responses by natural polymorphism in the IE1(315-324) epitope of human cytomegalovirus. *J Immunol* **170**(4), 2030-6.
- Prod'homme, V., Retiere, C., Valtcheva, R., Bonneville, M., and Hallet, M. M. (2003b). Cross-reactivity of HLA-B*1801-restricted T-lymphocyte clones with target cells

- expressing variants of the human cytomegalovirus 72kDa-IE1 protein. *J Virol* **77**(12), 7139-42.
- Provenzano, M., Lim, J. B., Mocellin, S., Monsurro, V., Bettinotti, M., Marincola, F. M., and Stroncek, D. F. (2003). The matrix protein pp65(341-350): a peptide that induces ex vivo stimulation and in vitro expansion of CMV-specific CD8⁺ T cells in subjects bearing either HLA-A*2402 or A*0101 allele. *Transfusion* **43**(11), 1567-74.
- Puchhammer-Stockl, E., Gorzer, I., Zoufaly, A., Jaksch, P., Bauer, C. C., Klepetko, W., and Popow-Kraupp, T. (2006). Emergence of multiple cytomegalovirus strains in blood and lung of lung transplant recipients. *Transplantation* **81**(2), 187-94.
- Rasmussen, L., Geissler, A., Cowan, C., Chase, A., and Winters, M. (2002). The genes encoding the gCIII complex of human cytomegalovirus exist in highly diverse combinations in clinical isolates. *J Virol* **76**(21), 10841-8.
- Rasmussen, L., Geissler, A., and Winters, M. (2003). Inter- and intragenic variations complicate the molecular epidemiology of human cytomegalovirus. *J Infect Dis* **187**(5), 809-19.
- Reeves, M. B., Davies, A. A., McSharry, B. P., Wilkinson, G. W., and Sinclair, J. H. (2007). Complex I binding by a virally encoded RNA regulates mitochondria-induced cell death. *Science* **316**(5829), 1345-8.
- Retiere, C., Prod'homme, V., Imbert-Marcille, B. M., Bonneville, M., Vie, H., and Hallet, M. M. (2000). Generation of cytomegalovirus-specific human T-lymphocyte clones by using autologous B-lymphoblastoid cells with stable expression of pp65 or IE1 proteins: a tool to study the fine specificity of the antiviral response. *J Virol* **74**(9), 3948-52.
- Ribbert (1904). Ueber protozoenartige Zellen in der Niere eines syphilitischen Neugeborenen und in der Parotis von Kindern. *Zbl All Pathol* **15**:945-948.
- Rice, P., Longden, I., and Bleasby, A. (2000). EMBOS: the European Molecular Biology Open Software Suite. *Trends Genet* **16**(6), 276-7.
- Roby, C., and Gibson, W. (1986). Characterization of phosphoproteins and protein kinase activity of virions, noninfectious enveloped particles, and dense bodies of human cytomegalovirus. *J Virol* **59**(3), 714-27.
- Rowe, W. P., Hartley, J. W., Waterman, S., Turner, H. C., and Huebner, R. J. (1956). Cytopathogenic agent resembling human salivary gland virus recovered from tissue cultures of human adenoids. *Proc Soc Exp Biol Med* **92**(2), 418-24.
- Rowshani, A. T., Bemelman, F. J., van Leeuwen, E. M., van Lier, R. A., and ten Berge, I. J. (2005). Clinical and immunologic aspects of cytomegalovirus infection in solid organ transplant recipients. *Transplantation* **79**(4), 381-6.
- Sanchez, V., Greis, K. D., Sztul, E., and Britt, W. J. (2000). Accumulation of virion tegument and envelope proteins in a stable cytoplasmic compartment during human cytomegalovirus replication: characterization of a potential site of virus assembly. *J Virol* **74**(2), 975-86.
- Schmolke, S., Kern, H. F., Drescher, P., Jahn, G., and Plachter, B. (1995). The dominant phosphoprotein pp65 (UL83) of human cytomegalovirus is dispensable for growth in cell culture. *J Virol* **69**(10), 5959-68.
- Slezak, S. L., Bettinotti, M., Selleri, S., Adams, S., Marincola, F. M., and Stroncek, D. F. (2007). CMV pp65 and IE-1 T cell epitopes recognized by healthy subjects. *J Transl Med* **5**, 17.
- Slobedman, B., Mocarski, E. S., Arvin, A. M., Mellins, E. D., and Abendroth, A. (2002). Latent cytomegalovirus down-regulates major histocompatibility complex class II expression on myeloid progenitors. *Blood* **100**(8), 2867-73.

- Smith, M. G. (1956). Propagation in tissue cultures of a cytopathogenic virus from human salivary gland virus (SGV) disease. *Proc Soc Exp Biol Med* **92**(2), 424-30.
- Staden, R. (1994). Staden: introduction. *Methods Mol Biol* **25**, 9-26.
- Stanton, R., Westmoreland, D., Fox, J. D., Davison, A. J., and Wilkinson, G. W. (2005). Stability of human cytomegalovirus genotypes in persistently infected renal transplant recipients. *J Med Virol* **75**(1), 42-6.
- Strong, R. K., Holmes, M. A., Li, P., Braun, L., Lee, N., and Geraghty, D. E. (2003). HLA-E allelic variants. Correlating differential expression, peptide affinities, crystal structures, and thermal stabilities. *J Biol Chem* **278**(7), 5082-90.
- Sylwester, A. W., Mitchell, B. L., Edgar, J. B., Taormina, C., Pelte, C., Ruchti, F., Sleath, P. R., Grabstein, K. H., Hosken, N. A., Kern, F., Nelson, J. A., and Picker, L. J. (2005). Broadly targeted human cytomegalovirus-specific CD4⁺ and CD8⁺ T cells dominate the memory compartments of exposed subjects. *J Exp Med* **202**(5), 673-85.
- Tamouza, R., Rocha, V., Busson, M., Fortier, C., El Sherbini, S. M., Esperou, H., Filion, A., Socie, G., Dulphy, N., Krishnamoorthy, R., Toubert, A., Gluckman, E., and Charron, D. (2005). Association of HLA-E polymorphism with severe bacterial infection and early transplant-related mortality in matched unrelated bone marrow transplantation. *Transplantation* **80**(1), 140-4.
- Tarrago, D., Quereda, C., and Tenorio, A. (2003). Different cytomegalovirus glycoprotein B genotype distribution in serum and cerebrospinal fluid specimens determined by a novel multiplex nested PCR. *J Clin Microbiol* **41**(7), 2872-7.
- Thoma, C., Borst, E., Messerle, M., Rieger, M., Hwang, J. S., and Bogner, E. (2006). Identification of the interaction domain of the small terminase subunit pUL89 with the large subunit pUL56 of human cytomegalovirus. *Biochemistry* **45**(29), 8855-63.
- Tomasec, P., Braud, V. M., Rickards, C., Powell, M. B., McSharry, B. P., Gadola, S., Cerundolo, V., Borysiewicz, L. K., McMichael, A. J., and Wilkinson, G. W. (2000). Surface expression of HLA-E, an inhibitor of natural killer cells, enhanced by human cytomegalovirus gpUL40. *Science* **287**(5455), 1031.
- Trivedi, D., Williams, R. Y., O'Reilly, R. J., and Koehne, G. (2005). Generation of CMV-specific T lymphocytes using protein-spanning pools of pp65-derived overlapping pentadecapeptides for adoptive immunotherapy. *Blood* **105**(7), 2793-801.
- Ulbrecht, M., Martinozzi, S., Grzeschik, M., Hengel, H., Ellwart, J. W., Pla, M., and Weiss, E. H. (2000). Cutting edge: the human cytomegalovirus UL40 gene product contains a ligand for HLA-E and prevents NK cell-mediated lysis. *J Immunol* **164**(10), 5019-22.
- Varnum, S. M., Streblow, D. N., Monroe, M. E., Smith, P., Auberry, K. J., Pasa-Tolic, L., Wang, D., Camp, D. G., 2nd, Rodland, K., Wiley, S., Britt, W., Shenk, T., Smith, R. D., and Nelson, J. A. (2004). Identification of proteins in human cytomegalovirus (HCMV) particles: the HCMV proteome. *J Virol* **78**(20), 10960-6.
- Vauloup-Fellous, C., Picone, O., Cordier, A. G., Parent-du-Chatelet, I., Senat, M. V., Frydman, R., and Grangeot-Keros, L. (2009). Does hygiene counseling have an impact on the rate of CMV primary infection during pregnancy? Results of a 3-year prospective study in a French hospital. *J Clin Virol* **46 Suppl 4**, S49-53.
- Verbraak, F. D., Bruinenberg, M., van den Horn, G. J., Meenken, C., van der Lelij, A., Hoyng, C. B., Kijlstra, A., and Peek, R. (1998). Cytomegalovirus (CMV) strain differences between the eye and blood in AIDS patients with CMV retinitis. *Aids* **12**(7), 713-8.
- Vossen, M. T., Westerhout, E. M., Soderberg-Naucler, C., and Wiertz, E. J. (2002). Viral immune evasion: a masterpiece of evolution. *Immunogenetics* **54**(8), 527-42.
- Walling, D. M., Perkins, A. G., Webster-Cyriaque, J., Resnick, L., and Raab-Traub, N. (1994). The Epstein-Barr virus EBNA-2 gene in oral hairy leukoplakia: strain

- variation, genetic recombination, and transcriptional expression. *J Virol* **68**(12), 7918-26.
- Wang, E. C., McSharry, B., Retiere, C., Tomasec, P., Williams, S., Borysiewicz, L. K., Braud, V. M., and Wilkinson, G. W. (2002). UL40-mediated NK evasion during productive infection with human cytomegalovirus. *Proc Natl Acad Sci U S A* **99**(11), 7570-5.
- Wang, X., Huang, D. Y., Huong, S. M., and Huang, E. S. (2005). Integrin alphavbeta3 is a coreceptor for human cytomegalovirus. *Nat Med* **11**(5), 515-21.
- Wang, X., Kenyon, W. J., Li, Q., Mullberg, J., and Hutt-Fletcher, L. M. (1998). Epstein-Barr virus uses different complexes of glycoproteins gH and gL to infect B lymphocytes and epithelial cells. *J Virol* **72**(7), 5552-8.
- Weller, T. H., Hanshaw, J. B., and Scott, D. E. (1960). Serologic differentiation of viruses responsible for cytomegalic inclusion disease. *Virology* **12**, 130-2.
- Wilkinson, G. W., Tomasec, P., Stanton, R. J., Armstrong, M., Prod'homme, V., Aicheler, R., McSharry, B. P., Rickards, C. R., Cochrane, D., Llewellyn-Lacey, S., Wang, E. C., Griffin, C. A., and Davison, A. J. (2007). Modulation of natural killer cells by human cytomegalovirus. *J Clin Virol*.
- Wilkinson, G. W., Tomasec, P., Stanton, R. J., Armstrong, M., Prod'homme, V., Aicheler, R., McSharry, B. P., Rickards, C. R., Cochrane, D., Llewellyn-Lacey, S., Wang, E. C., Griffin, C. A., and Davison, A. J. (2008). Modulation of natural killer cells by human cytomegalovirus. *J Clin Virol* **41**(3), 206-12.
- Wolinsky, S. M., Korber, B. T., Neumann, A. U., Daniels, M., Kunstman, K. J., Whetsell, A. J., Furtado, M. R., Cao, Y., Ho, D. D., and Safrit, J. T. (1996). Adaptive evolution of human immunodeficiency virus-type 1 during the natural course of infection. *Science* **272**(5261), 537-42.
- Wreghitt, T. G., Teare, E. L., Sule, O., Devi, R., and Rice, P. (2003). Cytomegalovirus infection in immunocompetent patients. *Clin Infect Dis* **37**(12), 1603-6.
- Zhao, J., Sinclair, J., Houghton, J., Bolton, E., Bradley, A., and Lever, A. (2010). Cytomegalovirus beta2.7 RNA transcript protects endothelial cells against apoptosis during ischemia/reperfusion injury. *J Heart Lung Transplant* **29**(3), 342-5.
- Zong, J. C., Arav-Boger, R., Alcendor, D. J., and Hayward, G. S. (2007). Reflections on the interpretation of heterogeneity and strain differences based on very limited PCR sequence data from Kaposi's sarcoma-associated herpesvirus genomes. *J Clin Virol* **40**(1), 1-8.
- Zweyberg Wirgart, B., Brytting, M., Linde, A., Wahren, B., and Grillner, L. (1998). Sequence variation within three important cytomegalovirus gene regions in isolates from four different patient populations. *J Clin Microbiol* **36**(12), 3662-9.

Figure legends

Table 1:

Amplification of HCMV UL123 gene: sequence of sense and anti-sense primers, cycle conditions and additional primers for sequencing.

Table 2:

Variability of UL18, UL40, UL111a, UL123 and US3 genes in the twenty-one HCMV clinical strains.

Fig. 1:

Phylogenetic representations of HCMV nucleotidic variation, for the five studied genes individually and for the concatenated sequence. Samples were identified as follows: patient's number (1 to 17 for renal transplant or 101 to 106 for lung transplant)-Bx (Bordeaux, France)-year (03, 04, 05 or 06) and month of collection (01 to 12).

Fig. 1A, 1B, 1C, 1D and 1E illustrate the UL18, UL40, UL111a, UL123 and US3 gene sequences.

Fig. 1F illustrates HCMV variation in the concatenated sequence (4804 nucleotides).

Fig. 2:

Reticulate pattern representations of SplitsTree analysis for concatenated sequences, for UL40 and UL123 gene sequences and for UL40 cloned sequences. Phi test values are indicated for each analysis.

Fig. 2A illustrates SplitsTree analysis for the concatenated sequences of the five genes. Sequences from lung transplant recipients are indicated by dotted arrows, and reference strains underlined. The two strains presented in Fig. 3 are indicated by full arrows.

Fig. 2B and 2C illustrate SplitsTree analysis for UL40 and UL123 gene sequences.

Fig. 2D shows SplitsTree analysis for the concatenated sequences obtained from 12 HCMV positive samples totally devoid of nucleotidic double populations. The corresponding patients were 3 D+/R+, 1D-/R+, 8 D+/R-.

Fig. 2E: SplitsTree analysis of 14 1Bx UL40 cloned sequences: the phi test did find statistically significant evidence for recombination.

Fig. 2F: SplitsTree analysis of 22 9Bx UL40 cloned sequences: the phi test did not find any statistical evidence for recombination.

Fig. 3:

Examples of bootscan analysis illustrating either intragenic (106Bx0411, Fig. 3A) or intergenic (10Bx0502, Fig. 3B) probable recombination events in concatenated sequences. Bootscan indicates which sequence resembles mostly the target sequence, in which proportion (percentage shown in the y axis), all along the nucleotidic sequence. The rule at the bottom of the graphs indicates nucleotidic lengths of the studied HCMV genes.

Fig. 3A: 106Bx0411 (D+/R+) sequence is compared with 103Bx0607 (D+/R+), 7Bx0506 (D-/R+) and 14Bx0507 (D+/R-) sequences. The 5' and 3' termini of gene UL40 are similar in 106Bx0411 and 103Bx0607; UL40 central sequence is alike in 106Bx0411 and 7Bx0506. SimPlot software considers this bootscan pattern as highly suggestive of intra-genic recombination events. The recombination sites in UL40 are indicated by arrows at the top of the figure.

Fig. 3B: 10Bx0502 (D+/R-) sequence is compared with 8Bx312 (D+/R-), 12Bx0501 (D+/R-), 17Bx0508 (D+/R-) and 104Bx0604 (D+/R+) sequences. UL18 sequence is similar in 10Bx0502 and 104Bx0604. UL40, UL111a, UL123 and US3 sequences are closer in 10Bx0502 and 8Bx0312. SimPlot software considers this bootscan pattern as highly suggestive of recombination events which have taken place in HCMV genome sequences located between UL18 and UL40. The recombination site between UL18 and UL40 is indicated by an arrow at the top of the figure.

Fig. 3C: 15Bx0506 (D+/R-) sequence is compared with 9Bx0411 (D+/R-), 14Bx0411 (D+/R-) and 106Bx0411 (D+/R+) sequences. 15Bx0506 is similar all along the concatenated sequence with 106Bx0411 and Simplot does not find any indication of recombination event.

Table 1:

	Sense primer	Anti-sense primer	Cycle conditions
1 st PCR	5'- Agg CgT gAC ACg TTT ATT gAg TAg g -3'	5'- TAA CTC CYg TTg Cgg TgC TgT TAA C -3'	94°C 12 min 94°C 30 s - 60°C 30 s - 72°C 2 min : 50 cycles 72°C 7 min
Nested PCR	5'- AgT ATA CAA TAG TgA CgT ggg ATC C -3'	5'- TAG CTg ACA gAC TAA CAg ACT gTT C -3'	94°C 12 min 94°C 30 s - 60°C 30 s - 72°C 2 min : 40 cycles 72°C 7 min
Additional primers	5'- Tgg CCA AAg TgT Agg CTR CAA TAg C -3' 5'- AAg gTC TTT gCC CAg TAC ATT CTg g -3' 5'- Agg CTg AgT TCT Tgg TAA AgA ACT C -3'	5'- gAg ATg Tgg ATg gCT TgT ATT AAg g -3' 5'- ACA TAC AYA AgT AAA TCC RCg TCC C -3' 5'- AgT gAC CgA gga TTg CAA CgA gAA C -3'	

Table 2 :

	Sequence length (nt)	Nb (%) of variable sites	Nb (%) of variable sites		Maximum % strain to strain variability	
			missense	silent	nucleotides	amino-acids
UL18	1107	95 (8.6)	49 (52)	46 (48)	5.9	7.6
UL40	666	80 (12)	43 (54)	37 (46)	6.3	10
UL111a	687-702	67 (9.7)	36 (69) *	16 (31) *	3.4	3.9
UL123	1757-1762	208 (11.8)	56 (39) *	88 (61) *	6.3	6.6
US3	561	37 (6.6)	20 (54)	17 (46)	3.4	6.3

* in the coding sequence (spliced gene)

nt : nucleotides

Fig. 1:

FIG 1A :

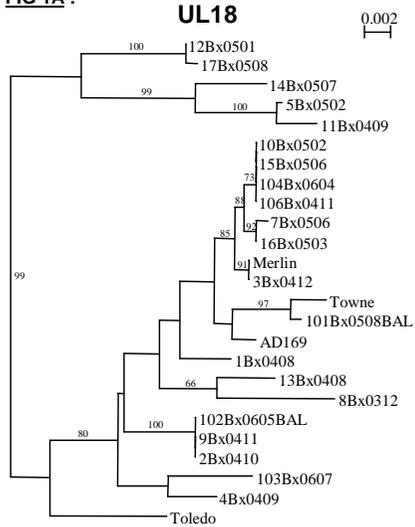


FIG 1D :

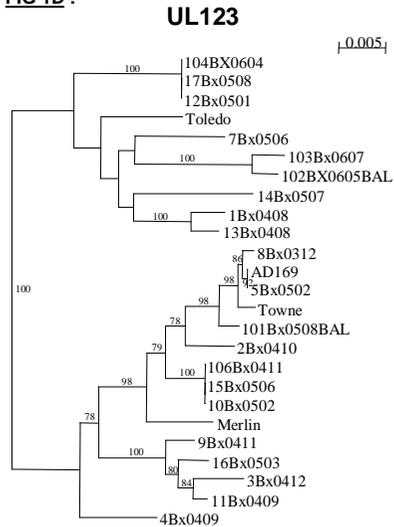


FIG 1B :

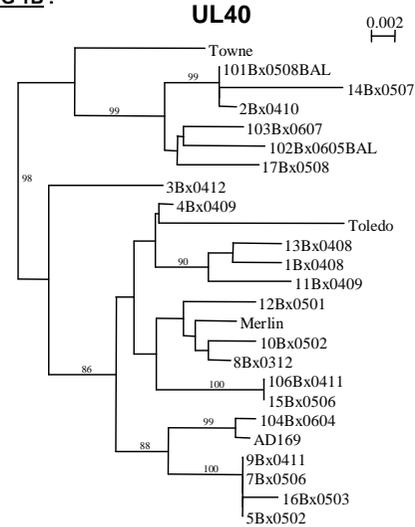


FIG 1E :

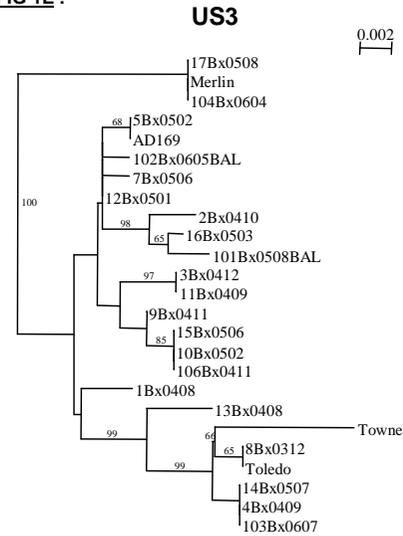


FIG 1C :

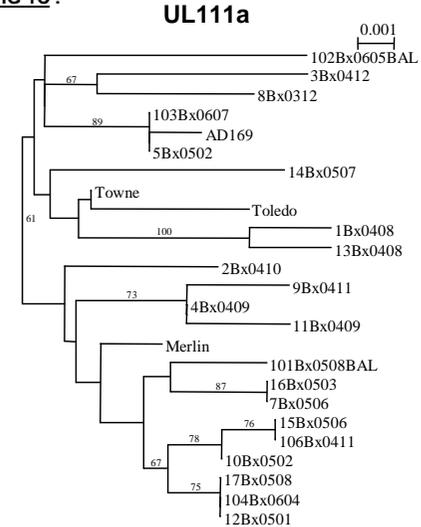


FIG 1F :

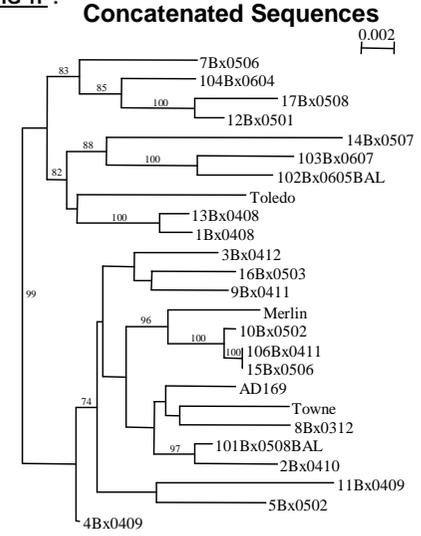


Fig. 2A:

Concatenated sequences

—|0.0010

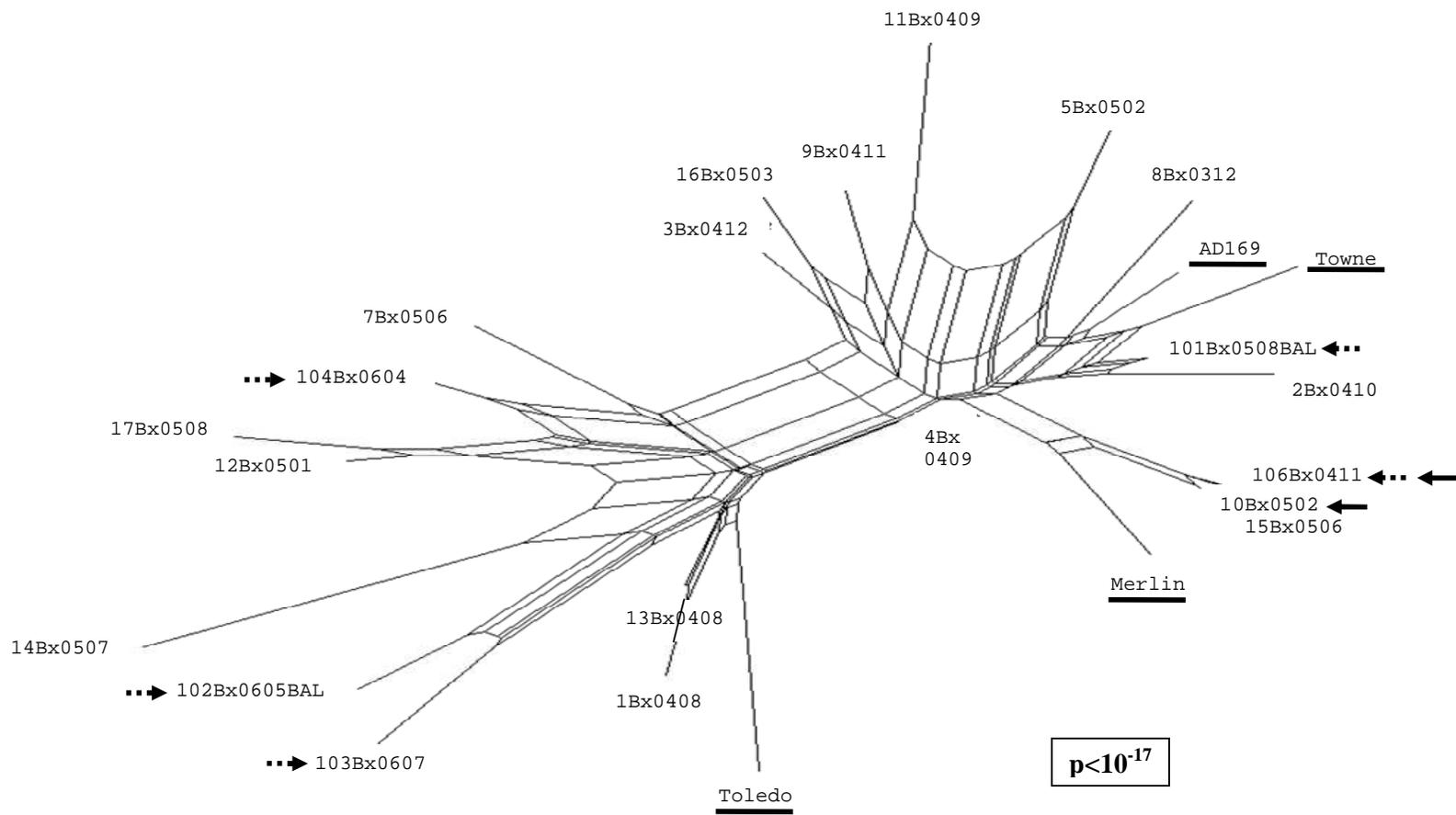


Fig. 2B:

UL40

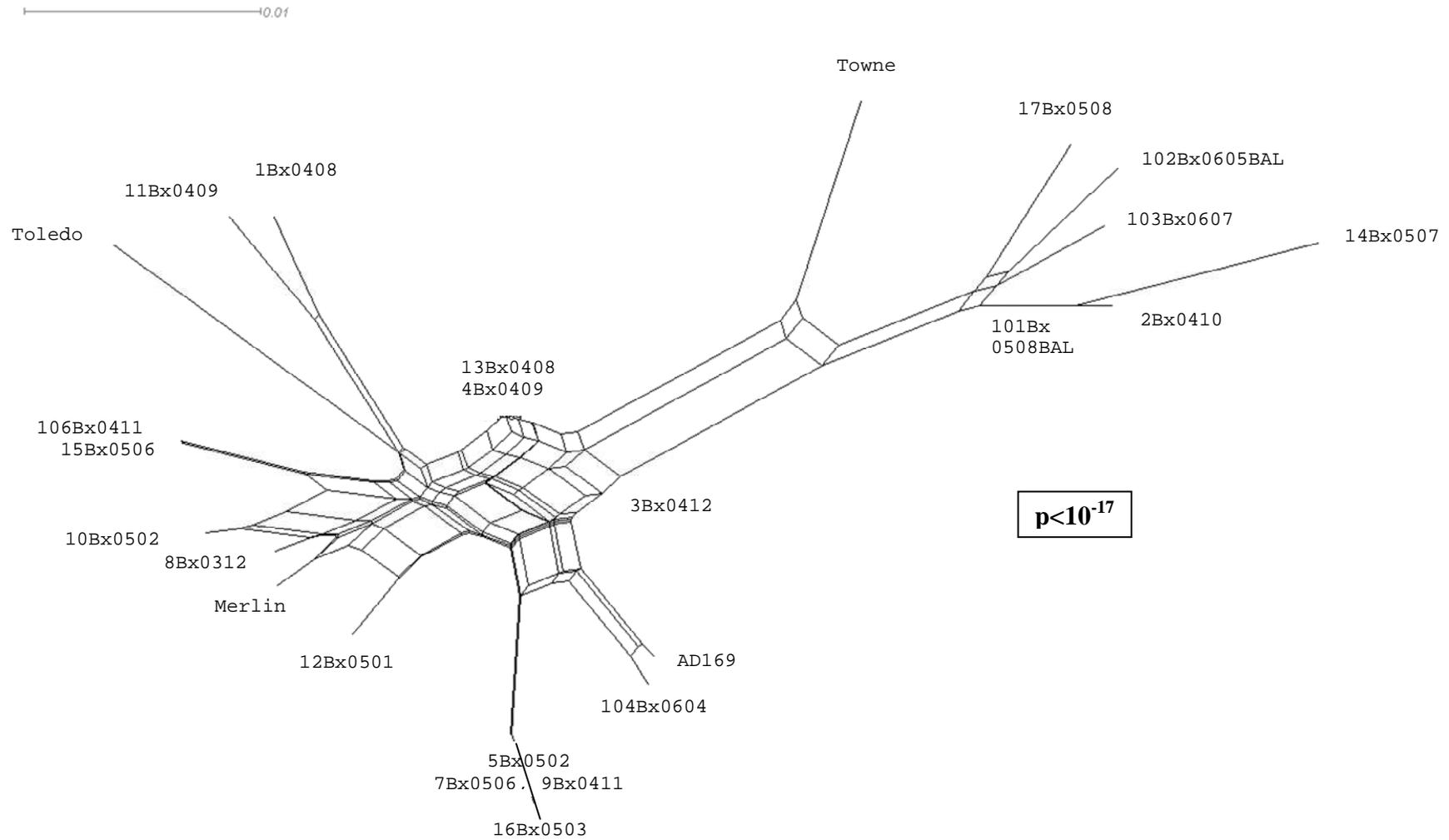


Fig. 2C:

1:10,0010

UL123

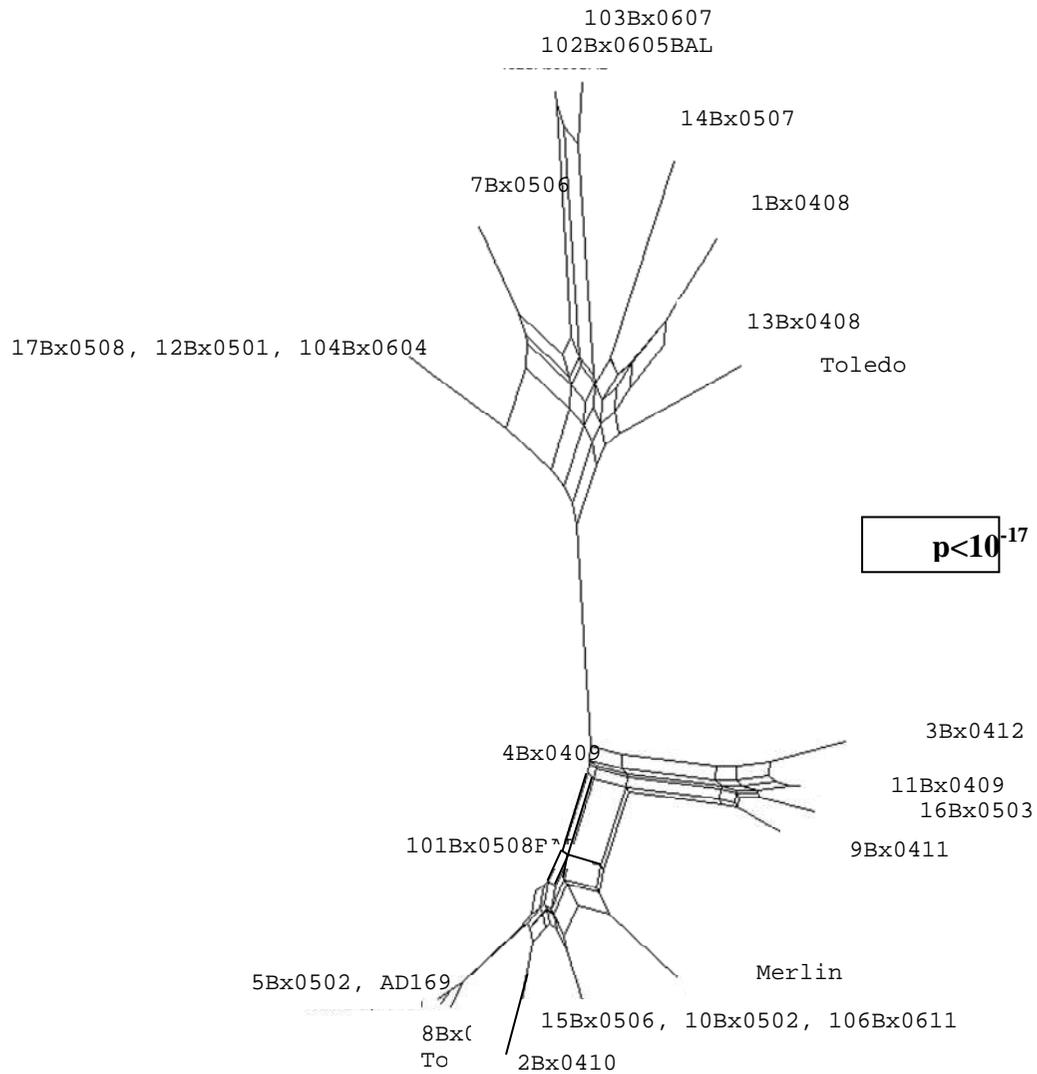


Fig. 2D:

Concatenated sequences free from double populations

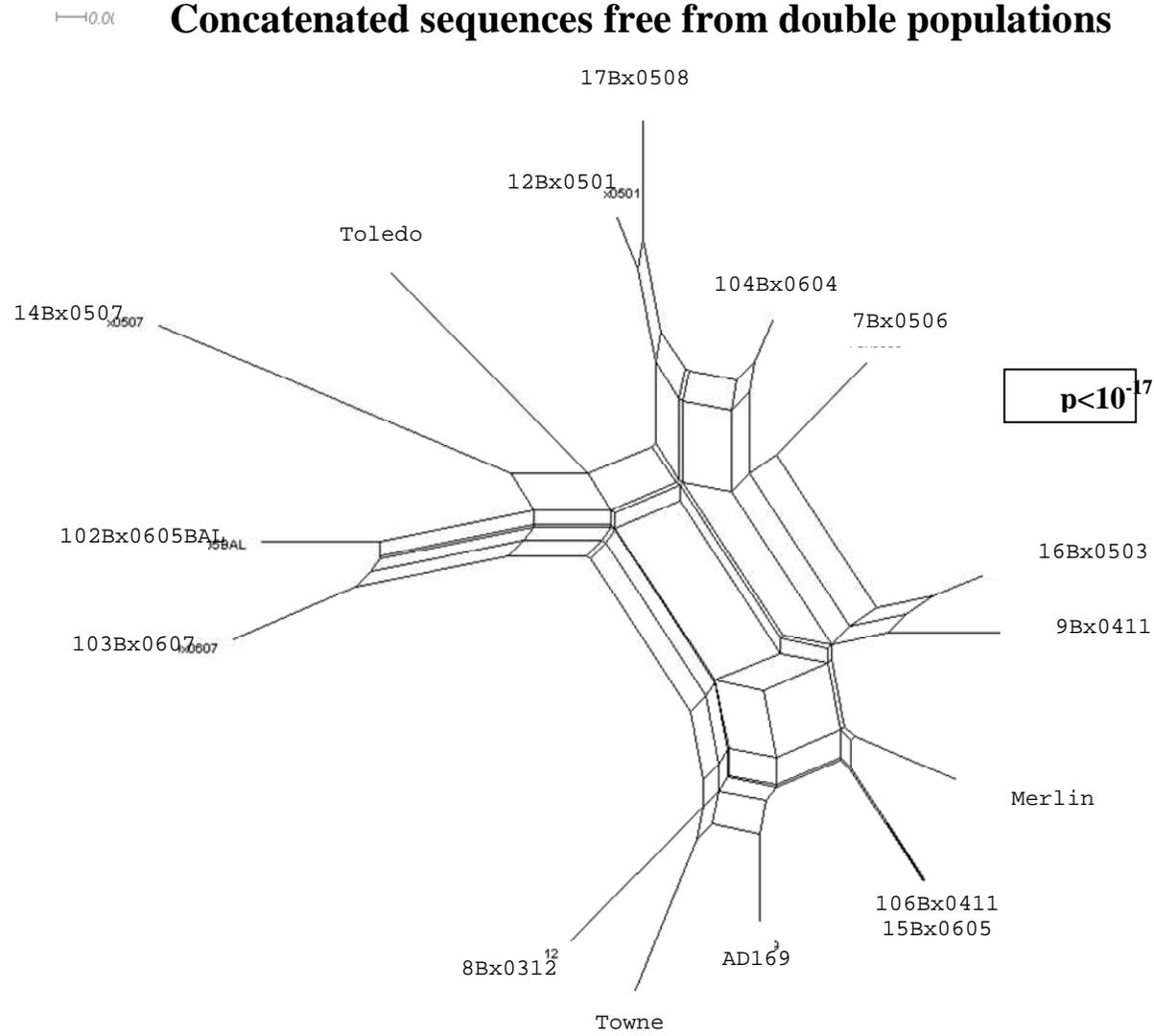


Fig. 2E:

Clones 1Bx UL40

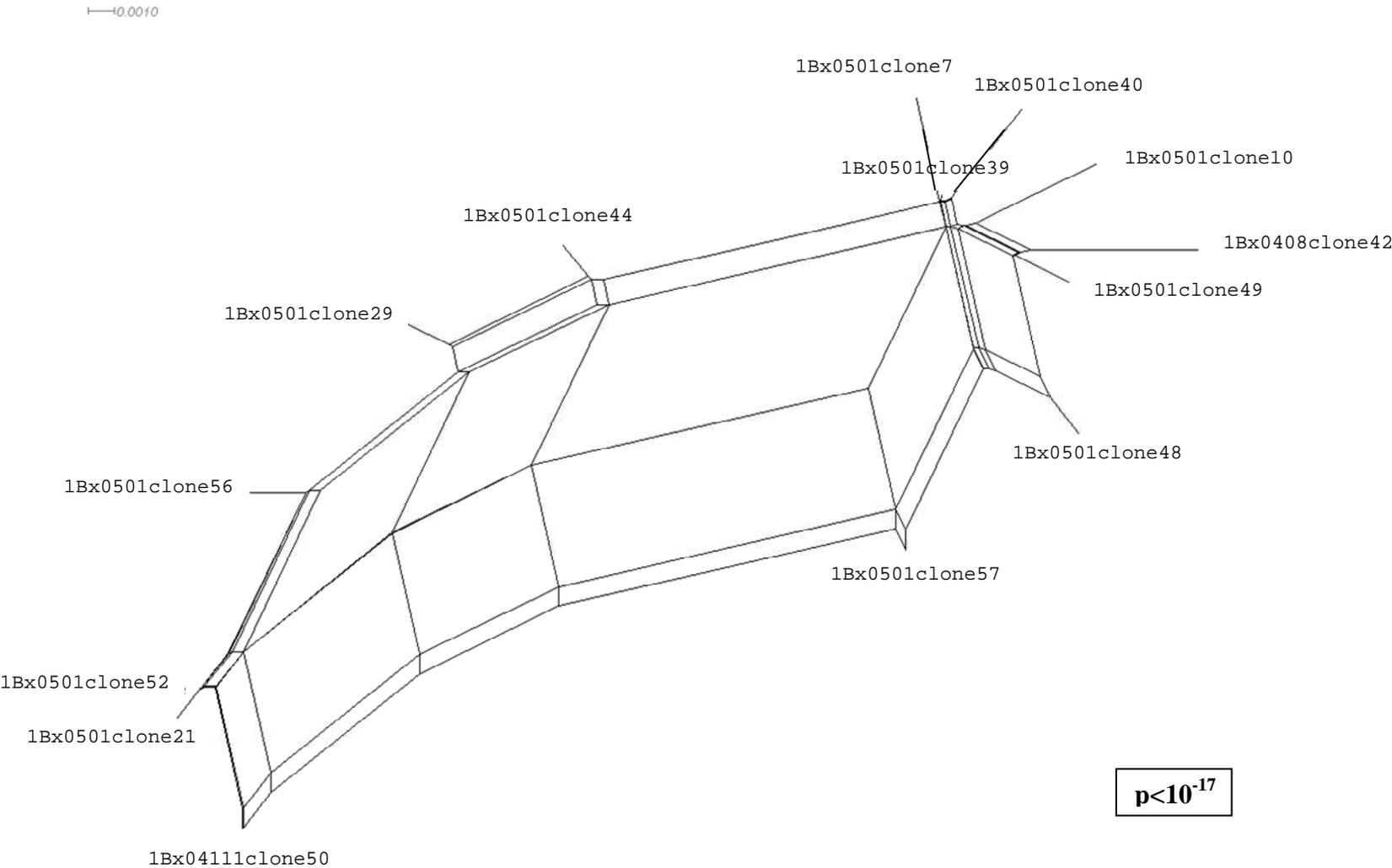


Fig. 2F:

Clones 9Bx UL40

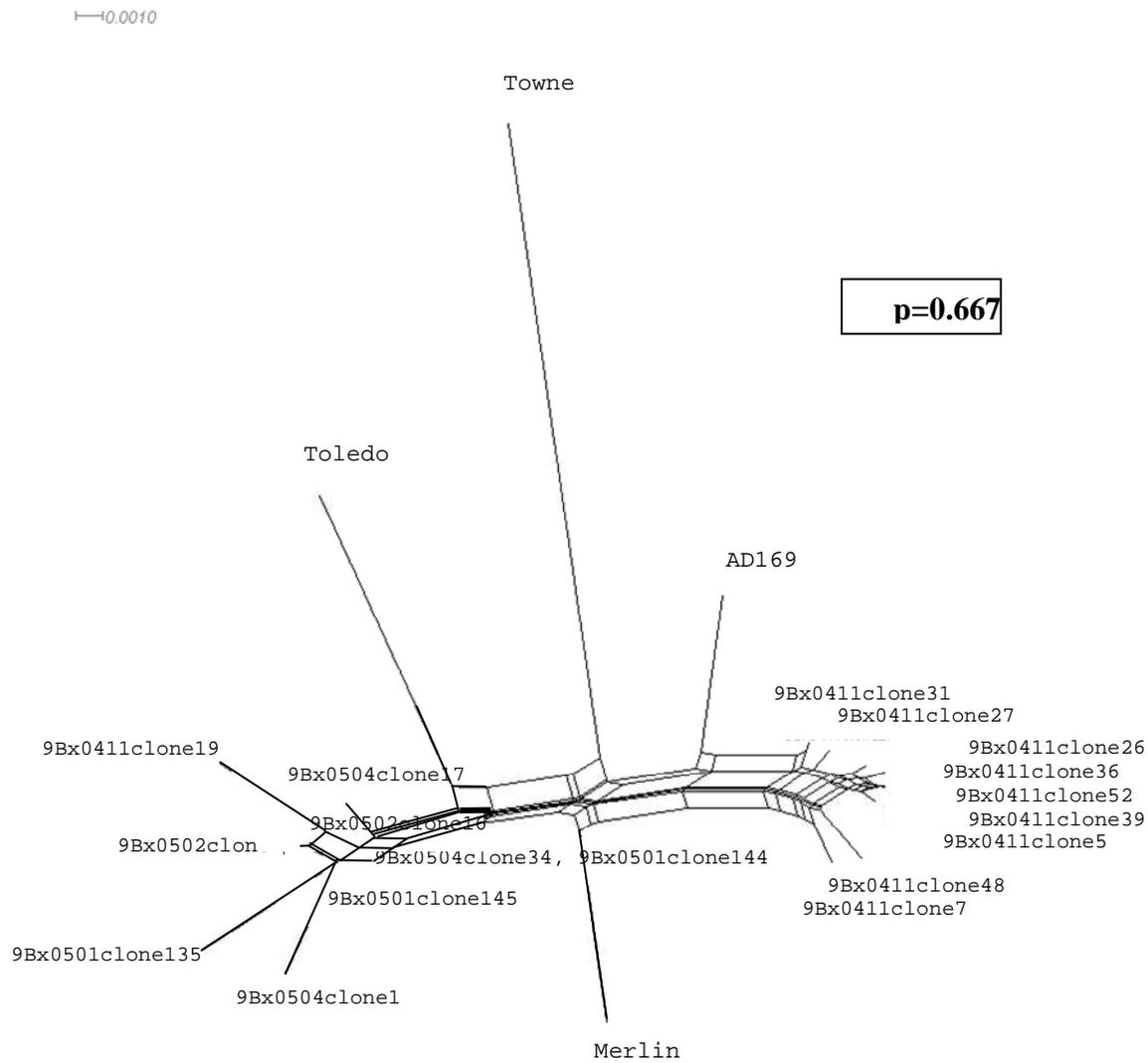


Fig. 3A:

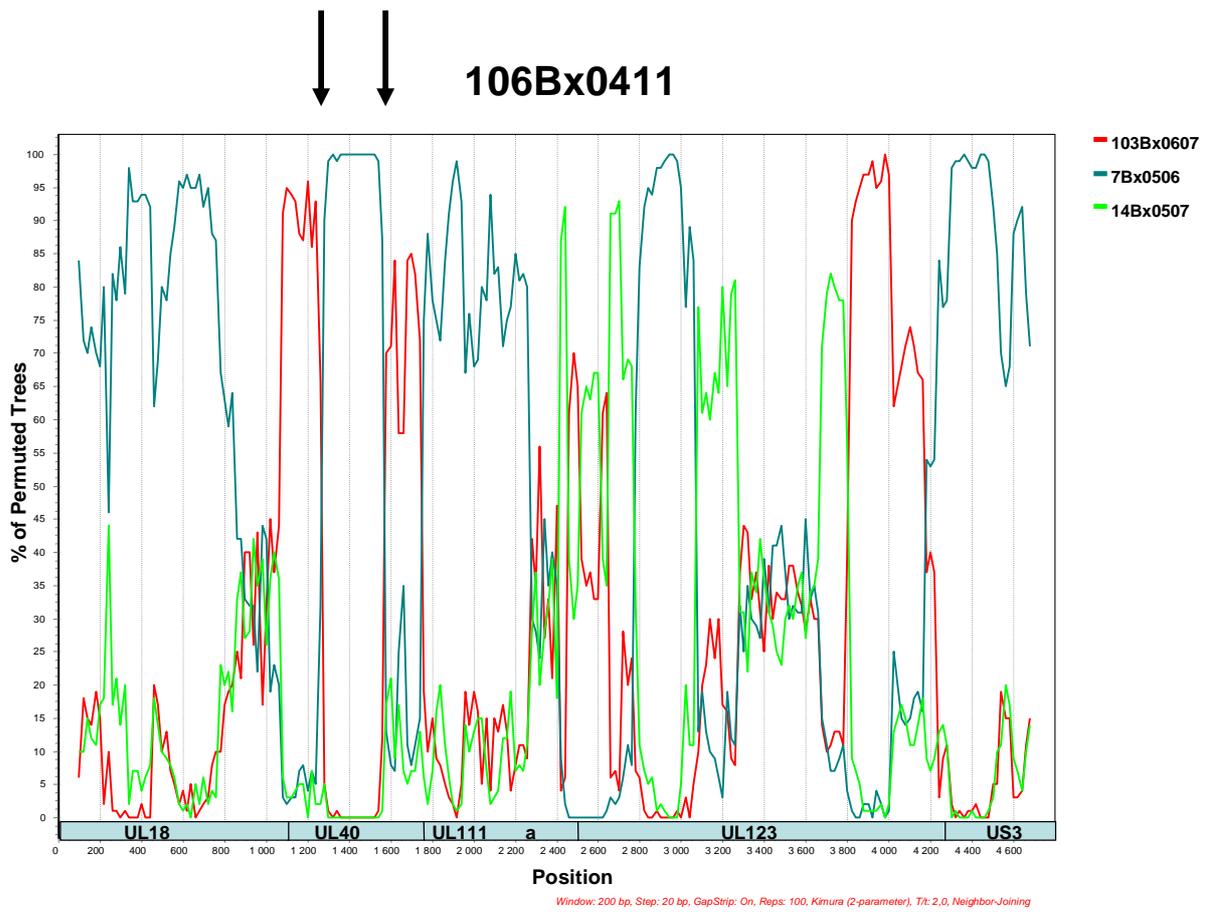


Fig. 3B:

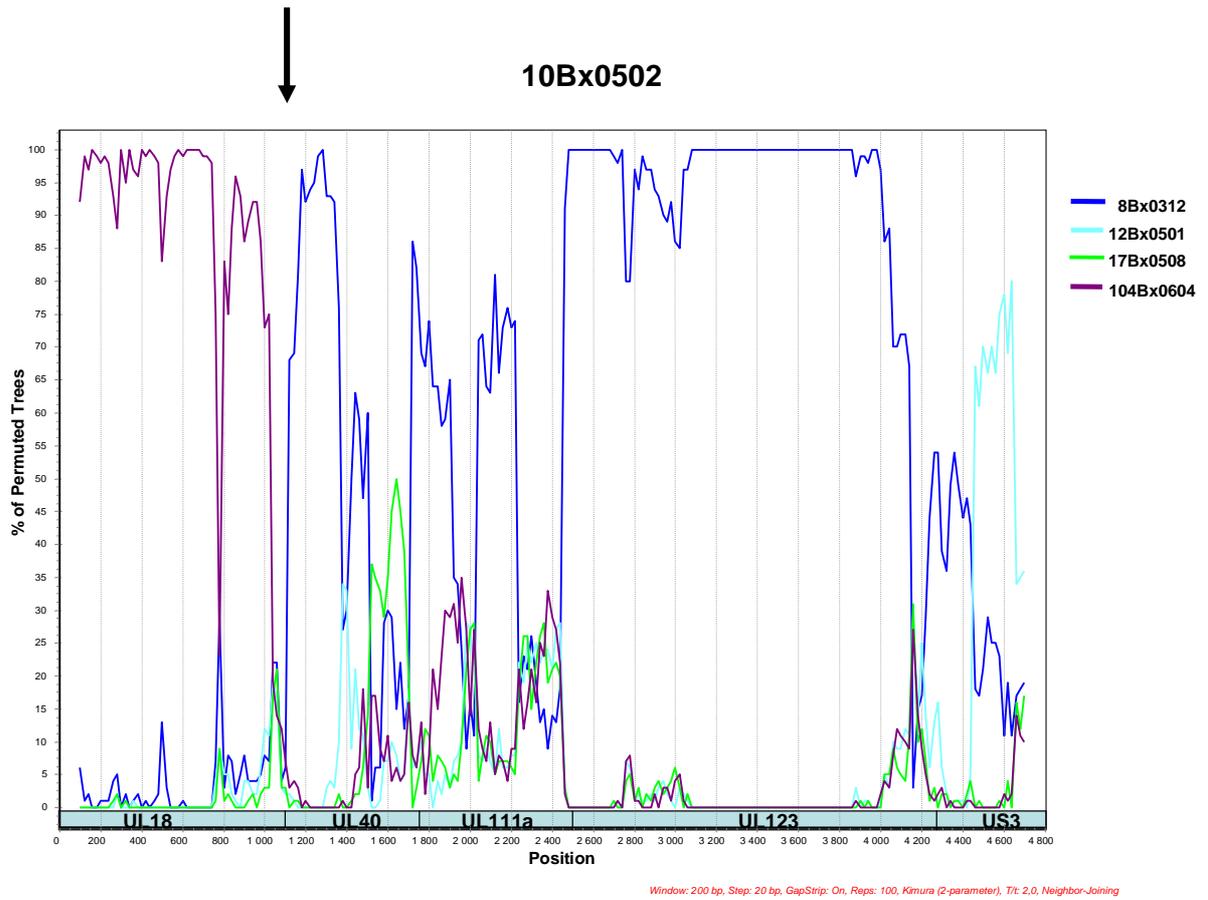


Fig. 3C:

