HAL
open science

# Études structurales de nouvelles cibles thérapeutiques du CMVH révélées par les dérivés benzimidazolés ribonucléosides.

Anthony Couvreux

## To cite this version:

Anthony Couvreux. Études structurales de nouvelles cibles thérapeutiques du CMVH révélées par les dérivés benzimidazolés ribonucléosides.. Biologie structurale [q-bio.BM]. Université Paris 5, 2009. Français. NNT : . tel-01122855

**HAL Id: tel-01122855**

**https://theses.hal.science/tel-01122855v1**

Submitted on 4 Mar 2015

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

*UNIVERSITE PARIS DESCARTES*

FACULTE DES SCIENCES PHARMACEUTIQUES ET BIOLOGIQUES

SCIENCES DE LA VIE ET DE LA SANTE

ECOLE DOCTORALE DU MEDICAMENT

**Année 2009**

**N°2009PA05P631**

**DOCTORAT DE L'UNIVERSITE PARIS DESCARTES**

COUVREUX Anthony

# Études structurales de nouvelles cibles thérapeutiques du CMVH révélées par les dérivés benzimidazolés ribonucléosides

Thèse dirigée par Serge BOUAZIZ

soutenue publiquement le 13 Novembre 2009

Membres du jury:

| | |
|---|---|
| Mme. Catherine ETCHEBEST, | Rapporteur |
| M. Christian ROUMESTAND, | Rapporteur |
| Mme. Sophie ALAIN, | Examinateur |
| M. Frédéric DARDEL, | Président |
| M. Patrice MORAND, | Examinateur |
| M. Daniel SCHERMAN, | Examinateur |
| M. Serge BOUAZIZ, | Directeur de thèse |

Titre de la thèse: **Etudes structurales de nouvelles cibles thérapeutiques du CMVH révélées par les dérivés benzimidazolés ribonucléosides**

Résumé :

**Le Cytomégalovirus Humain est un virus touchant toutes les populations quelque soit leur milieu socioprofessionnel. Inoffensif chez des sujets sains, les dégâts causés par le virus chez l'hôte se révèlent désastreux dès que celui-ci devient immunodéprimé. Ainsi, les porteurs de VIH, les personnes ayant subi une greffe d'organe ou les nourrissons sont particulièrement exposés à CMVH. Les premiers inhibiteurs formulés visant la polymérase virale se sont avérés inefficaces suite à l'observation de multiples mutations de résistance. Une nouvelle classe de molécule, les dérivés benzimidazolés ribonucléosides, visant les étapes de maturation et d'encapsidation, a donc émergé visant les protéines du complexe terminase, pUL89 et pUL56, et la phosphotransférase pUL97 ainsi que sa protéine accessoire, pUL27.**

**Grâce à l'utilisation de la résonance magnétique nucléaire, des spectroscopies UV/Visible et de dichroïsme circulaire, et de la modélisation par homologie, nous avons collecté des informations structurales relatives aux cibles des dérivés ribobenzimidazolés. Nous avons pu proposer des hypothèses sur le mode d'action de ces molécules. Par ailleurs, la découverte d'un domaine capable d'interagir avec l'ADN viral double brin de CMVH a été décrite.**

---

Title: **Structural studies of new therapeutic targets of HCMV revealed by benzimidazoles derivatives ribonucleoside**

Summary :

**The Human Cytomegalovirus is a virus affecting all populations regardless of their socio-professional environment. Harmless in healthy subjects, the damage caused by the virus in the host proved disastrous when it becomes immunocompromised. Thus, people with HIV, people who have undergone organ transplants or infants are particularly susceptible to HCMV. The first inhibitors formulated for the viral polymerase have been ineffective due to the observation of multiple resistance mutations. A new class of molecule, benzimidazoles ribonucleoside derivatives, for stages of maturation and encapsidation, has emerged for the terminase complex proteins, pUL89 and pUL56, and the phosphotransferase pUL97 and its accessory protein, pUL27. Through the use of nuclear magnetic resonance, UV / Visible spectroscopy and circular dichroism, and homology modelling, we collected structural information on target of derivatives ribobenzimidazoles. We have proposed hypotheses about the mode of action of these molecules. Moreover, the discovery of a domain capable of interacting with the viral double-stranded DNA of HCMV has been described.**

---

**Discipline: Biologie structurale**

---

**Mots-clefs: Cytomégalovirus Humain, terminase, pUL89, pUL56, dérivés ribobenzimidazolés, RMN, modélisation moléculaire**

---

**Adresse de l'auteur:**

**Adresse électronique: anthony.couvreux@wanadoo.fr**

*A mes grand-parents,*

*A Bernard, mon beau-père,*

*Remerciements*

En premier lieu, je remercie les membres du jury de me faire l'honneur d'évaluer mon travail de thèse.

Je tiens à remercier le Professeur Daniel Scherman pour m'avoir accueilli dans l'Unité de Pharmacologie Chimique et Génétique durant ces quatre ans et demi passés au sein de son laboratoire.

J'exprime ma reconnaissance envers le personnel administratif (Evelyne Lucas, Gounasségarin Virapin et Christine Dupuis) pour leur bonne humeur et leur disponibilité. En particulier, un grand merci à Christine pour avoir réussi, avec l'aval du Professeur Scherman, à m'obtenir des financements lors de mes derniers mois de thèse.

Je salue toute l'équipe de chercheurs, techniciens, thésards, stagiaires des équipes de biologie et physicochimie les discussions enrichissantes stimulées par des problèmes techniques.

Je pense également à Dina, Lisa et Marie pour les moments de détente passés ensemble

Je n'oublie pas les différentes personnes qui ont contribué aux investigations entreprises pour mener à bien l'avancée de mes recherches. Je suis reconnaissant envers Sophie Alain, notre collaboratrice, du Centre National Français de Référence du Cytomégalovirus Humain (Limoges) qui nous a proposé cette thématique constituant au départ mon sujet de stage de M2 et qui a finalement débouché sur cette aventure de quatre années supplémentaires.

Je ne peux dissocier ces remerciements de ceux dus à ses étudiants de thèse, Gaël Champier et Sébastien Hantz, pour tous les échanges fructueux et les (très) longues conversations téléphoniques passées à la révision des manuscrits des publications et à la méticuleuse dissection des alignements de séquences.

Je formule aussi mes remerciements aux différents stagiaires pris en charge par Sophie qui se sont succédés tout au long de ma thèse. Peu de leurs visages

me sont connus mais je n'oublie pas pour autant toute l'énergie qu'ils ont développée pour tenter d'exprimer, non sans mal, les peptides et protéines ayant fait l'objet de ces études.

Je pense aussi à Marie Claude Menet de l'unité d'Analyse Physicochimique de Paris 5 pour les analyses de spectrométrie de masse qui se sont déroulées dans une atmosphère studieuse mais néanmoins conviviale.

Je salue l'équipe de chimistes, Armand Blommaert, Cyrille Richard et en particulier Hélène Dhotel, qui a contribué de manière significative à la synthèse des inhibiteurs, ainsi que tous les étudiants qui se sont succédés (Thomas, Ronan, Rémi et Nicolas) dans ce bureau où j'ai passé un nombre incalculable d'heures à parler et plaisanter sur toutes sortes de sujets.

Il est évident que je ne peux oublier cette formidable équipe de RMN qui a partagé au quotidien mon enthousiasme, quelquefois, mais qui a réussi aussi à me motiver lors des moments difficiles de remise en question succédant aux échecs qu'implique, j'imagine, la profession de chercheur. Je mesure la chance que j'ai eue de faire partie de cette équipe mêlant rigueur, honnêteté et humilité.

Pour les nommer: merci à Marie-José Penouilh et à Sarah Boubigot ; merci à Gilmar Salgado qui m'a encadré pour les expériences de Biochimie et qui a réussi à instaurer, dès son arrivé au labo, le rituel des « conférences » autour d'une bière; merci à Pascale Coric et Serge 'Tovaritch' Turcaud pour leur sympathie et leur sens de l'humour. Merci à mon 'poto' Rodrigue Marquant, pour toute les heures de réflexion auxquelles s'est parfois joint Nicolas Inguimbert de la Faculté des Saints-Pères, de synthèses, de purifications passées sur les peptides analysés en RMN ; merci à Nelly Morellet pour son rôle d'excellent Second, toujours disponible, pour la RMN et la modélisation moléculaire ; et enfin merci à mon directeur de thèse Serge Bouaziz pour sa franchise, sa disponibilité et surtout sa patience.

Finalement, je remercie profondément mes amis qui ont dû composer avec mes sautes d'humeur. Evidemment une mention spéciale à mes chers parents et frères qui m'ont soutenu, malgré la distance, pendant cette période 'parisienne' de ma vie.

Enfin, le plus grand mérite revient à ma compagne, Gaëlle, qui m'a soutenu et a partagé au quotidien toutes les préoccupations que cette expérience passionnante et enrichissante a généré dans mon esprit.

## Abréviations

| | |
|---|---|
| ACV | Acidofovir |
| ALPHA | α-herpesvirus |
| BDCRB | 2-bromo-5,6-dichloro-1-(bêta-D-ribofuranosyl)benzimidazole |
| BÊTA | β-herpesvirus |
| BTCRB | 2-bromo-4,5,6-trichloro-1-(2,3,5-tri-*O*-acetyl-β-d-ribofuranosyl) |
| CDV | Cidofovir |
| $CL_4RB$ | 2,4,5,6-tetrachloro-1-(2,3,5-tri-*O*-acetyl-β-d-ribofuranosyl) benzimidazole |
| CMVH | Cytomégalovirus Humain |
| CSI | Chemical Shift Index |
| DC | Dichroïsme circulaire |
| EBV | Epstein-Barr Virus |
| ET | Evolutionnary Trace |
| GAMMA | γ-herpesvirus |
| GCV | Ganciclovir |
| HSV | Herpes Simplex Virus |
| HHV | Human Herpesvirus |
| HSQC | Heteronuclear Single Quantum Coherence |
| MBV | Maribavir |
| NOE | Nuclear Overhauser Effect |
| NOESY | Nuclear Overhauser Effect SpectroscopY |
| PFA | Foscarnet |
| RCI | Random Coil Index |
| RMN | Résonance Magnétique Nucléaire |
| RMSD | Root Mean Square Deviation |
| TCRB | 2,5,6-Trichloro-1-(bêta-D-ribofuranosyl)benzimidazole |
| TOCSY | TOtal Correlation SpectroscopY |
| VGCV | Valganciclovir |
| VZV | Varicella Zooster Virus |

ABRÉVIATIONS.....
I. INTRODUCTION GÉNÉRALE.....
II. BUT DE L'ÉTUDE.....
III. HERPESVIRUS.....
A. CLASSIFICATION ET PHYLOGÉNIE DES HERPESVIRUS.....
B. STRUCTURE DES HERPESVIRUS.....
B.1. Les protéines de l'enveloppe.....
B.2. Les protéines du tégument.....
B.3. Les protéines de la capside.....
B.4. L'ADN viral.....
IV. CIBLES THÉRAPEUTIQUES ET ANTIVIRAUX DIRIGÉS CONTRE LE CYTOMÉGALOVIRUS HUMAIN.....
A. INHIBITION DE LA VOIE DE SYNTHÈSE DE L'ADN VIRAL.....
A.1. pUL54: polymérase de CMVH et sa protéine accessoire pUL44.....
A.2. Les traitements visant la synthèse de l'ADN viral et Formiversen.....
A.3. Les résistances aux inhibiteurs de l'ADN polymérase virale.....
B. LES AUTRES CIBLES THÉRAPEUTIQUES EN DÉVELOPPEMENT.....
B.1. Maturation et encapsidation.....
B.1.a. Le complexe Terminase: pUL89 et pUL56.....
B.1.b. La protéine portail: pUL104.....
B.1.c. Assemblage et translocation de l'ADN.....
B.2. Rôle central de pUL97.....
B.2.a. Généralités sur les phosphotransférases.....
B.2.b. La phosphotransférase pUL97 et sa protéine accessoire pUL27.....
B.2.c. Les différents substrats de pUL97.....
B.3. Nouveaux inhibiteurs de CMVH.....
B.3.a. Les benzimidazoles -D et L- ribonucléosides.....
B.3.b. BAY 38-4766 et ses dérivés.....
B.3.c. Les indolocarbazoles.....
V. PRÉDICTION DES INTERFACES PROTÉINE-PROTÉINE.....
A. REPLIEMENT MOLÉCULAIRE.....
B. PROCESSUS DE L'ÉVOLUTION MOLÉCULAIRE.....
C. LA RECHERCHE DES DOMAINES D'INTERACTIONS.....
C.1. Alignements de séquences.....
C.2. Trace de l'évolution.....
C.3. Analyses des Hot spots.....
VI. OUTILS THÉORIQUES ET EXPÉRIMENTAUX.....
A. CHOIX DES TECHNIQUES UTILISÉES POUR LA RÉSOLUTION DE LA PROBLÉMATIQUE.....
B. DÉMARCHE BIOINFORMATIQUE POUR L'ANALYSE DE LA SÉQUENCE PRIMAIRE.....
C. SYNTHÈSE PEPTIDIQUE.....
D. RÉSONANCE MAGNÉTIQUE NUCLÉAIRE, ATTRIBUTION ET MODÉLISATION SOUS CONTRAINTES.....
E. MODÉLISATION THÉORIQUE DE PROTÉINES.....
VII. ETUDE DE LA PROTÉINE PUL89.....
A. DOMAINE N-TERMINAL DE PUL89.....

*B. Domaine hélicase de pUL89*.....
*B.1. Résidus critiques pour la fonction hélicase déduits de l'analyse de séquences*.....
*B.2. Modélisation du domaine hélicase*.....
B.2.a. Obtention du modèle par PHYRE.....
B.2.b. Amélioration du modèle.....
B.2.c. Fixation de l'ATP dans le site actif.....
B.2.d. Surface d'interaction avec l'ADN.....
B.2.e. Mutants aux dérivés D-ribobenzimidazolés.....
*B.3. Bilan concernant le domaine hélicase de pUL89*.....
*C. Domaine d'interaction entre pUL89 et pUL56: implication dans l'activité endonucléasique du complexe terminase*.....
*C.1. Étude du peptide pUL89(580-600)*.....
C.1.a. Dichroïsme circulaire.....
C.1.b. Résonance Magnétique Nucléaire.....
*C.2. Étude du peptide pUL89(568-635)*.....
C.2.a. Dichroïsme circulaire.....
C.2.b. RMN de pUL89-XI.....
*C.3. Domaine endonucléasique de pUL89: pUL89-endo*.....
C.3.a. Modélisation de pUL89-endo.....
C.3.b. Qualités des structures théoriques.....
C.3.c. Analyse de pUL89-endo.....
*C.4. Bilan de l'étude du domaine d'interaction avec pUL56*.....
*D. Conclusion sur l'étude de pUL89*.....
**VIII. ÉTUDES DE LA PROTÉINE PUL56**.....
*A. Alignement de séquences*.....
*B. Doigt de zinc de pUL56 de HCMV*.....
*C. Étude structurale de pUL56(507-548): bZIP domaine*.....
*C.1. Dichroïsme circulaire*.....
*C.2. Cross-linking*.....
*C.3. pUL56(506-549): un peptide de type bZIP ?*.....
*C.4. Étude de la dimérisation de pUL56(506-548) par HPLC et RMN*.....
*D. Bilan de l'étude des domaines de pUL56*.....
**IX. CONCLUSION GÉNÉRALE ET PERSPECTIVES**.....
**X. RÉFÉRENCES**.....
**XI. ANNEXES**.....
*A. Annexe A : Analyse théorique des cibles du Maribavir*.....
*A.1. Annexe A1: Etude préliminaire de la phosphotransférase pUL97*.....
A.1.a. Obtention d'un modèle théorique de pUL97.....
A.1.b. Analyse de la structure pUL97Δ (590-607)_2nruA.....
A.1.c. Substrat naturel et de synthèse à pUL97 et perspectives.....
A.1.d. Bilan de l'étude de pUL97.....
*A.2. Annexe A2: Protéine accessoire pUL27*.....
A.2.a. Multimère de pUL27 ?.....
A.2.b. Sites de phosphorylation.....
A.2.c. Mutations au Maribavir.....
A.2.d. Bilan de l'étude de pUL27.....
*B. ANNEXE B: Alignements de séquences*.....

*B.1. ANNEXE B1: Alignement de séquence de pUL89 et 17 de ses homologues herpesviraux.....*
*B.2. ANNEXE B2: Alignement de séquence de pUL56 et 22 de ses homologues herpesviraux.....*
*B.3. ANNEXE B3: Alignement de séquence de pUL89 et gp17 (bactériophage T4).....*
*B.4. ANNEXE B4: Alignement de séquence de pUL89(161-570) et du domaine hélicase de la famille des snf2/rad54 de Sulfolobus solfataricus (PDB 1Z63).....*
*B.5. ANNEXE B5: Sous domaines conservés à travers quelques phosphotransférases herpesvirales comparés avec celles de Cdk2 et PKA.....*
*B.6. ANNEXE B6: Alignement de séquences entre pUL27 et ses homologues issus de bêta herpesvirus.....*
*C. ANNEXE C: FICHIERS UTILISÉS POUR LE CALCUL DU MODÈLE DU SOUS-DOMAINE HÉLICASE DE PUL89.....*
*C.1. Annexe C1: Fichier de contraintes (.rstrnt).....*
*C.2. Annexe C2: Fichier entrée pour minimsation sous contraintes (.inp).....*
*D. ANNEXE D: MODÉLISATION SOUS CONTRAINTES RMN ET TABLES DE DÉPLACEMENTS CHIMIQUES .....*
*D.1. Annexe D1: Création du fichier de contraintes NOEs (.tbl).....*
*D.2. Annexe D2: pUL89(580-600).....*
*D.3. ANNEXE D3: RMN de pUL89-XI tables de déplacements chimiques.....*
*D.4. ANNEXE D4: pUL56(506-548).....*
*E. ANNEXE E : PUL89-ENDO PAR MODELLER.....*
*E.1. Annexe E1: Fichiers d'alignement de séquence pour le calcul des formes libre et lié au duplexe(.ali).....*
E.1.a. Forme libre.....
E.1.b. Forme liée.....
*E.2. Annexe E2: Fichier de contraintes (.py).....*

# I. INTRODUCTION GÉNÉRALE

Le cytomégalovirus humain est un virus de la famille des herpesvirus découvert en 1904 par Jesionek et Kielemenoglou après l'observation de cellules volumineuses à inclusions intranucléaires et cytoplasmiques dans les poumons, le foie et les reins d'enfants mort-nés (Jesionek and Kielemenoglou 1904). En 1921, Goodpasture et Talbot proposent le nom de cytomégalie pour désigner cette pathologie. Cinq ans plus tard, Cole et Kuttner proposent le terme de maladie à inclusions cytomégaliques (Cole and Kuttner 1926). Il a fallu attendre 1956 pour identifier l'agent pathogène dans des fibroblastes humains (Rowe, Hartley et al. 1956; Smith 1956). C'est finalement en 1960 que le terme cytomégalovirus est proposé par Weller et adopté par la communauté scientifique (Weller, Hanshaw et al. 1960).
Le cytomégalovirus humain (CMVH) est un pathogène majeur après transplantation et la première cause d'infections congénitales virales. Comme les autres formes d'herpesvirus, CMVH est sans conséquence pour l'organisme tant que celui-ci reste immunocompétent. Cependant, après des greffes d'organes ou de moelle osseuse, ou chez des sujets atteints par le VIH (80% des cas) celui-ci peut entraîner des complications considérables (pneumopathie, rétinite, hépatite, ulcération des muqueuses gastro-intestinales, conséquences neurologiques et perte de l'audition chez le nourrisson). Du fait de sa faible résistance à l'air, les vecteurs principaux de l'infection par CMVH sont principalement les secrétions génitales, le lait maternel, la salive et le sang. Ainsi, la transmission du virus est principalement favorisée par les rapports sexuels et l'allaitement.

Face à l'étendue de ce virus, des traitements dirigés contre l'ADN polymérase virale pUL54 ont été développés. Ces traitements, Ganciclovir et sa prodrogue Valganciclovir, le Cidofovir et le Foscarnet, ont généré un nombre important de résistances sur pUL54 et la phosphotransférase de CMVH, pUL97, après traitements isolés ou combinés sur des sujets infectés. D'autres traitements ont donc émergé: les dérivés D-ribobenzimidazolés visant les étapes de maturation et d'encapsidation assurées par les unités terminases pUL89 et pUL56 ainsi que la protéine portail pUL104, et leur analogues L-ribobenzimidazolés visant l'étape d'exportation extracellulaire de la capside impliquant pUL97 et sa protéine accessoire pUL27. Jusqu'à présent peu d'informations structurales sont disponibles concernant ces diverses cibles thérapeutiques. Par conséquent, le mode d'action des dérivés ribobenzimidazolés reste peu clair et leur optimisation peu rationnelle.

Dans ce manuscrit, nous présentons les différents résultats issus d'études par spectroscopies optiques, résonance magnétique nucléaire et modélisation moléculaire.

Un premier axe de recherche a concerné l'étude de la petite sous-unité du complexe terminase pUL89. Les domaines hélicase, endonucléase et d'interaction avec son partenaire pUL56 ont fait l'objet d'une recherche particulière.

Puis, deux domaines de la grande sous-unité pUL56 présumés capables d'interagir avec l'ADN ont été étudiés par spectroscopie UV/visible, dichroïsme circulaire et résonance magnétique nucléaire: un domaine de type doigt de zinc et un domaine pouvant permettre la dimérisation de pUL56 et la constitution d'un cluster basique de type *basic leucine zipper* (*bZIP*).

Finalement, des études préliminaires sur l'obtention d'un modèle de la phosphotransférase pUL97 et une analyse d'alignement de séquence opéré entre pUL27 et ses homologues herpesviraux ont constitué deux axes de recherches supplémentaires qui resteront à développer en profondeur pour l'élucidation des modes d'actions des dérivés L-ribobenzimidazolés.

## II. BUT DE L'ÉTUDE

Le travail présenté ici a pour objectif premier de dégager de nouveaux axes d'investigation à visées thérapeutiques concernant les différentes interactions impliquant les terminases du cytomégalovirus humain. Comme nous l'avons souligné pendant l'introduction, de nombreux traitements sont disponibles mais quels que soit les cas, divers problèmes sont rencontrés. Dans l'optique d'un traitement visant en premier lieu l'ADN polymérase de CMVH, pUL54, plusieurs résistances sont apparues après un traitement prolongé aux différents antiviraux à la fois sur les gènes *UL54* et *UL97*.Cet inconvénient majeur a ainsi débouché sur le ciblage de nouvelles protéines par de nouvelles molécules. Une des voies envisagées est l'inhibition du complexe terminase. On peut imaginer deux types de voies d'inhibition de ce complexe: soit en abolissant sa formation, c'est-à-dire la reconnaissance mutuelle des deux sous-unités pUL56 et pUL89; soit en interférant avec les interactions que le complexe terminase peut établir avec ses partenaires et en particulier avec l'ADN viral.

Les mutations de résistances au niveau des protéines pUL56 et pUL89 après traitement par les dérivés benzimidazolés permettent d'établir un lien entre l'étape d'encapsidation de l'ADN au cours du cycle viral et ce nouveau type de molécules qui ont débouchés sur la proposition de nombreuses autres molécules ayant notamment une meilleure biodisponibilité.

Contrairement aux DNA-polymérases qui constituent une classe de protéine retrouvée chez tout type d'organisme, y compris l'homme, les terminases ne rencontrent d'équivalent que chez un nombre limité de systèmes vivants (entre autres les bactériophages).

Ceci entraîne une difficulté supplémentaire et non négligeable pour la conception de nouveaux inhibiteurs car très peu d'informations structurales sont accessibles sur les terminases et en particulier celles des herpesvirus.

L'enjeu de cette étude est donc de répertorier un maximum de données structurales afin de pouvoir décrire à l'échelle atomique de nouvelles cibles thérapeutiques et ainsi, par les méthode de *drug design in silico*, de concevoir des inhibiteurs hautement spécifiques des interactions prenant place au sein du complexe terminase ou entre ce dernier et ses partenaires.

La RMN étant limitée par la taille des systèmes étudiés à cause de la multiplication croissante de signaux et de l'élargissement de la largeur de raie due aux phénomènes de relaxation transversale en rapport avec la taille, il nous faut dans un premier temps définir les

domaines accessibles par cette spectroscopie grâce à des alignements de séquences entre un maximum d'homologues.

Cette étude théorique permettra ainsi de répertorier les régions ayant des rôles structuraux ou catalytiques indispensables au complexe terminase en partant du dogme selon lequel la conservation implique la fonction. De cette manière, cette étude préliminaire permettra d'orienter les différents axes d'investigation telle que la formation du complexe terminase et ses interactions avec ses partenaires telle que la protéine portail pUL104 par exemple, ses substrats à savoir l'ADN viral, ou encore la compréhension de mécanisme d'action de molécules inhibitrices de la maturation et de l'encapsidation de l'ADN du cytomégalovirus humain.

Nous développerons dans ce manuscrit deux axes principaux.

Le premier axe est l'interaction entre les deux sous-unités du complexe terminase. Du côté de pUL89, l'identification de la partie minimum requise pour la reconnaissance à pUL56 a été découverte par l'équipe de Bogner en 2006 (Thoma, Borst et al. 2006).

Cependant, la structure de ce fragment de 21 acides aminés identifié était encore inconnue. Nous nous sommes donc proposé d'en résoudre la structure par RMN, puis, par homologie avec des RNase de type I, de reconstruire le domaine auquel il appartient.

Par ailleurs, la modélisation de son domaine hélicase a pu être réalisée par homologie avec son homologue, gp17, du bactériophage T4.

Le deuxième axe a visé l'étude des domaines potentiels d'interactions avec l'ADN *via* un motif de type *basic leucine zipper* (Schumacher, Goodman et al. 2000) situé au milieu de sa séquence, et un motif en doigt de zinc localisé dans sa partie N-terminal et hautement conservés chez tous les homologues herpesviraux de pUL56.

En dernier lieu, des études préliminaires ont été engagées sur l'étude des protéines pUL97 et pUL27, ciblées par le Maribavir. Ces travaux préliminaires sont décrits en Annexe A de ce manuscrit.

## III. HERPESVIRUS

### A. *Classification Et Phylogénie Des Herpesvirus*

La classification des différents herpesvirus à travers le règne animal a été établie en fonction des données de phylogénie moléculaire recueillies après séquençage des ADN viraux. Ce qui a donné lieu à l'arbre phylogénétique des herpesvirus de la Figure 1 proposé par Davison (Davison 2002).

Depuis leur ancêtre commun, les herpesvirus se sont différenciés en 3 grands groupes (1, 2 et 3). Le premier comprend les virus associés aux mammifères (humains, murins, équins, bovins…), le second inclut le groupe des poissons (salmonidé SalHVs et poisson-chat CCV) et des batraciens (grenouille RaHVs) et le troisième est représenté, ici, par l'herpesvirus de l'huître (OsHV-1).

Chez l'homme en particulier, huit formes différentes ont été identifiées. Les différentes sous famille d'herpesvirus se distinguent, entre autre, par leurs modes d'infection récapitulés dans le Tableau 1 (pour revue Weir 1998 (Weir 1998) et page internet (Conseil 2006)).

En plus des sous familles alpha-, bêta- et gamma- herpesvirinae, l'étude de Davison fait apparaître des genres différents à l'intérieur de ces sous-familles. Chez les alpha-herpesvirinae, les simplexvirus constituent le genre α1, les varicellovirus le genre α2 et les "Marek's disease-like" virus le genre α3. La sous-famille bêta-herpesvirinae est subdivisée en deux genres prédominants (et un non attribué) que sont les cytomégalovirus/muromégalovirus pour le genre β1 et les roseolovirus pour β2. Comme pour les bêta-, les gamma- sont principalement classés en deux genres γ1 pour les lymphocryptovirus et γ2 pour les rhadinovirus (Tableau 1).

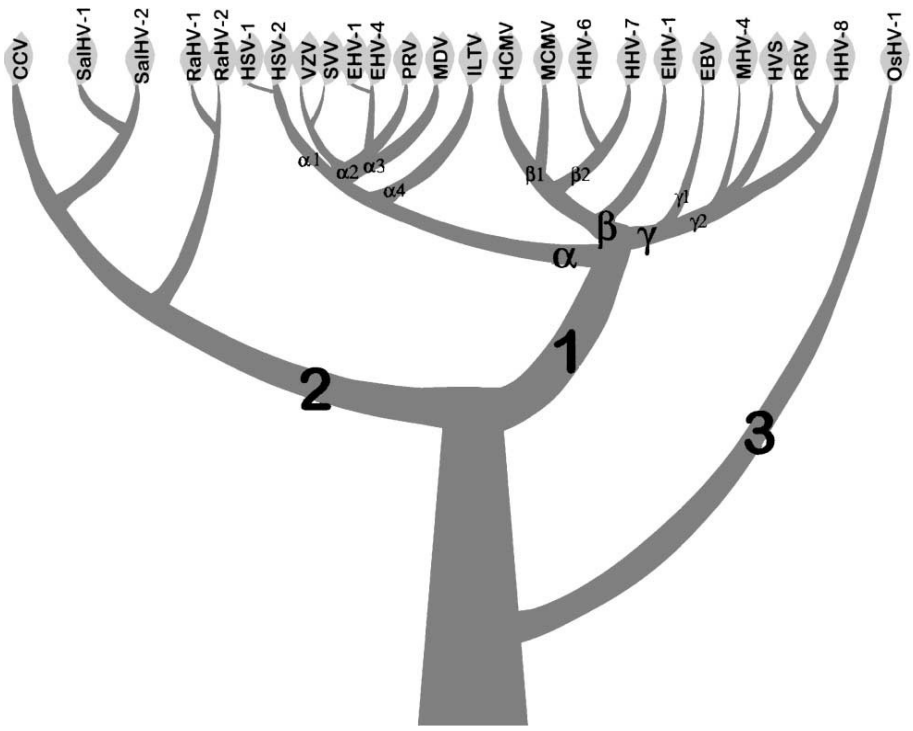



**Figure 1: Arbre représentant l'évolution des herpesvirus.** Les grandes branches séparent la famille des mammifères (1), poissons et batraciens (Walter, Ealick et al.) et mollusques (Biron, Harvey et al.). La famille 1 est subdivisée en sous-familles α, β et γ, elles mêmes contenant des genres différents (α1-4, β1-2 et γ1-2). HCMV appartient au genre β1. D'après Davison 2002.

| Sous famille | Spectre d'hôte | Virus | Non usuel | Abbréviation | Taille du génome | Site de latence |
|---|---|---|---|---|---|---|
| α herpesvirinae | large | HHV-1 | Herpes Simplex Virus 1 | HSV-1 | 152 kpb | Neurones |
| | | HHV-2 | Herpes Simplex Virus 2 | HSV-2 | 152kpb | |
| | | HHV-3 | Varicella-zooster virus | VZV | 125kpb | |
| β herpesvirinae | restreint | HHV-5 | Cytomégalovirus Humain | CMVH | 248kpb | Cellules hématopoïétiques |
| | | HHV6-A | Roseolovirus | HHV6-A | 159kpb | Monocytes |
| | | HHV6-B | Roseolovirus | HHV6-B | 162kpb | |
| | | HHV-7 | HHV-7 | HHV-7 | 145kpb | Lymphocytes T (CD4+) |
| □herpesvirinae | très restreint | HHV-4 | Epstein-Barr virus | EBV | 172kpb | Lymphocytes B |
| | | HHV-8 | Sarcome de Kaposi | HHV-8 | 170kpb | Cellules épithéliales |

**Tableau 1: Classification des herpesvirus humains.**

### *B. Structure Des Herpesvirus.*

Historiquement, plusieurs critères physiques ont permis de classer des virus dans la famille des herpesvirus. Ainsi, tous les herpesvirus sont constitués d'un ADN double-brin linéaire (dont la taille varie avec le type de virus) contenu dans une capside à symétrie icosaédrique elle-même baignant dans une substance amorphe appelée tégument contenant les protéines utiles à la primoinfection. Le tout est enveloppé d'une enveloppe à bicouche lipidique (Figure 2).

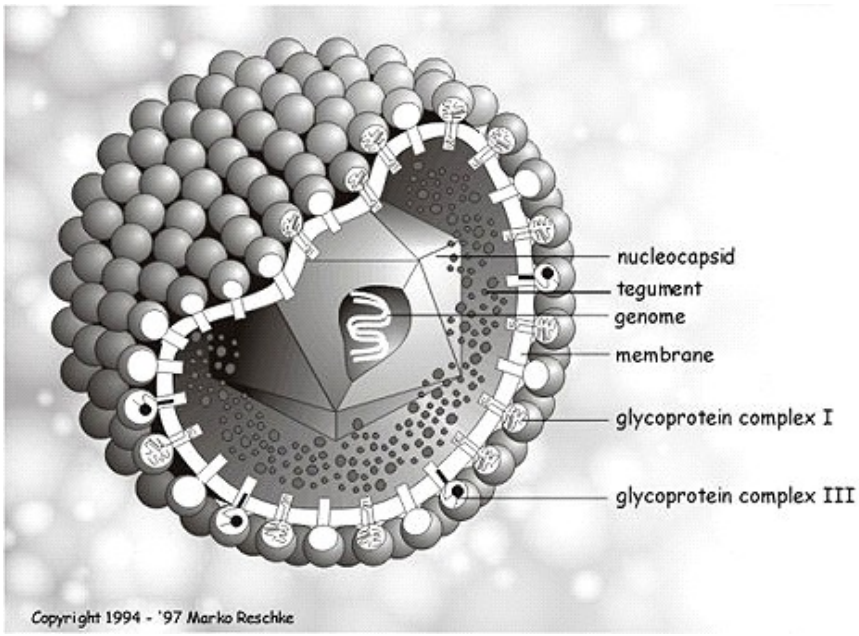



**Figure 2: Structure des herpesvirus.** Le double-brin d'ADN viral est renfermé dans une nucléocapside à symétrie icosaédrique baignant dans un tégument. Le virion est enveloppé d'une bicouche lipidique portant les différents complexes glycoprotéiques.

#### B.1. Les protéines de l'enveloppe

Chaque herpesvirus porte à peu près une douzaine de glycoprotéines différentes sur sa membrane lipidique issue de la fusion du virion avec la membrane d'une cellule. Trois d'entre elles sont très conservées à travers les sous-familles: gB, gH et gL (gH et gL sont présentes sous forme d'un hétérodimère et parfois d'hétérotrimère). Chez CMVH, trois complexes de glycoprotéines liées par des ponts disulfures peuplent la plus grande partie de l'enveloppe du virion (Gretch, Kari et al. 1988). Ils sont appelés gCI formé par un homodimère de la protéine gB (Lopper and Compton 2002), gCII complexe des protéines gM et gN (Mach, Kropff et al. 2005), et enfin gCIII qui est l'association de gL, gH et gO. En plus de ces protéines, HSV-1

et HSV-2 (mais pas VZV) ainsi que d'autres alpha herpesvirus possèdent une protéine supplémentaire, gD (Shukla and Spear 2001; Rey 2006). Cependant, de manière spécifique à CMVH d'autres glycoprotéines sont observées comme gpTRL10, gpTRL11 et gpUL132 dont le rôle n'est pas encore avéré.

gB est une glycoprotéine N-glycosilée à hauteur de 50 à 60 kDa de chaîne polysaccharidique. De plus c'est une protéine phosphorylée sur sa sérine 900 qui la localise dans le réseau trans golgien Elle est indispensable à une réplication virale efficace (Jarvis, Jones et al. 2004).

gM et gN, produits des gènes *UL100* et *UL73* respectivement, forment le complexe gCII. Contrairement au cas des alpha-herpesvirus la formation du complexe gM/gN est indispensable à l'infection par le virus. La formation de ce complexe se fait grâce au pont disulfure impliquant la cystéine 44 de gM et la cystéine 90 de gN. Cependant, la formation de ce complexe n'exige pas la formation de ce lien covalent puisqu'il fait intervenir une interaction plus importante impliquant la partie C terminale de gM avec gN. L'importance de ce complexe est multiple puisqu'on présume qu'il est à la fois impliqué dans l'assemblage du virion ou à la fixation et l'entrée du virus pendant le stade précoce de l'infection (Mach, Kropff et al. 2005). Ce dernier point est suggéré par la longue chaîne glycosidique portée par gN qui équivaut à 200% de son poids moléculaire (Kari and Gehrz 1992).

Le complexe gCIII est un hétérotrimère formé par l'association des protéines gH (gpUL75), gL (gpUL115) et gO (gpUL74). Un premier hétérodimère gH/gL va être constitué via un pont disulfure indispensable à la stabilité du complexe. Celui-ci va s'associer avec gO pour former le précurseur de gCIII. Après modification post-traductionnelle dans le reticulum endoplasmique le complexe est mature (Huber and Compton 1999) (Figure 9). Il a été montré par Paterson et al en 2002 que gO est très impliquée dans l'infection par son rôle dans la fusion membranaire (Paterson, Dyer et al. 2002).

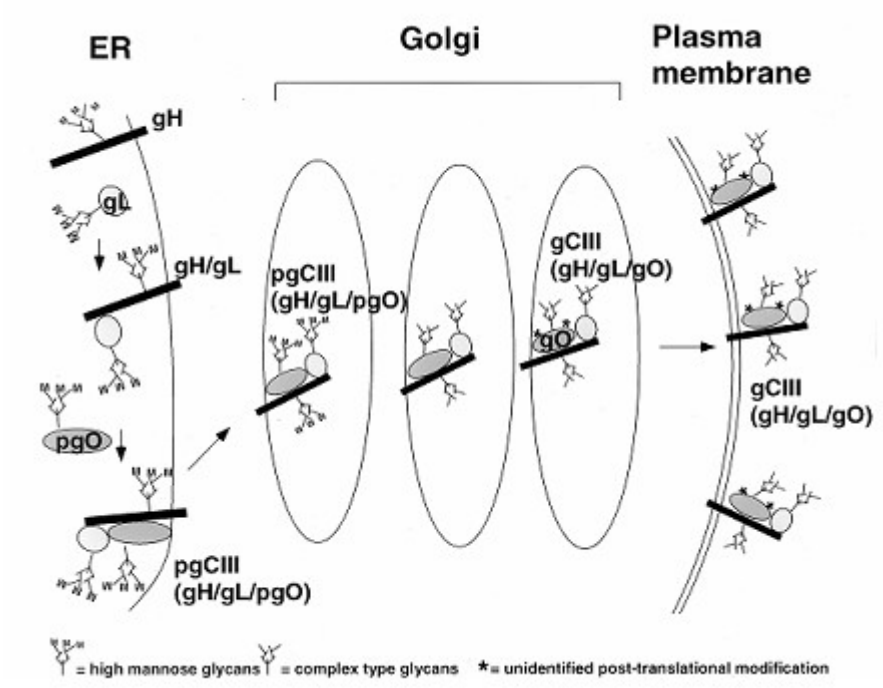



**Figure 3: Cycle de maturation du complexe gCIII.** D'après Huber et Compton, 1999

### B.2. Les protéines du tégument

Le tégument est définit comme étant compris entre la membrane lipidique et la capside. Il contient un grand nombre de protéines (environ une trentaine) qui vont être larguées à l'intérieur de la cellule après fusion des membranes cellulaire et virale. Ces protéines vont encadrer toutes les étapes de l'infection virale. Leur dénombrement et leur description sont résumés dans le Tableau 2, d'après Kalejta, 2008.

Tout d'abord les protéines pUL47 (*HMWP-Binding Protein*) et pUL48 (*Hight Molecular Weight Protein*) sont deux protéines intimement liées à la capside contenant l'ADN viral. Le complexe formé va guider la capside le long des microtubules jusqu'au noyau. En plus de cette fonction de guide, pUL48 (Wang, Loveland et al. 2006) est une protéase qui va veiller à éliminer les molécules d'ubiquitines afin que le système viral ne se fasse reconnaître puis dégrader par le système immunitaire de l'hôte.

Lors de l'infection, le virus apporte également au sein de son tégument les protéines qui vont lancer la machinerie de traduction de l'ADN. Cette fonction est principalement imputable à la phosphoprotéine pp71 (car de masse 71 kDa) produit du gène *UL82.* C'est en effet pp71 qui va activer la production des protéines très précoces. La protéine ppUL35, même si son rôle est encore peu clair, semble faciliter l'expression des gènes précoces en s'associant avec pp71. Cette interaction va mener à la dégradation de Daxx -qui est chargée

de la répression de l'expression virale- par pp71. Quand le virus est au stade latent, pp71 reste stockée dans le cytoplasme de l'hôte attendant le moment adéquate pour pénétrer dans le noyau. Enfin une troisième protéine intervenant dans l'expression et la traduction est une autre phosphoprotéine, ppUL69, dont le rôle est de transporter des ARNm non-épissés (c'est à dire correspondant aux gènes précoces et tardifs) du noyau vers le cytoplasme pour qu'ils soient traduits en protéines (Toth and Stamminger 2008). Il semblerait également que cette protéine soit responsable d'un blocage de la division cellulaire en phase G1.

Pour que l'infection se fasse sans pertes trop significatives pour le virus, celui-ci possède des protéines chargées de neutraliser la réponse immunitaire. L'une d'elle, la plus représentée dans le tégument, est la phosphoprotéine ppUL83 (ou pp65). En complément de ppUL83 les protéines pIRS1 et pTRS1 ont aussi un rôle dans la neutralisation de la réponse immunitaire.

pUL32 (ou pp150) est la deuxième protéine la plus représentée dans le tégument après ppUL83. Son rôle est clairement associé à l'acquisition du tégument, en interagissant avec la procapside néoformée après la sortie du noyau. D'un autre côté, la kinase pUL97 va faciliter l'incorporation des protéines dans le tégument en les phosphorylant et pUL99 (pp28) va se charger de terminer la formation des particules virales.

**Tableau 2 (page suivante): Protéines du tégument du CMVH.** Les phénotypes sont listés selon que la protéine considérée soit essentielle (E), dispensable (D) ou augmente (A) la réplication lytique dans les cellules de fibroblastes humain in vitro. La fonction de chacune de ces protéines y est sommairement décrite. Les protéines dont la fonction n'est pas encore élucidée sont réduites à un blanc dans la colonne « Fonction(s) » (d'après Kalejta, 2008).

| Gene | Phenotype | Fonction(s) |
|---|---|---|
| UL23 | D | |
| UL24 | D | |
| UL25 | D | Colocalisée avec pp28 |
| UL26 | A | Augmente la stabilité des protéines du vrion |
| UL32 (pp150) | E | Dirige la capside vers le site d'enveloppement final |
| UL35 | A | Active l'expression des gènes viraux |
| UL36 | D | Inhibe l'apoptose |
| UL38 | A | Inhibe l'apoptose |
| UL43 | D | |
| UL44 | E | Facteur de processivité/transcription de la DNA polymérase UL54 |
| UL45 | A | Sous unité inactive ( ?) de la ribonucléotide réductase |
| UL47 | A | Libération de l'ADN viral de la capside |
| UL48 | E | Déubiquitine protéase<br>Libération de l'ADN viral de la capside |
| UL50 | E | Sortie nucléaire de la capside |
| UL53 | E | Sortie nucléaire ou assemblage de la capside |
| UL54 | E | DNA polymérase virale |
| UL57 | E | Protéine de liaison à l'ADN simple brin |
| UL69 | A | Exportation du noyau des mRNA non-épissés<br>Arrêt du cycle cellulaire en phase G1 |
| UL71 | A/E | |
| UL72 | A | dUTPase inactive ( ?) |
| UL76 | A/E | Fonction nucléaire ? |
| UL77 | E | Probable pyruvoyl décarboxylase |
| UL79 | E | |
| UL82 (pp71) | A | Dégrade Daxx, facilite l'expression des gènes IE<br>Dégrade Rb, stimule la progression du cycle cellulaire<br>Empêche l'expression des MHC à la surface des cellules |
| UL83 (pp65) | D | Activité kinase endogène<br>Associée à l'activité kinase<br>Neutralise les défenses immunitaires de l'hôte |
| UL84 | E | Facteur de la réplication/transcription de l'ADN viral |
| UL88 | D | |
| UL93 | E | |
| UL94 | A/E | Probable protéine de liaison à l'ADN<br>Similaire à un autoantigène dans des cas de sclérose systémique |
| UL96 | A/E | |
| UL97 | A | Kinase phosphorylant le Ganciclovir<br>Stimule la replication de l'ADN viral<br>Assemblage/exportation<br>Fonctions analogues aux cyclin-dependent kinases<br>Perturbe les aggrésomes nucléaires |
| UL99 (pp28) | E | Guide l'empaquetage des particules enveloppées |
| UL103 | A | |
| UL112 | A | Facteur de réplication de l'ADN viral |
| IRS1/TRS1 | A/E | Inhibe la réponse antiviral des PKR<br>Participe à l'assemblage du virion |
| US22 | D | |
| US23 | A | |
| US24 | A | Active l'expression des gènes viraux |

### B.3. Les protéines de la capside

Cinq protéines essentielles et communes aux différents herpesvirus constituent la capside (Chen, Jiang et al. 1999).

960 copies de la protéine majeure de capside (*Major Capsid Protein*, *MCP* 151kDa) s'organisent en 150 hexamères et 12 pentons. Selon le type d'herpes leur masse moléculaire est comprise entre 120 kDa et 155 kDa. pUL86 en est son représentant chez CMVH.

Deux copies de la protéine mineure d'encapsidation (*minor Capsid Protein*, *mCP*), vont s'associer avec une copie de la protéine mC-BP (*minor Capsid Binding Protein*) pour former un triplexe s'intercalant entre les hexamères et pentones de MCP. pUL85 et pUL46 sont respectivement la mCP et la mC-BP chez le cytomégalovirus humain.

pUL48.5 (ou pUL48/49) est le représentant chez CMVH de la plus petite protéine de capside (SCP, *Smallest Capsid Protein*) qui a elle aussi une fonction de cohésion.

Enfin, le gène UL80 code pour une protéine (pUL80) qui après clivages successifs engendrera trois protéines complémentaires: pUL80a qui constitue la partie N-terminal de pUL80 et clive la partie C-terminal pour délivrer pUL80.5 ou AP (*Assembly protein*) qui peut être inhibé par la dernière protéine dérivée de pUL80 (Britt and Boppana 2004; Mettenleiter, Klupp et al. 2006). Une image tirée d'études de microscopie électronique est représentée en Figure 4 (Chen, Jiang et al. 1999).

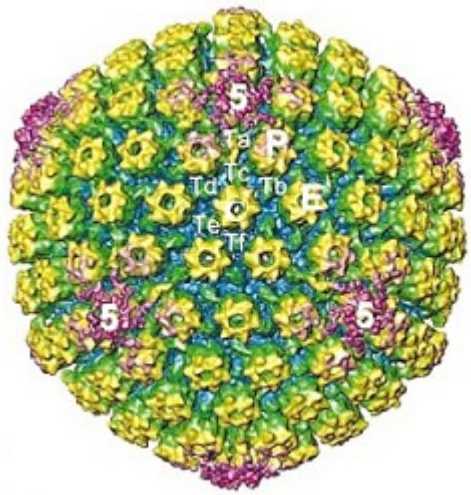



**Figure 4: Structure tridimensionnelle de la capside de HSV-1 obtenue par Microscopie Électronique à une résolution de 20 Å (d'après Chen et al., 1999).** Chacune des 16 unités asymétriques de la capside contient un penton (noté 5), 2 ½ hexons (notés P,E,C) et 5 1/3 triplexes (notés Ta-f). L'image est représentée suivant un axe d'ordre 3.

### B.4. L'ADN viral

L'ADN de CMVH est un ADN double-brin de 230 kpb environ, codant pour plus de 200 protéines, ce qui en fait le plus gros des herpesvirus. La topologie de l'ADN de CMVH est organisée en deux grandes régions appelées unique longue (*UL*) et unique courte (*US*). Ces deux régions sont encadrées par des séquences répétées inversées TRL et TRS pour *Terminal Repeat Long* et *Terminal Repeat Short*, respectivement, localisées aux extrémités du génome et IRL et IRS pour *Internal Repeat Long* et *Internal Repeat Short* situés à la jonction des deux fragments. Le fragment UL code pour les produits de gènes *UL1* à *132* et le fragment court pour les produits de gènes *US1* à *34*. A ses deux extrémités et à la jonction IRL/IRS, le génome contient des séquences *pac* (*cis-acting packaging motifs*) essentielles pour la reconnaissance et le clivage de l'ADN viral au moment de l'empaquetage (Figure 5).

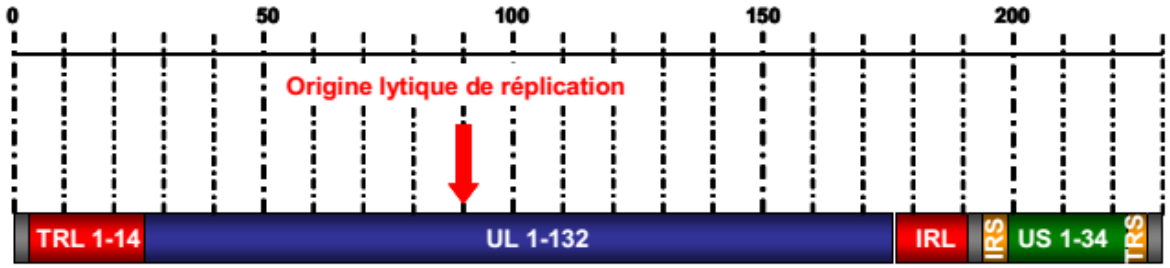



**Figure 5: Représentation schématique du génome de CMVH.** Long de 230 kpb, il est constitué de deux segments de gènes, Unique Long (UL) et Unique court (US), codant respectivement pour 132 et 34 protéines. UL est encadré par un segment terminal répété long (TRL) et un segment interne répété long (IRL). US est quant à lui compris entre les segments IRS et TRS (respectivement, segment interne et terminal répété court). La graduation en haut est en kpb. L'origine lytique de réplication est désignée par une flèche rouge. Les séquences *pac* sont représentées par des rectangles gris.

| Fonction/Structure | PhaseIE | PhaseE | PhaseL |
|---|---|---|---|
| **Inhibition de l'apopotose** | pUL36<br>vMIA | | |
| **Régulation et transactivation de l'expression des gènes viraux** | IE1-72, IE2-86<br>IE2-55<br>pIRS1, pTRS1<br>gpUS3 | ppUL69 | γIE2<br>pp71 |
| **Synthèse de l'ADN viral** | | DNA-pol<br>ssDNA-BP<br>UL112,113<br>UL114 | ppUL44<br>helicase primase<br>ppUL84<br>pUL97 |
| **Enveloppe** | | gp48, gB | gpUL33<br>gH, gL, gM<br>gN, gO |
| **Capside** | | mC-BP | SCP, MCP<br>pUL80.a, AP |
| **Tégument** | | | pp150, pp65<br>pp28 |

**Tableau 3: Classement et apparition chronologique de quelques protéines du CMVH.** L'axe horizontale représente un axe du temps jalonné par les phases IE (*Immediately Early*), E (*Early*) et L (*Late)*. Certaines protéines sont produites entre deux phases (Adapté de Landolfo, 2003).

La traduction des gènes répond à une chronologie bien définit se subdivisant en trois étapes: immédiatement précoce (*Immediatly early*, *IE*), précoce (*Early*, *E*) et tardive (*Late*, *L*) (Tableau 3). L'initiation de cette cascade est assurée par la protéine apportée par le tégument, pp71.

La phase immédiatement précoce correspond à l'expression des gènes *IE1* et *IE2*. Ceux-ci sont des transactivateurs des gènes viraux et cellulaires. Ainsi, ils détournent la machinerie cellulaire au profit de la production des particules virales préparant ainsi l'entrée en phase précoce.

La phase précoce est une production en plusieurs exemplaires des différentes protéines indispensables à la synthèse de l'ADN viral. Cette synthèse est dépendante de la reconnaissance de l'origine lytique de réplication par la phosphoprotéine ppUL84. L'ADN viral se circularise pendant cette phase. Il en suit un déroulement de l'hélice d'ADN par le complexe primase/hélicase constitué par pUL70, pUL102 et pUL105. ppUL57 va empêcher

la réhybridation des brins d'ADN en se fixant sur les ADN simple-brin, d'où son nom de *single-stranded DNA Binding protein* (*ssDNA-BP*). C'est alors que la polymérase virale pUL54 et sa protéine accessoire pUL44 initient la synthèse de l'ADN selon le modèle du cercle roulant. Il en résulte une succession d'unités virales sous forme de concatémère.

Enfin, la phase tardive correspond à la production des protéines de la capside et du tégument, à la maturation de l'ADN viral assurée par le complexe terminase pUL89/pUL56, à son encapsidation grâce à la protéine portail pUL104, et finalement à son excrétion hors du noyau.

## IV. Cibles thérapeutiques et antiviraux dirigés contre le Cytomégalovirus Humain

La Figure 6 résume les grandes étapes de l'infection par le cytomégalovirus humain. On y retrouve les différents éléments développés précédemment (entrée du virus, rôle des protéines du tégument dans l'invasion nucléaire et la neutralisation des défenses cellulaires, expression des protéines selon un ordre défini par leur rôle dans la formation du virus néoformé).

Afin de contrer l'infection de CMVH plusieurs mécanismes moléculaires du cycle viral ont été l'objet de cibles thérapeutiques. A ce jour trois stratégies ont été entreprises. Celles-ci touchent l'inhibition de la polymérase virale UL54 (étape 4a), la traduction des gènes (étape 3) par une approche oligonucléotidique antisens et enfin la maturation et l'encapsidation des unités virales dans la procapside (étape 5).

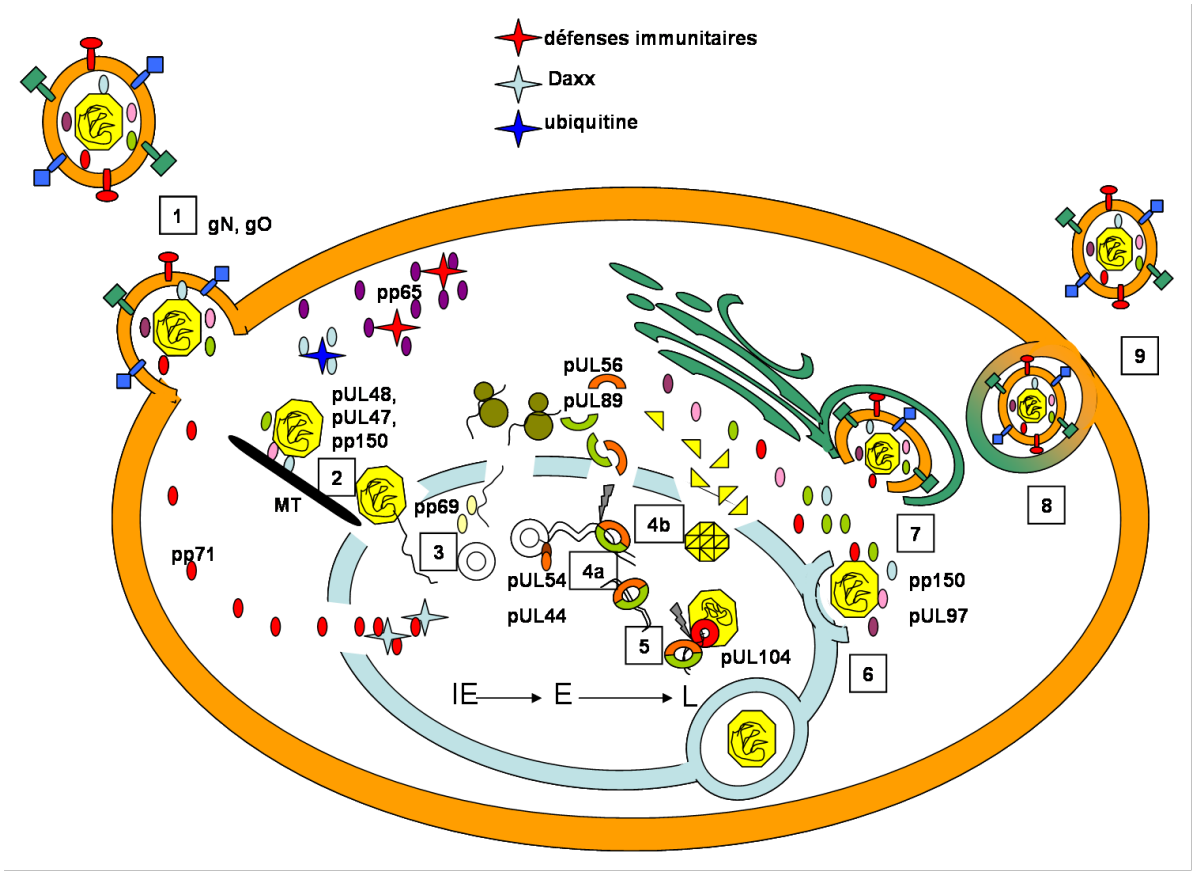



**Figure 6: Modèle simplifié du cycle d'infection du CMVH.** 1. Entrée du virus par endocytose par l'intermédiaire des complexes glycoprotéiques et largage des protéines du tégument dans le noyau. 2. Exportation de la capside le long des microtubules (MT) et entrée de l'ADN viral dans le noyau. 3. Expression des gènes IE et circularisation de l'ADN viral. 4a. Synthèse de l'ADN selon le modèle du cercle roulant, clivage et exportation des unités

virale vers la capside néoformée (4b). 5. Entrée de l'ADN viral dans la capside et deuxième coupure de l'ADN pour enlever l'excédant. 6. Excrétion du noyau de la capside (nuclear egress) et acquisition du tégument. 7. Entrée dans le Golgi et acquisition de l'enveloppe. 8. Fusion et 9. Exportation du virus hors de la cellule. Les protéines dont le nom est précisé remplissent une fonction importante pour l'étape considérée.

# A. Inhibition De La Voie De Synthèse De L'ADN Viral.

## A.1. pUL54: polymérase de CMVH et sa protéine accessoire pUL44

L'ADN polymérase, pUL54 est une protéine de 1242 acides aminés trouvant des homologues chez plusieurs bactériophages, archaebactéries, eucaryotes, polymérases cellulaires humaines et herpesvirus (Larder, Kemp et al. 1987; Braithwaite and Ito 1993; Wang, Sattar et al. 1997) constituant la famille des polymérases B. La séquence de pUL54 est décrite selon sept domaines conservés numérotés par ordre décroissant d'homologie avec la polymérase α humaine, et un domaine δ-C spécifique aux herpesvirus (Kouzarides, Bankier et al. 1987; Larder, Kemp et al. 1987; Ye and Huang 1993) toujours présents dans le même ordre séquentiel IV-II-(δ-c)-VI-III-I-VII-V. Les polymérases B possèdent deux fonctions distinctes: 5'-3' polymérase ADN-dépendante et 3'-5' exonucléase. Bien que la structure de pUL54 demeure encore inconnue, les structures de la polymérase du bactériophage RB69 seule (PDB :1WAF) (Wang, Sattar et al. 1997) et en complexe avec l'ADN (PDB :1IG9) (Figure 13) (Franklin, Wang et al. 2001) ou de HSV-1 (PDB :2GV9) (Liu, Knafels et al. 2006) sont connues.

Le domaine polymérase de pUL54, comme de toutes les polymérases B, est constitué de trois domaines appelés *fingers* (doigts), *thumb* (pouce) et *palm* (paume) (représentés en bleu, vert et rose, respectivement dans la Figure 7A. et C), tant leur arrangement spatiale est analogue à une main. La polymérisation de l'ADN s'effectue dans la paume de la main. Le nucléotide prêt à être accroché au brin d'ADN en élongation est capté par trois résidus basiques contenus dans les régions conservées des doigts, puis les doigts se rabattent sur la paume où des interactions entre ses phosphates et des ions divalents coordonnés au niveau des Motifs A (DxxSLYPS) et C (YGDTDS) (Delarue, Poch et al. 1990) vont maintenir la queue polyphosphate. Le ribose est quant à lui maintenu par une tyrosine par empilement hydrophobe. Le brin matriciel est calé dans un sillon dans lequel se loge un tour entier d'hélice et est maintenu par de nombreuses interactions électrostatiques via des résidus non conservés à travers les α- polymérases. Le primer en élongation est surtout maintenu par des interactions de types liaisons hydrogènes.

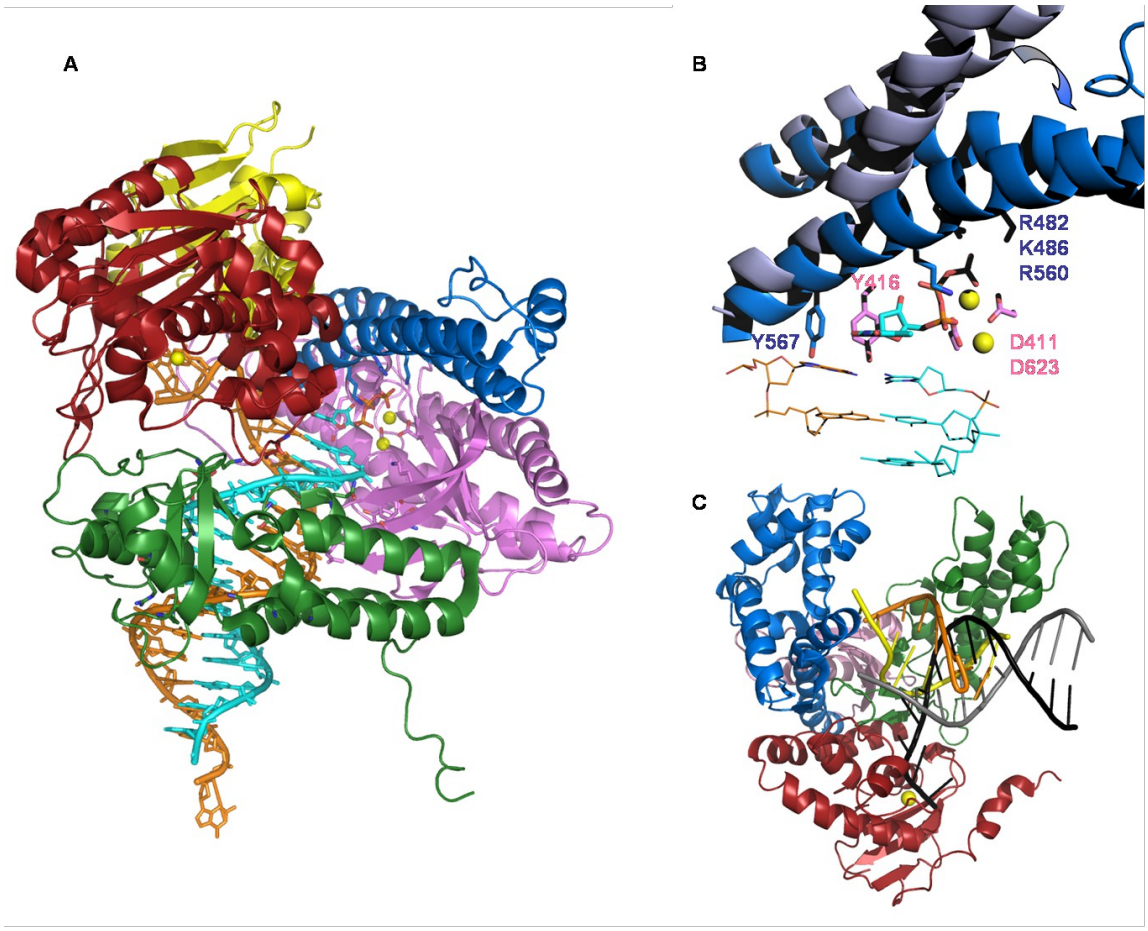



**Figure 7: Structure et activité des polymérases.** A. Structure complexée de la polymérase du bactériophage RB69 cocristalisé avec le primer néosynthétisé (cyan) et le brin matrice en orange (PDB :1IG9). Sont représentés en rouge, jaune bleu, rose et vert, les domaines 3'-5' exonucléasique, N-terminal, doigts (*fingers*), paume (*palm*) et pouce (*thumb*), respectivement. Les sphères jaunes représentent les ions calcium. B. Mécanisme de fixation d'un dTTP (sticks cyan) sur le primer (lignes cyan) grâce aux résidus basiques du doigt, les aspartates chélatant les ions $Ca^{2+}$ et la tyrosine 416 de la paume qui va stabiliser le ribose. La tyrosine 567 contrôle le bon appariement des bases. La flèche indique le mouvement du doigt avant (gris, PDB 1WAF) et pendant la polymérisation (bleu, PDB :1IG9) C. Structure de la polymérase de Taq (PDB: 1TAU) (Eom, Wang et al. 1996) dont le brin d'ADN (primer orange, matrice jaune) a été superposé avec celui de la structure PDB 1KLN (primer noir, matrice grise) (Beese, Derbyshire et al. 1993). Les mésappariements impliquent un basculement du primer du site de polymérisation au site exonucléasique par modification de l'affinité protéine ADN entraînant la fixation de l'extrémité 3' du primer dans le site catalytique exonucléasique.

Le contrôle des appariements est assuré par une autre tyrosine très conservée qui, se logeant dans le petit sillon de l'ADN double-brin, est sensible à l'encombrement stérique résultant de mésappariements (Franklin, Wang et al. 2001) générant une déstabilisation qui s'accompagne d'un basculement de l'extrémité 3' du brin d'ADN néosynthétisé vers le site exonucléasique (Figure 7B). Un cluster constitué de résidus de type carboxylate va ainsi procéder à l'excision du nucléotide faisant défaut (Figure 7C) (Steitz 1999). Le site actif du domaine 3'-5' exonucléasique chez pUL54 est constitué des domaines ExoI ($^{379}$VxEFxSEx$_3$Lx$_2$F$^{392}$), ExoII ($^{405}$TGYNx$_3$FD$^{413}$) et ExoIII ($^{536}$GxYCxQDx$_2$LV$^{546}$) (Ducancelle and Mazeron 2006).

Le rendement de polymérisation de telles enzymes est assuré par la fixation d'une protéine accessoire, pUL44 dans le cas de CMVH, qui stabilise le complexe ADN-polymérase/ADN matriciel lors de la catalyse.

Le complexe pUL44/pUL54 se forme par l'interaction du domaine C-terminal (1223-1242) de pUL54 avec sa protéine accessoire (PDB :1YYP) (Appleton, Brooks et al. 2006) tout comme pUL42, analogue de pUL44 chez HSV-1, se lie au domaine C-terminal de la polymérase virale pUL30 (PDB :1DML) (Zuccola, Filman et al. 2000) (Figure 8).

Cette association requiert dans un premier temps la phosphorylation de pUL44 par la phosphotransférase virale pUL97 (Krosky, Baek et al. 2003; Marschall, Freitag et al. 2003). pUL44 ainsi phosphorylée formera le complexe avec pUL54 capable de maintenir l'ADN matriciel pour une polymérisation efficace. La structure de pUL44 est constituée de deux monomères connectés par une boucle hydrophobe dont les résidus Phe$^{121}$ et Leu$^{86}$ jouent un rôle prépondérant (Appleton, Loregian et al. 2004). Chaque monomère est constitué d'un domaine N terminal (résidus 9-128) et C terminal (résidus 143-271) liés par une boucle de connexion. L'association avec pUL54 s'effectue au niveau de cette boucle par l'établissement d'un feuillet bêta. Il est suggéré que, du fait qu'une boucle sur deux soit le site de liaison à pUL54, le deuxième site pourrait servir de plateforme pour d'autres interactions tant il est avéré que cette protéine peut établir des interactions avec d'autres protéines (Park, Kim et al. 2006). L'interaction avec l'ADN requiert cette dimérisation et la présence de résidus basiques qui interagissent directement avec l'ADN (Komazin-Meredith, Petrella et al. 2008).

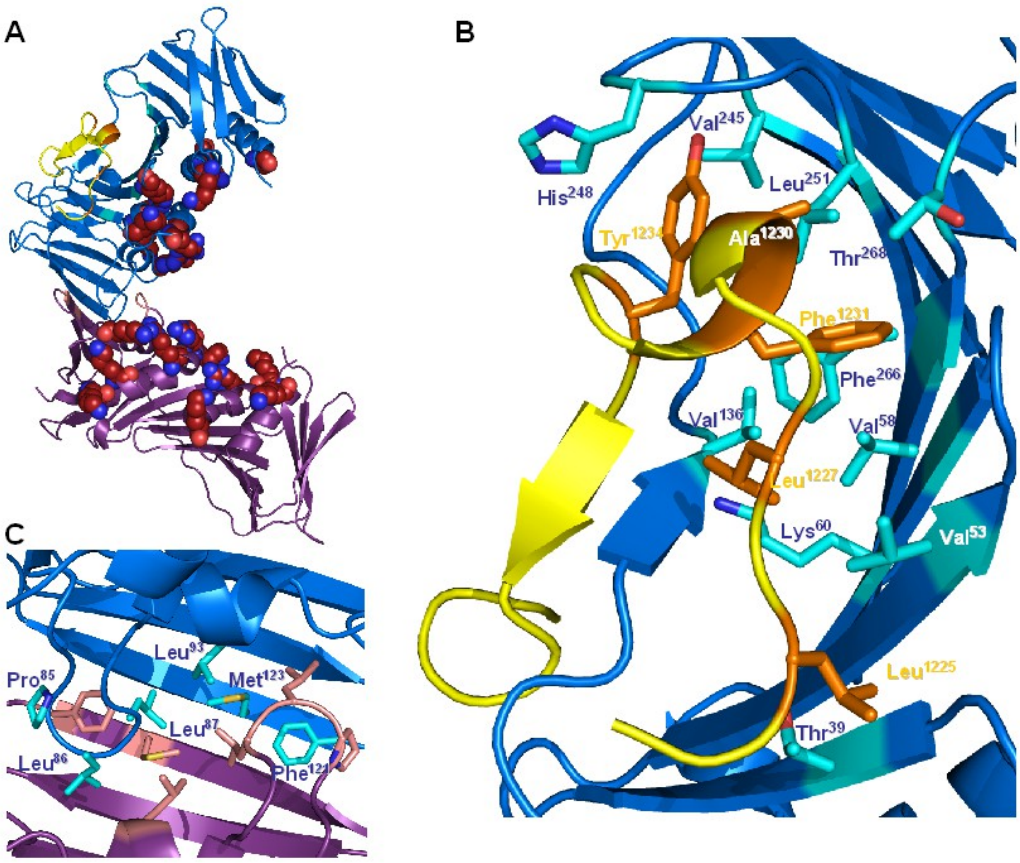



**Figure 8: pUL44.** A. Dimère de pUL44. La sous-unité 1 est représentée en bleu et la 2 en mauve. L'extrémité C terminale de pUL54 formant un feuillet bêta avec la boucle de jonction de pUL44 est représentée en jaune. Les chaînes latérales des arginines et lysines peuplant l'intérieur de la structure pour interagir avec l'ADN sont représentées en sphères (les carbones en rouge et les azotes en bleu). A partir de la superposition des structures PDB :1YYP (Appleton, Brooks et al. 2006) et 1T6L (Appleton, Loregian et al. 2004). B. Interaction entre pUL44 et pUL54 par la formation d'un feuillet antiparallèle et stabilisée par un réseau d'interactions hydrophobes. C. Interface hydrophobe homodimérique de pUL44 par formation d'un feuillet antiparallèle. Seuls les noms des résidus du monomère 1 sont précisés.

### A.2. Les traitements visant la synthèse de l'ADN viral et Formiversen

Ils impliquent des molécules analogues de nucléosides comme le Ganciclovir, (GCV), et sa prodrogue le Valganciclovir (VGCV), de nucléotide à l'instar du Cidofovir (CDV), de pyrophosphate dont le Foscarnet (ou PFA) est un représentant, ou encore un oligonucléotide de 21mer, le Formivirsen (De Clercq 2004). Ces trois types de molécules ont un mode d'action différent (Figure 9).

Le GCV, 9-(1,3-hydroxy-2-propoxymethyl)guanine, analogue acyclique de la désoxyguanosine, et le VGCV, valylester du GCV, doivent, pour être efficaces, être monophosphorylés par la phosphotransférase virale pUL97 (Littler, Stuart et al. 1992) avant d'être phosphorylés deux fois par des kinases cellulaires (Matthews and Boehme 1988). Une fois triphosphorylés, le GCV et le VGCV vont rentrer en compétition avec la désoxyguanosine. Leur spectre d'activité s'applique au CMV, HSV-1 & 2 et d'autres herpesvirus. Ils sont administrés en particulier chez les patients immunodéprimés atteints de rétinite en traitement curatif ou préventif. L'efficacité du GCV est discutable car son administration par voie orale offre une biodisponibilité de seulement 6%. Il doit être administré soit par intraveineuse soit par voie intraoculaire. Afin de pallier le manque de biodisponibilité, le VGCV a alors été développé puisqu'il élève cette valeur à 60% par voie orale. Cependant, ces deux molécules ne constituent pas des terminateurs définitifs de la réplication.

Contrairement aux GCV et VGCV, le Cidofovir, (S)-1-(3-hydroxy-2-phosphoryl-methoxypropyl) cytosine, analogue de la désoxycitidine, est déjà monophosphaté. De ce fait, le mécanisme d'action du CDV est légèrement différent de celui de GCV et VGCV puisqu'il n'implique pas un passage par pUL97 mais simplement sa double phosphorylation par les kinases cellulaires (Cihlar and Chen 1996). Le CDV a un spectre d'action extrêmement large qui s'applique aux herpesvirus en général, aux papilloma-, polyoma-, adeno- et poxvirus. Cependant, cet avantage ne suffit pas à dépasser l'efficacité du GCV puisque comme ce dernier, le CDV ne constitue pas un terminateur définitif de la polymérisation de l'ADN viral. Cet inconvénient a été par la suite levé par l'incorporation d'une deuxième molécule de CDV qui va mettre fin à toute activité enzymatique résidulle (Xiong, Smith et al. 1997). Il est administré par voie intraveineuse.

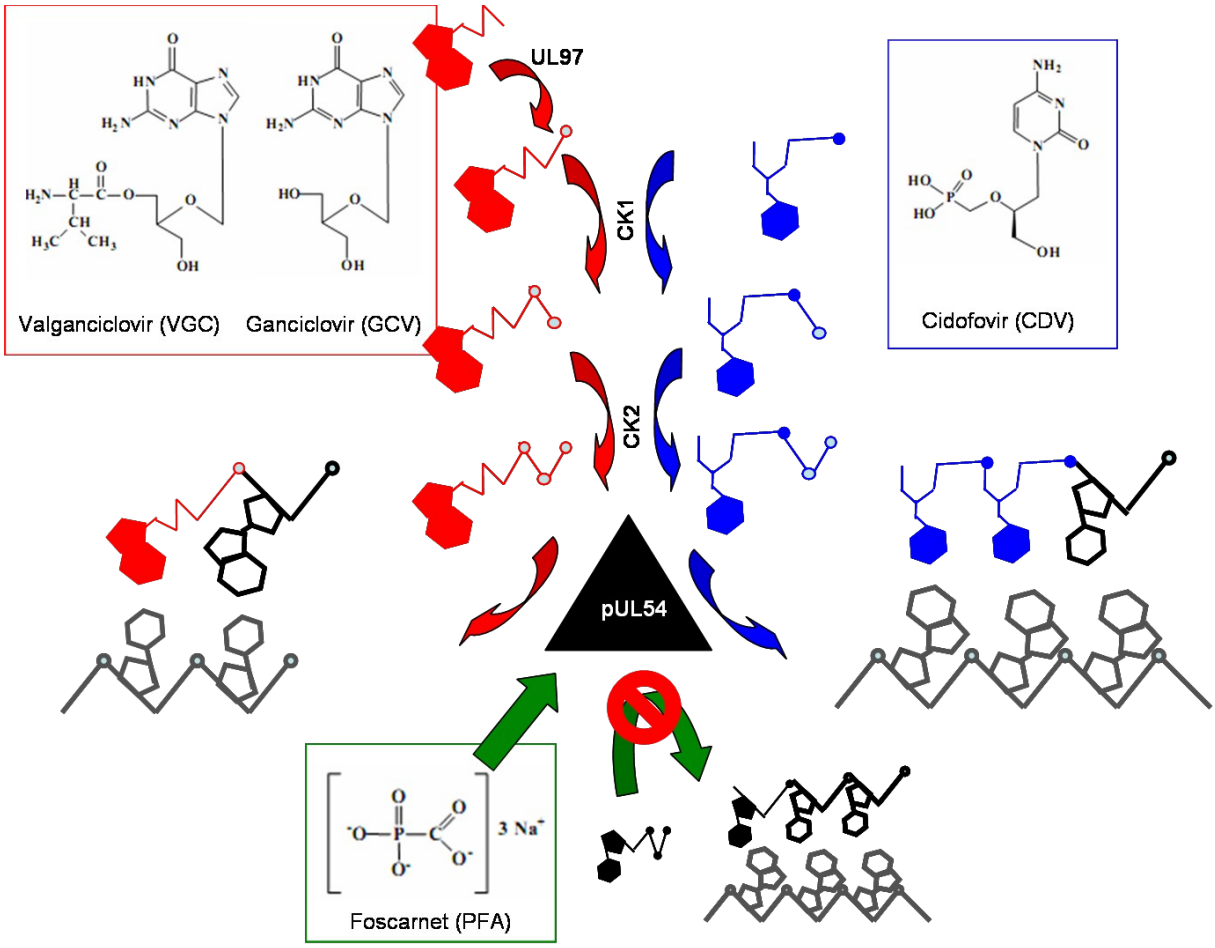



**Figure 9: Structures et mécanismes d'action d'inhibiteurs de la synthèse d'ADN viral.** En rouge, le mécanisme d'action du Ganciclovir (et du Valganciclovir) commence par leur primophosphorylation par la phosphotransférase virale pUL97, puis sont phosphorylés deux fois supplémentaires par des kinases cellulaires. Ils sont finalement intégrés au génome viral afin d'en bloquer la synthèse après compétition avec la désoxyguanosine. En bleu, le Cidofovir est déjà un composé monophosphorylé (disque bleu plein sur sa structure schématique) ce qui lui permet d'arriver plus vite vers pUL54 sans passer par pUL97 mais toutefois en se faisant diphosphoryler par les kinases cellulaires. Deux molécules de Cidofovir, compétiteur de désoxycytosine, sont indispensables pour une efficacité complète. En vert, le Foscarnet agit directement en se liant au domaine de fixation des queues polyphosphatées se situant sur le domaine *finger* de pUL54 (Figure 7B) empêchant toute incorporation de nucléotide supplémentaire (représenté en plein noir) au brin d'ADN en cours d'élongation. Les brins matriciels et néosynthétisés sont représentés en traits épais gris et noir, respectivement).

Le Foscarnet, trisodium phosphonoformate, inhibe la réaction dNTP -> dNDP +Pi. Lors de l'incorporation du nouveau nucléotide dans le primer en se fixant sur le site de liaison de la queue pyrophosphate sur la polymérase (Reusser 1996), il empêche ainsi le maintien et l'incorporation de tout nucléotide supplémentaire. Le Foscarnet est administré par voie intraveineuse aux patients infectés par les herpesvirus (HSV-1 & 2, VZV, CMVH) et/ou le VIH principalement dans le cas de rétinites et dans le cas de patients portant des souches résistantes à l'acyclovir (HSV et VZV).

Dû à une forte toxicité rénale, le PFA, bien qu'efficace, n'est prescrits qu'en cas d'intolérance ou d'échec aux traitements aux GCV et VGCV (Gilbert and Boivin 2005).

Le Formivirsen est un oligonucléotide de 21 unités dont les liens phosphates sont substitués par des groupements phosphorothioates ce qui limite sa dégradation par les nucléases cellulaires: **5'-d-[G*C*G*T*T*T*G*C*T*C*T*T*C*T*T*C*T*T*G*C*G]-3** (***** représentent les groupements phosphorothioates**)**. Cette séquence vise, par complémentarité, l'ARNm codant pour la protéine IE2 (*Immediatly Early protein 2*, protéine codée par le gène *UL123*) qui, rappelons-le, est un transactivateur positif et négatif des gènes viraux et cellulaires, respectivement. IE2 détourne ainsi la machinerie cellulaire au profit du cycle infectieux de CMVH. En bloquant sa synthèse, le Formivirsen bloque toute la cascade d'évènements qui en découle. Il est strictement réservé aux patients infectés par CMVH et administré par voie intravitréenne uniquement.

### A.3. Les résistances aux inhibiteurs de l'ADN polymérase virale

Les résistances croisées aux différents analogues nucléosidiques touchent aussi bien la polymérase pUL54 que la phosphotransférasepUL97 (Erice 1999; Baldanti, Lilleri et al. 2004; Ducancelle and Mazeron 2006; Marfori, Exner et al. 2007; Chou 2008).

Nous représentons dans la Figure 10 les différentes résistances observées sur pUL54 etpUL97 après traitement avec les différents inhibiteurs présentés.

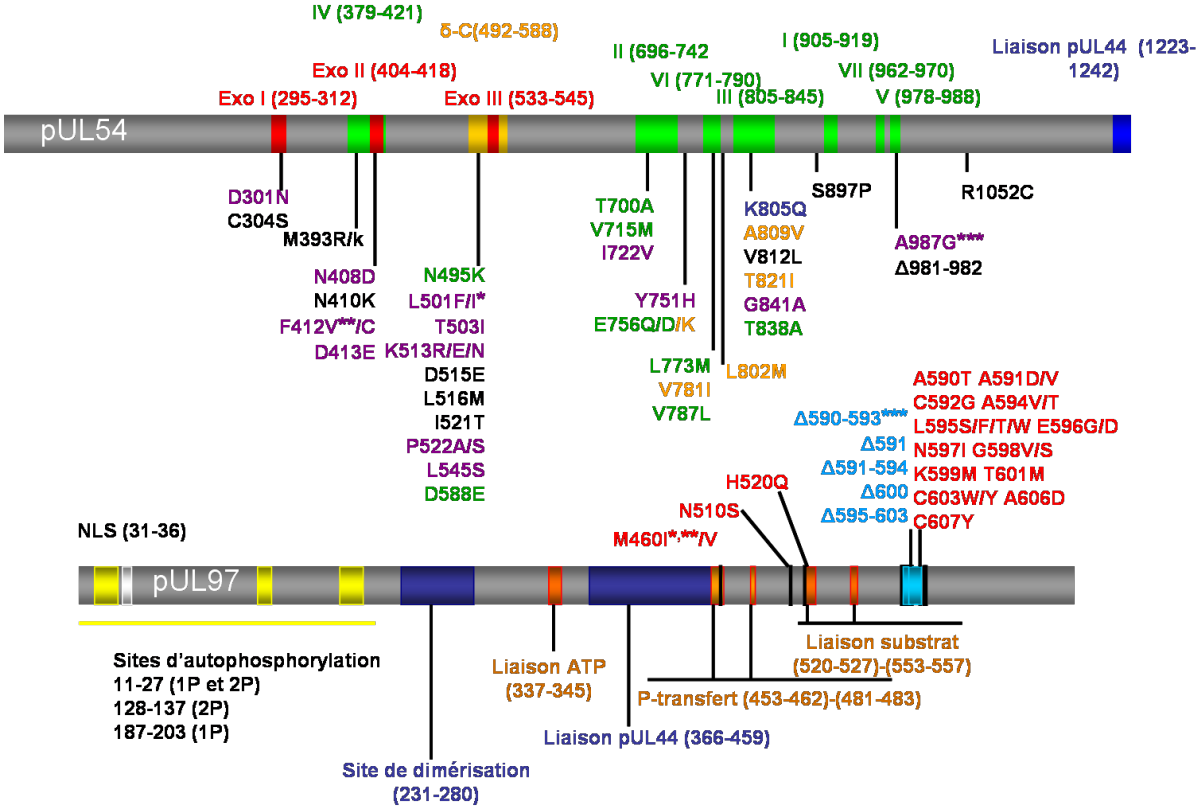



**Figure 10: Récapitulatif des domaines fonctionnels et mutations résistantes aux inhibiteurs de la synthèse virale localisés sur les séquences de pUL54 et pUL97.** Les domaines fonctionnels de pUL54 et pUL97 sont représentés au dessus et en dessous de leurs séquences, respectivement. Les résistances observées sont annotées entre les séquences. En rouge les résistance liés à GCV, en bleu à CDV, en vert à PFA, en violet à GCV+CDV, en orange à GCV+PFA, en noir à GCV+CDV+PFA. Δ symbolise les délétions observées chez les souches résistantes au GCV chez pUL97 (en cyan) et chez pUL54. * correspond aux mutation croisée M460I (pUL97) et L501I (pUL54); **M406I (pUL97) et F412V (pUL54) ; et ***Δ590-593 (pUL97) et A987G (pUL54).(Michel, Pavic et al. 1996; Erice 1999; Baldanti, Lilleri et al. 2004; Ducancelle and Mazeron 2006; Chou 2008)

Les premières mutations à apparaître sont localisées sur pUL97, indispensable à la phosphorylation du Ganciclovir. Toutes les études menées montrent une localisation très intense des mutations au GCV dans la partie 590-607 de pUL97 comprenant à la fois des mutations ponctuelles ou des délétions allant de un à près de 10 résidus. Ces souches résistantes empêchent la fixation du GCV pour sa primophosphorylation. 60 à 80% des souches résistantes au GCV se répartissent entre les mutants M460V/I, A594V et L595S et ne représentent à ce stade qu'un faible niveau de résistance.

Par suite d'un traitement prolongé, les mutations touchent pUL54 sur l'ensemble de sa séquence n'épargnant guère que les domaines I et VII. Celles-ci confèrent une résistance élevée au Ganciclovir souvent doublée d'une résistance au Cidofovir, localisées dans les domaines Exo I, II et III portant l'activité exonuclésique et le domaine δ-C commun aux herpesvirus (14 mutations dont 1 dans le domaine Exo III). Les mutations communes à GCV, CDV et PFA sont elles aussi plutôt concentrées entre les régions Exo I et δ-C. Des domaines distincts de résistances au Foscarnet sont les domaines II à III. Celles-ci peuvent apparaître à long terme après arrêt de son traitement et perte d'une résistance à GCV (Iwasenko, Scott et al. 2007). De manière surprenante, des résistances communes uniquement au PFA et au GCV, et non au CDV, sont observées dans ces mêmes régions. L'apparition de cette diversité de résistances peut être due à la cœxistence simultanée de plusieurs souches résistantes.

La conséquence de ces résistances est variable. Pour pUL97, il semble que ces mutations aient une faible répercution sur son activité enzymatique alors que celles affectant le domaines polymérasique de pUL54 pénalisent la vitesse de réplication virale (Alain, Mazeron et al. 2004) .

L'accumulation de ces résistances à la fois sur pUL54 et pLU97 à la suite de ces traitements a incité à identifier de nouvelles cibles thérapeutiques. Celles-ci visent préférentiellement le rôle central de pUL97 qui intervient dans plusieurs étapes de l'infection virale, sa protéine accessoire pUL27 vraisemblablement impliquée dans l'exportation nucléaire de la capside et un groupe de protéines impliquées dans la maturation et l'encapsidation de l'ADN viral. Cette dernière catégorie de cible thérapeutique est connue sous le nom de « terminases » dont pUL56 et pUL89 sont les représentants chez le cytomégalovirus humain et auxquelles s'ajoute la protéine portail, pUL104. Elles interviennent dans un processus biochimique qui ne trouve pas d'équivalent dans l'organisme humain ce qui constitue un avantage thérapeutique. Ces nouvelles stratégies ont de fait conduit au développement de nouvelles molécules spécifiques n'agissant pas sur le stade de réplication de l'ADN viral et donc pouvant être actives sur des souches résistantes aux antiviraux actuellement disponible.

## *B. Les Autres Cibles Thérapeutiques En Développement*

### B.1. Maturation et encapsidation

Le mécanisme de maturation et d'encapsidation commun aux herpesvirus et aux bactériophages prend place au centre de l'infection virale. Dans le CMVH, il fait intervenir

sept protéines: pUL51, pUL52, pUL56, pUL77, pUL89 et pUL104. Grâce aux études menées à la fois sur les bactériophages et les herpesvirus il a été possible de décrire cette étape-clé de la réplication virale (Visalli and van Zeijl 2003).

Le scénario qui va conduire à l'encapsidation de l'ADN viral débute par la synthèse en cercle roulant de l'ADN via le complexe pUL54/pUL44. Ainsi, commence la synthèse d'un double-brin d'ADN concatémérique mettant bout à bout une multitude d'unités virales.

En parallèle, les protéines de la capside sont exprimées dans le cytoplasme puis importées dans le noyau où elles vont s'assembler pour donner dans un premier temps une capside sphérique (capside de type A), puis icosaédrique vide (capside de type B), prête à loger une unité d'ADN viral. L'ADN concatémérique néosynthétisé va être reconnu, au niveau de ses séquences *pac* (*cis-acting packaging signal*) spécifiques encadrant ces unités d'ADN viral, puis débitée par le complexe terminase.

Le complexe terminase de CMVH est un hétérocomplexe formé par la grande sous-unité pUL56 et la petite sous-unité pUL89 (Scheffczik, Savva et al. 2002). Le complexe terminase va opérer un premier clivage de l'ADN (Bogner, Radsak et al. 1998) nécessitant la protéine pUL52 dont le rôle est encore indéterminé (Borst, Wagner et al. 2008). Puis il va exporter l'unité virale vers la protéine portail pUL104 (Dittmer, Drach et al. 2005) et transloquer l'ADN dans la capside par le biais d'une machinerie ATPasique (Lebedev, Krause et al. 2007). Enfin, l'excédant d'ADN va être excisé afin de terminer l'encapsidation. La capside contenant l'ADN est appelé type C.

La capside est alors dirigée vers la membrane interne du noyau précédant sa sortie du noyau. Le mécanisme analogue présenté dans le cas du bactériophage P22 est illustré à la Figure 11.

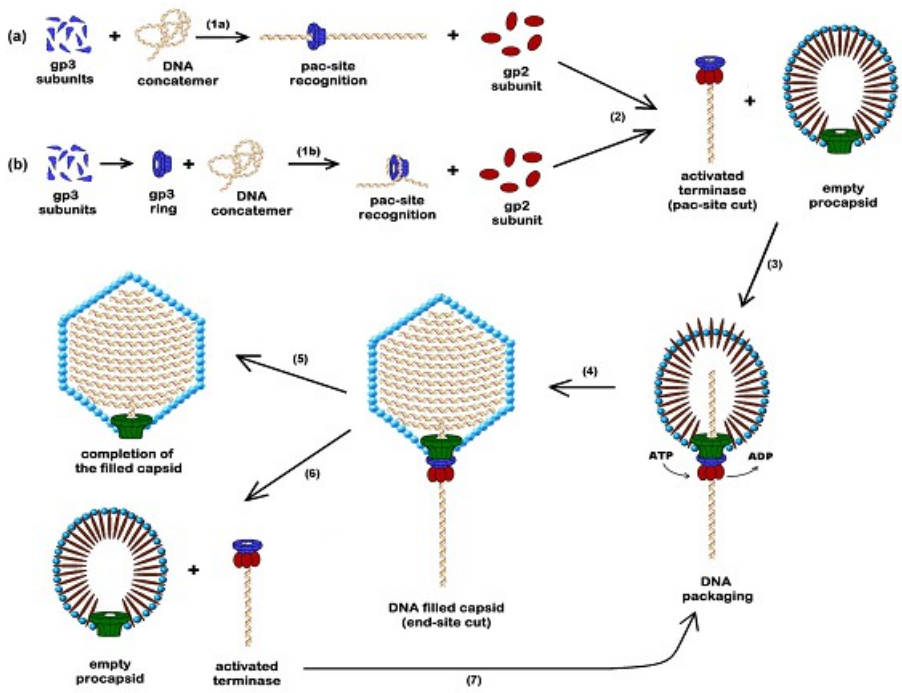



**Figure 11: Maturation et Encapsidation de l'ADN viral.** Exemple du bactériophage P22. Des monomères (**a**) de gp3 (petite sous-unité) ou une forme préassemblée en anneau (**b**) vont reconnaître l'ADN concatémérique (**1a** et **1b**). Une ou plusieurs unités de gp2 (grande sous-unité) sont recrutées par le complexe ADN/gp3 pour cliver l'ADN concatémérique en unité virale (Walter, Ealick et al.). Le complexe gp2/gp3/ADN se lie au complexe gp1 (protéine portail)/capside de type A (sphérique vide) et entame la procédure d'encapsidation sous l'effet de mécanismes ATP-dépendant (Biron, Harvey et al.). La capside devient de type B (icosaédrique vide) puis C (icosaédrique pleine) et remplie d'ADN compacté (104% de la taille du génome) (**4**). L'excédant d'ADN est coupé (**5**) et le complexe terminase relargué pour une nouvelle association avec une capside A (**6** & **7**) (Nemecek, Gilcrease et al. 2007).

### B.1.a. Le complexe Terminase: pUL89 et pUL56

| | Petite sous-unité | | Grande sous-unité | | |
|---|---|---|---|---|---|
| Phage/virus | Gene product | taille | Gene product | taille | ratio |
| λ | gpNu1 | 181 | gpA | 641 | 3,541 |
| T4 | gp16 | 164 | gp17 | 613 | 3,738 |
| T5 | ? | | gp155 | 438 | |
| φ29 | gp3 | 266 | gp16 | 332 | 1,248 |
| T3/T7 | gp18 | 89 | gp19 | 586 | 6,584 |
| P22 | gp3 | 162 | gp2 | 499 | 3,08 |
| SPP1 | gp1 | 184 | gp2 | 422 | 2,293 |
| HSV1 | pUL15 | 735 | pUL28 | 785 | 1,068 |
| HSV2 | pUL15 | 734 | pUL28 | 785 | 1,069 |
| CeHV1 | gp16 | 738 | gp30 | 785 | 1,064 |
| CeHV2 | gUL15 | 735 | gUL28 | 784 | 1,067 |
| EHV1 | ORF44 | 734 | ORF32 | 775 | 1,056 |
| BHV5 | pUL15 | 737 | ? | | 0 |
| HHV3 | ORF42/gp44 | 747 | ORF30/gp32 | 770 | 1,031 |
| MeHV1 | gp22 | 738 | gp36 | 787 | 1,066 |
| GaHV1 | gp22 | 764 | gp43 | 764 | 1 |
| GaHV2 | MD5V_027 | 737 | MD5V_041 | 793 | 1,076 |
| GaHV3 | gp28 | 748 | gp43 | 811 | 1,084 |
| SuHV1 | gp41 | 735 | gp13 | 724 | 0,985 |
| HCMV | pUL89 | 674 | pUL56 | 850 | 1,261 |
| PoHV4 | gp80 | 672 | gp58 | 839 | 1,249 |
| RhCMV | UL89 protein | 671 | pR56 | 893 | 1,331 |
| TuHV1 | T89 | 673 | T56 | 782 | 1,162 |
| HHV6-A | U66/U60 | 667 | ? | | 0 |
| HHV6-B | U66 | 666 | ? | | 0 |
| HHV7 | gp59 | 663 | gp36 | 721 | 1,087 |
| PL_HV1 | ORF 29 | 684 | ? | | 0 |
| PL_HV3 | ORF29 | 683 | ? | | 0 |
| HHV4 | BGRF1/BDRF1 | 690 | BALF3 | 685 | 0,993 |
| HHV8 | | 687 | ORF 7 | 695 | 1,012 |

**Tableau 4: Liste non exhaustive des sous-unités du complexe terminase chez quelques bactériophages et herpesvirus d'origines humaine et animale.** La taille des différentes protéines est exprimée en acides aminés. Les points d'interrogation signifient que les protéines n'ont pas été identifiées.

Bien que les terminologies « grande » et « petite » sous-unités restent discutables (Tableau 4), des homologies flagrantes ont été mises en évidence entre les différents représentants des terminases chez les bactériophages et les herpesvirus. Du fait même d'une fonction commune, clivage et empaquetage de l'ADN viral, des redondances structurales et fonctionnelles

réparties sur l'une ou l'autre des sous-unités restent hautement conservées quelle que soit la distance phylogénique les séparant.

La caractérisation de la famille des terminases réside dans la reconnaissance, le clivage et la translocation dans la procapside de l'ADN viral. Chez les bactériophages, la reconnaissance de l'ADN est assurée par la petite sous-unité alors que son clivage (fonction nucléasique) et sa translocation (fonction hélicase) sont accomplis par la grande sous-unité. Les deux mécanismes que sont le clivage de l'ADN concatémérique et sa translocation sont ATP-dépédendants. Ainsi, chez un grand nombre de bactériophages des motifs de type boîte de Walker sont systématiquement localisés proche des sites nucléase et hélicase (Rentas and Rao 2003; Mitchell and Rao 2006). Les terminases des bactériophages sont jusqu'à présent les plus étudiées. Par conséquent, la description fonctionnelle des terminases d'une manière générale est largement basée sur la connaissance des représentants des phages T4, λ et SPP1.

### *1.a.i. Reconnaissance de l'ADN*

Le phénomène d'encapsidation de l'ADN viral débute tout d'abord par sa reconnaissance alors qu'il est encore sous sa forme concatémérique après synthèse en cercle roulant. La structure du domaine N-terminal de gpNu1, sous-unité du bactériophage λ, a été résolue par RMN en 2002 par deBeer et al. (de Beer, Fang et al. 2002) (Figure 12). Cette structure (PDB: 1J9I) révèle des aspects intéressants tel qu'un repliement de type en hélice ailée (*winged helix protein*) analogue à celui rencontré dans le domaine de liaison à l'ADN de MuC repressor (Ilangovan, Wojciak et al. 1999) (PDB 1QPM).

La reconnaissance de l'ADN requiert la formation d'un dimère stable impliquant des résidus hydrophobes. Des résidus basiques situés dans un domaine HTH hélices contenus dans les hélices αA et αB vont quant à eux lier de manière spécifique l'ADN viral. Pour que gpNu1 reconnaisse les séquences R, celui-ci nécessite la présence de l'IHF (Integration Host Factor) cellulaire dont la fonction est de courber l'ADN viral au niveau de la séquence médiane entre

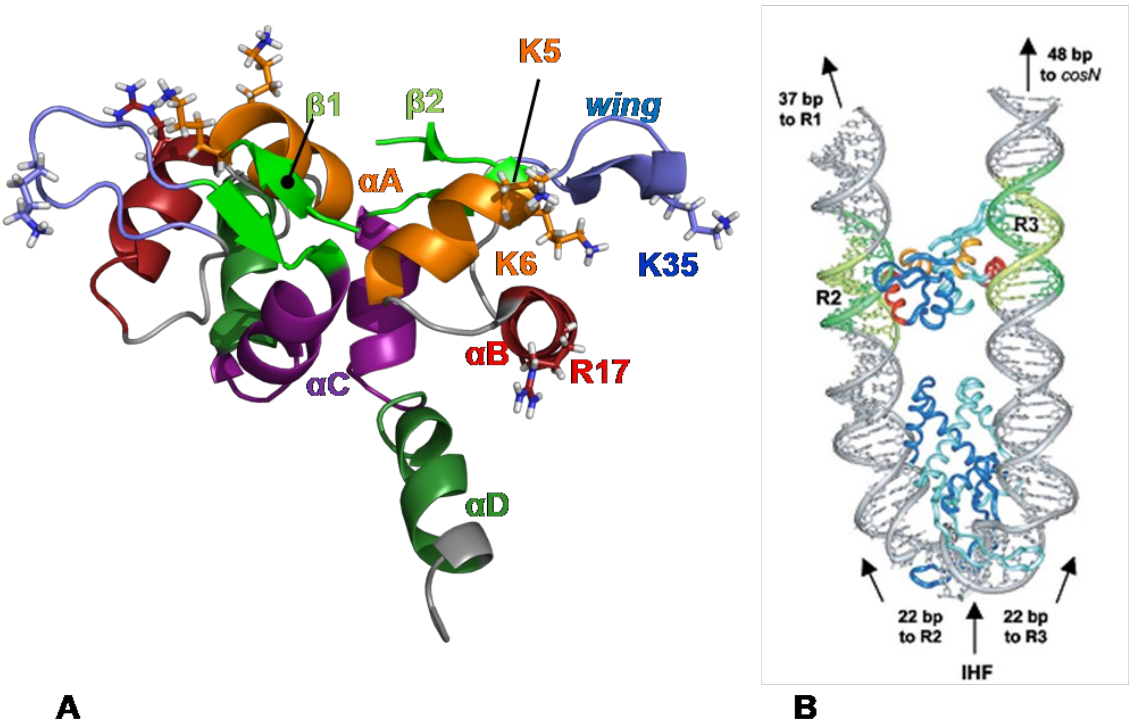


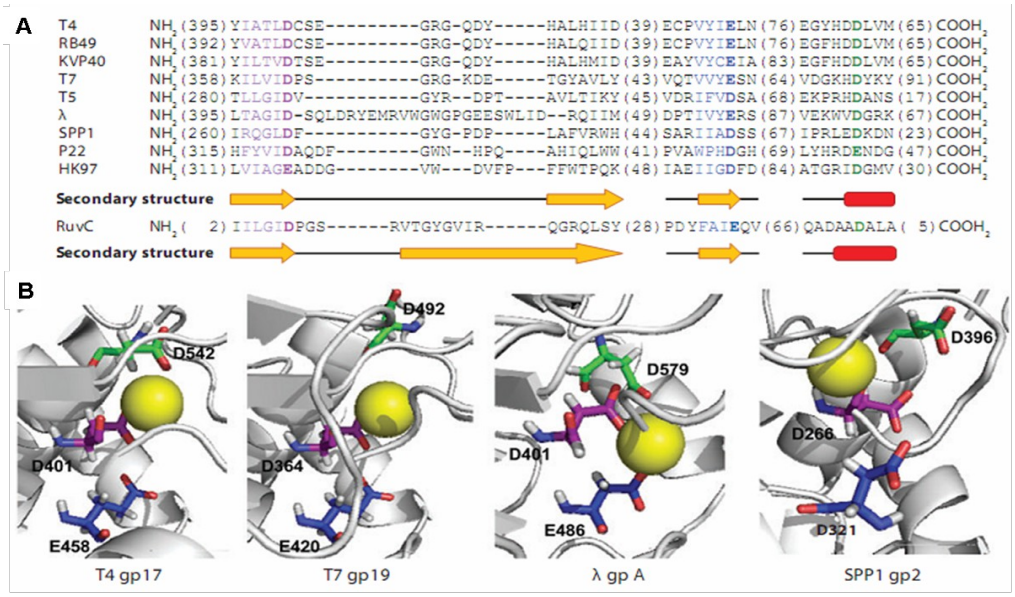

**Figure 12: A. Dimère de gpNu1.** Représentation des différents éléments de structures secondaires constituant le monomère de gpNu1: hélice αA (AA 5-12) en orange, hélice αB (AA 16-25) en rouge, hélice αC (AA 43-53) en violet, hélice αD (AA 53-60) en vert foncé, brins β1 (1-4) et β2 (40-43) en vert clair et domaine *wing* (AA 31-39) en bleu. La dimérisation se fait via des résidus aliphatiques des hélices αA et αC. Les résidus basiques impliqués dans la reconnaissance de l'ADN sont étiquetés et représentés en bâtons. Notons la présence de la lysine 35 située sur le domaine *wing* qui se loge dans le petit sillon de l'ADN. **B. Modèle de la reconnaissance de l'ADN par le dimère de gpNu1.** Grâce à l'intervention de la protéine de l'hôte Integration Host Factor qui va courber l'ADN viral au niveau de la séquence médiane entre R2 et R3, gpNu1 va reconnaître simultanément R2 et R3 en mettant en jeu un monomère pour chaque séquence.

R2 et R3 (Kosturko, Daub et al. 1989). L'intervention de cette protéine a aussi été montré dans le mécanisme de maturation du phage P21 (Feiss, Frackman et al. 1985).

De manière surprenante, ce site de liaison à l'ADN n'a montré aucun motif canonique de type P-loop malgré une activité ATP-asique faible mais détectable. Ceci implique deux

hypothèses, soit ce domaine est contenu sur un autre domaine de gpNu1 soit il est contenu dans son partenaire gpA.

Par ailleurs les études faites sur l'homologue de gpNu1 chez le bactériophage T4, gp16, confirme cette étrange constatation sur l'absence d'un motif de type Walker A ou B (Mitchell, Matsuzaki et al. 2002). Dans ce dernier cas, il a aussi été montré une oligomérisation de 8 unités de gp16 via un centre hydrophobe, et une reconnaissance de l'ADN par le même motif HTH décrit ci-dessus. Nemecek et al. ont quant à eux reconstruit le nonamère de gp3, petite sous-unité du phage P22, par microscopie électronique (Nemecek, Lander et al. 2008). Cette protéine semble s'auto assembler en forme d'anneau au centre de laquelle l'ADN double-brin entrerait en interaction de manière aspécifique (*in vitro*) via des interactions entre charges complémentaires impliquant un fragment C-terminal 143-152 riche en résidus basiques et les charges négatives des phosphates de l'ADN.




*1.a.ii. Fonction nucléase*

**Figure 13: Domaine nucléasique chez différents bactériophages. A. conservation de la triade Asp/Glu à travers les homologues de gp17 analogues à RuvC.** Le résidu aspartate conservé (en violet) est localisé à la pointe d'un brin bêta conservé (structure secondaire représentée sous l'alignement de séquence), le second Asp (en bleu) est localisé 60 acides aminés en aval, à l'extrémité d'un court brin bêta conservé, et le troisième Asp (en vert) est proche d'une hélice. **B. Modèles générés sur chaque protéine précisée grâce à l'alignement de séquence des 150 acide aminés comprenant dans chaque cas les trois résidus acides, basés sur la structure 1HJR de RuvC.** En bâtonnets sont représentés les résidus acides et en sphère jaune l'ion Mg2+ impliqué dans la catalyse (Rao and Feiss 2008).

La reconnaissance de l'ADN étant effectuée, il reste à le couper et à le transloquer. Ces deux fonctions vont être remplies par la « grande sous-unités ». Là encore les exemples des bactériophages T4 et λ sont des systèmes permettant d'illustrer notre propos.

L'activité nucléasique du complexe terminase est localisée chez le phage T4 au niveau de l'extrémité C-terminal de gp 17 au sein du motif $D^{401}X_2E^{404}X_3D^{409}$ situé entre deux brins bêta et à proximité d'un motif de liaison à un métal riche en histidines (région $H^{382}$-$H^{436}$) (Mitchell, Matsuzaki et al. 2002). Rentas et al. montrent d'une part que l'activité nucléasique met principalement en jeu l'aspartate 401 et que les résidus H436 et C402 sont indispensables pour l'interaction avec l'ADN (Rentas and Rao 2003). Ce dernier point explique l'observation de Kuebler et al qui ont observé un packaging et un clivage défectueux des souches comportant la mutation H436R (Kuebler and Rao 1998).

Ce domaine reste largement conservé à travers les bactériophages comme le montre l'alignement de séquence présenté en Figure 13A et semble appartenir à la même famille que RuvC, à savoir la famille des RNAseH/ résolvase de type feuillet bêta déjà suggéré par ailleurs (Ponchon 2005).

### *1.a.iii. Fonction hélicase*

Pour revues (Tuteja and Tuteja 2004; Singleton, Dillingham et al. 2007)

L'évènement suivant, la translocation de l'ADN, fait intervenir la fonction hélicase conservée des terminases (Figure 14). Celle-ci est reconnaissable par la présence de plusieurs éléments qui agissent synergétiquement pour l'accomplissement de la translocation de l'ADN viral dans la capside.

Les hélicases sont des protéines (ou sous domaines de protéines) impliquées dans différents mécanismes cellulaires ou infectieux mettant en jeu des interactions protéines/acides nucléiques. Aussi, cette fonction est elle représentée lors de processus de réparation, réplication, transcription, traduction de l'ADN, maturation et épissage d'ARN ou encore dans l'exportation nucléaire. Le nom d'hélicase se rapporte à la capacité de ces enzymes à aider à la formation ou la déformation des hélices d'ADN par le biais d'un moteur ATPasique. Cependant, à l'aide d'un même domaine catalytique, les translocases -enzyme aidant à la translocation- sont capables de transformer l'énergie chimique produite par l'hydrolyse d'une liaison phosphodiester d'ATP en énergie mécanique se traduisant par des changements conformationnels induisant le déplacement de l'ADN par rapport à l'enzyme. La question du classement des hélicases a été abordée dans un premier temps par Gorbalenya &

Koonin en 1993 qui ont put distinguer à partir de l'analyse de séquences trois familles et 2 sous familles d'hélicases. Il s'est finalement avéré indispensable de revoir cette classification au regard des résultats obtenus pour des protéines de type hélicase n'étant pas chargées de débobiner l'ADN mais de le transloquer.

De ce fait, en 2007 une nouvelle classification a été établie au regard des différentes informations structurales et biochimiques collectées depuis lors. Sept superfamilles ont découlé de cette analyse. Dans les deux cas, hélicase pure et translocase, la nature du substrat (ADN, ARN, hybride ADN/ARN, double ou simple-brin) et la directionnalité de l'enzyme (3'-> 5' ou 5'-> 3') sont variables selon le mécanisme biologique dans lequel est impliquée l'enzyme. La nomenclature adoptée définit de type $\alpha$ et $\beta$ les hélicases reconnaissant des ADN (ou ARN) simple et double-brin, respectivement ; et A et B selon les sens 3'-> 5' et 5'-> 3', respectivement.

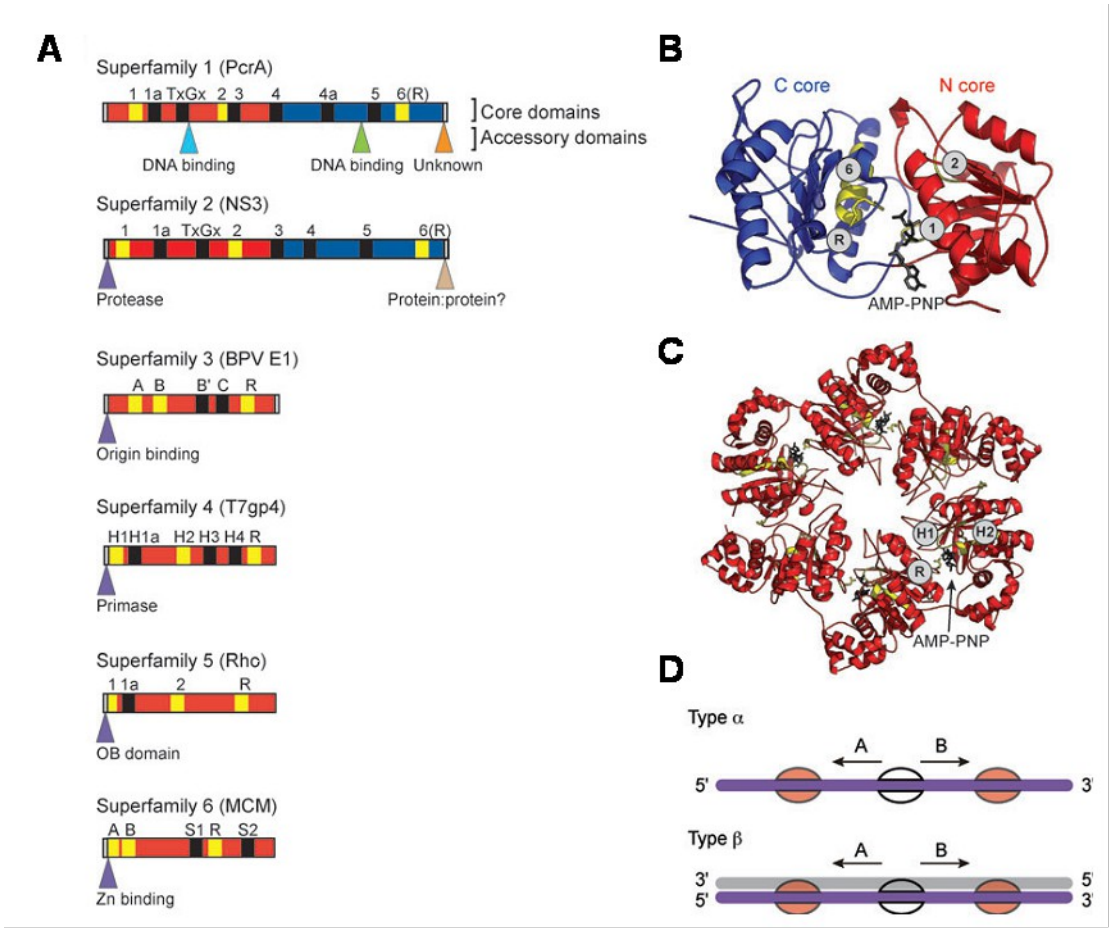



**Figure 14: Les différentes superfamilles d'hélicases.** A. Les différents motifs de six super familles d'hélicases. Les sous domaines RecA-like sont colorés en rouge et bleu respectivement et les différents motifs caractéristiques à la fonction hélicase annotées. B. PcrA le code couleur utilisé est celui de A. En stick noir est représentée une molécule d'ATP.

C. Exemple de translocase toroïdale hexamèrique de la superfamille SF4. D. Nomenclature utilisée pour les hélicases: type α pour les hélicases de mono-brin et β pour les double-brin, type A dans le sens 3'->5' et B dans le sens 5'->3' (Singleton, Dillingham et al. 2007).

Le type de repliement canonique est constitué de deux sous-domaines RecA-Like – alternation de brins bêta et d'hélices alpha formant un feuillet bêta flanqué d'hélices de part et d'autre- constituant la poche de liaison au NTP. Les motifs la constituant, en particulier les boîtes de Walker A et B, ainsi qu'un doigt d'arginine (*arginine finger*) sont retrouvés dans toutes les sous-familles d'hélicase/translocase. Quelques distinctions entre les superfamilles sont tout de même faites.

La superfamille 1 (SF1) se décompose en deux sous familles, SF1A et SF1B, toutes deux de type α (sens 3'->5'). PcrA, représentant de SF1A, déstabilise le duplexe d'ADN par le rapprochement de deux sous-domaines après hydrolyse d'ATP ce qui va conduire à la constitution d'une surface négativement chargée déstabilisatrice des appariements de paires de bases menant au débobinage du duplex. Ce mécanisme est appelé *inchworm model*. Malgré les données structurales disponibles concernant RecD, et biochimiques au sujet de l'hélicase Dda, représentantes de la famille SF1B, le mécanisme des SF1B reste peu clair.

SF2, adoptant majoritairement la directionnalité 3'->5' (type A) comprend le plus grand nombre d'hélicases présentes dans un nombre varié de processus cellulaires. Elle contient les sous-familles DEAD-box RNA hélicase, Rec-like et Snf2-like. Elles peuvent transloquer de manière ATP-dépendante des substrats ADN simple ou double-brin. L'exemple de NS3 de l'hépatite C, illustre la capacité de cette enzyme conçue strictement pour lier un ADN simple-brin à la translocation. Cette hélicase de type SF2Aα met en jeu ses motifs 1a, TxGx, 4 et 5 (Figure 32A) dans la liaison de l'ADN en conformation B en travers des deux sous-unités RecA-like via les groupements phosphates du *squelette* ainsi que deux résidus hydrophobes qui vont agripper cinq plateaux de bases. L'hydrolyse de l'ATP va ici encore provoquer l'éloignement des deux sous-domaines RecA-like induisant la translocation du brin d'ADN. La translocation d'un duplexed'ADN pour des hélicases (translocase dans ce cas) SF2β trouve un exemple dans la PDB avec la structure de Rad54 de *Sulfolbus solfataris*, membre de la sous-famille Snf2-like. Sa structure fait apparaître un tandem de domaines RecA-like (N et C cores) et deux insertions de fonctions inconnues. Le N core reconnaît les deux brins d'ADN par le petit sillon. Un des deux brins est toutefois lié de manière équivalente au simple-brin décrit dans le cas précédent de NS3, et subi la majorité des interactions avec l'enzyme. Un

mécanisme semblable à PcrA (*inchworm model*) permettrait d'expliquer la rotation d'un duplexesans son déroulement conduisant à sa translocation.

Les autres superfamilles d'hélicase (SF3 à7) sont pour la plupart toroïdales et constituées d'hexamères (Figure 14C), contrairement aux hélicase appartenant à SF1 et SF2 qui sont monomériques ou dimériques, s'auto assemblant autour de l'ADN et peuvent être responsable du déroulement de quelques kilos bases à plusieurs méga bases. Cependant le contrôle de l'assemblage des hexamères est important puisque une activité hélicase trop intense peut être catastrophique pour la double hélice d'ADN. Ces différentes familles se distinguent chacune par des boîtes qui leur sont propres. Par exemples, A, B, A', B' et C sont attribués à SF3 ; H1, H1a, H2, H3 et H4 pour SF4 (Figure 14A). La symétrie de l'anneau a permit d'envisager plusieurs mécanismes de translocations (Figure 15).

L'un d'eux mentionne une alternance de sites contenant NTP, NDP et vide et impliquant un positionnement différent de boucles basiques interagissant avec l'ADN: boucle proche de l'extrémité 5' pour les unités contenant un NTP, à un niveau intermédiaire pour les porteuses de NDP et proche de l'extrémité 3' pour les sites vides. Avec l'alternance des sites chaque boucle monte d'un cran selon l'état d'avancement de la réaction d'hydrolyse du NTP créant un mouvement cyclique poussant l'ADN dans une direction. Ce mécanisme a deux déclinaisons faisant intervenir soit 3 sites sur 6 (Figure 15A) soit les 6 sites disponibles (Figure 15B). Un autre mécanisme, concerté, fait évoluer les 6 sites en même temps de la même façon: tout-NTP->tout-NDP->tout-vide (Figure 15E).

Les différents mécanismes sont représentés dans la Figure 15 et l'adoption de tel ou tel model dépend de la nature (RNA ou DNA) et du type (simple ou double-brin) de substrat.

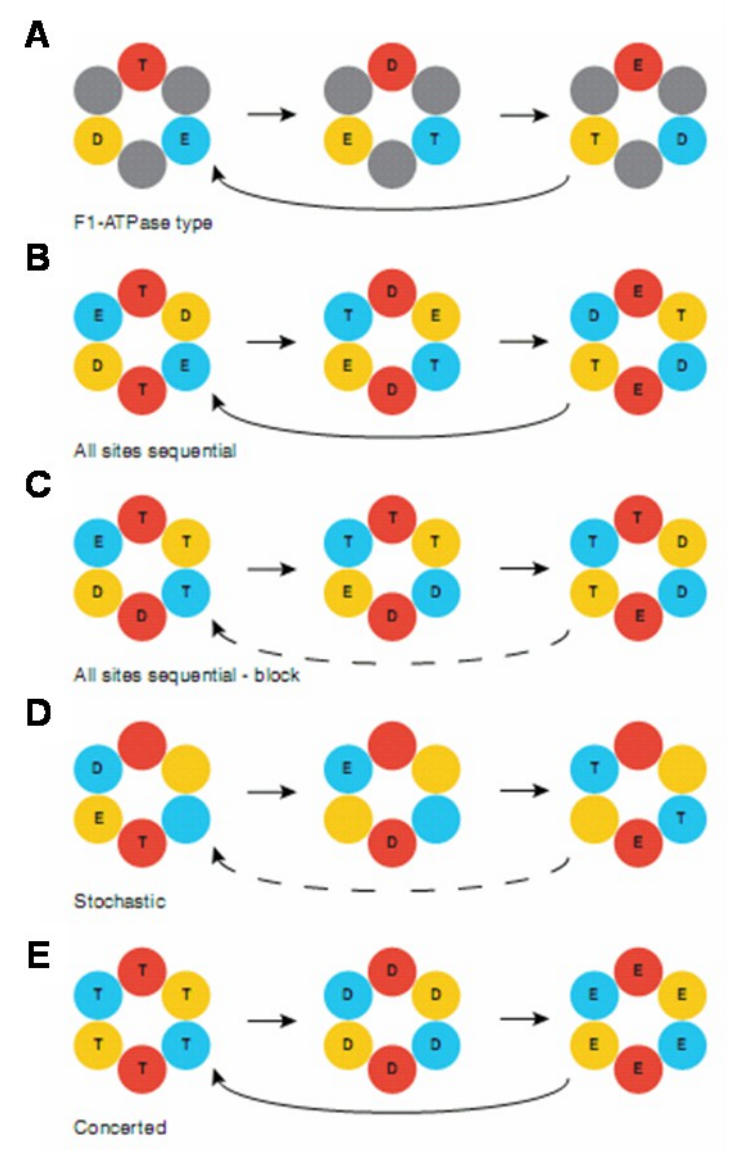



**Figure 15: Modes de translocation des hélicases/translocases hexamèriques.** Dans lez différents mécanisme la notation suivante est adoptée: E pour un site vide, D pour un site contenant du NDP et T pour un site contenant du NTP. Une alternace de coloration représente les différentes unités de l'hexamère. Les flèches pleines correspondent à de smécanisme à 3 étapes relmiant le dernier et le premier état (A, B et E). Les flèches en pointillés indiquent des mécanismes contenant des étapes intermédiaires (C et D). (Singleton, Dillingham et al. 2007)

**Figure 16: Alignement de séquence de plusieurs grandes sous-unités de divers bactériophages montrant la conservation des motifs impliqués dans l'activité hélicase.** Notez la présence de l'homologue de HSV-1 (10° ligne, au dessus de la prédiction de structure secondaire)(Rao and Feiss 2008).

Le domaine hélicase comprend une boîte de Walker A correspondant au motif consensus G/A-X4-GK-T/S. Chez gp17, la lysine 166 maintient les groupements phosphate β et γ de la queue polyphosphate de l'ATP alors que le groupement hydroxyle de la thréonine 167 chélate l'ion Mg2+ apporté par l'ATP.

Le site de liaison à l'adénine de l'ATP situé environ 20 à 30 acides aminés avant la lysine de la boîte de Walker A, est composé du doublet YQ où la tyrosine peut être remplacée par un autre aromatique qui va interagir par π-stacking avec la base nucléotidique alors que la glutamine semble strictement conservée pour former une liaison hydrogène entre l'oxygène de son squelette avec les azotes N6 et N7 de l'adénine.

Une boîte de Walker B identifiable par le cluster ZZZZD (où Z représente des résidus hydrophobes contenus dans un brin bêta) est rencontrée 50 à 130 résidus après la lysine de la boîte de Walker A. L'aspartate complètement conservé va activer une molécule d'eau qui va coordonner à son tour l'ion Mg2+ de l'ATP. Directement adjacent à l'extrémité C-ter de ce

domaine, un résidu glutamate est parfaitement conservé. Son rôle est d'activer une autre molécule d'eau qui, par attaque nucléophile sur le phosphate γ, va procéder à l'hydrolyse de la liaison phosphodiester. Cette hydrolyse est rendue possible par la présence d'une arginine ou d'une glutamine situé dans la boîte de Walker A appelé *switch résidu* chargé de tendre la queue polyphosphate par liaison ionique avec son groupement γ.

Enfin un troisième motif de la forme T/S-G/A-T/S situé entre 20 et 30 acides aminés en aval de la boîte de Walker B fait le lien entre la molécule d'ATP en cours d'hydrolyse et l'ADN double-brin prêt à être transloqué dans la capside (Velankar, Soultanas et al. 1999) (Figure 16).

En 2007, le double mutant de gp17 ayant subit l'inversion ED au niveau des résidus 255-256 (en violet Figure 17A et B), a été cristallisé sous ses formes apo-, complexé avec l'ATP et complexé avec l'ADP (Sun, Kondabagil et al. 2007) . D'une manière générale, la structure obtenue est constituée de deux domaines. Le premier correspond au folding canonique des hélicases contenant un feuillet bêta de topologie β3-β2-β4-β5-β1-β6. β1 précède la boîte de Walker A, β2 et β3 contiennent la boîte de Walker B et β4 précède le motif III (boîte $^{285}$TTT$^{287}$) le tout formant le cœur du centre ATP-asique. Le sous domaine II est potentiellement une composante du centre nucléasique de gp17.

Dû aux artefacts inhérents à la mutation effectuée afin de rendre la protéine et ses complexes cristallisables, l'ATP dans ces conditions n'est pas hydrolysée en ADP + Pi mais la différence de flexibilité au niveau de la boucle 125-140 (en vert pour la forme liée et en rouge pour la forme libre sur la Figure 17A) reflète la capacité à lier l'ATP de gp17. Sun et al. émettent l'hypothèse que le mouvement de cette boucle cause la conversion d'énergie chimique (hydrolyse) en énergie mécanique après relargage du substrat à l'instar d'une catapulte. Cette énergie mécanique serait alors utilisée pour la translocation de l'ADN dans la procapside. Par ailleurs, l'inactivité du double mutant ED est expliquée par l'inversion des longueurs des chaînes latérales des résidus 255 et 256 normalement impliquées dans des ponts hydrogènes avec des molécules d'eau, importantes pour l'activation du substrat lors de la catalyse (Figure 17B et C). Cette mutation a pour conséquence le décalage de 3 Å de la molécule d'ATP par rapport à son site de fixation théorique.

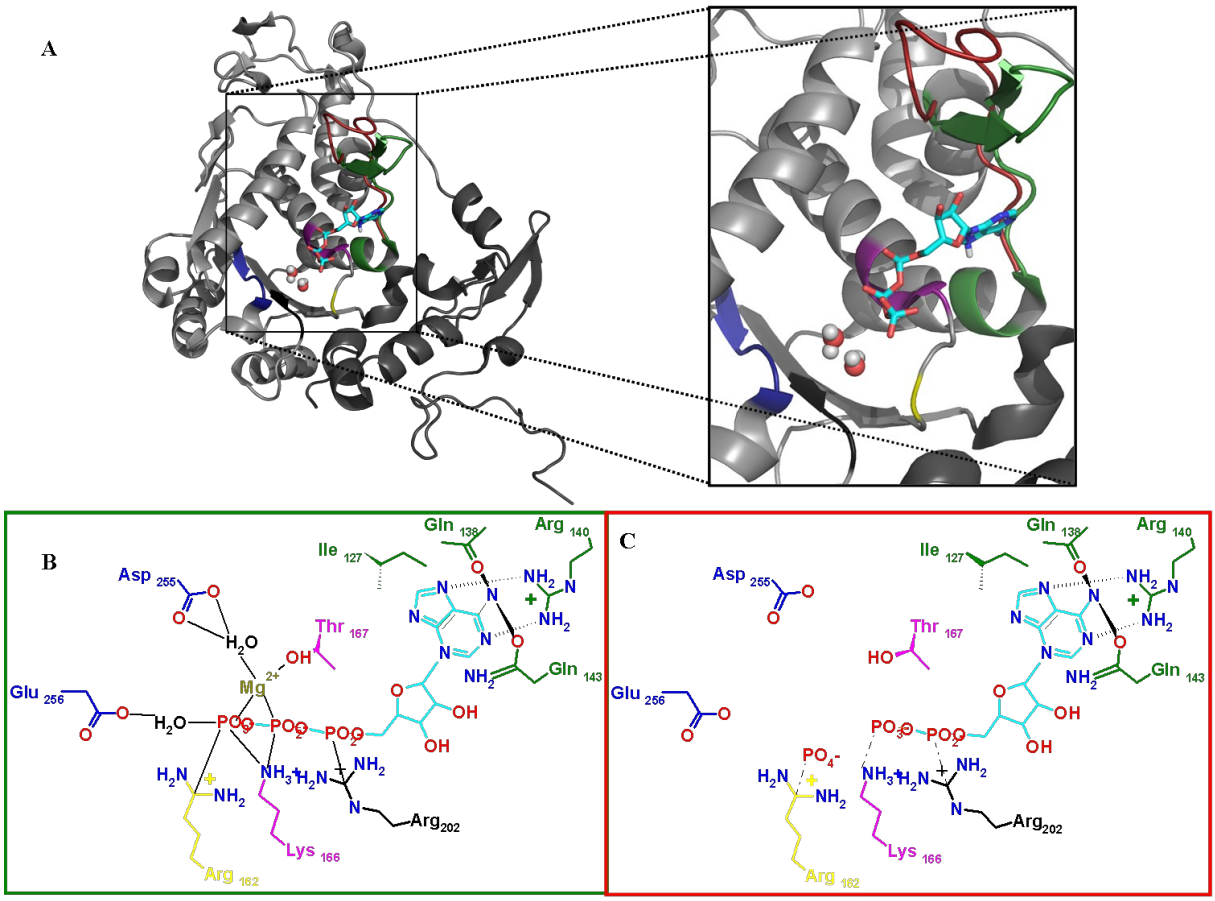



**Figure 17: structure du double mutant DE255ED de gp17. A. Structure de gp17.** Les deux sous-domaines sont représentés en gris clair et gris foncé, respectivement. A droite, zoom du domaine ATP-asique: l'ATP en cyan, les molécules d'eau en sphères, la boîte de Walker A en violet (arrière plan), la boîte de Walker B en bleu, le motif III en noir, le *switch* residue en jaune et le site de liaison à l'adénine comprenant la boucle 125-140 mobile en vert pour la forme liée et en rouge pour la libre. **B et C. Mécanisme réactionnel de l'hydrolyse de l'ATP correspondant aux formes liée et libre**, **respectivement.**

Nous venons de présenter les fonctions imputées jusqu'alors aux terminases des bactériophages. Comme le montre, pour le moins, l'alignement de séquences des différents domaines hélicases identifiés chez les bactériophages (Figure 16), une homologie est identifiée avec pUL15 de HSV-1. De plus, l'activité terminase étant observée chez les deux types d'organismes il semble donc que les bactériophages et les herpesvirus descendent d'un ancêtre commun. Par extension, ceci présage des fonctions similaires chez les autres sous-unités terminases des herpesvirus en général, et de CMVH en particulier. Comme le montre le Tableau 4, les rapports des tailles des sous-unités chez les bactériophages d'une part et chez les herpesvirus d'autre part sont bien différentes. En effet, chez les bactériophages, la grande sous-unité est environ 2 à 3 fois plus grande que la petite, alors que chez les herpesvirus se rapport est très proche de 1 voir inférieur. Ceci pourrait être expliqué par une répartition différente des fonctions de reconnaissance et de catalyse entre les deux sous-unités du complexe terminase. Par conséquent, il n'est pas étonnant que les fonctions décrites chez les bactériophages se retrouvent sur une sous-unité différente chez les herpesvirus.

Le complexe terminase de CMVH est constitué des sous-unités pUL56 et pUL89 portant chacune une fonction ATPasique et nucléasique.

La protéine pUL56 est une protéine de 850 acides aminés (95,8 kDa) homologue de pUL28 (ICP18.5) de HSV-1 au niveau des régions N- et C- terminales à hauteur de 43.2%. Elle a été colocalisée avec des protéines de la nucléocapside et du tégument sous une forme suggérant une modification post traductionnelle (autre qu'une glycosylation) du fait d'un poids moléculaire observé de 130 kDa. Ainsi, il fut proposé que pUL56 établissait des interactions avec les dites protéines de la nucléocapside (Bogner, Reschke et al. 1993). Par homologie avec pUL28 de HSV-1, il a été par la suite proposé que pUL56 devait avoir une fonction commune avec celle-ci, incluant le clivage et l'empaquetage de l'ADN (Newcomb, Homa et al. 1994; Thomsen, Roof et al. 1994).

La reconnaissance de l'ADN, pour son clivage postérieure, est assurée par la présence de séquences *a* situées entre les unités longue (UL) et courte (US) du génome. Ces séquences *a* sont conservées chez les herpesvirus (Spaete and Frenkel 1985) et sont constituées de deux motifs *pac* (*cis-acting packaging signal)* (Spaete and Mocarski 1985) riches en nucléotides AT encadrés par des régions GC: *pac1* (5'-TAAAAA-3') et *pac2* (5'-TTTTAT-3'). Effectivement, la reconnaissance de ces séquences d'ADN par pUL56 fut prouvée et accompagnée de la caractérisation d'une fonction nucléasique utile à la conversion de l'ADN concatémérique en unités virales (Bogner, Radsak et al. 1998).

L'implication de pUL56 dans les étapes de la réplication virale a été clairement mise en évidence en 2000 par Giesen et al. réussissant à colocaliser pUL56 avec les protéines pUL112-113 et à la coimmunoprécipiter avec pUL44 déjà connues comme étant impliquées dans la réplication virale (Giesen, Radsak et al. 2000). La même année, la même équipe souligne l'importance de son domaine C-terminal en vue de sa translocation dans le noyau. Ce mécanisme fait intervenir la voie de signalisation de l'importine α (Gorlich 1997) reconnaissant son segment $^{816}$RRVRATRKRPRR$^{827}$ riche en résidus basiques et contenant une séquence consensus de la forme K(K/R)X(K/R) (Chelsky, Ralph et al. 1989). pUL56 est donc importée grâce à son seul motif de signalisation nucléaire et n'a donc pas besoin d'autre partenaire pour se lier à l'importine. Des expériences d'alanine scanning ont indiqué le rôle crucial des acides aminés $R^{822}$ et $K^{823}$ dans ce mécanisme (Giesen, Radsak et al. 2000). De la même façon, chez HSV-1, le complexe terminase se formant au niveau du cytoplasme va être incorporé dans le noyau grâce au NLS $^{185}$KKR$^{186}$ porté par pUL15 qui entraîne alors ses deux autres partenaires, pUL28 et pUL33 (Yang, Homa et al. 2007).

Hwang et al montrent en 2002 que pUL56 contient une activité ATPasique qui peut être exacerbée par son association avec pUL89 via sa moitié C-terminal (Hwang and Bogner 2002). Cet aspect renvoie à une dépendance énergétique du clivage de l'ADN concatémérique par le complexe terminase déjà décrit dans le cas des bactériophages (Mitchell, Matsuzaki et al. 2002) ; et à la translocation des unités d'ADN viral dans la procapside. L'activité ATPasique de pUL56 a tout d'abord été présumée être localisée dans les segments $^{462}$RARGGGKK$^{470}$ (ATP-bs1) et $^{709}$YNETFGKQ$^{716}$ (ATP-bs2). Il s'est finalement avéré que des mutations multiples dans ATP-bs1 n'altéraient en rien la dégradation de l'ATP par pUL56. Au contraire, la simple mutation G714A contenue dans ATP-bs2 provoquait une perte significative de l'activité ATPasique (65%), chose non observée pour le mutant K715N. Le double mutant G714A/K715N provoque une perte de 75% de l'activité. A l'inverse, le mutant E711A amplifie l'hydrolyse de l'ATP. Par la suite, en fusionnant le seul domaine ATP-bs2 à la GST en présence d'ATP, il fut observé une hydrolyse du substrat indiquant bien que ce motif seul est suffisant à l'activité ATPasique *in vitro* de pUL56 (Scholz, Rechter et al. 2003).

Concernant le rôle de pUL56 dans la reconnaissance de l'ADN, une hypothèse basée sur des alignements de séquences, a été émise par Krosky et al. (Krosky, Underwood et al. 1998). En effet, la comparaison de pUL56 avec quatre homologues provenant de HSV-1, VZV, HHV-6 et HHV-7 a montré la présence d'un motif parfaitement conservé $CX_2CX_{24}CXH$. Cette observation soulève alors l'idée probable de l'existence d'un doigt de zinc pouvant

jouer un rôle dans la reconnaissance de l'ADN, et en particulier au niveau des séquences *pac*. De plus une souche résistante Q204R aux benzimidazole-D-ribonucléosides, inhibiteurs en développement de l'étape de maturation de l'ADN, suggère pour le moins un rôle de ce domaine dans le mécanisme d'action de ce type de molécule.

Par ailleurs, pUL56 a été coimmunoprécipitée avec la protéine portail pUL104. Cette dernière interagit préférentiellement avec la moitié C-terminal (404-850) de pUL56 plutôt qu'avec pUL89 entier (Dittmer, Drach et al. 2005). Entre temps, la première structure de pUL56 a été obtenue en 2004 par microscopie électronique (Figure 18) et a révélé une structure toroïdale en accord avec la forme annulaire redondante des protéines impliquées dans l'empaquetage des ADN viraux (Scheffczik, Savva et al. 2002; Savva, Holzenburg et al. 2004).

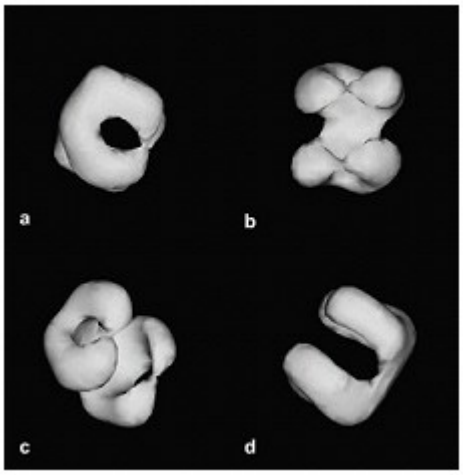



**Figure 18: Reconstruction tridimensionnelle par microscopie électronique de pUL56** (d'après Savaa et al., 2004).

La « petite » sous-unité du complexe terminase de CMVH, pUL89 est une protéine de 650 acides aminés (75kDa), homologue de gp17 du bactériophage T4 et de pUL15 de HSV-1 (Underwood, Harvey et al. 1998). C'est après une exposition à des inhibiteurs de la maturation et de l'encapsidation que le rôle de pUL89 dans ces étapes a été avéré. Les travaux de Scheffczick puis de Thoma ont montré, tout comme pour pUL56, que pUL89 était localisée au niveau du centre de réplication et ce dès la phase précoce.

Il semblerait que pUL89 existe aussi sous forme de dimère adoptant une structure toroïdale (Scheffczik, Savva et al. 2002; Thoma, Borst et al. 2006). De plus, son interaction avec puL56 a été vérifiée et montrée comme se faisant grâce à son extrémité C-terminal tout comme son homologue pUL28 vis-à-vis de pUL15 chez HSV-1 (Thoma, Borst et al. 2006;

Yang, Homa et al. 2007). Afin de préciser le domaine d'interaction de pUL89 avec pUL56, différents mutants de délétions fusionnés à la GST ont été exprimés et soumis à une filtration sur colonne contenant des billes greffées en pUL56-Cter (450-850) marqué en méthionine-[ $S^{35}$]. Les radiographies résultantes de cette expérience montrent que le fragment $^{580}$GRDKALAVEQFISRFNSGYIK$^{600}$ constitue le domaine minimum assurant l'interaction entre les deux sous-unités du complexe terminase. Toutefois, le domaine 580-674 montre une affinité encore plus élevée avec pUL56 (Thoma, Borst et al. 2006).

### B.1.b. **La protéine portail: pUL104**

La protéine pUL104 est une protéine de 697 acides aminés (75kDa) très conservée chez les herpesvirus. Cette protéine joue un rôle central dans le mécanisme d'encapsidation des unités d'ADN viral et colocalisée avec pUL56 (Dittmer, Drach et al. 2005). pUL104 est l'homologue de la protéine pUL6 de HSV-1.

Sa structure tridimensionnelle reste inconnue mais d'autres protéines portail notamment des bactériophages Φ29 (Guasch, Pous et al. 2002) et SPP1 (Lebedev, Krause et al. 2007) ont été résolues par diffraction des rayons X. Ces structures montrent l'association multimérique d'une douzaine d'unités autour d'un axe de symétrie formant ainsi un énorme édifice moléculaire de forme toroïdale comme représentée dans les Figures 19A et B. L'association de ces unités se fait grâce à des interactions hydrophobes au travers de leucines zippers (appariement d'hélices par leurs faces hydrophobes) comme cela a été observé pour pUL6 (Nellissery, Szczepaniak et al. 2007). Cependant, les études menées sur les structures 3D de gp6 de SPP1 et gp10 de Φ29 contrastent avec cette conclusion.

En effet il a été observé que la majorité des interactions entre sous-unités du complexe étaient largement localisées au niveau du *clip* sous l'aspect d'un enchevêtrement de deux brins bêta d'une sous-unité avec un brin de l'unité adjacente. Le reste de l'interaction est assurée par l'appariement de charges des deux unités voisines D'autres part l'interface entre ces sous-unités est légèrement inclinée par rapport à l'axe de symétrie ce qui provoque une interpénétration des monomères.

Un motif particulièrement conservé est le domaine formé par la longue hélice α6 contenant une torsion et reliée à l'hélice α5 par la boucle du tunnel (*tunnel loop*). L'hélice α5 est elle-même en interaction avec α3 par des interactions hydrophobes qui aident à stabiliser cette conformation particulière (Figure 19C). La conservation de ce domaine très particulier entre gp6 et gp10 (Figure 19D) ne fait aucun doute sur l'importance de son rôle dans la translocation ce qui est confirmé par différentes expériences de mutagenèse dirigée (Isidro,

Henriques et al. 2004; Oliveira, Henriques et al. 2006). Par ailleurs, ce motif semble conservé à travers les bactériophages et l'EBV.

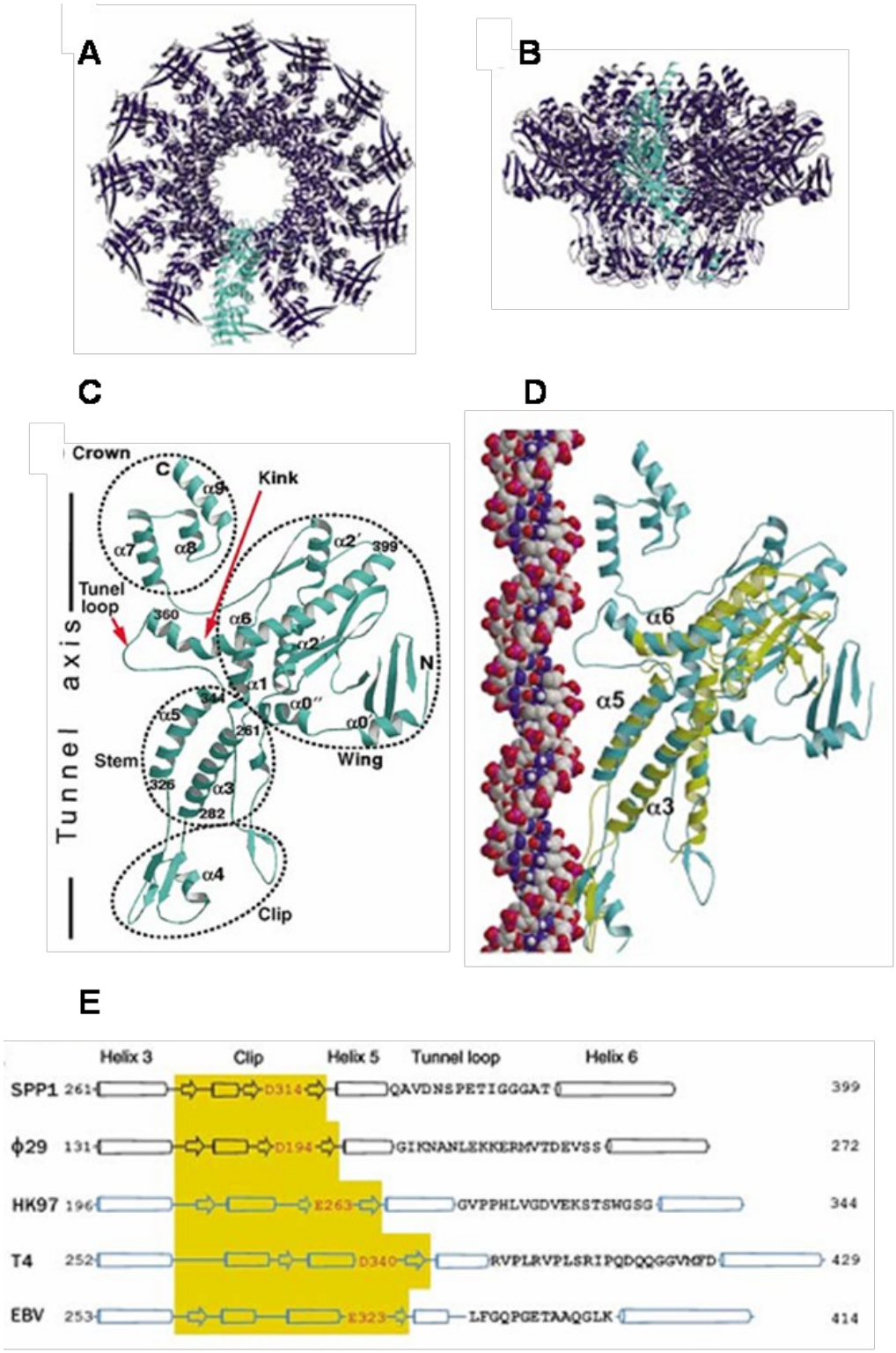



**Figure 19 (ci-contre): Protéine portail.** Vue axiale (A) et transversale (B) de la structurale toroïdale de la protéine portail du bactériophage SPP1. L'unité structurale est colorée en cyan. (C) Détails de l'unité structurale. (D) Conservation de la structure des unités de protéines portail du bactériophage SPP1 (cyan) et Φ29 (vert) en interaction avec l'ADN

(sphères). (E) Conservation de structures secondaires à travers les bactériophages et herpesvirus (Lebedev, Krause et al. 2007).

### B.1.c. **Assemblage et translocation de l'ADN**

La translocation de l'ADN viral dans la procapside nécessite l'interaction directe entre le complexe terminase apportant le double-brin d'ADN et la protéine portail servant de porte d'entrée de la capside. Chez les bactériophages λ, T4, P21 et T3, les domaines du coté terminase, impliqués dans cette interaction ont été clairement identifiés. Ils sont dans tous ces cas portés par les grande sous-unités et n'excède pas dix acides aminés, gpA (LSGEDE) (Wu, Christiansen et al. 1988; Yeo and Feiss 1995), gp17 (central :LYNDEDIFDD et C-ter: DYADKDD) (Lin, Rao et al. 1999), gp2 (VNSGYSR) (Yeo and Feiss 1995) et gp19 (LYWEDD) (Morita, Tasaka et al. 1995) respectivement. Ces séquences présentent toute des résidus acides probablement importants pour l'interaction, toujours situés à l'extrémité C-terminal de ces protéines. Le complexe terminase/protéine portail résulte en un vortex constitué par les trois structures en anneau des grandes et petites sous-unités avec la protéine portail comme représenté en Figure 11.

Un mécanisme probable de la translocation de l'ADN dans la capside a été proposé par Lebedev et al. et se décompose en cinq étapes: i. la boucle du tunnel reconnaît l'ADN au niveau de son grand sillon (Figure 19D) et une permutation cyclique des boucles s'opère (boucle de l'unité 1 s'interchange avec la 3, la 2 avec la 4…la 11 avec la 1 et la 12 avec la 2), ii. translocation de l'ADN par pas d'une paire de base, iii. rotation du portail de 12° par rapport à l'ADN, iv. hydrolyse d'une molécule d'ATP par les terminases pour fournir l'énergie suffisante à un cycle supplémentaire (Figure 20). Le caractère négativement chargé du domaine de la protéine portail interagissant avec l'ADN facilite la translocation de celui-ci propulsé dans la capside par les forces de répulsion de charges. Seule une ceinture interne de résidus positifs s'apparie avec les charges négatives de l'ADN pour s'y lier temporairement pendant le mécanisme d'entrée (Guasch, Pous et al. 2002).

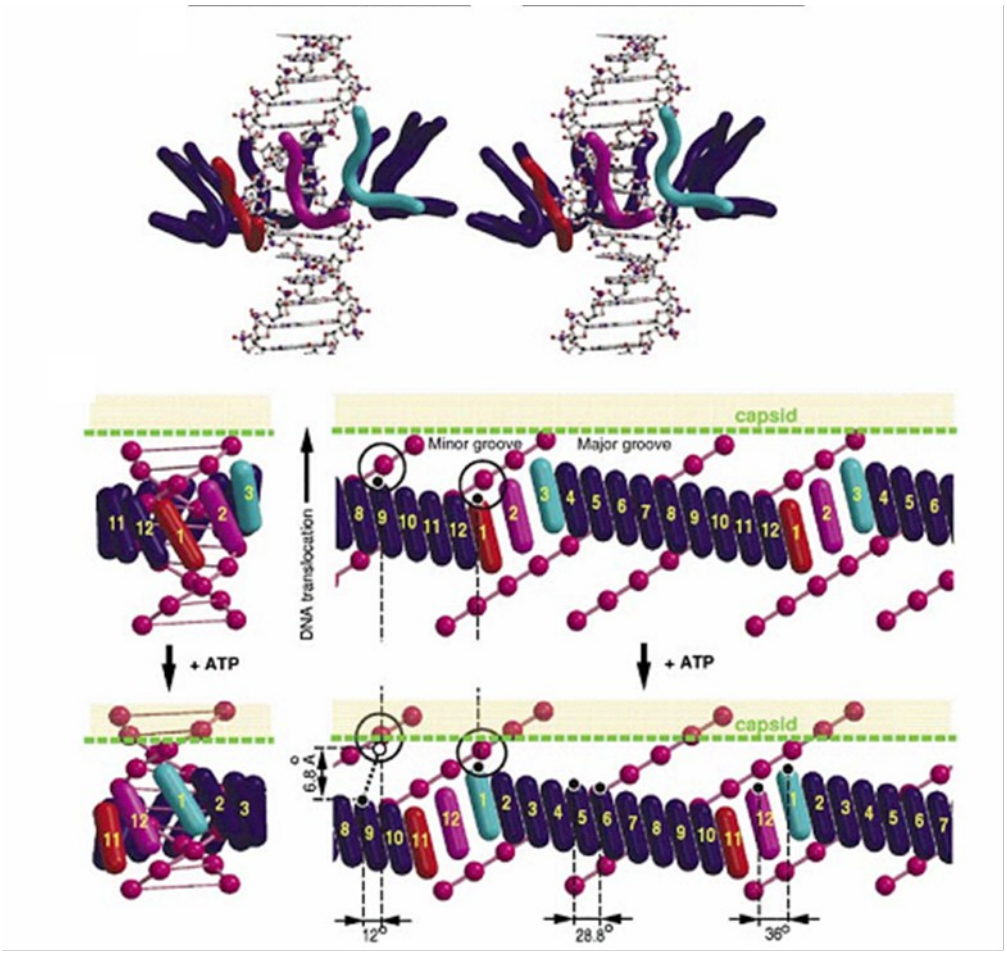



**Figure 20: Mouvement relatif des différentes unités de la protéine portail durant la translocation de l'ADN.**

## B.2. Rôle central de pUL97.

### B.2.a. Généralités sur les phosphotransférases.

pUL97 fait partie de la famille des protéines kinases, capables de transférer le groupement phosphate γ d'une molécule d'ATP sur le groupement hydroxyle d'une sérine ou d'une thréonine chez la sous famille des Thr/Ser-kinases, ou d'une tyrosine dans le cas des Tyr-kinases. Les protéines kinases sont essentielles dans les phénomènes de signalisation. Après

phosphorylation, le substrat devient actif ou inactif selon sa place dans un mécanisme biologique donné. Ce groupe d'enzyme se distingue par une conservation de plusieurs domaines impliqués dans la fixation de l'ATP, le transfert du groupement phosphate et la fixation du substrat. Chez les herpesvirus, la phosphotransférase virale rencontre des homologues chez HSV-1 (pUL13), EBV (BGLF4), VZV (ORF47), HHV-6 (U69), HHV-8 (ORF36), MCMV(M97), RCMV(R97) et enfin CMVH (pUL97) (Michel and Mertens 2004). L'alignement de leurs séquences avec des kinases cellulaires d'origines diverses a souligné l'existence de domaines conservés, signature de leur fonction commune (Chee, Lawrence et al. 1989). Neuf domaines plus ou moins conservés (numérotés de I à IX) ont été identifiés (Figure 21).

Les transférases se décomposent en deux lobes, un grand et un petit. L'ATP va se fixer dans une poche hydrophobe du petit lobe alors que le peptide-substrat va se caler à l'interstice des deux lobes.

Le domaine de liaison au nucléotide adopte un repliement spécifique formé d'un feuillet bêta incluant les sous domaines I, II V et IV permettant la proximité de résidus clés pour la fixation et la catalyse de l'ATP. La boucle de SDI contenant le motif $GxGx_2GxV$ et un résidu hydrophobe en position –1 crée des liaisons électrostatiques avec les phosphates α et β de l'ATP. Ce motif spécifique forme une boucle qui va entrer en contact à la fois avec le phosphate β et le ribose de l'ATP par ponts hydrogènes.

Environ 20 acides aminés plus loin, le domaine II se présente sous la forme AXKXO (O est un résidu hydrophobe) où la lysine conservée est chargée de maintenir l'ATP par interactions électrostatiques avec les phosphates α et β.

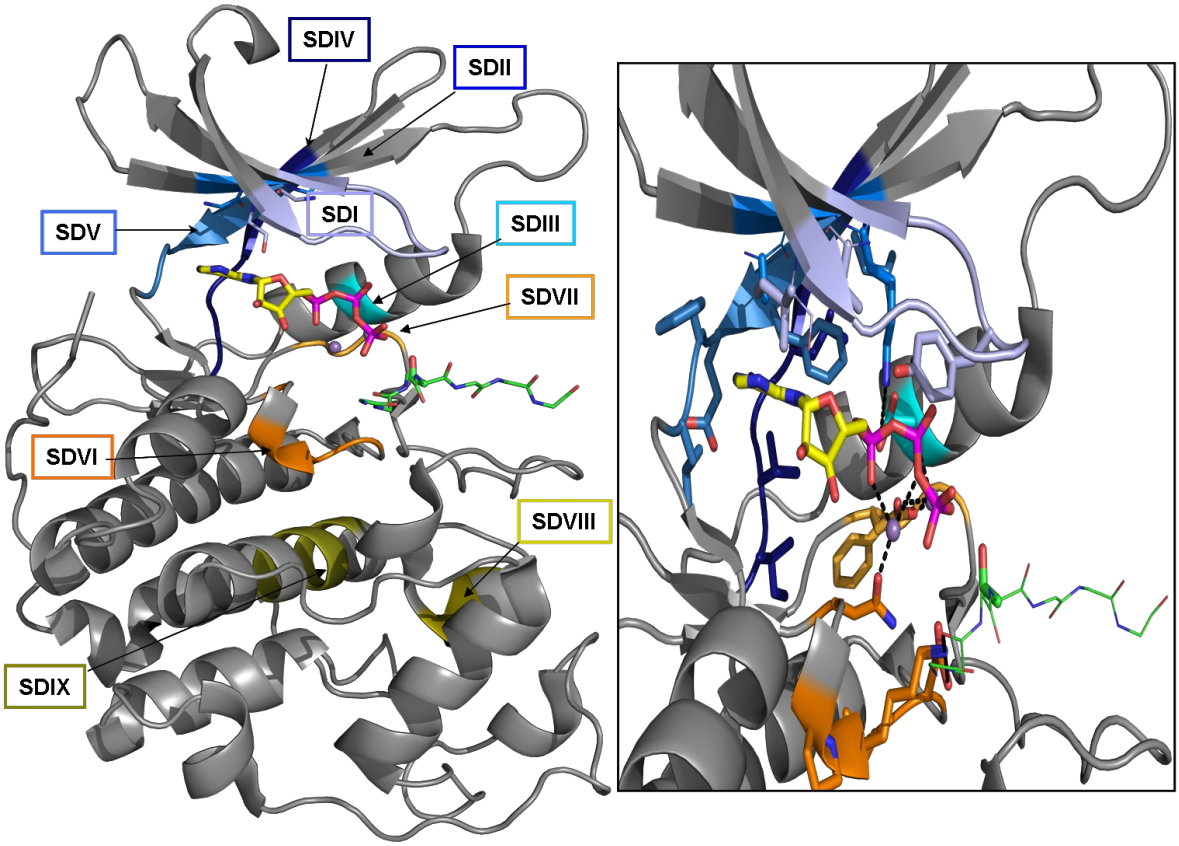



**Figure 21: La phosphotransférase cellulaire Cdk2.** A gauche, les différents sous domaines (SD) numérotés de I à IX communs aux phosphotransférases sont représentés sur la structure de Cdk2 (code PDB 1VYW). A droite, zoom sur le site de fixation de l'ATP (en sticks jaune). Les SDI, IV et V inclus dans le feuillet bêta créent une poche hydrophobe stabilisant le nucléotide. Le résidu carboxylate contenu dans SDIII oriente la lysine catalytique de SDII pour sa fixation aux phosphates α et β de l'ATP. Les résidus de SDVI pontent la sérine du substrat (en sticks vert) vers le phosphate γ. L'aspartate contenu dans SDVII chélate un ion magnésium stabilisant l'état de transition.

### B.2.b. <u>La phosphotransférase pUL97 et sa protéine accessoire pUL27</u>

Cette protéine de 707 acides aminés, est localisée dans le noyau et exprimée entre les stades précoce et tardif (Michel, Pavic et al. 1996).

Par homologie de séquence (Figure 22), il est aisé d'identifier les domaines contenus dans la structure primaire de pUL97.

```
                                  I         II   III IV/V  VI     VII   VIII  IX     X  XI
UL97   (1) ----------------- 337 GQGSFGEV - VVK - GEQ  -  DITPMN - DYS - REC - ELSALG - FRR 651 ---  (707)
 R97   (1)      ------------ 243 GKGSFGQV - ALK - LDA  -  DVTPMN - DYS - TDL - DLCALG - FRC 557 ---  (612)
UL13   (1)          -------- 158 GSGGYGDV - AVK - GEC  -  DIKCAN - DFS - TAL - DLYALG - DRC 476 ---  (530)
BGLF4  (1)              ----  85 GRGSYGAV - TVK - AED  -  DISPSN - DYG - REP - DLQSLG - CRS 387 --   (429)
Cdk2   (1)                 -  10 GEGTYGVV - ALK - REI  -  DLKPQN - DFG - PEI - DIWSLG - KRI 275 -    (298)
c-Src  (1)        ---------- 202 GKGEFGDV - AVK - AEA  -  DLAARN - DFG - PEA - DVWSFG - MRP 431 -    (450)
 p38   (1)                --  31 GSGAYGSV - AVK - REL  -  DLKPSN - DFG - PEI - DIWSVG - FRG 276 ---- (364)
PKC α  (1) ----------------- 346 GKGSFGKV - AIK - VEK  -  DLKLDN - DFG - PEI - DWWAYG - KRL 581 -----(672)
```

**Figure 22: Alignement de séquence entre pUL97, ses homologues herpesviraux et d'autres protéines kinases cellulaires.** (pUL97 de HCMV, pR97 de RCMV, pUL13 de HSV-1, BGLF4 de EBV, Cdk2, c-Src, p38 et PKC cellulaires) (Romaker, Schregel et al. 2006)

Le domaine I correspond au fragment $^{337}$LGQGSFGEV$^{345}$.

Le domaine II contient la lysine 355 catalytique très conservée, située dix résidus après le site de liaison au nucléotide. Son importance a été montrée par mutation qui supprime l'activité de l'enzyme. Expérimentalement, les mutants Lys355Met (Marschall, Stein-Gerlach et al. 2002; Prichard, Britt et al. 2005) ou Lys355Gln (Baek, Krosky et al. 2002; Krosky, Baek et al. 2003) sont utilisés comme témoins négatifs de l'activité kinase de pUL97.

Le domaine III contient un glutamate (Glu$^{380}$ pour pUL97). Alors que la lysine stabilise le phosphate gamma de l'ATP, le carboxylate va être un accepteur de proton lors de la ligation du groupement phosphate γ sur l'hydroxyle du substrat (sérine ou thréonine) (Hanks and Hunter 1995). Le mécanisme réactionnel de telles enzymes est décrit à la Figure 23.

Les domaines IV, V et X sont quant à eux peu conservés.

Les sous domaines VI et VII sont les sous domaines les plus conservés chez tous les types de kinases herpes virales, cellulaires ou bactériennes. Il est présumé que ces domaines soient impliqués dans le transfert de phosphate. Ils sont constitués de résidus hydrophobes et de carboxylates (aspartates ou glutamates).

Des modèles basés sur la structure de la kinase cellulaire Cdk2 ( code PDB 1VYW) (Pevarello, Brasca et al. 2004) ont été générés afin de décrire les interactions entre pUL97 et quelques uns de ses inhibiteurs (Romaker, Schregel et al. 2006; Chou 2008).

La protéine pUL27 est la protéine accessoire de pUL97. Cette protéine de 608 amino acides reste encore très énigmatique puisque peu de choses sont connues sur sa structure et sa fonction.

Cependant, comme nous le verrons plus tard, pUL27 semble influencée par les inhibiteurs originellement dirigés contre pUL97 (Komazin, Ptak et al. 2003; Chou 2009).

Partant de l'hypothèse que pUL97 est impliqué directement ou indirectement dans la synthèse de l'ADN viral et dans l'exportation nucléaire de la capside, il a été suggéré que pUL27 était elle aussi impliquée dans l'une ou l'autre de ces étapes (Komazin, Ptak et al. 2003).

Il semblerait que cette protéine ne soit pas indispensable au virus puisque des mutants *UL27*-déficients semblent se développer normalement. Peut être cela explique-t-il la présence d'homologues de pUL27 uniquement chez les β-herpesvirus (Komazin, Ptak et al. 2003) ?

**Figure 23: Mécanisme réactionnel de la phosphotransférase (Madhusudan, Trafny et al. 1994) .** En particulier, chez pUL97, l'équivalent de Lys 168 est Lys 355 et celui de Asp 166 est Glu 380.

### B.2.c. <u>Les différents substrats de pUL97</u>

Outre son interaction avec pUL44 via son domaine 366-459 (Marschall, Freitag et al. 2003), il a été montré que pUL97 intervenait dans l'assemblage de la capside (Wolf, Courcelle et al. 2001), la tégumentation (van Zeijl, Fairhurst et al. 1997; Prichard, Britt et al. 2005) ou encore la sortie du noyau (Krosky, Baek et al. 2003).

Pour ce faire, pUL97 phosphorylerait des protéines cellulaires comme p32 afin de former le complexe pUL97/p32. p32 interagissant naturellement ave les récepteurs de la lamine B, il provoque l'accumulation de pUL97 à l'abord de ces récepteurs. Après phosphorylation de ces derniers, la membrane nucléaire se disloque permettant ainsi la sortie de la capside (Marschall, Marzi et al. 2005).

Par ailleurs, la fonction prépondérante de pUL97, au moment de la tégumentation, a été démontré par Prichard et al. (Prichard, Britt et al. 2005) et plus récemment confirmé par Kamil et Coen (Kamil and Coen 2007). Cela a notamment été mis en évidence en observant l'agrégation de corps d'inclusion dans le noyau composés des protéines pp65, ppUL25 (protéines du tégument) et MCP (Protéine Majeur de la Capside).

En résumé, pUL97 intervient dans au moins trois étape-clés de la réplication virale: la synthèse de l'ADN en phosphorylant pUL44 pour finalement former le complexe pUL44/pUL54, la tégumentation en interagissant avec la phosphoprotéine du tégument pp65 afin d'en éviter l'agrégation, et enfin la sortie du noyau en déstabilisant les membranes nucléaires.

Finalement, pUL97 est un substrat pour elle-même (van Zeijl, Fairhurst et al. 1997; Baek, Krosky et al. 2002) et existe sous forme de dimère à travers sa région 231-280 (Schregel, Auerochs et al. 2007) (Figure 10). Ces deux derniers points sont des prérequis pour une efficacité maximum du virus.

Notons, enfin que des expériences de phosphorylation par pUL97 sur l'histone HB2 ont révélé une sensibilité particulière de pUL97 pour un motif peu conventionnel T/S$x_4$R/K avec une importance particulière pour la position P+5 occupée par un acide aminé basique, d'où l'idée d'employer ce motif comme inhibiteur potentiel (Baek, Krosky et al. 2002).

Une interaction avec l'ARN polymérase II a aussi été mise en évidence soulevant l'hypothèse d'un rôle mineur de pUL97 dans la régulation et l'expression des gènes viraux (Baek, Krosky et al. 2004).

## B.3. Nouveaux inhibiteurs de CMVH

### B.3.a. Les benzimidazoles -D et L- ribonucléosides.

Comme nous l'avons souligné précédemment l'utilisation des molécules visant l'ADN polymérase pUL54 a engendré un grand nombre de mutations les rendant finalement inactives.

En 1995, Townsend et al. synthétisent et testent une nouvelle classe de molécules s'avérant particulièrement efficaces contre les infection à CMVH et présentant une faible cytotoxicité: les 2,5,6-trihalo-1-(bêta-D-ribofuranosyl)benzimidazoles (Townsend, Devivar et al. 1995). De cette famille de molécule, deux en particuliers se distinguent par leur efficacité: le 2,5,6-Trichloro-1-(bêta-D-ribofuranosyl) benzimidazole (TCRB) et son analogue bromé le 2-bromo-5,6-dichloro-1-(bêta-D-ribofuranosyl)benzimidazole (BDCRB).

Très vite, il a été montré que le BDCRB vise pUL89 en empêchant l'étape de maturation de l'ADN (Underwood, Harvey et al. 1998). Il sera démontré plus tard que le BDCRB affecte la fonction nucléase associée à pUL89 (Scheffczik, Savva et al. 2002) et l'activité ATPasique associé à pUL56 (Scholz, Rechter et al. 2003).

Dans le même temps des mutants résistants au dérivés benzimidazolés sur les protéines pUL89 (D344E et A355T) et pUL56 (Q204R) ont été identifiées mettant ainsi en relief la relation entre cette classe de molécules et les stades de maturation et d'encapsidation (Krosky, Underwood et al. 1998; Underwood, Harvey et al. 1998).

Chaque protéine mutante isolée s'est montrée 10 fois plus résistante que la souche sauvage et 30 fois plus en combinant les deux. Cependant, ces molécules se sont vus

rapidement dégradées *in vivo* au niveau de leurs liaisons glycosidiques par la N-methylpurine DNA glycosylase et la 8-oxoguanine DNA glycosylase (Lorenzi, Landowski et al. 2006).

Cet inconvénient à donc stimulé l'élaboration de composés plus stable (Townsend, Gudmundsson et al. 1999) et plusieurs équipes n'ont pas tardé à s'engager dans l'optimisation de ces composés (Chan, Chamberlain et al. 2000; Gudmundsson, Freeman et al. 2000; Zou, Kawashima et al. 2000).

Finalement, un analogue du BDCRB incorporant un sucre –L, le 2-isopropylamino-5,6-dichloro-1-(bêta-L-ribufuranosyl) benzimidazole, 1263W94 ou Maribavir (MBV), a montré toutes les qualités d'un inhibiteur de choix: une efficacité comparable à celle du BDCRB même contre les souches résistantes aux anti-pUL54 (GCV et PFA), une cytotoxicité faible et une sélectivité intéressante puisque ce composé viserait directement la phosphotransférase pUL97 (Biron, Harvey et al. 2002). Sa biodisponibilité a par ailleurs été testée. Elle s'élève à 90% et 50% chez le rat et le singe, respectivement (Koszalka, Johnson et al. 2002).

L'activité *in vitro* des benzimidazoles D- et L- ribonucléosides semble s'appliquer uniquement à CMVH et EBV, mais pas aux HSV-1, HSV-2, VZV, HHV-6 et HHV-8 (Williams, Hartline et al. 2003).

Par la suite, une nouvelle mutation de résistance aux dérivés benzimidazolés au niveau du gène codant la protéine pUL27 (L335P) a été détectée sur des souches de laboratoire exposées à long terme au MBV. Ces résultats suggèrent donc l'implication de pUL27 dans le mécanisme d'action de MBV ou d'un rôle probable dans l'étape d'exportation du noyau de la capside (Komazin, Ptak et al. 2003).

Une autre molécule a été développée à partir de 2004, un D-ribonucléoside tout comme le *lead* BDCRB, baptisé GX275175X ou 175X. Celui-ci révèle une efficacité comparable à ses précurseurs, BDCRB et TCRB, et une biodisponibilité intéressante. Du fait, qu'il soit d'une part efficace contre des souches résistantes au GCV, CDV, PFA, ACV (Aciclovir, inhibiteur de la polymérase virale chez les simplex virus) et MBV, et d'autre part inerte contre les souches résistantes au BDCRB, mènent à croire que cette molécule agit selon un mécanisme proche de celui de ses précurseurs mais différent du MBV.

De plus, les souches portant la double mutation résistante sur pUL56 et sur pUL89 sont deux fois plus résistantes vis-à-vis du GW275175X (Underwood, Ferris et al. 2004).

Des études *in vivo* confrontant l'efficacité relative du GCV, BDCRB, 175X et MBV simultanément sur des modèles murins, montrent l'efficacité des composés benzimidazolés dans des conditions mimant des cas de rétinites ou de greffe d'organes. Cependant les dérivés

D-ribonucléosides sont efficaces sur les modèles infectés tels que le foie et le thymus mais inefficaces sur des cas d'infections oculaires. Par contre, le MBV a montré une efficacité dans les deux cas. Cette étude confirme que des cas de greffes peuvent trouver plusieurs alternative thérapeutique alors que le cas des rétinites propose un choix plus restreint de molécules (Kern, Hartline et al. 2004).

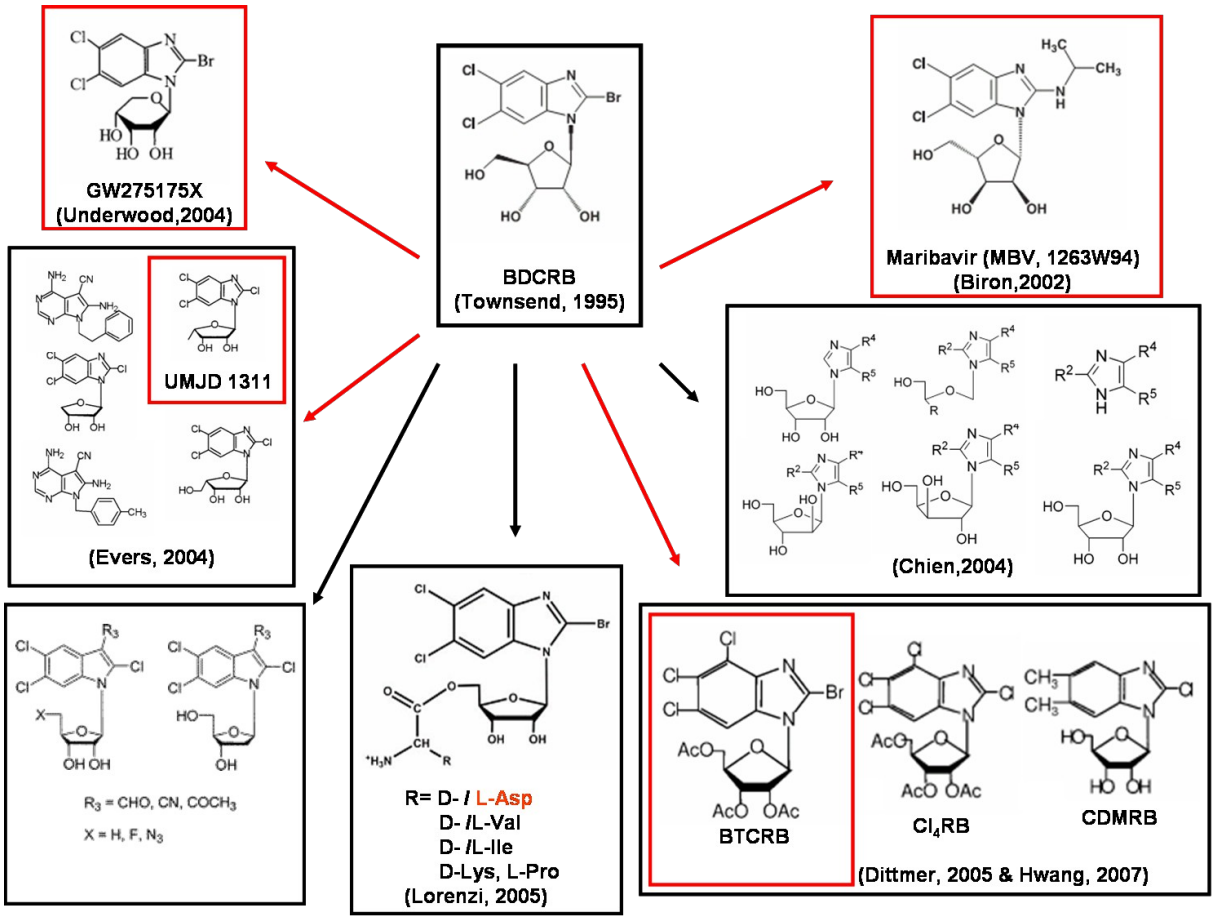



**Figure 24: Dérivés des 2,5,6-trihalo-1-(bêta-D-ribofuranosyl)benzimidazoles: du BDCRB au BTCRB.** En rouge sont représentées les molécules ayant manifestées une activité pharmacologique.

Une troisième classe de dérivés benzimidazolés a été testée conjointement avec le MBV et le BDCRB. Cette classe est constituée d'analogue α-5'-deoxyfuranosyl du TCRB, dont le composé UMJD1311 en est le représentant.

Une étude montre que cette famille de molécules inhiberait une étape antérieure à la maturation de l'ADN et à sa synthèse. Il montre une efficacité à la fois contre les souches résistantes au GCV, CDV, PFA, MBV et BDCRB (Evers, Komazin et al. 2004).

Entre 2004 et 2005, la chimiothèque initiée par la découverte des TCRB et BDCRB s'agrandie de plus en plus en opérant différentes modifications chimiques susceptibles

d'optimiser leurs efficacités (Chien, Saluja et al. 2004; Williams, Ptak et al. 2004; Chen, Wei et al. 2005; Song, Vig et al. 2005).

Une nouvelle alternative a été exploitée faisant intervenir le greffage d'L-amino acides ester et comparée au BDCRB. Alors que le dérivé d'isoleucine montre une efficacité 40 fois plus faible que celle de son parent, le lysyl ester exhibe une toxicité élevée. La meilleure alternative semble être l'aspartyl ester, L-Asp-BDCRB, qui fait preuve d'une haute activité antivirale, d'une faible toxicité et d'une stabilité intéressante de ses liaisons ester et glycosidique, ce qui a été confirmé en tests *in vivo* (Lorenzi, Landowski et al. 2005).

Un nouvel indice sur l'élucidation du mécanisme d'action des dérivés D-ribonucléosides du BDCRB a été publié par l'équipe de Elke Bogner qui a montré une interaction directe entre pUL56 et pUL104 inhibée par ces composés et en particulier par le dérivé tétrachloro du BDCRB, le 2,4,5,6-tetrachloro-1-(2,3,5-tri-*O*-acetyl-β-d-ribofuranosyl)benzimidazole ($Cl_4RB$) (Dittmer, Drach et al. 2005).

Le $Cl_4RB$ a récemment donné naissance a un autre composé tétrahalogéné *O*-acétylé le 2-bromo-4,5,6-trichloro-1-(2,3,5-tri-*O*-acetyl-β-d-ribofuranosyl) benzimidazole (BTCRB) montrant une activité intéressante en inhibant l'activité ATP-asique de pUL56 nécessaire au clivage des concatémères d'ADN néosynthétisés. Sa biodisponibilité *in vivo* doit être testé dans l'avenir afin de déterminer son potentiel de développement en tant que médicament potentiel (Hwang, Kregler et al. 2007).

Les divers dérivés du BDCRB sont représentés à la Figure 24.

### B.3.b. BAY 38-4766 et ses dérivés

D'autres inhibiteurs du cycle de la maturation et de l'encapsidation de l'ADN viral de type non nucléosidique ont été développés. En particulier, une molécule, le 9-(1,3-dihydroxy-2-propoxymethyl-N-[4({[5-dimethylamino)-1-naphtyl[sulfonyl}amino)-phenylpropanamide ou, « plus simplement » BAY 38-4766. Ce *lead* a été développé par le groupe pharmaceutique BAYER à la fin des années 90, tant il a pu montré d'effet sur le cycle de maturation de l'ADN (Contrafouris 1999; Hoffman and Skiest 2000).

Les efficacités de BAY 38-4766 et de sa prodrogue glycyl-ester, le BAY 40-1007, ont été comparées au Ganciclovir *in vitro* et *in vivo* sur des modèles murins infectés par MCMV (Cytomégalovirus Murin) et CMVH.

L'administration de ces molécules a ainsi montré leur efficacité *in vitro* avec toutefois une efficacité 10 fois supérieure sur les souches MCMV par rapport à leur effet sur CMVH mais

tout de même 600 fois plus actif que le GCV contre ces mêmes souches de MCMV y compris celles résistantes à GCV.

*In vivo*, les résultats sur le traitement d'organes (yeux, foie, glandes salivaires) se sont avérés encourageant pour la poursuite des investigations du BAY 38-4766 et de ses prodrogues (Buerger, Reefschlaeger et al. 2001).

Le BAY 38-4766 ne semble pas influencer la transcription des gènes ou leur synthèse mais manifeste toutefois un effet sur la surproduction de capside non infectieuse, c'est-à-dire vide. Bien qu'influençant aussi les protéines pUL89, pUL56 et pUL104, comme le prouve la présence de mutants résistants à BAY 38-4766, il est proposé que cette classe de molécules non nucléosidiques agisse selon un mécanisme différent des dérivés benzimidazolés-D-ribonucléoside.

En dehors du fait que le TCRB ou BDCRB restent efficaces contre les résistances au BAY 38-4766 *et vice versa*, l'action du BAY 38-4766 contre le MCMV a déjà été montrée, chose qui n'a pas été vérifiée pour les BDCRB et TCRB. Puisque mécanismes d'action différents, ceci suggère donc, à l'échelle atomique, une poche de liaison différente pour les deux classes d'inhibiteurs. Ceci reste surprenant tant les sites de mutations imputés aux BDCRB et au BAY 38-4766 sont proches à la fois sur pUL56 (acides aminés 202 et 208 pour BAY 38-4766, 204 pour BDCRB) et pUL89 (358 et 344 pour BAY 38-4766 et BDCRB, respectivement) (Buerger, Reefschlaeger et al. 2001; Evers, Komazin et al. 2002).

McSharry et al. et Reefschlaeger et al. confirment l'effet inhibiteur de BAY 38-4766 et un de ses métabolites, BAY 43-9695 sur plusieurs souches de HCMV y compris certaines résistantes ou sensibles au GCV, au GCV/CDV et GCV/PFA (McSharry, McDonough et al. 2001; Reefschlaeger, Bender et al. 2001).

Cependant, bien que visant *a priori* des cibles similaires, le BAY 38-4766 et le BDCRB ont des effets antagonistes et synergétiques (observation difficilement explicable) alors que l'association de l'un ou l'autre avec des inhibiteurs de la synthèse de l'ADN ont des effets uniquement synergétiques (Evers, Komazin et al. 2002).

Sa biodisponibilité et sa faible toxicité ont été finalement testées sur le cochon d'Inde porteur de GPCMV (Guinea Pig Cytomegalovirus). Les résultats encourageants, notamment sur la facilité de ce composé à traverser les barrières placentaires, en font un candidat de choix pour le traitement des infections à CMV chez la femme enceinte afin d'empêcher sa transmission au fœtus (Schleiss, Bernstein et al. 2005).

### B.3.c. Les indolocarbazoles.

BAY 38-4766 et ses dérivés ne sont pas les seules molécules non nucléosidiques dont le pouvoir inhibiteur contre le développement du CMV ait été découvert. La classe chimique des indolocarbazoles a ainsi fait son apparition dans la littérature traitant du cytomégalovirus entre les années 1999 et 2000.

En criblant une chimiothèque d'inhibiteurs de protéines kinase sur pUL97, le composé Arcyriaflavin A a montré son efficacité et a motivé la synthèses de dérivés pour des études de SAR (*Structure/Activity relationship*) (Slater, Cockerill et al. 1999; Slater, Baxter et al. 2001). A partir d'un set de molécules connues comme étant des inhibiteurs de protéines kinase (PKI), des indolocarbazoles ont été remarqués pour leur faible cytotoxicité et une efficacité redoutable se traduisant par une $IC_{50}$ de l'ordre du nanomolaire comparée à celle du GCV dans les mêmes conditions qui était de l'ordre du micromolaire.

Ces molécules nommées Gö6976, K252a et K252c s'avèrent par contre inefficaces contre des souches de HSV-1 traduisant ainsi leur spécificité vis-à-vis de pUL97. D'un autre côté, une des molécules testées, Gö6850, montrée comme ayant un fort pouvoir inhibiteur sur la cycline dépendante kinase cdk2a, est inefficace contre pUL97. Ceci montre donc que le motif indolocarbazole contient une géométrie particulièrement intéressante pour le développement d'inhibiteurs de la phosphotransférase pUL97 (Zimmermann, Wilts et al. 2000). Romaker et al ont prédits par *docking* le site de liaison d'un des indolocarbazole (NGIC-I) qui se loge dans la poche catalytique au niveau des résidus $His^{360}$, $Glu^{362}$, $Arg^{413}$-$Phe^{419}$, $Met^{460}$ et $Cys^{480}$ (Romaker, Schregel et al. 2006).

# V. PRÉDICTION DES INTERFACES PROTÉINE-PROTÉINE

Avant de rentrer dans le vif du sujet, il semble nécessaire de revenir sur les principes fondamentaux concernant les protéines. Nous allons nous intéresser dans un premier temps aux mécanismes cellulaires engendrant la synthèse des protéines, les briques élémentaires les constituant, et les différentes forces permettant le repliement et la cohésion moléculaire amenant à la fonction.

La synthèse des protéines est clairement divisée en deux étapes majeures: la transcription et la traduction. La transcription fait intervenir l'ARN-polymérase (ARN-pol) qui va se fixer sur l'ADN au niveau de sites spécifiques, précédant un gène, identifiable par les boîtes CAAT et TATA constituant le « promoteur » du gène considéré. Ainsi, l'ARN-pol va dérouler la double hélice d'ADN puis par complémentarité de bases et en se servant d'un brin d'ADN comme matrice va procéder à la synthèse de l'ARN-messager (ARNm). La synthèse de l'ARNm prend fin dès que l'ARN-pol va reconnaître, de la même manière qu'elle reconnaît les séquences constituant le promoteur, la terminaison de la séquence du gène. L'ARN-pol se détache de l'ADN qui reprend alors sa forme hélicoïdale alors que l'ARNm est libéré pour que commence l'étape de traduction du gène. La traduction des gènes dépend dans un premier temps du type d'organisme. Chez les procaryotes, l'ARNm est traduit directement alors que chez les eucaryotes, une étape intermédiaire, l'épissage va consister à garder les parties codantes du gène (exons) et enlever les non codantes (introns). L'ARNm mature peut être exporté hors du noyau pour être traduit en protéine au sein du cytoplasme.

La traduction démarre toujours par la reconnaissance par le ribosome du codon (triade de nucléotides) initiateur AUG codant pour une méthionine. Alors que le ribosome va lire l'ARNm codon par codon, un ARN de transfert (ARNt) spécifique à chaque acide aminé va apporter le résidu considéré pour continuer l'élongation de la chaîne polypeptidique jusqu'à ce que le ribosome rencontre le codon stop qui va marquer la fin de la synthèse de la protéine. La protéine est ensuite exportée hors de la cellule ou pas selon la fonction moléculaire qui lui est attribuée.

## *A. Repliement Moléculaire.*

Chaque acide aminé est codé par un ou plusieurs codons différents au niveau de l'ARN messager. On distingue vingt résidus différents constitué par le même squelette: une fonction

amine et une fonction acide d'où leur nom d'acides aminés. Ils se différencient au niveau de la chaîne latérale porté par le carbone alpha situé sur le squelette.

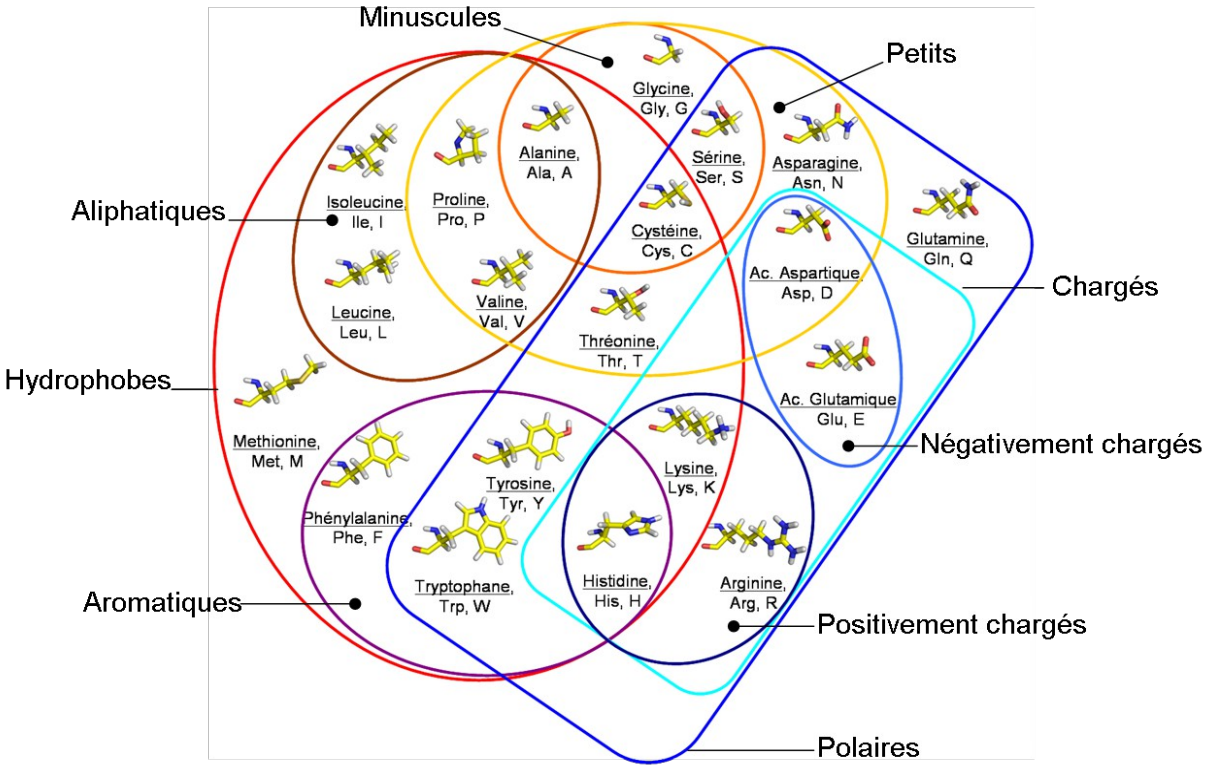



**Figure 25: Diagramme de Venn des 20 acides aminés naturels.** Le cercle rouge contient les résidus hydrophobes, l'ellipse marron les aliphatiques, l'ellipse violette les aromatiques, l'ellipse jaune les petits, l'ellipse orange les minuscules, le rectangle bleu foncé les polaires, le rectangle cyan les chargés contenant les positifs (cercle bleu marine) et négatifs (ellipse bleue).

Les acides aminés naturels se décomposent en plusieurs classes chimiques pouvant intervenir dans la structure et la fonction des protéines en y permettant la formation des différents types de forces intermoléculaires (Figure 25). D'après le code à une lettre, on distingue ainsi les aliphatiques (A, V, L, I) et les aromatiques (F, Y et W) capables de former des interaction hydrophobes par interactions des méthyles pour les uns ou par « *pi stacking* » pour les autres (empilement des cycles aromatiques par mise en commun de leurs électrons pi). Les interactions électrostatiques se mettent en place avec des acides aminés chargés ou dipolaires non chargés. Le caractère chargé des acides aminés dépend de leurs caractéristiques acido-basiques intrinsèques (Figure 26).

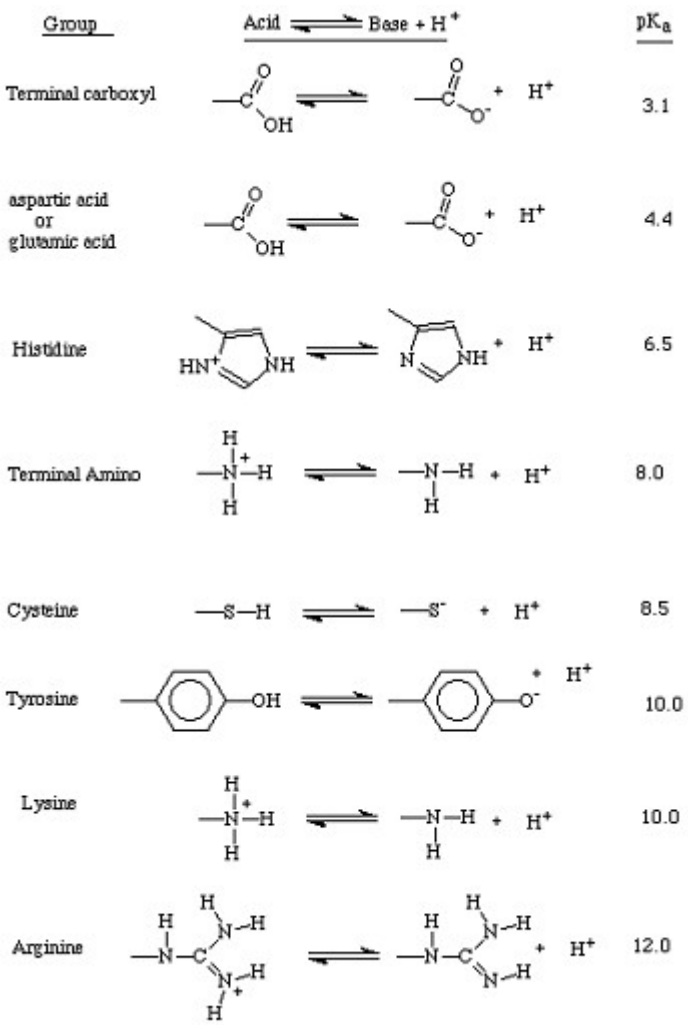



**Figure 26: État de charge des acides aminés en fonction de leur pKa.**

C'est pourquoi on remarque que les acides aminés portant des fonctions acide carboxylique dont les $pK_A$ des protons labiles sont de l'ordre de 3,7 et 4,3 pour les acides aspartique et glutamique, respectivement, sont représentés, à pH physiologique (neutre), sous forme de carboxylate faisant apparaître une charge négative.

Par contre, les charges positives sont elles représentées par l'intermédiaire des résidus lysine -portant un groupement amine primaire ($pK_A$ = 9,5) - et arginine -un groupement guanidino ($pK_A$ = 12) - quelque soit le pH. D'autres résidus telle que l'histidine peuvent exister sous la forme cationique de leur groupement imidazole en deçà de pH = 6,5. Le caractère chargé des chaînes latérales est largement déterminé par leur exposition au solvant et en l'occurrence à l'eau. Suivant ce paramètre, les $pK_A$ d'une histidine peuvent varier de 5,4 à la surface d'une protéine à 7,4 au centre actif (Liu, Ryan et al. 1997). Les interactions entre charges sont fondamentales pour la cohésion de l'édifice protéique. Les interactions charge-

charge peuvent former des réseaux tels que la mutation d'un résidu chargé situé à la surface de la protéine (à 20 Å du site catalytique) peut entraîner une baisse considérable de l'activité biologique de la protéine. Cet effet traduit la dépendance de plusieurs régions éloignées d'une protéine à travers un réseau étendu de forces électrostatiques (Loewenthal, Sancho et al. 1993; de Kreij, van den Burg et al. 2002).

D'autres résidus ne sont pas chargés mais du fait de la présence d'hétéroatomes au niveau de leurs chaînes latérales sont le siège de dipôles électriques. Ainsi, les asparagine et glutamine par leurs groupements amides, le tryptophane par son groupement indole, les sérines, thréonines et tyrosines par leurs groupes hydroxyles forment des dipôles; ainsi que proline, cystéine (thiol) et méthionine (mercapto) (Zimmerman, Eliezer et al. 1968; Grantham 1974). Dans l'absolu, les charges jouent un rôle si important dans la fonction des protéines qu'une étude statistique diligentée par Thornton en 2002 a montré que les sites catalytiques sont constitués à 65% de résidus chargés, de 27% de polaires et seulement 8% d'apolaires. En particulier on soulignera que quatre acides aminés représentent à eux seuls plus de la moitié des sites actifs: 18% d'histidines, 15% d'aspartates, 11% d'arginines et autant de glutamates(Bartlett, Porter et al. 2002).

Pour résumer, la nature a mis à la disposition des divers organismes vivants vingt acides aminés faisant un étalage non exhaustif des fonctions chimiques possibles mais pouvant à eux seuls subvenir aux besoins de la diversité des organismes vivants. On peut faire un distinguo assez clair entre les résidus hydrophobes (y compris proline et glycine) plutôt voués au repliement pur des protéines alors que les résidus chargés et polaires sont en général destinés aux processus catalytiques.

Nous venons de traiter ici des fonctions chimiques des chaînes latérales mais le squelette carboné composé de fonctions carbonyles et amides est le siège de multiples ponts hydrogènes contribuant en très grande partie au repliement moléculaire.

Il existe quatre types de structure intervenant dans la description des protéines. Celles-ci sont définies selon le degré de repliement considéré.

La structure primaire d'une protéine est constituée par la succession d'acides aminés appelée aussi « séquence ». Les différents résidus sont reliés les uns aux autres par une liaison peptidique. Par convention, le premier acide aminé de la séquence est celui portant une amine libre. C'est aussi le premier acide aminé qui a été traduit lors de la synthèse de la chaîne polypeptidique: c'est l'extrémité N-terminal (N-ter). Le résidu portant une fonction carboxylique constitue, quant à lui, l'extrémité C-terminal (C-ter).

Les structures secondaires se décompose en trois grande classes: hélices, feuillets β et coudes. Elles mettent en jeu des liaisons hydrogènes impliquant les atomes du squelette protéique dépendant des couples d'angle φ et ψ définis comme étant les angles dièdres formés par les atomes $C'_i$-$NH_{i+1}$-$C\alpha_{i+1}$-$C'_{i+1}$ et $NH_i$-$C\alpha_i$-$C'_i$-$NH_{i+1}$, respectivement (C'est la convention utilisée pour nommer le carbonyle du squelette). L'ensemble des couples de valeurs observées expérimentalement peuvent être reportées sur un diagramme appelé diagramme de Ramachandran (Ramachandran, Ramakrishnan et al. 1963).

Les hélices sont soit α, $3_{10}$ ou π. L'hélice α est caractérisée par un pas de 3,6 résidus par tour. Le tout est solidement maintenu par un réseau de ponts hydrogènes impliquant les oxygènes des carbonyles du squelette avec l'hydrogène porté par la fonction amide du i+4$^{\text{ième}}$ résidu. Ainsi la distance typique de la liaison hydrogène formée atteint 2,86 Å c'est-à-dire une longueur énergétiquement optimale pour ce type de liaison H. Tous les acides aminés peuvent participer à ce type de structure secondaire avec une propension plus ou moins importante selon le résidu. Tous sauf un: la proline qui, du fait du repliement de sa chaîne latérale sur l'azote de son squelette, est déficiente en atome donneur d'hydrogène et donc ne peut participer à une telle structure secondaire. En revanche elle peut former des hélices de type poly-pro I ou II (Walter, Ealick et al.) suivant qu'elles soient constituées de prolines *cis* ou *trans*, respectivement. Les deux autres types d'hélices, $3_{10}$ et π, sont des variantes de l'hélice α en ceci qu'elles sont respectivement plus ou moins compactes. L'hélice $3_{10}$ est d'ailleurs observée seulement en fin d'hélice α alors que l'hélice π n'est jamais rencontrée tant son couple φ/ψ est défavorable.

Les feuillets β sont des types de repliements locaux formés par l'association d'au moins deux brins bêta. Ceux-ci se disposent de manière parallèle ou anti-parallèle et là encore mettent en jeu des liaisons hydrogènes entre les groupements carbonyles et amides du squelette. Les feuillets sont des structures étendues qui ne sont pas planes mais légèrement distordues. Pour la même raison évoquée précédemment (carence de proton amide) les prolines ne peuvent être impliquées dans de telles structures.

Enfin, les coudes sont des structures secondaires extrêmement localisées pontant d'autres structures secondaires. Ils impliquent trois à quatre résidus et certains résidus comme la glycine (du fait de sa chaîne latérale extrêmement peu volumineuse) y sont fréquemment rencontrés. Le Tableau 5 regroupe les différentes valeurs caractérisant les différentes structures secondaires.

| | Angles de liaisons | | | résidues par tour | translation par résidue |
|---|---|---|---|---|---|
| | φ | ψ | ω | | |
| feuillet β antiparallèle | -139 | 135 | -178 | 2 | 3,4 |
| feuillet β parallèle | -119 | 113 | 180 | 2 | 3,2 |
| hélice α droite | -57 | -47 | 180 | 3,6 | 1,5 |
| hélice $3_{10}$ | -49 | -26 | 180 | 3 | 2 |
| hélice π | -57 | -70 | 180 | 4,4 | 1,15 |
| poly(Pro)I | -83 | 158 | 0 | 3,33 | 1,9 |
| poly(Pro)II | -78 | 149 | 180 | 3 | 3,12 |

**Tableau 5: caractéristiques géométriques des différents éléments de structure secondaire des protéines.**

La structure tertiaire est le repliement tridimensionnel de la chaîne polypeptidique de protéines monomériques. Il est à présent largement admis que la force motrice de ce repliement est constituée par des interactions hydrophobes entre domaines (structures secondaires) (Anfinsen 1973; Tsai and Nussinov 2001; Kmiecik and Kolinski 2007). En milieu aqueux, il y aura formation d'un cœur hydrophobe alors que les résidus hydrophiles seront répartis à la surface de la molécule. Les processus de repliement sont certes dirigés principalement par ces interactions mais la présence de ponts disulfures entre chaînes latérales de cystéines ou l'isomérisation de prolines tout aussi indispensable à l'intégrité structurale des protéines doit être contrôlés. *In vivo* ce sont des enzymes spécifiques chargées du repliement ou de la dégradation de protéines défectueuses qui vont aider à accomplir cette fonction. Ces enzymes sont connues sous le nom de chaperonnes principalement localisées dans le réticulum endoplasmique (Lee and Tsai 2005; Ni and Lee 2007). Enfin, certains ions métalliques de différents degrés d'oxydation issus du manganèse, du fer, du nickel, du cuivre, du zinc, du calcium, du magnésium, et dans une moindre mesure du cobalt, du molybdène, du tungstène ou du cadmium peuvent contribuer au repliement tertiaire d'une protéine en rapprochant par liaison ionique des chaînes latérales d'acides aminés situés à plusieurs dizaines de résidus d'intervalle dans la séquence. Du fait de leur changement potentiel de degré d'oxydation, ils interviennent notamment dans des réactions de transferts d'électrons mais aussi dans la reconnaissance de l'ADN comme les motifs liant le zinc (Bavoso, Ostuni et al. 1998; Lachenmann, Ladbury et al. 2004).

Le dernier degré de structuration des protéines est l'assemblage de plusieurs chaînes polypeptidiques qui va conduire à des édifices moléculaires de tailles considérables comme

les ribosomes, des anticorps ou des récepteurs membranaires. Nous développerons plus en détails la structure quaternaire lors de la description des domaines d'interactions.

## B. Processus De L'évolution Moléculaire

*D'après (Creighton 1984)*

Le principe d'évolution peut être appréhendé de deux manières différentes selon l'échelle à laquelle on se place mais l'une découle de l'autre. La première est celle développée par Charles Darwin qui a créé des liens de parentés entre les différentes formes de vie à partir d'observations faites sur leurs similitudes anatomiques. La deuxième est celle concernant les molécules de la vie. *A priori*, si l'on se base uniquement sur la chimie des différents organismes vivants rien ne les différencie du point de vue moléculaire: ils sont tous constitués des mêmes nucléotides, des mêmes acides aminés, des mêmes saccharides. Bref, des mêmes briques élémentaires à tel point que les mêmes isomères sont utilisés quelle que soit l'espèce (L-amino acides et D-ribonucléotides). De plus, la proportion de chaque acide aminé dans toutes les protéines, de tous les êtres vivants, est comparable. Ceci est largement dû au nombre de codons pouvant coder un même acide aminé (Tableau 6). De cette manière, la leucine est très représentée car codée par six codons différents alors que le tryptophane est beaucoup plus rare car codé par un seul codon. L'arginine fait toutefois exception. Ceci s'explique par le fait qu'elle provient du cycle de l'urée, voie de synthèse de l'ornithine, originellement le vingtième acide aminé. Il semblerait que l'arginine soit finalement rentrée en compétition avec l'ornithine et l'ait finalement emporté sur son précurseur. Mais malgré cette victoire, il n'en reste pas moins que l'arginine est moins représentée que ce que l'on pourrait prévoir. Ceci est sûrement dû au fait que la nature a préféré privilégier la lysine pour compenser l'ornithine (se différenciant de la lysine par un –CH2 surnuméraire). Malgré tous ces éléments communs aux différentes espèces du règne animal, il s'est bien passé quelque chose pour constater des différences évidentes entre la bactérie et l'éléphant.

Pour essayer de comprendre comment la vie a pu se diversifier à ce point, remontons à quelques 4,5 milliards d'années. La Terre se forme et n'est alors qu'un milieu hostile à la vie telle que nous la connaissons aujourd'hui. Les seules molécules qui prédominent sont le $CO_2$, CO, $N_2$ et $H_2O$. Par suite de bombardements ultraviolet et électrique (provenant de la foudre) les premières réactions photochimiques ont lieu pour donner les acides aminés, les nucléotides, les saccharides et les autres molécules constituant la Vie. Comme nous venons de l'évoquer, la conservation d'un seul isomère par type de molécule est troublante et bien qu'elle soit probablement le fruit du hasard on peut malgré tout parler de sélection du point de

vue de la Chimie. Puis, par un processus encore inconnu, la Chimie est devenu Vie. Le premier organisme vivant est apparu avec tout ce qui caractérise la vie à savoir la naissance, la croissance et la mort en passant par la procréation. Mais, la grande question vient du fait que pour se développer l'organisme a besoin de protéines pouvant accomplir différentes fonctions biologiques dont la réplication de l'ADN mais ces mêmes protéines sont en étroite relation avec le code génétique. On en revient à la grande question de l'humanité: lequel est venu avant l'autre: l'œuf ou la poule ?

De cet organisme ancestral -qui devait être un « extremophile » au vu des conditions climatiques et atmosphériques qui prédominaient en ces temps reculés, ont découlé toutes les formes de vie, fruits de la spéciation. La spéciation a certainement comme facteur prépondérant la localisation géographique des espèces. Celles-ci ont chacune muté et laissé survivre les individus viables au détriment des formes ne pouvant se reproduire à cause de mutations les privant d'une ou plusieurs fonctions vitales. Les différentes mutations génétiques sont de trois types: délétion, insertion ou substitution.

Dans le premier cas, le moindre mal est celui d'une délétion de trois codons consécutifs qui prive la protéine associée d'un acide aminé. Si cet acide aminé n'est pas primordial pour la conservation de la fonction, l'organisme survivra. Si cette mutation entraîne de graves conséquences sur le repliement et donc la fonction, il y a de fortes chances que l'organisme ne survive pas pour pouvoir transmettre à sa descendance ce gène mutant. Si la délétion intervient sur le premier ou le deuxième nucléotide c'est tout le cadre de lecture qui se trouve décalé donnant naissance à une séquence complètement différente de l'originale et donc n'assurant pas la viabilité de l'organisme. De la même manière, l'insertion d'un ou deux nucléotides aura un effet similaire sur le produit de la traduction. Enfin, le dernier cas est celui de la substitution. Comme le sous-entend le Tableau 6 de correspondance entre codon et acide aminé, c'est finalement le deuxième nucléotide qui conditionne la physico chimie de l'acide aminé.

| premier nucléotide | | second nucléotide: U | C | A | G | troisième nucléotide |
|---|---|---|---|---|---|---|
| | U | Phe | Ser | Tyr | Cys | U |
| | | | | | | C |
| | | Leu | | Stop | Stop | A |
| | | | | | Trp | G |
| | C | Leu | Pro | His | Arg | U |
| | | | | | | C |
| | | | | Gln | | A |
| | | | | | | G |
| | A | Ile | Thr | Asn | Ser | U |
| | | | | | | C |
| | | | | Lys | Arg | A |
| | | Met | | | | G |
| | G | Val | Ala | Asp | Gly | U |
| | | | | | | C |
| | | | | Glu | | A |
| | | | | | | G |

**Tableau 6: Codons des acides aminés dans l'ARN.**

Alors qu'une substitution du premier nucléotide donnera un acide aminé très proche chimiquement, un changement du troisième n'a souvent aucun effet sur le résidu résultant. Ainsi, on peut affirmer que des espèces proches auront des séquences de protéines très proches qui donneront des repliements comparables. Une mutation *a priori* peut toucher un résidu sans conservation de propriété si celui-ci a un rôle négligeable dans la structure ou la fonction, avec conservation de la physicochimie si cela optimise la fonction biologique de la protéine, mais elle se produira à une fréquence quasi-nulle sur des positions impliquées directement dans la fonction (ex: dans le cytochrome c les résidus impliqués dans la fixation de l'hème). Plus des protéines proviennent d'espèces proches plus leur homologie de séquence sera élevée du fait d'un long passé en commun.

Un autre événement génétique est à l'origine de la complexité croissante des organismes: la duplication de gènes. Ces phénomènes se produisant au moment de la mitose, vont avoir un impact direct sur l'augmentation du génome et donc du nombre de protéines coexistant dans les cellules. La duplication de gènes donnera naissance à des motifs répétés dans une même protéine, chaque domaine subissant indépendamment les lois de divergences par mutations successives. Une autre possibilité est la production en double exemplaire d'une même protéine. L'une conservera la fonction initiale alors que la seconde pourra muter à volonté pour acquérir d'autres fonctions ou compléter celles préexistant dans la protéine initiale. Un autre phénomène expliquant la diversité moléculaire est le crossing-over entre deux gènes

portés par deux chromosomes différents qui va mener à deux protéines hybrides dans lesquelles chacune portera un domaine initialement portée par l'autre (Vogel, Bashton et al. 2004).

Il faut bien rappeler que l'existence de telle ou telle protéine, suite à un processus génétique, qu'il soit de la simple mutation d'un nucléotide au crossing-over entre des portions de chromosomes, n'a été permis par la Nature que parce qu'il ne contrarie pas la viabilité et la pérennité d'une espèce. Au contraire, ces modifications auraient plutôt tendance à augmenter les performances de la machinerie cellulaire.

### *C. La Recherche Des Domaines D'interactions*

Les interactions entre protéines sont au centre des processus biologiques où des protéines vont s'associer à leurs partenaires pour remplir une fonction. La connaissance de ces complexes est d'une importance capitale dans le développement de nouvelles molécules à viser thérapeutique. En particulier, en virologie, cancérologie ou bactériologie l'intérêt est d'identifier les différents partenaires pour comprendre les mécanismes de développements de pathogènes d'une part, et de formuler des inhibiteurs pouvant perturber ces interactions, d'autre part. La connaissance de la topologie des interfaces s'avère indispensable pour formuler des molécules agissant de manière hautement spécifique sur ces complexes supramoléculaires: ceci constitue la base de la conception rationnelle d'inhibiteur aussi appelée *drug design*.

#### C.1. Alignements de séquences.

Toute étude phylogénétique débute par la comparaison de plusieurs séquences polypeptidiques. Les travaux de Margaret Oakley Dayhoff, publiés dans son *Atlas of Protein Sequence and Structure* (Dayhoff 1978), constituent la base de la recherche bioinformatique pour relier des protéines par des liens de parenté. Les principaux résultats sont basés sur 71 comparaisons entre groupes de séquences de protéines connues pour être de proches parents , c'est-à-dire identiques d'au moins 85%, afin de minimiser à la fois les ambiguïtés lors de l'alignement et le nombre de mutation coïncidentes. 71 arbres phylogénétiques ont ainsi pu être reconstitués en déterminant le nombre d'acides aminés divergents, d'une espèce à une autre, à l'intérieur d'un même groupe. L'exemple du cytochrome C est illustré en Figure 27.

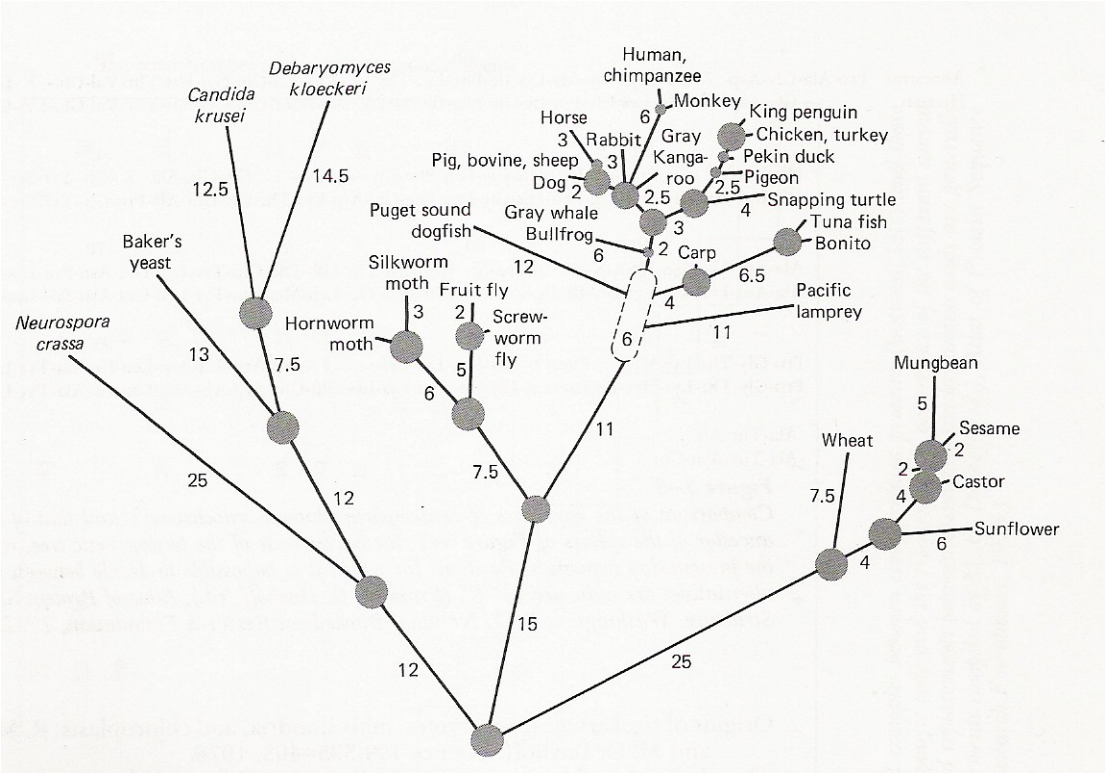



**Figure 27: Arbre phylogénétique du cytochrome C.**

Ces différents alignements ont donné naissance aux matrices PAM (Point Accepted Mutation) qui donnent la probabilité pour qu'un acide aminé soit muté par un autre tout en conservant la fonction de la protéine. La première matrice 1PAM (Percent Point Mutation) donne la probabilité qu'une substitution apparaisse sur 100 acides aminés.

Les matrices PAM donnent des probabilités exprimées en logarithmes décimaux ce qui permet une extrapolation des fréquences de mutations pour de plus longs intervalles de temps d'évolution. Ainsi la matrice 2PAM=1PAM x 1PAM ; 3PAM=2PAM x 1PAM,… A chacune de ces matrices XPAM correspond une matrice de mutation de Dayhoff appelée PAMX utilisée dans les alignements de séquences.

| | | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| C | 12 | | | | | | | | | | | | | | | | | | | |
| S | 0 | 2 | | | | | | | | | | | | | | | | | | |
| T | -2 | 1 | 3 | | | | | | | | | | | | | | | | | |
| P | -3 | 1 | 0 | 6 | | | | | | | | | | | | | | | | |
| A | -2 | 1 | 1 | 1 | 2 | | | | | | | | | | | | | | | |
| G | -3 | 1 | 0 | -1 | 1 | 5 | | | | | | | | | | | | | | |
| N | -4 | 1 | 0 | -1 | 0 | 0 | 2 | | | | | | | | | | | | | |
| D | -5 | 0 | 0 | -1 | 0 | 1 | 2 | 4 | | | | | | | | | | | | |
| E | -5 | 0 | 0 | -1 | 0 | 0 | 1 | 3 | 4 | | | | | | | | | | | |
| Q | -5 | -1 | -1 | 0 | 0 | -1 | 1 | 2 | 2 | 4 | | | | | | | | | | |
| H | -3 | -1 | -1 | 0 | -1 | -2 | 2 | 1 | 1 | 3 | 6 | | | | | | | | | |
| R | -4 | 0 | -1 | 0 | -2 | -3 | 0 | -1 | -1 | 1 | 2 | 8 | | | | | | | | |
| K | -5 | 0 | 0 | -1 | -1 | -2 | 1 | 0 | 0 | 1 | 0 | 3 | 5 | | | | | | | |
| M | -5 | -2 | -1 | -2 | -1 | -3 | -2 | -3 | -2 | -1 | -2 | 0 | 0 | 6 | | | | | | |
| I | -2 | -1 | 0 | -2 | -1 | -3 | -2 | -2 | -2 | -2 | -2 | -2 | -2 | 2 | 5 | | | | | |
| L | -8 | -3 | -2 | -3 | -2 | -4 | -3 | -4 | -3 | -2 | -2 | -3 | -3 | 4 | 2 | 8 | | | | |
| V | -2 | -1 | 0 | -1 | 0 | -1 | -2 | -2 | -2 | -2 | -2 | -2 | -2 | 2 | 4 | 2 | 4 | | | |
| F | -4 | -3 | -3 | -5 | -4 | -5 | -4 | -6 | -5 | -5 | -2 | -4 | -5 | 0 | 1 | 2 | -1 | 9 | | |
| Y | 0 | -3 | -3 | -5 | -3 | -5 | -2 | -4 | -4 | -4 | 0 | -4 | -4 | -2 | -1 | -1 | -2 | 7 | 10 | |
| W | -8 | -2 | -5 | -6 | -6 | -7 | -4 | -7 | -7 | -5 | -3 | 2 | -3 | -4 | -5 | -2 | -6 | 0 | 0 | 17 |
| | C | S | T | P | A | G | N | D | E | Q | H | R | K | M | I | L | V | F | Y | W |

**Tableau 7: Matrice de substitution PAM250.**

Par exemple, la matrice PAM250 (Tableau 7) donne la probabilité que 250 mutations différentes aient eu lieu pour une séquence de 100 acides aminés. Du fait des mutations silencieuses et des reversions de mutations, cette matrice correspond à des séquences homologues à 20%. Une valeur faible (négative) correspond à une substitution peu probable (ex: W/C = -8) alors qu'une valeur élevée correspond à une forte probabilité de substitution (ex: Y/F = 7).

Les matrices PAM sont un peu moins utilisées maintenant au profit des matrices BLOSUM (BLOcks SUbstitution Matrix). Les raisons sont les suivantes:

1. l'hypothèse de départ des matrices PAM est que tous les acides aminés ont une probabilité de mutation égale
2. le nombre de séquences protéiques à l'origine des matrices PAM (en 1978) était restreint en regard de l'échantillon de séquences protéiques à l'origine des matrices BLOSUM. Il est cependant à noter qu'en 1992, le nombre de séquences et de familles ont été actualisé par l'analyse de 16130 séquences pour 2621 familles de protéines.
3. l'échantillon de séquences protéiques à l'origine des matrices PAM était biaisé car la majorité des séquences sont celles des petites protéines globulaires.

Les matrices PAM ont donc été supplantées par les matrices développées par Henikoff & Henikoff (Henikoff and Henikoff 1992) au début des années 1990. Comme leur nom l'indique, ces matrices sont construites à partir de 2000 blocks provenant de 200 familles de

protéines différentes. Les blocks sont en fait des régions parfaitement conservées ne contenant ni insertions ni délétions. Une banque de données reliée à ces matrices est tenue à jour régulièrement pour y inclure de nouveaux blocks sur http://blocks.fhcrc.org/ (Henikoff, Greene et al. 2000).

La procédure consiste premièrement en un alignement de séquences très conservées pour obtenir des BLOCKs. Puis à partir de ces BLOCKs sont constitués des sous-ensembles contenant des portions de séquences à un degré d'identité donné. Ainsi, on obtient une probabilité de substitution pour chaque pair d'acides aminés dont découle la matrice BLOSUM. La matrice BLOSUM60 est obtenue pour 60% d'identité. Ces matrices se révèlent très efficaces pour des alignements locaux de séquences similaires d'où leur utilisation dans des logiciels de type BLAST (Altschul, Madden et al. 1997).

En pratique, il est préférable d'utiliser une matrice BLOSUM élevée ou une PAM faible pour des séquences très similaires. Inversement, pour des séquences plus éloignées on choisira une PAM élevée ou une BLOSUM faible.

### C.2. Trace de l'évolution.

La phylogénie par l'intermédiaire des alignements de séquences est d'une importance capitale pour l'élucidation des fonctions biologiques. Dans ce but, plusieurs équipes ont tenté depuis deux décennies de déterminer les régions essentielles qui confèrent la fonctionnalité des protéines. L'analyse des séquences de familles de protéines basée sur l'argument de la conservation impliquant la fonction a été d'autant plus consolidée par la considération des milliers de structures disponibles dans la Protein Data Bank (http://www.rcsb.org/pdb/statistics/holdings.do). A ce jour, plus de 50000 structures de protéines et d'acides nucléiques y sont réunies grâce à la contribution de la diffraction des rayons X (85%), de la Résonance Magnétique Nucléaire (14%) et par la microscopie électronique et autres techniques (1% restant, diffraction des électrons, neutrons et fibres, diffusion en solution, spectroscopie infrarouge...). Cette base de données permet de corréler des données tridimensionnelles redondantes au sein d'une famille de structures reliées entre elles par des liens phylogénétiques. C'est donc sur ce principe que Lichtarge (Lichtarge, Bourne et al. 1996) établit le principe de la Trace de l'Evolution (*evolutionnary trace* en anglais, ET) à partir d'alignements de séquences desquels des informations de conservations de certains épitopes sont déduites afin d'expliquer une fonction catalytique ou structurale. Ces résidus sont ensuite resitués dans l'espace, grâce aux structures disponibles dans la PDB, et

des cibles thérapeutiques sont alors identifiés (Innis, Shi et al. 2000; Sowa, He et al. 2001). Le principe de l'ET repose sur deux observations :

4. Les structures de protéines ayant un ancêtre commun divergent très faiblement au niveau des domaines « vitaux » impliqués dans la fonction qu'elle soit catalytique ou de reconnaissance moléculaire (Chothia and Lesk 1986).
5. Le corollaire de la première hypothèse est qu'il a été observé beaucoup moins de mutations durant l'évolution au sein de ces sites critiques (Zvelebil, Barton et al. 1987).

Par conséquent la méthode décrite par l'ET vise à identifier des sites importants pour les protéines faisant l'hypothèse que les domaines importants pour la protéine sont situés dans des environnements tridimensionnels similaires et que le taux de mutation à leur niveau est bien plus faible que partout ailleurs dans la protéine. Ceci a été mis en évidence par l'étude de familles de protéines dont les structures tridimensionnelles étaient connues. Le processus de détermination des traces s'initie par un alignement de séquences de plusieurs protéines appartenant à une même classe. Des sous-familles vont être générées en fonction du taux d'homologie entre les séquences et va donner lieu à la reconstruction d'un dendrogramme ou arbre phylogénique. Pour chaque sous-famille une première trace est estimée en considérant uniquement les résidus parfaitement conservés. Les séquences consensus identifiées sont à leur tour alignées entre elles pour ne faire ressortir que les résidus hypothétiquement importants pour les protéines. Une nuance est cependant introduite concernant des positions conservées à travers toute la famille de protéines mais divergeant dans leur nature: ces positions sont alors appelées classe-spécifiques. Finalement, la trace est reportée (par un code couleur) sur les structures tridimensionnelles et les domaines fonctionnels identifiés (Figure 28).

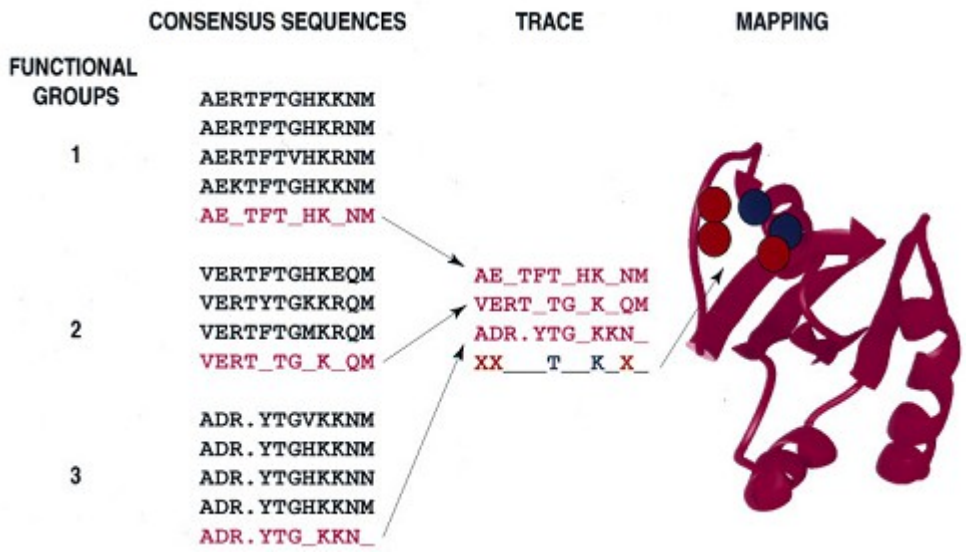



**Figure 28: Principe de la détermination de la trace de l'évolution selon Lichtarge** (Lichtarge, Bourne et al. 1996)**.** A partir de trois alignements de séquences propres à une sous famille de protéines une première séquence consensus pour chacune d'elle est déterminée. Ces séquences consensus sont à leur tour alignées afin d'en extraire la Trace de l'évolution. Les « X » représentent les résidus classe-spécifiques propres à chaque sous-famille, les lettres en majuscule sont les résidus parfaitement conservés. Le repositionnement de ces résidus dans une structure tridimensionnelle reflète la concentration topographique des résidus parfaitement conservés et classe spécifiques.

Mais la trace de l'évolution ne peut à elle seule permettre une prédiction des domaines d'interaction des protéines puisque la conservation de certains résidus implique deux possibilités à savoir la constitution d'un site catalytique ou l'identification d'un épitope lié à la reconnaissance moléculaire. Il est donc indispensable de pouvoir différencier ces deux classes par une analyse plus fine des interfaces protéine-protéine impliquées dans de nombreux mécanismes biomoléculaires.

### C.3. Analyses des Hot spots

L'analyse en composition des interfaces protéiques est un champ de recherche très intense tant l'enjeu est important. En effet, les processus moléculaires sont connus pour faire intervenir plusieurs chaînes polypeptidiques. Celles-ci peuvent s'assembler de manière permanente ou temporaire. Dans tous les cas, ces interactions doivent être spécifiques ce qui fait appel à la notion de reconnaissance moléculaire. Du point de vue pharmaceutique, l'intérêt de comprendre les règles physicochimiques qui régissent ces interactions peut aboutir

à la conception d'inhibiteurs spécifiques pouvant bloquer la formation de complexes protéiques d'origines virales notamment.

Dans ce but, plusieurs équipes se sont lancées dans la rationalisation des règles régissant les interactions protéine-protéine. Toutes se sont basées sur des études statistiques impliquant des échantillons de quelques dizaines (Janin and Chothia 1990; Young, Jernigan et al. 1994; Bogan and Thorn 1998) à plusieurs centaines (Tsai, Lin et al. 1997; Rodier, Bahadur et al. 2005) de structures tridimensionnelles de complexes contenus dans la PDB. A ce jour, l'étude employant le plus grand nombre de complexe est celle de Ofran et Rost, menée en 2003 (Ofran and Rost 2003). La façon de classer les structures d'un échantillon statistique mène à la différentiation des structures selon divers critères. Ceux-ci incluent le caractère temporaire ou permanent des complexes (Glaser, Steinberg et al. 2001; Ansari and Helms 2005), les familles biologiques de protéines (anticorps/antigène, immunoglobulines, serine protéase/inhibiteurs…) (Janin and Chothia 1990; Jones and Thornton 1996; Lo Conte, Chothia et al. 1999; Hu, Ma et al. 2000), les structures secondaires impliquées dans les interfaces (Eilers, Patel et al. 2002; Ansari and Helms 2005) ou en combinant une distinction sur la nature homo- ou hétérodimérique du complexe avec leur caractère plus ou moins permanent (Ofran and Rost 2003). Après avoir défini les groupes d'interfaces il reste à déterminer la distance à partir de laquelle deux résidus de deux chaînes différentes interagissent. Cette distance est généralement prise entre 3,5 Å pour des interactions entre les atomes du squelette et jusqu'à 6 Å pour les atomes lourds des chaînes latérales (Hu, Ma et al. 2000; Glaser, Steinberg et al. 2001; Ansari and Helms 2005; Keskin, Ma et al. 2005). En recoupant les différentes informations on peut distinguer deux familles de complexes comme les ont décrits Ofran & Rost: hétéro- et homo- obligomères et hétéro- et homo- complexes.

Les homo- et hétéro- obligomères sont formés de deux chaînes identiques ou différentes, respectivement, liées de manière permanente pour assurer continuellement une fonction biologique. L'hémoglobine est un exemple d'hétéro- obligomères.

Les homo- et hétéro complexes sont décrits selon les auteurs comme des édifices moléculaires formés de manière ponctuelle comme les protéases à sérine avec leurs substrats. Une fois le clivage de la chaîne peptidique effectué le substrat est séparée de l'enzyme pour que cette dernière accomplisse une nouvelle catalyse (Janin and Chothia 1990; Jones and Thornton 1996).

La classification des interfaces sur le critère d'homologie de séquence entre les chaînes entrant en contact ne montre pas toujours de différences flagrantes (Ofran and Rost 2003;

Rodier, Bahadur et al. 2005). Par contre, le caractère permanent ou temporaire d'une interaction peut être mis en équation par la considération de plusieurs facteurs physicochimiques. D'une manière générale, on peut distinguer trois éléments que sont la taille, la forme et la balance existante entre résidus hydrophobes et polaires présents à l'interface.

1. La différenciation des différents complexes en fonction de leur taille est largement dépendante de l'échantillon statistique de structures considérées. En 1990, Janin et al. définissent quinze interfaces comme petites quand celles-ci n'excèdent pas les 600Å² et grandes au-delà de 1600 Å² (Janin and Chothia 1990). Bogan, en 1998, constate sur une vingtaine de dimères et monomères que la taille minimum d'une interface est généralement supérieure à 600 Å² (Bogan and Thorn 1998). En 2001, l'équipe de Glaser les considère petites en deçà de 1000 Å² et grandes au-dessus de 5000 Å² et ce sur 621 interfaces homo- ou hétérodimériques (Glaser, Steinberg et al. 2001). En confrontant, d'une part, des études menées sur des complexes temporaires (Janin and Chothia 1990; Ansari and Helms 2005) montrant en particulier l'implication en moyenne de 30 à 70 acides aminés aux interfaces et, d'autre part, celles sur l'observation d'obligomères qui met en relief une redondance de taille d'interface croissante avec le poids moléculaire des partenaires (Jones and Thornton 1996) ; on arrive à la conclusion que la taille de l'interface va de pair avec la longévité du complexe.
2. En plus de la taille, la forme des interfaces est aussi déterminante. Jones & Thornton ont montré en particulier que les hétérocomplexes se forment à travers des interfaces planes alors que les homocomplexes exhibent des formes s'enchevêtrant les unes dans les autres stabilisant ainsi fortement le complexe (Jones and Thornton 1996). Plusieurs autres exemples mentionnent des interactions multisegmentielles c'est-à-dire faisant intervenir plusieurs épitopes éloignés dans l'espace (Lo Conte, Chothia et al. 1999; Ofran and Rost 2003; Keskin, Ma et al. 2005). Par contre, le lien établi par Jones & Thornton entre la nature homo- ou hétéro dimère semble fortuite probablement due au fait que l'ensemble statistique analysé comportait des obligomères principalement homodimériques et des complexes temporaires plutôt hétérodimériques. Une interpénétration bloquera le complexe dans une conformation stable alors qu'une interface plane facilitera sa dislocation par gain d'entropie déstabilisante (Janin and Chothia 1990). Par contre, une équiprobabilité des différents types de structures secondaires a été observée aux interfaces avec toutefois des appariements préférentiels

entre hélices d'une part et entre feuillets d'autre part. Les fragments dépourvus de structures secondaires (*random coil*), du fait de leur flexibilité, ont quant à eux une faculté d'adaptation de leur conformation envers n'importe quel type de partenaire (Jones and Thornton 1996; Ansari and Helms 2005).

3. Le caractère permanent ou temporaire d'un complexe dépend du nombre et de la nature des interactions siégeant au sein de l'interface. Ceci se répercute donc sur sa composition en acide aminé. Comme pour le repliement, la liaison des chaînes polypeptidiques est influencé par l'environnement aqueux dans lequel elles baignent comme le sous-tend le modèle de la goutte d'huile (Tsai, Lin et al. 1997; Purkiss, Skoulakis et al. 2001). La répartition des molécules d'eau est intimement liée à celle des résidus hydrophobes qui par définition chassent l'eau. Cependant, en mettant à contribution les doublets non-liants de son atome d'oxygène, l'eau peut se loger dans des cavités et établir deux liaisons hydrogènes afin de ponter deux résidus pour favoriser une conformation. Rodier et al. ont estimé que l'eau participe à hauteur d'un tiers des liaisons hydrogènes dans lesquels elle est impliquée comme donneuse d'hydrogènes avec les carbonyles du squelette peptidique et environ 12% en tant qu'accepteur avec les amides du squelette (Rodier, Bahadur et al. 2005). Sinon, l'eau forme des liaisons hydrogènes avec Asp, Glu, Arg, Ser et Tyr mais beaucoup moins avec Lys (Purkiss, Skoulakis et al. 2001; Rodier, Bahadur et al. 2005). Cependant, la présence d'eau peut s'avérer indésirable pour la stabilisation des complexes. C'est pourquoi dans les complexes permanents 80% à 90% de la surface d'interaction est déshydratée (Figure 29) (Janin and Chothia 1990; Purkiss, Skoulakis et al. 2001).

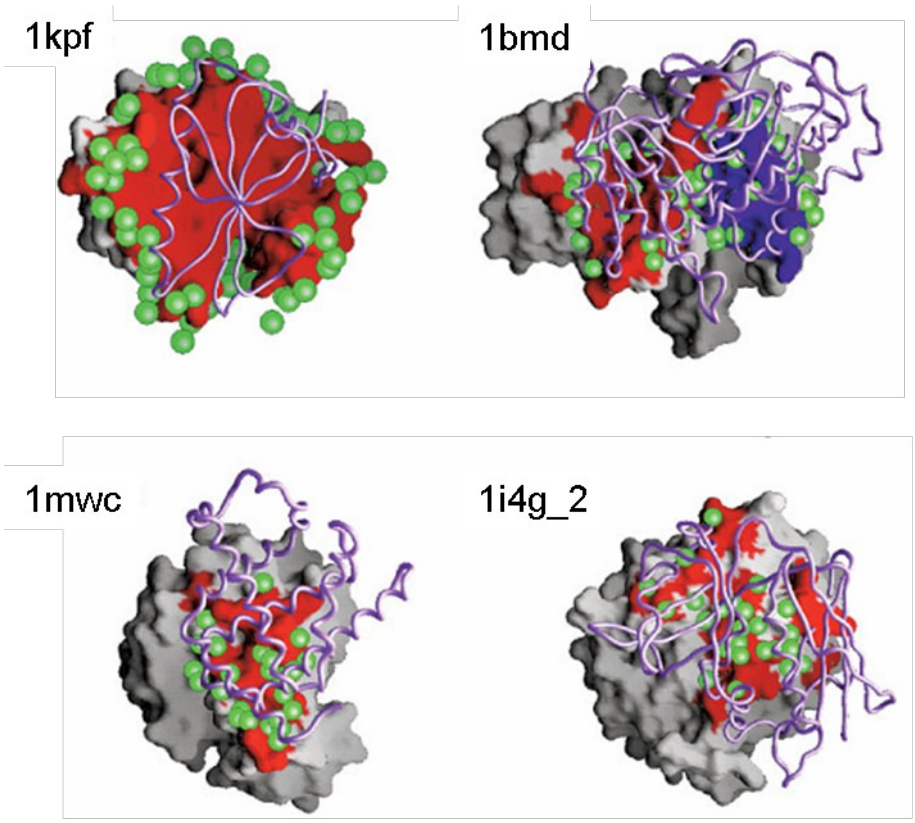



**Figure 29: Répartition des molécules d'eau dans différentes interfaces.** Dans chaque image une protéine est représentée en surface et l'autre en ruban. Les sphères vertes représentent les molécules d'eau. Les interfaces sont colorées en rouge ou bleu. Les codes PDB des complexes sont précisés dans chaque cas (Rodier, Bahadur et al. 2005).

Du point de vue de la composition en acides aminés, ce phénomène se traduit par une large proportion de résidus hydrophobes (60% environ) (Janin and Chothia 1990; Purkiss, Skoulakis et al. 2001; Ofran and Rost 2003; Fooks, Martin et al. 2006). Ces hydrophobes (Leu, Ile, Val, Ala) ont, du fait de leurs chaînes latérales flexibles, la capacités de favoriser un *packing* efficace des interfaces. Pour des complexes de grandes tailles l'hydrophobicité est importante toujours dans le même but: stabiliser le complexe (Janin and Chothia 1990; Tsai, Lin et al. 1997; Ansari and Helms 2005). Le taux d'hydratation d'une surface est souvent exprimé en ASA (*Accessible Surface Area*) c'est-à-dire la surface globale ou partielle d'un polypeptide qui rentre en contact direct avec l'eau. Plus la zone observée est hydrophobe plus l'ASA sera faible. Dans le but de favoriser l'interaction, l'ASA va diminuer progressivement de l'extérieur vers l'intérieur de l'interface (Keskin, Ma et al. 2005). En terme d'ASA, les petites interfaces sont accessibles à 25% et les grandes à 5% (ce qui concorde avec les chiffres de Rodier énumérés plus haut). On s'aperçoit que les résidus importants pour l'interaction protéine-protéine sont protégés de l'eau par un gradient d'hydrophobicité croissante depuis

l'intérieur de l'interface vers l'extérieur. On appelle ses résidus *hot spots* car ils contribuent de manière significative à la reconnaissance et à la stabilisation du complexe qu'il soit temporaire ou permanant (Bogan and Thorn 1998). La quantification de leur implication dans l'interaction est mesurée via *alanine scanning* (mutation ponctuel d'un résidu en alanine) en mesurant la différence d'énergie entre la forme native et la forme mutée du complexe (Figure 30). On estime généralement qu'un hot spot apporte une contribution à l'énergie de stabilisation du complexe d'au moins 2 Kcal/mol (Bogan and Thorn 1998; Moreira, Fernandes et al. 2007). Cependant, cette mesure peut être faussée par une contribution concertée de plusieurs résidus comme le font les hydrophobes par la coopérativité de plusieurs interactions de Van Der Walls isolément faibles. C'est pourquoi un autre type d'expérience a émergé: l'*alanine shaving*. Cette expérience consiste à muter en alanine non pas un mais plusieurs résidus en même temps (Bogan and Thorn 1998; DeLano 2002).

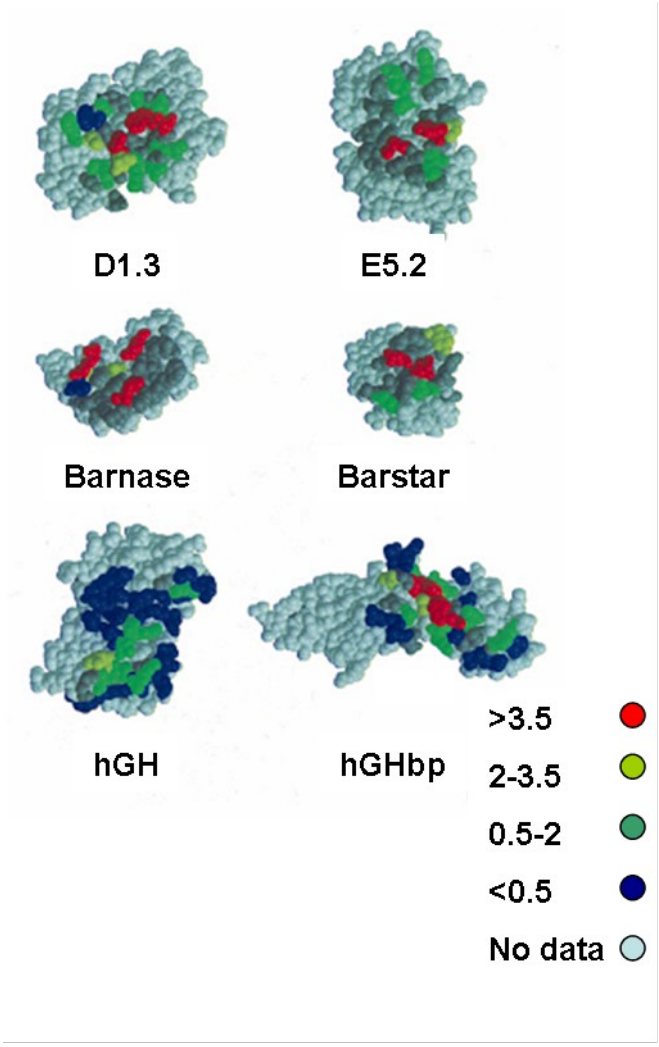



**Figure 30: Contribution énergétique des résidus au sein de l'interface**. La contribution des acides aminés impliqués dans l'interaction avec le partenaire est représentée par le code

couleur précisé, exprimée en Kcal/mol mesurée par *alanine scanning* (Bogan and Thorn 1998).

Une corrélation a été établie entre le taux de conservation d'un acide aminé et sa propension à être un hot spot. Il semblerait que chaque famille de protéines conserve tout au long de son arbre phylogénique une « empreinte digitale » qui lui est propre (Hu, Ma et al. 2000; Keskin, Ma et al. 2005). L'étendue de la conservation le long de l'interface est cependant variable. Par exemple, chez les anticorps le taux de conservation est homogène sur la totalité du site de contact et en particulier concernant les résidus aromatiques qui remplissent la fonction de reconnaissance moléculaire (Janin and Chothia 1990; Hu, Ma et al. 2000). Par contre, dans le cas des protéases à sérine la conservation suit un gradient dégressif de l'intérieur vers l'extérieur de l'interface (Keskin, Ma et al. 2005). On remarque que ce gradient est inversé par rapport à celui suivi par l'hydrophobicité. Ces comportements croisés du facteur hydrophobe avec le caractère phylogénique sont une illustration physique du rôle de chaque type de force (covalente, ionique, dipolaire et hydrophobe) dans la réalisation de l'interaction protéine-protéine. L'idée selon laquelle les hot spots sont protégés de l'eau par un cortège d'interactions hydrophobes impliquant des résidus hydrophobes modérément conservés (Hu, Ma et al. 2000) constitue la base de l'hypothèse de l'anneau (*O-ring hypothesis*) (Bogan and Thorn 1998; Rodier, Bahadur et al. 2005; Moreira, Fernandes et al. 2007). Ceci rejoint clairement les études entreprises sur les traces de l'évolution établie par Lichtarge (Sowa, He et al. 2001; Zhu and Tytgat 2004). Un troisième gradient qui suit remarquablement celui de la conservation est celui concernant les énergies d'interaction mises en jeu dans l'interface comme l'illustre Bogan qui montre la relation entre $\Delta\Delta G°$ (différence d'énergie entre forme sauvage et mutée par alanine scanning) et ASA.

Hu explique que les interactions hydrophobes alentours sont importantes pour la stabilité du complexe et les interactions électrostatiques conditionnent l'orientation des chaînes pour la reconnaissance moléculaire spécifique (Hu, Ma et al. 2000). La stratification des interactions le long de la surface explique pourquoi différents types de résidus sont localisés de manière redondante au cœur, pour certains, et aux alentours, pour d'autres, de l'interface. On comprend donc pourquoi on rencontre au centre des arginines, des aromatiques (Bogan and Thorn 1998) et des polaires tant ils peuvent offrir de possibilités de former des liaisons hautement énergétiques (Glaser, Steinberg et al. 2001; Ofran and Rost 2003; Ansari and Helms 2005). Les cystéines sont représentées dans tous les types de complexes grâce à leur faculté exceptionnelle à pouvoir former des ponts disulfures et donc lier de manière covalente

deux molécules (Tsai, Lin et al. 1997; Glaser, Steinberg et al. 2001; Ofran and Rost 2003; Ansari and Helms 2005). L'arginine est exceptionnelle elle aussi car elle peut offrir cinq types d'interactions. En effet, elle peut former i. cinq liaisons hydrogènes simultanément grâce à son groupement guanidino, ii. des ponts salins avec des aspartates ou des glutamates, iii. des interactions dipolaires du fait de l'asymétrie marquée de son nuage électronique sous l'impulsion de son groupement électroattracteur guanidino, iv. des *π-stacking* grâce à la délocalisation de sa charge en bout de chaîne, et enfin de manière plus surprenante v. des interactions hydrophobes *via* sa chaîne aliphatique à trois méthylènes. Ce dernier point explique pourquoi des appariements Lys/Arg sont observés malgré des charges de même signe. C'est en fait une véritable interaction hydrophobe qui y est formée (Bogan and Thorn 1998; Glaser, Steinberg et al. 2001; Ansari and Helms 2005). On note aussi un appariement $Trp^{Arom.}/Arg^{aliphatique}$ au détriment de $Trp^{Arom.}/Arg^{guanido}$. Les ponts salins impliquent aussi les appariements Lys/Asp ou Lys/Glu. Les interactions dipolaires impliquent, en plus des Arg, Lys, Asp et Glu déjà cités, Ser, Tyr et Thr via leurs groupes hydroxyles, Asn et Gln via leurs groupements amides. Là encore, une nuance est faite sur une abondance plus importante des aspartates et asparagines par rapport aux glutamates et glutamines qui présentent un $CH_2$ surnuméraire, leur donnant un gain d'entropie supplémentaire défavorable à la stabilité -même temporaire- du complexe (Bogan and Thorn 1998). Les résidus aromatiques Tyr, Trp, Phe et His (en fonction du pH) sont aussi très représentés grâce aux interactions de *π-stacking*, ponts hydrogènes ou dipolaires via leurs hétéroatomes respectifs. Ce dernier point explique justement pourquoi les tyrosines sont plus représentées que les phénylalanines car elles offrent deux possibilités supplémentaires de créer une interaction avec un partenaire. Pour finir, les interactions hydrophobes entourant la zone d'interaction font intervenir des résidus aliphatiques (y compris la proline) ou polaires à caractère hydrophobes (Met et Thr).

Pour conclure cette partie, dressons le portrait type d'un complexe permanent et d'un complexe temporaire. Un complexe permanent présente une large interface hydrophobe, multisegmentielle qui fait s'imbriquer les chaînes les unes dans les autres. On peut comparer le complexe formé comme le produit d'une seule protéine originelle dont le gène la codant aurait été coupé en deux au cours de l'évolution et les deux nouvelles entités qui en découlent auraient été programmées pour s'associer afin de reformer le cœur hydrophobe tout comme celui formé par leur ancêtre. Un complexe temporaire est constitué d'une interface de petite taille (facilement dissociable dès que la fonction biologique est accomplie) au cœur duquel serait logés des aminoacides sélectionnés par l'évolution pour leurs capacités à former diverses interactions simultanées, afin d'augmenter la spécificité et l'orientation d'une

molécule par rapport à son partenaire. Ces résidus très polaires, voir chargés, sont entourés de résidus hydrophobes plus modérément conservés dont le rôle est de chasser les molécules d'eau capable de perturber l'interaction.

La prédiction des sites d'interactions génère un engouement d'une grande ampleur dans le domaine de la bioinformatique mais à ce jour il existe peu de logiciels capables de prédire ces domaines en s'affranchissant de structures tertiaires contenues dans la PDB (Neuvirth, Raz et al. 2004; Counterman Burba, Lenhert et al. 2006; Ofran and Rost 2007). La prédiction à partir de la seule structure primaire est un but encore jamais atteint.

## VI. OUTILS THÉORIQUES ET EXPÉRIMENTAUX

### *A. Choix Des Techniques Utilisées Pour La Résolution De La Problématique*

Comme nous venons de le préciser, les systèmes faisant l'objet de notre étude sont de taille considérable. Les outils aujourd'hui disponibles pour la résolution de structures à l'échelle atomique sont au nombre de trois ayant chacun leurs avantages et leurs limites. Deux d'entre elles sont expérimentales et de ce fait sont dès le départ dépendantes de la production en quantité élevées (quelques milligrammes, en fonction de la masse molaire de l'échantillon) à un degré de pureté supérieur à 95%.

La cristallographie par diffraction des rayons X est une technique nécessitant la périodicité régulière de l'échantillon s'organisant en un réseau cristallin. Les cristaux obtenus sont alors soumis aux rayons X permettant l'obtention d'une carte de densité électronique de chaque atome de la molécule. L'avantage majeur de la diffraction des RX est la possibilité de résoudre sans limite de taille tout système pouvant s'organiser en cristaux. L'obtention des cristaux demande une étude systématique de conditions de cristallisation s'opérant dans des plaques 96 puits (et plus) passant en revue différentes combinaisons de sels et de pH. Il est parfois utile de tronquer des séquences contenant des portions hydrophobes empêchant la cristallisation (problème des protéines transmembranaires) ou opérer à des mutations ponctuelles (voir le cas de gp17 présenté dans la partie traitant des domaines hélicases des bactériophages, page 62). Les inconvénients reposent sur des considérations liées à l'influence de la phase cristalline dense et figée qui empêche l'étude dynamique de processus biologiques et qui peut parasiter le *packing* des entités observées. Un autre inconvénient réside dans la résolution de la densité électronique détectée. Celle-ci équivaut à la taille d'un hydrogène et implique donc que les hydrogènes, et par extension les ponts hydrogènes, doivent être déduits intuitivement et non observés.

La Résonance Magnétique Nucléaire permet une étude en solution des échantillons reproduisant de manière plus fidèle le milieu du vivant (phase liquide aqueuse) et de ce fait d'étudier la dynamique des entités chimique observées. Contrairement à la diffraction des rayons X, la RMN est quant à elle limitée en taille (200 résidus maximum en utilisant des marquages isotopiques divers et des séquences d'impulsions élaborées). Quelques exemples surpassant cette limite sont présents dans la littérature mais ils restent encore rares (Sprangers and Kay 2007). Ainsi, il est possible de reconstruire des structures de sous-domaines de

protéines ou des petites protéines (éventuellement en complexe aves des acides nucléiques suivant le nombre de signaux présent sur les cartes RMN).

Une dernière approche théorique, est la reconstruction *in silico* de modèles moléculaires. Les prédictions de structures tertiaires s'opèrent de trois manière différentes: modélisation comparative ou par homologie (Sali and Blundell 1993), par reconnaissance de repliement *(fold recognition,* FR) aussi appellé *threading* (enfilement), et repliement *ab initio*. Les différents serveurs sont classés en fonction de la qualité des modèles générés pour les symposiums biennales du Critical Assessment of Structure Prediction (CASP) (Cozzetto, Giorgetti et al. 2008) (http://predictioncenter.org/casp8/index.cgi)

Le CASP est une compétition faisant s'affronter, sur la prédiction de plusieurs protéines de structures inconnues, les différentes équipes du monde entier développant des algorithmes de reconstruction de protéines. Les qualités des structures finales sont estimées, les alignements de séquences et le champ d'action de la méthodologie employée (taille des systèmes prédits) sont examinés.

La reconstruction par homologie de séquence et de *threading* construisent des modèles à partir d'alignement de séquences entre une séquence cible et une séquence de protéine dont la structure entière ou en partie est connue et déposée dans la PDB. A partir du ou des alignement(s) de séquence(s), le squelette peptidique sera reconstitué puis les chaînes latérales rajoutées. La différence notable entre la modélisation comparative et la méthode d'enfilage réside dans le pourcentage d'homologie entre la séquence cible et les structure-supports.

La méthode *ab initio* est quant à elle réservée dans le cas où aucun homologue de la protéine cible est connue. L'inconvénient majeur de cette méthode est le temps de calcul qui peut être considérable (plusieurs mois) (Helles 2008).

Évidemment, la qualité des modèles générés dépend des structures disponibles dans la Protein Data Bank notamment. La base même de la reconstruction par homologie de séquence implique que la prédiction sera seulement possible par l'existence d'un homologue proche – d'un point de vue évolutif. Dans le cas contraire, et en particulier lorsqu'une protéine contient plusieurs domaines catalytiques indépendants, il est préférable de soumettre isolément ces domaines puis de les relier entre eux. Dans le cas où la structure primaire ne permet pas de retrouver un homologue dans la banque de données, les prédictions de structures secondaires prennent le relais et permettent de reconnaître des homologues structuraux par reconnaissance de repliement (*Fold recognition*) correspondant à des enchaînements particuliers d'éléments

de structures secondaires (brins bêta ou hélice alpha) correspondant à une famille de repliement protéique. De cette manière deux protéines éloignées sur le plan de la structure primaire peuvent mener à des repliements identiques.

Cet aspect renvoie au paradoxe de Leventhal qui aboutit à la conclusion que des chemins de repliements préférentiels ont été adoptés par la nature menant à un nombre limité de famille de repliement en comparaison avec l'infinité de famille statistiquement possible si l'on considère l'ensemble des degrés de liberté présents dans une protéine.

Comme nous venons de le sous-entendre, la prédiction théorique de modèles est sans limite de taille. La qualité de la prédiction est quant à elle largement dépendante des homologues de la protéine cible présents dans les banques de données. Dans tous les cas, l'inconvénient majeur de cette technique est qu'elle reste théorique mais elle permet néanmoins d'envisager des hypothèses de travail pour guider une analyse expérimentale ultérieure.

### *B. Démarche Bioinformatique Pour L'analyse De La Séquence Primaire*

Le protocole décrit ci-dessous a été appliqué aux différentes protéines qui ont fait l'objet de l'étude décrite dans ce manuscrit.

A travers toute la littérature, nous avons dans un premier temps sélectionné un maximum de séquences d'homologues de nos protéines d'intérêt, pUL89 et pUL56. Cette première étape est déterminante dans l'analyse des domaines conservés qui doivent refléter les fonctions communes aux terminases de l'ensemble des herpesvirus et donc vitaux pour leur dévelopement. Nous nous sommes donc efforcé de rassembler les homologues de pUL89 et par la suite de pUL56 en nous basant sur la classification décrite dans la littérature (Davison 2002; McGeoch, Rixon et al. 2006).

Par ailleurs, toujours en se basant sur la littérature, nous avons « préclassé » les différentes séquences de protéine en fonction de leur genre respectif (α, β, ou γ) afin de tenir compte de la distance évolutive entre les différents homologues. Les différentes séquences ont été par la suite alignées en utilisant le programme d'alignement de séquences CLUSTAL W disponible, entre autres, sur le Network Protein Sequence Analysis du Pôle Bio Informatique Lyonnais (http://npsa-pbil.ibcp.fr/cgi-bin/npsa_automat.pl?page=/NPSA/npsa_server.html). Nous avons utilisé pour notre étude les paramètres par défaut.

Une fois alignée, chaque position a été vérifiée manuellement pour éviter les sur- ou sous-estimation de conservation. Pour se faire, nous nous sommes rapportés au diagramme de Venn présenté en introduction (Figure 25) et adopté la convention suivante.

Pour une position donnée, un acide aminé est dit **parfaitement conservé** lorsqu'il est retrouvé à l'identique dans 100% des séquences ; **hautement conservé** lorsqu'une propriété physicochimique appartenant à l'un des petits sous-ensembles *Aromatique*, *Aliphatique*, *Positivement Chargé*, *Négativement Chargé* ou *Minuscule* est conservée dans 100% des séquences ; et **moyennement conservé** lorsque un acide aminé appartenant à l'un des grands sous-ensembles *Hydrophobe*, *Dipolaire* ou *Petit* est systématiquement rencontré. Dans tous les autres cas (gap ou conservation inferieure à 100%) la position sera considérée **non conservée**.

Afin de rendre l'analyse physicochimique plus visuelle, nous avons par la suite revu la notation en nous efforçant d'adopter une notation la plus intuitive possible (la lettre grecque représentant un domaine ou sous domaine correspond à la première lettre ou syllabe du nom de ce groupe en français). Ainsi, en vertu du diagramme de Venn, les A**ro**matique (Y, F, H, W) sont notés **ρ**, les **A**liphatiques (I, P, L, A,V) **α**, les chargés **P**ositivement (R,K,H) **π**, les chargés **N**égativement (E et D) **ν**, les minuscules -**T**iny en anglais- (A, G, C, S) **τ**, les Hydro**ph**obes **φ**, les petits –**S**mall en anglais- **σ** et les **D**ipolaires **δ**. A partir de ce code, les traces pour chaque genre (**ALPHA**, **BÊTA** et **GAMMA**) puis, la trace de l'évolution (**ET**) de tous les homologues ont été générées.

Dans **ET**, les parfaitement conservés sont colorés en rouge, les hautement en orange et les moyennement en vert. Les positions « classe-spécifique » -i.e. conservées dans un genre et variant d'un genre à un autre- sont notées **x** lorsqu'une propriété physicochimique varie d'un genre à un autre et **X** lorsqu'une position est parfaitement conservée dans tous les genres pris isolement mais varie d'un genre à un autre (sans relation physicochimique évidente entre eux).

Finalement, pour une meilleure lisibilité nous avons numéroté les positions de l'alignement en nous référant aux séquences d'intérêt, en l'occurrence pUL89 et pUL56.

En complément de l'alignement de séquences, nous avons jugé utile d'expliciter la prédiction de structure secondaire grâce à PSI-PRED (http://bioinf.cs.ucl.ac.uk/psipred/). La prédiction de la structure secondaire est un élément clé dans la prédiction de la structure tridimensionnelle d'une protéine par la méthode de *threading* qui permet de surmonter la distance évolutive entre une protéine de structure inconnue et celles utilisées pour reconstruire

un modèle théorique. Le *threading* consiste en la reconnaissance de repliements tertiaires locaux d'une protéine par la reconnaissance d'enchaînements de structures secondaires prédites qui vont être comparés aux différentes familles de repliement répertoriées dans les bases de données.

Un algorithme visant à reconstruire des modèles en se servant de cette logique porte le nom évocateur de Frankenstein monsters. L'analogie avec le croquemitaine de Mary Shelley est en effet saisissante. Dans son roman le docteur Frankenstein reconstruit un corps à l'aide de fragments d'autres et donne vie à celui-ci à l'aide de la foudre. La démarche du *threading* génère différents fragments théoriques (membres) d'une protéine, les assemble pour donner un corps inerte et minimise ce dernier pour lui conférer une intégrité structurale.

Le découpage des zones conservées est défini par l'accumulation de positions conservées appartenant à un même élément de structure secondaire et/ou contenu dans une région dont le caractère hydrophobe est le même (plutôt hydrophile ou plutôt hydrophobe). Ce dernier critère prend en considération l'accessibilité au solvant des différentes régions de la séquence primaire.

L'estimation de ce caractère hydrophobe est déduite de graphiques d'hydrophobicité calculés *via* Protscale, avec deux échelles différentes (Eisenberg et Kite & Doolitlle). En effet, nous partons du principe que l'hydrophobicité reflète la capacité d'un domaine à s'exposer ou au contraire à se cacher de l'eau. Ainsi, nous pouvons faire l'approximation raisonnable qu'un domaine hydrophile sera situé à la surface de la protéine alors qu'un domaine hydrophobe appartiendra au cœur de la protéine ou participera à la formation d'un complexe avec ses partenaires polypeptidiques. En considérant un complexe protéine – protéine dans son ensemble comme une protéine à part entière, l'interface entre les sous-unités reconstitue partiellement ou complètement le cœur hydrophobe selon que l'interaction soit temporaire ou permanente.

Précisément dans notre cas, il ne faut pas perdre de vue que les protéines étudiées ici sont des chaînes polypeptidiques de tailles importantes (674 et 850 acides aminés pour pUL89 et pUL56, respectivement) dont l'une d'entre elles – pUL56- dimérise. De plus chez les herpesvirus ces deux partenaires sont colocalisés pendant plusieurs étapes du cycle infectieux de CMVH. En effet, ils travaillent de concert tout au long d'un processus biologique faisant impliquant i. l'assemblage de ces deux protéines dans le cytoplasme, ii. l'importation dans le noyau du complexe, iii. la reconnaissance de l'ADN, iv. la translocation de l'ADN dans la procapside et le clivage de l'ADN par deux fois. Ainsi, on peut s'attendre à ce que l'interface

entre ces deux protéines fasse intervenir plusieurs segments déssiminés dans la structure primaire afin de stabiliser sa constitution.

### *C. Synthèse Peptidique*

La conception d'inhibiteurs hautement spécifiques de l'interaction entre ces deux protéines nécessite tout d'abord la connaissance des interactions existant à l'interface entre elles grâce à la résolution de leur structure. Cependant, la production d'échantillon polypeptidique n'est possible que par deux voies: la synthèse chimique et l'expression chez des systèmes biologiques (bactéries, insectes…).

La synthèse chimique a l'avantage de pouvoir contrôler le marquage ponctuel dans la séquence primaire ($^{15}N$, $^{13}C$) qui peut s'avérer utile pour l'attribution des signaux RMN d'acides aminés, ou introduire des acides aminés modifiés incorporant éventuellement des sondes fluorescentes pour le traçage des peptides dans des organismes vivants ou des groupements biotinylés.

Cependant, les inconvénients de la synthèse chimique sont inhérents à la nature (proportion d'hydrophobes, introduction de glycines ou prolines) et la taille de la séquence qui conditionnent le rendement final de la synthèse. En effet, une synthèse avec un rendement de l'ordre de 80% peut être attendu pour des peptides jusqu'à 20 résidus (séquence dépendant) mais décroit fortement avec la taille du peptide. Il est fastidieux d'aller au-delà de 100 acides aminés et les rendements sont souvent de l'ordre de 10% ce qui reste intéressant quand on ne peut obtenir d'échantillon par une autre voie de synthèse.

En contre partie, la deuxième voie de synthèse utilisant des systèmes bactériens dont seule la séquence d'ADN codant est nécessaire pour la production du peptide (ou de la protéine désirée) qui peut alors être récupérée avec des rendements variables selon sa toxicité pour la bactérie. Dans toute notre étude sur pUL89 ou pUL56, nous avons été confronté à ces problèmes de toxicité et c'est pour cette raison que nous nous sommes tournés vers la synthèse peptidique en phase solide pour l'obtention de nos échantillons analysés par RMN.

### *D. Résonance Magnétique Nucléaire, Attribution Et Modélisation Sous Contraintes*

La résonance magnétique nucléaire (RMN) est une spectroscopie non destructive permettant l'analyse de transitions énergétiques des spins nucléaires s'opérant dans le domaine des ondes radio. Ses applications sont diverses en chimie, biologie, physique, métabolomique et médecine.

Les transitions sont générées simultanément sur l'ensemble des noyaux dont les fréquences de résonnances sont comprises dans la gamme de fréquence générée par le champ radio fréquence. La fréquence de chaque noyau observé dépend de sa nature chimique (H, N, C…) et de son environnement chimique.

C'est grâce à la différentiation de ces fréquences du aux influences électromagnétiques de leur entourage propre (liaison hydrogène, proximité d'un cycle aromatique agissant comme une micro bobine, effets électro attracteurs) qu'il est possible de reconstruire des structures de macromolécules.

Dans notre cas nous avons suivi une démarche classique faisant intervenir deux expériences bidimensionnelles: TOCSY et NOESY.

La première étape consiste en l'acquisition de cartes RMN par des expériences TOCSY (TOtal Corelation SpectroscopY). Celles-ci vont permettre l'identification de systèmes de spins identifiables grâce au transfert d'aimantation relayé à travers les liaisons jusqu'à « buter » contre un atome ne portant pas d'hydrogène (carbone quaternaire, carbonyle..). Finalement, l'acide aminé portant le système de spin peut être identifié compte tenu du nombre de signaux et des déplacements chimiques caractéristiques à chaque proton.

L'enchaînement des résidus est attribué par l'acquisition d'expérience NOESY (Nuclear Overhauser Enhancement SpectroscopY). Les NOESY permettent d'observer les corrélations à travers l'espace entre noyaux distants de 5 Å maximum. Les volumes de ces signaux sont proportionnels à la distance entre les atomes corrélés. La calibration des volumes à partir de signaux correspondant à la corrélation entre protons de distance fixe (protons aromatiques par exemple) est un pré requis pour une bonne mesure des distances. Ainsi la collecte de signaux NOEs est autant de contraintes de distances qui serviront au calcul des structures par recuit simulé (Annexe D1).

Les structures finales doivent satisfaire l'ensemble des contraintes. Il est commun d'observer des artefacts durant les expériences NOESY appelés « diffusion de spin » qui peuvent polluer le jeu de contraintes et doivent par conséquent être retiré. L'identification de ces artefacts et l'optimisation des mesures de distances nécessitent l'enregistrement de plusieurs expériences NOESY en faisant varier le temps de mélange (temps pendant lequel s'opère le transfert d'aimantation dans l'espace). Une régression linéaire est finalement opérée sur chaque signal pour une mesure de distance optimisée: le *build-up*.

L'inconvénient majeur de la RMN est la limite de taille qui peut difficilement excéder 100 acides aminés en proton homonucléaire ou 200 résidus pour des séquences d'impulsions

élaborées après marquage isotopiques divers ($^{15}N$ et $^{13}C$ en particulier) mais cette limite tend à être élevée avec des expériences de types TROSY (Sprangers and Kay 2007). L'avantage le plus intéressant de cette technique est l'étude en solution d'échantillon biologique pouvant rendre compte de leur dynamique à plusieurs échelles de temps, de mesures de constantes d'affinité ou d'échange entre plusieurs formes.

### *E. Modélisation Théorique De Protéines*

Alors que le nombre de structures expérimentales déposées dans la PDB explose d'année en année de façon exponentielle (Figure 31A), le nombre de repliements protéiques et de nouvelles superfamilles commencent à plafonner (Figure 31B et C).

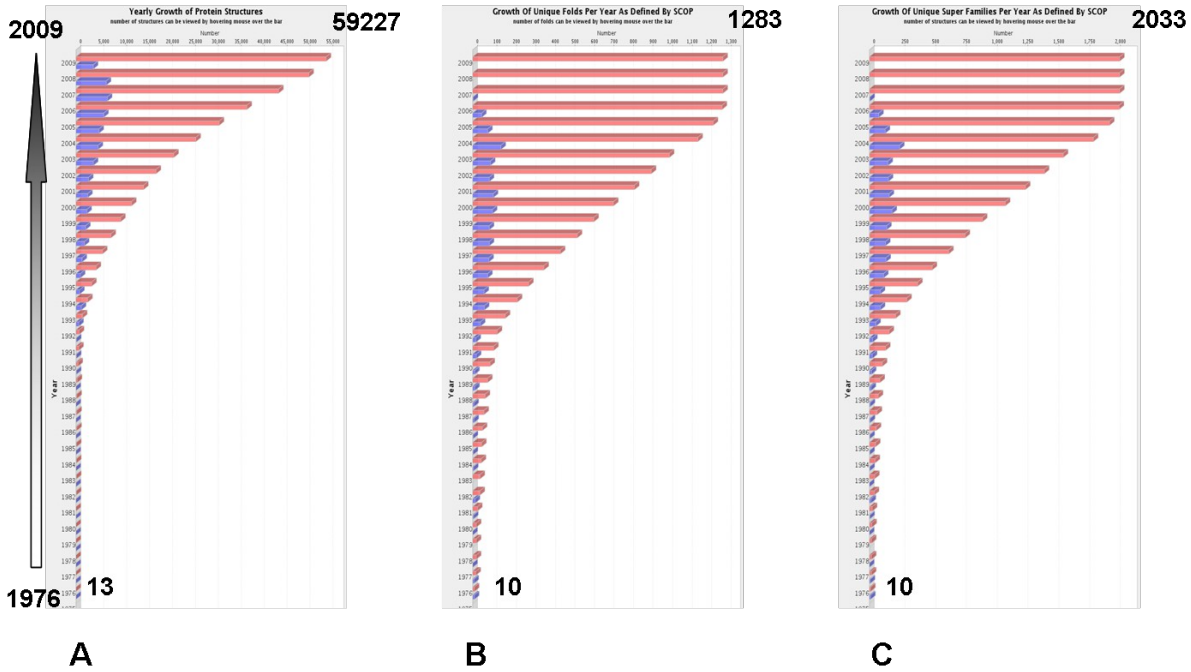



**Figure 31: Statistiques de la PDB au 29/07/09.** A. Nombre de structures résolues. B. Nombre de repliements différents identifiés (selon SCOPE). C. Nombre de superfamilles différentes découvertes (selon SCOPE). Dans les trois histogrammes, les bars bleues représentent les statistiques annuelles et les rouge les totaux cumulés. Les valeurs pour les années 1976 et 2009 sont précisées.

Ceci illustre le fait que les protéines adoptent des chemins de repliements préférentiels traduisant des stabilités énergétiques plus favorables d'un nombre limité de conformations quelque soit la fonction. En effet, on peut noter en comparant le nombre de repliements et de superfamilles que le second est deux fois plus abondant que le premier. Un repliement n'est alors plus spécifique d'une seule fonction. Cette dernière est spécifiée par la répartition des

différents types d'acides aminés (acides, basiques, hydrophobes…) à l'intérieur du repliement de la protéine.

Lorsqu'il est difficile d'accéder à la structure expérimentale d'une protéine à cause de la difficulté à la produire (cytotoxicité lors de l'expression dans un système bactérien, séquence trop grande pour la synthèse chimique), il est toutefois possible de tirer profit des structures contenues dans la PDB pour tenter de reconstruire un modèle théorique soit par le biais de serveurs disponible sur Internet (voir le site de ExPasy pour quelques exemples, http://www.expasy.org/) ou grâce à des logiciels tels que MODELLER (Eswar, Eramian et al. 2008).

Ainsi, trois branches de la modélisation théorique des protéines ont émergées, chacune inhérente au degré d'homologie existant entre une protéine cible et l'une ou plusieurs structures expérimentales (Zhang 2009).

La *modélisation comparative* ou par *homologie* est issue de la reconstruction d'un modèle 3D après alignement de séquence entre une séquence cible et un proche homologue.

La seconde branche regroupe les modèles issus de la reconnaissance de repliement par superposition d'enchaînements d'éléments de structures secondaires prédits pour la séquence cible et ceux observés expérimentalement pour une protéine n'ayant pas de lien phylogénique évident. Cette démarche est appelée *threading* (enfilage) ou *fold recognition* (FR, reconnaissance de repliement).

Enfin, la dernière branche inclus les modèles obtenus par le biais de méthode *ab initio* ne reposant sur aucune structure expérimentale comme structure-support. Cette dernière méthode est limitée à la prédiction de petites protéines (<120 résidus).

De ces trois méthodes résultent des degrés de confiance croissant avec l'homologie avec la structure-support utilisée et de fait, les informations exploitables de ces modèles théoriques va de la simple détermination du repliement général d'une protéine (*ab initio*) à la possibilité d'effectuer du *docking* pour l'élaboration de nouveaux inhibiteurs quand des modèles atteigne une résolution comparable à des structures expérimentales (*modélisation comparative*) (Figure 32).

La démarche générale consiste en un alignement de séquences duquel va découler un premier positionnement dans l'espace des carbones alpha de chaque position en prenant comme contrainte les fragments alignés sur le principe du calque.

Les hélices et brins bêta vont être générés et les boucles, souvent l'objet de mutations, d'insertions ou de délétions, positionnées en conséquence.

Le problème reste à déterminer la conformation relative des acides aminés constituant ces boucles. Ce problème peut être levé après consultation des banques de données répertoriant les boucles expérimentalement observées en fonction de leurs compositions en acides aminés et de leur longueur.

Le second problème, réside dans le positionnement des chaînes latérales qui ont été positionnées arbitrairement par rapport à leur seul carbone alpha respectif, et non en prenant en considération le tissu de forces électrostatiques primordiale au bon repliement d'une protéine.

Pour lever les différents encombrements stériques générés et établir les appariements dipolaires favorables au repliement tertiaire, une minimisation est nécessaire.(Forster 2002)

L'évaluation de la justesse d'un modèle est faite par le calcul de RMSD avec la structure support utilisée pour son obtention. Cependant, celui-ci peut être pollué par les boucles bien que le repliement général du cœur est correct.

D'où l'utilisation de nouveaux outils de mesure tel que le TM-score (Zhang and Skolnick 2004) qui pondère les erreurs de conformation selon la distance entre les différents atomes. Une erreur à courte distance aura un poids négligeable par rapport à des erreurs à longue distance favorisant ainsi plutôt la détermination du repliement global par rapport au repliement locale. Le TM-score est compris dans une fourchette de 0 à 1 avec un bon indice de confiance pour TM-score > 0,5.

D'autres outils tels que ProSA (Wiederstein and Sippl 2007) calculent les énergies locales et globales de chaque carbone alpha en se référant à des potentiels standards considérant les appariements et l'exposition au solvant. L'énergie globale est donnée sous la forme d'un Z-score replacé dans un graphique réunissant les résultas obtenus pour des structures expérimentales en fonction de leur taille en amino acides.

D'autres outils tels que la fonction QmeanClust (Benkert, Kunzli et al. 2009) permettent d'analyser localement et globalement la qualité d'un ensemble de structure simultanément ce qui peut s'avérer intéressant pour sélectionner un modèle parmi d'autres issus de structure-supports différentes, ou de choisir le meilleur parmi un nombre important de structures basées sur la même structure-support mais générés après exploration des espaces conformationnels possibles (problème des boucles).

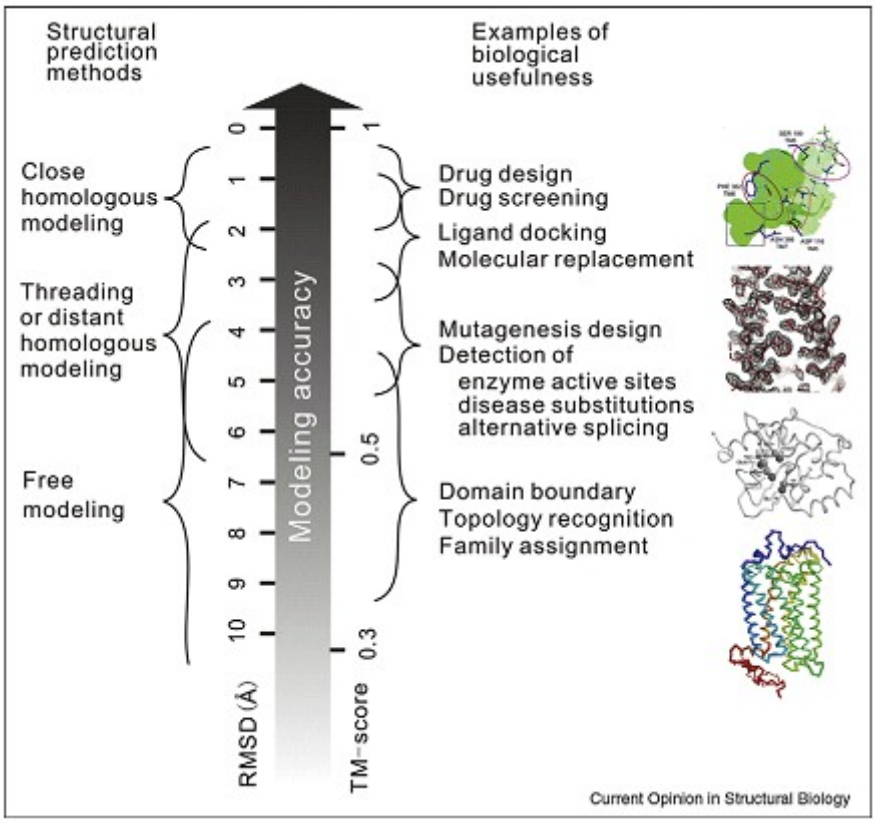



**Figure 32: Types d'informations exploitables selon la méthode de modélisation utilisée.** D'après Zhang, 2009.

## VII. ETUDE DE LA PROTÉINE PUL89

Pour générer l'alignement entre pUL89 et ses homologues, nous avons utilisé les séquences listées dans le Tableau 8. Les codes qui y sont cités sont ceux qui leur ont été attribués dans la banque de données du NCBI (http://www.ncbi.nlm.nih.gov/sites/entrez?db=protein).

| genre | virus | protéine | taille (AA) | accession number |
|---|---|---|---|---|
| alpha | HSV-1 | pUL15 | 735 | NP_044616 |
| | HSV-2 | pUL15 | 734 | NP_044484 |
| | CeHV-1 | pUL15 | 735 | AAP41433 |
| | CeHV-2 | pUL15 | 738 | YP_164457 |
| | EHV-1 | ORF44 | 734 | YP_053090 |
| | BHV-5 | pUL15 | 737 | NP_954931 |
| | HHV-3 | ORF42 | 747 | NP_040165 |
| | MDV2 | pUL15 | 737 | BAA82909 |
| beta | HCMV | pUL89 | 674 | CAA35363 |
| | CCMV | pUL89 | 672 | AAM00728 |
| | RHCMV | pUL89 | 671 | AAT07420 |
| | TuHV-1 | T89 | 673 | AAK57138 |
| | HHV6-B | U66 | 666 | AAD49667 |
| | HHV7 | U60 | 663 | YP_073802 |
| gamma | PL_HV3 | ORF29 | 684 | AAO12336 |
| | PL_HV1 | ORF29 | 683 | AAM22132 |
| | HHV8 | ORF29 | 687 | AAB62677 |
| | HHV4 | BDRF1 RF | 690 | CAA24834 |

**Tableau 8: Séquences de pUL89 et de ses homologues utilisées pour l'alignement de séquences.**

L'alignement de séquences présenté en Annexe B1 souligne une grande conservation à travers la totalité de la séquence.

L'analyse de la séquence de pUL89 après alignement avec ses homologues herpesviraux, conduit à la distinction de plusieurs domaines. Un domaine N-terminal entre les acides aminés 1 à 150, un domaine contenant tous les attributs de l'activité hélicase nécessaire à la translocation de l'ADN s'étendant des résidus 150 à 340, un doigt de zinc potentiel dans la partie centrale de la séquence (367-385), un domaine endonucléasique contenant le site d'interaction avec pUL56 et un domaine C-terminal (627-674) (Figure 33).

Cette alignement de séquence a été publié (Champier, Hantz et al. 2007).

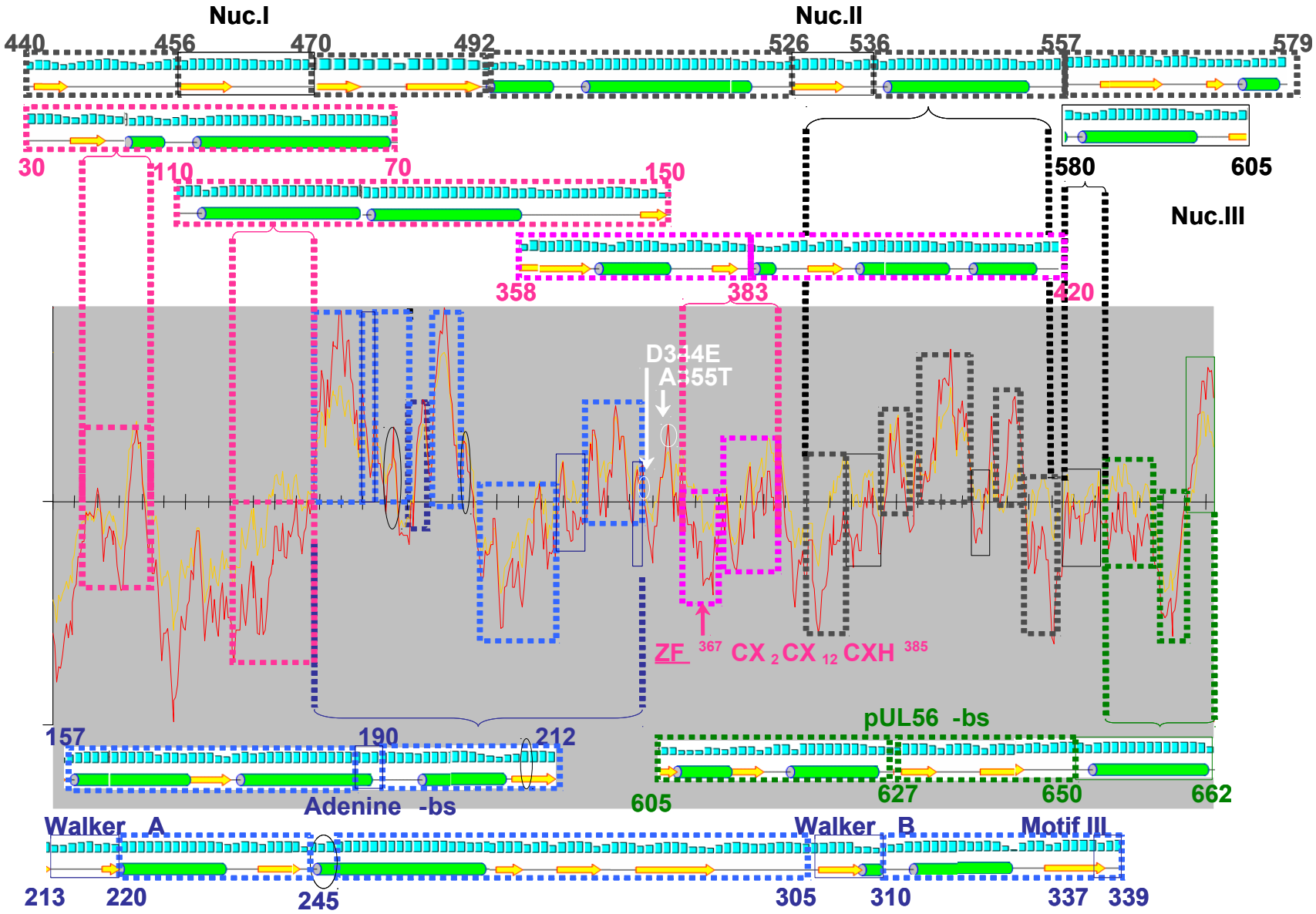

Nuc.I
Nuc.II
Nuc.III
440
456
470
492
526
536
557
579
30
110
70
150
580
605
358
383
420
D354E
A355T
ZF 367 CX 2 CX 12 CXH 385
157
190
212
Adenine -bs
pUL56 -bs
605
627
650
662
Walker A
Walker B
Motif III
213
220
245
305
310
337
339

**Figure 33: Récapitulatif des régions conservées à travers les herpesvirus sur le graphique d'hydrophobicité de pUL89.** Deux échelles d'hydrophobicité ont été utilisées, Kyte & Doolittle et Eisenberg colorés en rouge et jaune, respectivement. Les zones dont les fonctions sont présumées d'importance catalytique sont encadrées en traits pleins alors que celles dont le rôle seraient purement structurales ou inconnues en pointillés. Quatre domaines sont identifiables. En bleu, le domaine hélicase contenant le site de liaison à l'adénine (Adenine-bs), les boîtes de Walker A et B, le motif III et une région riche en brins bêta En noir et gris le domaine nucléasique dont les résidus acides catalytiques sont dénotés par Nuc. I, II et III. Le domaine d'interaction à pUL56, en vert. D'autres domaines, N-ter et doigt de zinc sont colorés en rose. Les microdomaines basiques sont mis en évidence par des ellipses noires, les mutations de résistances aux dérivés D-benzimidazolés par des ellipses blanches. Nous avons précisé les structures secondaires selon PSI-PRED (flèches jaunes et cylindres verts pour les brins bêta et hélices alpha, respectivement). Les barres en cyan représentent le taux de confiance se référant à la structure secondaire pour chaque position. Les bornes de chaque segment en fonction de leur hydrophobicité ont été indiquées.

Nous allons dans la partie qui suit présenter les résultats obtenus pour quelques uns de ces domaines. Une analyse de la conservation du domaine N-terminal issue d'alignement de séquences de l'Annexe B1 est abordée brièvement.

Puis, une seconde partie traite du domaine hélicase mettant en relief le rôle de plusieurs amino acides dans la fixation de l'ATP et la reconnaissance de l'ADN par modélisation par homologie avec gp17 du bactériophage T4. Un début d'explication sur le mécanisme de résistance aux dérivés D-benzimidazolé est proposé ainsi que le rôle probable du doigt de zinc adjacent au domaine hélicase pour la reconnaissance de l'ADN.

Enfin, une troisième partie présente les résultats obtenus par RMN sur les peptides 580-600 et 568-635. Cette étude a été complétée par un travail de modélisation du domaine endonucléasique par reconnaissance de repliement basé sur des homologies avec les RNase de type HI. A partir de ce modèle, nous présentons les résidus importants à la reconnaissance et au clivage de l'ADN, et l'étroite relation entre l'activité endonucléasique et l'interaction abvec pUL56.

### *A. Domaine N-terminal De PUL89*

Le domaine N-terminal de pUL89 (pUL89(1-150)) est la région de la séquence la plus variable. Cependant, en Annexe B1, nous y avons surligné en noir les résidus basiques (arginine, histidine et lysine) dont les positions sont parfaitement alignées à l'intérieure de même sous-famille d'herpes.

Il en découle ainsi trois motifs correspondant aux homologues alpha, bêta et gamma respectivement. Plus explicitement, les alphas conservent un motif $\boldsymbol{\pi}$-$x_{27\text{-}33}$-**RH**-$x_{10}$-**H**-$x_{14}$-**R**, les bêtas se distinguent par $\boldsymbol{\pi}$-$x_{19\text{-}20}$-**KR**-$x_{11}$-**H**-$x_{14\text{-}15}$- $\boldsymbol{\pi}$-$x_{1\text{-}2}$-**R** et les gammas présentent un consensus un peu plus étendu de la forme $\boldsymbol{\pi}$-$x_{26\text{-}31}$-**K**-$x_2$-**R**-$x_2$-**H**- $x_{20\text{-}21}$-$\boldsymbol{\pi}$ ; où $\boldsymbol{\pi}$ représente un résidus basique parmi R, K et H.

La richesse de ce domaine en résidus basique, le fait que les terminases sont exprimées dans le cytoplasme de l'hôte durant la phase précoce et qu'elles soient ensuite localisées au sein du centre de réplication virale dans le noyau nous laisse penser que les différents motifs cités ci-dessus sont peut-être les signatures de signaux de localisation nucléaire (Lange, Mills et al. 2007; Sorokin, Kim et al. 2007). Cependant, à ce jour, aucune étude n'a montré une entrée dans le noyau de pUL89 par ce moyen. D'un autre côté, par analogie avec le complexe terminase de HSV-1 (Yang, Homa et al. 2007), il serait plus probable que pUL89 s'associe avec pUL56 dans le cytoplasme et que le complexe ainsi formé soit transporté dans le noyau *via* l'interaction entre pUL56 et hSRP1 α (Giesen, Radsak et al. 2000).

Outre ce domaine basique, deux motifs contenus dans des hélices potentielles sont très conservés à travers les herpesvirus. Le premier est au centre du domaine basique cité ci-dessus et contient en particulier la proline 48 et la glycine 51 situées en N-terminal d'une hélice hydrophobe. Le deuxième est centré sur la phénylalanine 130, elle aussi incluse dans une hélice d'apparence amphipathique dont la plupart des acides aminés conservés seraient situés sur la face hydrophobe. Aucune fonction n'a à ce jour pu être attribuer à ces deux motifs.

### *B. Domaine Hélicase De PUL89*

pUL89 est la sous-unité du complexe terminase du CMVH qui rencontre chez les bactériophages un grand nombre d'homologues. Contrairement à ce que nous avons rapporté dans le Tableau 7 -qui classe les sous-unités en fonction de leurs tailles et non de leurs fonctions-, pUL89 possède des similitudes avec les grandes sous-unités terminase des bactériophages. Celles-ci contiennent en particulier deux fonctions au centre du phénomène

de maturation et d'encapsidation de l'ADN viral: les activités nucléase et hélicase. Avant d'aborder pleinement la présentation et l'analyse des résultats obtenus au cours de cette étude, nous allons revenir sur des généralités sur la fonction hélicase qui comme sont nom ne l'indique pas, est assimilable à une activité translocase chez certaines protéines.

### B.1. Résidus critiques pour la fonction hélicase déduits de l'analyse de séquences.

L'analyse des alignements de séquences des homologues herpesviraux révèle que le domaine hélicase de pUL89 se situerait entre les acides aminés 150 et 340 avec une parfaite conservation de toutes les signatures caractéristiques (Figure 34) décrites dans la littérature pour le cas des bactériophages en particulier (Mitchell, Matsuzaki et al. 2002).

Le site contenant l'activité hélicase de pUL89 est inséré dans une large région où alternent plusieurs sous-domaines hydrophiles et hydrophobes conservés dans leur composition en acides aminés dont la fonction de chacun est encore peu évidente à prédire. Cependant, on peut noter la présence d'une région hydrophobe constituée d'hélices alpha et de 2 brins bêta prédits par PSI-PRED avec un taux de confiance maximum.

En particulier, une triade basique hautement conservée entre les positions 204 à 206 précède un brin bêta $^{209}$VFL$^{211}$ parfaitement conservé, lui-même adjacent à la boîte de Walker A $^{213}$PRRHGKTW$^{220}$. On peut insister sur le fait que, en dehors du motif GKT, cette boîte contient deux arginines consécutives dont l'une ou l'autre pourrait être impliquée dans la captation du pyrophosphate relargué après hydrolyse de la liaison phosphodiester βγ de l'ATP, aussi appelé *switch residue (Sun, Kondabagil et al. 2007)*.

Succédant à la boîte de Walker A, viennent une hélice alpha suivie d'un brin bêta eux aussi très conservés mais dont la fonction reste inconnue à ce stade de notre analyse. Pour clôturer cette région hydrophobe, un deuxième microdomaine basique $^{243}$HQKHV$^{247}$ contenu vraisemblablement en N-terminal d'une hélice alpha d'une vingtaine de résidus initie l'entrée dans une zone conservée hydrophile riche en brins bêta s'étendant sur une cinquantaine d'acides aminés. Le dernier de ces brins contient la boîte de Walker B canonique $^{305}$LLLVDE$^{310}$. Une fois de plus, un domaine hydrophobe constitué d'une hélice alpha et d'un brin bêta fait son apparition et se termine par la présence du Motif III $^{337}$SST$^{339}$.

Suivent les deux sites de résistances aux dérivés D-benzimidazolés, D344E et A355T identifiés en 1998 par l'équipe de Drach et Townsend (Krosky, Underwood et al. 1998; Underwood, Harvey et al. 1998). Bien que l'aspartate 344 soit substitué par une alanine chez

les alpha herpesvirus, par une histidine chez ses plus proches homologues issus de bêta herpesvirus, ou par une arginine ou une glutamine chez les gamma-, la position 355 est quant à elle moins variable puisque seule une sérine est observable chez des homologues alpha- et bêta herpesviraux à la place de l'alanine présente chez pUL89.

Le site de liaison à l'adénine de l'ATP indispensable à son maintien dans une poche hydrophobe est quant à lui prédit dans le segment [156]Leu-Asn[162] contenant en particulier la paire [158]FQ[159]. Nous nous sommes basés là encore sur l'analyse de Mitchell pour la déduction du rôle de ce fragment. Cependant, Mitchell émettant des incertitudes sur la constitution exacte du site de liaison à l'adénine, nous ne pouvons certifier *a fortiori* l'exactitude de notre hypothèse.

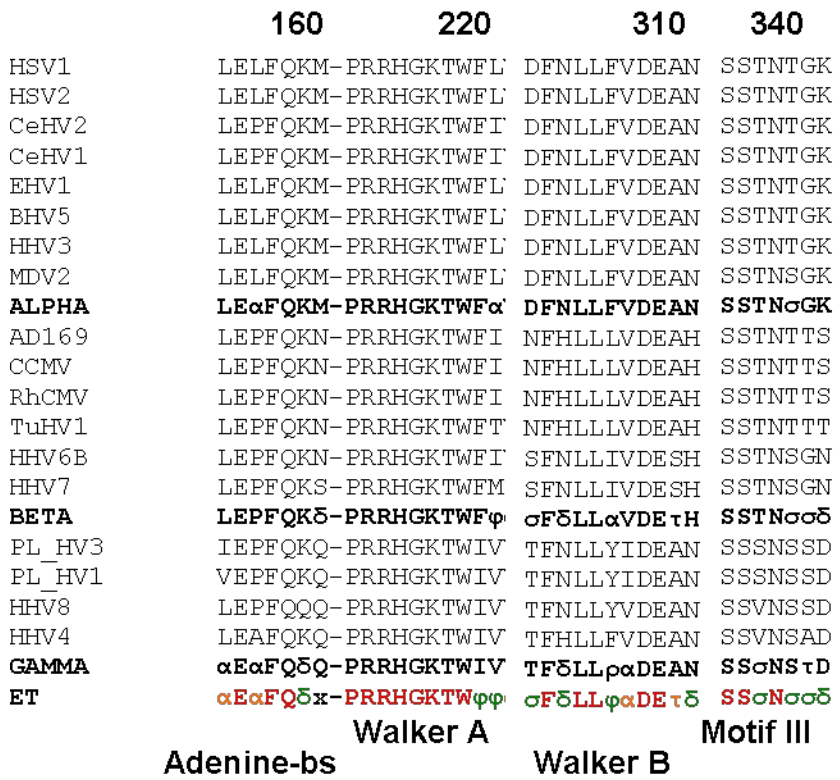



**Figure 34: Boîtes caractéristiques des domaines hélicases de pUL89 et de 17 de ses homologues.** En haut, sont précisées les numérotations de la glutamine 160 du site potentiel de liaison de l'adénine de l'ATP, du tryptophane 220 de la boîte de Walker A, du glutamate 310 de la boîte de Walker B et de l'asparagine 340 en aval de la triade [337]SST[339] du Motif III, d'après la numérotation de la séquence de pUL89.

La Figure 35 est un extrait de la revue de Rao et Feiss. L'homologue d' HSV-1 y est représenté. Il semblerait que l'analyse des auteurs diffère de la notre concernant en particulier

le site de liaison à l'adénine que nous avions prédit situé autour de la phénylalanine 204 de l'homologue de HSV-1 alors que Rao et Feiss le prédise autour du glutamate 225 aligné avec l'asparagine 180 de pUL89 (Annexe B1).

Pour pouvoir apporter une réponse à ces incertitudes et développer l'analyse de la conservation de ce domaine, nous avons généré un modèle théorique tridimensionnel du domaine hélicase de pUL89.

**Figure 35: Conservation des boîtes hélicases à travers les bactériophages et comparaison avec des représentants de la superfamille des hélicases SF2, des enzymes de restriction de type 1 et 3, et des translocases** (Rao and Feiss 2008)**.**

### B.2. Modélisation du domaine hélicase

Pour obtenir un modèle théorique du domaine hélicase nous avons soumis la séquence complète de pUL89 à PHYRE (**P**rotein **H**omology/analog**Y** **R**ecognition **E**ngine) conçu par l'équipe de Sternberg de l'Imperial College à Londres (Kelley and Sternberg 2009) et I-TASSER du laboratoire de Zhang de l'université du Kansas (Zhang 2008).

Ces deux serveurs opèrent de manières légèrement différentes. PHYRE retourne dix résultats, chacun basé sur une structure-support différente, pour les fragments dont la fiabilité de prédiction est maximum après reconnaissance de repliement. I-TASSER est quant à lui un serveur procédant à une prédiction *ab initio* qui va raccrocher plusieurs fragments chacun issus de différents *supports* puis minimiser l'ensemble. I-TASSER est disponible sur http://zhang.bioinformatics.ku.edu/. Ce site contient aussi d'autres outils de type reconnaissance de repliement comparable à PHYRE, tels que MUSTER que nous évoquerons plus tard dans la partie traitant du domaine endonucléasique de pUL89.

### B.2.a. Obtention du modèle par PHYRE

Au-delà des alignements de séquences et des prédictions de structures secondaires, nous avons tiré profit des structures de gp17 cristallisées avec ATP, ADP et en forme libre, il nous a été possible de modéliser par les serveurs PHYRE (Kelley and Sternberg 2009) et I-TASSER (Zhang 2008) une structure théorique du domaine hélicase de pUL89. Celui-ci a été exploité afin de confirmer nos hypothèses émises au paragraphe précédent sur l'attribution des fonctions catalytiques, d'émettre des propositions sur la fixation de l'ADN, de localiser les mutations de résistance décrites par ailleurs dans la littérature et d'en expliquer les mécanismes éventuels.

PHYRE a généré dix modèles à partir de dix structures PDB listées dans le Tableau 9 toutes construites sur des domaines hélicases. Les différents modèles calculés par I-TASSER ont quant à eux été obtenu à partir de dix alignements de séquences parallèles qui ont permis de générés différents fragments de la protéine. Ces fragments ainsi identifiés ont été mis bout à bout pour reconstruire 5 modèles de pUL89 entier.

Au terme des prédictions établies par PHYRE ou I-TASSER, seule une région contenant le domaine brin déjà décrit chez les bactériophages a pu être généré consensuellement par ces deux serveurs en grande partie grâce à l'identification de l'homologue gp17 du bactériophage T4 décrit en introduction (PDB: 2o0j, 2o0k et 2o0k).

En particulier, I-TASSER a malgré tout généré cinq modèles de pUL89 dans son intégralité en concaténant différents domaines obtenus par différentes structure-supports (à mi-chemin entre le *threading* et la reconstruction *ab initio*). Ce serveur quantifie la qualité des structures par le C-score qui est compris entre 2 et -5 (2 étant un taux de confiance maximum du modèle généré). Malheureusement, le meilleur modèle généré n'excède pas un C-score de -1,94 ne permettant pas de discuter avec certitude de l'intégralité de la structure générée. De

plus, les cinq modèles générés divergent énormément sur la globalité de leurs structures puisque les RMSDs sur l'ensemble du *squelette* des structures prises deux à deux sont compris entre 30 Å et 50 Å. Même localement, le RMSD entre les cinq modèles est supérieur à 10 Å au niveau des motifs hélicases présumés (boîtes de Walker A et B, et motif III).

| Modèle | PDB template | protéine | E-value | précision estimée | segment généré |
|---|---|---|---|---|---|
| 1 | 2o0j_A | gp17/ T4 | 6,6 E-13 | 100% | L138-S364 |
| 2 | 1rif_A | UvsW/ T4 | 1,4 E-4 | 95% | F169-S354 |
| 3 | 2oca_A | UvsW/ T4 | 4,9 E-4 | 95% | N128-T381 |
| 4 | 2fz4_A1 | XPB/ Archaeoglobus Fulgidus | 1,1 E-3 | 95% | F190-T339 |
| 5 | 2eyq_A3 | TRCF/ E.coli | 1,2 E-3 | 95% | V175-T339 |
| 6 | 1gm5_A3 | RecG/ Thermotoga maritima | 1,2 E-3 | 95% | A171-S338 |
| 7 | 1wp9_A1 | Hef/ Pyrococcus furiosus | 1,6 E-3 | 95% | I187-H372 |
| 8 | 2fwr_A2 | XPB/ Archaeoglobus Fulgidus | 1,7 E-3 | 95% | F190-S364 |
| 9 | 3b6e_A | MDA5/ Homo Sapiens | 2,7 E-3 | 95% | F190-T330 |
| 10 | 1w36_D1 | RecD/ E.coli | 3,2 E-3 | 95% | L89-L360 |

**Tableau 9: Synthèse des modèles générés par PHYRE.**

Les modèles calculés par PHYRE (Tableau 9) exhibent des RMSDs deux à deux compris entre 5 Å et 20 Å sur l'ensemble du *squelette* et entre 1,3 Å et 4,4 Å sur les motifs hélicases uniquement.

Nous avons donc fait le choix de travailler par la suite avec le seul domaine hélicase généré par PHYRE en se basant sur la structure de gp17 (2o0jA). Ce modèle est construit grâce à une homologie élevée entre la séquence cible et la séquence-support (47,14% de similarité dont 13,07 % d'identité sur la totalité de la séquence, Annexe B3) conduisant à un RMSD de 3 Å sur la totalité du *squelette*. On peut parler alors de reconstruction de modèle

par homologie et non plus par simple reconnaissance de repliement. Ceci implique un degré de confiance élevé dans le modèle généré.

La superposition de ce premier modèle a permis, comme le souligne les Figures 36A et C, de confirmer les prédictions faites sur les boîtes de Walker A et B et du motif III dont les noms des résidus les constituant sont rappelés en Figure 36C. Cependant, le site de liaison à l'adénine semble différent de celui que nous avions envisagé ($F^{158}Q^{159}$) au terme des alignements de séquences et basée sur les hypothèses de Mitchell (Mitchell, Matsuzaki et al. 2002). De même, la prédiction de Rao et Feiss c'est à dire la paire $N^{180}R^{181}$ chez pUL89, est certes proche du site de liaison à l'ATP mais n'est pas en interaction directe avec l'adénine. Par contre notre modèle, permet d'envisager que la paire $F^{190}Q^{191}$ est susceptible de créer des interactions hydrophobes avec l'adénine. Il est intéressant de noter que contrairement à $N^{180}$, $F^{190}$ est parfaitement conservée ce qui appuie notre hypothèse sur le rôle de liaison de l'ATP par ce résidu.

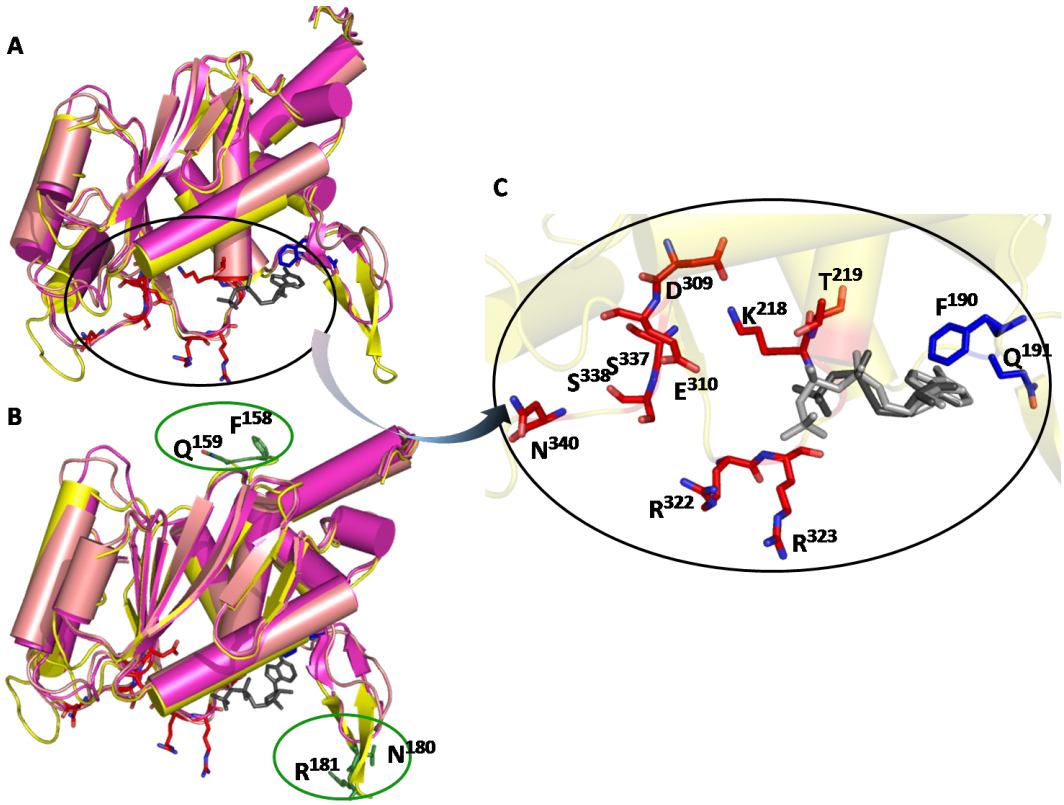



**Figure 36: Modèle 3D préliminaire de pUL89(138-364) obtenu par homologie avec la structure 2O0J de gp17 du bactériophage T4 par PHYRE.** A. Le modèle de pUL89 est coloré en jaune et superposé au complexe gp17/ADP qui a servi de structure-support (magenta et ADP en sticks noir) et la forme libre de gp17 (structure 2O0H, saumon). Les résidus appartenant aux boîtes de Walker A et B et le motif III sont colorés en rouge, la paire

$F^{190}Q^{191}$ est représentée en bleu. B. rotation de 90° de la structure représentée en A. La paire $F^{158}Q^{159}$ que nous avons prédite et la paire $N^{180}R^{181}$ prédite par Rao et Feiss comme liant l'adénine de l'ATP sont colorées en vert. C. Zoom sur l'ellipse de A. Les différents résidus identifiés par alignements de séquences sont mentionnés. La paire $F^{180}Q^{191}$ semble plus favorable à interagir avec l'adénine de l'ATP. L'ATP issu de la structure du complexe gp17/ATP (2o0K) est représentée en sticks gris et met en évidence une interaction possible entre $R^{322}$ et le phosphate γ.

### B.2.b. Amélioration du modèle

Rappelons que les structures de gp17 qui ont servi de structure-support pour la construction de notre modèle ont pu être obtenues à partir de cristaux du double mutant DE255ED de gp17 ce qui a influencé la fixation de l'ATP dans le site actif (Sun, Kondabagil et al. 2007).

De ce fait, une analyse directe des interactions par simple superposition de notre modèle avec la structure-support ne peut rendre compte entièrement et avec cohérence des différentes interactions entre pUL89 et l'ATP.

C'est pourquoi nous avons calculé un nouveau modèle en figeant le *squelette* pour ne pas perdre le repliement global du domaine modélisé et en incluant une molécule d'ATP, deux molécules d'eau et un ion magnésium $Mg^{2+}$.

Diverses contraintes ont été introduites pour stabiliser la queue phosphate de l'ATP. En particulier, la chélation du magnésium a été imposée par la définition de contraintes entre $Mg^{2+}$ et les phosphates β et γ de l'ATP, les oxygènes des deux molécules d'eau et le groupement hydroxyle de la thréonine 219 de la boîte de Walker A. Les distances entre les molécules d'eau avec les groupements carboxylates des résidus de la boîte de Walker B ($Asp^{309}$ et $Glu^{310}$) et le phosphate γ, ainsi que celle entre le groupement guanidinium de $Arg^{214}$ (*switch residue*) et le phosphate γ. Le fichier de contraintes utilisé est rappelé en Annexe C1.

Nous avons procédé à la réparation des chaînes latérales et imposé les charges en considérant un pH neutre.

La structure réparée a été soumise à une première minimisation, une dynamique moléculaire à 300 K puis une minimisation finale dans le champ de force AMBER implémenté dans le pack Insight II (Accelerys, San Fransisco, USA), suivant le protocole décrit en Annexe C2. L'analyse dans TM-align (http://zhang.bioinformatics.ku.edu/TM-align/) a donné un TM-score de 0,57, un RMSD de 0,80 Å pour une identité de 18% entre la

structure finale et le fragment Ile[98]-Glu[309] de gp17 utilisé par PHYRE. Aussi bien le TM-score (qui doit être supérieur à 0,5 pour une fiabilité notable du modèle) que le RMSD permettent de juger de la bonne qualité de ce dernier modèle.

La Figure 37 illustre comment l'ATP s'insère dans une poche pour moitié positivement chargée maintenant la queue polyphosphate de l'ATP, et pour moitié hydrophobe pour la stabilisation du groupement adénine. Le détail des interactions est explicité à la Figure 38.

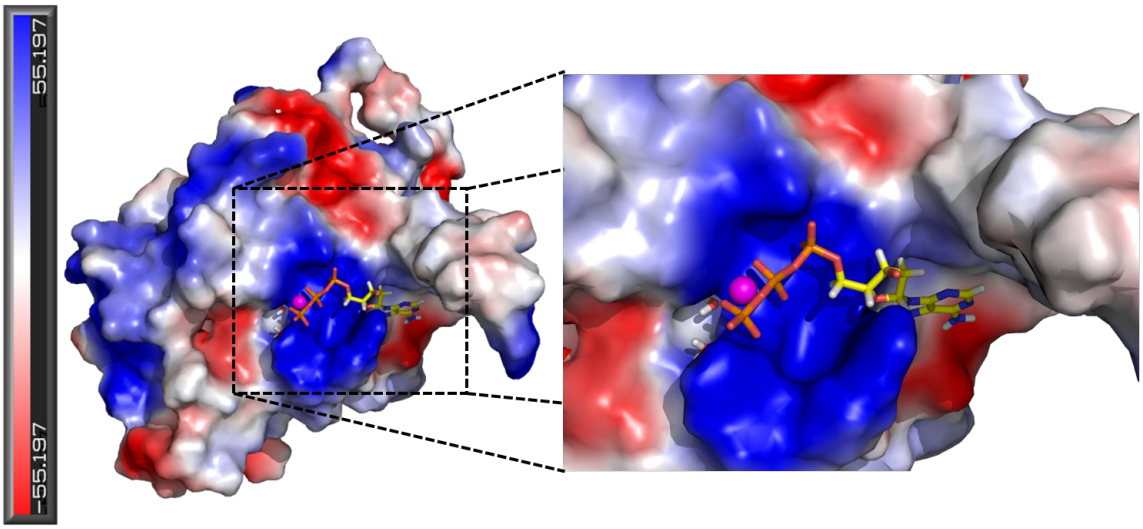



**Figure 37: Fixation de l'ATP dans le site actif du domaine hélicase de pUL89.** Le modèle obtenu après minimisation est représenté en surface colorée en fonction de la répartition de charge (échelle à gauche). La queue polyphosphate est insérée dans une poche positivement chargée et stabilisée par la présence de l'ion Mg2+ (sphère magenta) et de deux molécules d'eau (sticks). Le nucléotide est contenu dans une poche apolaire.

### B.2.c. Fixation de l'ATP dans le site actif

Les différentes contraintes imposées pour le maintien de la queue polyphosphate par les molécules d'eau, le magnésium et les différents acides aminés de pUL89 (Arg[214], Lys[218], Thr[219], Lys[253], Asp[309] et Glu[310]) conformément au mécanisme proposé par Sun et al. (Figure 17 B), ont été respectées.

On peut souligner la participation du tryptophane 220, parfaitement conservé (Figure 34), à la fois dans la formation d'une liaison hydrogène avec l'oxygène O5' via son Nε et des interactions hydrophobes avec le désoxyribose.

La stabilisation du sucre est assuré en plus de $Trp^{220}$ par un autre résidu parfaitement conservé, $Arg^{215}$ adjacent au *switch résidu* ($Arg^{214}$) qui établit via son groupement guanido une liaison hydrogène avec O4'.

La fixation du nucléotide est assurée par des interactions hydrophobes avec le phényle de $F^{190}$ (parfaitement conservé) et $Pro^{176}$. $Pro^{176}$ est conservé chez les β-herpesvirus, substitué par Gly, Arg ou Trp chez les α et remplacé par Leu chez l'homologue à pUL89 de HHV-4.

De façon intéressante, le résidu 193 est systématiquement représenté par un résidu acide chez les β-herpesvirus (proline chez les α, variable chez les γ). L'intervention de ce résidu dans la fixation de l'ATP est sûrement propre aux β-herpesvirus. Sa conservation est expliquée ici par un pont hydrogène s'établissant entre le carboxylate de $D^{193}$ et l'un des hydrogènes portés par N6 de l'adénine, le second hydrogène de N6 établit une liaison H avec le carbonyle du *squelette* de $Q^{191}$ (contrainte imposée dans notre calcul, Annexe C1).

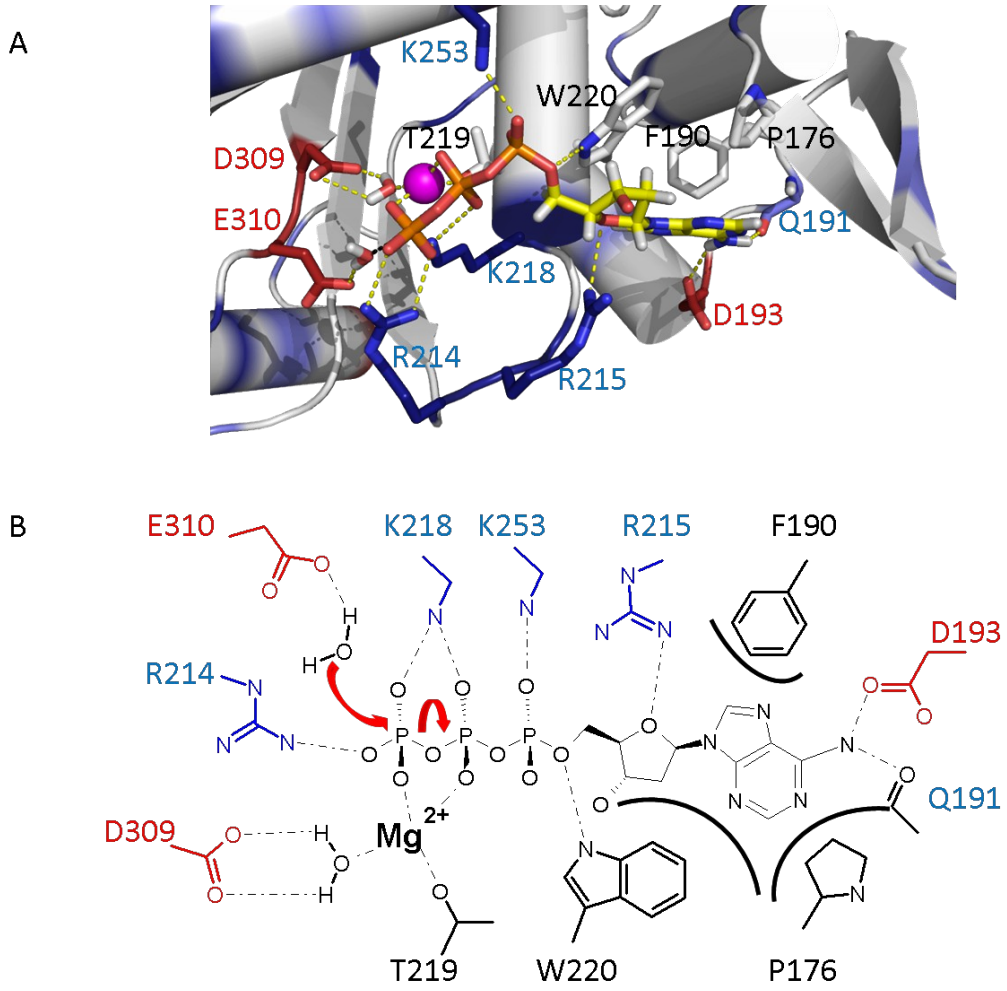



**Figure 38: Interactions établies par l'ATP dans le site actif du sous domaine hélicase de pUL89.**

B.2.d. Surface d'interaction avec l'ADN

Les hélicases de types DEXX, dont font parties les homologues herpesviraux, appartiennent à la famille SF2 (Figure 35, (Durr, Flaus et al. 2006)). Nous avons aussi rappelé que cette sous famille contient des représentants de translocase dont un seul d'entre eux a été cristallisé en complexe avec un duplexed'ADN: le domaine SWI2/SNF2 ATPase de *Sulfolobus solfataricus* (PDB 1Z36), Rad54, appelé par les auteurs SsoRad54cd (*catalytic domain of the Sulfolobus solfataricus Rad 54 homolog*) (Durr, Flaus et al. 2006).

SsoRad54cd est composé de deux domaines avec un repliement RecA-*like* (1A et 2A) et deux autres sous-domaines, 1B et 2B (Figure 39).

Les sept motifs canoniques de la famille des hélicases SF2 impliqués dans l'activité ATPasique ou de liaison à l'ADN sont répartis sur 1A et 2A.

1A comporte les motifs I (Walker A), Ia, II (DEXX) et III alors que 2A contient les motifs IV, V et VI. Deux autres sous domaines (1B et 2B) semblent être propre à SsoRad54cd.

1B est constitué d'un brin β complétant le feuillet préexistant de 1A et d'un motif hélice-boucle-hélice.

2B est formé d'un faisceau de six hélices constituant le C-ter de SsoRad54cd. La liaison à l'ADN se produit par la reconnaissance du squelette phosphodiester le long des petits sillons par le sous-domaine 1A et dans une moindre mesure 2A.

SsoRad54cd semble par ailleurs spécifique aux ADN double-brins en conformation B (Kd= 0,10 ± 0,02 µM pour les dsDNa contre Kd= 11 ± 5 µM pour les ssDNA). Le brin 3'-5' (déterminé selon le sens présumé de translocation, hélicase de type A) interagit avec les deux boucles Ia et Ib situées entre β2 et αC, et β3 et αD, respectivement. Le brin 5'-3' est reconnu par la boucle située entre le motif DEXX et αF (motif IIa), et par le motif Ic *via* une arginine très conservée située 20 résidus en amont du motif II. Le domaine 2 aussi fait participer deux résidus basiques interagissant avec des phosphates du petit sillon de l'ADN mais avec une intensité plus faible.

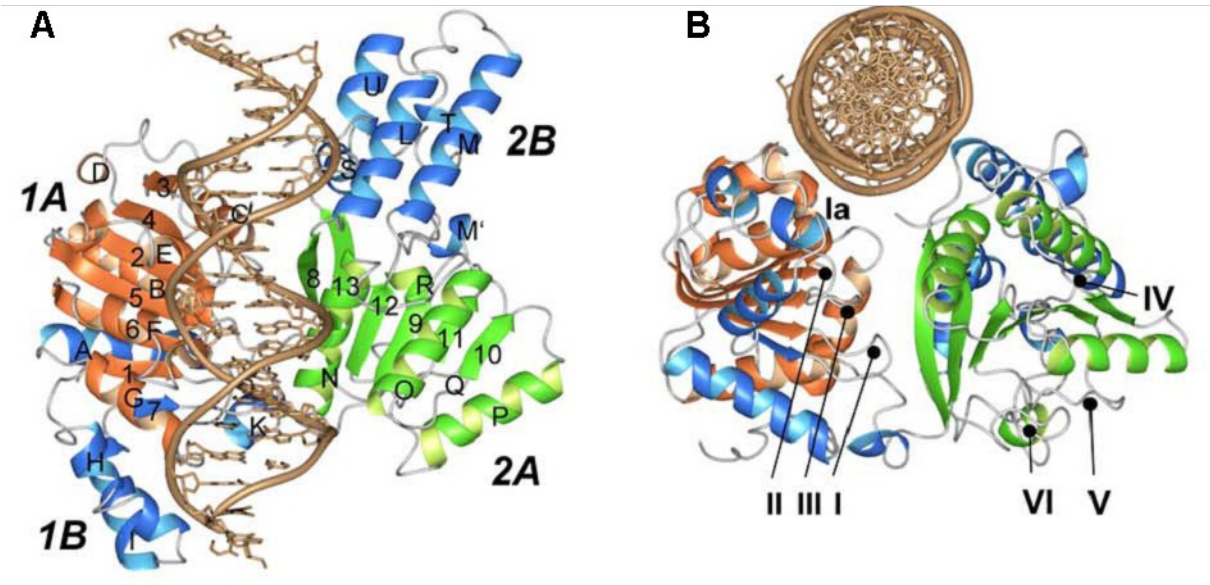



**Figure 39: Structure de SsoRad54cd et interaction avec un duplexed'ADN (PDB 1Z63).** A. Les sous-domaines 1A, 2A et, 1B et 2B sont colorés en orange, vert, et bleu, respectivement. Les différents éléments de structure secondaires sont désignés par des lettres et des nombres pour les hélices alpha et les brins bêta, respectivement, en accord avec leur

ordre d'apparition dans la structure primaire. Le duplexe d'ADN est représenté en ruban marron et s'insère à l'interface entre les RecA-*like* sous domaines 1A et 2A. B. Annotation de la topologie relative des sept motifs caractéristiques de la famille des hélicases SF2. Ia est directement impliqué dans l'interaction avec l'ADN. D'après Dürr et al, 2005.

Les terminases de bactériophages ou herpesvirales sont impliquées dans des phénomènes comparables de translocation de l'ADN double-brin de leur génome respectif en vu de l'encapsidation de ce dernier. pUL89 est de ce fait un homologue fonctionnel, et par extension structural de SsoRad54cd . Il est donc raisonnable de déduire de la structure du complexe ADN/ SsoRad54cd, l'attribution des motifs Ia, Ib, Ic et IIa chez pUL89 à partir du modèle théorique obtenu par homologie avec gp17.

Nous avons donc procédé à la superposition de notre modèle avec la structure de SsoRad54cd, en prenant comme fragments d'alignement 3D les boîtes de Walker A et B de chacune des deux protéines. Le RMSD obtenu est de 0,56 Ǻ. La Figure 40 représente la comparaison de ces deux structures après superposition. Nous avons pu ainsi attribuer les différentes boucles par homologie tridimensionnelle et identifier les motifs suivants.

La boucle Ia est un cluster de quatre acides aminés très conservé de la forme $^{242}$AHQKH$^{246}$. Le motif Ib est plus variable et correspond au fragment Glu$^{270}$-Ser$^{277}$.

Le motif Ic est parfaitement conservé à travers tous les homologues herpesviraux de pUL89 et correspond au segment $^{291}$SCYNT$^{295}$. Enfin, le motif IIa, aussi appelé *Switch* est assuré par le fragment $^{311}$AHFIKK$^{316}$. On peut remarquer que la boucle Ib de notre modèle n'est pas, dans l'état actuel de l'avancée des calculs, en interaction directe avec le duplexe (Figure 40B).

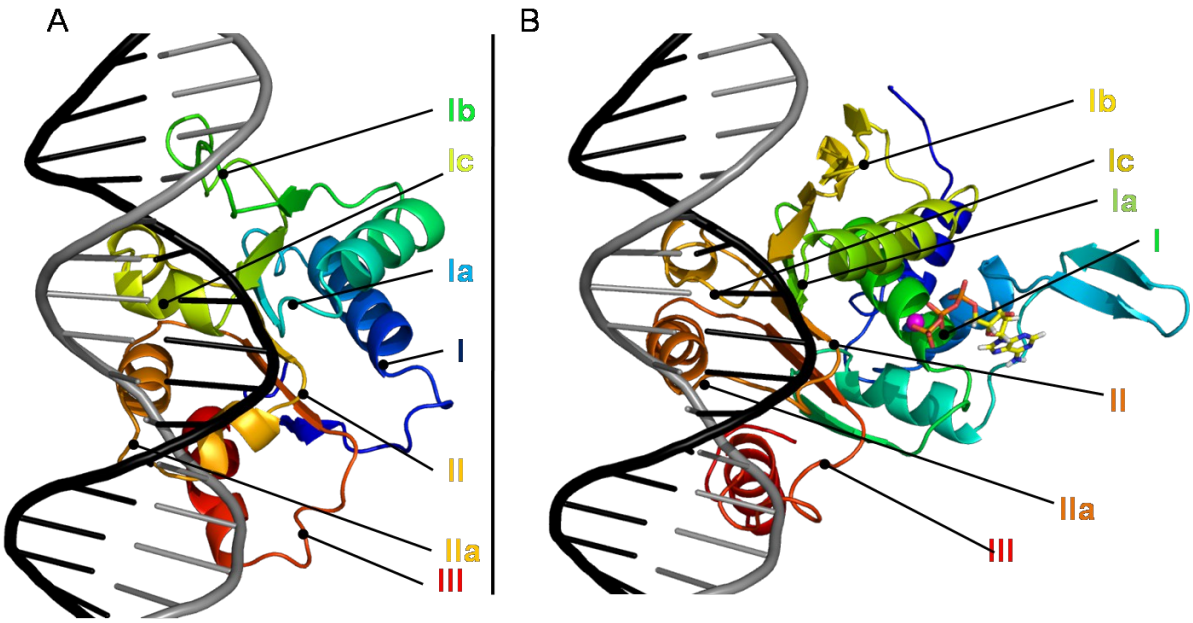



**Figure 40: Annotation des domaines d'interaction de pUL89 avec l'ADN par comparaison avec le complexe SsoRad54cd/ADN.** Chaque chaîne polypeptidique est colorée du N-ter au C-ter, du bleu au rouge, respectivement. Le duplexe d'ADN est coloré en noir et gris pour les brins 3'-5' et 5'-3', respectivement. A. complexe expérimental de SsoRad54cd/ADN (1Z63). B. Complexe théorique du domaine hélicase de pUL89 avec un duplexe d'ADN par superposition de sa structure à celle de SsoRad54cd au niveau des motifs I et II (Walker A et B). La déduction par superposition de l'annotation des domaines Ia, Ib, Ic et IIa correspond aux fragments $Ala^{242}$-$His^{246}$, $Glu^{270}$-$Ser^{277}$, $Ser^{291}$-$Ser^{297}$ et $Ala^{311}$-$Lys^{316}$, respectivement. Les domaines I, II et III sont aussi rappelés.

La Figure 41 présente le domaine de pUL89 complexé artificiellement avec le duplexe d'ADN selon la même vue qu'à la Figure 40. Le taux de conservation y est codifié en dégradé de vert (duvert foncé au blanc pour les résidus conservés à non conservés, respectivement) dans la Figure 41A. La Figure 41B cartographie la répartition des caractères physicochimiques sur cette même structure (bleu, bleu clair, rouge et blanc pour les résidus basiques, asparagines et glutamines, acides et tous les autres, respectivement).

On remarque que la zone d'interaction avec l'ADN concerne un patch très conservé Figure 41A) riche en résidus basiques (Figure 41B) contenus principalement dans les motifs Ia et IIa ($His^{243}$, $Lys^{245}$, $Lys^{315}$ et $Lys^{316}$) et en asparagine contenues sur Ic ($Asn^{294}$ et $Asn^{296}$). On peut noter par ailleurs la présence d'aromatiques ($Tyr^{291}$ et $Phe^{313}$) pouvant éventuellement interagir avec les plateaux de bases de l'ADN. Le motif III et la boucle le prolongeant ($Thr^{342}$-

Thr[346]) sont quant à eux situés au pied du duplex. Ceci est en accord avec son rôle potentiel dans le couplage de l'hydrolyse de l'ATP avec la translocation de l'ADN. Par le biais de divers ponts hydrogènes, le motif III relie le site ATPasique au duplexe ce qui va permettre le transfert de l'énergie chimique produite pendant l'hydrolyse de l'ATP en énergie mécanique en particulier *via* l'asparagine 340 (Graves-Woodward, Gottlieb et al. 1997; Caruthers and McKay 2002).

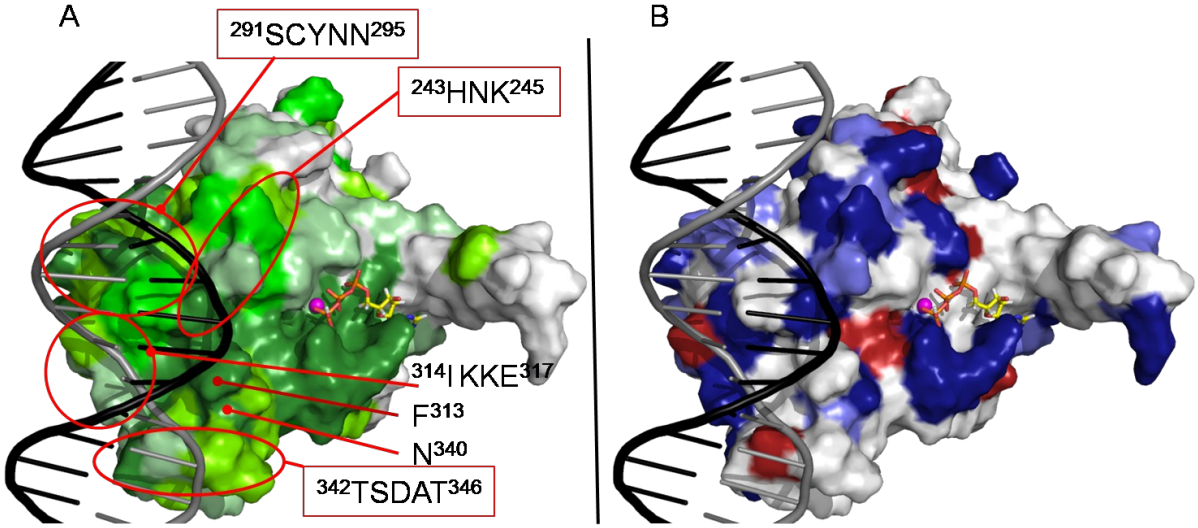



**Figure 41: Répartition du taux de conservation et des caractère physicochimique à l'interface avec le duplexe d'ADN.** A. Le taux de conservation est codifié par nuance de vert. Les résidus dont les atomes lourds sont situés à moins de 5Å du duplexe y sont précisés. Ils sont principalement inclus dans les boucles Ia, Ic, IIa et III. B. Même vue qu'en A. Les positivement (H, K et R), négativement chargés (D et E), glutamines et asparagines, et tous les autres résidus sont colorés en bleu, rouge, bleu clair et blanc respectivement ce qui montre non seulement une bonne conservation des résidus à l'interface avec l'ADN mais aussi un grand nombre d'asparagines et résidus basiques.

### B.2.e. Mutants aux dérivés D-ribobenzimidazolés

Les deux sites de mutations D344E et A355T observés sur pUL89 après traitement aux dérivés D-ribobenzimidazolés (Krosky, Underwood et al. 1998; Underwood, Harvey et al. 1998) peuvent eux aussi être localisés sur notre modèle (Figure 42).

L'aspartate 344 est peu conservé chez l'ensemble des homologues de pUL89. Il est systématiquement remplacé par une alanine chez les alpha herpesvirus, peut être remplacé par une histidine chez des homologues de bêta herpesvirus et par une arginine ou une glutamine chez les gamma. Ainsi, une corrélation de caractère physicochimique ou d'encombrement

stérique est difficile à élaborer pour attribuer une fonction particulière à ce résidu. Pourtant, notre modèle met en exergue sa proximité avec le duplexe d'ADN. Comme nous l'avons vu en Figure 41, cet acide aminé est aligné avec les motifs Ic et IIa qui sont probablement impliqués dans l'interaction avec le brin 5'-3' (Figure 41A, brin ADN gris).

Par contre, l'alanine 355 est, elle, très conservée à travers les herpesvirus et n'est substituée par une sérine (autre résidu *minuscule*, cf. diagramme de Venn, Figure 25) que chez quelques homologues de alpha et bêta herpesvirus. Cependant, son positionnement, dans notre modèle, par rapport au site ATPasique ou au site de liaison à l'ADN ne permet pas de conclure au sujet de sa participation dans l'un ou l'autre de ces processus moléculaire.

Il est intéressant de noter que ces mutations de résistances observées chez pUL89 semblent influencer des contraintes géométriques et non physicochimiques. En effet, l'aspartate et le glutamate comportent tous les deux une fonction carboxylate et se différencient par un méthylène, et la thréonine comporte un méthyle comme l'alanine et une fonction hydroxyle comme la sérine mais fait apparaître une liaison C-C supplémentaire.

Par ailleurs Underwood et al. (Underwood, Harvey et al. 1998) ont montré que le mutant D344E exhibe un degré de résistance moindre aux inhibiteurs D-benzimidazolés par rapport au double mutant A355T/D344E.

L'analyse de l'alignement de séquences présenté en Annexe B1 montre la présence d'un doigt de zinc potentiel parfaitement conservé de la forme $CX_2HX_{12}CXC$ entre les cystéines 367 et 385. Ces motifs sont fréquemment rencontrés dans les mécanismes de reconnaissance d'oligonucléotides (Laity, Lee et al. 2001; Krishna, Majumdar et al. 2003). Le fait que ce motif soit adjacent, dans la séquence de pUL89, au domaines III, lui aussi impliqué dans la reconnaissance d'ADN, permet de formuler l'hypothèse que les dérivés D-benzimidazolés visent à déstabiliser cette reconnaissance (éventuellement en s'intercalant entre les plateaux de bases de par leur homologie chimique avec les nucléotides) et ainsi faire échouer les phénomènes de translocation et/ou de clivage de l'ADN.

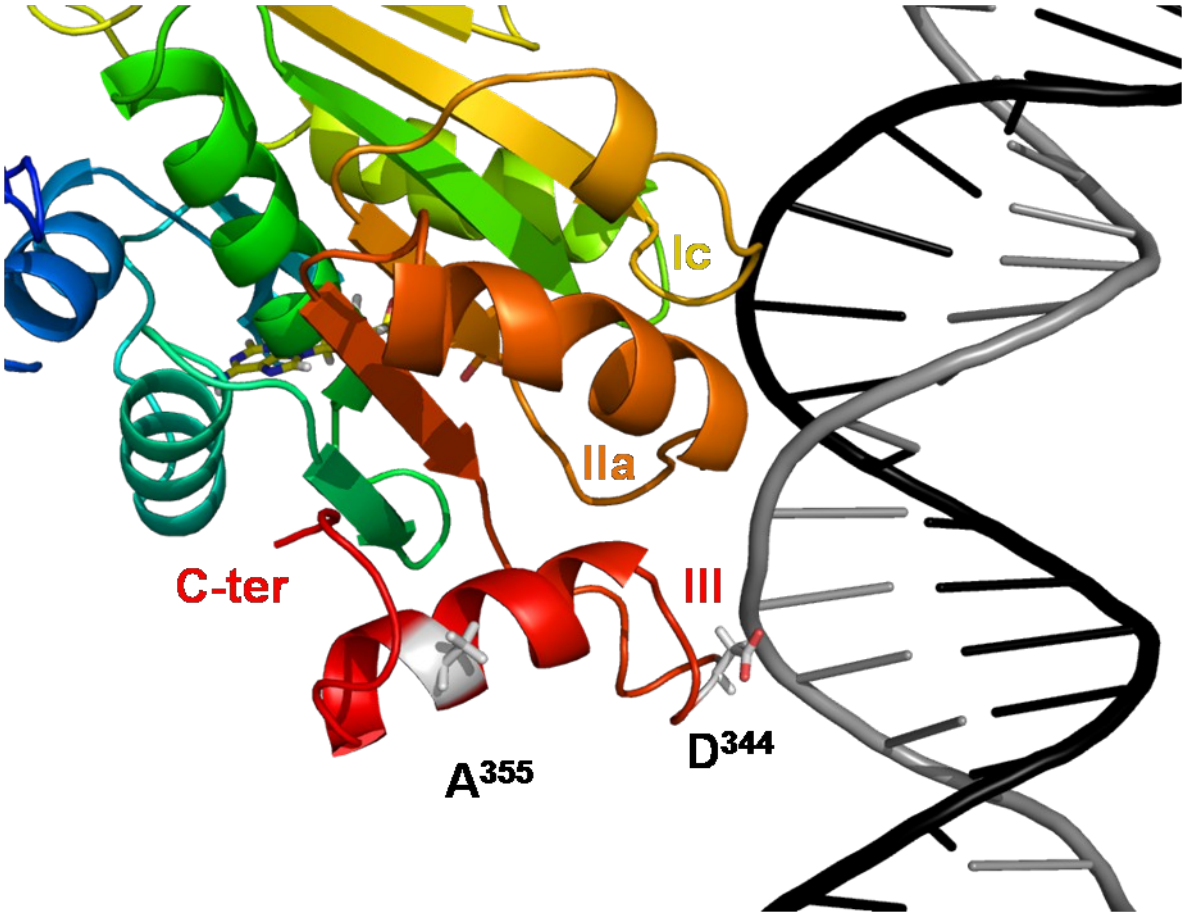



**Figure 42: Sites de mutation aux dérivés benzimidazolés, $A^{355}$ et $D^{344}$.** pUL89 est coloré du bleu au rouge du N-ter au C-ter, respectivement. Les sites de mutations sont colorés en blanc. $Asp^{344}$ est favorablement placé pour interagir avec l'ADN (cf. Figure 41). La fonction de $Ala^{355}$ est peu évidente mais peut-être qu'une mutation à cette position peut modifier l'hélice à laquelle elle appartient.

En résumé de cette partie traitant le domaine hélicase de pUL89, nous avons reconstruit le domaine $Leu^{138}$-$Ser^{364}$ par homologie avec gp17 du bactériophage T4. Ceci a été rendu possible grâce à l'utilisation de serveurs informatiques en ligne. Toutefois, divers améliorations ont été apportés au modèle généré par PHYRE (réparation des chaînes latérales, ajout de molécules d'eau, d'ion magnésium, minimisation et dynamique moléculaire). Par la suite, nous avons déduit d'une homologie avec le domaine 1A de SsoRad54cd complexé avec l'ADN les fonctions d'autres motifs conservés à travers les homologues de pUL89 jusqu'alors non révélées.

### B.3. Bilan concernant le domaine hélicase de pUL89

Ce travail a permis de :

1. confirmer l'attribution des motifs I, II (boîte de Walker A et B) et III (motif C) à partir des alignements de séquences
2. revenir sur l'attribution du motif de liaison à l'adénine de l'ATP qui semble plutôt être le motif $F^{190}Q^{191}$
3. définir le rôle propre de plusieurs acides aminés dans l'hydrolyse de l'ATP
4. identifier les motifs supplémentaires Ia, Ib, Ic et IIa impliqués dans la liaison à l'ADN.
5. proposer un mode d'action des dérivés D-ribobenzimidazolés

Le domaine complexé artificiellement à l'ADN devra être sujet à d'autres calculs intégrant le duplexe qui pourrait influencer de par les potentiels électrostatiques, le positionnement des chaînes latérales. Ceci pourrait éventuellement rendre compte du rôle du domaine Ib dans la liaison de l'ADN, chose non observée dans notre modèle préliminaire. Par ailleurs, le mode d'action des D- ribobenzimidazolés suggère l'implication d'un doigt de zinc potentiel dans ce mécanisme. Ainsi, il serait préférable d'intégrer l'étude théorique ou, de préférence, expérimentale du domaine contenant le motif $CX_2HX_{12}CXC$ entre les cystéines 367 et 385 afin d'avoir une vision plus globale de l'interaction avec l'ADN et son inhibition probable par les dérivés D-ribobenzimidazolés.

La seconde zone d'ombre concerne la méconnaissance du domaine N-terminal pUL89(1-150). Le domaine pUL89(1-150) est variable à travers l'ensemble des homologues de pUL89 mais contient dans sa séquence des motifs basiques dont le positionnement dépend du genre d'herpesvirus (alpha, bêta ou gamma). Sa topologie, selon PSI-PRED, serait constituée principalement d'hélices. Ce domaine aurait lui aussi une influence sur la liaison à l'ADN par le biais de ces patches basiques

## *C. Domaine D'interaction Entre PUL89 Et PUL56: Implication Dans L'activité Endonucléasique Du Complexe Terminase.*

L'analyse de la séquence de pUL89 s'est poursuivie, motivée en cela par la découverte du domaine d'interaction de pUL89 avec la grande sous-unité pUL56 par Thoma et al en 2006 (Thoma, Borst et al. 2006). Les résultats de cette étude sont représentés à la Figure 43.

En effet, l'interaction entre pUL89 et pUL56 est primordiale car les deux sous-unités de CMVH se partagent les différentes fonctions ATPasique, reconnaissance des séquences *pac* spécifiques, clivage de ces dernières et interaction avec la protéine portail pour l'encapsidation finale du génome de CMVH. Bien que la reconnaissance de l'ADN et son clivage spécifique au niveau des séquences *pac* aient été attribuées à pUL56 (Bogner, Radsak et al. 1998), une activité nucléasique aspécifique propre à pUL89 exacerbée par son association à pUL56 a également été mise en évidence (Scheffczik, Savva et al. 2002).

Ce dernier point révèle donc un intérêt majeur dans le ciblage de l'interaction entre les deux sous-unités du complexe terminase dans un but thérapeutique. Etant donné que pUL56 reconnaît spécifiquement les séquences *pac*, que pUL89 porte l'activité nucléase, et que le clivage de l'ADN viral concatémérique est un prérequis pour son encapsidation future, théoriquement, empêcher l'interaction pUL56/pUL89 permettrait de perturber la maturation et donc l'encapsidation de l'ADN viral pour ainsi mettre fin à l'infection à CMVH.

Ici, une première étape a été franchie en étudiant par résonance magnétique nucléaire la structure expérimentale du fragment pUL89(580-600) identifié comme étant la portion minimum de pUL89 critique au maintien du complexe pUL89/pUL56.

Par la suite, nous avons rallongé ce peptide en N et C-terminal (pUL89(568-635)) pour essayer d'extraire de nouvelles informations sur le repliement et l'influence éventuelle de celui-ci autour de pUL89(580-600).

Une ultime étape a consisté à modéliser une structure théorique du domaine pUL89(456-612) afin d'émettre des hypothèses sur l'influence potentielle de pUL56 sur la liaison de l'ADN et son clivage par pUL89.

Ces résultats ont été publié dans le journal *PROTEINS: Structure, Function and Bioinformatics*.

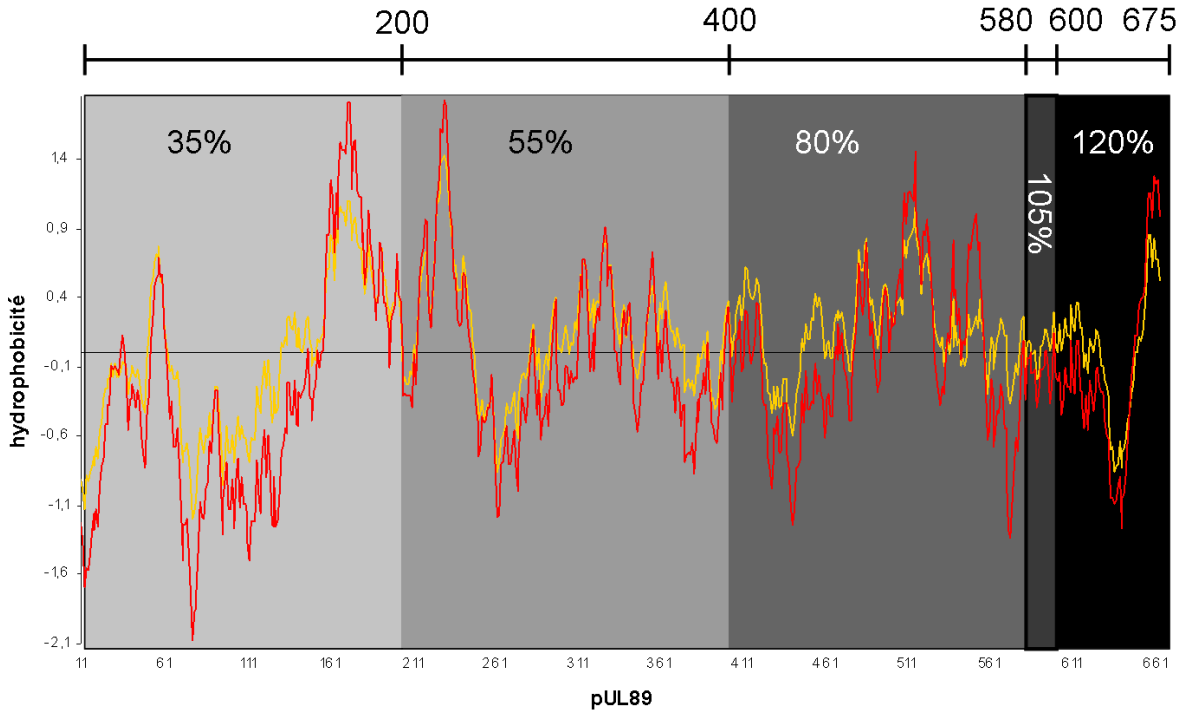



**Figure 43: Résultats des affinités relatives pour pUL56 de fragments de pUL89 fusionnés à la GST.** Adapté de Thoma et al. 2006. La séquence de pUL89 est représentée en fonction du profil d'hydrophobicité selon les échelles de Kyte & Doolittle et Eisenberg, en rouge et jaune respectivement. L'affinité exprimée en % de chaque portion est normalisée par rapport à pUL89 entier et symbolisée en dégradé de gris. L'affinité de la GST pour pUL56 est de l'ordre de 20%.

## C.1. Étude du peptide pUL89(580-600)

Conformément aux résultats de l'équipe de Elke Bogner (Thoma, Borst et al. 2006), nous avons synthétisé en stratégie Fmoc le peptide pUL89(580-600) correspondant à la séquence $^{580}$GRDKALAVEQFISRFNSGYIK$^{600}$. Après purification sur HPLC en phase inverse greffée C4, 28 mg de peptide avec une pureté de 98% ont été obtenus.

Nous avons donc analysé la structuration de ce peptide dans différentes conditions par dichroïsme circulaire puis par RMN.

### C.1.a. Dichroïsme circulaire

Les spectres de dichroïsme circulaire ont été enregistrés dans quatre conditions différentes: eau pure à pH 3,5, 30% de 2,2,2-trifluoroethanol à pH 3,3 et 6,8 et dans 120

équivalents de dodecyl phosphocholine (DPC) à pH 3,7. La concentration de peptide dans chaque échantillon a été fixée à 40 µM. Chaque spectre a été enregistré entre 185 nm et 260 nm par pas de 0,5 nm, un temps d'intégration de 2s et une bande passante de 2 nm. Quatre scans dans chaque condition ont été moyennés et la résultante lissée sur 25 points.

La première condition que nous avons testée est le milieu aqueux pour sa qualité à mimer au moins partiellement (absence de sels) les conditions physiologiques. La montée du pH jusqu'à 7 a été effectuée mais a présenté l'inconvénient de faire précipiter le peptide.

Nous avons donc poursuivi le reste des analyses de dichroïsme circulaire et par la suite de RMN à pH acide (3,5). Comme le montre la Figure 44, le peptide dans l'eau semble très peu structuré (minima intense à 195 nm) mais fait apparaître un large épaulement autour de 220 nm laissant prévoir une propension à la structuration en hélice.

La cause de tel résultat a été imputée à la taille du système qui est très inférieure à la taille de la protéine native (21 amino acides contre 674 pour pUL89 entier). Ainsi, la taille a vraisemblablement une influence sur son repliement propre. Cette hypothèse est reliée à la notion de force hydrophobe régissant le repliement des protéines par l'intermédiaire d'interactions préférentielles entre domaines ou grâce à l'intervention de protéines chaperonnes dans les systèmes biologiques.

Pour pallier à ce biais introduit par la taille, nous avons pris le parti de vérifier l'influence d'un environnement hydrophobe (organique ou lipidique) sur le repliement de pUL89(580-600). Précisons que de tels tests subsidiaires sont généralement opérés sur des protéines ou des domaines de protéines de nature transmembranaire et donc établissant *in vivo* une interaction réelle avec un environnement hydrophobe tel que des membranes lipidiques. Or le complexe terminase n'a *a priori* pas de caractère transmembranaire puisqu'il est sécrété dans le cytoplasme puis transloqué dans le noyau jusqu'au centre de réplication virale.

Le but est donc ici de vérifier le repliement de ce petit peptide sous l'influence d'un milieu hydrophobe à la manière de celui apporté par des amino acides apolaires dans la protéine native.

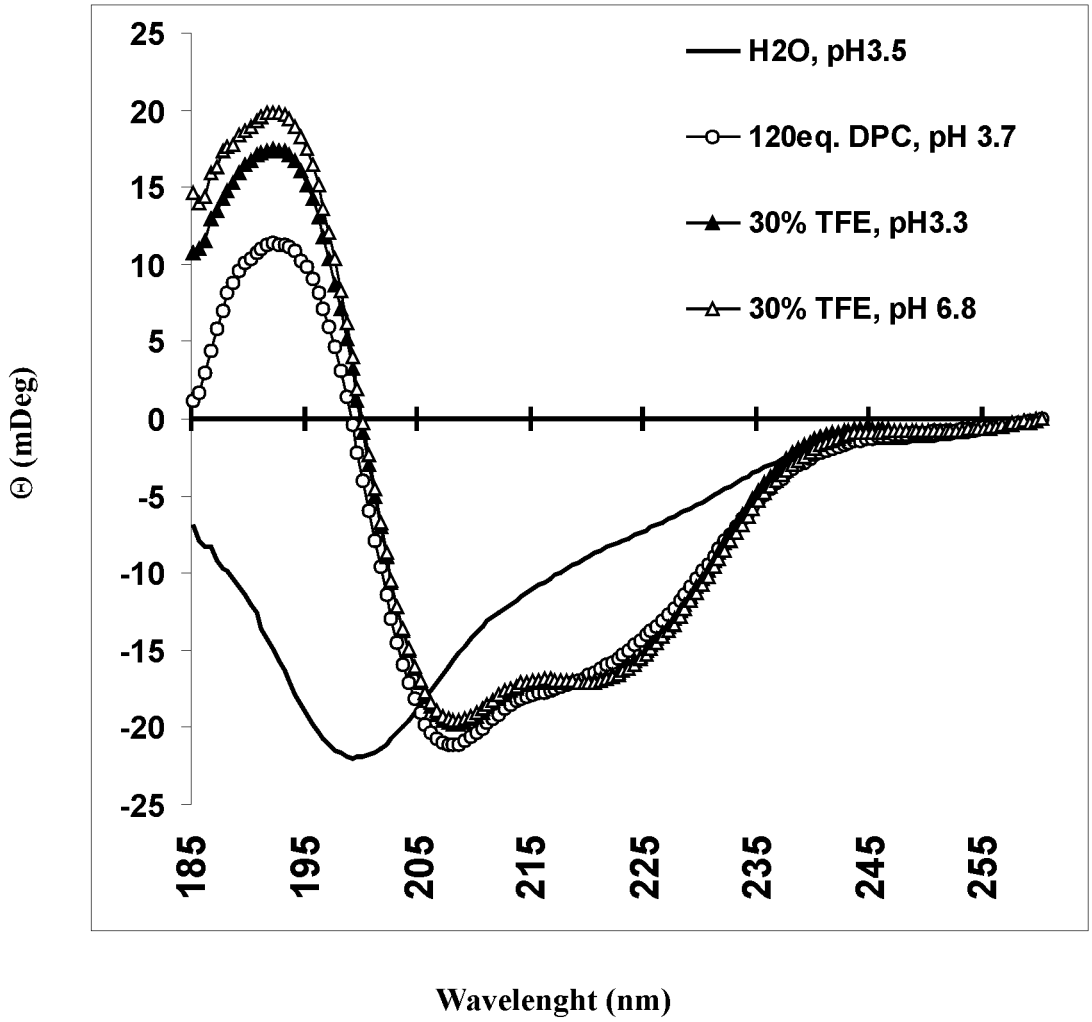



**Figure 44: Courbe de dichroïsme circulaire obtenues pour le peptide pUL89(580-600) dans diverses conditions.** Chaque courbe a été enregistrée entre 185 nm et 260 nm par pas de 0,5 nm, un temps d'intégration de 2 s et une bande passante de 2 nm sur quatre scans moyennés. Les courbes moyennes ont été lissées sur 25 points. Les courbes en ligne continue, cercles, triangles noirs et triangles blancs correspondent aux expériences dans H2O pH 3,5, 120 équivalents de DPC pH 3,7, 70% H2O/ 30% TFE pH 3,3 et 6,8, respectivement.

L'utilisation du 3,3,3-trifluoroéthanol (TFE) a donc été envisagée. Ce solvant organique est bien connu pour sa capacité à induire des hélices en chassant des molécules d'eau parasitant la formation des liaisons hydrogènes indispensables à l'intégrité des hélices.

Ce sujet fait l'objet d'une grande controverse dans le milieu de la biologie structurale où il est question de l'observation de structures qui peuvent être très influencées par le TFE, voir avoir un effet dénaturant sur les protéines (Arunkumar, Kumar et al. 1997). Il a effectivement été montré que des protéines de topologie tonneau bêta en milieu aqueux pouvaient adopter

des structures très largement composées d'hélice alpha. Cependant, il faut préciser que de telles observations ont été faites en présence de proportion excessive de TFE (60% à 90% en volume) (Jayaraman, Kumar et al. 1996). Par ailleurs, des études ont montré qu'un ratio 70% H2O/ 30% TFE (v/v) ne modifie pas de manière significative le repliement des éléments de structures secondaires. Par contre, il empêche les interactions hydrophobes entre domaines et donc modifie le repliement tertiaire des protéines (Morellet, Bouaziz et al. 2003). Nous n'avons pas dépassé cette proportion de TFE.

Le TFE a permis, de par son pouvoir solvatant, d'analyser le peptide à pH acide (3,3) et neutre (6,8). Les expériences ont été mené en incrémentant le TFE par pas de 5% un mélange H2O/TFE et en gardant la concentration en peptide constante. Les courbes présentées en Figure 44 (triangles noirs et blancs à pH acide et neutre, respectivement) ont été obtenues à l'équilibre. Les courbes de dichroïsme circulaire montrent une superposition totale des deux courbes dans la région 200 nm-260 nm. Cette région des spectres contient les deux signaux caractéristiques d'hélices alpha à 208 nm et 222 nm. Le rapport des intensités $[\theta]^{222nm}/[\theta]^{208nm}$ aux deux pH est de 0,79.

Finalement, afin d'introduire un milieu plus en adéquation avec la chimie du vivant, nous avons réalisé une incrémentation en dodecylphosphocholine (DPC) sur notre peptide initialement. Ce lyso-lipide est constitué d'une chaîne à 12 carbones et d'une tête choline. Sa concentration micellaire critique (c.m.c.) est de 1,5 mM.

Là encore, la solution a été incrémentée par pas de 20 équivalents en DPC (quantité de matière) en gardant une concentration en peptide de 40 µM constante. La courbe présentée en Figure 44 a été obtenue pour 120 équivalents de DPC c'est-à-dire une concentration supérieure à la c.m.c.: les lipides sont en phase micellaire. Les expériences ont été réalisées à pH acide exclusivement (3,7) car comme pour l'expérience dans l'eau le peptide précipite à pH neutre. Comme le TFE, elle présente les deux minima caractéristiques d'hélice $\alpha$ mais avec une ellipticité légèrement plus faible ($[\theta]^{222nm}/[\theta]^{208nm} = 0,74$) traduisant une structuration de notre peptide un peu moins prononcée que dans le cas du TFE.

Ces tests préliminaires ont permis d'entreprendre une étude structurale grâce à la reconstruction sous contraintes RMN des structures tridimensionnelles de pUL89(580-600) dans les différentes conditions présentées ci-dessus.

### C.1.b. Résonance Magnétique Nucléaire

Trois échantillons ont été préparés pour l'étude RMN de pUL89(580-600): 1,4 mM de peptide dans 90% $H_2O$/ 10%$D_2O$ à pH 3,5, 1,4 mM dans 70% $H_2O$/ 30%TFE-$d_3$ à pH 3,3 (puis par la suite pH 6,8) et enfin 0,8 mM en présence de 100 mM de DPC-$d_{38}$ (120 équivalents) dans un volume de 600 µL.

Des spectres 2D TOCSY et NOESY ont été enregistrés dans chaque cas pour établir la structure du peptide dans les différents milieux sur un spectromètre Avance Bruker accordé à la fréquence du proton à 600,13 MHz, avec 2048 points en dimension $t_2$, 512 incréments en $t_1$ et une largeur spectrale de 5787,04 Hz dans les deux dimensions.

Le signal de l'eau a été supprimé par présaturation. Les expériences TOCSY ont été réalisées avec $\tau_m$= 70ms.

Les spectres NOESY ont été enregistrés à cinq temps de mélange (50 ms, 100 ms, 150 ms, 200 ms et 250 ms) pour les expériences dans TFE à pH acide et dans le DPC afin d'effectuer un build-up pour une meilleure mesure des distances obtenue par l'intégration des valeurs des NOEs collectées après attribution.

Pour les expériences dans l'eau et le TFE à pH neutre, les NOESY ont été enregistré au seul temps de mélange de 250 ms.

Afin de compléter le jeu de contraintes pour le calcul des structures, nous avons réalisé des expériences d'échange de protons après lyophilisation des échantillons et réhydratation de ceux-ci dans de l'eau deutérée. Des spectres TOCSY et NOESY successifs (4 scans) dans le TFE et le DPC, respectivement, ont été enregistrés avec la même résolution en t1 et t2 que précédemment. L'extinction progressive des différents signaux des protons amides marque l'implication des différents résidus dans l'hélice et permet donc d'en déduire la présence de pont hydrogènes qui seront introduits sous forme de contraintes entre le proton amide *i* identifié par ces expériences et le carbonyle du *squelette* du résidu *i*-4. La température a été fixée à 280 K pour les expériences dans l'eau et le TFE et 293 K dans les DPC.

Le traitement des cartes a été opéré sur XWIN NMR. Une fonction d'apodisation cos π/2 a été appliquée aux FID et la résolution a été artificiellement améliorée (*zero filling*) pour une dimension finale des spectres de 2048 points dans les deux dimensions $\omega_1$ et $\omega_2$. Les spectres ont été calibrés par rapport au signal de l'eau en fonction de la température.

L'attribution complète des cartes RMN a été effectuée sous FELIX (Accelrys, San Diego, CA, USA) (Annexe D1) et les volumes mesurés après calibration sur les tâches de corrélation

$H^{\varepsilon}$/ $H^{\delta}$ des protons aromatiques des phénylalanines 590 et 594 et de la tyrosine 598 de notre peptide. Les volumes ont finalement été convertis en contraintes de distances avec une incertitude sur la mesure des volumes de ± 20% pour être utilisées pour le calcul des structures dans X-PLOR 3.84 (Brünger 1992) (Annexe D1)).

Les différents jeux de contraintes propres aux structures secondaires obtenues dans les quatre conditions décrites ci-dessus sont représentés dans les tables de connectivités Figure 45 (à gauche).

En accord avec les expériences de dichroïsme circulaire, pUL89(580-600) est faiblement structuré dans l'eau pure. Cependant, plusieurs contraintes de structure secondaires de faibles intensités ($\alpha\beta$ (i, i+3) et $\alpha N$ (i,i+3)) ont été collectées pour les résidus compris entre $Ala^{584}$ et $Phe^{590}$ (Figure 45A, gauche). D'autres NOEs de structures secondaires de type $\alpha\beta$ (i, i+3), $\alpha N$ (i,i+3) et $\alpha N$ (i,i+4) sont observées pour $Ile^{591}$. Au total, seules 30 NOEs de moyenne portée (|i-j|≤5) et de faibles intensités (distance ≥ 3,8Å ± 0,75) ont été collectées sur un total de 274 (Tableau 10).

Les structures correspondantes et notamment la structure moyenne représentée en ruban fait toutefois apparaître la formation d'un tour d'hélice entre $Val^{587}$ et $Phe^{590}$ (Figure 45A, droite). Il faut préciser ici que la distinction de faibles signaux pris en compte comme tels et non comme de la diffusion de spins ou du bruit ont été guidés par l'analyse des déplacements chimiques des protons alpha et amides (Figure 46). Cette analyse repose sur l'information contenue dans le déplacement chimique ($\delta$) qui résulte de l'influence des effets électromagnétiques locaux induits par des groupements chimiques à proximité d'un proton donné. Ainsi, des protons impliqués dans des liaisons hydrogènes ou situés à proximité de systèmes $\pi$ vont être déblindés alors que des protons exposés au solvant et contenus dans des régions de haut degré de liberté (partie flexible) vont avoir un déplacement chimique s'approchant d'une valeur standard. Les valeurs standard ont été déterminées à partir de petits peptides de synthèse en conformation *random* coil (glycine-rich).

Pour la détermination d'éléments de structure secondaire, l'analyse des déplacements chimiques des protons constituant le *squelette* peut être un bon indicateur du niveau de structuration du peptide.

Ainsi, Wishart a développé deux notions: le Chemical Shift Index (CSI) qui prend en compte $\delta(H\alpha)$ (Wishart, Sykes et al. 1992) et le Random Coil Index (RCI) qui est une forme généralisé du CSI et qui est une combinaison linéaire de la forme :

$$RCI=[(A\,|\Delta\delta C\alpha|+B\,|\Delta\delta CO|+C\,|\Delta\delta H\beta|+D|\Delta\delta N|+E\,|\Delta\delta NH|+F\,|\Delta\delta H\alpha|)/n]^{-1}$$

Où $|\Delta\delta C\alpha|$, $|\Delta\delta CO|$, $|\Delta\delta H\beta|$, $|\Delta\delta N|$, $|\Delta\delta NH|$ et $|\Delta\delta H\alpha|$ sont les variations entre les déplacements chimiques observés et standards pour les différents atomes constituant le *squelette*, les coefficients A, B, C, D, E et F sont des coefficients empiriques propres à chaque type d'atome et dépendant de la valeur de n qui se réfère au nombre de type de déplacements chimiques disponibles. Dans notre cas, l'attribution a été faite pour des échantillons en abondance isotopique naturelle et avec des expériences homonucléaires protons. Donc, le calcul du RCI a pris en compte dans notre cas les protons Hα, Hβ et NH d'où n=3. Il faut noter que plus n est important plus le calcul du RCI sera fiable car prenant en compte plusieurs informations le long du squelette de la chaîne polypeptidique (Berjanskii and Wishart 2006; Berjanskii and Wishart 2007).

L'avantage des méthodes reposant sur l'analyse de données brutes (déplacement chimique) est de fournir une information sur l'implication de chaque acide aminé sur le degré de structuration générale du peptide en se dispensant de l'attribution de tous les NOEs ceci simplement grâce à l'attribution des déplacements chimiques des atomes constituant le *squelette*. Par extension, la connaissance du degré de structuration implique l'estimation du paramètre d'ordre à chaque position du peptide étudié.

**Figure 45: Récapitulatif des structures obtenues pour pUL89(580-600) dans différentes conditions.** A gauche sont représentées les **t**ables de connectivités des NOEs de structures secondaires dans A. 90% H2O/ 10%D2O (v/v), pH 3,5, B. 120eq. DPC, pH 3,7, C. 70% H2O/ 30% TFE (v/v), pH 3,3 and D. 70% H2O/ 30% TFE (v/v) at pH 6,8. La séquence est précisée en haut de chaque panneau. Les résidus ayant des NH en échange lents déterminés expérimentalement sont soulignés. Les connectivités NOEs propre aux structures secondaires précisées à gauche et reliant deux résidus sont représentées par des lignes horizontales dont les épaisseurs sont proportionnelles à l'intensité des pics de corrélation correspondant (fort, moyen, faible, très faible). Adroite, les dix meilleures structures de chaque calcul sont superposées sur la structure moyenne (ruban) et les acides aminés colorés en fonction de leur caractère hydrophobe respectif. Les résidus hydrophobes et hydrophiles sont colorés en rouge et bleu, respectivement. Les différents jeux de structure sont représentés avec la même orientation (C-ter et N-ter précisés en A). Quelques résidus sont annotés en B et mettent en exergue le caractère amphipathique de l'hélice dans le TFE et les DPC sur deux

tours d'hélice (Ala[586]-Phe[594]). Pour la clarté des images, seuls les *squelettes* des structures calculées dans l'eau ont été représentés (A) et les hydrogènes enlevés partout ailleurs (B-D).

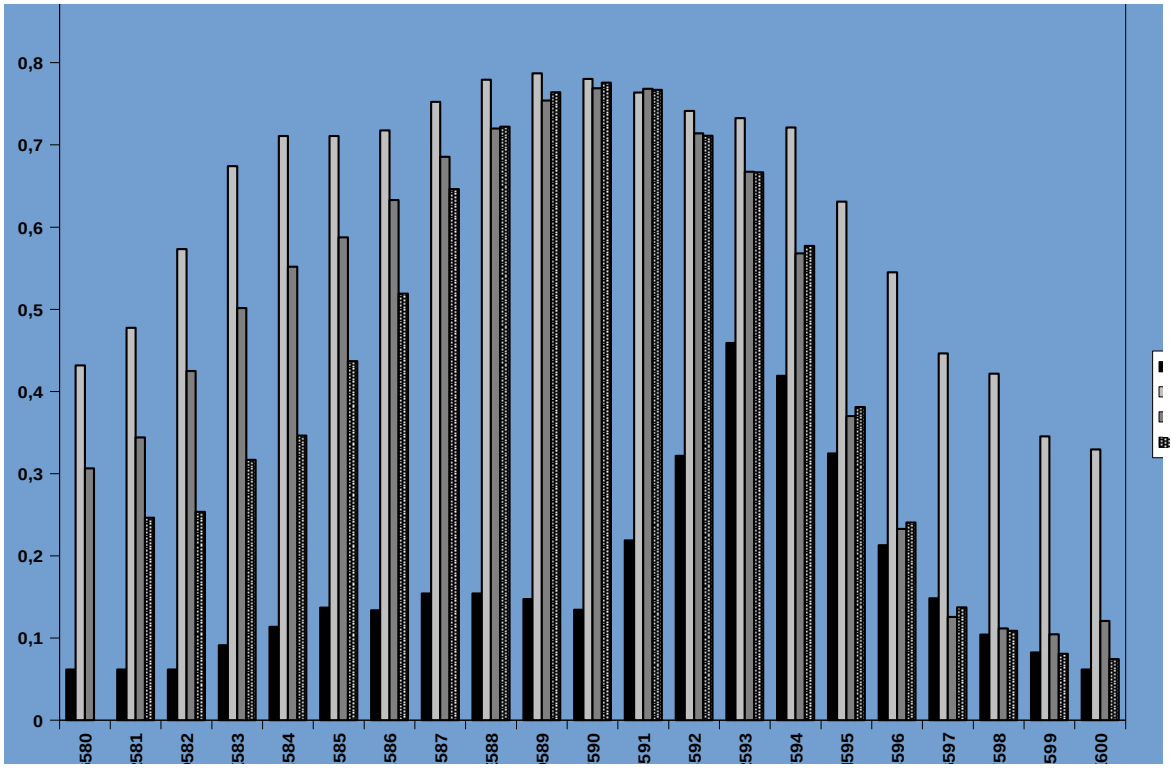



**Figure 46: Calcul du paramètre d'ordre ($S^2$) pour chaque acide aminé le long de la séquence de pUL89(580-600).** L'estimation a été faite sur le serveur http://wishart.biology.ualberta.ca/rci. en prenant en compte les déplacements chimiques Hα, Hβ et NH. La structure dans l'eau est peu stable alors que dans le TFE et le DPC elle est très ordonnée entre Ala[584] et Phe[595], principalement.

L'histogramme obtenu après calcul du paramètre d'ordre du peptide dans l'eau (Figure 46, barres noires) reflète effectivement l'existence des différents NOEs représentés en Figure 45A avec un domaine légèrement ordonné ($0,1<S^2<0,15$) entre Ala[584] et Phe[590] et un fragment Ile591-Ser[596] plus ordonné ($0,15<S^2<0,45$) en accord avec les différents NOEs *α,β (i,i+3)*, *α,N (i,i+2)*, *(i,i+3)* et *(i,i+4)* récoltés dans cette zone du peptide (Figure 45A).

Ces premiers éléments montrent donc que le peptide pUL89(580-600) a bien une propension à se replier en hélice. L'ajout de solvant organique ou de lipides est donc

justifiable par le fait qu'ils vont stabiliser la formation de l'hélice préexistante et non forcer sa formation.

De légères différences sont toutefois observées si l'on compare l'étendue de l'hélice dans les trois cas restants (DPC, TFE pH acide et neutre). Dans les DPC, l'hélice s'étend entre les résidus $Leu^{585}$-$Phe^{594}$ dont sept protons amides sont impliqués dans des liaisons hydrogènes d'après les expériences d'échange. Dans le TFE, l'hélice s'étend de manière différente selon le pH puisqu'elle contient les fragments $Lys^{583}$-$Phe^{594}$ et $Leu^{585}$-$Asn^{595}$ à pH acide et neutre, respectivement, en accord avec le RCI calculé pour ces deux jeux de structures (Figure 46, barres gris clair et gris foncé). Pour autant, certains ponts hydrogènes impliquant des protons amides ont été révélés par les expériences d'échange mais n'ont pas été utilisé comme contraintes car l'identification des oxygènes auxquels ils sont appariés n'était pas évidente. En effet ces résidus ne sont pas contenus dans l'hélice au vu des différents NOEs collectés ($Ile^{599}$ à pH acide et la triade $Tyr^{598}$-$Lys^{600}$ à pH neutre) ou initient l'hélice ($Asp^{582}$ et $Lys^{583}$ à pH neutre). Dans ce deuxième cas, ces résidus font participer leur carbonyle du *squelette* avec le proton amide du résidu *i*+4 et non l'inverse pour la formation de l'hélice.

L'analyse des RMSDs par rapport aux paramètres géométriques idéaux permet de conclure que les quatre jeux de structures sont de qualité comparable (Tableau 10). La prise en compte des différents NOEs et liaisons hydrogènes collectés ont ainsi permis de calculer des jeux de structures convergeant de manière significative en énergie et en géométrie (Tableau 10).

Les différents jeux de structures ont tous des énergies dont les médianes sont situées autour de 60 Kcal/mol (eau, DPC, TFE pH acide) ou environ 47 Kcal/mol pour le TFE à pH neutre. Par contre, la formation d'hélices en milieu hydrophobe quel qu'il soit permet une meilleure convergence des énergies des dix meilleures structures qui sont comprises dans un intervalle de 8 Kcal/mol dans le TFE à pH neutre, 11 Kcal/mol dans le TFE à pH acide et 15 Kcal/mol dans les DPC alors qu'il est de presque 30 Kcal/mol dans l'eau. La formation de l'hélice est donc bien stabilisante.

La superposition des structures deux à deux et par rapport à la structure moyenne sur les atomes lourds du *squelette*, dans chaque cas, montre une meilleure convergence en milieu hydrophobe (DPC et TFE) que dans l'eau.

L'analyse groupée des angles dièdres φ/ψ pour chaque jeu de structure montre elle aussi une conformité de nos structures aux canons géométriques des protéines au niveau du

*squelette* puisque au moins 97% des résidus de chaque jeu de structure sont compris dans les zones autorisées du diagramme de Ramachandran.

Nous avons donc obtenu dans le TFE et les DPC des structures semblables en hélice alpha sur des fragments de taille comparable et centrés sur les résidus Val$^{587}$-Gln$^{589}$ qui ont été identifiés par ailleurs comme étant impliqués dans la formation d'une structure hélicoïdale dans l'eau pure (Figure 45A, droite).

Ce résultat permet de conclure avec un haut degré de confiance que la structure du peptide pUL89(580-600) à l'intérieur de la protéine native de pUL89 est en hélice alpha. Par ailleurs, l'analyse de cette hélice basée sur la conservation des acides aminés et leur caractère physicochimique permet d'émettre des hypothèses sur l'implication de certains résidus dans une interaction probable avec pUL56 qui est la fonction originellement imputée à ce peptide par l'équipe de Bogner (Thoma, Borst et al. 2006).

L'alignement de séquences de pUL89 dans la région 580-600 (Annexe B1) montre la conservation totale de trois acides aminés (Lys$^{583}$, Ala$^{586}$ et Asn$^{595}$). Des résidus hydrophobes sont eux aussi hautement conservés (Phe$^{590}$, Ile$^{591}$, Phe$^{594}$ et Ile$^{599}$). Au niveau des hydrophiles seuls Asp$^{582}$ (substituable par un résidu polaire) et Glu$^{588}$ (conservé chez tous les homologues ou substitué par un aspartate chez quelques homologues alpha-herpesviraux) sont conservés. Comme le montrent les structures représentées en Figure 45B-D, Ala$^{586}$, Val$^{587}$, Phe$^{590}$, Ile$^{591}$, Phe$^{594}$ constituent une face hydrophobe et Glu$^{588}$, Gln$^{589}$, Ser$^{592}$ et Arg$^{593}$ une face hydrophile. La comparaison de la conservation des différents acides aminés constituant les deux faces permet d'envisager un rôle important des résidus hydrophobes dans la reconnaissance de pUL56 pour la constitution du complexe terminase.

Afin de pouvoir observer la formation éventuelle d'un domaine répondant aux critères des interfaces de par la répartition et la nature physicochimique des différents acides aminés, et ainsi appuyer l'hypothèse selon laquelle la face hydrophobe de l'hélice contenue dans pUL89(580-600) est impliquée dans l'interaction avec pUL56, nous avons synthétisé au laboratoire un peptide plus grand de pUL89 encadrant le fragment 580-600: pUL89(568-635).

| | H2O pH 3.5 | 120 eq DPC pH 3.7 | 30% TFE pH 3.3 | 30% TFE pH 6.8 |
|---|---|---|---|---|
| **Restraints for calculation** | | | | |
| Total NOE restraints | 274 | 386 | 458 | 440 |
| Intraresidue | 180 | 220 | 232 | 218 |
| Sequential (\|i-j\|=1) | 64 | 79 | 91 | 100 |
| Medium range (\|i-j\|≤5) | 30 | 87 | 135 | 122 |
| Hbonds | n.a. | 7 | 9 | 7 |
| **Structure statistics** | | | | |
| R.M.S.D. | | | | |
| Bonds (Á) | (2.71-3.25)E-03 | (2.59-3.08)E-03 | (2.59-3.08) E-03 | (2.17-2.6)E-03 |
| Bond angles (°) | 0.52-0.64 | 0.55-0.59 | 0.54-0.58 | 0.50-0.53 |
| Improper torsions (°) | 0.37-0.43 | 0.36-0.45 | 0.34-0.38 | 0.36-0.39 |
| NOE restraints (Á) | (3.14-4.13)E-02 | (2.65-3.32)E-02 | (2.59-2.94)E-02 | (2.10-2.61)E-02 |
| **Final Energies (kcal/mol)** | | | | |
| Total | 48.997-76.3745 | 53.44-68.02 | 52.76-63.24 | 43.73-51.37 |
| Bonds | 2.53-3.65 | 2.31-3.08 | 2.32-3.27 | 1.63-2.33 |
| Angles | 25.88-38.03 | 29.04-33.31 | 27.06-31.47 | 23.96-26.46 |
| Improper angles | 3.87-5.2 | 3.63-5.15 | 3.22-4.03 | 3.54-4.32 |
| Van Der Waals | 3.08-8.08 | 4.37-8.63 | 3.79-6.9 | 4.11-5.84 |
| NOE | 13.42-23.23 | 14.03-21.89 | 15.45-20.00 | 9.87-15.22 |
| **Ramachandran plot** | | | | |
| Most allowed (%) | 53.2 | 66.2 | 67.2 | 69.4 |
| Additionally allowed (%) | 38.7 | 29.8 | 26.7 | 26.7 |
| Generously allowed (%) | 8.1 | 1.5 | 4.4 | 3.3 |
| Disallowed (%) | 0.0 | 2.5 | 1.7 | 0.6 |
| **RMSD from mean structure (Á)** | | | | |
| Backbone atoms (overall) | 3.86±1.35 | 2.19±0.24 | 1.35±0.23 | 1.35±0.25 |
| Backbone atoms(2° structure) | 0.32±0.21 | 0.32±0.02 | 0.65±0.05 | 0.8±0.08 |
| **RMSD from pairwise (Á)** | | | | |
| Backbone atoms (overall) | 4.906±2.25 | 2.69±0.66 | 1.31±0.37 | 1.09±0.36 |
| Backbone atoms(2° structure) | 0.23±0.23 | 0.25 ±0.07 | 0.21±0.09 | 0.33±0.2 |

**Tableau 10: Statistiques des structures calculées de pUL89(580-600) dans différentes conditions.**

### C.2. Étude du peptide pUL89(568-635)

pUL89(568-635) a été sélectionné à partir de la détermination de douze domaines conservés à travers les homologues de pUL89 et a fait l'objet d'une publication (Champier, Hantz et al. 2007). Nous avons donc identifié pUL89(580-600) comme faisant parti de la région conservée XI (pUL89-XI) s'étendant entre les résidus Pro$^{575}$ et Ile$^{624}$. Afin de pouvoir observer le repliement de ce peptide sans que les extrémités ne soient trop exposées au

solvant, nous avons synthétisé un peptide rallongé de 7 et 11 acides aminé en N- et C-ter, respectivement: pUL89(568-635) que nous assimilerons pour des commodités d'annotation à pUL89-XI. pUL89-XI synthétisé correspond à la séquence suivante :

$^{568}$DQNHIEQPFYL*MGRDKALAV*EQFISRFNSGYIKASQEL*V*SYTIKLSHDPIEY LL*EQIQNL*HRVTLAEG$^{635}$

dont des résidus marqués 95% $^{15}$N, 15% $^{13}$C (suivi par * dans la séquence ci-dessus) ont été introduits pour faciliter l'attribution des signaux RMN. Le protocole de synthèse peptidique en phase solide de ce peptide est en cours d'optimisation mais a toutefois permis de recueillir, aussi faible en soit la quantité (0,8mg), suffisamment de matériel pur à plus de 95% pour en étudier la structure en solution par RMN.

A la différence de pUL89(580-600), pUL89-XI est très peu soluble dans l'eau. Il a donc fallu introduire 40% d'acétonitrile et agiter pendant 15 minutes au sonicateur.

### C.2.a. Dichroïsme circulaire

Tout comme pUL89(580-600), pUL89-XI a été préalablement analysé par dichroïsme circulaire. Sur un échantillon de pUL89-XI de 120µM (concentration mesurée à 280 nm par UV-visible) solubilisé dans 350µL (40% acétonitrile/ 60% H2O), quatre scans ont été enregistrés entre 190 et 260 nm par pas de 0,5nm avec un temps d'intégration de 2 s et une bande passante de 2 nm, et finalement moyennés. La courbe moyenne résultante a été lissée sur 25 points. Les enregistrements ont été répétés jusqu'à atteinte de l'équilibre du signal de dichroïsme circulaire. Les courbes sont représentées sur la Figure 47.

Comme dans le cas de pUL89(580-600), le signal est majoritairement dominé par la composante d'hélices alpha. Les différentes courbes montrent la solubilisation progressive du peptide avec le temps (signaux à 195 nm). Le rapport des intensités $[\theta]^{222nm}/[\theta]^{208nm}$ décroît légèrement avec le temps (de 0,88 à 0,84 au bout de 80 minutes) ce qui peut être expliqué par la formation de plusieurs hélices au cours du temps et non de l'élongation d'une seule hélice, chose qui va être confirmée en RMN.

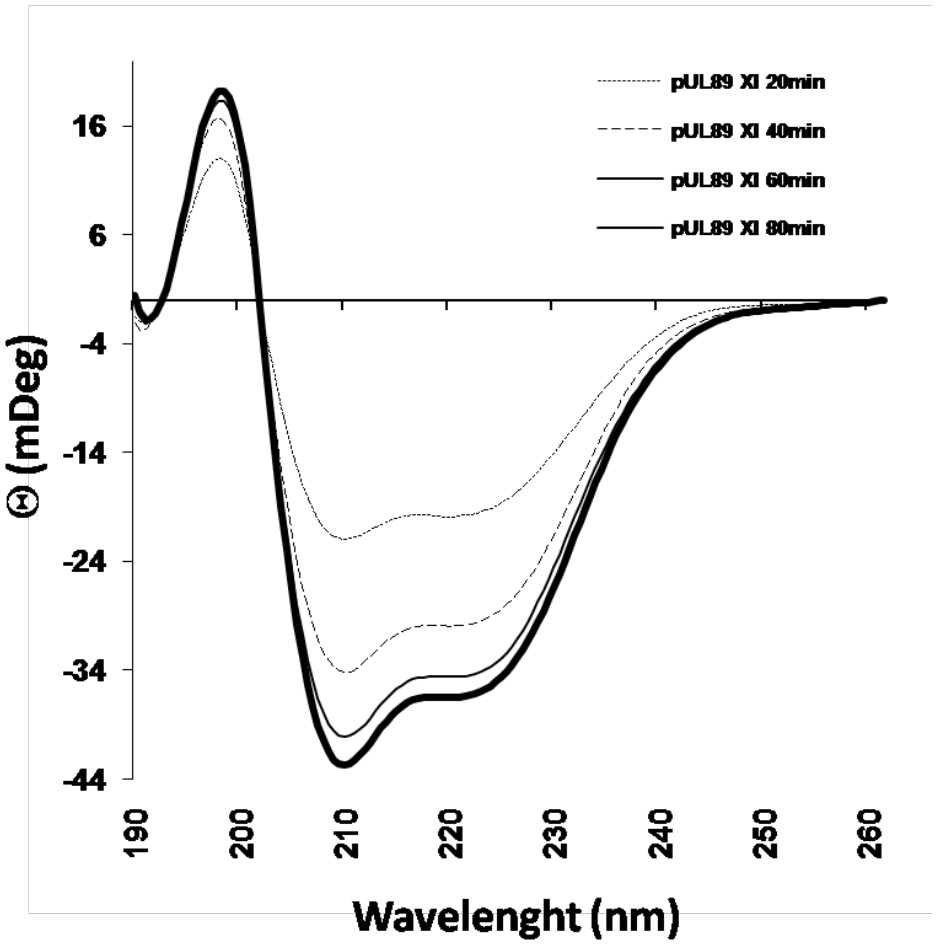



**Figure 47: Courbes de dichroïsme circulaire sur pUL89-XI dans 40% acétonitrile/ 60% H20 (v/v).** Les courbes ont été enregistrées sur un échantillon à 120µM entre 190 nm et 260 nm par pas de 0,5 nm, un temps d'intégration de 2s et une bande passante de 2nm. Quatre scans ont été enregistrés, moyennés et la résultante lissée sur 25 points. L'analyse a été répétée jusqu'à équilibre du signal.

### C.2.b. RMN de pUL89-XI

La faible quantité d'échantillon purifié nous a incité à analyser notre peptide en solution dans un mélange 50% d'acétonitrile/ 50% $H_2O$ pour un volume total de 450 µL soit une concentration de 800 µM, pH 3,2, 283 K. Les expériences ont été réalisées sur un spectromètre Bruker 600 MHz équipé d'une cryosonde. En plus des expériences homonucléaires TOCSY et NOESY, des expériences hétéronucléaires ont été réalisées pour faciliter l'attribution.

### *2.b.i. Expériences hétéronucléaires $^{15}N/^{1}H$*

Afin de faciliter l'étude structurale par RMN du domaine pUL89-XI, plusieurs acides aminés marqués $^{15}N$ et $^{13}C$ ont été incorporés lors de la synthèse peptidique. L'incorporation de ces acides aminés particuliers, $Leu^{578}$, $Val^{587}$, $Leu^{605}$, $Val^{606}$, $Leu^{621}$ et $Leu^{627}$, a permis la réalisation de spectres hétéronucléaires 2D 15N HSQC, 3D NOESY-HSQC et 3D TOCSY-HSQC. Cette dernière n'a pas permis d'obtenir un spectre satisfaisant. Ces différentes expériences ont permis d'apporter une aide à l'étape d'attribution du domaine pUL89-XI. En effet ce peptide étant structuré majoritairement en hélice alpha, un faible éclatement de résonances des protons amides a été observé, entraînant ainsi une superposition des signaux RMN dans cette région. L'introduction d'acides aminés marqués à permis de lever cette dégénérescence de déplacement chimique par l'introduction d'une nouvelle dimension en azote 15.

Les expériences hétéronucléaires font intervenir des constantes de couplage plus intenses permettant un transfert d'aimantation efficace entraînant une bonne sensibilité de ces expériences. L'expérience 2D 15N-1H HSQC (Heteronuclear Single Quantum Correlation) permet d'observer les pics de corrélation dus au couplage scalaire hétéronucléaire (90 Hz) entre l'hydrogène amide et l'azote de chaque résidu (Figure 48). Chaque pic observé sur le spectre correspond à une corrélation entre un proton amide et l'azote auquel il est directement lié. Un pic de corrélation est observé par acide aminé, en dehors des chaînes latérales.

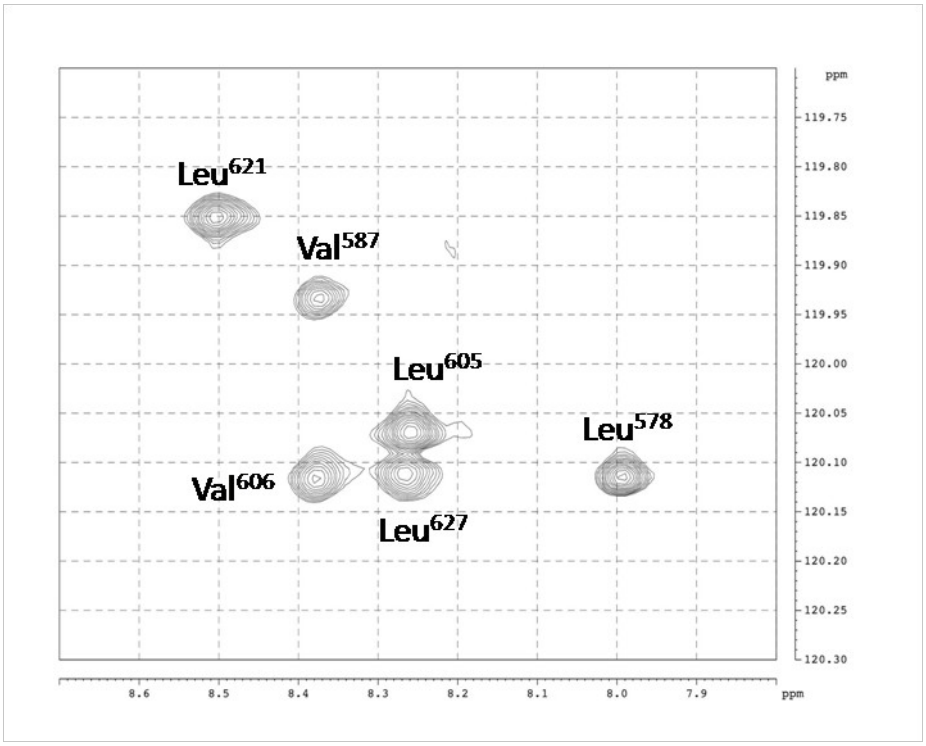



**Figure 48: Spectre 2D 15N HSQC du domaine pUL89-XI.** Chaque pic correspond à une corrélation entre un proton amide et l'azote qui le porte à travers leur couplage hétéronucléaire de 90 Hz. Un pic par acide aminé enrichi est observé. Les autres acides aminés n'apparaissent donc pas sur le spectre.

Les expériences tridimensionnelles hétéronucléaires en $^{15}N$ permettent quant à elles d'éclater l'information dans une troisième dimension et de lever ainsi les incertitudes d'attribution dues aux recouvrements des signaux. L'expérience 3D TOCSY-HSQC permet d'observer le système de spins auquel appartient chaque proton amide. L'expérience 3D NOESY-HSQC permet d'observer, pour chaque hydrogène amide, les proximités à travers l'espace, avec les autres atomes d'hydrogènes du peptide. Le principe d'attribution repose sur l'exploitation en parallèle de ces deux expériences, mais dans notre cas l'expérience 3D-NOESY-HSQC a permis de lever les indéterminations sur le spectre 2D NOESY.

L'expérience 2D HSQC permet de repérer les différents résidus enrichis de la molécule et de relever leurs déplacements chimiques $^{15}N$ et $^{1}HN$. Ces fréquences sont utilisées pour extraire les plans correspondants dans l'expérience 3D (Figure 49). Dans notre cas, nous avons mené une attribution en parallèle utilisant les expériences 2D et 3D.

Les plans ont ensuite été extraits de l'expérience 3D NOESY-HSQC et ont été attribuées. Ces bandes Figurent côte à côte et ne sont pas classées en fonction du numéro de l'acide aminé dans la séquence mais en fonction du déplacement chimique $^{15}N$ auquel est extrait le plan (Figure 50).

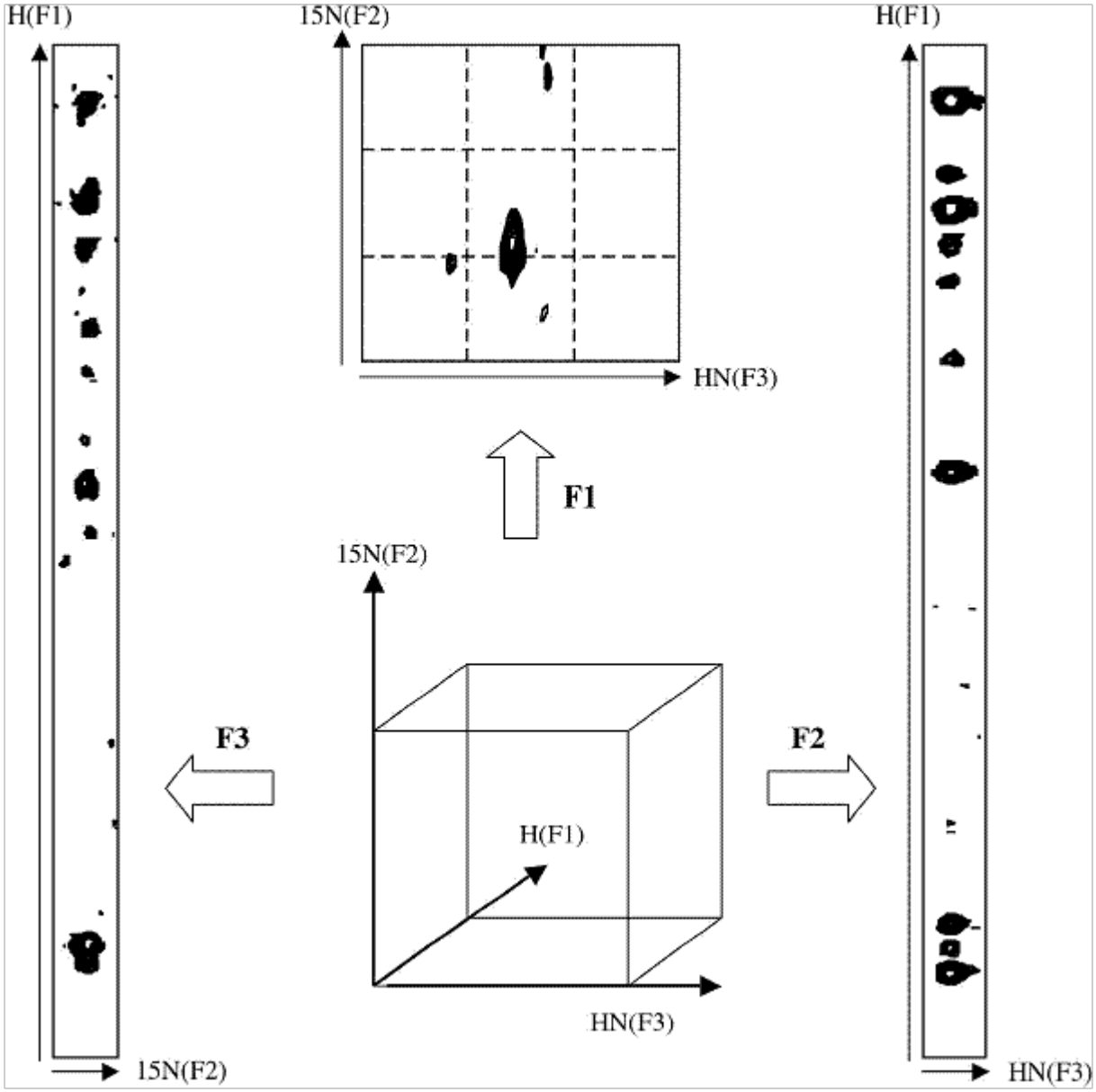



**Figure 49: Spectre 3D NOESY-HSQC.** Le spectre 3D est représenté sous la forme d'un cube dont les dimensions F1, F2 et F3 représentent respectivement les déplacements chimiques de tout proton, de l'azote 15 et du proton amide, respectivement. Un exemple de plan extrait de la 3D perpendiculairement à chaque dimension est donné en exemple.

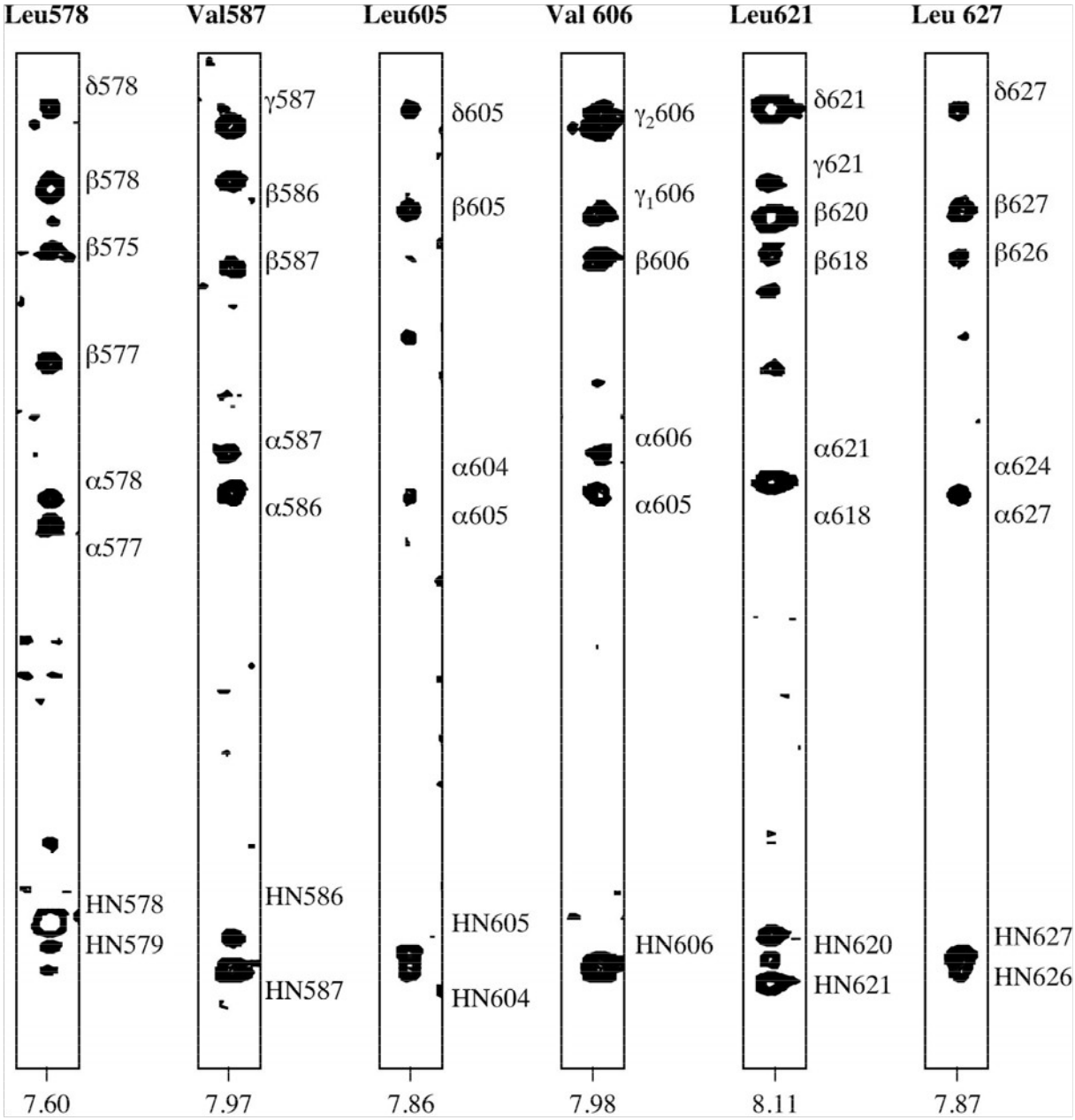



**Figure 50: Bandes extraites des plans de la 3D 15N NOESY-HSQC perpendiculairement à la dimension 15N.** Dans la dimension horizontale (F3), dimension de l'acquisition apparaît le déplacement chimique du proton amide lié à l'azote 15 à la fréquence duquel a été extrait le plan. Dans la dimension verticale (F1), apparaît le déplacement chimique des protons proches de ce proton amide à travers l'espace. Une partie des attributions a été reportée. La calibration a été effectuée en fonction du déplacement chimique du CD3CN (1.94 ppm) à 293 K.

### *2.b.ii. Expériences homonucléaires $^{1}H/^{1}H$*

Grâce au marquage isotopique des résidus Leu$^{578}$, Val$^{587}$, Leu$^{605}$, Val$^{606}$, Leu$^{621}$ et Leu$^{627}$ plusieurs points de départ pour l'attribution ont pu être envisagés et ont menés à l'attribution des tâches de corrélation (Annexe D3). Dû à un élargissement des signaux quelques protons n'ont pu être attribués (Lys$^{600}$ (Hα), par exemple). Cependant, nous avons réussi à collecter 598 contraintes NOEs dont 128 de moyenne portée (Tableau 11).

Quatre hélices ont finalement pu être modélisées en accord avec les expériences de dichroïsme circulaire: Pro$^{575}$-Gly$^{580}$, Asp$^{582}$-Phe$^{594}$, Asn$^{595}$-Ile$^{610}$ et Pro$^{616}$-Leu$^{632}$ (Figure 51). L'utilisation de l'acétonitrile s'est avérée indispensable à cause de la faible solubilisation de l'échantillon dans l'eau pure. L'acétonitrile a permis d'outrepasser cet inconvénient et de stabiliser les éléments de structures secondaires. Comme il a été montré dans le cas de Vpr, protéine de VIH 1(Morellet, Bouaziz et al. 2003), contrairement au TFE, l'acétonitrile ne déstabilise pas le repliement tertiaire des protéines.

Par ailleurs, nous n'avons pas utilisé de DPC dans cette étude pour des raisons pratiques et techniques. La première raison est que nous avons du tester plusieurs conditions pour solubiliser notre échantillon (eau, mélange méthanol/chloroforme, acétonitrile). Rappelons que nous disposions de peu d'échantillon (0,8 mg). Pour chaque nouveau milieu, il a fallu évaporer ou lyophiliser (suivant la proportion d'eau) l'échantillon. Ceci est chose aisée avec des solvants organiques volatiles. Par contre, l'introduction de DPC n'aurait autorisé aucun test postérieur.

La deuxième raison est la prise en compte de la taille du système. L'ajout de DPC qui se trouve sous forme micellaire aux concentrations utilisées augmente considérablement la taille de l'objet observé et donc son temps de corrélation. Ce paramètre est d'une importance crucial pour l'enregistrement d'expériences TOCSY. En effet, avec la taille du système croissante, le transfert d'aimantation à travers les liaisons se fait de façon moins efficace ce qui se traduit sur les spectres par des élargissements de signaux de faible intensité et par conséquent une attribution moins aisée.

En comparant les structures obtenues pour pUL89-XI et celles de pUL89(580-600), on remarque que le fait d'avoir allongé la séquence peptidique a permis de faire apparaître l'implication d'amino acides supplémentaires pour la formation de l'hélice comprise dans le fragment Gly$^{580}$-Lys$^{600}$ qui était comprise entre Ala$^{584}$et Phe$^{590}$ dans le peptide court dans les DPC et qui s'étend à présent entre Asp$^{582}$ et Phe$^{594}$.

De plus des hélices supplémentaires de part et d'autre du fragment 580-600 sont mise en évidence et implique d'une part Gly$^{580}$ et d'autre part les résidus Asn$^{596}$-Lys$^{600}$ qui initient une hélice se poursuivant jusqu'à Ile$^{610}$.

Prises individuellement, les différentes hélices sont formées de manière consensuelle sur l'ensemble du jeu de structures calculé appuyé par des RMSDs sur les atomes du *squelette* de l'ordre de l'angström (Tableau 11). Cependant, faute de signaux de structure tertiaire, leur orientation relative est peu définie ce qui aboutit à une grande disparité des conformations (Figure 51B). Un exemple est donné de cette mauvaise convergence dans le Tableau 11. La superposition sur les atomes du *squelette* et à plus forte raison sur l'ensemble des atomes lourds entre Asp$^{582}$ et Ile$^{610}$ aboutit à des RMSDs de 2,7 Å et 4,08 Å, respectivement.

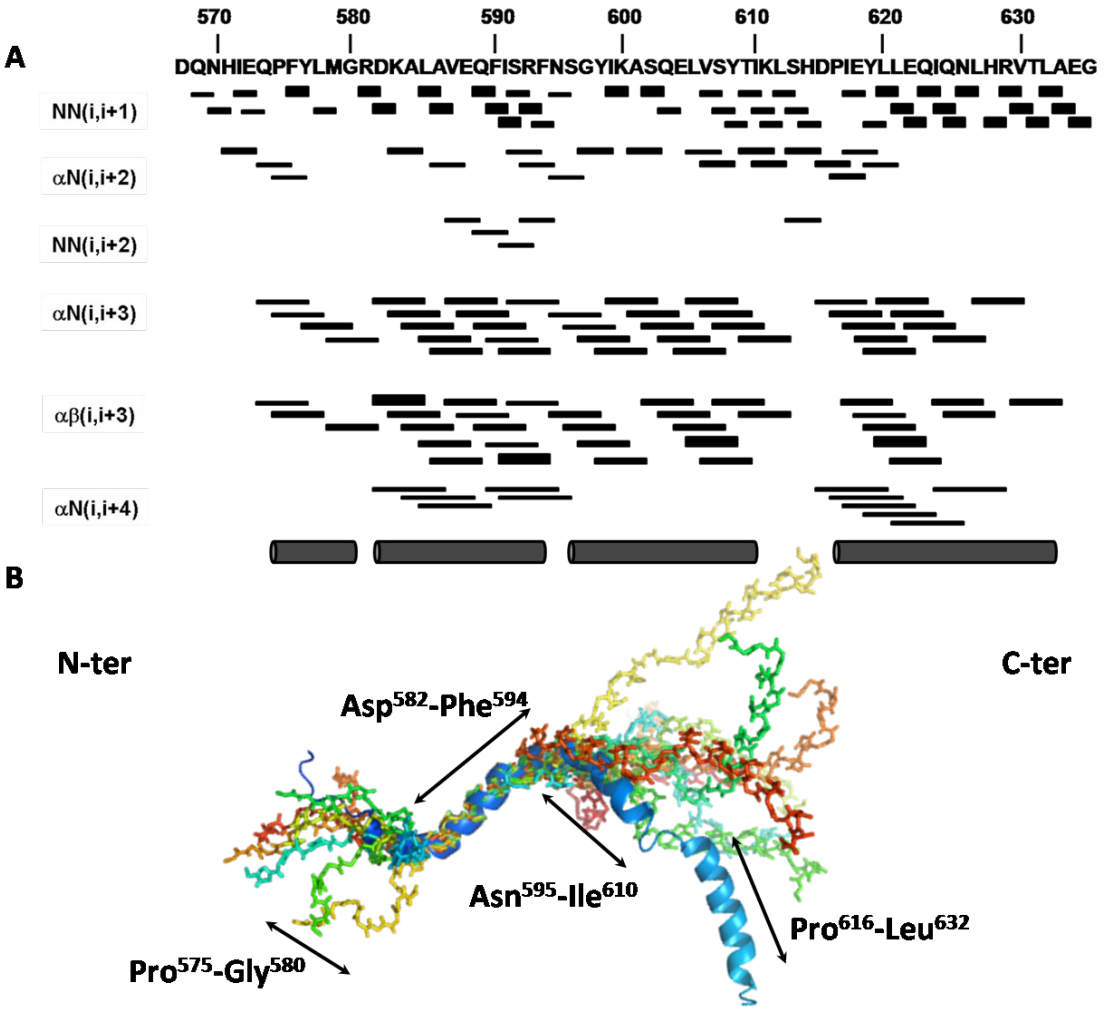



**Figure 51: Modélisation sous contraintes RMN de pUL89-XI.** A. La séquence est précisée en haut. Les connectivités NOEs propres aux structures secondaires précisées à

gauche et reliant deux résidus sont représentées par des lignes horizontales dont les épaisseurs sont proportionnelles à l'intensité des pics de corrélation observés (fort, moyen, faible, très faible). Les cylindres en bas d'image symbolisent les hélices observées après calcul. B. Superposition des dix meilleurs modèles sur la région $Asp^{582}$-$Phe^{594}$ en fonction de leur énergie avec 0 violations (≥ 0,2 Å) colorés par chaînes. L'orientation des flèches décrit les quatre hélices se référant à la structure de plus basse énergie représentée en ruban bleu.

| | | |
|---|---|---|
| **Restraints for calculation** | | |
| Total NOE restraints | 598 | |
| Intraresidue | 268 | |
| Sequential (\|i-j\|=1) | 202 | |
| Medium range (\|i-j\|≤5) | 128 | |
| **Structure statistics** | | |
| R.M.S.D. | | |
| Bonds (Å) | 1.68E-03 ± 8E-04 | |
| Bond angles (°) | 0.47 ± 8E-03 | |
| Improper torsions (°) | 0.33 ± 1E-03 | |
| NOE restraints (Å) | 2.3E-02 ± 2E-03 | |
| **Final Energies (kcal/mol)** | | |
| Total | 101.51 ± 9.03 | |
| Bonds | 3.31±0.58 | |
| Angles | 68.8±3.21 | |
| Improper angles | 9.61±0.11 | |
| Van Der Waals | 5.58±2.46 | |
| NOE | 13.94 ± 3.17 | |
| **Ramachandran plot** | | |
| Most allowed (%) | 81.1 | |
| Additionally allowed (%) | 16.1 | |
| Generously allowed (%) | 1.9 | |
| Disallowed (%) | 0.8 | |
| **RMSD from pairwise (Å)** | **Backbone atoms** | **Heavy atoms** |
| Pro575-Met579 | 0.37±0.19 | 1.34±0.61 |
| Asp582-Phe594 | 0.61±0.3 | 1.64±0.57 |
| Asp582-Ile610 | 2.7±1.24 | 4.08±1.83 |
| Tyr598-Ile610 | 0.94±0.48 | 1.82±0.7 |
| Pro616-Thr631 | 1.22±0.62 | 2.00±0.78 |

**Tableau 11: Statistiques sur les structures de pUL89-XI calculées dans 50% acétonitrile/ 50%$H_2O$, pH 3,2, 283K.**

Cette faible convergence des structures sur leur ensemble ne permet donc pas de discuter du packing de cette région de pUL89. Nous pouvons tout de même analyser la conservation des hélices en rapport avec les caractéristiques physicochimiques des acides aminés.

1. Pro$^{575}$-Gly$^{580}$ est une courte hélice de 6 acides aminés très conservée (Figure 52) à travers les herpesvirus et composée exclusivement de résidus avec un fort caractère hydrophobe (Pro, Phe, Tyr, Leu, Met). Les NOEs observés entre Tyr$^{577}$ et Gly$^{580}$ (α,N (i,i+3)) et entre Met$^{579}$ et Asp$^{582}$ (α,N (i,i+3) et α,β (i,i+3)) (Figure 51A) permet la formation d'un coude à la jonction entre cette hélice et l'hélice Asp$^{582}$-Phe$^{594}$.
2. Asp$^{582}$-Phe$^{594}$ est l'hélice contenue dans le fragment 580-600 étudié précédemment, rallongée en N-ter sous l'effet d'une augmentation de la taille de notre système. La majorité des résidus conservés sont localisés sur la même face de l'hélice qui est principalement composée de résidus hydrophobes (Ala$^{586}$, Val$^{587}$, Phe$^{590}$, Phe$^{594}$) et de Lys$^{583}$ qui, de part ces quatre méthylènes, contient un fort caractère hydrophobe en plus de son groupement amine polaire. L'autre face moins conservée, contient Asp$^{582}$, Leu$^{585}$ et Glu$^{588}$ (totalement conservé à travers les homologues de pUL89 ou remplacé par un aspartate, noté ν dans ET, Figure 52).
3. Asn$^{595}$-Ile$^{610}$ semble elle aussi très conservée puisque seulement 3 résidus sur 15 ne sont pas conservés (Tyr$^{598}$, Glu$^{604}$ et Tyr$^{608}$). Contrairement à Asp$^{582}$-Phe$^{594}$, cette hélice ne semble pas être amphipathique. On remarquera une parfaite conservation de plusieurs acides aminés: Asn$^{596}$, la triade $^{601}$Ala-Ser-Gln$^{603}$, Ser$^{607}$ et Thr$^{609}$. On peut donc envisager un rôle structural ou fonctionnel important de cette hélice en vertu du principe de conservation/fonction.
4. Pro$^{616}$-Leu$^{632}$, hélice très hydrophile (1 Asp, 2 Glu, 1 Asn, 2 Gln, 1 His, 1 Arg et 1 Thr), contient quatre résidus parfaitement conservés situés sur une même face de l'hélice (Asp615, Pro616, Tyr619 et Gln623). Il semble que la fin de l'hélice soit classe-spécifique (acides aminés notés x dans la ligne « ET » de l'alignement de séquence, Figure 52).

Le fait d'avoir rallongé la taille du système a entraîné une stabilisation de l'hélice insérée dans le fragment pUL89(580-600), et a permis de mettre en évidence l'existence d'une face hydrophobe conservée et une face hydrophile contenant entre autre le glutamate 588 dont le rôle reste à déterminer.

La comparaison des résultats expérimentaux (surlignage rouge, Figure 52) avec la prédiction de structure secondaire de PSI-PRED montre que les hélices Pro$^{575}$-Gly$^{580}$ et Asp$^{582}$-Phe$^{594}$ observées en RMN sont bel et bien retrouvées dans la prédiction. Par contre, les hélices Asn$^{595}$-Ile$^{610}$ et Pro$^{616}$-Leu$^{632}$ sont plus longues que celles prédites. En effet PSI-PRED a envisagé l'existence de trois brins bêta s'intercalant entre les hélices avec des taux de confiance élevés. Pourtant, nous considérons que l'expérience prévaut sur la théorie (même avec l'utilisation d'acétonitrile qui n'est pas un fort inducteur d'hélice) et par conséquent nous allons utiliser les données de structures secondaires obtenues par RMN pour la construction du domaine dans lequel est inséré pUL89-XI: le domaine endonucléase de pUL89 (pUL89-*endo*).

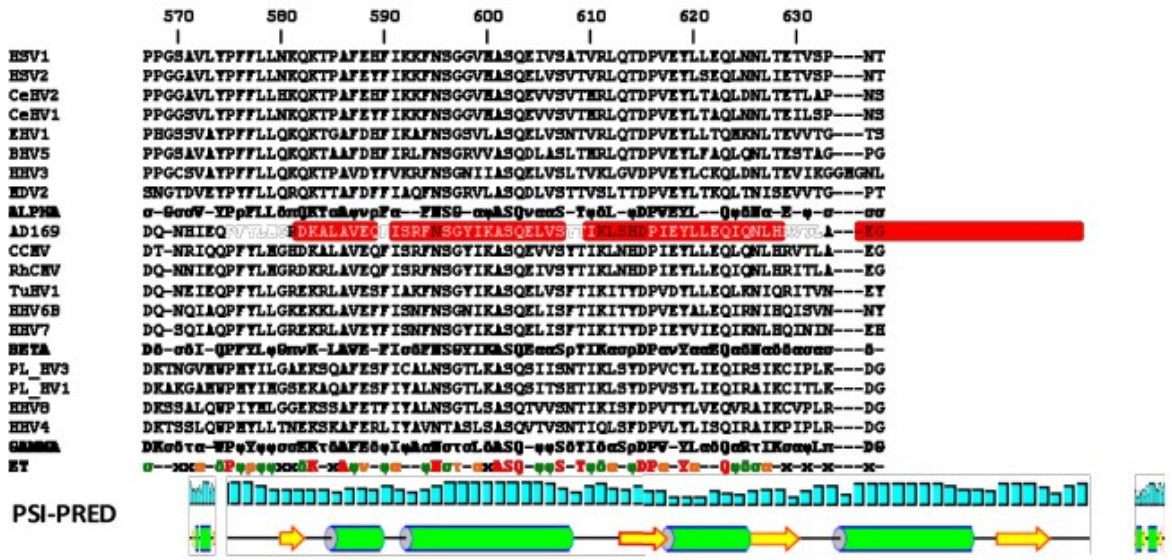



**Figure 52: Alignement de séquences de pUL89-XI avec ses homologues herpesviraux.** Extrait de l'annexe B1. Les hélices expérimentalement observées sont encadrées en rouge sur la séquence de pUL89 (souche de référence AD169). ALPHA, BÊTA et GAMMA sont les traces pour un même genre et ET la trace de l'évolution de l'ensemble des classes. Les résidus parfaitement conservés sont en majuscule (rouge dans ET), les hautement conservés (orange dans ET) correspondent à α (aliphatique), ρ (aromatique), ν (négativement chargé), π (positivement chargé) et τ (*tiny* ou minuscule), les moyennement conservés (vert dans ET) pour δ (polaire), φ (hydrophobe) et σ (*small* ou petit), en vertu du diagramme de Venn présenté en Figure 25. En bas de la figure est représentée la prédiction de structures secondaires obtenue par PSI-PRED (flèche et cylindre pour brin bêta et hélice alpha, respectivement) avec le taux de confiance schématisé par des barres cyans.

### C.3. Domaine endonucléasique de pUL89: pUL89-*endo*

Après avoir étudié par RMN le peptide pUL89(580-600) puis pUL89-XI, nous n'avons pas réussi à mettre en évidence un repliement tertiaire propre à ce domaine dû certainement au fait de sa taille trop faible comparé à la protéine entière. Nous avons donc exploré un domaine de la séquence encore plus grand de pUL89 auquel appartient le fragment 580-600 identifié par l'équipe de Bogner pour son rôle crucial dans l'interaction avec pUL56 (Thoma, Borst et al. 2006). Afin de sélectionner la séquence à étudier, nous avons croisé des informations présentes dans la littérature concernant les terminases au sens large et celles concernant pUL89 en particulier.

Comme nous l'avons déjà rappelé, les terminases sont chargées de reconnaître, cliver et transloquer l'ADN viral dans la procapside. Ces fonctions sont en grande partie assurées par les domaines hélicase et endonucléase d'une sous-unité et ATPasique située sur l'autre sous-unité du complexe terminase. Ce mécanisme est commun aux bactériophages et aux herpesvirus (Mitchell and Rao 2006; Ponchon, Boulanger et al. 2006; Rao and Feiss 2008) et nous avons donc tiré profit des informations provenant des bactériophages pour en déduire les équivalents chez pUL89.

Précédemment, nous avons réussi à identifier et à reconstruire par homologie avec gp17 du bactériophage T4 le domaine hélicase de pUL89 (chapitre 6.2). Nous avons aussi identifié un motif en doigt de zinc situé en C-terminal du domaine hélicase. Par déduction, nous avons émis l'hypothèse que la fonction endonucléase aspécifique imputée à pUL89 (Scheffczik, Savva et al. 2002) devait être contenue dans la partie C-terminal –i.e. au-delà du doigt de zinc potentiel $C^{367}X_2H^{370}X_{11}C^{383}XC^{385}$ (pUL89-ZF).

Comme il a été décrit par ailleurs pour les bactériophages, ce domaine contiendrait un site actif de la forme ββα-Me, c'est-à-dire contenant deux résidus acides portés par des brins bêta et un porté par une hélice alpha, le tout liant un ou plusieurs métaux ($Mg^{2+}$ ou $Mn^{2+}$) (Friedhoff, Franke et al. 1999; Yang, Lee et al. 2006; Rao and Feiss 2008; Yang 2008). Ponchon et al.(Ponchon, Boulanger et al. 2006) ont montré dans le cas des bactériophages que ce domaine était similaire à celui rencontré chez les ribonucléases H, les Holliday junction résolvases et les intégrases.

Nous avons donc envisagé de modéliser à notre tour le domaine endonucléasique de pUL89 à l'aide de MUSTER (Wu and Zhang 2008), serveur de reconnaissance de repliement, puis par l'ajustement des alignements de séquence que nous utiliserons finalement pour la reconstruction du modèle grâce à la version *standalone* de MODELLER (Eswar, Eramian et al. 2008).

### C.3.a. Modélisation de pUL89-endo

Afin d'identifier une ou plusieurs structure-support indispensable à la construction du domaine endonucléasique de pUL89, nous avons d'abord dû délimiter un domaine potentiel contenant le domaine endonucléasique. Pour ce faire, nous avons admis l'hypothèse que le site catalytique doit contenir trois résidus acides qui, de par leur fonction cruciale, doivent être très conservés à travers les homologues herpesviraux de pUL89. De notre alignement de séquences (Annexe B1), plusieurs aspartates et glutamates contenu en C-terminal de pUL89 (au-delà de pUL89-ZF) présentent un fort taux de conservation: $Asp^{463}$, $Glu^{534}$, $GLu^{588}$, $Asp^{615}$, $Asp^{650}$ et $Asp^{651}$. Ainsi, la séquence $Lys^{456}$-$Leu^{612}$ (157 acides aminés) a été soumise à MUSTER (http://zhang.bioinformatics.ku.edu/MUSTER). Les résultats de la recherche d'homologues structuraux par MUSTER et des modèles générés en conséquence sont répertoriés dans le Tableau 12 obtenu dans le fichier de sortie de l'analyse (http://zhang.bioinformatics.ku.edu/MUSTER/output/S2100/, résultats consultables sur le serveur sur une période de un an).

La fonction d'évaluation des modèles utilisée par MUSTER est représentée par le Z-score (colonne 5, Tableau 12). Celle-ci prend en compte plusieurs composantes qui vont être estimées pour chaque position de la séquence cible.

| Rank | Template | Align_length | Coverage | Zscore | Seq_id | Type | Target-Template alignments | 3-D models from threading alignments | Full-length models by MODELLER |
|---|---|---|---|---|---|---|---|---|---|
| 1 | 2ehgA | 135 | 0.859 | 3.404 | 0.141 | Bad | alignment_1 | threading_1 | model_1 |
| 2 | 1hjra | 132 | 0.840 | 3.375 | 0.159 | Bad | alignment_2 | threading_2 | model_2 |
| 3 | 1k1dA2 | 140 | 0.891 | 3.332 | 0.150 | Bad | alignment_3 | threading_3 | model_3 |
| 4 | 1h72c1 | 143 | 0.910 | 3.322 | 0.084 | Bad | alignment_4 | threading_4 | model_4 |
| 5 | 1ril_ | 131 | 0.834 | 3.183 | 0.115 | Bad | alignment_5 | threading_5 | model_5 |
| 6 | 1ejba | 144 | 0.917 | 3.175 | 0.097 | Bad | alignment_6 | threading_6 | model_6 |
| 7 | 1jl1a | 128 | 0.815 | 3.168 | 0.125 | Bad | alignment_7 | threading_7 | model_7 |
| 8 | 1euca2 | 146 | 0.929 | 3.149 | 0.137 | Bad | alignment_8 | threading_8 | model_8 |
| 9 | 1hqka | 131 | 0.834 | 3.129 | 0.130 | Bad | alignment_9 | threading_9 | model_9 |
| 10 | 1rvv1 | 131 | 0.834 | 3.085 | 0.168 | Bad | alignment_10 | threading_10 | model_10 |

**Tableau 12: Résultats de la recherche de structure-support par MUSTER de pUL89(456-612).** Les dix modèles générés par MUSTER sont classés (colonne 1) par leur Z-score (colonne 5). Les codes PDB (colonne 2), la taille du fragment aligné de la séquence de pUL89(456-612) exprimé en acides aminé (colonne 3) et en pourcentage (colonne 4), le taux de conservation (colonne 6), le type d'alignement en fonction du Z-score (colonne 7, « Bad » pour Z-score <0,75), les alignements de séquences (colonne 8), le fichier PDB contenant les Cα de la structure-support pour la phase de *threading* (colonne 9) et le modèle reconstruit par MODELLER après ajout des autres atomes du *squelette* et des chaînes latérales et dynamique moléculaire (colonne 10) sont représentés par chaque résultat.

1. Le premier terme est issu d'une recherche par homologie de séquences dans PSI-BLAST (Altschul, Madden et al. 1997). Il est décomposé en une fraction de similarité proche obtenue pour E-value fixé à 0,001 et une composante de similarité éloignée obtenue avec E-value à 1,0.
2. Le deuxième terme compare la prédiction de structures secondaires obtenue par PSI-PRED (Jones 1999) de la séquence cible et celle de la structure-support.
3. Le troisième terme mesure la déviation (par le biais du RMSD) obtenue sur des fragments de 9 amino acides (fragments graines) de la structure-support aligné (sans insertion de *gap*) avec d'autres structures non redondantes PDB homologues sélectionnées par PISCES (Wang and Dunbrack 2003).
4. Le quatrième terme compare l'accessibilité du solvant entre le résidu *i* de la séquence cible et celle observée pour le résidu aligné *j* de la structure-support calculé par STRIDE (Frishman and Argos 1995).
5. Les cinquième et sixième termes comparent les angles du squelette peptidique φ et ψ, respectivement, prédits par SVR-ANGLE (Wu and Zhang 2008) et calculés après *threading*.
6. Le septième terme compare les similarités de profil d'hydrophobicité, intimement liées au repliement moléculaire, par l'utilisation d'une matrice d'hydrophobicité (Silva 2008).
7. L'ajustement des *gaps* dans l'alignement de séquence est optimisé par l'utilisation de l'algorithme dynamique de Needleman-Wunsch (Needleman and Wunsch 1970). Pour prendre en compte l'éventualité qu'une séquence peut contenir plusieurs domaines

indépendants, MUSTER calcule le rapport de la taille du domaine aligné et de la structure-support utilisé plutôt que la prise en considération de la taille totale de la séquence cible par rapport à celle de la structure-support utilisé.

Ainsi, après la prise en considération de ces différentes informations pour chaque position de chaque modèle généré à partir des dix structure-supports sélectionnés, MUSTER n'a pas réussi à générer de modèle avec une grande fiabilité compte tenu des Z-scores inférieurs à 3,5. Les tests d'entraînement du serveur ont montré qu'un modèle est dit « *Easy* » (facile et donc crédible) pour un Z-score au moins égal à 7,5. Pour autant, MUSTER a permis de retrouver une homologie entre pUL89 et plusieurs ribonucléases H (codes PDB 2EHG, 1RIL et 1JL1) et une Holliday junction resolvase (code PDB 1HJR) concordant avec des observations similaires faites avec les bactériophages (Ponchon, Boulanger et al. 2006).

Le meilleur modèle obtenu à partir de la structure-support issue du fichier PDB 2EHG, c'est-à-dire la ribonucléase H de type 1 de *Solfolobus Tokodaii* (*Sto*-RNase HI) permet d'identifier au moins deux acides aminés susceptibles de participer à la poche de liaison à un ou plusieurs métaux.

Ces deux acides aminés sont $Asp^{463}$ et $Glu^{534}$ contenus tous deux dans des brins bêta. Reste donc à identifier l'acide aminé porté par une hélice alpha s'il on se réfère à la topologie canonique ββα-Me (Friedhoff, Franke et al. 1999). La Figure 53 montre que le glutamate 588 est dans ce calcul préliminaire proche des deux autres résidus catalytiques. Cependant, sa chaîne latérale n'est à ce stade pas tournée vers l'intérieur de la poche. Nous imputons ce défaut d'orientation à un mauvais alignement de séquence entre la séquence de pUL89-*endo* et la structure-support 2EHG. Afin de corriger cette erreur, nous avons entrepris de calculer nous même ce modèle par MODELLER.

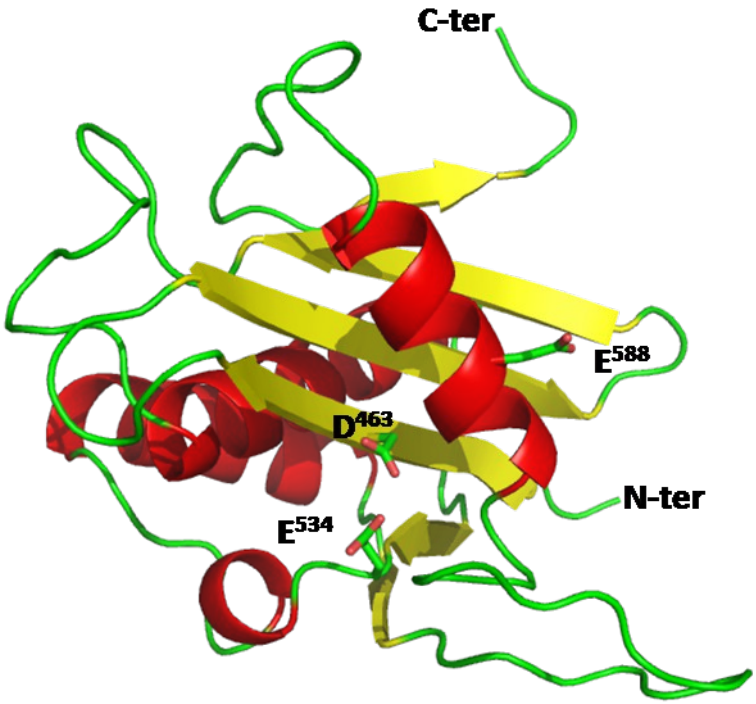



**Figure 53: Modèle de pUL89-*endo* calculé par MUSTER à partir de la structure PDB de la ribonucléase H de type 1 de *Solfolobus Tokodaii* (2EHG)**. Le squelette est coloré en fonction des éléments de structure secondaires (brins β jaunes et hélices α rouges). Les chaînes latérales d'Asp$^{463}$, Glu$^{534}$ et Glu$^{588}$ sont représentées en sticks et font apparaître un positionnement entre elles favorables à la formation d'une poche de liaison à un ou plusieurs métaux.

Pour construire nous même le modèle de pUL89(456-612) grâce à MODELLER version 9.6, nous avons ajusté l'alignement de séquences généré par MUSTER. Nous avons d'abord dû procéder à la recherche de structure-supports supplémentaires pour incorporer dans nos calculs les molécules d'eau, de magnésium et d'ADN indispensables à l'intégrité de notre modèle. Ceci a dans un premier temps nécessité la recherche dans la PDB d'une ribonucléase cocristallisée avec un duplexe pour pouvoir introduire le *squelette* de cette nouvelle structure-support dans l'alignement de séquence.

Dans un second temps, un réajustement manuel des alignements de séquences à du être opéré pour optimiser le modèle. Deux critères majeurs ont été considérés à cet effet: l'alignement des différents éléments de structures secondaires basé sur l'alignement des acides aminés catalytiques avérés dans les structure-supports et supposés dans la séquence de pUL89-*endo*.

Par recherche par mot-clé, nous avons réussi à trouver dans la PDB une RNase cocristallisée avec un duplexe hybride ARN/ADN issu de *Bacillus Hallodurans* (*Bachd*-RNase HI, PDB 1ZBI) (Nowotny, Gaidamakov et al. 2005; Nowotny, Gaidamakov et al. 2007). Du fait qu'aucun homologue plus proche de pUL89 n'a été identifié par MUSTER nous avons fait le choix d'étudier l'interaction pUL89-*endo* / acides nucléiques en se basant sur l'analogie avec cette RNase et son substrat. Les éléments de structures secondaires ont été alignés comme présenté en Figure 54.

Les acides aminés observés expérimentalement dans la formation des poches catalytiques de *Sto*-RNase et *Bachd*-RNase sont $Asp^{7}$, $Asp^{76}$ et $Asp^{125}$, et $Asp^{71}$, $Asp^{132}$Asn et $Asp^{192}$, respectivement. Pour la séquence de *Bachd*-RNase nous avons pris en compte la position mutée au niveau du deuxième résidu catalytique ($Asp^{132}$Asn) et tronqué la séquence au-delà de $Thr^{171}$ dont le repliement est très différent (*random coil*) de *Sto*-RNase et de la structure secondaire prédite de pUL89-*endo*. Finalement, l'alignement de séquences a reposé sur l'alignement d'$Asp^{463}$, $Glu^{534}$ et $Glu^{588}$ avec d'une part $Asp^{7}$, $Asp^{76}$ et $Asp^{125}$ de *Sto*-RNase, et d'autre part avec $Asp^{71}$et $Asn^{132}$ de *Bachd*-RNase (Figure 54A). Comme nous l'avons précisé à la fin de la partie traitant de pUL89-XI, nous avons pris en considération les éléments de structures secondaires observés en RMN. Nous avons exclus la partie C-terminal de pUL89-XI ($Asn^{613}$-$Gly^{635}$) qui, selon MUSTER, ne s'aligne pas avec les RNases. Peut être ce fragment appartient il a un autre domaine de pUL89. Nous avons réalisé les calculs de la forme libre et liée au duplexe séparément selon les alignements de séquence de l'annexe E1. Le fichier de contraintes utilisé pour le calcul de pUL89-*endo* lié et libre indifféremment est rappelé à l'Annexe E2.

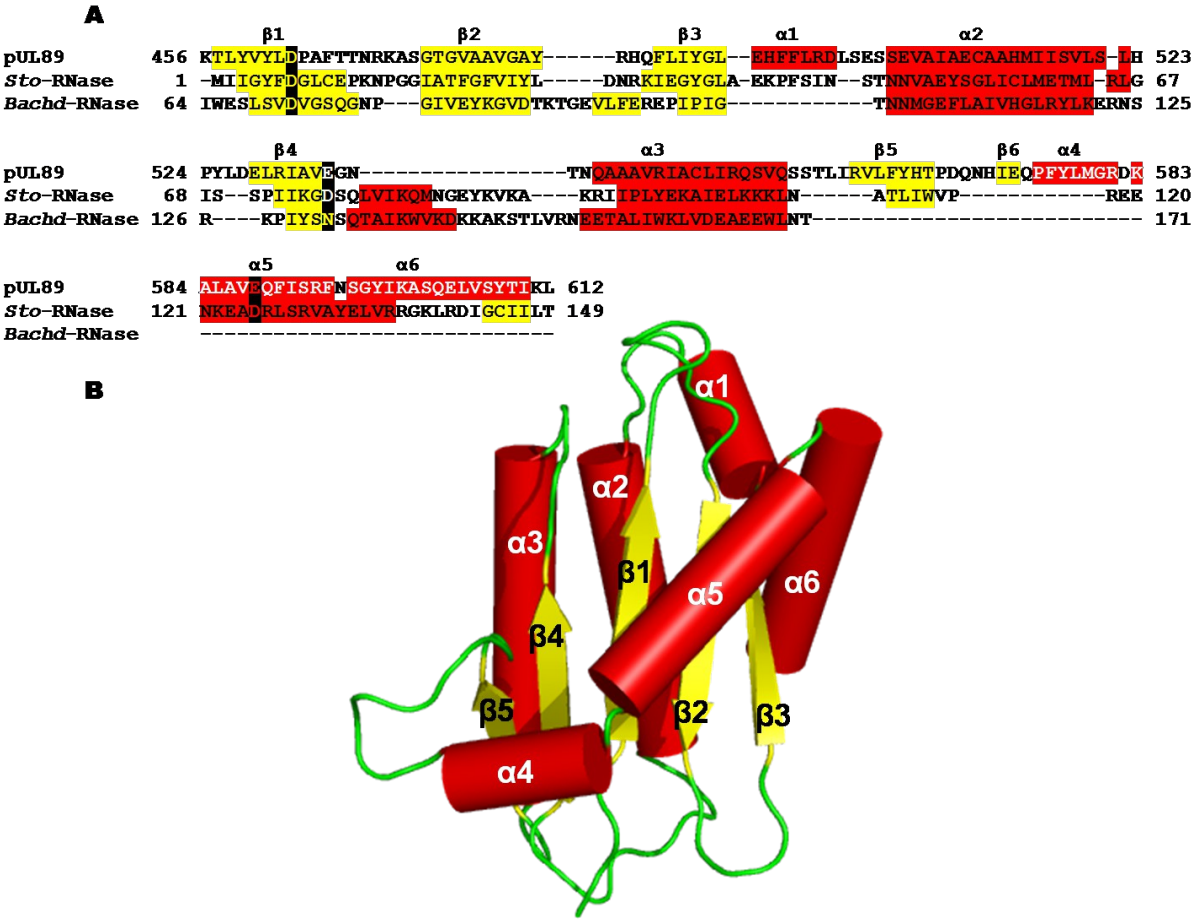



**Figure 54: Construction du modèle pUL89-*endo* par MODELLER.** A. Alignement de séquences entre pUL89-*endo* avec *Sto*-RNase (2EHG, chaîne A, fragment 1-149) et *Bachd*-RNase (1ZBI, chaîne A, fragment 64-171). Les brins β et hélices α prédits pour pUL89-*endo* et observés expérimentalement pour les structure-supports sont colorés en jaune et rouge, respectivement. Les hélices observées en RMN sont représentées avec des lettres blanches et imposées comme contraintes. Les résidus catalytiques alignés sont encadrés en noir. B. Modèle de pUL89-*endo* sans le duplexe ARN/ADN issu de l'alignement de A. L'annotation des éléments de structures secondaires se réfère à celle utilisée en A.

En plus des données RMN, nous avons forcé la formation des hélices et l'appariement des brins prédits par PSI-PRED. L'introduction des molécules d'eau et de magnésium ont nécessité pour leur positionnement dans le site actif, l'ajout de contraintes de distances relevées sur les fichiers PDB 2O6M (Eastberg, Eklund et al. 2007), 3C6A (Sun, Kondabagil et al. 2008) et 2QNC (Biertumpfel, Yang et al. 2007) et en accord avec la littérature (Harding

1999). De cette façon, les contraintes entre $O^{\delta 1}$ de $Asp^{463}$ et $O^{\varepsilon 1}$ de $Glu^{534}$ avec $Mg_1$, $O^{\delta 2}$ de $Asp^{463}$ et $O^{\varepsilon 1}$ de$Glu^{588}$ avec $Mg_2$ ont été imposées à des distances de 2,6 ± 0,5 Å. Le duplexe hybride ARN/ADN a dû être introduit comme objet figé (limite de MODELLER) ce qui a tout de même permis l'appariement préférentiel de chaînes latérales de la protéine avec les charges et dipôles portés par les chaînes d'acides nucléiques. 100 modèles ont été calculés pour la forme liée (pUL89-*endo*$^{liée}$) et la forme libre (pUL89-*endo*$^{libre}$).

### C.3.b. Qualités des structures théoriques

L'estimation de la qualité des structures est un défi dans la modélisation de structures théorique du fait de leur caractère spéculatif. Plusieurs outils disponibles sur Internet sont disponibles permettant l'analyse de structures (théoriques ou expérimentales). Parmi celles-ci, nous avons considérés deux outils en particulier: QMEANclust (http://swissmodel.expasy.org/qmean/cgi/index.cgi) et Prosa (http://www.came.sbg.ac.at/typo3/index.php?id=prosa).

QMEANclust (Benkert, Kunzli et al. 2009) dérive de la fonction QMEAN qui est une combinaison linéaire de plusieurs termes énergétiques calculée pour chaque structure soumise à l'analyse. QMEAN inclut des potentiels sur les angles dièdres, sur la solvatation individuelle de chaque amino acide, sur les interactions à longues distances, mais aussi l'estimation de l'accessibilité au solvant et la comparaison entre les structures secondaires prédites et calculées. Chaque terme est pondéré par des coefficients empiriques et le QMEAN score finalement calculé. A la fonction QMEAN, QMEANclust introduit un dernier paramètre qui traduit la convergence locale d'un ensemble de modèles (cluster) pour chaque position de la séquence. Cette composante est dérivée de la comparaison entre un modèle donné du cluster et la structure médiane de ce cluster. QMEANclust introduit donc la notion de densité structurale. Plus un élément de structure sera redondant plus celui-ci sera statistiquement correct. Le score donné par QMEANclust est compris entre 0 pour un mauvais, et 1 pour un bon modèle. L'avantage pratique de QMEANclust réside dans l'analyse simultanée de plusieurs structures (dans notre cas 100 modèles pour chaque cluster pUL89-*endo*$^{liée}$ et pUL89-*endo*$^{libre}$) ce qui offre un gain de temps considérable.

Le deuxième outil que nous avons utilisé est ProSA (Wiederstein and Sippl 2007). Celui-ci calcule les énergies des carbones alphas en fonction des potentiels d'interactions avec les atomes voisins appartenant à la protéine et d'exposition au solvant. Deux informations vont découler de ce calcul d'énergie: un *z-score* et un graphique d'énergie résidu par résidu. Le *z-*

*score* traduit la qualité du modèle dans son ensemble en comparant l'énergie totale calculée pour le modèle à celles obtenues pour des structures expérimentales (RMN et RX). Le *z-score* est placé dans un graphique regroupant les résultats en fonction de la taille de la séquence en amino acides obtenus par RMN et RX. Il est donc aisé de vérifier la cohérence de chaque modèle par rapport à des systèmes dont les structures ont été résolues expérimentalement.

Le graphique énergétique local permet d'identifier les contributions énergétiques le long de la séquence calculé sur une fenêtre de 40 résidus. Contrairement à QMEANclust, ProSA ne permet pas l'analyse simultanée de plusieurs modèles. Nous avons donc analysé avec ce serveur les dix meilleures structures de chaque cluster, pUL89-*endo*$^{liée}$ et pUL89-*endo*$^{libre}$.

L'estimation de la qualité des deux clusters a donc été faite dans un premier temps par QMEANclust. Les dix meilleures structures (QMEANscore les plus élevés) ont révélé des qualités similaires pour les deux jeux de structures générés. Ceux-ci sont compris dans les intervalles 0,866-0,873 et 0,822-0,835 pour pUL89-*endo*$^{liée}$ et pUL89-*endo*$^{libre}$, respectivement (Tableau 13, colonne 4). L'analyse postérieure opérée par ProSA révèle que les dix meilleures structures de chaque cluster appartiennent toutes à l'intervalle -1<*z-score*<-9 calculé pour des structures expérimentales de tailles égales à celle de pUL89-*endo* (i.e. 157 acides aminés).

En effet, pour les dix meilleures structures (selon QMEANclust) des clusters pUL89-*endo*$^{liée}$ et pUL89-*endo*$^{libre}$, les *z-scores* sont compris dans les intervalles tels que -2,58 <z-score <-3,58 et -3,05 <z-score < -3,75, respectivement (Tableau 13, colonne 5).

Une analyse complémentaire de la qualité des angles dièdres du *squelette* a été réalisée par PROCHECK et révèle dans les deux jeux de structures des valeurs standard rencontrées dans les protéines en vertu du diagramme de Ramachandran (Tableau 13, colonne 6 à 9).

Nous pouvons donc conclure de ces analyses que bien que MUSTER n'ait pas réussi à détecter d'homologue proche du domaine pUL89-*endo*, l'utilisation de RNase HI comme structure-support pour la modélisation théorique de sa structure a conduit à deux jeux de structures de qualité raisonnable puisque les topologies (PROCHECK) et énergies (ProSA) sont comparables à celles observées pour des structures expérimentales. L'introduction d'un duplexe hybride ARN/ADN ne semble pas affecter la qualité des structures calculées mais permet au contraire, par rapport à la forme libre, une meilleure convergence des structures sous l'influence des interactions électrostatiques formées entre la protéine et le duplexe (QMEANclust scores) et de la rigidité introduite.

Tableau 13: Résultats de l'estimation des structures obtenues par MODELLER pour pUL89-*endo*$^{liée}$ et pUL89-*endo*$^{libre}$. Les numéros des modèles (colonne 2) sont ceux attribués par ordre de calcul par MODELLER pour les deux jeux de structures (colonne 1). Les modèles sont classés pour chaque cluster (colonne 3) par rapport à leurs QMEANclust scores (colonne 4). Chaque structure a été soumise à une analyse par ProSA dont le *z-score* calculé est précisé à la colonne 5. Une analyse supplémentaire des angles dièdres φ et ψ placés dans un graphique de Ramachandran a été effectuée par PROCHECK (colonnes 6 à 9). Les valeurs indiquées par PROCHECK sont exprimées en pourcentage.

Nous allons à présent exploiter les résultats de ces deux jeux de structures. Pour ce faire, nous ne prendrons en considération que la meilleure structure obtenue dans chaque calcul, c'est à dire le modèle 77 de la forme libre et le modèle 5 de la forme liée (numérotation par ordre de structure calculée pat MODELLER). Nous allons discuter de trois aspects du domaine pUL89-*endo* reflétés par nos calculs: premièrement, le site de liaison aux métaux, deuxièmement, la surface d'interaction entre pUL89-*endo* et le duplexe, et troisièmement le site d'interaction potentiel avec pUL56.

### C.3.c. Analyse de pUL89-endo

#### *3.c.i. Poche de liaison aux métaux*

Les acides aminés importants pour la chélation des magnésiums ont été déduits des résultats de MUSTER comme cela a été expliqué précédemment (Figure 53). Cependant, les endonucléases peuvent contenir au sein de leur site actif de un à trois ions métalliques différents (magnésium ou manganèse) (Pingoud, Fuxreiter et al. 2005; Yang 2008). Pour pouvoir évaluer dans notre cas combien de magnésium potentiels sont logés *a priori* dans le site actif, nous avons procédé à une inspection de la conservation des acides aminés à proximité de la triade acide.

Dans le cas d'un mécanisme à un métal, la triade acide est complétée par un résidu basique très conservé qui va orienter les chaînes latérales des résidus acides et d'une molécule d'eau qui va accomplir l'hydrolyse du lien phosphodiester.

Le mécanisme à deux métaux substitue ce résidu basique par un deuxième métal. Dans notre cas, aucun résidu basique conservé n'a été identifié à proximité de la triade acide. De

plus, les RNases qui ont servi de support pour l'obtention des modèles calculés requièrent la fixation de deux magnésiums pour accomplir la catalyse (Yang, Lee et al. 2006). De ce fait, nous avons fait l'hypothèse que cela devait aussi être le cas pour pUL89-*endo*. Par analogie avec les RNases, nous proposons le mécanisme suivant: chélation d'un magnésium par les carboxylates des résidus $Asp^{463}$ et $Glu^{588}$ ainsi que par les oxygènes du lien phosphodiester au niveau des atomes $O_{3'}$ et $OP_1$ permettant la stabilisation de l'état de transition lors de la catalyse, pontage des deux magnésiums par l'intermédiaire de $Asp^{463}$ qui va maintenir à une distance d'environ 4Å les deux ions métalliques, chélation du deuxième magnésium par deux molécules d'eau et la chaîne latérale de $Glu^{534}$ qui va optimiser le positionnement favorable d'une des deux molécules d'eau pour l'attaque nucléophile sur le phosphate du squelette du brin à cliver (Figure 55).

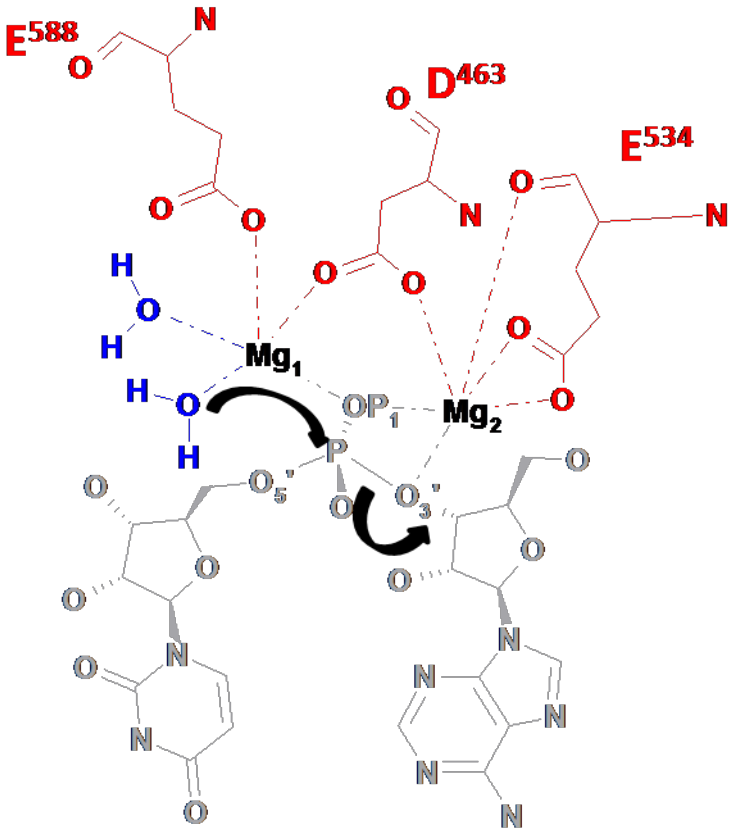



**Figure 55: Chélation des magnésiums et Mécanisme de l'hydrolyse de la liaison phosphodiester au sein du site catalytique de pUL89-*endo*.** Les résidus acides, les molécules d'eau, les acides nucléiques et les magnésiums sont colorés en rouge, bleu, gris et noir respectivement. Les flèches noires explicitent le mécanisme réactionnel conduisant à la lyse du lien phosphodiester.

### *3.c.ii. Surface d'interaction avec le duplexe*

Le modèle calculé en présence du duplexe hybride ARN/ADN, bien que non conforme avec la préférence naturelle de pUL89 à cliver l'ADN double-brin (Scheffczik, Savva et al. 2002), nous a permis d'identifier un grand nombre d'interactions pouvant s'établir entre l'enzyme et le duplexe. Notre étude a été conduite par la comparaison de l'interface observée expérimentalement chez *Bachd*-RNase (code PDB 1ZBI) et celle observée dans le cas de pUL89-*endo* après calculs. *Bachd*-RNase interagit avec le duplexe hybride en conformation B au niveau de son petit sillon par l'intermédiaire de plusieurs résidus polaires ou chargés (Nowotny, Gaidamakov et al. 2005) (Figure 56A).

L'analyse de notre structure en complexe avec le duplexe fait apparaître une grande similarité dans la composition des interfaces entre *Bachd*-RNase et pUL89-*endo*. En effet, plusieurs amino acides de la RNase trouvent un équivalent chez pUL89. De plus, la grande majorité des résidus identifiés comme ayant un rôle critique dans la liaison à l'ADN sont très conservés à travers les homologues herpesviraux de pUL89 (*squelette* en dégradé de vert dans la Figure 56B). Nous pouvons donc prédire les interactions suivantes.

Pour le brin d'ARN (brin subissant le clivage), à l'instar de $Gly^{73}$, $Ser^{74}$, $Lys^{180}$ et $Thr^{183}$ de la RNase, les résidus $Pro^{464}$, $Thr^{467}$, $His^{565}$ et $His^{571}$ (et/ ou $Gln^{574}$) interagiraient avec les groupements phosphates alors que, comme $Gly^{76}$, $Asn^{105}$, $Gln^{134}$ et $Lys^{180}$ de *Bachd*-RNase, $Thr^{475}$, $Ser^{504}$, $Asn^{536}$et $His^{565}$ créeraient des interactions avec les sucres. Nous n'avons pas considéré ici les acides aminés de la RNase interagissant avec les oxygènes 2'-OH des riboses car non rencontrés dans le système natif pUL89/dsDNA.

Le brin d'ADN est quant à lui maintenu par des contacts polaires avec les hydroxyles des sucres par l'intervention de $Asn^{469}$et $Ser^{504}$ (contreparties de $Asn^{77}$et $Asn^{105}$ de *Bachd*-RNase) et avec les groupements phosphates *via* $Ser^{503}$, $Asn^{538}$, $Arg^{470}$ et/ou $Lys^{471}$, et $Gln^{539}$ similaires à $Thr^{104}$, $Trp^{139}$, $Ser^{147}$ et $Thr^{148}$, respectivement.

La stabilisation du duplexe implique donc la participation d'un nombre important de résidus incluant les fragments $Gly^{535}$-$Gln^{539}$ et $Pro^{464}$-$Val^{477}$ directement adjacents aux résidus catalytiques $Glu^{534}$ et $Asp^{463}$ et situés au centre de la surface d'interaction implique un rôle crucial de ces résidus pour la fixation du duplexe et son clivage. Ces deux fragments semblent être accompagnés de la triade $^{595}$Asn-Ser-$Gly^{597}$. Cette triade dont nous avons parlé aux paragraphes traitant de la RMN de pUL89(580-600) et pUL89-XI peut avoir un rôle dans le maintien de l'ADN.

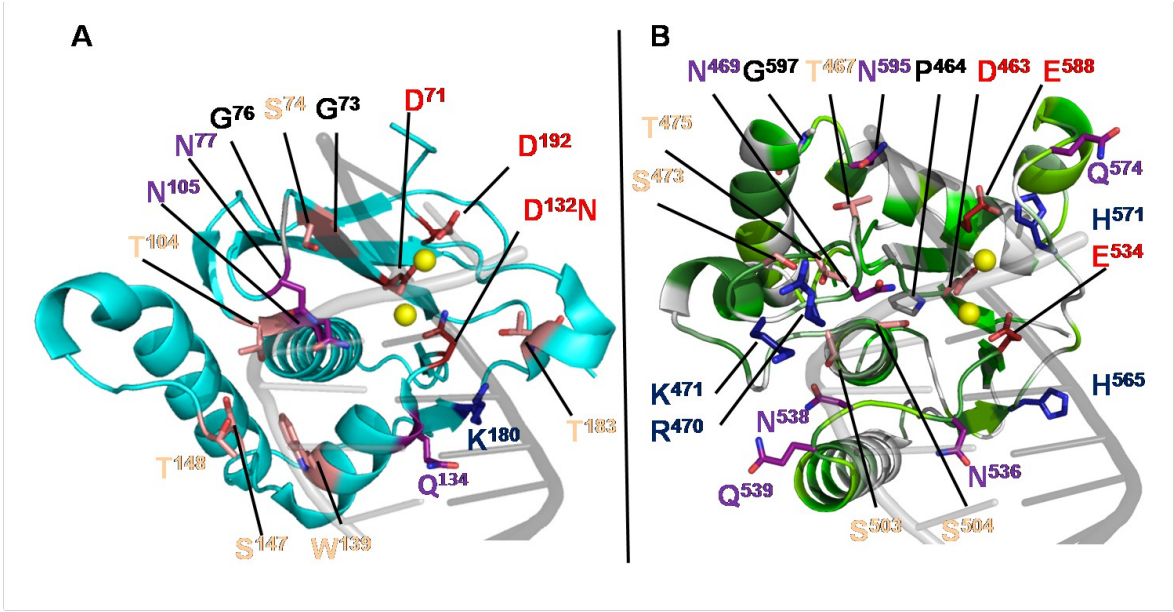



**Figure 56: Comparaison des interfaces avec le duplexehybride ARN/ADN lié à A. *Bachd*-RNase (code PDB 1ZBI) et B. pUL89-*endo* (modèle 5 du cluster pUL89-*endo*$^{liée}$).** Les chaînes latérales des acides aminés interagissant avec le duplexe ou les ions magnésiums (sphères jaunes) sont représentés en sticks et colorés selon leur caractère physicochimique: rouge, bleu, violet, saumon et gris pour négativement chargé, positivement chargés, asparagine et glutamine, donneurs de liaisons hydrogène et apolaires, respectivement. Le brin d'ARN est représenté en ruban transparent gris foncé et le brin d'ADN en ruban transparent gris clair. En B. le squelette de pUL89-*endo* est coloré avec une échelle de vert fonction du taux de conservation de chaque acide aminé. La majorité des résidus potentiels interagissant avec le duplexe sont conservés à travers les homologues herpesviraux de pUL89.

### *3.c.iii. Impact de l'interaction avec pUL56*

Rappelons que la reconstruction théorique des modèles de pUL89-*endo* sous les formes liée et libre a été motivée par le désir d'expliquer l'influence de l'interaction entre les deux sous-unités du complexe terminase de CMVH sur l'activité endonucléasique propre à pUL89 révélée par Scheffczik et al. (Scheffczik, Savva et al. 2002).

Grâce aux modèles obtenus, nous avons réussi à prédire quels acides aminés de pUL89 forment la poche de liaisons aux métaux et quels sont ceux qui sont chargés de maintenir le substrat durant la catalyse. Nous avons ainsi émis l'hypothèse que le glutamate Glu$^{588}$ lie un ion magnésium. Précédemment nous avions observé par RMN que le peptide pUL89(580-600) est constitué d'une hélice amphipathique entre Asp$^{582}$ et Phe$^{594}$. Nous avions aussi souligné une grande conservation de la face hydrophobe alors que la face hydrophile ne

contenait que peu de résidus conservés dont Glu$^{588}$. Par conséquent, le rôle crucial de pUL89(580-600) décrit par Thoma et al. (Thoma, Borst et al. 2006) peut s'expliquer par la participation de Glu$^{580}$ dans la liaison à un magnésium, facteur indispensable au clivage de l'ADN lors de la phase de maturation. Son interaction avec pUL56 est donc, par déduction, assurée par la face hydrophobe de l'hélice.

Grâce au modèle du domaine endonucléasique présenté ici, nous pouvons replacer le fragment pUL89(580-600) dans le contexte du domaine endonucléase de pUL89. La Figure 57A montre pUL89-*endo* selon une vue dans l'axe du duplex. Les différents amino acides annotés sont ceux dont les chaînes latérales sont exposées de manière similaire aux résidus Ala$^{586}$ et Phe$^{590}$ portés par la face hydrophobe de l'hélice Asp$^{582}$-Phe$^{594}$.

Ici encore, nous faisons le parallèle entre la conservation à travers les homologues herpesviraux de pUL89 des résidus annotés et le caractère physicochimique.

Nous observons alors que les résidus aliphatiques Leu$^{458}$, Val$^{480}$, Ala$^{482}$, Ile$^{599}$, Val$^{606}$, Ile$^{610}$ et les aromatiques Tyr$^{461}$, Tyr$^{490}$ et His$^{523}$ forment un patch hydrophobe très conservé. Ce patch hydrophobe est encadré de résidus hydrophiles très conservés dont Lys$^{583}$, Gln$^{603}$ et Ser$^{607}$. La surface créée par cet ensemble d'amino acides est donc en accord avec la topologie rencontrée dans le cas des interfaces protéine-protéine. En effet, la densité importante de résidus hydrophobes peut être capable d'assécher la surface d'interaction avec pUL56 afin de stabiliser le complexe. Les résidus hydrophiles peuvent jouer un rôle dans la spécificité de l'interaction avec pUL56 (Bogan and Thorn 1998; Hu, Ma et al. 2000; Glaser, Steinberg et al. 2001; DeLano 2002).

Rappelons que pUL56 et pUL89 se comportent de manière différente avec l'ADN. Lorsque l'ADN viral est encore sous forme concatémérique à la suite de sa synthèse en cercle roulant, pUL56 est capable de reconnaître spécifiquement les séquences *pac* encadrant les unités virales. pUL89 est quant à lui charger de cliver de manière aspécifique l'excès ADN à la fin de l'étape de translocation. L'association avec pUL56 a été montrée comme étant décisive dans l'efficacité du clivage par pUL89. Par ailleurs, le complexe terminase composé par le dimère de pUL56 et le monomère de pUL89 est formé avant sa translocation dans le noyau.

La Figure 57B apporte une hypothèse pouvant rendre compte de l'influence de pUL56 sur l'activité endonucléasique de pUL89. Nous avons superposé les meilleures structures issues des calculs de pUL89-*endo* libre (magenta) et liée (bleue) à l'ADN. Sachant que la

reconnaissance de l'ADN ne puisse s'effectuer qu'en présence de pUL56, nous pouvons en déduire que la structure calculée de pUL89-*endo* liée au duplexe d'ADN (complexe binaire) serait en fait une structure théorique de pUL89-*endo* dans le complexe ternaire pUL89/pUL56/ADN. Au contraire, la forme libre de pUL89-*endo* représenterait l'état conformationnel de celui-ci sans interagir avec pUL56 et donc ne pouvant cliver l'ADN.

Finalement, la superposition des deux structures à la Figure 57B montre une divergence des structures surtout au niveau de l'hélice α4 et de la boucle la précédant. En effet ce fragment du domaine contient les acides aminés $His^{565}$, $His^{571}$ et $Gln^{574}$ dont nous avons évoqué plus haut l'hypothèse de leur rôle dans la stabilisation du brin-substrat (ARN dans notre modèle) pour son clivage ultérieur. En effet, ce fragment subit une rotation de 135° en présence du duplexe.

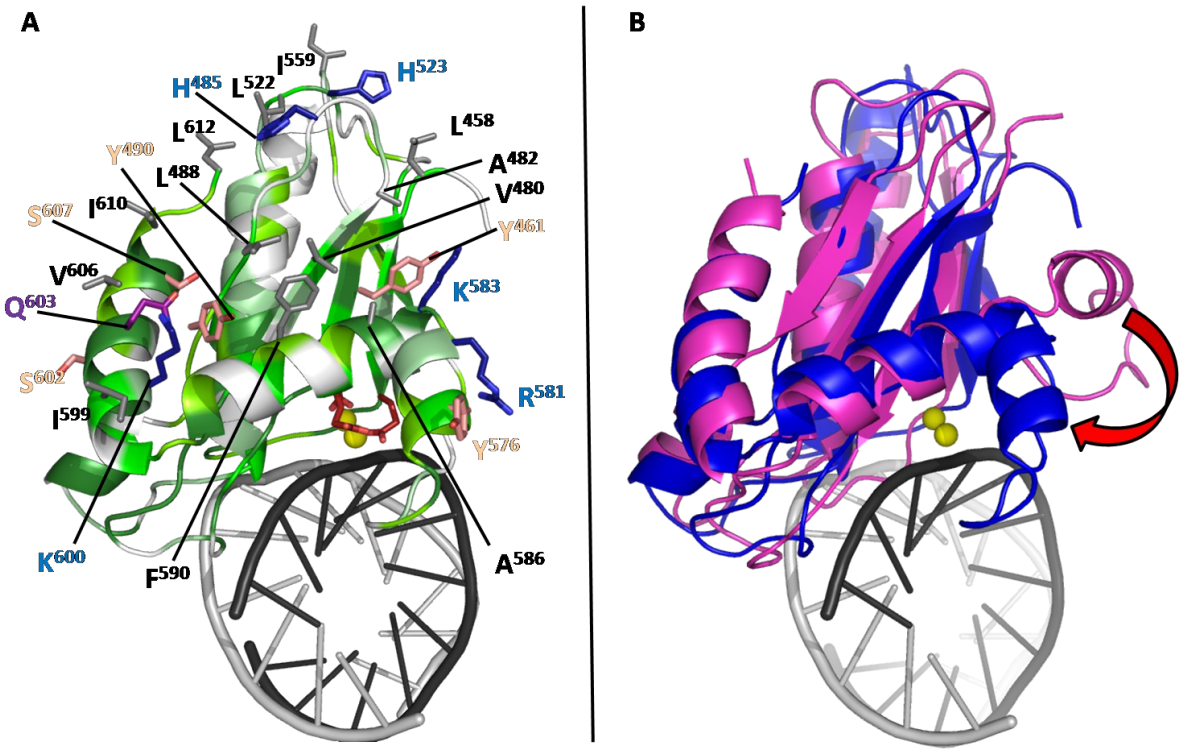



**Figure 57: Surface d'interaction avec pUL56 et impact sur la liaison à l'ADN.** A. Description des différents résidus constituant la surface d'interaction avec pUL56. Le *squelette* est représenté en ruban coloré en fonction de la conservation à travers les homologues herpesviraux de pUL89. Les chaînes latérales des acides aminés potentiellement impliqués dans l'interaction avec pUL56 et la triade catalytique acide sont représentés en sticks et colorés selon leurs propriétés physico-chimique (i.e. rouge, bleu, violet, saumon et

gris pour négativement chargé, positivement chargés, asparagine et glutamine, donneurs de liaisons hydrogène et apolaires, respectivement). Un grand nombre de ces résidus sont hydrophobes. Les brins d'ARN et d'ADN sont représentés en rubans noir et gris, respectivement. B. superposition des meilleurs modèles (selon QMEANclust) calculés pour pUL89-*endo*$^{libre}$ (modèle numéro 77) et pUL89-*endo*$^{liée}$ (modèle numéro 5) colorés en magenta et bleu respectivement. La flèche rouge décrit la rotation d'environ 135° de l'hélice α5 pour la liaison au duplex.

*3.c.iv. Comparaison avec les terminases de bactériophages: validation des hypothèses*

Récemment, des structures de domaines endonucléasiques de terminases ont été résolues par diffraction des rayons X: gp17 des phages T4 et RB49 (codes PDB 3CPE et 3C6A, respectivement) et G2P du phage SPP1 (code PDB 2WBN). La comparaison de notre modèle avec ces structures expérimentales souligne un repliement similaire à celui que nous avons calculé à partir de l'homologie avec des RNases. Malheureusement, MUSTER n'a identifié aucune de ces trois structures lors de sa recherche d'homologues structuraux. Ceci est peut être dû à une mise à jour de la banque de données antérieure à la résolution de l'une de ces trois structures PDB.

Smits et al. (Smits, Chechik et al. 2009) ont résolu la structure de la grande sous-unité terminase, G2P du phage SPP1. Une comparaison de ce domaine avec les structures des sous-unités gp17 issus des bactériophages RB49 et T4 résolues auparavant (Sun, Kondabagil et al. 2008) a été décrite.

Des modèles théoriques des domaines nucléases des protéines gp19 du phage T7 et gpA du phage λ ont été construits à partir de la structure RX de RB49 comme structure-support. Les structures théoriques et expérimentales ont toutes aboutie à la description d'un repliement RNase H-*like* constitué d'un feuillet bêta flanqué d'hélices alpha. Par homologie avec les RNases, ces différentes sous-unités terminases sont supposées chélater deux ions magnésium. Ceci a été vérifié expérimentalement pour G2P.

Les triades acides impliquées dans la chélation des métaux sont variables mais toutes contenues dans des motifs ββα-Me: $Asp^{401}$ $Glu^{458}Asp^{539}$ pour gp17 de T4, $Asp^{398}$ $Glu^{455}Asp^{539}$ pour gp17 de RB49, $Asp^{266}$ $Asp^{321}Asp^{403}$ pour G2P de SPP1. Les séquences de G2P et gp17 sont peu similaires (12% d'identité) mais adoptent des repliements tertiaires identiques.

Proches du repliement des RNases H et des Holliday junction resolvases, elles se différencient de ces classes de protéines par l'existence d'un β hairpin complétant le feuillet

bêta commun aux RNases H. Ce motif supplémentaire a été retrouvé dans les structures de gp17 de RB49 et T4 et est supposé être spécifique des terminases. Son implication dans la liaison à l'ADN a été suggérée. La structure de G2P a été modélisée avec un duplexe d'ADN en conformation B. Le modèle théorique résultant a été superposé avec la structure d'une RNase H humaine complexé avec un duplexe hybride ARN/ADN (Nowotny, Gaidamakov et al. 2007; Nowotny, Cerritelli et al. 2008). L'analyse des facteurs B des cartes de diffraction de la structure 2WBN corrèle avec une flexibilité importante de ce fragment qui est supposé moduler le clivage de l'ADN sous l'effet de la petite sous-unité G1 P. Ce β hairpin est proche de l'histidine 400 et l'aspartate 403 montré par mutagenèse dirigée comme cruciaux pour l'activité nucléase de G2P.

Pour conclure, un repliement analogue aux RNases a été retrouvé dans notre modèle théorique et dans les structures résolues par RX de terminases de bactériophages. De plus, les triades acides identifiées respectent le motif ββα-Me comparable aux observations faites chez les RNases, Holliday junction resolvases et intégrases. Plus spécifiquement aux terminases, un β hairpin a été identifié dans le cas des bactériophages. Ce motif précède directement l'hélice portant le troisième résidu de la triade. Dans pUL89-*endo* ce motif est équivalent à l'hélice α4 et la boucle la précédant (en bleu dans la Figure 58). Il est alors intéressant de remarquer la conjonction des mouvements du β hairpin chez G2P et boucle-hélice chez pUL89-*endo* observés expérimentalement et théoriquement, respectivement. Les conclusions que nous avions émises sur le rôle de ce mouvement dans le clivage de l'ADN concorde avec celles de Smits et al.

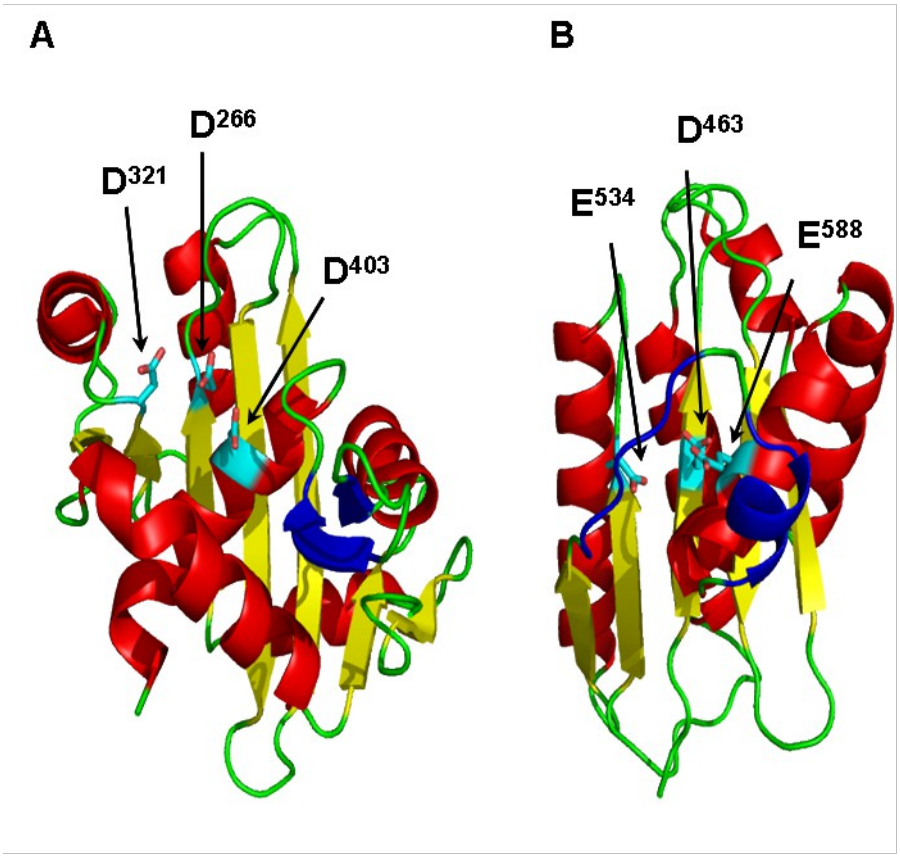



**Figure 58: Comparaison des structures de G2P du phage SPP1 (A) et pUL89-*endo* (B).** Les structures sont colorées en fonction des éléments de structures secondaires (brins β jaunes et hélices α rouges). Les triades acides de chaque protéine sont colorées en cyan et les chaînes latérales représentées en sticks. Le β hairpin de G2P et le motif boucle-hélice α4 de pUL89-*endo* sont colorés en bleu et montrent une orientation similaire par rapport à leur poche catalytique respective.

### C.4. Bilan de l'étude du domaine d'interaction avec pUL56

Partant des observations de Thoma et al., nous avons entrepris l'étude du peptide $^{580}$Gly-Lys$^{600}$ montré comme ayant une grande affinité avec pUL56 et indispensable à l'activité du complexe terminase. Par ailleurs, l'interaction avec pUL56 a été montrée comme modulant l'interaction de pUL89 avec l'ADN.

Les analyses RMN sur ce peptide de 21 acides aminés ont montré une propension de ce fragment à adopter une structure en hélice. Cette hélice, stabilisée en milieu hydrophobe, est amphipathique. La face hydrophobe semble très conservée alors que la face hydrophile ne contient qu'un résidu conservé: Glu$^{588}$.

En vue de pouvoir résoudre le repliement tertiaire du domaine contenant ce peptide, nous avons allongé le système entre $Asp^{568}$-$Leu^{635}$ et résolu sa structure dans l'acétonitrile 40%. Aucun signal de repliement tertiaire n'a été observé. Cependant, nous avons tiré profit des éléments de structures secondaires observés expérimentalement que nous avons introduites comme contraintes expérimentales pour compléter la construction du domaine $Lys^{456}$-$Leu^{612:}$ pUL89-*endo*.

Ce fragment a été révélé par MUSTER comme homologue structural de RNases de type H et d'Holliday junction resolvase. La modélisation de ce domaine par MODELLER 9.6 a finalement permis de souligner la constitution d'une poche de liaison à deux magnésiums par homologie avec les RNases. Cette triade acide est constituée des résidus $Asp^{463}$, $Glu^{534}$ et $Glu^{588}$ appartenant à l'hélice contenue dans pUL89(580-600). Cette première observation est suffisante pour faire le lien entre l'interaction pUL56/pUL89 et activité nucléasique de ce dernier.

Nous avons poursuivi nos investigations en comparant les formes liée et libre de pUL89-*endo*. Les résidus impliqués dans la fixation à l'ADN ont été déduits par l'homologie structurale entre pUL89-*endo* et les RNases. Certains résidus potentiellement impliqués dans la liaison à l'ADN sont situés sur un motif boucle hélice précédant pUL89(580-600).

La comparaison des deux formes de pUL89-*endo* révèle un large mouvement de ce fragment en présence d'ADN. Des observations similaires ont été observées expérimentalement chez les terminases de bactériophages et donc appuient le rôle de ce motif dans la régulation du clivage de l'ADN. Or, la reconnaissance de l'ADN se fait *via* pUL56. Par conséquent, le mouvement du motif boucle-hélice est induit ou son positionnement stabilisé pour interagir avec l'ADN par l'interaction avec pUL56 ce qui explique les résultats observés concernant l'influence de cette interaction sur le rendement catalytique du complexe terminase.

## D. Conclusion Sur L'étude De PUL89

Un alignement de séquence entre pUL89, sous-unité du complexe terminase, et 17 homologues herpesviraux a été effectué. Celui-ci a révélé l'existence de cinq domaines distincts. Un domaine N-terminal riche en résidus basiques dont la fonction est encore peu claire, un domaine hélicase chargé de transloquer le duplexe d'ADN sous l'impulsion de la lyse de molécules d'ATP, un doigt de zinc dont la structure est inconnue mais l'implication dans la reconnaissance de l'ADN est suggérée, un domaine endonucléasique couplée à une

surface d'interaction avec pUL56 et un domaine C-terminal dont la fonction est elle aussi peu claire.

Afin de récolter des informations sur la structure de ces différents domaines, à défaut de pouvoir en résoudre la structure complète par diffraction des rayons X à cause d'un protocole d'expression qui reste encore à optimiser, nous avons couplé recherche bioinformatique et RMN.

Ces études ont permis de modéliser le domaine hélicase de pUL89 grâce à la structure de gp17 du bactériophage T4. Les résidus impliqués dans la liaison à l'ATP et à l'ADN ont étaient suggérés par comparaison de notre modèle avec des structures expérimentales et en prenant en considération la répartition croisée des caractères physicochimiques et des taux de conservation des résidus présents à la surface du modèle généré. Un mécanisme de résistance aux dérivés D-benzimidazolés a été proposé. L'hypothèse retenue consiste en la déstabilisation de la reconnaissance de l'ADN en présence de ces molécules peut être capable de s'intercaler dans les duplexes d'ADN du fait de leur grande similarité structurale avec des nucléotides. Les mutations de résistance montrent une infime modification de la géométrie et non des fonctions chimiques des acides aminés mutés. Nous suggérons que ces mutations créent des encombrements stériques empêchant les inhibiteurs à s'intercaler au niveau des plateaux de bases de l'ADN.

Nous avons ensuite étudié le domaine endonucléasique. Ce travail a débuté par l'étude du fragment pUL89(580-600) indispensable à l'interaction avec pUL56. Le système a été étendu de part et d'autre pour atteindre une taille de 68 acides aminés. Le repliement tertiaire de ce domaine rallongé n'a pas pu être observé mais quatre hélices ont été identifiées.

La construction du domaine pUL89-*endo* à partir d'une homologie avec des RNases a permis d'expliquer comment pUL89(580-600) interagit avec pUL56 *via* une face hydrophobe et assure la liaison aux magnésium catalytique *via* son résidu conservé $Glu^{588}$. La comparaison des structures liée et libre avec un duplexe hybride ADN/ARN en conformation B permet d'expliquer l'influence de pUL56 sur le clivage de l'ADN par pUL89 par réarrangement structurale.

Ce travail théorique reste à être confirmé et complété par la résolution de la protéine entière si elle est exprimable en grande quantité et cristallisable pour des études structurales par rayon X (compte tenu de la taille considérable du système, i.e. 674 acides aminés). Si cette possibilité n'est pas envisageable, la synthèse chimique (ou l'expression) et l'étude

RMN domaine par domaine peuvent constituer une alternative. Cependant, même fragmentés, les différents domaines de pUL89 sont en moyenne long de 150 résidus ce qui est un facteur limitant pour la synthèse peptidique en phase solide et incite à l'enrichissement en isotopes stables des échantillons ($^{15}N$, $^{13}C$) pour aider à l'attribution des signaux RMN et donc à la reconstruction des structures. L'étude des mutants D344E et A355T doit aussi être envisagé pour élucider le mécanisme d'action des dérivés D-ribobenzimidazolés.

# VIII. ÉTUDES DE LA PROTÉINE PUL56

Ce chapitre traite de pUL56 dont les données structurales sont actuellement quasi inexistantes dans la littérature. Nous avons commencé à étudier les aspects phylogéniques en nous basant sur les alignements de séquences. Des études de spectroscopie (UV/visible, dichroïsme circulaire et RMN) ont été initiées pour pUL56, cible thérapeutique révélé par traitements aux dérivés D-ribobenzimidazolés.

La protéine pUL56 constitue la grande sous unité du complexe terminase de CMVH avec une séquence de 850 acides aminés. Sa structure observée en microscopie électronique est en forme de selle à cheval composée de deux chaînes (Scheffczik, Savva et al. 2002; Savva, Holzenburg et al. 2004). La protéine pUL56 interagit avec la petite sous-unité pUL89 et pUL104, la protéine portail. Montrée comme portant une activité ATPasique, cette protéine remplit un rôle particulièrement important dans la reconnaissance de l'ADN, au niveau des séquences *pac* délimitant les unités d'ADN viral, après synthèse en cercle roulant de l'ADN concatémérique.

Dans cette partie, nous allons présenter les travaux réalisés sur deux fragments de pUL56 pouvant avoir un rôle dans la reconnaissance de l'ADN: un domaine de type *basic leucine zipper* (*bZIP*) identifié par PHYRE le motif en doigt de zinc .

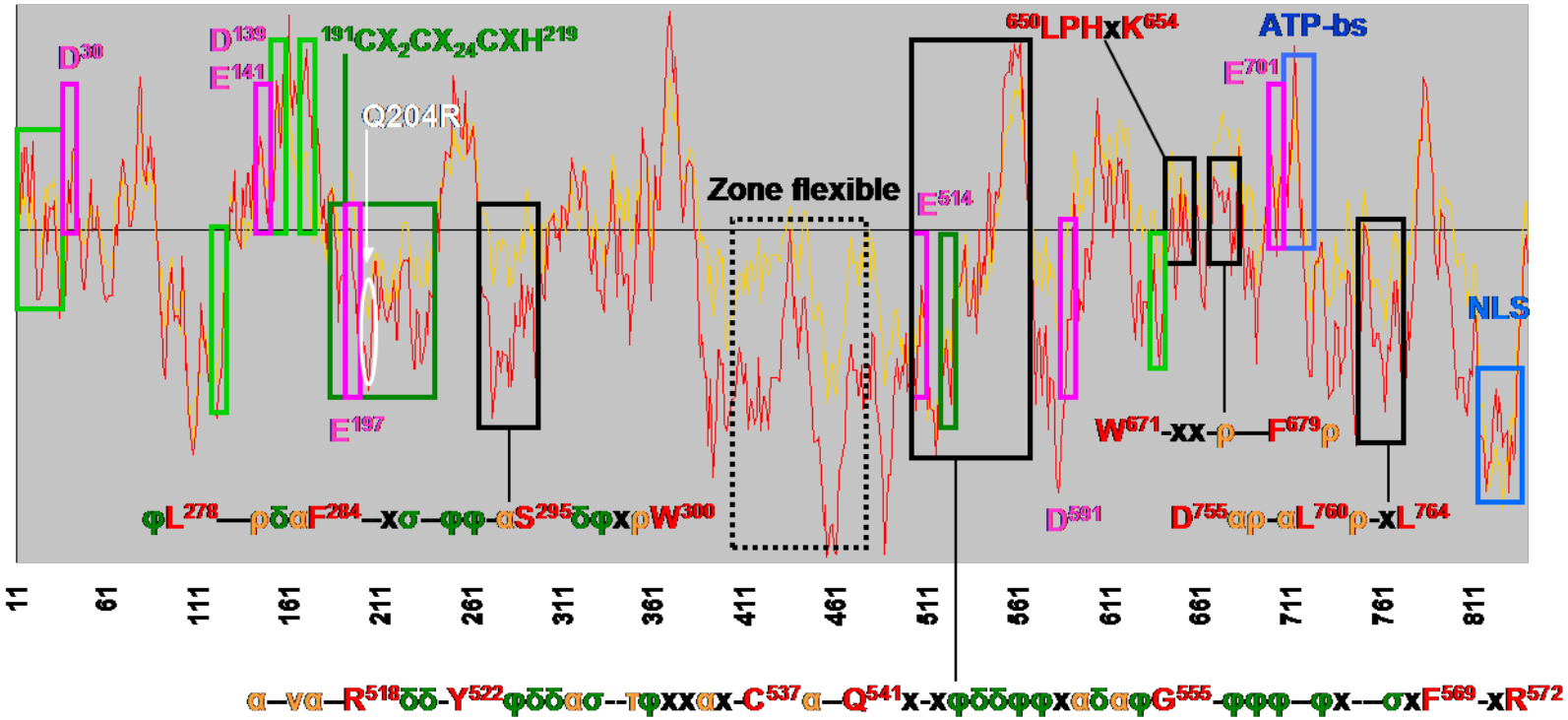



**Figure 59: Récapitulatif des régions conservées à travers les herpesvirus sur le graphique d'hydrophobicité de pUL56.** Deux échelles d'hydrophobicité ont été utilisées, Kyte & Doolittle et Eisenberg colorées en rouge et jaune, respectivement. Différentes boîtes très conservées ont été précisées et la trace correspondante rapportée (cf. Annexe B2). Les fragments identifiés expérimentalement sont représentés en bleu (site de liaison à l'ATP et signal de localisation nucléaire notés ATP-bs et NLS respectivement). Les cystéines classe-spécifiques et complètement conservées sont colorées en vert clair et vert foncé, respectivement. En particulier, trois d'entre elles sont totalement conservées et appartiennent au doigt de zinc potentiel. Une autre appartient au fragment suivant directement la zone flexible peu conservée (pointillées). Les résidus acides totalement et hautement conservés sont colorés en magenta.

## A. Alignement De Séquences

L'alignement de séquences entre pUL56 et 22 homologues herpesviraux d'origines humaine et animale est présenté en Annexe B2 et a été fait l'objet d'une publication (Champier, Couvreux et al. 2008).

Nous avons récapitulé dans la Figure 59 les différents éléments remarquables issus de cet alignement et les boîtes caractéristiques identifiées expérimentalement: le signal de localisation nucléaire (NLS) indispensable au passage de pUL56 dans le noyau par la voie de l'importine α qui entraîne dans son sillage pUL89 (Giesen, Radsak et al. 2000) et le site de liaison à l'ATP (Hwang and Bogner 2002). Nous avons aussi indiqué quelques boîtes (la trace pour chacune est indiquée et extraite de l'Annexe B2) qui semblent être très conservées à travers les homologues herpesviraux de pUL56.

Afin de proposer des pistes pour l'identification ultérieure par mutagenèse dirigée (par exemple) des résidus importants à l'activité endonucléasique, nous avons relevé huit résidus acides conservés à travers tous les homologues de pUL56 alignés (encadrés en magenta en Figure 59). Quatre d'entre eux sont situés dans des régions conservées: $Glu^{197}$ situé dans le doigt de zinc potentiel, $Glu^{514}$ après la région flexible, $Glu^{701}$ adjacent à l'ATP-bs et $Asp^{755}$.

La protéine pUL56 existant sous forme de dimère (Savva, Holzenburg et al. 2004), il nous a semblé intéressant d'annoter les différentes cystéines conservées (surlignées en noir en Annexe B2) soit uniquement chez les bêta herpesvirus soit complètement conservées à travers tous les herpesvirus notées en vert clair et vert foncé, respectivement, en Figure 59. Il apparaît donc que le N-terminal contient une région riche en cystéines conservées chez les homologues de bêta herpesvirus ($Cys^{10}$, $Cys^{13}$, $Cys^{16}$, $Cys^{25}$ et $Cys^{29}$). Cinq autres cystéines sont disséminées dans la séquence dont quatre ($Cys^{115}$, $Cys^{145}$, $Cys^{163}$ et $Cys^{174}$) précédant le doigt de zinc $CX_2CX_{24}CXH$ parfaitement conservé proposé par Krosky (Krosky, Underwood et al. 1998). Une autre cystéine parfaitement conservée est présente à la position 537 et appartient à une région que nous avons nommée domaine *bZIP* (*basic leucine zipper)* identifié par homologie avec les protéines CREB-*like* domaine (c-AMP Responsive Element Binding protein).

La reconnaissance spécifique de l'ADN viral au niveau des séquences *pac* étant déterminante pour sa maturation et son encapsidation, nous nous sommes intéressés à l'étude

de deux domaines pouvant potentiellement remplir cette fonction: le doigt de zinc contenu dans le fragment $^{186}$Val-Ile$^{233}$ et le domaine $^{506}$Arg-Ile$^{548}$ de type *bZIP*.

## B. Doigt De Zinc De PUL56 De HCMV

L'expression de ce peptide s'étant avéré difficile, la séquence $^{186}$Val-Ile$^{233}$ (pUL56-ZF) contenant le doigt de zinc a été synthétisée par synthèse peptidique en phase solide au laboratoire selon une stratégie Fmoc et sa masse de 5308,04 Da vérifiée par spectroscopie de masse MALDI-TOF.

L'étude du doigt de zinc a été réalisée en utilisant plusieurs spectroscopies. En UV/visible nous avons tenté de vérifier la propension du peptide à lier le métal, en dichroïsme circulaire nous avons voulu observer l'effet structurant de l'ajout de zinc, et en RMN nous avons enregistré des spectres TOCSY et NOESY afin d'en calculer la structure.

La spectroscopie UV/visible est une technique de choix pour l'observation de doigt de zinc puisqu'elle nécessite peu d'échantillon (concentration micromolaire), elle est non destructive et permet d'observer les transitions spécifiques sous l'effet de l'ajout de métal. Ainsi, un grand nombre d'études portant sur les doigts de zinc ont été réalisées (Burke, Sanyal et al. 1992; Worthington, Amann et al. 1996; Bavoso, Ostuni et al. 1998; Westman, Perdomo et al. 2004). Les spectres ont été enregistrés sur un spectromètre Beckman-DU 530 entre 250 et 800 nm. Dans une cuve en quartz de 1 cm de large, 150µM de pUL56-ZF sont solubilisés dans 170 µM de DTT en tampon phosphate 7,5 préalablement dégazé sous argon. La titration avec $Co^{2+}$ a été réalisée par ajout de 30-300 µM de $CoCl_2$. Le spectre sans ajout de $CoCl_2$ a été soustrait aux spectres obtenus par incrémentation de $CoCl_2$. Le doigt de zinc de Ncp7(1-30) (Morellet, Jullian et al. 1992) a été analysé dans les mêmes conditions pour pouvoir valider le protocole.

Les résultats présentés en Figure 60A révèlent que pUL56-ZF lie effectivement le métal. On observe les transitions internes au métal sous l'effet Jhan-Teller autour de 650 nm et les dons de paires d'électrons des ligands au métal autour de 300 nm. En comparant les courbes obtenues avec le premier doigt de zinc de Ncp7(1-30) (Figure 60A, droite), nous pouvons juger de l'instabilité du peptide. Les intensités des transitions observées pour pUL56-ZF augmentent de façon régulière jusqu'à environ 0,5 équivalent. Au-delà de cette valeur, les raies d'absorption vers 300 nm deviennent moins bien définies.

Les courbes de dichroïsme circulaire (Figure 60B) ont été enregistrées avec un spectromètre JobinYvon-CD6 sur une solution de 80 µM de pUL56-ZF à pH 7,5 et

incrémentée avec une solution de $ZnCl_2$ (16-140µM). On peut observer l'influence du métal induisant une légère structuration en hélice du peptide par l'apparition de signaux à 214 et 222nm.

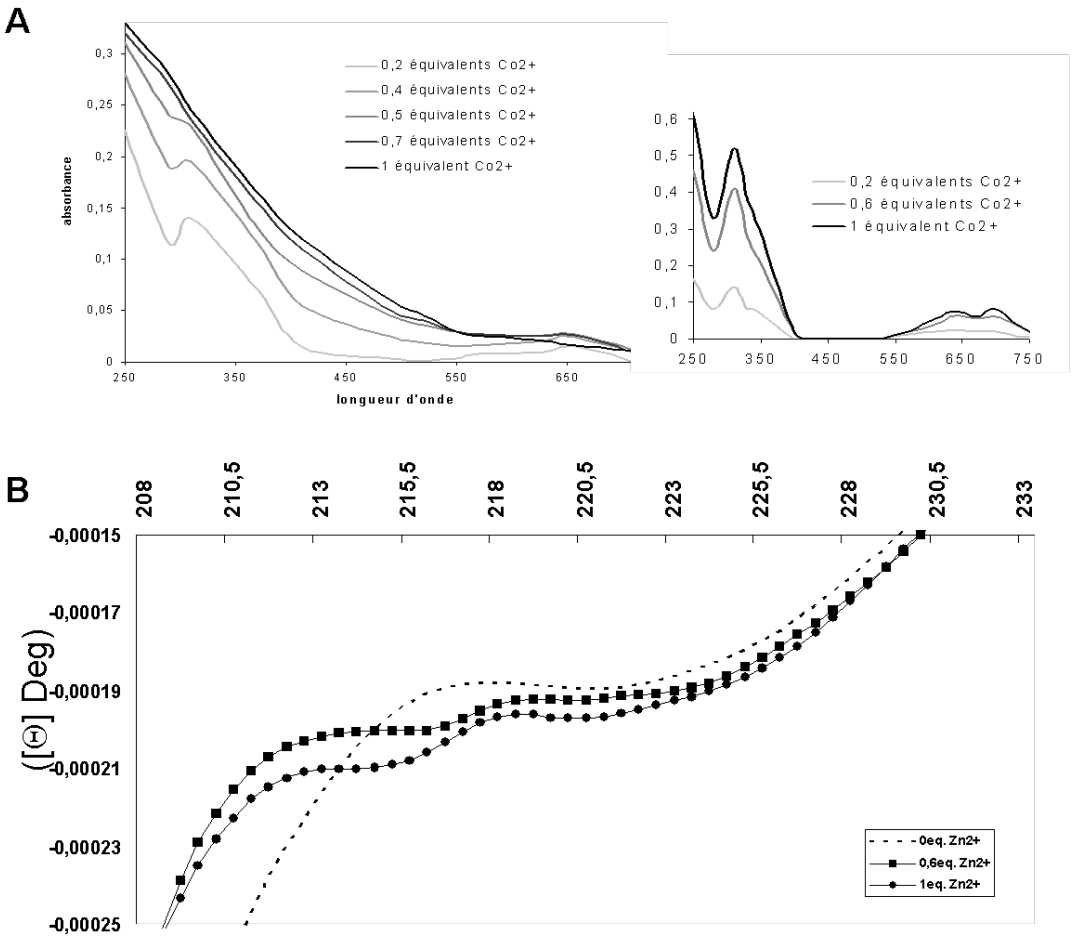



**Figure 60: Analyse de pUL56-ZF par spectroscopie UV/visible et dichroïsme circulaire.** A. Une solution de 150µM de pUL56-ZF en tampon phosphate 7,5 et en présence de DTT a été incrémenté en $CoCl_2$ (gauche). La même expérience avec Ncp7(1-30) du VIH a été effectuée (droite) pour valider notre protocole expérimentale. B. Zoom des courbes de dichroïsme circulaire dans la région 208-235nm obtenu pour puL56-ZF incrémenté en $ZnCl_2$. Les courbes obtenues pour 0, 0,6 et 1 équivalent sont représentées en pointillés, carrés et disques noirs, respectivement.

Nous avons par la suite tenté de résoudre la structure de ce peptide (530 µM, 1,5mM $ZnCl_2$) par RMN. Sa faible solubilité au-delà de pH 5,5 a limité son étude au-delà de cette valeur. Les spectres ont été enregistrés avec et sans zinc à pH 5,5 mais n'ont révélé aucune

différence. La formation du doigt de zinc n'a donc pas été observée. Ceci a sûrement été défavorisé par le pH inférieur au pKa des hydrogènes portés par les fonctions thiol de la cystéine (8,5) et imidazole de l'histidine (6,5), dans lequel le peptide a été solubilisé.

Dû aux difficultés de mise en solution pour les expériences RMN, les spectroscopies optiques ont quand même permis d'observer une influence du métal sur le repliement de pUL56-ZF. En particulier, les transitions observées en UV/visible prouvent, sans ambigüité, la propension des résidus cystéines et histidine à chélater le métal jusqu'à 0,5 équivalent, valeur après laquelle le signal devient plus difficile à interpréter. Ces larges bandes d'absorption traduisent selon nous une instabilité du doigt de zinc pouvant être expliquée de différentes façons.

La Figure 61 montre les différentes possibilités pouvant rendre compte des proportions de 0,5 équivalents observés en UV/visible.

Les deux premières possibilités (Figure 61A et B) sont la chélation d'un métal par deux peptides mettant chacun en jeu deux ligands.

En B nous avons représenté la formation éventuelle de ponts disulfures intermoléculaire après oxydation des thiols (traits rouges) pouvant expliquer la stabilité d'un tel dimère. Cependant, nous n'avons trouvé aucun exemple dans la littérature similaire à l'un ou l'autre de ces cas.

En C, l'instabilité du doigt de zinc est expliquée par l'entropie élevée de la boucle de 24 acides aminés qui diminue statistiquement le rapprochement des quatre ligands potentiels pour la chélation du zinc.

Enfin, en D, nous mettons en jeu la présence de $His^{223}$ non conservée pouvant entrer en compétition avec $His^{219}$.

La formation de ponts disulfures a été évitée par dégazage et ajout de DTT. De surcroît, les expériences avec Ncp7 sont cohérentes avec la structure en doigt de zinc. Par conséquent, notre protocole expérimental semble être adapté à la prévention de la formation de ponts disulfures. Donc, la possibilité présentée en B parait peu probable.

Afin de vérifier la possibilité C, il serait intéressant de rallonger le peptide pouvant faire apparaître des éléments de structures secondaires supplémentaires afin de faciliter la proximité des résidu-ligands.

Pour vérifier la possibilité D, impliquant un équilibre entre une forme mettant en jeu $His^{219}$ et une seconde impliquant $His^{223}$, nous avons synthétisé le peptide homologue $^{186}Val$-$Ile^{233}$

d'origine murine ne contenant pas d'histidine autre que celle alignée avec His[219] de pUL89 du CMVH. Celui-ci a été obtenu par synthèse peptidique et seul quelques milligrammes ont pu être purifiés. Ce peptide s'est finalement trouvé très difficile à solubiliser y compris à pH acide. De fait, nous n'avons pu vérifier notre hypothèse

Ce fragment de pUL56 semble avoir une importance dans le mécanisme d'encapsidation du fait de sa propension à former un doigt de zinc et donc à son potentiel à interagir avec l'ADN. Comme dans le cas de pUL89, où nous avions localisé les positions mutées (Asp[344] et Ala[355]) proches de la surface d'interaction avec l'ADN et du doigt de zinc potentiel $CX_2HX_{12}CXC$, la mutation de résistance aux dérivés D-ribobenzimidazolés observée sur pUL56 est située au centre de la boucle entre Cys[194] et Cys[217] (Q204R). Ce résidu est encadré par des résidus parfaitement conservés (Asn[203] et Gly[205]) ayant un rôle potentiel dans une interaction avec l'ADN.

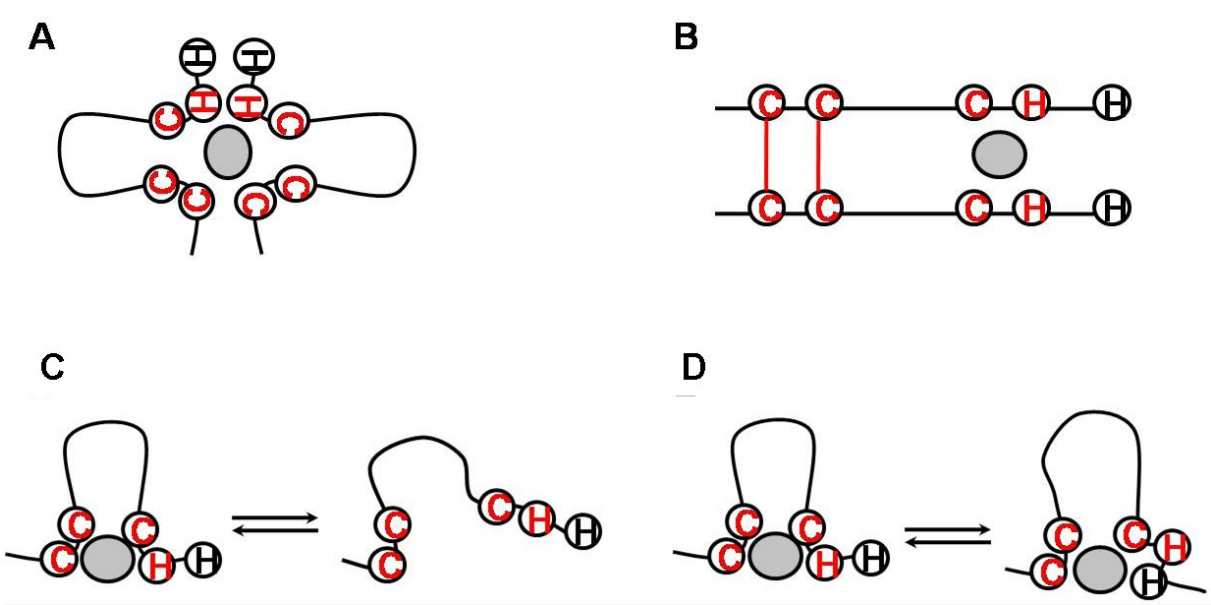



**Figure 61: Cas expliquant une déstabilisation du signal observé en UV/visible au-delà de 0,5 équivalents.** Le disque gris schématise l'atome de zinc, les résidus conservés (Cys[191], Cys[194], Cys[217], His[219]) sont annotés en rouge et l'histidine non conservée (His[223]) en noir. A et B montrent la possibilité de dimérisation à travers la chélation du zinc. C et D illustrent des échanges conformationnels entre deux formes.

Après avoir optimisé sa synthèse chimique ou son expression en système bactérien, il serait intéressant de poursuivre les investigations sur la structuration de ce doigt de zinc rallongé de part et d'autre de sa séquence pour faire apparaître des éléments de structures

secondaires supplémentaires pouvant stabiliser son repliement tertiaire et éventuellement faciliter sa solubilisation. Une fois ces conditions remplies, son étude en interaction, si elle existe, avec les séquences *pac* spécifiques et l'étude du mutant Q204R en présence de BDCRB offriraient une meilleure compréhension des mécanismes de reconnaissance de l'ADN et d'action des dérivés D-ribobenzimidazolés.

## C. Étude Structurale De PUL56(507-548): BZIP Domaine

Nous avons synthétisé au laboratoire un autre peptide de pUL56: pUL56(507-548).

Celui-ci a été sélectionné empiriquement afin de mettre en évidence une interaction entre ce peptide et pUL89(580-600). Nous nous sommes basés sur les différentes données statistiques disponibles sur les interfaces protéine-protéine (composition en acides aminés, appariements de structures secondaires, conservation) et sur les données propres au système pUL56/pUL89.

Nous avons de cette façon sélectionné un fragment composé de résidus conservés (Hu, Ma et al. 2000) typiques des interfaces (arginines et aromatiques) (Lo Conte, Chothia et al. 1999; Ofran and Rost 2003; Moreira, Fernandes et al. 2007) et prédit en hélice du fait des appariements préférentiels entre éléments de structures secondaire de même nature (Jones and Thornton 1996; Ansari and Helms 2005). Ce dernier critère a été orienté par les résultats expérimentaux obtenus par RMN pour pUL89(580-600) montrant une structure en hélice alpha de ce peptide. Par ailleurs, les tests sur l'activité ATPasique de pUL56 ont été effectués par Scheffczick et al. sur le fragment pUL56(404-850) en présence de pUL89. Il a été prouvé que pUL89 influence l'activité ATPasique de ce fragment de pUL56 (Scheffczik, Savva et al. 2002). Par conséquent, pUL89 interagit avec un fragment de pUL56 situé dans la région 404-850.

En entrecoupant ces différentes informations statistiques et expérimentales, nous avons sélectionné deux peptides de pUL56: $Arg^{506}$-$Ile^{548}$ et $Glu^{663}$-$Ala^{720}$. Ces deux peptides ont été synthétisés par SPPS. pUL56(663-720) a été produit en très faible quantité à cause d'une séquence peptidique défavorable aux couplages entre amino acides successifs (<5mg). Par ailleurs, il s'est montré très peu soluble dans l'eau, l'acétonitrile ou le TFE ne permettant pas son étude, même à faible concentration. Nous nous sommes donc focalisé sur l'étude de pUL56(506-548).

### C.1. Dichroïsme circulaire

Le peptide pUL56(506-548) a été purifié en grande quantité (120mg) et plusieurs conditions de mise en solution ont été testées afin de stabiliser son repliement ont été étudiées par dichroïsme circulaire (Figure 62).

Dans l'eau, le peptide est majoritairement non structuré.

Les effets du TFE et du méthanol à 10% sont sensiblement les mêmes et entraînent une légère structuration en hélice.

30% de méthanol ou un mélange 10% MeOH/10% TFE/80% H2O augmentent la proportion d'hélice mais le peptide reste majoritairement *random coil*.

Finalement, 30% de TFE stabilisent fortement la structuration en hélice en accord avec la prédiction établit par PSI-PRED (Annexe B2).

Nous avons donc choisi ces dernières conditions pour faire l'étude de notre peptide par RMN.

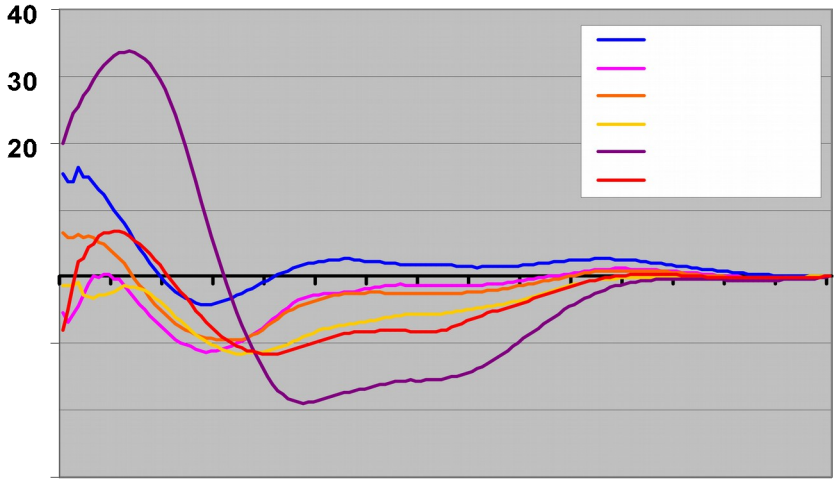



**Figure 62: Dichroïsme circulaire dans différents solvants sur le peptide pUL56(506-549).** Pour chaque condition, trois scans moyennés et lissés sur 25 points ont été enregistrées entre 185 nm et 260 nm par pas de 0,5 nm avec une bande passante de 2 nm et un temps d'intégration de 2s.

### C.2. Cross-linking

Le but premier dans l'étude de ce peptide fut de mettre en évidence une interaction éventuelle entre celui-ci et pUL89(580-600). Pour ce faire, nous avons procédé à des expériences de cross-linking selon le protocole suivant (Back, de Jong et al. 2003; Sinz 2006):

1. Pour la préparation des solutions mères, les peptides ont été mis en solution dans H2O milliQ à une concentration de 1 mM déterminée par la mesure de l'absorbance à 280 nm (équation de Beer-Lambert). pUL56(506-549) a été mis en solution dans 5 équivalents de DTT pour empêcher la formation d'un pont disulfure *via* sa cystéine 536.
2. Le nombre d'amines primaires (chaînes latérales des lysines et N-ter) susceptibles de réagir avec l'agent de réticulation a été calculé afin de déterminer la quantité de cross-linker à ajouter en solution.
3. Dans le milieu réactionnel, la concentration de chaque peptide a été fixée à 50 µM et le volume total ajusté à 350µL avec une solution de tampon phosphate 100 mM, pH 7,0.
4. Les expériences de réticulations des peptides seuls et mélangés ont été conduites en parallèles.
5. Une solution mère de $BS^3$ (Bis-sulfosuccinimidyl) suberate, Figure 63) a été préparée. Le temps de demi-vie d'une solution de $BS^3$ dans l'eau est de 1 heure. Nous avons donc préparé cette solution au dernier moment afin d'éviter son hydrolyse. La quantité de $BS^3$ ajoutée dans chaque milieu réactionnel a été calculée pour un ratio [$BS^3$]/ [amines libres total en solution] = 0,5 car le $BS^3$ est bifonctionnel et donc réagit avec 2 amines primaires pour former des liaisons peptidiques avec elles (Figure 63).
6. Les réactions ont été contrôlées sous agitation faible, à température ambiante, pendant 3 à 5 minutes puis stoppées par ajout de 350 µL de tampon Tris 100 mM, pH 7,5.
7. Les différentes expériences ont été analysées sur gel d'acrylamide (Tricine 16% adapté aux faibles poids moléculaires).

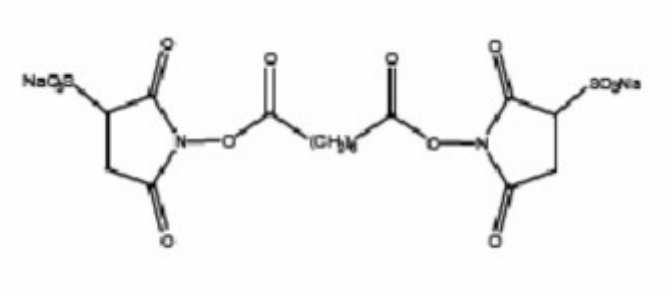


**Figure 63: Formule du Bis-sulfosuccinimidyl) suberate.**

Nous avons mis en présence les deux peptides dans l'eau d'une part (Figure 64, puit 7), et dans 30% de TFE d'autre part (Figure 64, puit 6). En comparant les résultats obtenus pour les peptides étudiés isolement (Figure 64, puits 1-5 et 8), nous sommes arrivés à la conclusion que, dans les conditions d'expériences, les deux peptides n'interagissent pas.

Trois hypothèses peuvent expliquer ces résultats. Premièrement, dans l'eau, ni pUL56(506-548) ni pUL89(580-600) ne sont structurés de façàn stable. Ce défaut de structuration peut empêcher l'interaction. Deuxièmement, dans le TFE, les deux peptides sont structurés mais le solvant peut faire obstacle aux interactions hydrophobes et par conséquent empêcher lui aussi l'interaction entre les deux peptides. Troisièmement, le domaine pUL56(506-548) n'est pas impliqué dans l'interaction avec pUL89.

Par contre, aux concentrations utilisées dans cette expérience, nous pouvons observer clairement la formation d'un dimère de pUL56(506-548) dans le tampon phosphate (Figure 64, puits 1 et 5). Au contraire, pUL89(580-600) n'interagit ni avec lui-même (Figure 64, puit 3) ni avec pUL56(506-548) quel que soit le solvant (Figure 64, puits 6 et 7).

Notons tout de même que le dimère de pUL56(506-548) n'est pas observé, ou très faiblement, dans les mélanges de peptide. Dans le TFE, nous pouvons penser que le solvant empêche la formation du dimère nécessitant l'établissement d'interactions hydrophobes (puit 6). Dans le tampon, l'explication la plus évidente serait que bien que n'interagissant pas avec pUL89(580-600) ce dernier semble toutefois interférer avec la formation du dimère de pUL56(506-548) (puit 7).

Au vue des résultats de cross-linking obtenus sur les peptides seuls, nous pouvons affirmer que pUL56(506-548) a la capacité de dimériser. Nous avons donc poursuivi les investigations pour calculer la structure de ce dimère par RMN.

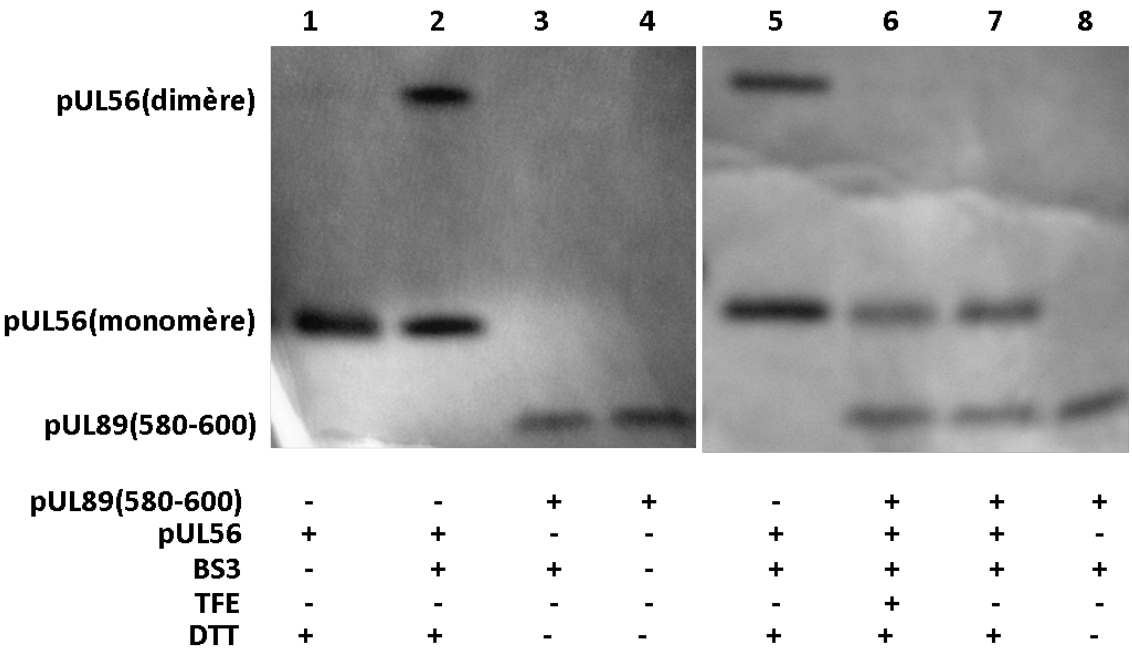



**Figure 64: Expériences de cross-linking entre pUL56(506-548) et pUL89(580-600).** Chaque peptide a été introduit avec une concentration de 50 µM. 0,5 équivalent de $BS^3$ par rapport au nombre d'amines primaires (chaîne latérale des lysines et N-terminals) présentes en solution est introduit (2, 3,5-8). En 1 et 4, les peptides sont seuls en solutions. En 2, 3,5 et 8, chaque peptide isolé est mis en présence de $BS^3$. Le mélange de peptides en présence de $BS^3$ est solubilisé dans 30% TFE (puit 6) et dans le tampon seul (puit 7). En 6, une légère bande correspondante au dimère de pUL56(506-548) est présente mais de faible intensité. Les différentes conditions sont précisées sous chaque puit.

### C.3. pUL56(506-549): un peptide de type *bZIP* ?

Afin de pouvoir proposer un rôle à pUL56(506-549), nous avons soumis sa séquence à PHYRE. Celui-ci a identifié une similitude avec la structure PDB 1DH3 définie comme un domaine CREB (*cAMP response element-binding protein*) (Schumacher, Goodman et al. 2000), avec un taux de confiance de 40% bien que le taux d'identité entre pUL56(506-549) et cette protéine soit seulement de 9%.

Les protéines de type CREB sont caractérisées par une capacité à dimériser *via* un leucine zipper parallèle (structure type *coil coiled*) et, un site d'interaction spécifique avec l'ADN constitué de résidus basiques d'où la notation associée *bZIP* pour *basic leucine zipper*. Ces protéines sont impliquées dans des mécanismes de régulation de la transcription (Konig and Richmond 1993; Schumacher, Goodman et al. 2000). Ces deux caractéristiques, dimérisation et interaction avec l'ADN, sont en accord avec les observations faites dans le cas de pUL56.

Nous avons aligné pUL56(506-548) avec trois protéines CREB présentes dans la PDB et observé de grandes similitudes entre elles notamment par une composition en résidus basiques (triade $^{517}$RKR$^{519}$) et hydrophobes (Val$^{510}$, Val$^{514}$, Tyr$^{521}$, Leu$^{522}$, Val$^{525}$, Val$^{533}$, Ile$^{537}$) (Figure 65).

En particulier, la cystéine parfaitement conservée à travers les homologues herpesviraux de pUL89 est alignée avec des leucines impliquées dans la dimérisation des protéines CREB. Nous formulons l'hypothèse que la conservation de ce résidu est justifiée par son implication dans la stabilisation du dimère de pUL56.

```
                    510        520        530        540      548
                     |          |          |          |        |
ET      -------------α--να--Rδ-δ-Yφδδασ--τφxxαx-Cα--Qx-xφδδφφ-------
BETA    ---------δ-Hδα-δEαδαRK-πAYφδKαSE-GYσπVφ-CIπ-QE-LφδKLφ-------
pUL56   ---------RIHNVTREVNVRK-RAYLQKVSEVGYGKVIRCIKTQERLTSKLI-------

1DGC    IVPESSDPAALKRARNTEAARRSRARKLQRMKQLEDKVEELLSKNYHLENEVARLKKLVG
1DH3    -------KREVRLMKNREAARECRRKKKEYVKCLENRVAVLENQNKTLIEELKALKDLYC
1NWQ    ---GNSNEYRVRRERNNIAVRKSRDKAKQRNVETQQKVLELTSDNDRLRKRVEQLSRELD
```

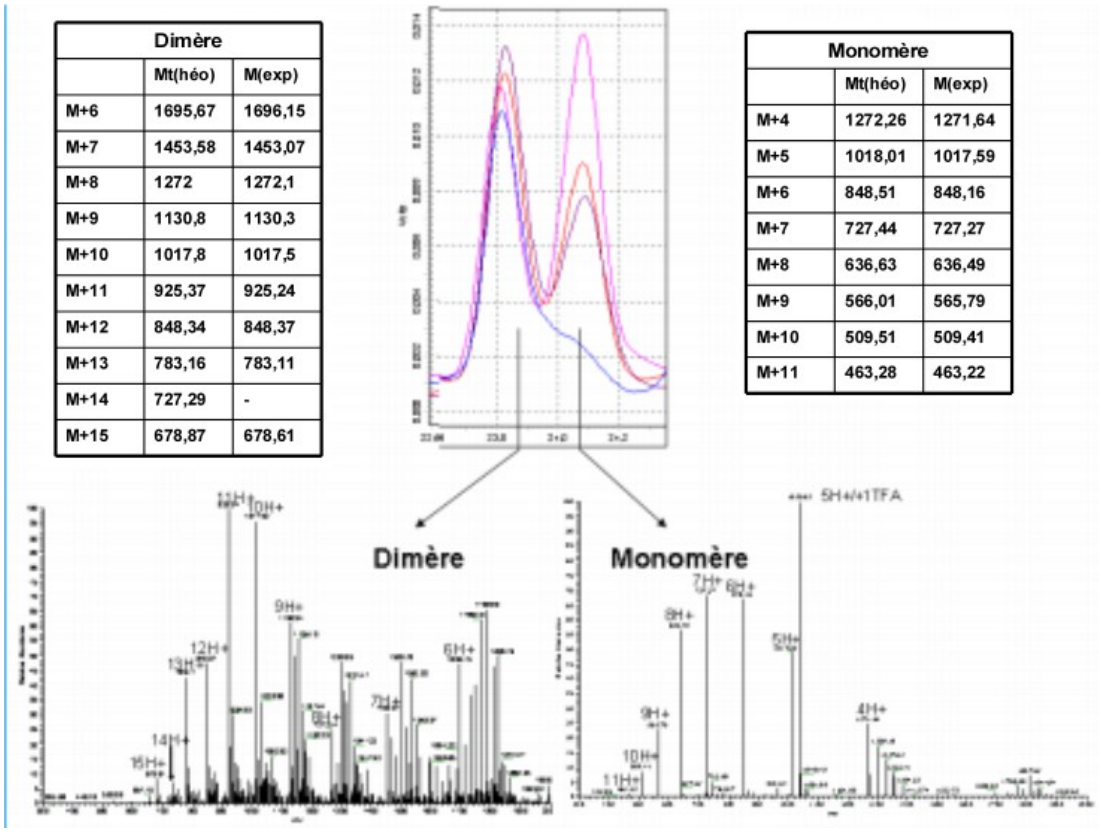
**Figure 65: Alignement de pUL56(506-548) avec 3 protéines de type CREB.** Les première et deuxième lignes indiquent la trace de l'évolution chez l'ensemble des homologues herpesviraux de pUL56 et la trace propre aux homologues bêta herpesviraux, respectivement. Les résidus hydrophobes et basiques alignés avec pUL56(506-548) sont colorés en blancs et jaunes, respectivement. Les trois séquences de protéines CREB sont issues de GCN4_YEAST (accession number: P03069, PDB 1DGC), CREB1_MOUSE (accession number: Q01147, PDB 1DH3), et CREBPA_RAT (accession number: P05554, PDB 1NWQ).

### C.4. Étude de la dimérisation de pUL56(506-548) par HPLC et RMN

Les hypothèses sur la formation d'un dimère et d'une structure de type *bZIP* issue de la recherche d'homologues par PHYRE nous ont incité à déterminer la structure de ce peptide par RMN. Dans ce but, nous avons contrôlé la formation du pont disulfure par HPLC en phase inverse avec une colonne greffée C4.

La forme dimérique est éluée près d'une minute avant la forme monomérique comme le montre la Figure 66. Ainsi, du fait d'une taille deux fois plus importante qui impliquerait

intuitivement un temps d'élution plus grand, seule une surface hydrophobe permet d'expliquer une élution plus rapide que celle du monomère. Ceci permet d'émettre l'hypothèse que la dimérisation se fait à travers une interface hydrophobe complétée par le pont disulfure dont nous avons contrôlé la formation.




| Dimère | | |
|---|---|---|
| | Mt(héo) | M(exp) |
| M+6 | 1695,67 | 1696,15 |
| M+7 | 1453,58 | 1453,07 |
| M+8 | 1272 | 1272,1 |
| M+9 | 1130,8 | 1130,3 |
| M+10 | 1017,8 | 1017,5 |
| M+11 | 925,37 | 925,24 |
| M+12 | 848,34 | 848,37 |
| M+13 | 783,16 | 783,11 |
| M+14 | 727,29 | - |
| M+15 | 678,87 | 678,61 |

| Monomère | | |
|---|---|---|
| | Mt(héo) | M(exp) |
| M+4 | 1272,26 | 1271,64 |
| M+5 | 1018,01 | 1017,59 |
| M+6 | 848,51 | 848,16 |
| M+7 | 727,44 | 727,27 |
| M+8 | 636,63 | 636,49 |
| M+9 | 566,01 | 565,79 |
| M+10 | 509,51 | 509,41 |
| M+11 | 463,28 | 463,22 |

**Figure 66: Suivi de la formation du pont disulfure par HPLC et spectrométrie de masse ESI.** L'empilement des chromatogrammes montre le passage de la forme monomérique (15,15 minutes) à la forme dimérique (15,65 minutes). Les masses expérimentales (M(exp)) des spectres de chaque espèce ont été attribués et vérifient les masses théoriques (M(théo)) des formes mono- et dimérique à leurs différents états de charges (Tableaux).

Les peptides avec et sans pont disulfure ont finalement été analysés dans 30% de TFE à 303 K par RMN à pH 6,1 à différents temps de mélange (50 ms, 100 ms, 150 ms, 200 ms et 250 ms). L'attribution des déplacements chimiques a été réalisée dans les deux cas (Annexe D4) et peu de différences ont été relevées, traduisant un environnement similaire des acides aminés dans les deux conditions.

Le jeu de contraintes de structures secondaires présenté à la Figure 67 a été utilisé pour reconstruire le dimère.

L'ensemble des signaux recueillis permet d'envisager la formation d'un dimère parallèle dû à la présence de corrélations entre protons de bouts de chaînes latérales de résidus adjacents, comme cela a déjà été rencontré par ailleurs (Junius, O'Donoghue et al. 1996). En particulier, les distances mesurées par *buid-up* entre les protons aromatiques de $Tyr^{521}$ et les méthyles de $Leu^{522}$ ont été mesurées à 4,37 Å ± 0,87 ($Tyr^{521}$ (δ*)/ $Leu^{522}$(δ2)), 4,21 Å ± 0,84 ($Tyr^{521}$ (δ*)/ $Leu^{522}$(δ1)), 4,58 Å ± 0,92 ($Tyr^{521}$ (ε*)/ $Leu^{522}$(δ2)) et 4,42 Å ± 0,88 ($Tyr^{521}$ (ε*)/ $Leu^{522}$(β*)) (Figure 68, A). Ces contraintes ont été introduites comme contraintes non ambiguës en plus du pont disulfure.

Les statistiques obtenues pour le calcul du dimère sont résumées dans le Tableau 14. Faute de contraintes intermoléculaires suffisantes, les RMSDs calculés montrent la faible convergence des dix structures calculées sur l'ensemble du dimère.

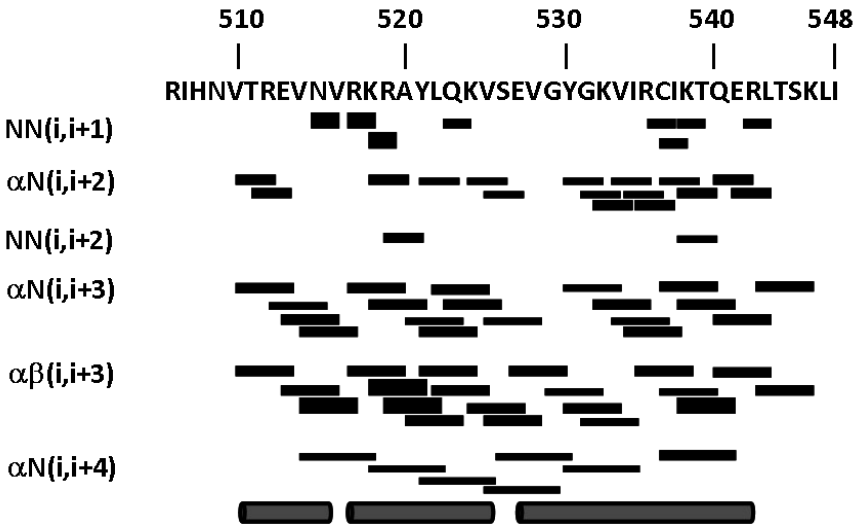



**Figure 67: Table de connectivités des NOEs recueillies pour la modélisation du dimère de pUL56(506-548).**

L'ensemble du spectre dans les deux conditions se superpose parfaitement excepté pour certains résidus.

Même si aucune contrainte intermoléculaire supplémentaire n'a été collectée en plus de celles relevées entre $Tyr^{521}$ et $Leu^{522}$, les déplacements chimiques de résidus aliphatiques et conservés (Figure 65) s'en voit fortement affectés traduisant leurs implications probables dans

la formation du *leucine zipper* dont nous avons supposé l'existence (Annexe D4, en rouge). Ces acides aminés sont $Leu^{522}$, $Val^{525}$, $Val^{528}$, $Val^{533}$, $Ile^{534}$ et $Thr^{544}$. Tous ces résidus sont des résidus hydrophobes appuyant ainsi notre hypothèse sur leur implication dans la dimérisation en accord avec les expériences d'HPLC.

Les deux glutamines $Gln^{523}$ et en particulier $Gln^{540}$, parfaitement conservée (Figure 65) et située un tour d'hélice avant $Thr^{544}$, sont très influencées par la formation du dimère (Figure 68A'). Aucune contrainte n'a été relevée mais la double fonction de donneur et d'accepteur de liaison hydrogène permet de formuler l'hypothèse d'un appariement antisymétrique de leur chaînes latérales pour la formation de deux liaisons hydrogènes qui pourrait stabiliser le dimère dans son extrémité C-terminal (King 1996). Il est suggéré que cette acide aminé pourrait, en plus de la cystéine, conférer une spécificité pour la reconnaissance des deux monomères stabilisée par le leucine zipper (Harbury, Zhang et al. 1993; Junius, Mackay et al. 1995). De plus ce résidu est conservé dans les familles des dystrophine, utrophine, bZIP et bHLH-ZIP (basic helix-loop-helix leucine zipper) (Hurst 1994; Blake, Tinsley et al. 1995).

| | | |
|---|---|---|
| **Restraints for calculation** | | |
| Total NOE restraints | 581 | |
| Intraresidue | 346 | |
| Sequential (\|i-j\|=1) | 114 | |
| Medium range (\|i-j\|=5) | 112 | |
| intermoléculaire | 9 | |
| | | |
| **Structure statistics** | | |
| R.M.S.D. | | |
| Bonds (?) | (1.32-1.7)E-03 | |
| Bond angles (°) | 0.46-0.49 | |
| Improper torsions (°) | 0.33-0.35 | |
| NOE restraints (?) | (2.86-3.47)E-02 | |
| | | |
| **Final Energies (kcal/mol)** | | |
| Total | 135.75-160.18 | |
| Bonds | 2.60-4.33 | |
| Angles | 88.13-98.47 | |
| Improper angles | 12.29-13.89 | |
| Van Der Waals | 4.3-9.83 | |
| NOE | 24.95-36.60 | |
| | | |
| **Ramachandran plot** | | |
| Most allowed (%) | 69.4 | |
| Additionally allowed (%) | 22.8 | |
| Generously allowed (%) | 4.7 | |
| Disallowed (%) | 3.1 | |
| | | |
| **RMSD from pairwise (?)** | **Backbone atoms** | **Heavy atoms** |
| R517-E541 | 3.87±1.27 | 7.92±2.82 |
| R517-V520/V523-K538 | 3.06±0.93 | 8.16±2.98 |
| R517-V520 | 2.73±0.77 | 8.24±3.00 |
| V523-K538 | 2.55±1.11 | 8.13±2.87 |

**Tableau 14: Table de statistiques des dix meilleures structures calculées pour pUL56(506-548) dans 30% TFE à 303 K.**

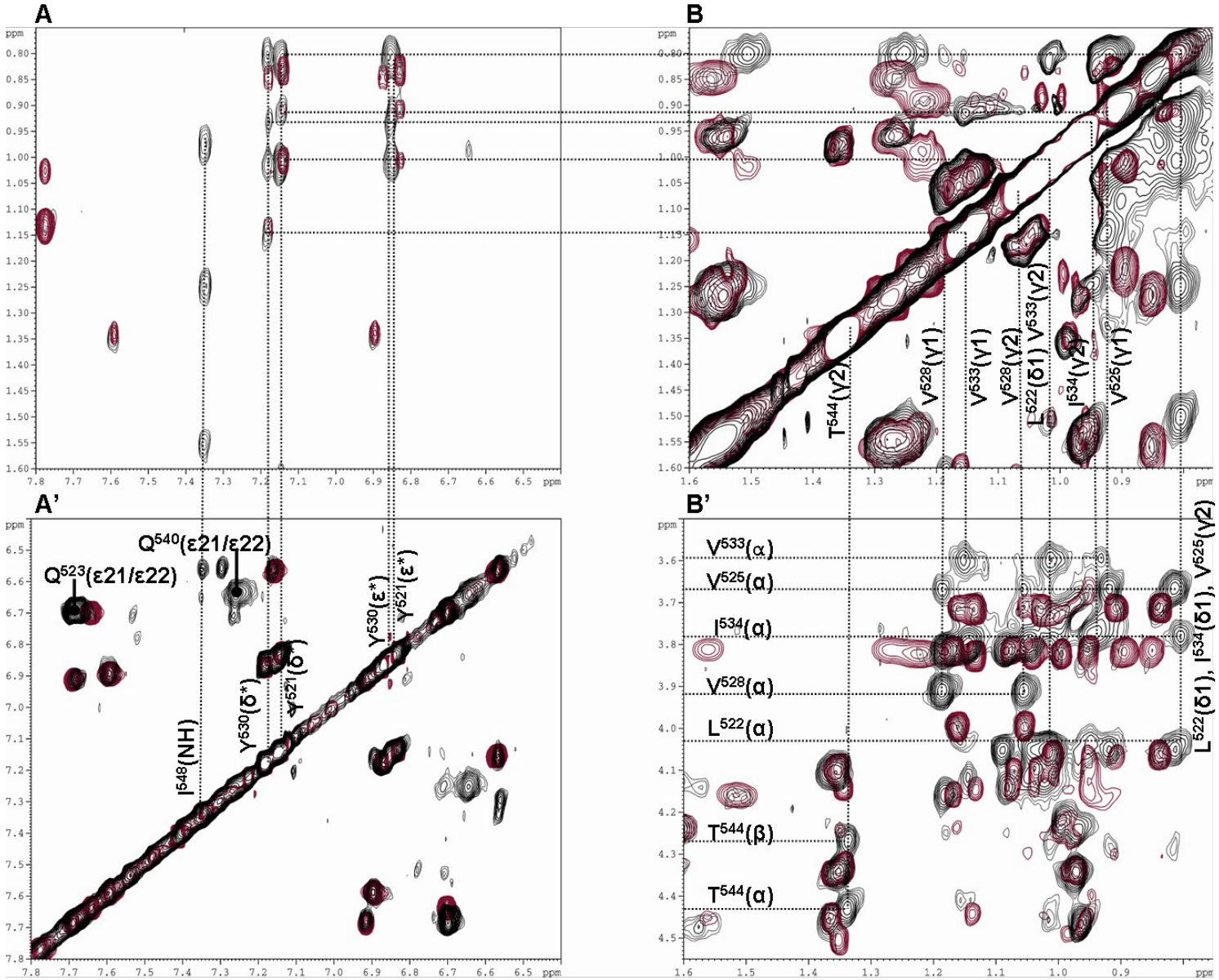
La structure obtenue est composée de trois éléments: un pont disulfure, une interface hydrophobe et un cluster basique. Le pont disulfure est formé par l'appariement des cystéines 536 dont nous avons contrôlé la formation par HPLC (Figure 66).

L'interface hydrophobe est formée principalement par Tyr$^{521}$, Leu$^{522}$, Val$^{525}$ et Val$^{533}$, résidus conservés (ET, Figure 65), en accord avec les observations faites sur le chromatogramme (Figure 66) et les variations de déplacements chimiques décrites en Figure 68.

Enfin le cluster basique contient la triade $^{517}$Arg-Lys-Arg$^{519}$. La superposition avec la structure PDB 1DH3 (Figure 69) permet de prédire une reconnaissance probable de l'ADN au niveau de son grand sillon.

Cette analyse reste hypothétique mais permet d'expliquer la parfaite conservation à travers les homologues herpesvirauux de pUL56 des résidus Arg$^{517}$ pour la liaison de l'ADN, et de Tyr$^{521}$, Cys$^{536}$ et Gln$^{540}$ pour sa dimérisation avec une orientation comparable à celle observée dans les structures de type *bZIP* consignées dans la PDB.

**Figure 68 (page suivante): Comparaison des spectres RMN de pUL56(506-548) entre la forme oxydée et réduite du pont disulfure.** Les spectres des formes réduite et oxydée sont colorés en rouge et noir, respectivement. A, A', B et B' sont des extraits des spectres correspondants au zone des protons aromatique/aliphatique, aromatique/aromatique, aliphatique/aliphatique et proton alpha/aliphatique, respectivement. Les corrélations entre Tyr$^{521}$ et Leu$^{522}$ ont été introduites comme contraintes intermoléculaire pour la formation du dimère (A). Gln$^{523}$ et en particulier Gln$^{540}$ sont influencés par la formation du pont disulfure (A'). Les méthyles de Leu$^{522}$, Val$^{525}$ et Ile$^{534}$ et les protons alpha de Leu$^{522}$, Val$^{525}$, Val$^{528}$, Val$^{533}$, Ile$^{534}$ et Thr$^{544}$ sont influencés par la formation du pont disulfure (B et B').


A
B
A'
B'
Q523(ε21/ε22)
Q540(ε21/ε22)
I548(NH)
Y530(δ*)
Y521(δ*)
Y530(ε*)
Y521(ε*)
T544(γ2)
V528(γ1)
V533(γ1)
V528(γ2)
L522(δ1) V533(γ2)
I534(γ2)
V525(γ1)
V533(α)
V525(α)
I534(α)
V528(α)
L522(α)
T544(β)
T544(α)
L522(δ1), I534(δ1), V525(γ2)
ppm

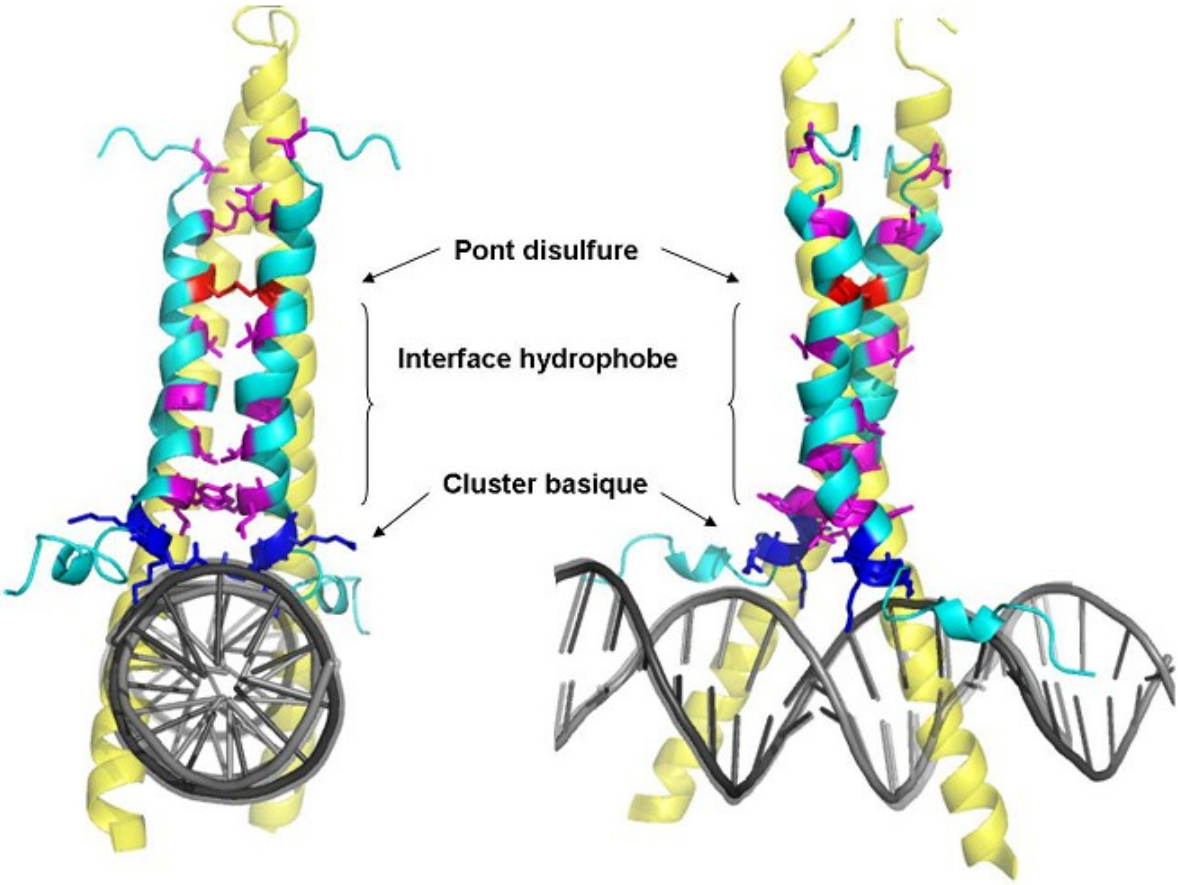



**Figure 69: Comparaison de la structure de plus faible énergie calculée par RMN du dimère de pUL56(506-548) avec la structure PDB 1DH3.** La structure de pUL56(506-549) est coloré en ruban cyan. Le cluster basique ($^{517}$Arg-Lys-Arg$^{519}$) est coloré en bleu foncé, les chaînes latérales des résidus constituant l'interface hydrophobe sont colorées en magenta (Tyr$^{521}$, Leu$^{522}$, Val$^{525}$, Val$^{533}$, Gln$^{540}$ et Thr$^{544}$) et le pont disulfure en rouge (Cys$^{536}$). La structure 1DH3 est colorée en ruban jaune. Le positionnement relatif des deux unités monomériques est orienté de manière comparable dans les deux cas. Le duplexe d'ADN issu de la structure PDB 1DH3 est représenté en ruban gris.

L'étude de ce domaine de pUL56 a été initiée par la recherche de manière empirique d'un site d'interaction pour pUL89(580-600).

Pour vérifier cette hypothèse, nous avons procédé à des expériences de *cross-linking*. Bien que nous n'ayons pu observer de réticulation entre les deux peptides, nous avons mis en évidence une forte propension du fragment pUL56(506-548) à dimériser.

Une recherche d'homologues structuraux a été réalisée par PHYRE et a révélé une similarité avec les protéines de types CREB capables d'interagir avec l'ADN grâce à un cluster basique après constitution d'un dimère par une interface hydrophobe.

Nous avons poursuivi nos investigations en forçant la formation d'un pont disulfure probable dû à la présence d'une cystéine parfaitement conservée à travers les homologues herpesviraux de pUL56. La formation de ce pont disulfure a été suivie par HPLC en phase inverse. Bien que deux fois plus gros, le dimère est élué 30 secondes avant le monomère. Nous avons donc suggéré que dans ces conditions les résidus hydrophobes sont cachés du solvant pour former une interface hydrophobe stabilisant le dimère.

Ces observations ont été confortées par l'observation en RMN de signaux entre extrémités de chaînes latérales de résidus adjacents, d'une part, et le blindage électronique de plusieurs méthyles de résidus hydrophobes conservés et situés sur une même face de l'hélice formée, d'autre part. En plus du pont disulfure, le dimère formé fait apparaître une interface impliquant des résidus conservés à travers les homologues herpesviraux de pUL56.

De plus, la glutamine 540 est elle aussi fortement influencé par la formation du pont disulfure. Il est alors envisagé qu'elle aussi occupe un rôle prépondérant dans la dimérisation de ce peptide.

La superposition de ce dimère avec une protéine CREB souligne le rôle du cluster basique conservé dans l'interaction potentielle avec le grand sillon d'un duplexe d'ADN.

Ceci nous permet de définir ce peptide comme appartenant à la famille des structures de type *bZIP* compatible à la fois avec le rôle de reconnaissance spécifique de l'ADN et de la dimérisation observée en microscopie électronique.

## D. Bilan De L'étude Des Domaines De PUL56

L'identification des domaines fonctionnels de pUL56 après alignement de séquences a révélé la présence d'un grand nombre de cystéines, en particulier dans son extrémité N-terminal, conservées à travers les homologues de bêta herpesvirus de pUL56. Cette concentration importante de tels résidus peut traduire la faculté à dimériser de pUL56 à travers ce domaine.

Par ailleurs, nous avons tenté de mettre en évidence la formation d'un doigt de zinc de type $CX_2CX_{24}CXH$ parfaitement conservé chez les homologues herpesviraux de pUL56 et probablement impliqué dans le mode d'action des dérivés D-benzimidazolés car contenant une mutation résistante à ces inhibiteurs (Q204R) au milieu de sa boucle.

Nous avons donc synthétisé le peptide pUL56(186-233) que nous avons analysé par divers spectroscopies. Par dichroïsme circulaire, nous avons mis en évidence une légère influence du zinc sur la structuration en hélice du peptide. En UV/visible, nous avons observé les transitions caractéristiques de la chélation du métal par des résidus cystéines et histidines. Malheureusement, au-delà de 0,5 équivalent de métal, les signaux deviennent moins bien définis suggérant une instabilité du doigt de zinc potentiel.

Nous avons émis plusieurs hypothèses qui restent à vérifier pour expliquer ce phénomène. En particulier, nous proposons d'allonger la taille du système pour faire éventuellement apparaître des éléments de structures secondaires supplémentaires capables de stabiliser la large boucle de 24 acides aminés. Nous proposons aussi de muter une histidine non conservée ($His^{223}$) qui peut interférer avec l'histidine conservée $His^{219}$ et ainsi déstabiliser la formation du doigt de zinc.

Les expériences de RMN ont été entreprises pour l'étude de ce peptide mais sa faible solubilisation à pH neutre a empêché l'observation d'un repliement sous l'effet d'ajout de zinc.

Un autre domaine capable d'interagir avec l'ADN a été étudié. Celui-ci a été sélectionné empiriquement comme partenaire potentiel dans une interaction éventuelle avec pUL89(580-600).

Par *cross-linking*, cette hypothèse n'a pas été vérifiée mais une dimérisation de ce peptide a été observée.

La comparaison des temps d'élution en HPLC phase inverse de ce peptide avec et sans pont disulfure, formé à travers la cystéine 536, a suggéré une dimérisation à travers une interface hydrophobe.

En RMN, les attributions des déplacements chimiques ont été réalisées avec et sans pont disulfure. Les différences de déplacements chimiques ont été principalement observées pour des résidus hydrophobes, conservés de surcroît. En plus de ces résidus hydrophobes, la glutamine 540, complètement conservé, est elle aussi très influencé par la formation du pont disulfure impliquant donc un rôle dans la dimérisation. Par ailleurs, des NOEs présentes pour la forme réduite, et exacerbés dans la forme oxydée, entre $Tyr^{521}$ et $Leu^{522}$ permettent d'envisager un rôle crucial de ces deux résidus dans l'amorce de la dimérisation.

La structure RMN de ce dimère avec pont disulfure a été calculée.

La comparaison de pUL56(506-548) avec des protéines de type CREB montre la constitution d'une interface hydrophobe de type *leucine zipper* permettant la dimérisation de deux monomères pour former, à sa base, un cluster basique capable d'interagir avec le grand sillon d'un ADN double-brin.

La reconnaissance spécifique de l'ADN est attribuée à pUL56 dans l'étape de maturation et de ce fait peut constituer une cible thérapeutique intéressante. C'est pour cette raison que nos hypothèses doivent être vérifiées afin d'aboutir à l'élucidation des mécanismes de reconnaissance de l'ADN viral par pUL56, en particulier au niveau des séquences *pac*, et des dérivés D-benzimidazolés au niveau du doigt de zinc potentiel.

Il sera donc indispensable d'étudier les complexes pUL56-ZF (sauvage puis mutant Q204R)/*pac* en présence de D-ribobenzimidazolé d'une part, et pUL56-*bZIP*/*pac* d'autre part.

## IX. CONCLUSION GÉNÉRALE ET PERSPECTIVES

Les travaux exposés dans cette thèse ont ciblé l'étude structurale de plusieurs protéines du Cytomégalovirus humain considérées comme de nouvelles cibles thérapeutiques. Celles-ci ont été découvertes suite à la conception d'inhibiteurs autres que les traitements actuellement disponibles visant pUL54, l'ADN polymérase de CMVH (Ganciclovir, Valganciclovir, Cidofovir et Foscarnet). En effet, les traitements visant pUL54 ont fait émerger un nombre important de mutations de résistances au niveau de pUL54 et de la phosphotransférase virale pUL97. Depuis environ 15 ans, une nouvelle classe de molécules offrant une alternative thérapeutique intéressante est en cours de développement: les dérivés ribobenzimidazolés. Les premiers à avoir été développés sont les D-ribobenzimidazolés BDCRB et TCRB. Ceux-ci ont montré des activités *in vitro* très intéressantes dirigées vers le complexe terminase ((pUL89 et pUL56) de CMVH et sa protéine portail (pUL104). Cependant, *in vivo*, ces molécules se sont vues dégradées au niveau de leur liaison glycosidique par les glycosylases cellulaires. Pour remédier à ce défaut de biodisponibilité, plusieurs autres molécules ont été développées après optimisation des BDCRB et TCRB par l'incorporation de groupements hydroxyles, de chaînes aliphatiques, d'halogénation supplémentaire ou de la formulation de leurs homologues L-ribobenzimidazolés. Cette dernière classe a comme représentant principal le Maribavir visant la phosphotransférase pUL97 et sa protéine accessoire pUL27.

A l'heure actuelle, malgré le *screening* de nombreuses molécules visant les terminases ou pUL97, peu d'informations structurales sont actuellement disponibles concernant ces protéines-cibles. Afin d'élucider le mécanisme d'action de ces nouvelles molécules et de rationaliser leur optimisation, la connaissance de leur structure tridimensionnelle s'avère indispensable. Par conséquent, nous avons initié une étude structurale des terminases pUL89 et pUL56. Celles-ci a reposé sur des analyses bioinformatiques qui nous ont permis de déterminer des domaines conservés ayant par conséquent des rôles importants dans des fonctions catalytiques ou dans l'imbriquement des différentes sous-unités terminases menant à la maturation et à la translocation de l'ADN viral dans la capside.

L'étude de ces domaines a nécessité leur production en masse et en grande pureté dans la perspective de leur étude par résonance magnétique nucléaire. Cependant, les protocoles d'expression dans des systèmes bactériens restent à optimiser puisque nous avons dû faire face à des problèmes de toxicité ne permettant pas l'obtention de matériel suffisant pour des études expérimentales.

Historiquement, l'Unité de Pharmacologie Chimique et Génétique où ont été effectués ces travaux est riche d'une grande expérience dans le domaine de la synthèse peptidique en phase solide (SPPS). Ceci a donc constitué une alternative pour la production d'échantillons de grandes puretés. Néanmoins, plusieurs inconvénients inhérents à cette technique et à la nature des séquences peptidiques sélectionnées ne nous ont pas permis d'explorer expérimentalement les différentes hypothèses structurales émises par la simple analyse des alignements de séquence.

En effet, la SPPS est basée sur le couplage chimique d'acides aminés par l'alternance d'étapes de déprotection et de couplage à l'intérieur d'un réacteur. Le rendement de ces étapes est intimement lié à la nature de la séquence peptidique (domaines riches en résidus hydrophobes, présence de glycines ou de prolines) qui peut aboutir à des rendements faibles. Par ailleurs, avec la longueur de la séquence l'accessibilité des acides aminés à coupler diminue et par conséquent le rendement en est affecté tout autant. L'utilisation de résines spéciales pour limiter la charge en peptide, le *capping* pour bloquer des positions dont les déprotections auraient été inefficaces et l'augmentation des temps de couplage pour augmenter les chances d'enchaîner les acides aminés ont été optimisés pour chaque séquence synthétisée. Ainsi, nous avons pu synthétiser quelques domaines de pUL89 et de pUL56 en grandes quantités (de 50 mg à plus de 100 mg pour certains).

Cependant, après obtention de ces échantillons, restent les étapes de solubilisation de ces peptides pour leurs études en RMN (dans l'eau de préférence pour être au plus proche des conditions physiologiques). Là encore, des problèmes récurrents ont été rencontrés notamment pour l'étude du doigt de zinc potentiel de pUL56 (de HCMV ou MCMV) ou des domaines d'interaction de pUL89 (pUL89 (568-635)). En dernier recours des solvants organiques (acétonitril, TFE) ou des milieux lipidiques (DPC) ont dû être utilisés pour résoudre ce genre de problèmes.

Une ultime étape reste à surmonter lorsque l'on étudie des protéines: leur structuration. C'est la structure qui va répartir efficacement les acides aminés pour le repliement stable d'un domaine menant dans le cas idéal à la constitution d'un cœur hydrophobe, d'un site catalytique ou un motif de fixation à un métal. Cette dernière étape a été difficile à surmonter pour l'étude du doigt de zinc de pUL56 en particulier. Pour pUL89(580-600), l'utilisation de TFE ou de DPC a mené à la formation d'une hélice en accord avec les prédictions de structures secondaires et retrouvé lors de l'allongement de la séquence (pUL89(568-635)). Comme nous l'avons rappelé, la SPPS est limitée par la taille du peptide à synthétiser. De la

même manière, la RMN dans son utilisation classique est elle aussi peu adaptée pour l'étude de peptides excédant 100 acides aminés en abondance isotopique naturelle.

Nous nous sommes donc tournés vers des méthodes théoriques pour étudier différents domaines de pUL89. La modélisation n'est pas limitée par la taille mais par l'enrichissement des banques de données telle que la PDB. La connaissance d'homologues structuraux déjà résolus par RMN ou diffraction des rayons X est à la base de la modélisation moléculaire. Cependant la fiabilité des modèles générés dépend grandement du pourcentage de similarité entre une structure-support et la séquence cible. Plus celui-ci sera élevé, plus la structure théorique sera plausible et donc exploitable pour l'analyse fine du repliement et pouvant servir d'outil de travail pour l'optimisation d'inhibiteurs.

Les modèles générés doivent cependant être scrupuleusement examinés pour éliminer les clashs stériques après réparation des chaînes latérales, compenser l'insertion de *gaps* éventuels, optimiser les alignements de séquences ou ajouter des cofacteurs (molécules d'eau, métaux..).

Les modèles générés restent théoriques et par conséquent peuvent être sujet à controverse tant que l'expérience ne vient pas soutenir la fiabilité de ces modèles.

Nous avons ainsi étudié les domaines hélicase, d'une part, et endonucléase couplé au site d'interaction avec pUL56, d'autre part, contenus dans la séquence de pUL89.

Le domaine hélicase modélisé à partir de la structure de son homologue gp17 du bactériophage T4 a permis de rendre compte de la conservation de certains acides aminés impliqués dans la fixation de l'ATP dans son site actif et la reconnaissance de l'ADN pour sa translocation. Ce modèle inclus les positions $Asp^{344}$ et $Ala^{355}$ sensibles aux dérivés D-ribobenzimidazolés et sont situés à proximité d'un doigt de zinc potentiel (d'après les alignements de séquences) qui n'a pas pu être modélisé mais pourrait être impliqué dans la reconnaissance du duplexe d'ADN pour sa translocation.

Le domaine endonucléase a quant à lui été modélisé grâce à une homologie structurale avec des RNases H de type I. Par ailleurs, la résolution de la structure du peptide 568-635 établie par RMN a permis l'observation de quatre hélices qui ont été utilisées comme contraintes pour l'obtention de ce modèle. Celui-ci contient un site de fixation aux métaux constitué par trois résidus acides conservés. L'un d'eux appartient au fragment 580-600 identifié par l'équipe d'Elke Bogner comme étant indispensable à l'interaction avec pUL56. Notre modèle peut donc expliquer la dépendance de l'activité endonucléasique de pUL89 sous l'effet de son association avec pUL56. Un deuxième élément pouvant rendre compte de

cette dépendance est le mouvement d'un fragment à proximité du fragment 580-600 observé en comparant les structures calculées en présence et en absence d'un duplexe hybride ADN/ARN. La reconnaissance spécifique de l'ADN au niveau des séquences *pac* est imputée à pUL56. Le mouvement observé peut être induit par l'assemblage de pUL89 avec pUL56 afin d'optimiser l'interaction avec l'ADN pour le clivage ultérieur de l'excès d'ADN après sa translocation dans la capside. Cette hypothèse semble être cohérente avec des observations faites sur les structures expérimentales de divers homologues de pUL89 issus de plusieurs bactériophages.

Les structures et fonctions des domaines N-terminal (1-140), central (365-455) contenant le doigt de zinc $^{365}CX_2HX_{11}CXC^{385}$ et C-terminal (613-674) restent à élucider.

L'étude de pUL56 s'est avérée plus laborieuse que celle de pUL89 faute d'homologue proche ou éloigné dans les banques de données. Nous avons tout de même réussi à montrer une propension à lier le zinc au niveau du fragment $^{186}$Val-Ile$^{233}$ en accord avec l'hypothèse sur l'existence d'un doigt de zinc de type $CX_2CX_{24}CXH$ contenant au centre de sa boucle le site de mutation résistante Q204R aux BDCRB et TCRB. Cependant, la RMN n'a pas permis de confirmer ce résultat à cause d'une faible solubilité de ce peptide à pH physiologique.

Par ailleurs, nous avons recherché le site d'interaction de pUL56 pour pUL89 en nous basant sur des études statistiques concernant les interfaces protéine-protéine. Les expériences de *cross-linking* n'ont pas permis de vérifier cette hypothèse. Cependant elles ont révélées la propension du fragment 506-548 à dimériser. D'après une recherche d'homologues opérée par PHYRE, cette hypothèse s'avère plausible puisque ce domaine a montré des similitudes structurales avec des protéines de type CREB capable d'interagir avec l'ADN *via* un cluster basique. Cette hypothèse reste à vérifier expérimentalement mais plusieurs signatures dans la séquence primaire concordent avec cette idée et notamment la conservation d'un cluster basique, de divers résidus hydrophobes et d'une cystéine parfaitement conservée. La structure RMN de ce peptide a été réalisée dans le TFE et la superposition de cette structure avec une protéine CREB tirée de la PDB permet effectivement d'observer une interaction probable avec l'ADN à travers le cluster basique de pUL56(506-548).

Enfin, nous nous sommes intéressés à l'étude préliminaire de pUL97 et pUL27, toutes deux ciblées par le Maribavir.

Concernant pUL97, la modélisation entreprise sur son domaine kinase reste à optimiser car plusieurs incohérences avec les résultats expérimentaux ont été observées dans notre modèle. L'utilisation de Modeller pourra remédier à cet inconvénient. Une fois le modèle

obtenu il sera possible de mieux comprendre comment pUL97 reconnaît spécifiquement des motifs de type S/T-$X_4$-K/R. Ainsi, la formulation d'un inhibiteur basé sur ce motif capable d'interagir spécifiquement avec pUL97 et ainsi inhiber son activité au centre du cycle de réplication viral de CMVH pourra être envisagé.

L'étude sur pUL27, protéine accessoire de pUL97, s'est limitée à l'analyse de sa séquence primaire. Il semblerait que seuls les bêta herpesvirus soient pourvus d'une telle protéine. Grâce à un alignement de séquences, nous avons décrit quatre domaines conservés à travers les différents homologues de pUL27 contenant des signatures compatibles avec une multimérisation éventuelle de pUL27. Nous avons par ailleurs identifié deux fragments situés en N et C-terminaux riches en résidus basiques pouvant contenir des signaux de localisation nucléaire. Aucun homologue structural n'a été trouvé par PHYRE et donc nous n'avons pu proposer de modèle théorique pour cette protéine.

En conclusion, les problèmes rencontrés pour la production, la solubilisation et la structuration des fragments étudiés ont été des facteurs limitant non négligeables pour l'élucidation expérimentale de leurs structures.

pUL89, pUL56, pUL97 et pUL27 sont des protéines de grandes tailles (674, 850, 707 et 608 acides aminés). Par conséquent, la RMN dans son utilisation conventionnelle n'est pas une technique adaptée pour la résolution complète de leurs structures. Il ne reste alors que deux possibilités: la diffraction des rayons X ou la modélisation par homologie. Nous l'avons vu, la modélisation est limitée par l'enrichissement des banques de données en homologues structuraux et reste une méthode théorique. Il subsiste donc toujours un doute sur la fiabilité et l'exploitation tirées de ces structures. Il est donc indispensable d'étudier ces différents systèmes par diffraction des rayons X dès que les protocoles de surexpression seront optimisés.

Il sera alors possible d'envisager la cocristallisation des unités pUL89, pUL56 et pUL104 (702 acides aminés) en présence de double-brin d'ADN pour mieux comprendre les interactions régissant la constitution de cet énorme édifice moléculaire pour les étapes de maturation et de translocation de l'ADN viral. Le mode d'action des dérivés D-ribobenzimidazolés et lui aussi au stade d'hypothèses bien que son rôle dans la perturbation de la reconnaissance entre le complexe terminase et l'ADN soit une idée qui nous semble plausible au regard des mutations de résistance situés systématiquement au niveau de doigts de zinc potentiels de pUL56 et pUL89. Les protocoles de synthèses chimiques des dérivés benzimidazolés ont été optimisés et les chimistes de notre équipe ont réussi à produire

quelques unes de ces molécules en grande pureté. L'étude des mutants de pUL89, pUL56 et pUL104 sont des étapes indispensables à l'élucidation de leur mécanisme d'action et donc de leur optimisation future pour augmenter leur spécificité et réduire les effets secondaires éventuels. Des études similaires, toujours par RX devront être menées sur pUL97 et pUL27. En particulier, pour cette dernière l'élucidation de sa structure pourra affirmer ou infirmer l'hypothèse que nous avons formulée sur sa multimérisation.

Afin de résoudre la problématique posée, cette thèse s'est appuyée sur l'utilisation de la résonance magnétique nucléaire qui a pu être complétée par la modélisation grâce à l'accumulation croissante des structures RMN et RX dans les banques de données. A long terme, les différentes informations collectées par les différentes équipes de biostructuralistes du monde entier, pourront peut être permettre de se dispenser de techniques expérimentales pour la résolution de structures de protéine dont leur modélisation pourra s'effectuer grâce à la seul connaissance de leur séquence.

## X. RÉFÉRENCES


Alain, S., M. C. Mazeron, et al. (2004). "Résistance du cytomégalovirusnext term au ganciclovir ." Antibiotiques **6**: 113-119

Altschul, S. F., T. L. Madden, et al. (1997). "Gapped BLAST and PSI-BLAST: a new generation of protein database search programs." Nucleic Acids Res **25**(17): 3389-402.

Anfinsen, C. B. (1973). "Principles that govern the folding of protein chains." Science **181**(96): 223-30.

Ansari, S. and V. Helms (2005). "Statistical analysis of predominantly transient protein-protein interfaces." Proteins **61**(2): 344-55.

Appleton, B. A., J. Brooks, et al. (2006). "Crystal structure of the cytomegalovirus DNA polymerase subunit UL44 in complex with the C terminus from the catalytic subunit. Differences in structure and function relative to unliganded UL44." J Biol Chem **281**(8): 5224-32.

Appleton, B. A., A. Loregian, et al. (2004). "The cytomegalovirus DNA polymerase subunit UL44 forms a C clamp-shaped dimer." Mol Cell **15**(2): 233-44.

Arunkumar, A. I., T. K. Kumar, et al. (1997). "Non-specific helix-induction in charged homopolypeptides by alcohols." Biochim Biophys Acta **1338**(1): 69-76.

Back, J. W., L. de Jong, et al. (2003). "Chemical cross-linking and mass spectrometry for protein structural modeling." J Mol Biol **331**(2): 303-13.

Baek, M. C., P. M. Krosky, et al. (2002). "Relationship between autophosphorylation and phosphorylation of exogenous substrates by the human cytomegalovirus UL97 protein kinase." J Virol **76**(23): 11943-52.

Baek, M. C., P. M. Krosky, et al. (2002). "Specific phosphorylation of exogenous protein and peptide substrates by the human cytomegalovirus UL97 protein kinase. Importance of the P+5 position." J Biol Chem **277**(33): 29593-9.

Baek, M. C., P. M. Krosky, et al. (2004). "Phosphorylation of the RNA polymerase II carboxyl-terminal domain in human cytomegalovirus-infected cells and in vitro by the viral UL97 protein kinase." Virology **324**(1): 184-93.

Baldanti, F., D. Lilleri, et al. (2004). "Human cytomegalovirus double resistance in a donor-positive/recipient-negative lung transplant patient with an impaired CD4-mediated specific immune response." J Antimicrob Chemother **53**(3): 536-9.

Bartlett, G. J., C. T. Porter, et al. (2002). "Analysis of catalytic residues in enzyme active sites." J Mol Biol **324**(1): 105-21.

Bavoso, A., A. Ostuni, et al. (1998). "Metal ion binding to a zinc finger peptide containing the Cys-X2-Cys-X4-His-X4-Cys domain of a nucleic acid binding protein encoded by the Drosophila Fw-element." Biochem Biophys Res Commun **242**(2): 385-9.

Beese, L. S., V. Derbyshire, et al. (1993). "Structure of DNA polymerase I Klenow fragment bound to duplex DNA." Science **260**(5106): 352-5.

Benkert, P., M. Kunzli, et al. (2009). "QMEAN server for protein model quality estimation." Nucleic Acids Res **37**(Web Server issue): W510-4.

Berjanskii, M. and D. S. Wishart (2006). "NMR: prediction of protein flexibility." Nat Protoc **1**(2): 683-8.

Berjanskii, M. V. and D. S. Wishart (2007). "The RCI server: rapid and accurate calculation of protein flexibility using chemical shifts." Nucleic Acids Res **35**(Web Server issue): W531-7.

Biertumpfel, C., W. Yang, et al. (2007). "Crystal structure of T4 endonuclease VII resolving a Holliday junction." Nature **449**(7162): 616-20.

Biron, K. K., R. J. Harvey, et al. (2002). "Potent and selective inhibition of human cytomegalovirus replication by 1263W94, a benzimidazole L-riboside with a unique mode of action." Antimicrob Agents Chemother **46**(8): 2365-72.

Blake, D. J., J. M. Tinsley, et al. (1995). "Coiled-coil regions in the carboxy-terminal domains of dystrophin and related proteins: potentials for protein-protein interactions." Trends Biochem Sci **20**(4): 133-5.

Bogan, A. A. and K. S. Thorn (1998). "Anatomy of hot spots in protein interfaces." J Mol Biol **280**(1): 1-9.

Bogner, E., K. Radsak, et al. (1998). "The gene product of human cytomegalovirus open reading frame UL56 binds the pac motif and has specific nuclease activity." J Virol **72**(3): 2259-64.

Bogner, E., M. Reschke, et al. (1993). "Identification of the gene product encoded by ORF UL56 of the human cytomegalovirus genome." Virology **196**(1): 290-3.

Borst, E. M., K. Wagner, et al. (2008). "The essential human cytomegalovirus gene UL52 is required for cleavage-packaging of the viral genome." J Virol **82**(5): 2065-78.

Braithwaite, D. K. and J. Ito (1993). "Compilation, alignment, and phylogenetic relationships of DNA polymerases." Nucleic Acids Res **21**(4): 787-802.

Britt, W. J. and S. Boppana (2004). "Human cytomegalovirus virion proteins." Hum Immunol **65**(5): 395-402.

Brünger, A. T. (1992). X-PLOR: Software Manual, version 3.1. NewHaven, CT.

Buerger, I., J. Reefschlaeger, et al. (2001). "A novel nonnucleoside inhibitor specifically targets cytomegalovirus DNA maturation via the UL89 and UL56 gene products." J Virol **75**(19): 9077-86.

Burke, C. J., G. Sanyal, et al. (1992). "Structural implications of spectroscopic characterization of a putative zinc finger peptide from HIV-1 integrase." J Biol Chem **267**(14): 9639-44.

Caruthers, J. M. and D. B. McKay (2002). "Helicase structure and mechanism." Curr Opin Struct Biol **12**(1): 123-33.

Champier, G., A. Couvreux, et al. (2008). "Putative functional domains of human cytomegalovirus pUL56 involved in dimerization and benzimidazole D-ribonucleoside activity." Antivir Ther **13**(5): 643-54.

Champier, G., S. Hantz, et al. (2007). "New functional domains of human cytomegalovirus pUL89 predicted by sequence analysis and three-dimensional modelling of the catalytic site DEXDc." Antivir Ther **12**(2): 217-32.

Chan, J. H., S. D. Chamberlain, et al. (2000). "Synthesis and evaluation of a series of 2'-deoxy analogues of the antiviral agent 5,6-dichloro-2-isopropylamino-1-(beta-L-ribofuranosyl)-1H-benzimidazole (1263W94)." Nucleosides Nucleotides Nucleic Acids **19**(1-2): 101-23.

Chee, M. S., G. L. Lawrence, et al. (1989). "Alpha-, beta- and gammaherpesviruses encode a putative phosphotransferase." J Gen Virol **70 ( Pt 5)**: 1151-60.

Chelsky, D., R. Ralph, et al. (1989). "Sequence requirements for synthetic peptide-mediated translocation to the nucleus." Mol Cell Biol **9**(6): 2487-92.

Chen, D. H., H. Jiang, et al. (1999). "Three-dimensional visualization of tegument/capsid interactions in the intact human cytomegalovirus." Virology **260**(1): 10-6.

Chen, J. J., Y. Wei, et al. (2005). "Design, synthesis, and antiviral evaluation of some polyhalogenated indole C-nucleosides." Nucleosides Nucleotides Nucleic Acids **24**(10-12): 1417-37.

Chien, T. C., S. S. Saluja, et al. (2004). "Synthesis and antiviral evaluation of polyhalogenated imidazole nucleosides: dimensional analogues of 2,5,6-trichloro-1-(beta-D-ribofuranosyl)benzimidazole." J Med Chem **47**(23): 5743-52.

Chothia, C. and A. M. Lesk (1986). "The relation between the divergence of sequence and structure in proteins." Embo J **5**(4): 823-6.

Chou, S. (2008). "Cytomegalovirus UL97 mutations in the era of ganciclovir and maribavir." Rev Med Virol **18**(4): 233-46.

Chou, S. (2009). "Diverse cytomegalovirus UL27 mutations adapt to loss of viral UL97 kinase activity under maribavir." Antimicrob Agents Chemother **53**(1): 81-5.

Chou, S., G. I. Marousek, et al. (2004). "Mutations in the human cytomegalovirus UL27 gene that confer resistance to maribavir." J Virol **78**(13): 7124-30.

Cihlar, T. and M. S. Chen (1996). "Identification of enzymes catalyzing two-step phosphorylation of cidofovir and the effect of cytomegalovirus infection on their activities in host cells." Mol Pharmacol **50**(6): 1502-10.

Cole, R. and A. G. Kuttner (1926). "A filtrable virus present in the submaxillary glands of guinea pigs." J. Exp. Med. **44**: 855-73.

Conseil, J. (2006) "Les Herpesvirus : classification, génome et cycle viral. ." **Volume**, DOI:

Contrafouris, Y. (1999). "New antimicrobial agents." IDrugs **2**(12): 1250-2.

Counterman Burba, A. E., U. Lenhert, et al. (2006). "Helix Interaction Tool(HIT): a web-based tool for analysis of helix-helix interactions in proteins." Structural bioinformatics **22**: 2735-2738.

Cozzetto, D., A. Giorgetti, et al. (2008). "The evaluation of protein structure prediction results." Mol Biotechnol **39**(1): 1-8.

Creighton, T. E. (1984). PROTEINS Structures and Molecular Properties. New York, the United States of America.

Davison, A. J. (2002). "Evolution of the herpesviruses." Vet Microbiol **86**(1-2): 69-88.

Dayhoff, M. O. (1978). supplement 3. Atlas of Protein Sequence and Structure

Washington, Natl. Biomed. Res. Found. **5:** 345-352.

de Beer, T., J. Fang, et al. (2002). "Insights into specific DNA recognition during the assembly of a viral genome packaging machine." Mol Cell **9**(5): 981-91.

De Clercq, E. (2004). "Antiviral drugs in current clinical use." Journal of Clinical Virology **30**: 115-133.

de Kreij, A., B. van den Burg, et al. (2002). "The effects of modifying the surface charge on the catalytic activity of a thermolysin-like protease." J Biol Chem **277**(18): 15432-8.

DeLano, W. L. (2002). "Unraveling hot spots in binding interfaces: progress and challenges." Curr Opin Struct Biol **12**(1): 14-20.

Delarue, M., O. Poch, et al. (1990). "An attempt to unify the structure of polymerases." Protein Eng **3**(6): 461-7.

Dittmer, A., J. C. Drach, et al. (2005). "Interaction of the putative human cytomegalovirus portal protein pUL104 with the large terminase subunit pUL56 and its inhibition by benzimidazole-D-ribonucleosides." J Virol **79**(23): 14660-7.

Ducancelle, A. and M. C. Mazeron (2006). "L'ADN polymérase ducytomégalovirus humain: structure, inhibiteurs et mécanismes moléculaires de résistance.
." Virologie **10**: 83-93.

Durr, H., A. Flaus, et al. (2006). "Snf2 family ATPases and DExx box helicases: differences and unifying concepts from high-resolution crystal structures." Nucleic Acids Res **34**(15): 4160-7.

Eastberg, J. H., J. Eklund, et al. (2007). "Mutability of an HNH nuclease imidazole general base and exchange of a deprotonation mechanism." Biochemistry **46**(24): 7215-25.

Eilers, M., A. B. Patel, et al. (2002). "Comparison of helix interactions in membrane and soluble alpha-bundle proteins." Biophys J **82**(5): 2720-36.

Eom, S. H., J. Wang, et al. (1996). "Structure of Taq polymerase with DNA at the polymerase active site." Nature **382**(6588): 278-81.

Erice, A. (1999). "Resistance of human cytomegalovirus to antiviral drugs." Clin Microbiol Rev **12**(2): 286-97.

Eswar, N., D. Eramian, et al. (2008). "Protein structure modeling with MODELLER." Methods Mol Biol **426**: 145-59.

Evers, D. L., G. Komazin, et al. (2004). "Inhibition of human cytomegalovirus replication by benzimidazole nucleosides involves three distinct mechanisms." Antimicrob Agents Chemother **48**(10): 3918-27.

Evers, D. L., G. Komazin, et al. (2002). "Interactions among antiviral drugs acting late in the replication cycle of human cytomegalovirus." Antiviral Res **56**(1): 61-72.

Feiss, M., S. Frackman, et al. (1985). "Essential interaction between lambdoid phage 21 terminase and the Escherichia coli integrative host factor." J Mol Biol **183**(2): 239-46.

Fooks, H. M., A. C. Martin, et al. (2006). "Amino acid pairing preferences in parallel beta-sheets in proteins." J Mol Biol **356**(1): 32-44.

Forster, M. J. (2002). "Molecular modelling in structural biology." Micron **33**(4): 365-84.

Franklin, M. C., J. Wang, et al. (2001). "Structure of the replicating complex of a pol alpha family DNA polymerase." Cell **105**(5): 657-67.

Friedhoff, P., I. Franke, et al. (1999). "A similar active site for non-specific and specific endonucleases." Nat Struct Biol **6**(2): 112-3.

Frishman, D. and P. Argos (1995). "Knowledge-based protein secondary structure assignment." Proteins **23**(4): 566-79.

Giesen, K., K. Radsak, et al. (2000). "The potential terminase subunit of human cytomegalovirus, pUL56, is translocated into the nucleus by its own nuclear localization signal and interacts with importin alpha." J Gen Virol **81**(Pt 9): 2231-44.

Giesen, K., K. Radsak, et al. (2000). "Targeting of the gene product encoded by ORF UL56 of human cytomegalovirus into viral replication centers." FEBS Lett **471**(2-3): 215-8.

Gilbert, C. and G. Boivin (2005). "Human cytomegalovirus resistance to antiviral drugs." Antimicrob Agents Chemother **49**(3): 873-83.

Glaser, F., D. M. Steinberg, et al. (2001). "Residue frequencies and pairing preferences at protein-protein interfaces." Proteins **43**(2): 89-102.

Gorlich, D. (1997). "Nuclear protein import." Curr Opin Cell Biol **9**(3): 412-9.

Grantham, R. (1974). "Amino acid difference formula to help explain protein evolution." Science **185**(4154): 862-4.

Graves-Woodward, K. L., J. Gottlieb, et al. (1997). "Biochemical analyses of mutations in the HSV-1 helicase-primase that alter ATP hydrolysis, DNA unwinding, and coupling between hydrolysis and unwinding." J Biol Chem **272**(7): 4623-30.

Gretch, D. R., B. Kari, et al. (1988). "Identification and characterization of three distinct families of glycoprotein complexes in the envelopes of human cytomegalovirus." J Virol **62**(3): 875-81.

Guasch, A., J. Pous, et al. (2002). "Detailed architecture of a DNA translocating machine: the high-resolution structure of the bacteriophage phi29 connector particle." J Mol Biol **315**(4): 663-76.

Gudmundsson, K. S., G. A. Freeman, et al. (2000). "Synthesis of fluorosugar analogues of 2,5,6-trichloro-1-(beta-D-ribofuranosyl)benzimidazole as antivirals with potentially increased glycosidic bond stability." J Med Chem **43**(12): 2473-8.

Hanks, S. K. and T. Hunter (1995). "Protein kinases 6. The eukaryotic protein kinase superfamily: kinase (catalytic) domain structure and classification." Faseb J **9**(8): 576-96.

Hantz, S., A. Couvreux, et al. (2009). "Conserved domains and structure prediction of human cytomegalovirus UL27 protein." Antivir Ther **14**(5): 663-72.

Harbury, P. B., T. Zhang, et al. (1993). "A switch between two-, three-, and four-stranded coiled coils in GCN4 leucine zipper mutants." Science **262**(5138): 1401-7.

Harding, M. M. (1999). "The geometry of metal-ligand interactions relevant to proteins." Acta Crystallogr D Biol Crystallogr **55**(Pt 8): 1432-43.

Helles, G. (2008). "A comparative study of the reported performance of ab initio protein structure prediction algorithms." J R Soc Interface **5**(21): 387-96.

Henikoff, J. G., E. A. Greene, et al. (2000). "Increased coverage of protein families with the blocks database servers." Nucleic Acids Res **28**(1): 228-30.

Henikoff, S. and J. G. Henikoff (1992). "Amino acid substitution matrices from protein blocks." Proc Natl Acad Sci U S A **89**(22): 10915-9.

Hoffman, V. F. and D. J. Skiest (2000). "Therapeutic developments in cytomegalovirus retinitis." Expert Opin Investig Drugs **9**(2): 207-20.

Hu, Z., B. Ma, et al. (2000). "Conservation of polar residues as hot spots at protein interfaces." Proteins **39**(4): 331-42.

Huber, M. T. and T. Compton (1999). "Intracellular formation and processing of the heterotrimeric gH-gL-gO (gCIII) glycoprotein envelope complex of human cytomegalovirus." J Virol **73**(5): 3886-92.

Hurst, H. C. (1994). "Transcription factors. 1: bZIP proteins." Protein Profile **1**(2): 123-68.

Hwang, J. S. and E. Bogner (2002). "ATPase activity of the terminase subunit pUL56 of human cytomegalovirus." J Biol Chem **277**(9): 6943-8.

Hwang, J. S., O. Kregler, et al. (2007). "Identification of acetylated, tetrahalogenated benzimidazole D-ribonucleosides with enhanced activity against human cytomegalovirus." J Virol **81**(21): 11604-11.

Ilangovan, U., J. M. Wojciak, et al. (1999). "NMR structure and functional studies of the Mu repressor DNA-binding domain." Biochemistry **38**(26): 8367-76.

Innis, C. A., J. Shi, et al. (2000). "Evolutionary trace analysis of TGF-beta and related growth factors: implications for site-directed mutagenesis." Protein Eng **13**(12): 839-47.

Isidro, A., A. O. Henriques, et al. (2004). "The portal protein plays essential roles at different steps of the SPP1 DNA packaging process." Virology **322**(2): 253-63.

Iwasenko, J. M., G. M. Scott, et al. (2007). "Emergence and persistence of multiple antiviral-resistant CMV strains in a highly immunocompromised child." J Clin Virol **40**(2): 152-5.

Janin, J. and C. Chothia (1990). "The structure of protein-protein recognition sites." J Biol Chem **265**(27): 16027-30.

Jarvis, M. A., T. R. Jones, et al. (2004). "Phosphorylation of human cytomegalovirus glycoprotein B (gB) at the acidic cluster casein kinase 2 site (Ser900) is required for localization of gB to the trans-Golgi network and efficient virus replication." J Virol **78**(1): 285-93.

Jayaraman, G., T. K. Kumar, et al. (1996). "2,2,2-Trifluoroethanol induces helical conformation in an all beta-sheet protein." Biochem Biophys Res Commun **222**(1): 33-7.

Jesionek, A. and B. Kielemenoglou (1904). "Uber einen befund von protozoenartigen gebilden in

den organen eines heriditarluetischen fotus." Munch. Med. Wochenschr **51**: 1905-07.

Jones, D. T. (1999). "Protein secondary structure prediction based on position-specific scoring matrices." J Mol Biol **292**(2): 195-202.

Jones, S. and J. M. Thornton (1996). "Principles of protein-protein interactions." Proc Natl Acad Sci U S A **93**(1): 13-20.

Junius, F. K., J. P. Mackay, et al. (1995). "Nuclear magnetic resonance characterization of the Jun leucine zipper domain: unusual properties of coiled-coil interfacial polar residues." Biochemistry **34**(18): 6164-74.

Junius, F. K., S. I. O'Donoghue, et al. (1996). "High resolution NMR solution structure of the leucine zipper domain of the c-Jun homodimer." J Biol Chem **271**(23): 13663-7.

Kamil, J. P. and D. M. Coen (2007). "Human cytomegalovirus protein kinase UL97 forms a complex with the tegument phosphoprotein pp65." J Virol **81**(19): 10659-68.

Kari, B. and R. Gehrz (1992). "A human cytomegalovirus glycoprotein complex designated gC-II is a major heparin-binding component of the envelope." J Virol **66**(3): 1761-4.

Kelley, L. A. and M. J. Sternberg (2009). "Protein structure prediction on the Web: a case study using the Phyre server." Nat Protoc **4**(3): 363-71.

Kern, E. R., C. B. Hartline, et al. (2004). "Activities of benzimidazole D- and L-ribonucleosides in animal models of cytomegalovirus infections." Antimicrob Agents Chemother **48**(5): 1749-55.

Keskin, O., B. Ma, et al. (2005). "Hot regions in protein--protein interactions: the organization and contribution of structurally conserved hot spot residues." J Mol Biol **345**(5): 1281-94.

King, G. F. (1996). "NMR spectroscopy and X-ray crystallography provide complementary information on the structure and dynamics of leucine zippers." Biophys J **71**(2): 1152-4.

Kmiecik, S. and A. Kolinski (2007). "Characterization of protein-folding pathways by reduced-space modeling." Proc Natl Acad Sci U S A **104**(30): 12330-5.

Komazin-Meredith, G., R. J. Petrella, et al. (2008). "The human cytomegalovirus UL44 C clamp wraps around DNA." Structure **16**(8): 1214-25.

Komazin, G., R. G. Ptak, et al. (2003). "Resistance of human cytomegalovirus to the benzimidazole L-ribonucleoside maribavir maps to UL27." J Virol **77**(21): 11499-506.

Konig, P. and T. J. Richmond (1993). "The X-ray structure of the GCN4-bZIP bound to ATF/CREB site DNA shows the complex depends on DNA flexibility." J Mol Biol **233**(1): 139-54.

Kosturko, L. D., E. Daub, et al. (1989). "The interaction of E. coli integration host factor and lambda cos DNA: multiple complex formation and protein-induced bending." Nucleic Acids Res **17**(1): 317-34.

Koszalka, G. W., N. W. Johnson, et al. (2002). "Preclinical and toxicology studies of 1263W94, a potent and selective inhibitor of human cytomegalovirus replication." Antimicrob Agents Chemother **46**(8): 2373-80.

Kouzarides, T., A. T. Bankier, et al. (1987). "Sequence and transcription analysis of the human cytomegalovirus DNA polymerase gene." J Virol **61**(1): 125-33.

Krishna, S. S., I. Majumdar, et al. (2003). "Structural classification of zinc fingers: survey and summary." Nucleic Acids Res **31**(2): 532-50.

Krosky, P. M., M. C. Baek, et al. (2003). "The human cytomegalovirus UL97 protein kinase, an antiviral drug target, is required at the stage of nuclear egress." J Virol **77**(2): 905-14.

Krosky, P. M., M. C. Baek, et al. (2003). "The human cytomegalovirus UL44 protein is a substrate for the UL97 protein kinase." J Virol **77**(14): 7720-7.

Krosky, P. M., M. R. Underwood, et al. (1998). "Resistance of human cytomegalovirus to benzimidazole ribonucleosides maps to two open reading frames: UL89 and UL56." J Virol **72**(6): 4721-8.

Kuebler, D. and V. B. Rao (1998). "Functional analysis of the DNA-packaging/terminase protein gp17 from bacteriophage T4." J Mol Biol **281**(5): 803-14.

Lachenmann, M. J., J. E. Ladbury, et al. (2004). "Why zinc fingers prefer zinc: ligand-field symmetry and the hidden thermodynamics of metal ion selectivity." Biochemistry **43**(44): 13910-25.

Laity, J. H., B. M. Lee, et al. (2001). "Zinc finger proteins: new insights into structural and functional diversity." Curr Opin Struct Biol **11**(1): 39-46.

Lange, A., R. E. Mills, et al. (2007). "Classical nuclear localization signals: definition, function, and interaction with importin alpha." J Biol Chem **282**(8): 5101-5.

Larder, B. A., S. D. Kemp, et al. (1987). "Related functional domains in virus DNA polymerases." Embo J **6**(1): 169-75.

Lebedev, A. A., M. H. Krause, et al. (2007). "Structural framework for DNA translocation via the viral portal protein." Embo J **26**(7): 1984-94.

Lee, S. and F. T. Tsai (2005). "Molecular chaperones in protein quality control." J Biochem Mol Biol **38**(3): 259-65.

Lichtarge, O., H. R. Bourne, et al. (1996). "An evolutionary trace method defines binding surfaces common to protein families." J Mol Biol **257**(2): 342-58.

Lin, H., V. B. Rao, et al. (1999). "Analysis of capsid portal protein and terminase functional domains: interaction sites required for DNA packaging in bacteriophage T4." J Mol Biol **289**(2): 249-60.

Littler, E., A. D. Stuart, et al. (1992). "Human cytomegalovirus UL97 open reading frame encodes a protein that phosphorylates the antiviral nucleoside analogue ganciclovir." Nature **358**(6382): 160-2.

Liu, S., J. D. Knafels, et al. (2006). "Crystal structure of the herpes simplex virus 1 DNA polymerase." J Biol Chem **281**(26): 18193-200.

Liu, T., M. Ryan, et al. (1997). "Determination of pKa values of the histidine side chains of phosphatidylinositol-specific phospholipase C from Bacillus cereus by NMR spectroscopy and site-directed mutagenesis." Protein Sci **6**(9): 1937-44.

Lo Conte, L., C. Chothia, et al. (1999). "The atomic structure of protein-protein recognition sites." J Mol Biol **285**(5): 2177-98.

Loewenthal, R., J. Sancho, et al. (1993). "Long-range surface charge-charge interactions in proteins. Comparison of experimental results with calculations from a theoretical method." J Mol Biol **232**(2): 574-83.

Lopper, M. and T. Compton (2002). "Disulfide bond configuration of human cytomegalovirus glycoprotein B." J Virol **76**(12): 6073-82.

Lorenzi, P. L., C. P. Landowski, et al. (2006). "N-methylpurine DNA glycosylase and 8-oxoguanine dna glycosylase metabolize the antiviral nucleoside 2-bromo-5,6-dichloro-1-(beta-D-ribofuranosyl)benzimidazole." Drug Metab Dispos **34**(6): 1070-7.

Lorenzi, P. L., C. P. Landowski, et al. (2005). "Amino acid ester prodrugs of 2-bromo-5,6-dichloro-1-(beta-D-ribofuranosyl)benzimidazole enhance metabolic stability in vitro and in vivo." J Pharmacol Exp Ther **314**(2): 883-90.

Mach, M., B. Kropff, et al. (2005). "Complex formation by glycoproteins M and N of human cytomegalovirus: structural and functional aspects." J Virol **79**(4): 2160-70.

Madhusudan, E. A. Trafny, et al. (1994). "cAMP-dependent protein kinase: crystallographic insights into substrate recognition and phosphotransfer." Protein Sci **3**(2): 176-87.

Marfori, J. E., M. M. Exner, et al. (2007). "Development of new cytomegalovirus UL97 and DNA polymerase mutations conferring drug resistance after valganciclovir therapy in allogeneic stem cell recipients." J Clin Virol **38**(2): 120-5.

Marschall, M., M. Freitag, et al. (2003). "The protein kinase pUL97 of human cytomegalovirus interacts with and phosphorylates the DNA polymerase processivity factor pUL44." Virology **311**(1): 60-71.

Marschall, M., A. Marzi, et al. (2005). "Cellular p32 recruits cytomegalovirus kinase pUL97 to redistribute the nuclear lamina." J Biol Chem **280**(39): 33357-67.

Marschall, M., M. Stein-Gerlach, et al. (2002). "Direct targeting of human cytomegalovirus protein kinase pUL97 by kinase inhibitors is a novel principle for antiviral therapy." J Gen Virol **83**(Pt 5): 1013-23.

Matthews, T. and R. Boehme (1988). "Antiviral activity and mechanism of action of ganciclovir." Rev Infect Dis **10 Suppl 3**: S490-4.

McGeoch, D. J., F. J. Rixon, et al. (2006). "Topics in herpesvirus genomics and evolution." Virus Research **117**: 90-104.

McSharry, J. J., A. McDonough, et al. (2001). "Susceptibilities of human cytomegalovirus clinical isolates to BAY38-4766, BAY43-9695, and ganciclovir." Antimicrob Agents Chemother **45**(10): 2925-7.

Mettenleiter, T. C., B. G. Klupp, et al. (2006). "Herpesvirus assembly: a tale of two membranes." Curr Opin Microbiol **9**(4): 423-9.

Michel, D. and T. Mertens (2004). "The UL97 protein kinase of human cytomegalovirus and homologues in other herpesviruses: impact on virus and host." Biochim Biophys Acta **1697**(1-2): 169-80.

Michel, D., I. Pavic, et al. (1996). "The UL97 gene product of human cytomegalovirus is an early-late protein with a nuclear localization but is not a nucleoside kinase." J Virol **70**(9): 6340-6.

Mitchell, M. S., S. Matsuzaki, et al. (2002). "Sequence analysis of bacteriophage T4 DNA packaging/terminase genes 16 and 17 reveals a common ATPase center in the large subunit of viral terminases." Nucleic Acids Res **30**(18): 4009-21.

Mitchell, M. S. and V. B. Rao (2006). "Functional analysis of the bacteriophage T4 DNA-packaging ATPase motor." J Biol Chem **281**(1): 518-27.

Moreira, I. S., P. A. Fernandes, et al. (2007). "Hot spot occlusion from bulk water: a comprehensive study of the complex between the lysozyme HEL and the antibody FVD1.3." J Phys Chem B **111**(10): 2697-706.

Moreira, I. S., P. A. Fernandes, et al. (2007). "Hot spots--a review of the protein-protein interface determinant amino-acid residues." Proteins **68**(4): 803-12.

Morellet, N., S. Bouaziz, et al. (2003). "NMR structure of the HIV-1 regulatory protein VPR." J Mol Biol **327**(1): 215-27.

Morellet, N., N. Jullian, et al. (1992). "Determination of the structure of the nucleocapsid protein NCp7 from the human immunodeficiency virus type 1 by 1H NMR." Embo J **11**(8): 3059-65.

Morita, M., M. Tasaka, et al. (1995). "Analysis of the fine structure of the prohead binding domain of the packaging protein of bacteriophage T3 using a hexapeptide, an analog of a prohead binding site." Virology **211**(2): 516-24.

Needleman, S. B. and C. D. Wunsch (1970). "A general method applicable to the search for similarities in the amino acid sequence of two proteins." J Mol Biol **48**(3): 443-53.

Nellissery, J. K., R. Szczepaniak, et al. (2007). "A putative leucine zipper within the herpes simplex virus type 1 UL6 protein is required for portal ring formation." J Virol **81**(17): 8868-77.

Nemecek, D., E. B. Gilcrease, et al. (2007). "Subunit conformations and assembly states of a DNA-translocating motor: the terminase of bacteriophage P22." J Mol Biol **374**(3): 817-36.

Nemecek, D., G. C. Lander, et al. (2008). "Assembly architecture and DNA binding of the bacteriophage P22 terminase small subunit." J Mol Biol **383**(3): 494-501.

Neuvirth, H., R. Raz, et al. (2004). "ProMate: a structure based prediction program to identify the location of protein-protein binding sites." J Mol Biol **338**(1): 181-99.

Newcomb, W. W., F. L. Homa, et al. (1994). "Cell-free assembly of the herpes simplex virus capsid." J Virol **68**(9): 6059-63.

Ni, M. and A. S. Lee (2007). "ER chaperones in mammalian development and human diseases." FEBS Lett **581**(19): 3641-51.

Nowotny, M., S. M. Cerritelli, et al. (2008). "Specific recognition of RNA/DNA hybrid and enhancement of human RNase H1 activity by HBD." Embo J **27**(7): 1172-81.

Nowotny, M., S. A. Gaidamakov, et al. (2005). "Crystal structures of RNase H bound to an RNA/DNA hybrid: substrate specificity and metal-dependent catalysis." Cell **121**(7): 1005-16.

Nowotny, M., S. A. Gaidamakov, et al. (2007). "Structure of human RNase H1 complexed with an RNA/DNA hybrid: insight into HIV reverse transcription." Mol Cell **28**(2): 264-76.

Ofran, Y. and B. Rost (2003). "Analysing six types of protein-protein interfaces." J Mol Biol **325**(2): 377-87.

Ofran, Y. and B. Rost (2003). "Predicted protein-protein interaction sites from local sequence information." FEBS Lett **544**(1-3): 236-9.

Ofran, Y. and B. Rost (2007). "Protein-Protein Interaction Hotspots Carved into Sequences." PLoS Comput Biol **3**(7): e119.

Oliveira, L., A. O. Henriques, et al. (2006). "Modulation of the viral ATPase activity by the portal protein correlates with DNA packaging efficiency." J Biol Chem **281**(31): 21914-23.

Park, M. Y., Y. E. Kim, et al. (2006). "Interactions among four proteins encoded by the human cytomegalovirus UL112-113 region regulate their intranuclear targeting and the recruitment of UL44 to prereplication foci." J Virol **80**(6): 2718-27.

Paterson, D. A., A. P. Dyer, et al. (2002). "A role for human cytomegalovirus glycoprotein O (gO) in cell fusion and a new hypervariable locus." Virology **293**(2): 281-94.

Pevarello, P., M. G. Brasca, et al. (2004). "3-Aminopyrazole inhibitors of CDK2/cyclin A as antitumor agents. 1. Lead finding." J Med Chem **47**(13): 3367-80.

Pingoud, A., M. Fuxreiter, et al. (2005). "Type II restriction endonucleases: structure and mechanism." Cell Mol Life Sci **62**(6): 685-707.

Ponchon, L., P. Boulanger, et al. (2006). "The endonuclease domain of bacteriophage terminases belongs to the resolvase/integrase/ribonuclease H superfamily: a bioinformatics analysis validated by a functional study on bacteriophage T5." J Biol Chem **281**(9): 5829-36.

Prichard, M. N., W. J. Britt, et al. (2005). "Human cytomegalovirus UL97 Kinase is required for the normal intranuclear distribution of pp65 and virion morphogenesis." J Virol **79**(24): 15494-502.

Purkiss, A., S. Skoulakis, et al. (2001). "The protein-solvent interface: a big splash." Phil. Trans. R. Soc. London. **359**: 1515-1527.

Ramachandran, G. N., C. Ramakrishnan, et al. (1963). "Stereochemistry of polypeptide chain configurations." J Mol Biol **7**: 95-9.

Rao, V. B. and M. Feiss (2008). "The bacteriophage DNA packaging motor." Annu Rev Genet **42**: 647-81.

Reefschlaeger, J., W. Bender, et al. (2001). "Novel non-nucleoside inhibitors of cytomegaloviruses (BAY 38-4766): in vitro and in vivo antiviral activity and mechanism of action." J Antimicrob Chemother **48**(6): 757-67.

Rentas, F. J. and V. B. Rao (2003). "Defining the bacteriophage T4 DNA packaging machine: evidence for a C-terminal DNA cleavage domain in the large terminase/packaging protein gp17." J Mol Biol **334**(1): 37-52.

Reusser, P. (1996). "Herpesvirus resistance to antiviral drugs: a review of the mechanisms, clinical importance and therapeutic options." J Hosp Infect **33**(4): 235-48.

Rey, F. (2006). "Molecular gymnastics at the herpesvirus surface." EMBO **7**(10): 1000-1005.

Rodier, F., R. P. Bahadur, et al. (2005). "Hydration of protein-protein interfaces." Proteins **60**(1): 36-45.

Romaker, D., V. Schregel, et al. (2006). "Analysis of the structure-activity relationship of four herpesviral UL97 subfamily protein kinases reveals partial but not full functional conservation." J Med Chem **49**(24): 7044-53.

Rowe, W. P., J. W. Hartley, et al. (1956). "Cytopathogenic agent resembling human salivary gland virus recovered from tissue cultures of human adenoids." Proc Soc Exp Biol Med **92**(2): 418-24.

Sali, A. and T. L. Blundell (1993). "Comparative protein modelling by satisfaction of spatial restraints." J Mol Biol **234**(3): 779-815.

Savva, C. G., A. Holzenburg, et al. (2004). "Insights into the structure of human cytomegalovirus large terminase subunit pUL56." FEBS Lett **563**(1-3): 135-40.

Scheffczik, H., C. G. Savva, et al. (2002). "The terminase subunits pUL56 and pUL89 of human cytomegalovirus are DNA-metabolizing proteins with toroidal structure." Nucleic Acids Res **30**(7): 1695-703.

Schleiss, M. R., D. I. Bernstein, et al. (2005). "The non-nucleoside antiviral, BAY 38-4766, protects against cytomegalovirus (CMV) disease and mortality in immunocompromised guinea pigs." Antiviral Res **65**(1): 35-43.

Scholz, B., S. Rechter, et al. (2003). "Identification of the ATP-binding site in the terminase subunit pUL56 of human cytomegalovirus." Nucleic Acids Res **31**(5): 1426-33.

Schregel, V., S. Auerochs, et al. (2007). "Mapping of a self-interaction domain of the cytomegalovirus protein kinase pUL97." J Gen Virol **88**(Pt 2): 395-404.

Schumacher, M. A., R. H. Goodman, et al. (2000). "The structure of a CREB bZIP.somatostatin CRE complex reveals the basis for selective dimerization and divalent cation-enhanced DNA binding." J Biol Chem **275**(45): 35242-7.

Shukla, D. and P. G. Spear (2001). "Herpesviruses and heparan sulfate: an intimate relationship in aid of viral entry." J Clin Invest **108**(4): 503-10.

Silva, P. J. (2008). "Assessing the reliability of sequence similarities detected through hydrophobic cluster analysis." Proteins **70**(4): 1588-94.

Singleton, M. R., M. S. Dillingham, et al. (2007). "Structure and mechanism of helicases and nucleic acid translocases." Annu Rev Biochem **76**: 23-50.

Sinz, A. (2006). "Chemical cross-linking and mass spectrometry to map three-dimensional protein structures and protein-protein interactions." Mass Spectrom Rev **25**(4): 663-82.

Skamnaki, V. T., D. J. Owen, et al. (1999). "Catalytic mechanism of phosphorylase kinase probed by mutational studies." Biochemistry **38**(44): 14718-30.

Slater, M. J., R. Baxter, et al. (2001). "Synthesis of N-alkyl substituted indolocarbazoles as potent inhibitors of human cytomegalovirus replication." Bioorg Med Chem Lett **11**(15): 1993-5.

Slater, M. J., S. Cockerill, et al. (1999). "Indolocarbazoles: potent, selective inhibitors of human cytomegalovirus replication." Bioorg Med Chem **7**(6): 1067-74.

Smith, M. G. (1956). "Propagation in tissue culture of cytopathogenic virus human salivary gland virus disease." Proc. Soc. Exp. Biol. Med. **92**: 424-30.

Smits, C., M. Chechik, et al. (2009). "Structural basis for the nuclease activity of a bacteriophage large terminase." EMBO Rep **10**(6): 592-8.

Song, X., B. S. Vig, et al. (2005). "Amino acid ester prodrugs of the antiviral agent 2-bromo-5,6-dichloro-1-(beta-D-ribofuranosyl)benzimidazole as potential substrates of hPEPT1 transporter." J Med Chem **48**(4): 1274-7.

Sorokin, A. V., E. R. Kim, et al. (2007). "Nucleocytoplasmic transport of proteins." Biochemistry (Mosc) **72**(13): 1439-57.

Sowa, M. E., W. He, et al. (2001). "Prediction and confirmation of a site critical for effector regulation of RGS domain activity." Nat Struct Biol **8**(3): 234-7.

Spaete, R. R. and N. Frenkel (1985). "The herpes simplex virus amplicon: analyses of cis-acting replication functions." Proc Natl Acad Sci U S A **82**(3): 694-8.

Spaete, R. R. and E. S. Mocarski (1985). "The alpha sequence of the cytomegalovirus genome functions as a cleavage/packaging signal for herpes simplex virus defective genomes." J Virol **54**(3): 817-24.

Sprangers, R. and L. E. Kay (2007). "Quantitative dynamics and binding studies of the 20S proteasome by NMR." Nature **445**(7128): 618-22.

Steitz, T. A. (1999). "DNA polymerases: structural diversity and common mechanisms." J Biol Chem **274**(25): 17395-8.

Sun, S., K. Kondabagil, et al. (2008). "The structure of the phage T4 DNA packaging motor suggests a mechanism dependent on electrostatic forces." Cell **135**(7): 1251-62.

Sun, S., K. Kondabagil, et al. (2007). "The structure of the ATPase that powers DNA packaging into bacteriophage T4 procapsids." Mol Cell **25**(6): 943-9.

Thoma, C., E. Borst, et al. (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.

Thomsen, D. R., L. L. Roof, et al. (1994). "Assembly of herpes simplex virus (HSV) intermediate capsids in insect cells infected with recombinant baculoviruses expressing HSV capsid proteins." J Virol **68**(4): 2442-57.

Toth, Z. and T. Stamminger (2008). "The human cytomegalovirus regulatory protein UL69 and its effect on mRNA export." Front Biosci **13**: 2939-49.

Townsend, L. B., R. V. Devivar, et al. (1995). "Design, synthesis, and antiviral activity of certain 2,5,6-trihalo-1-(beta-D-ribofuranosyl)benzimidazoles." J Med Chem **38**(20): 4098-105.

Townsend, L. B., K. S. Gudmundsson, et al. (1999). "Studies designed to increase the stability and antiviral activity (HCMV) of the active benzimidazole nucleoside, TCRB." Nucleosides Nucleotides **18**(4-5): 509-19.

Tsai, C. J., S. L. Lin, et al. (1997). "Studies of protein-protein interfaces: a statistical analysis of the hydrophobic effect." Protein Sci **6**(1): 53-64.

Tsai, C. J. and R. Nussinov (2001). "The building block folding model and the kinetics of protein folding." Protein Eng **14**(10): 723-33.

Tuteja, N. and R. Tuteja (2004). "Unraveling DNA helicases. Motif, structure, mechanism and function." Eur J Biochem **271**(10): 1849-63.

Underwood, M. R., R. G. Ferris, et al. (2004). "Mechanism of action of the ribopyranoside benzimidazole GW275175X against human cytomegalovirus." Antimicrob Agents Chemother **48**(5): 1647-51.

Underwood, M. R., R. J. Harvey, et al. (1998). "Inhibition of human cytomegalovirus DNA maturation by a benzimidazole ribonucleoside is mediated through the UL89 gene product." J Virol **72**(1): 717-25.

van Zeijl, M., J. Fairhurst, et al. (1997). "The human cytomegalovirus UL97 protein is phosphorylated and a component of virions." Virology **231**(1): 72-80.

Velankar, S. S., P. Soultanas, et al. (1999). "Crystal structures of complexes of PcrA DNA helicase with a DNA substrate indicate an inchworm mechanism." Cell **97**(1): 75-84.

Visalli, R. J. and M. van Zeijl (2003). "DNA encapsidation as a target for anti-herpesvirus drug therapy." Antiviral Res **59**(2): 73-87.

Vogel, C., M. Bashton, et al. (2004). "Structure, function and evolution of multidomain proteins." Curr Opin Struct Biol **14**(2): 208-16.

Walter, R. L., S. E. Ealick, et al. (1996). "Multiple wavelength anomalous diffraction (MAD) crystal structure of rusticyanin: a highly oxidizing cupredoxin with extreme acid stability." J Mol Biol **263**(5): 730-51.

Wang, G. and R. L. Dunbrack, Jr. (2003). "PISCES: a protein sequence culling server." Bioinformatics **19**(12): 1589-91.

Wang, J., A. N. Loveland, et al. (2006). "High-molecular-weight protein (pUL48) of human cytomegalovirus is a competent deubiquitinating protease: mutant viruses altered in its active-site cysteine or histidine are viable." J Virol **80**(12): 6003-12.

Wang, J., A. K. Sattar, et al. (1997). "Crystal structure of a pol alpha family replication DNA polymerase from bacteriophage RB69." Cell **89**(7): 1087-99.

Weir, J. P. (1998). Genomic organization and evolution of the human herpesviruses. Virus Genes. **16:** 85-93.

Weller, T. H., J. B. Hanshaw, et al. (1960). "Serologic differentiation of viruses responsible for cytomegalic inclusion disease." Virology **12**: 130-2.

Westman, B. J., J. Perdomo, et al. (2004). "Structural studies on a protein-binding zinc-finger domain of Eos reveal both similarities and differences to classical zinc fingers." Biochemistry **43**(42): 13318-27.

Wiederstein, M. and M. J. Sippl (2007). "ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins." Nucleic Acids Res **35**(Web Server issue): W407-10.

Williams, J. D., R. G. Ptak, et al. (2004). "Synthesis, antiviral activity, and mode of action of some 3-substituted 2,5,6-trichloroindole 2'- and 5'-deoxyribonucleosides." J Med Chem **47**(23): 5773-82.

Williams, S. L., C. B. Hartline, et al. (2003). "In vitro activities of benzimidazole D- and L-ribonucleosides against herpesviruses." Antimicrob Agents Chemother **47**(7): 2186-92.

Wishart, D. S., B. D. Sykes, et al. (1992). "The chemical shift index: a fast and simple method for the assignment of protein secondary structure through NMR spectroscopy." Biochemistry **31**(6): 1647-51.

Wolf, D. G., C. T. Courcelle, et al. (2001). "Distinct and separate roles for herpesvirus-conserved UL97 kinase in cytomegalovirus DNA synthesis and encapsidation." Proc Natl Acad Sci U S A **98**(4): 1895-900.

Worthington, M. T., B. T. Amann, et al. (1996). "Metal binding properties and secondary structure of the zinc-binding domain of Nup475." Proc Natl Acad Sci U S A **93**(24): 13754-9.

Wu, S. and Y. Zhang (2008). "ANGLOR: a composite machine-learning algorithm for protein backbone torsion angle prediction." PLoS One **3**(10): e3400.

Wu, S. and Y. Zhang (2008). "MUSTER: Improving protein sequence profile-profile alignments by using multiple sources of structure information." Proteins **72**(2): 547-56.

Wu, W. F., S. Christiansen, et al. (1988). "Domains for protein-protein interactions at the N and C termini of the large subunit of bacteriophage lambda terminase." Genetics **119**(3): 477-84.

Xiong, X., J. L. Smith, et al. (1997). "Effect of incorporation of cidofovir into DNA by human cytomegalovirus DNA polymerase on DNA elongation." Antimicrob Agents Chemother **41**(3): 594-9.

Yang, K., F. Homa, et al. (2007). "Putative terminase subunits of herpes simplex virus 1 form a complex in the cytoplasm and interact with portal protein in the nucleus." J Virol **81**(12): 6419-33.

Yang, W. (2008). "An equivalent metal ion in one- and two-metal-ion catalysis." Nat Struct Mol Biol **15**(11): 1228-31.

Yang, W., J. Y. Lee, et al. (2006). "Making and breaking nucleic acids: two-Mg2+-ion catalysis and substrate specificity." Mol Cell **22**(1): 5-13.

Ye, L. B. and E. S. Huang (1993). "In vitro expression of the human cytomegalovirus DNA polymerase gene: effects of sequence alterations on enzyme activity." J Virol **67**(11): 6339-47.

Yeo, A. and M. Feiss (1995). "Specific interaction of terminase, the DNA packaging enzyme of bacteriophage lambda, with the portal protein of the prohead." J Mol Biol **245**(2): 141-50.

Young, L., R. L. Jernigan, et al. (1994). "A role for surface hydrophobicity in protein-protein recognition." Protein Sci **3**(5): 717-29.

Zhang, Y. (2008). "I-TASSER server for protein 3D structure prediction." BMC Bioinformatics **9**: 40.

Zhang, Y. (2009). "Protein structure prediction: when is it useful?" Curr Opin Struct Biol **19**(2): 145-55.

Zhang, Y. and J. Skolnick (2004). "Scoring function for automated assessment of protein structure template quality." Proteins **57**(4): 702-10.

Zhu, S. and J. Tytgat (2004). "Evolutionary epitopes of Hsp90 and p23: implications for their interaction." Faseb J **18**(9): 940-7.

Zimmerman, J. M., N. Eliezer, et al. (1968). "The characterization of amino acid sequences in proteins by statistical methods." J Theor Biol **21**(2): 170-201.

Zimmermann, A., H. Wilts, et al. (2000). "Indolocarbazoles exhibit strong antiviral activity against human cytomegalovirus and are potent inhibitors of the pUL97 protein kinase." Antiviral Res **48**(1): 49-60.

Zou, R., E. Kawashima, et al. (2000). "Design, synthesis, and antiviral evaluation of 2-deoxy-D-ribosides of substituted benzimidazoles as potential agents for human cytomegalovirus infections." Nucleosides Nucleotides Nucleic Acids **19**(1-2): 125-53.

Zuccola, H. J., D. J. Filman, et al. (2000). "The crystal structure of an unusual processivity factor, herpes simplex virus UL42, bound to the C terminus of its cognate polymerase." Mol Cell **5**(2): 267-78.

Zvelebil, M. J., G. J. Barton, et al. (1987). "Prediction of protein secondary structure and active sites using the alignment of homologous sequences." J Mol Biol **195**(4): 957-61.

# XI. ANNEXES

## A. Annexe A : Analyse Théorique Des Cibles Du Maribavir

Après avoir traité des cibles des dérivés D-ribobenzimidazolés, nous nous attachons ici à l'étude préliminaire des cibles du Maribavir, représentant des L-ribobenzimidazolés. Ainsi, deux protéines ont fait l'objet de notre attention: la phosphotransférase pUL97 et sa protéine accessoire pUL27. Les différents éléments tirés de ces études restent pour la plupart des hypothèses afin d'amorcer de plus amples investigations pour l'élucidation de leurs structures et de leurs fonctions.

### A.1. Annexe A1: Etude préliminaire de la phosphotransférase pUL97

La phosphotransférase de CMVH est au centre d'un grand nombre d'interactions durant le cycle d'infection du virus. Celles-ci incluent la maturation de protéines de CMVH telles que pUL44 pour sa reconnaissance par pUL54, mais aussi pp65 qui va être modifiée afin de s'agglutiner autour de la néocapside pour en former le tégument. La protéine pUL97 est aussi responsable de la déstabilisation de la membrane nucléaire après phosphorylation de la lamina cellulaire conduisant à l'exportation extranucléaire de la capside.

Nous présentons ici, une étude bioinformatique par modélisation du domaine catalytique de pUL97 de CMVH, cible du Maribavir.La connaissance du fonctionnement générale des phosphotransférases (Madhusudan, Trafny et al. 1994; Skamnaki, Owen et al. 1999) nous a permis d'avoir un point de vue critique sur des modèles générés par PHYRE.

#### A.1.a. Obtention d'un modèle théorique de pUL97.

Les différents sous-domaines de pUL97 au niveau de son site catalytique sont répertoriés à l'Annexe B5. La séquence complète de pUL97 (AAS48981) a été soumise au serveur PHYRE. Dix modèles du site catalytique de pUL97 ont été générés par *fold recognition* avec un taux de confiance de 100% estimé par PHYRE pour chacune d'entre elles (Figure 70).

Aucune de ces 10 structures ne regroupe l'intégralité des domaines fonctionnels (site de fixation de l'ATP, site de fixation du substrat et P-loop). Dans le meilleur des cas, un modèle a été construit les domaines fonctionneles par l'introduction de *gaps* dans la séquence, sans considérer l'intégralité de la séquence (modèle 2NRU_D).

Le seul modèle considérant un continuum est celui basé sur la structure PDB 1WHA qui ne présente pas le site complet de liaison au substrat et est classé 7/10 en fonction de son E-value, illustrant l'homologie entre la séquence cible et la structure-support.

En dehors d'un très bon recouvrement des modèles au niveau du site de liaison à l'ATP et de la P-loop, aucun modèle ne permet de visualiser avec une grande confiance le site de liaison à son substrat.

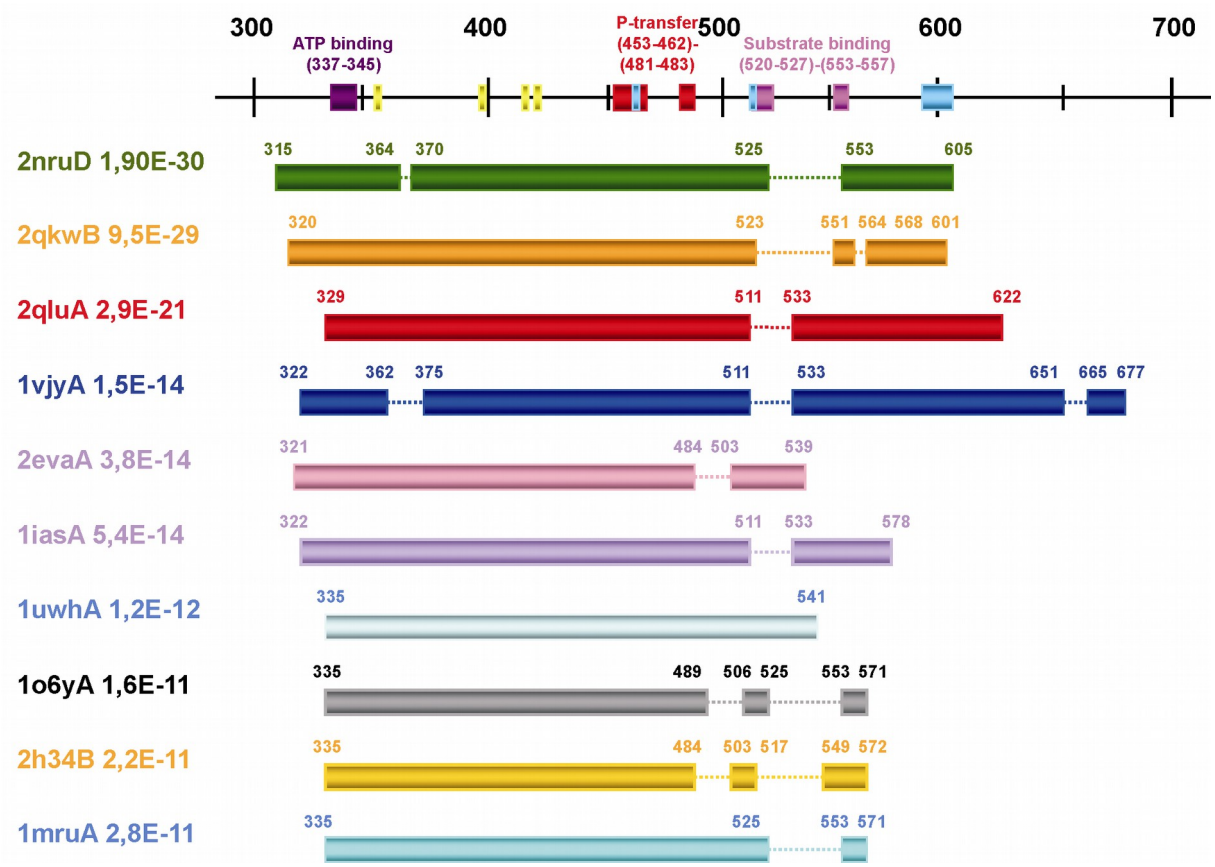



**Figure 70: Modèles générés par PHYRE.** En haut, la séquence de pUL97 avec en violet le site de liaison à l'ATP, en rouge, la boucle de transfert du phosphate (P-loop) et en magenta le site de liaison du substrat. Les segments jaunes et cyans correspondent au site de résistance au Maribavir et Ganciclovir, respectivement (Chou 2008). Les séquences des 10 modèles générés sont schématisées et les codes PDB des structures de référence précisées à gauche. Les pointillés représentent les parties de la séquence tronquées précisées par les numéros des acides aminés. Les séquences sont classées dela plus à la moins cohérente en fonction de leurs E-value précisées à coté des codes PDB des structure-supports.

Il a été montré que la délétion 590-607 rendait pUL97 résistante au Ganciclovir (Chou 2008). Ce qui signifie aussi qu'elle reste fonctionnelle sous cette forme. Donc, la séquence

complète de pUL97 tronquée du fragment 590-607 (pUL97Δ (590-607)) a été à son tour soumise à PHYRE. Là encore, et à partir des mêmes structures PDB, 10 modèles ont été générés (Figure 71). Contrairement au cas précédent, le meilleur modèle généré à partir de la structure- support 2NRU_D fait apparaître la totalité des sites fonctionnels. Ce modèle va donc servir de base pour les études structurales suivantes.

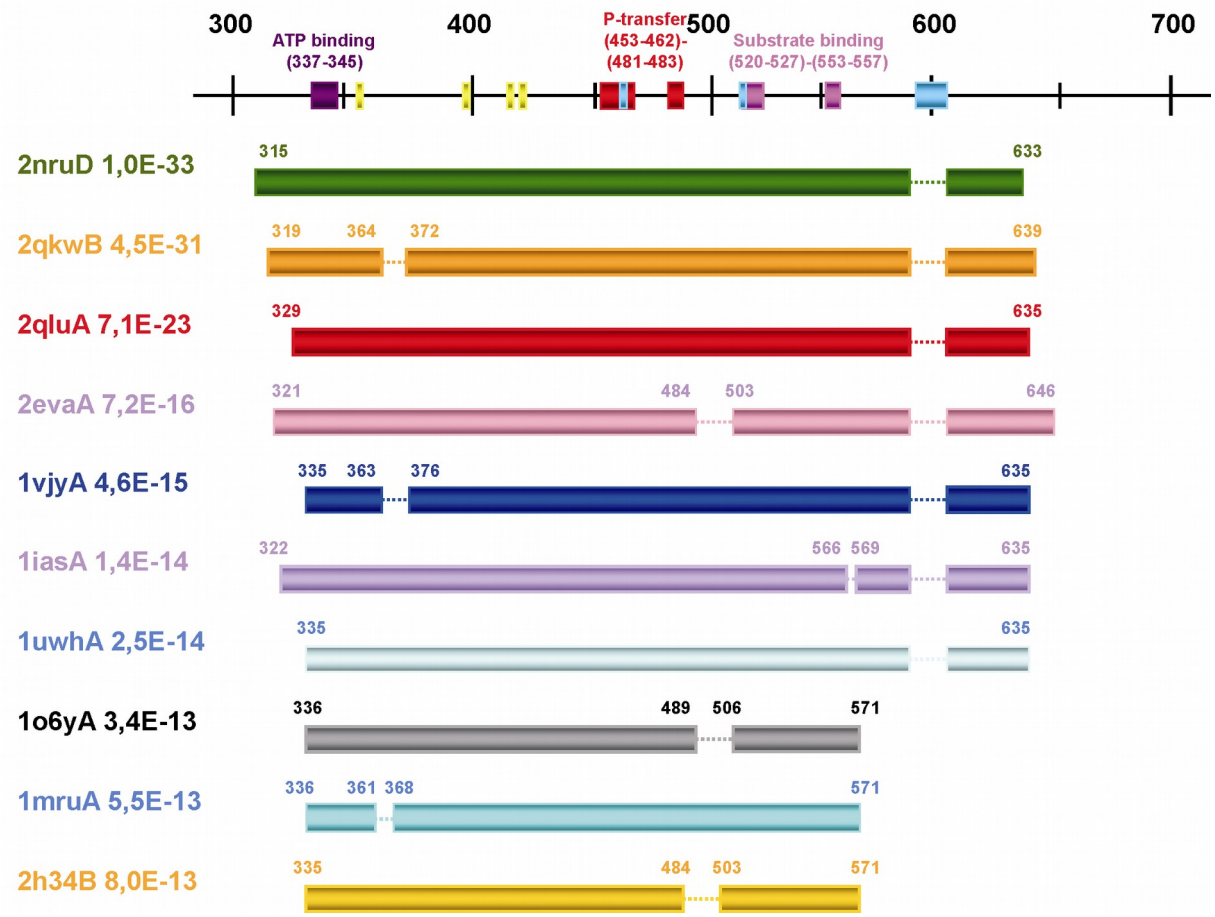



**Figure 71: Modèles générés par PHYRE sur le mutant de délétion pUL97Δ (590-607).** Même légende que précédemment. Les pointillés dans la zone 590-607 traduisent la délétion souhaitée qui n'est donc pas un artefact de reconstruction.

### A.1.b. Analyse de la structure pUL97Δ (590-607)_2nruA.

Comme l'ensemble des phosphotransférases, notre modèle est constitué d'un grand lobe portant le site de fixation du peptide-substrat et d'un petit lobe, poche de liaison de l'ATP (Figure 72A). Malgré les statistiques de PHYRE, plusieurs incohérences entre la littérature et notre modèle ne nous permettent pas d'exploiter cette structure théorique.

Tout d'abord, la proximité des grand et petit lobes peut empêcher le relargage de l'ATP et empêche toute étude de *docking* pour observer sa fixation ou celles des inhibiteurs de pUL97 dans le site actif (Figure 72A).

Deuxièmement, le repliement de pUL97 au niveau du site de fixation du substrat soulève lui aussi un problème. En effet, le sous-domaine IX (553-557) représentant ce site de reconnaissance du substrat semble très éloigné de la poche de fixation de l'ATP et donc ne peut rendre compte d'un mécanisme évident pour le transfert d'un groupement phosphate de l'ATP à une sérine ou une thréonine d'un peptide-substrat éventuel (Figure 72B).

Enfin, le positionnement de la lysine catalytique $Lys^{365}$ montrée expérimentalement comme cruciale dans la catalyse est lui aussi discutable. En effet, celle-ci est située au niveau du groupement adénine de l'ATP alors que son rôle est de ponter les phosphates β et γ pendant le transfert de ce dernier vers le substrat. Dans notre modèle c'est $Lys^{359}$ qui est favorablement située pour jouer ce rôle (Figure 72C).

Afin d'éviter ces problèmes, nous proposons d'ajuster manuellement les alignements de séquences afin de reconstituer le site actif (en particulier la lysine 365) en accord avec la bibliographie. Ceci pourra être contrôlé avec MODELLER. Enfin, le modèle généré devra être minimisé pour éliminer les contacts stériques éventuels.

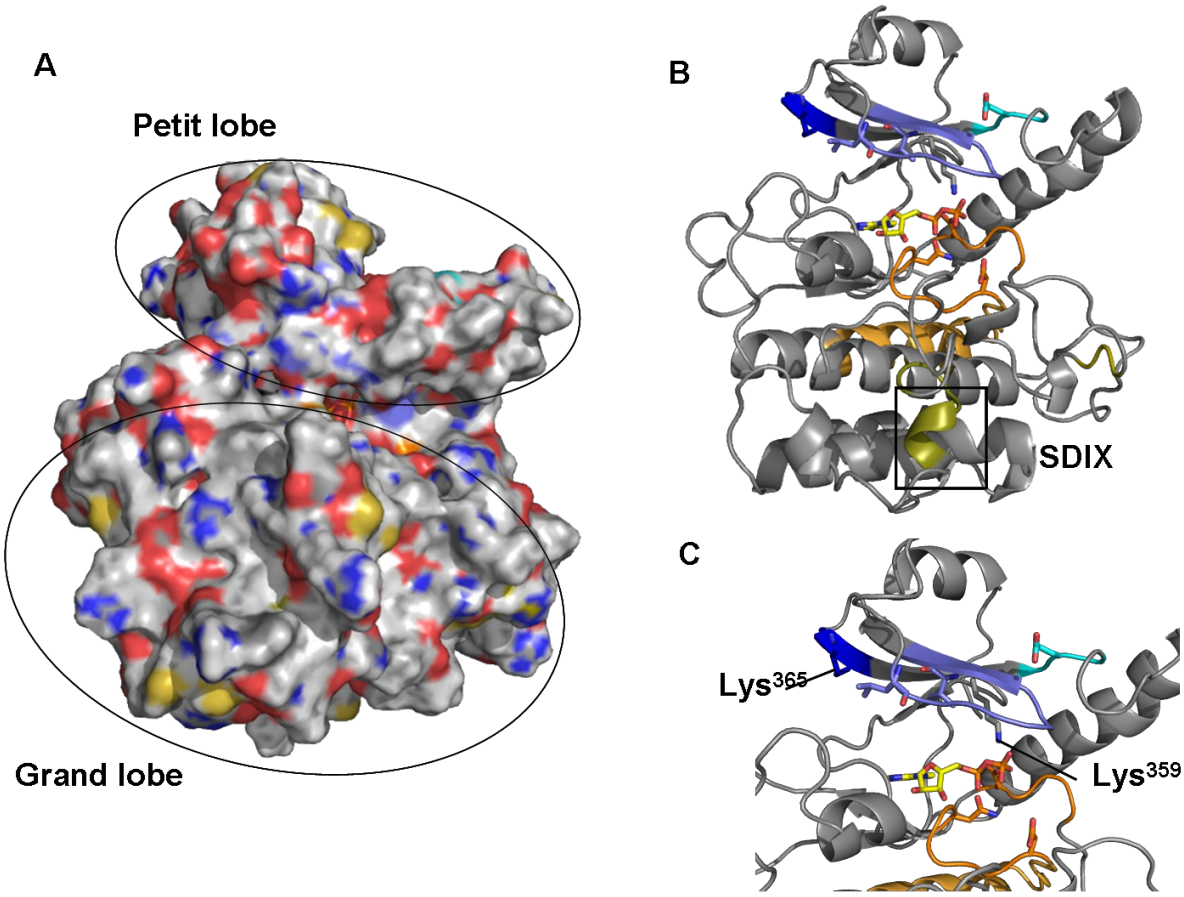



**Figure 72: Analyse du modèle préliminaire de pUL97 délété à partir de la structure PDB 2NRU.** A. Le modèle est représenté en surface. La proximité entre les deux lobes ne permet pas le relargage de l'ATP. B. Le domaine de fixation du substrat (SDIX) est éloigné de la poche de fixation de l'ATP (sticks jaune). C. La lysine catalytique Lys365 est située loin de la queue polyphosphate alors que la lysine non conservée 359 est positionnée favorablement pour la création de contacts polaires avec les phosphates β et γ.

### A.1.c. Substrat naturel et de synthèse à pUL97 et perspectives.

Nous proposons ici plusieurs pistes qu'il pourrait être intéressant de suivre pour l'élucidation des mécanismes de reconnaissance de pUL97 envers ses substrats dès que le modèle de pUL97 sera amélioré.

Il a été montré par Baek (Baek, Krosky et al. 2002) que pUL97 reconnaît spécifiquement les sites de phosphorylation correspondant au motif S/T-X4-R/K. Un peptide de synthèse SP38 (RKESY**S**VYVY**K**V) issu de l'histone H2B a montré une affinité particulièrement intéressante avec pUL97 ce qui peut être le point de départ de la formulation d'un peptidomimétique pouvant inhiber l'activité de pUL97.

Un substrat naturel de pUL97 a par ailleurs déjà été identifié. Il s'agit de la phosphoprotéine pUL44, protéine accessoire de la polymérase virale pUL54 (Krosky, Baek et al. 2003). Il a été montré que cette protéine est principalement phosphorylée au niveau des sérines bien que sa séquence contienne 38 sérines, 30 thréonines et 10 tyrosines. En effet, il a été observé 10 fois plus de phosphorylation sur les sérines par rapport aux thréonines (tyrosines non phosphorylées traduisant le caractère Ser/Thr- et non Tyr-kinase).

Afin de prédire ces sites de phosphorylations, nous avons soumis la séquence de pUL44 (**P16790**) à PROTSCAN (via NPS@ PBIL) pour la recherche du motif S/T-X4-R/K identifié par Baek et al. La Figure 73 montre la séquence de pUL44 colorée au niveau des résultats obtenus par PROTSCAN en représentant en magenta les sérines potentiellement phosphorylées par pUL97, en bleu les lysines et arginines en position i+5 indispensables à cette phosphorylation et en rouge les résidus situés entre les sérines et ces résidus basiques.

On observe donc 5 motifs S-X4-K. En effet, d'après la recherche par PROTSCAN seules des sérines sont propices à la phosphorylation ce qui est en accord avec les résultats de Krosky et al. De plus, le résidu basique en i+5 est systématiquement une lysine. Nous avons continué nos investigations en recherchant si pUL44 était disponible dans la PDB. En effet, sa structure est consignée dans la PDB sous le code 1T6L.

```
mdrktrlsep ptlalrlkpy ktaiqqlrsv iralkenttv tflptpslil qtvrshcvsk
itfnssclyi tdksfqpkti nnstpllgnf mylts skdlt kfyvqdisdl sakismcapd
fnmefssacv hgqdivrese nsavhvdldf gvvadllkwi gphtrvkrnv kkapcptgtv
qilvhagppa ikfiltngse leftannrvs fhgvknmrin vqlknfyqtl lncavtklpc
tlrivtehdt llyvasrngl favenfltee pfqrgdpfdk nyvgnsgksr gggggggsls
slanagglhd dgpgldndlm nepmglgglg gggggggkkh drgggggsgt rkmssggggg
dhdhglsske kyeqhkitsy ltskggsggg gggggggldr nsgnyfndak eesdsedsvt
fefvpntkkq kcg
```

**Figure 73: Recherche des motifs S/T-X4-R/K présents dans la séquence de pUL44.** Les motifs rencontrés sont colorés selon le code: sérine en magenta, résidu basique en i+5 en bleu et les acides aminés intermédiaires en rouge.

En respectant le même code couleur que dans la Figure 73, nous avons reporté les sites potentiels de phosphorylation qui sont bel et bien accessible ce qui appuie nos hypothèses (Figure 74).

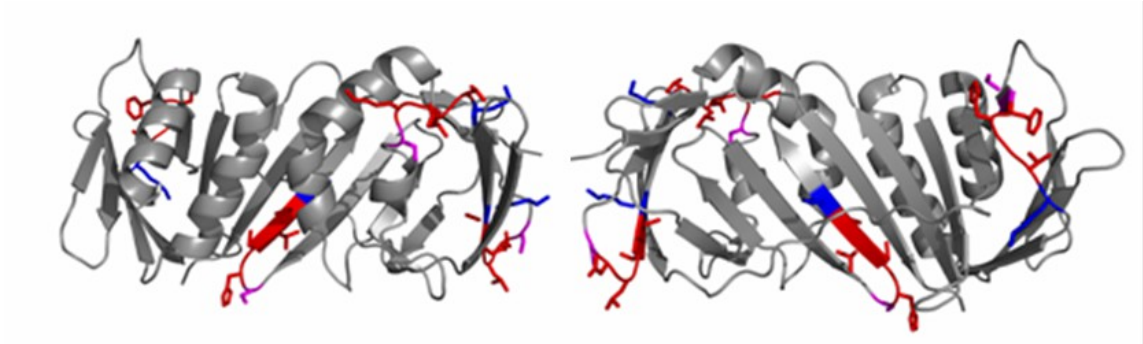


**Figure 74: Cartographie des sites potentiels de phosphorylation de pUL44 (code PDB 1T6L).** Les deux vues sont obtenues par rotation de 180°. Les différents sites identifiés sont tous situés à la surface de la protéine. Le code couleur utilisé correspond à celui de la Figure 73.

A.1.d. Bilan de l'étude de pUL97

En résumé, un modèle de pUL97Δ (590-607), résistant à GCV et donc fonctionnel, a été généré par PHYRE avec un haut indice de confiance si l'on se base sur les « E values » (plus élevé que le modèle basé sur la même structure-support PDB après soumission de la séquence non tronquée (Figures 70 & 71, modèles 2nurD en vert) et faisant apparaître tous les sites importants à son activité phosphotransférase. Cependant, ce modèle reste à optimiser pour faire apparaître une topologie pouvant reconstruire un site actif capable de loger une molécule d'ATP, et repositionner la lysine 365 afin de rendre compte de son rôle dans le transfert du phosphate. Par ailleurs, la poche de liaison au substrat est un autre point à optimiser pour pouvoir étudier *in silico* le mode de reconnaissance des substrats.

Nombres de choses commencent à être connues au sujet de cette protéine mais la reconnaissance d'un motif spécifique S/T-X4-K/R peu conventionnel reste à élucider. Les structures de MBV et GCV sont aisées à construire ou a récupérer dans la PDB (monophosphorylé et non phosphorylé pour GCV, codes PDB 2F3T et 1KI2, respectivement), la structure de pUL44 est connu et un peptide issu de l'hiostone HB2 ayant une grande affinité avec pUL97 a été identifié. Il serait donc intéressant de mener différentes calculs de *docking* pour tenter d'élucider les mécanismes moléculaires impliquant pUL97, ses partenaires naturels ou de synthèses, et ses inhibiteurs afin d'ouvrir d'autres voies pour le traitement du CMVH.

## A.2. Annexe A2: Protéine accessoire pUL27

Nous présentons dans ce chapitre un travail traitant de la protéine accessoire pUL27. Peu d'informations sont actuellement disponibles sur cette protéine. Pourtant, des traitements au Maribavir (MBV) ont mené à l'émergence de souches résistantes au niveau de sa séquence (Komazin, Ptak et al. 2003) suggérant ainsi un rôle de cette protéine dans le mécanisme d'action du MBV. Le MBV inhibe les étapes de sortie du noyau et de la réplication de l'ADN viral en visant en particulier la phosphotransférase pUL97.

La protéine pUL27 est colocalisée avec pUL97 mais ne semble pas indispensable à l'exportation extranucléaire de la capside et, à ce titre, a été définie comme étant la protéine accessoire de la phosphotransférase (Chou, Marousek et al. 2004). Plusieurs mutations et délétions ont été observées chez pUL27 chez des souches recombinantes de laboratoire contenant des séquences de pUL97 défectueuses avec un faible taux de résistance au MBV (Chou, Marousek et al. 2004; Chou 2009). Quelques unes de ces mutations ont été identifiées en association avec des mutations observées sur pUL97 et d'autres conjointement avec des mutants inefficaces de pUL97 suggérant un rôle régulateur de pUL27 vis-à-vis de l'activité de pUL97.

Ici, nous présentons une analyse de la structure primaire de pUL27. Nous avons formulé des hypothèses sur le rôle de certains domaines conservés et le rôle de quelques mutations pouvant rendre compte d'un mécanisme d'action de MBV sur pUL27.

La Figure 75 résume les différents domaines conservés issus de l'alignement de séquences présenté en Annexe B6.

En premier lieu, les homologues de pUL27 ont été identifiés uniquement chez les β-herpesvirus. Ce point permet d'émettre l'hypothèse d'un rôle de cette protéine propre au mode d'infection des bêta-herpesvirus. Ainsi, contrairement aux alignements de séquences générés pour les autres protéines ayant fait l'objet de cette thèse (pUL89 et pUL56), nous avons classé les séquences selon les genres cytomégalovirus, muromégalovirus, roséolovirus et l'homologue issu de l'herpesvirus Tuppaiid 1 constituant un genre à lui seul.

Nous avons identifié quatre domaines conservés communs aux différents genres de bêta herpesvirus (boîtes bleue, verte, noire et rouge, Figure 75), un domaine conservé chez tous les herpesvirus excepté l'homologue de HHV-6A (boîte violette en pointillées, Figure 75) et deux motifs riches en résidus basiques dont le motif est propre à chaque genre (surlignés en noir dans l'Annexe B6). Un domaine situé en N-terminal (Glu17-Glu50) est conservé à travers les

différents homologues cytomégalovirus et contient 16 résidus acides (acide aspartique ou glutamique) conservés. La fonction de ce fragment nous est difficile à prédire.

Aucune prédiction de structure tertiaire n'a pu être accomplie par PHYRE, faute d'homologue structural.

Cependant, bien que l'étude s'est basée uniquement sur cet alignement de séquence, plusieurs signatures permettent d'émettre des hypothèses sur une multimérisation potentielle de pUL27 et la présence de divers sites de phosphorylation probables par pUL97 de la forme S/T-$X_4$-K/R pUL97. Les résultats présentés ont fait objet d'une publication (Hantz, Couvreux et al. 2009).

### A.2.a. Multimère de pUL27 ?

Au regard de la conservation d'un grand nombre d'amino acides hydrophobes et de cystéines nous pouvons spéculer sur la nature multimérique de pUL27.

Le domaine conservé I ($Val^{163}$-$Thr^{222}$) contient quatre résidus parfaitement conservés ($Tyr^{188}$, $Arg^{207}$, $Leu^{209}$ et $Leu^{216}$) et plusieurs résidus hydrophobes dont deux aliphatiques ($Val^{163}$ et $Leu^{193}$) et deux aromatiques ($Tyr^{168}$ et $Tyr^{205}$) hautement conservés.

Le domaine conservé II ($Leu^{228}$-$Asp^{285}$) contient lui aussi des résidus aliphatiques ($Leu^{228}$ et $Ile^{236}$) et aromatiques ($Phe^{265}$ et $Tyr^{277}$) hautement conservés. En plus de ces résidus hydrophobes, deux acides aminés parfaitement conservés sont contenus dans ce domaine. Ces deux résidus sont des cystéines ($Cys^{281}$ et $Cys^{283}$) qui sont, de part leur capacité à former des liaisons covalentes, capables de jouer un rôle important dans le repliement de pUL27. Compte tenu de la proximité de ces deux cystéines dans un fragment prédit en hélice empêchant la formation d'un pont disulfure intramoléculaire, il semblerait que la liaison intermoléculaire soit plus probable.

Le domaine III ($Leu^{322}$-$Pro^{444}$) est prédit comme étant constitué de deux hélices alpha dont une longue de 55 acides aminés. Sept résidus appartenant à la première hélice sont parfaitement conservés ($Leu^{335}$, $Pro^{336}$, $His^{340}$, $Leu^{341}$, $Asp^{356}$, $Leu^{357}$ et $Gly^{331}$). Nous avons reconstruit grâce à Pymol ce fragment de 55 résidus prédit en hélice (Figure 76). On remarque que la face hydrophobe de cette hélice amphipathique est très conservée et contient les résidus parfaitement conservés $Asn^{393}$ et $Glu^{411}$. La taille de cette hélice et la conservation de sa face hydrophobe constituent des indices supplémentaires d'une multimérisation de pUL27.

Le domaine IV ($Phe^{471}$-$Arg^{509}$) contient les résidus parfaitement conservés $Phe^{471}$, $Glu^{498}$ et $His^{503}$ ainsi qu'un grand nombre de résidus hydrophobes.

La majorité des résidus conservés parmi les homologues de pUL27 sont aliphatiques ou à caractère hydrophobe. Le fait que ces résidus soient dépourvus de groupements fonctionnels justifie leur rôle plutôt dans le repliement des protéines ou aux interfaces dans des complexes permanents que dans des fonctions catalytiques. De plus, la présence de plusieurs cystéines parfaitement conservées est en accord avec la possibilité que pUL27 multimérise en particulier *via* les domaines III et IV. Cependant, il est pour le moment impossible de prédire l'importance de cette multimérisation dans la fonction de pUL27, à partir de cette analyse préliminaire.

### A.2.b. Sites de phosphorylation

Les études menées sur l'activité phosphotransférase de pUL97 ont permis l'identification d'un motif-substrat particulier: S/T-X4-R/K. La colocalisation et l'interdépendance des mutations au Maribavir observées sur pUL27 et pUL97 suggèrent une interaction probable entre ces deux protéines virales. Ici, nous avons découvert au moins trois motifs potentiels de phosphorylation de pUL27 par pUL97.

Le premier est contenu dans le domaine I et correspond à $^{171}$SRAAKR$^{176}$. Ce motif est prédit en hélice et est conservé à travers les homologues de cytomégalovirus. Le deuxième motif est $^{420}$SGDVQR$^{425}$ contenu dans le domaine III. Ce fragment est commun à pUL27 et son homologue de PoHV_4 uniquement. Enfin, un troisième motif est rencontré dans le domaine C-terminal: $^{555}$SNHANR$^{560}$. Comme le premier motif potentiel de phosphorylation celui-ci est conservé à travers les différents homologues de cytomégalovirus.

Les présences de ces trois motifs dont deux conservés à travers les homologues cytomégalovirus sont en faveur d'une interaction enzyme-substrat avec pUL97 pouvant permettre la maturation de pUL27 éventuellement pour l'exportation nucléaire de la capside.

### A.2.c. Mutations au Maribavir

Seize mutations ont été observées à des positions disséminées dans la séquence de pUL27 qui constituent des sites de résistances au Maribavir.

La majorité de ces résistances sont localisées dans les domaines conservés I (L193F et la délétion 218), II (R233S et A269T) et III (L335P, V353E, W362R, W362stop, A406V, C415stop et L426F). W153R est trouvée dans la région partiellement conservée. Deux autres enfin appartiennent à des régions non conservées: E22stop en N-terminal et la délétion 301-311 dans la région centrale de la séquence supposée flexible dûe à la présence d'un grand nombre de glycines et de sérines.

Il est intéressant de remarquer qu'un nombre important de mutations est rencontré aux niveaux de positions conservées chez les différents homologues ($Leu^{193}$, $Ala^{269}$, $Val^{353}$ et $Leu^{426}$) ou pour le moins conservées pour les cytomégalovirus ($Trp^{153}$, $Arg^{233}$, $Trp^{362}$, $Ala^{406}$ et $Cys^{415}$). De plus, la leucine 335 parfaitement conservée à travers tous les homologues de pUL27 est elle aussi sujette à une mutation.

Contrairement à pUL97 où un mécanisme d'action du MVB a pu être proposé intuitivement compte tenu de la localisation de mutation au niveau du site de fixation de l'ATP, il est moins évident de comprendre celui mis en jeu pour pUL27. Du fait que deux motifs potentiels de phosphorylation sont proches de plusieurs points de mutations, peut être ces mutations permettent-elles de contourner l'inhibition de l'activité phosphotransférase par MBV en modifiant quelques amino acides rendant possible la phosphorylation de pUL27 par pUL97.

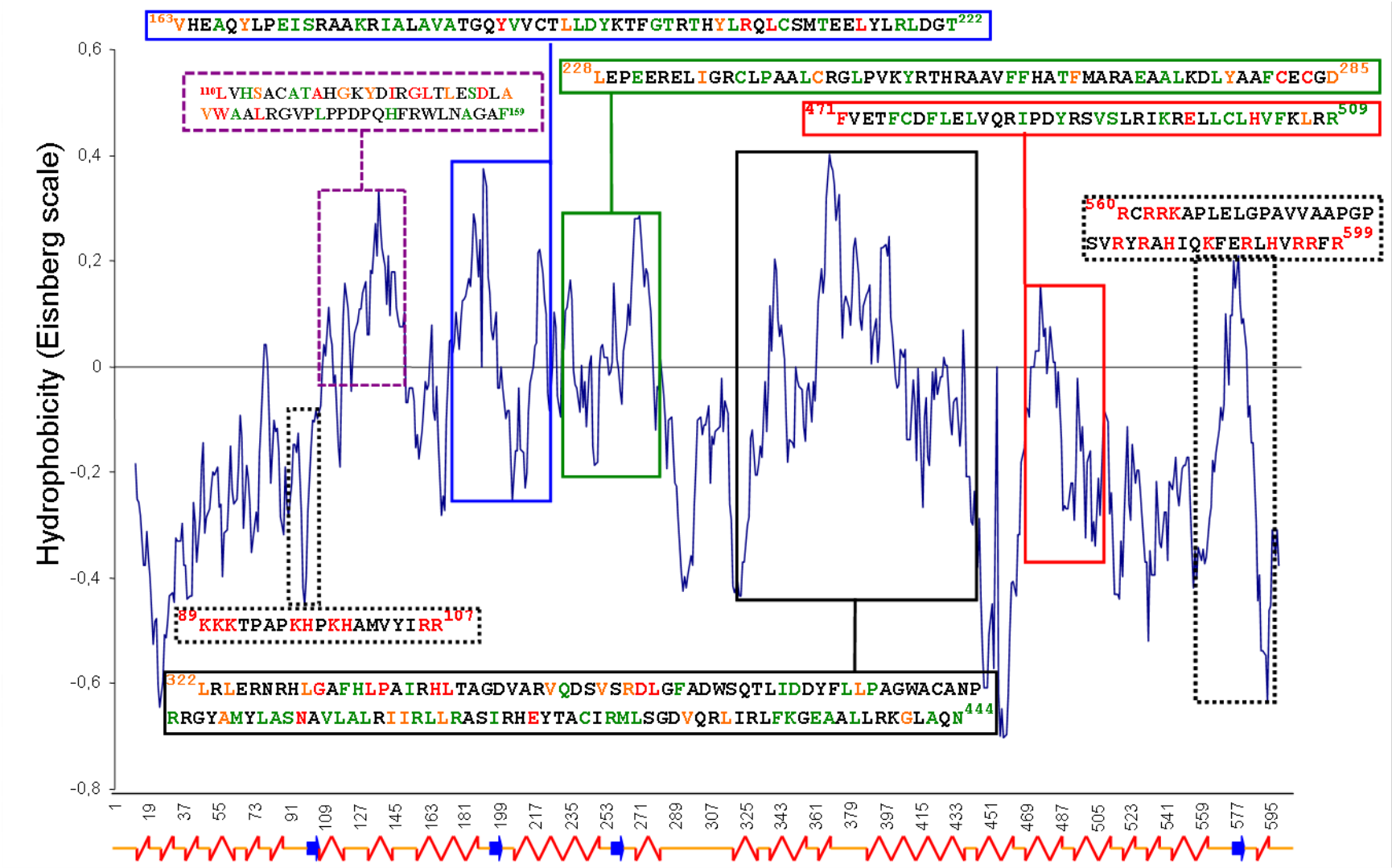

163VHEAQYLPEISRAAKRIALAVATGQYVVCTLLDYKTFGTRTHYLRQLCSMTEELYLRLDGT222
110LVHSACATAHGKYDIRGLTLESDLA
VWAALRGVPLPPDPQHFRWLNAGAF159
228LEPEERELIGRCLPAALCRGLPVKYRTHRAAVFFHATFMARAEAALKDLYAAFCECGD285
471FVETFCDFLELVQRIPDYRSVSLRIKRELLCLHVFKLRR509
560RCRRKAPLELGPAVVAAPGP
SVRYRAHIQKFERLHVRRFR599
89KKKTPAPKHPKHAMVYIRR107
322LRLERNRHLGAFHLPAIRHLTAGDVARVQDSVSRDLGFADWSQTLIDDYFLLPAGWACANP
RRGYAMYLASNAVLALRIIRLLRASIRHEYTACIRMLSGDVQRLIRLFKGEAALLRKGLAQN444
Hydrophobicity (Eisnberg scale)
0,6
0,4
0,2
0
-0,2
-0,4
-0,6
-0,8
1
19
37
55
73
91
109
127
145
163
181
199
217
235
253
271
289
307
325
343
361
379
397
415
433
451
469
487
505
523
541
559
577
595

**Figure 75: Domaines conservés de pUL27.** Les régions conservées issues de l'alignement de séquences présenté en Annexe B6 sont précisées suivant le même code de conservation (rouge, orange et vert pour parfaitement, hautement et moyennement conservé, respectivement). La prédiction de structures secondaires est schématisée en dessous (zig zag et flèches pour les hélices et brins bêta, respectivement).

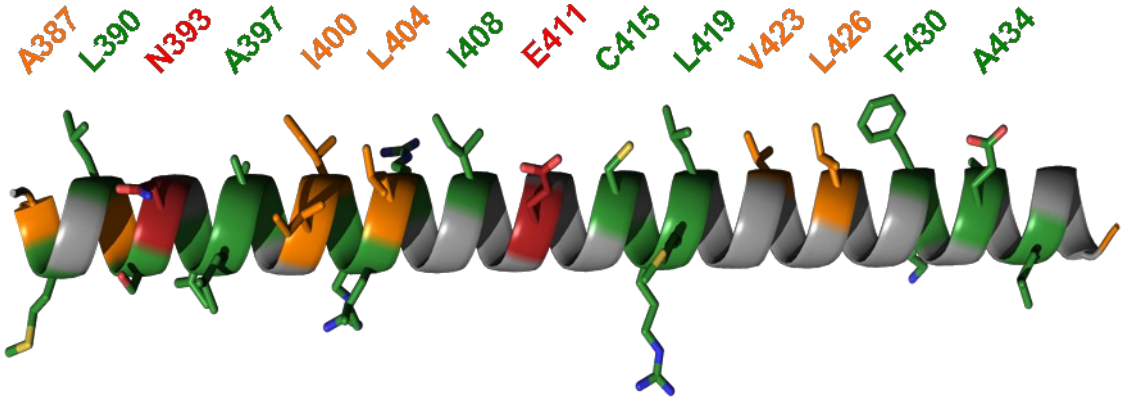



**Figure 76: Hélice 387-439 du domaine conservé III de pUL27 construite dans Pymol.** Les résidus sont colorés en fonction de leur taux de conservation en accord avec l'alignement de séquences de l'Annexe B6. La face hydrophobe est très conservée et contient les résidus parfaitement conservés $Asn^{393}$ et $Glu^{411}$.

### A.2.d. Bilan de l'étude de pUL27

La recherche de domaines fonctionnels à l'intérieur de la séquence de pUL27 a été initiée par alignement de séquences de celle-ci et de ses homologues herpesviraux. Comme Komazin et al (Komazin, Ptak et al. 2003), nous n'avons trouvé d'homologues de cette protéine que chez les bêta herpesvirus. Cette protéine se décompose en quatre domaines conservés notés I à IV) à travers tous ses homologues et deux deux fragments situés aux extrémités de sa séquence susceptibles de contenir des signaux de localisation nucléaire, car riches en résidus basiques.

La conservation des résidus à l'intérieur des domaines I à IV montre plusieurs signes nous faisant penser que cette protéine doit exister sous forme multimérique. Cette hypothèse est soutenue par la présence d'un grand nombre d'acides aminés hautement à parfaitement conservés constituant des domaines hydrophobes. De plus, deux cystéines parfaitement conservées appartenant au domaine IV distantes de deux résidus confortent cette hypothèse.

Par ailleurs, trois sites de phosphorylation potentiels, correspondant au motif caractéristique S/T-$X_4$-K/R reconnu par la phosphotransférase pUL97, ont été identifiés. Ceci est un premier élément compatible avec une interaction de type enzyme-substrat entre pUL27 et pUL97. Ainsi, les mutations croisées observées entre ces deux protéines soumises aux Maribavir peuvent trouver ici une explication. En mutant, pUL27 peut s'adapter aux formes résistantes de pUL97 et ainsi accomplir sa fonction biologique.

## B. *ANNEXE B: Alignements De Séquences*

### B.1. ANNEXE B1: Alignement de séquence de pUL89 et 17 de ses homologues herpesviraux

```
                                 1        10        20        30        40        50        60        70
                                 |         |         |         |         |         |         |         |
HSV-1        ----MFGQQLASDVQQYLERLEKQRQL VGADEASAGLTMGG-----DALRVPFLDFATATPKRHQTVVPGVGTL DCCEHSPLFSAVARRLLFNSLVPA
HSV-2        ----MFGQQLASDVQQYLERLEKQRQQ VGVDEASAGLTLGG-----DALRVPFLDFATATPKRHQTVVPGVGTL DCCEHSPLFSAVARRLLFNSLVPA
CeHV-2       ----MFGRQLADDVRLFFEGLEKQRQQ MGAADAPAGLP-AG-----DALRVPFLDFATATPRRHQTVVPGVGTL DCCEHAPLFAAVARRLLFNSLVPP
CeHV-1       ----MFGRQLADDVRLFFEGLEKQRQQ MGASDAPAGLA-AG-----DALRVPFLDFATAAPRRHQTVVPGVGTL DCCEHAPLFSAVARRLLFNSLVPP
EHV-1        ----MFGRVLGRETVQYFEALRREVQA RGAKNRAAEAQNGG----EDDAKTAFLNFAIPTPQRHQTVVPGVGTL DCCETAQIFASVARRLLFRSLSKW
BHV-5        ----MFGGALGEQSARYFQRLLRERRE AAGRGARAEGGGSGGARGEDDARAPFLDFAVAAPKRHQTVVPGVGTL GYCELAPLFAATASRLLLTSLARA
HHV-3        MSLIMFGRTLGEESVRYFERLKRRRDE FGTLESPTPCSTRQGSLG-NATQIPFLNFAIDVTRRHQAVIPGIGTL NCCEYIPLFSATARRAMFGAFLSS
MDV-2        ----MFGGLLGEETKRHFERLMKTKND LGASHRNERSIRDG-----DMVDAPFLNFAIPVPRRHQTVMPAIGIL NCCDSLGIYSAITTRMLYSSIACS
ALPHA        ----MFG—Lτσδδσ--ρφδ-L-πδ-δ-π-τσ----------------δ-φδφσFLσFAφσφσδRHQφVφPτaGφLHσδCν---αρττφσδRφφφ-τφ---
AD169        ----------------------MLRGDSAAKIQERYAELQKRKSHPTSCISTAFTNVATLCRKRYQMMHPELGLA SCNEAFLPLMAFCGRH -------
CCMV         ----------------------MLHGDSAAKIQERYEELQKRKNHPTSCISTAFTNVAALCRKRYQMMHPELGLA SCNEAFLPLMAFCGRY -------
RhCMV        ----------------------MLRGESASRIQERYQELLKRKNHPSSCISTSFTNVAALCRKRYQMMHPELGLA CCNEAFLPLTTFCGRY -------
TuHV-1       ----------------------MLSHETVHALRARYDELKRHREARPNVISTEFPDAATLCKKRFQTLSPELGLA ACNSVCLPLVRFC AY -------
HHV6-B       ----------------------MLRTCDITHIKNNYEAIIWKGERNCSTISTKYPNSAIFYKKRFIMLTPELGFA SYNQQVKPLYTFCEKQ -------
HHV-7        ----------------------MLRSCDIDAIQKAYQSIIWKHEQDVK-ISSTFPNSAIFCQKRFIILTPELGFT AYCRHVKPLYLFCDRQ -------
BÊTA         ----------------------MLδ-δσα--αδ—Y---α-δπ-δ---δ-ISσ-ρφσσAφφφδKRρ-φφδPELGφφHτδσδ--φPLφ-FC--δR-------
PL_HV3       ----------------------MLYAKILP LQKAYKDLKKRDSEKISTWENNPLVFSS-CKKAERWPHPFLGVLSGINQYSANLECYCRNFNP----YI
PL_HV1       ----------------------MLYAKMLP LHQAYKDLQIQDTEQIHKWENNPLVFCT-SKKAERWPHPFLGVLSGVNQYSSNLESYCRLFNP----YI
HHV-8        ----------------------MLLSRHRE LAANLQETAKDAGER---WELSAPTFTRHCPKTARMAHPFIGVVHRINSYSSVLETYCTRHHP----AT
HHV-4        ----------------------MLYASQRG LTENLRNALQQDSTTQGCLGAETPSIMYTGAKSDRWAHPLVGTIHASNLYCPMLRAYCRHYGP PVFVA
GAMMA        ----------------------MLφσδ---πL--σφδδφ--δσσδδ---φ--δφασφ---τφKσ-RφαHPφαGφαδ--N-Yτσ-LδσYCδ-ρ-P----φφ
ET           ----------------------xx-------------------------φ---φφσx---xxδx--φφxPxαGφφδ--σ-----φx-φσ--x--------
PSI-PRED     ----------------------cchhhhhhhhhhhhhhhhcccccccccccccccccccccchhhhcccchhhhhhhhhhhhhhhhhhhhhhhc-------
```

```
                          80        90                100       110         120       130       140       150
                           |         |                 |         |           |         |         |         |
HSV-1        QLKGRDFGGD----H--TAKLEFLAPELVRAVARLRFKECAPADVVPQRNAYYSVLNTFQALHRSEAFRQLVHFVRDFAQLLKTSFRASSLTE-TTGPPK
HSV-2        QLRGRDFGGD----H--TAKLEFLAPELVRAVARLRFRECAPEDAVPQRNAYYSVLNTFQALHRSEAFRQLVHFVRDFAQLLKTSFRASSLAE-TTGPPK
CeHV-2       QLRGRDFGGD----R--ATKLEFLAPELVRAVSRLRFAACAPGDVVPQRNAYYSVLNTFRALHRSEAFRQLVHFVRDFARLLRTSFRASSLSEEDAGPLK
CeHV-1       QLRGRDFGGD----R--ATKLEFLAPELVRAVARLRFAECAPGDVVPQRNAYYSVLNTFRALHRSEAFQQLVHFVRDFARLLRTSFRASSLAE-DAGPAK
EHV-1        QSGEARERLD----P--ASVEAYVDPKVRQALKTISFVEYSDDEARSCRNAYYSIMNTFDALRSSDAFHQVASFVARFSRLVDTSFNG-ADLDGDGQQAS
BHV-5        EAG-----LN----T--GTGEAHVSQELAGALGALRFAAHPPAEAAAHRNAYHSVMEALESMRASGAFAQVAAFVARFSRLVGTSFSHLGGDDGADPPRA
HHV-3        TGYNCTPNVVLKPWR--YSVNANVSPELKKAVSSVQFYEYSPEEAAPHRNAYSGVMNTFRAFSLSDSFCQLSTFTQRFSYLVETSFESIE----ECGSHG
MDV-2        EFDELRRDSVPRCYPRITNAQAFLSPMMMRVANSIIFQEYDEMECAAHRNAYYSTMNSFISMRTSDAFKQLTVFISRFSKLLIASFRDVNKL--DDHTVK
ALPHA        δ--------σ--------σ----ασ--φ--αα--α-F--δσ--νσ-σδRNAYδτφφδσφ-τφδ-S-τF-Qασ-Fφ-δFτδLα-σSFδ--------σ----
AD169        ------DYNSPEESQRELLFHERLKSALDK-------LTFRPCSEEQRASYQK--LDALTELYRDPQFQQINNFMTDFKKWLDGGFSTAVEGDAKA--IR
CCMV         ------DYHSPEE-KRELVFHERLKSALDE-------LAFRPCPEEQRASYQK--LDALTELYRDPQFVQINSFMTDFKKWLDGGFATA-EGDAKP--IR
RhCMV        ------DYHS-EEDRRTLLFHENLKTALDN-------LNFLPCPEDQRQSYQK--LDALTELYRDPQFVQINNFMTDFKRWLDGGFATM-EGDAKP--IR
TuHV-1       ------DYNQPLTRHATLNLSAAVRAGLDE-------ITFRPCSEEQRSQSHK--IAALTGLYKDPDFLQINSFLTDFRRWIDGDFDPSGGGDKSTTHIY
HHV6-B       ------HLKN----RKPLTILPSLTRKLQE-------MKFLPASDKSFESQYTEFLESFKILYREPLFLQIDGFIKDFRKWIKGEFNDFGDTRK----IQ
HHV-7        ------HVKS----K--IAICDPLNCALSK-------LKFTAIIEKNTEVQYQKHLELQTSFYRNPMFLQIEKFIQDFQRWICGDFENTNKKER----IK
BÊTA         ------δφδδ----δ--α-φ---αδ--Lδδ-------φ-F-αφ-νδδ---δδδ--α---δ-φYRδP-F-QIδ-FφδDFδπWαδG-F-σ----δ-----Iδ
PL_HV3       PTDFTSDVGTPRYMTTQ---RQMLIDGLAH--SIKSFFREDVPVEHDAKVEFQSALTTYQCSLKCPKFEELRNFIINLSSFLNGAYVSKTS--------S
PL_HV1       PFDVKCAMGTPRCISAQ---RQMLIDGFSD--SIRALFGTKIATDREVKVEFESALSTYTCSTQCRKFEELKNFIINLSSFLNGTYVSKSN--------S
HHV-8        PTSANPDVGTPRPSEDNVPAKPRLLESLSTYLQMRCVREDAHVSTADQLVEYQAARKTHDSLHACSVYRELQAFLVNLSSFLNGCYVPGVH--------W
HHV-4        SDESLPMFGASPALHTPVQVQMCLLPELRD--TLQRLLPPPNLEDSEALTEFKTSVSSARAILEDPNFLEMREFVTSLASFLSGQYKHKPA--------R
GAMMA        σ-δ--σ-φGσσ---δσ----δ--Lα--φ-δ--δφδ-φ-----α-δ-ν-φφEρδστ-δσφδτ-φ-δ--ρ-Eφδ-FαφδLσSFLσG-Yφ----------δ
ET           ---------x----------αx--φ----------x-x---xxδ---------x-δ--ρ-δφx-Fφ-δφxδφα-σ-ρ----------
PSI-PRED     ------cccccccccccccccchhhhhhhh--------hhccccchhhhhhhhh--hhhhhhhhhchhhhhhhhhhhhhhhhcccccccccccccc—ee
```

```
                                  160       170       180       190       200       210       220       230
                                   |         |         |         |         |         |         |         |
HSV-1        KRAKVDVATHGRTYGTLELFQKMILMHATYFLAAVLLGDHAEQVNTFLRLVFEIPLFSDAAVRHFRQRATVFLVPRRHGKTWFLVPLIALSLASFRGIKI
HSV-2        KRAKVDVATHGQTYGTLELFQKMILMHATYFLAAVLLGDHAEQVNTFLRLVFEIPLFSDTAVRHFRQRATVFLVPRRHGKTWFLVPLIALSLASFRGIKI
CeHV-2       KRAKVDVATHGRSHGTLEPFQKMILMHATYFLAAVLLGDHAEQVSTFLRLVFEIPLFSDAAVRHFRQRATVFLVPRRHGKTWFIVPLIALSLASFRGIKI
CeHV-1       KRAKVDVATHGRSHGTLEPFQKMILMHATYFLAAVLLGDHAEQVSTFLRLVFEIPLFSDAAVRHFRQRATVFLVPRRHGKTWFIVPLIALSLASFRGIKI
EHV-1        KRARVDVPTYGKQRGTLELFQKMILMHATYFIAAVILGDHADRIGAFLKMVFNTPEFSDATIRHFRQRATVFLVPRRHGKTWFLVPLIALALATFKGIKI
BHV-5        KRARVEPPS-DQTRGALELFQKMILMHATYFVAATLLRDHAERIGAFLRVAFNTPAFSDAAVAHFRQRATVFLVPRRHGKTWFLVPLIALALATFKGIKI
HHV-3        KRAKVDVPIYGRYKGTLELFQKMILMHTTHFISSVLLGDHADRVDCFLRTVFNTPSVSDSVLEHFKQKSTVFLVPRRHGKTWFLVPLIALVMATFRGIKV
MDV-2        KRARIDAPSYDKLHGTLELFQKMIFDACHLFCNFCFTWRSRRASERLLRVAFDTPHFSDIVTRHFRQRATVFLVPRRHGKTWFLVPLIALAMSSFEGIRI
ALPHA        KRAπαναα--σδ-δGσLEαFQKMIφ-φφφφFφσ-φφφ-δδ-δ----φLRφαFδφP-φSD-φφ-HFπQπσTVFLVPRRHGKTWFαVPLIALνφσδFδGIπα
AD169        ----------------LEPFQKNLLIHVIFFIAVTKIPVLANRVLQYLIHAFQIDFLSQTSIDIFKQKATVFLVPRRHGKTWFIIPIISFLLKHMIGISI
CCMV         ----------------LEPFQKNLLIHVIFFIAVTKIPVLANRVLQYLIHAFQIDFLSQTSIEIFKQKATVFLVPRRHGKTWFIIPIISFLLKHMIGISI
RhCMV        ----------------LEPFQKNLLIHVIFFIAVTKIPVLANRVLQYLVHAFEIDFLSQTSVEVFKQKATVFLVPRRHGKTWFIIPIICFLLKHMIGISI
TuHV-1       ----------------LEPFQKNLLMHVLVFIAVTKLPVLANRVAEYLLYMFDIDFLSAASLDLFKQKATVFLVPRRHGKTWFTVPIICFLLKNVQGISI
HHV6-B       ----------------LEPFQKNILIHVIFFIAVTKLPALANRVINYLTHVFDIEFVNESTLNTLKQKTNVFLVPRRHGKTWFIVPIISFLLKNIEGISI
HHV-7        ----------------LEPFQKSILIHIIFFISVTKLPTLANHVLDYLKYKFDIEFINESSVNILKQKASVFLVPRRHGKTWFMIPVICFLLKNLEGISI
BÊTA         ----------------LEPFQKδαLφHααφFIσVTKαPσLANπVαδYLφρφFδIνFασ-σσαδφφKQKσσVFLVPRRHGKTWFφαPαIσFLLKδφ-GISI
PL_HV3       ----------------IEPFQKQLILHTFYFLISIKAPASCVQLFNIFKHYFDMSDMITDTMEIFKQKASVFLIPRRHGKTWIVVAITSILLTSIEDLHI
PL_HV1       ----------------VEPFQKQLILHTFYFLISIKAPESCMKLFDIFKVYFDLSDMIPERMEVFKQKASVFLIPRRHGKTWIVVAIISILLTSIEDLHI
HHV-8        ----------------LEPFQQQLVMHTFFFLVSIKAPQKTHQLFGLFKQYFGLFETPNSVLQTFKQKASVFLIPRRHGKTWIVVAIISMLLASVENINI
HHV-4        ----------------LEAFQKQVVLHSFYFLISIKSLEITDTMFDIFQSAFGLEEMTLEKLHIFKQKASVFLIPRRHGKTWIVVAIISLILSNLSNVQI
GAMMA        ----------------αEαFQδQααφHσFρFLαSIKτα--σ-δφFσαFδ-φFσφ-νφφ-δ-φδφFKQKASVFLIPRRHGKTWIVVAIφSφαLσδαδσαδI
ET           ----------------αEαFQδxαφ-φxφφFφx-φφx---------φφx-φFxφ--φx---φ-φφπQπσσVFLαPRRHGKTWφφαααφσφαφxδφ-σαδα
PSI-PRED     ----------------echhhhhhhhhhhhhheeeeechhhhhhhhhhhhhhhcccccchhhhhhhhhhceeeeeccccccceehhhhhhhhhhhhcccccee
```

```
                240       250       260       270         280       290       300       310       320       330
                 |         |         |         |           |         |         |         |         |         |
HSV-1           GYTAHIRKATEPVFEEIDACLRGWFGSARVDHVKGE-TISFSFPDGSRSTIVFASSHNTNGIRGQDFNLLFVDEANFIRPDAVQTIMGFLNQANCKIIFV
HSV-2           GYTAHIRKATEPVFDEIDACLRGWFGSSRVDHVKGE-TISFSFPDGSRSTIVFASSHNTNGIRGQDFNLLFVDEANFIRPDAVQTIMGFLNQANCKIIFV
CeHV-2          GYTAHIRKATEPVFDEIDACLRGWFGAARVDHVKGE-TISFSFPDGSRSTIVFASSHNTNGIRGQDFNLLFVDEANFIRPDAVQTIMGFLNQANCKIIFV
CeHV-1          GYTAHIRKATEPVFDEIDACLQGWFGSSRVDHVKGE-TISFSFPDGSRSTIVFASSHNTNGIRGQDFNLLFVDEANFIRPDAVQTIMGFLNQANCKIIFV
EHV-1           GYTAHIRKATEPVFDEIGARLRQWFGNSPVDHVKGE-NISFSFPDGSKSTIVFASSHNTNGIRGQDFNLLFVDEANFIRPEAVQTIIGFLNQTNCKIIFV
BHV-5           GYTAHIRKATEPVFEEIVARLRQWFGSERVDHVKGE-VISFSFPDGARSTIVFASSHNTNGIRGQDFNLLFVDEANFIRPEAVQTIVGFLNQASCKIIFV
HHV-3           GYTAHIRKATEPVFEGIKSRLEQWFGANYVDHVKGE-SITFSFTDGSYSTAVFASSHNTNGIRGQDFNLLFVDEANFIRPDAVQTIVGFLNQTNCKIIFV
MDV-2           GYISHIRKAIEPVFEDIGDRLRRWFGAHRVDHVKGE-TITFSFPSGLKSTVTFASSHNTNSIRGQDFNLLFVDEANFIRPDAVQTIIGFLNQATCKIIFV
ALPHA           GYφσHIRKATEPVFE-I-σδLδ-WFGσ--VDHVKGE-σIδFSFφδG-δSTαφFASSHNTNσIRGQDFNLLFVDEANFIRPνAVQTIφGFLNQσδCKIIFV
AD169           GYVAHQKHVSQFVLKEVEFRCRHTFARDYVVENKDN--VISIDHRGAKSTALFASCYNTNSIRGQNFHLLLVDEAHFIKKEAFNTILGFLAQNTTKIIFI
CCMV            GYVAHQKHVSQFVLKEVEFRCRYTFARDYLVENKDN--VISIDHRGAKSTALFASCYNTNSIRGQNFHLLLVDEAHFIKKEAFNTILGFLAQNTTKIIFI
RhCMV           GYVAHQKHVSQFVLKEVEFRCRHTFAKDYVVENKDF--VISIDHKGAKSTALFASCYNTNSIRGQNFHLLLVDEAHFIKEKAFNTILGFLAQNTTKIIFI
TuHV-1          GYVAHQKHVSQFVLKEVEFRCRRLFGRTHTVENKDN--VISIDHPIAKSTALFASCYNTNSIRGQNFHLLLVDEAHFIKKDAFNTILGFLAQNTTKIIFI
HHV6-B          GYVAHQKHVSHFVMKEVEFKCRRMFPEK-TITCLDN--VITIDHQNIKSTALFASCYNTHSIRGQSFNLLIVDESHFIKKDAFSTILGFLPQASTKILFI
HHV-7           GYVAHQKHVSHFVMKDVEFKCRRFFPQK-NITCQDN--VITIEHETIKSTALFASCYNTHSIRGQSFNLLIVDESHFIKKDAFSTILGFLPQSSTKIIFI
BÊTA            GYVAHQKHVSδFVφKνVEFπCRδφFσδδ--αδσ-D---VIδIνH--αKSTALFASCYNTδSIRGQσFδLLαVDEτHFIKδδAFσTILGFLαQσδTKIαFI
PL_HV3          GYVAHQKHVSNAVFTEIMNTMYRFFPSKDIEVKRENGTILFKREGKKPSTLMCATCFNKNSIRGQTFNLLYIDEANFIKKDALPTILGFMLQKDAKLIFI
PL_HV1          GYVAHQKHVSNTVFTEIINTLNRFFPSKYIEIKKDNGTILFKADGKKPSTLMCATCFNKNSIRGQTFNLLYIDEANFIKKDALPTILGFMLQKDAKLIFI
HHV-8           GYVAHQKHVANSVFAEIIKTLCRWFPPKNLNIKKENGTIIYTRPGGRSSSLMCATCFNKNSIRGQTFNLLYVDEANFIKKDALPAILGFMLQKDAKLIFI
HHV-4           GYVAHQKHVASAVFTEIIDTLTKSFDSKRVEVNKETSTITFRHSGKISSTVMCATCFNKNSIRGQTFHLLFVDEANFIKKEALPAILGFMLQKDAKIIFI
GAMMA           GYVAHQKHVτδσVFσEIφδTφδπ-FσσKδαδαδπνστTIφρδ--G--σSσαMCATCFNKNSIRGQTFδLLραDEANFIKKνALPσILGFMLQKDAKαIFI
ET              GYφσHXππασδxVφx-α-xδφδ--Fσx---xxx-x---αxxx-----xSσαφφAσδρNδδσIRGQσFδLLφαDEτδFIπxδAφxσIφGFφxQxδσKααFα
PSI-PRED        eeecchhhhhhhhhhhhhhhhhhhcccccee eccccc--eeeeeccccee eeeeeeeeccccccccccccee eeeehhcccchhhhhhhhhhhcccceeeeee
```

```
                 340       350       360       370       380       390       400       410       420           430
                  |         |         |         |         |         |         |         |         |             |
HSV-1          SSTNTGKASTSFLYNLRGAADELLNVVTYICDDHMPRVVTHTNATACSCYILNKPVFITMDGAVRRTADLFLADSFMQEIIGG---QARETGDDRPVLTK
HSV-2          SSTNTGKASTSFLYNLRGAADELLNVVTYICDDHMPRVVTHTNATACSCYILNKPVFITMDGAVRRTADLFLPDSFMQEIIGG---QARETGDDRPVLTK
CeHV-2         SSTNTGKASTSFLYNLRGAADELLNVVTYICDDHMPRVVTHTNATACSCYILNKPVFITMDGALRRTADLFLADSFMQEIIGG---QSREAADDRPVLTK
CeHV-1         SSTNTGKASTSFLYNLRGAADELLNVVTYICDDHMPRVVTHTNATACSCYILNKPVFITMDGALRRTADLFLADSFMQEIIGG---QSREADDDRPVLTK
EHV-1          SSTNTGKASTSFLYNLKGAADDLLNVVTYICDEHMERVKAHTNATACSCYILNKPVFITMDGAMRNTAELFLPDSFMQEIIGGG--NVSGAHRDEPVFTK
BHV-5          SSTNTGKASTSFLYNLKGASDGLLNVVTYICNEHTPRVAAHGGATACSCYVLNKPVFITMDAAARSTAETFLPNSFMQEIIGGG--EVARRAEPAAVFTR
HHV-3          SSTNTGKASTSFLYNLRGSSDQLLNVVTYVCDDHMPRVLAHSDVTACSCYVLNKPVFITMDGAMRRTADLFMADSFVQEIVGGRKQNSGGVGFDRPLFTK
MDV-2          SSTNSGKASTSFLYGLKGSADDLLNVVTYICDEHMKHVTDYTNATSCSCYVLNKPVFITMDGAMRRTAEMFLPDSFMQEIIGGGVVDRTICQGDRSIFTA
ALPHA          SSTNσGKASTSFLYσLπGττD-LLNVVTYαCσνHφ-πVφσρσσαTτCSCYαLNKPVFITMDτAφRδTAνφFφσσSFφQEIαGG---δ------σ-σαφT-
AD169          SSTNTTSDSTCFLTRLNNAPFDMLNVVSYVCEEHLHSFTEKGDATACPCYRLHKPTFISLNSQVRKTANMFMPGAFMDEIIGGT--NKISQNTV--LITD
CCMV           SSTNTTSDATCFLTRLNNAPFDMLNVVSYVCEEHLHSFTEKGDATACPCYRLHKPTFISLNSQVRKTANMFMPGAFMDEIIGGT--NKIAQNNV--LITD
RhCMV          SSTNTTSDATCFLTRLNNAPFDMLNVVSYVCEEHLHTFTEKGDATACPCYRLHKPTFISLNSQVRKTANMFMPGSFMDEIIGGT--NKITESNV--LITD
TuHV-1         SSTNTTTDATCFLTKLNNSPFDMLNVISYVCEEHIQVFSEKPDATACPCYRLHKPTFITINSEVRQTANMFLAGSFMDEIMGGT--NKIVEDTP--LITD
HHV6-B         SSTNSGNHSTSFLMKLNNSPFEMLSVVSYVCEDHAHMLNERGNATACSCYRLHKPKFISINAEVKKTANLFLEGAFIHEIMGGA--TCNVINDV--LITE
HHV-7          SSTNSGNHSTSFLTKLSNSPFEMLTVVSYVCEDHVHILNDRGNATTCACYRLHKPKFISINADVKKTADLFLEGAFKHEIMGGS--LCNVVNDT--LITE
BÊTA           SSTNσσδδσTτFLφπLδNτPFνMLδVαSYVCEνHαδ-φσνπσσATσCσCYRLHKPφFIσαNτδVπδTAδφFφ-GτFφδEIφGGσ---δ-σ-σσφ--LITν
PL_HV3         SSSNSSDQSTSFLYKLKSADERLLNIVSYVCNEHKEDFNLQDKIISCPCYRLHIPSYITIDTTIKNTTNLFMDGVFSSELMGD---NDNQHLAAFQVVSD
PL_HV1         SSSNSSDQSTSFLYKLKSADEKLLNVVSYVCSEHKEDFSLQDKIISCPCYRLHIPSYITIDTSIKNTTNLFMDGVFATELMGD---ADNQPSVAFQIVSE
HHV-8          SSVNSSDRSTSFLLNLRNAQEKMLNVVSYVCADHREDFHLQDALVSCPCYRLHIPTYITIDESIKTTTNLFMEGAFDTELMGEG--AASSNATLYRVVGD
HHV-4          SSVNSADQATSFLYKLKDAQERLLNVVSYVCQEHRQDFDMQDSMVSCPCFRLHIPSYITMDSNIRATTNLFLDGAFSTELMGDT--SSLSQGSLSRTVRD
GAMMA          SSσNSτDδτTSFLφδLπσAδEπφLNαVSYVC-νHπδDFδφQD-φαSCPCρRLHIPσYITφDδσIπσTTNLFφνGσFσTELMGν---σσ-δ--σα-δφV-ν
ET             SSσNσσδxσTσFLφxLδστxX-φLδαασYαC-νHx--φxxxσ-φφσCσCρxLδφPxρIσφδxxφπxTφδφFφ-σσFxδEαφGx----------x--φφ--
PSI-PRED       ecccccccchhhhhhhccccccceeeeeeeechhhhhhhhcccccceeecchhccccceeeecchhhhhhhhhhccchhhhhhhccc--cccccccc--cccc
```

```
                 440       450       460       470       480           490       500       510       520
                  |         |         |         |         |             |         |         |         |
HSV-1        SAGERFLLYRPSTTTNS--GLMAPDLYVYVDPAFTANTRASGTGVAVVGRYRD---DYIIFALEHFFLRALTGSAPADIARCVVHSLTQVLALHPGAFRG
HSV-2        SAGERFLLYRPSTTTNS--GLMAPELYVYVDPAFTANTRASGTGIAVVGRYRD---DFIIFALEHFFLRALTGSAPADIARCVVHSLAQVLALHPGAFRS
CeHV-2       PAGERFLLYRPSTTTNG--GLMAPELYVYVDPAFTANTRASGTGIAAVGRYRD---DFIVFALEHFFLRALTGSAADDIARCAAHSLTQVLALHPGVFRS
CeHV-1       PAGERFLLYRPSTTTNG--GLMAPELYVYVDPAFTANTRASGTGVAAVGRYRD---DFIVFALEHFFLRALTGSAADDIARCAAHSLTQVLALHPGVFRS
EHV-1        TAQDRFLLYRPSTVANQ--DIMSSDLYVYVDPAFTTNAMASGTGVAVVGRYRS---NWVVFGMEHFFLSALTGSSAELIARCVAQCLAQVFAIHKRPFDS
BHV-5        AAGEQFLLYRPSTAATR--GPWPERLYMYIDPAFTSNARASGSGIAVVGRHRG---SWLVLGLEHFFLPALTGSSAAEIARCAVRCFAQVMAVHRRRLDG
HHV-3        TARERFILYRPSTVANC--AILSSVLYVYVDPAFTSNTRASGTGVAIVGRYKS---DWIIFGLEHFFLRALTGTSSSEIGRCVTQCLGHILALHPNTFTN
MDV-2        SAIDRFLIYRPSTVNNQ--DPFSQDLYVYVDPAFTANTKASGTGVAVIGKYGT---DYIVFGLEHYFLRALTGESSGSIGYCVAQCLIQICAIHRKRFGV
ALPHA        σA-νδFαaYRPSTφσδ---σaφσ--LYφYVDPAFTσNφ-ASGTGαAaaGπρ-σ---σρaaφσφEHρFL-ALTGδττ--IτδCaσδδφ-δφφAaH---φ-σ
AD169        QSREEFDILRYSTLNTNAYDYFGKTLYVYLDPAFTTNRKASGTGVAAVGAYRH---QFLIYGLEHFFLRDLSESSEVAIAECAAHMIISVLSLHP-YLDE
CCMV         QSREEFDILRYSTLNTNAYDYFGKTLYIYLDPAFTTNRKASGTGVAAVGTYRH---QFIIYGLEHFFLRDLSESSEVAISECAAHMAVSILSLHP-YLDE
RhCMV        QSRDEFDILRYSTLNVNAQEHFGKTIYVYLDPAFTTNRKASGTGIAAVGLYRH---QFIIYGLEHYFLRDLSESSETAIAECAAHMLTSVINLHP-YINE
TuHV-1       ESREEFDIFRYSTMNKQLHPHLDRCLYVYLDPAFTSNKKASGTGIVALGRYRE---QYLIYGLEHYFLKDLLESSETAIAECAAHMLLAVLQLHP-FFLE
HHV6-B       QGQTEFEFFRYSTINKNLIPFLGKDLYVYLDPAYTGNRRASGTGIAAIGTYLD---QYIVYGMEHYFLESLMTSSDTAIAECAAHMILSILDLHP-FFTE
HHV-7        QGLIEFDLFRYSTISKQIIPFLGKELYIYIDPAYTINRRASGTGVAAIGTYGD---QYIIYGMEHYFLESLLSNSDASIAECASHMILAVLELHP-FFTE
BÊTA         δσ--EFνφφRYSTφσφδα--ρφσπδαYαYαDPAρT-NππASGTGaaAaG-Y-δ---QρaaYGφEHρFLδσL-δσSνφτIτECAτHMaφτaaδLHP-ρφ-E
PL_HV3       IAISQFEMVRIDTLSDAVGKMLNPIVHMYIDPAFTNNTGASGTGI-SIIACLHNQTK-VLLGCEHYFLSELTGSATVDIARCALSLIKSVALLHP-VVKE
PL_HV1       VAISQFEMVRIDTSSDGVSKNIHPIVHMYIDPAYTNNTGASGTGI-SIIGCVQN-CK-ILLGCEHFFLNELTGSASLDIGRCAVSLIKSVSLLHP-SIKV
HHV-8        AALTQFDMCRVDTTAQQVQKCLGKQLFVYIDPAYTNNTEASGTGVGAVVTSTQTPTRSLILGMEHFFLRDLTGAAAYEIASCACTMIKAIAVLHP-TIER
HHV-4        DAINQLELCRVDTLNPRVAGRLASSLYVYVDPAYTNNTSASGTGIAAVTHDRADPNRVIVLGLEHFFLKDLTGDAALQIATCVVALVSSIVTLHP-HLEE
GAMMA        -AaσQφνφφRaDT-σ--V---α---αρφYαDPAρTNNT-ASGTGα-ταφ-σ--σ-σπ-aaLGφEHρFLδνLTGσAσφδIτδCaφσφaδτaσφLHP--αδ-
ET           -σ--δφxφφRφσT-σ------φ---αρφYαDPAρT-Nx-ASGTGα-xαx-x-----x-aaφσφEHρFL-xL-xxτx--IτδCaxxxφ-xφxxαH---φ--
PSI-PRED     cchhhcceeeeccccccccccccceeeeeeccccccccccccceeeeeeeecc---ceeeeechhhhhhhccccchhhhhhhhhhhhhhhhhhcc-ccce
```

```
                530       540       550                 560       570       580       590       600       610
                 |         |         |                   |         |         |         |         |         |
HSV-1           VRVAVEGNSSQDSAVAIATHVHTEMHRLLASEGADAGSGPELLFYHCEPPGSAVLYPFFLLNKQKTPAFEHFIKKFNSGGVMASQEIVSATVRLQTDPVE
HSV-2           VRVAVEGNSSQDSAVAIATHVHTEMHRILASAGAN-GPGPELLFYHCEPPGGAVLYPFFLLNKQKTPAFEYFIKKFNSGGVMASQELVSVTVRLQTDPVE
CeHV-2          VRVAVEGNSSQDSAVAIAARLHAELHRLLASGGAGVAA-PELLFYHCEPPGGAVLYPFFLLHKQKTPAFEHFIKKFNSGGVMASQEVVSVTMRLQTDPVE
CeHV-1          VRVAVEGNSSQDSAVAIAARLHAELHRLLASGGAGVAV-PELLFYHCEPPGGSVLYPFFLLNKQKTPAFEYFIKKFNSGGVMASQEVVSVTMRLQTDPVE
EHV-1           VRVAVEGNSSQDAAVAIATNIQLELNTLRRADVVPMPG--AVLFYHCTPHGSSVAYPFFLLQKQKTGAFDHFIKAFNSGSVLASQELVSNTVRLQTDPVE
BHV-5           LFVAVEGNSSQDSAVAIALGVRRELDSLAASGAVPMPA--ETRFYHCRPPGSAVAYPFFLLQKQKTAAFDHFIRLFNSGRVVASQDLASLTMRLQTDPVE
HHV-3           VHVSIEGNSSQDSAVAISLAIAQQFAVLEKGNVLSSA--PVLLFYHSIPPGCSVAYPFFLLQKQKTPAVDYFVKRFNSGNIIASQELVSLTVKLGVDPVE
MDV-2           IKIAIEGNSNQDSAVAIATRIAIEMISYMKAAVAPTPH--NVSFYHSKSNGTDVEYPYFLLQRQKTTAFDFFIAQFNSGRVLASQDLVSTTVSLTTDPVE
ALPHA           α-αταEGNSσQDτAVAIτφ-α--δφ--φ--σ-σασ-σ----φ-FYHτ-σ-GσσV-YPρFLLδπQKTσAφνρFα--FNSG-αφASQνααS-TφδL-φDPVE
AD169           LRIAVEGNTNQAAAVRIACLIRQSVQSSTLI---------RVLFYHTPDQ-NHIEQPFYLMGRDKALAVEQFISRFNSGYIKASQELVSYTIKLSHDPIE
CCMV            LRIAVEGNSNQAAAVRIACLIRQSVQSTKLI---------RVLFYHTPDT-NRIQQPFYLMGHDKALAVEQFISRFNSGYIKASQEVVSYTIKLNHDPIE
RhCMV           LRIAVEGNSNQAAAVRIACLIRQNVQCNSLI---------PVLFYHTPDQ-NNIEQPFYLMGRDKRLAVEQFISRFNSGYIKASQEIVSYTIKLNHDPIE
TuHV-1          VRVTIEGNSNQAAAVRIACAIKQNILCNRHV---------VILFYHTPDQ-NEIEQPFYLLGREKRLAVESFIAKFNSGYIKASQELVSFTIKITYDPVD
HHV6-B          VKIIIEGNSNQASAVKIACIIKENITANKSI---------QVTFFHTPDQ-NQIAQPFYLLGKEKKLAVEFFISNFNSGNIKASQELISFTIKITYDPVE
HHV-7           LKIIIEGNSNQSSAVKIACILKQTISVIRYK---------HITFFHTLDQ-SQIAQPFYLLGREKRLAVEYFISNFNSGYIKASQELISFTIKITYDPIE
BÊTA            απαφαEGNσNQττAVπIACαaπδσα-σ---φ----------αφFρHTαDδ-σδI-QPFYLφGπνK-LAVE-FIσδFNSGYIKASQEaαSρTIKασρDPαν
PL_HV3          IHISVEGNTSQDATVAIANIIDECSC----I---------PAKFYHHADKTNGVMWPMYILGAEKSQAFESFICALNSGTLKASQSIISNTIKLSYDPVC
PL_HV1          VHISVEGNTSQDAAVAIANIIDECSH----I---------PIKFYHHADKAKGAMWPMYIMGSEKAQAFESFIYALNSGTLKASQSITSHTIKLSYDPVS
HHV-8           VNAAVEGNSSQDSGVAIATVLNEICP----L---------PIHFLHYTDKSSALQWPIYMLGGEKSSAFETFIYALNSGTLSASQTVVSNTIKISFDPVT
HHV-4           VKVAVEGNSSQDSAVAIASIIGESCP----L---------PCAFVHTKDKTSSLQWPMYLLTNEKSKAFERLIYAVNTASLSASQVTVSNTIQLSFDPVL
GAMMA           αδατVEGNσSQDτσVAIAσαασE-τ-----α---------PφφFφHφφDKσδτα-WPφYφφσσEKτδAFEδφIφAαNστσLδASQ-φφSδTIδαSρDPV-
ET              α-αxαEGNσσQστσVxIτx-α---x-----x----------φ-FφHx-σ--xxα-δPφρφφxxδK-xAφν-φα--φNστ-αxASQ-φφS-Tφδα-φDPα-
PSI-PRED        eeeeeccccchhhhhhhhhhhhhhcccccc---------eeeeeeeccc-cceecchhhhcchhhhhhhhhhhhhcccceeeehhhhhheeeeccchhh
```

```
                 620       630          640        650       660       670
                  |         |            |          |         |         |
HSV-1            YLLEQLNNLTETVSP---NTDVRTYSGKRNGA--SDDLMVAVIMAIYLAAQAGPPHTFAPITRVS-----
HSV-2            YLSEQLNNLIETVSP---NTDVRMYSGKRNGA--ADDLMVAVIMAIYLAAPTGIPPAFFPITRTS-----
CeHV-2           YLTAQLDNLTETLAP---NSDARSYSGKRNGA--SDDLMVAVIMAVYLAATPHAHHAFAPITRAAP----
CeHV-1           YLTAQLNNLTEILSP---NSDARSYSGKRNGA--SDDLMVAVIMAVYLAATPHAHHAFAPVTRAGQALER
EHV-1            YLLTQMKNLTEVVTG---TSETRVFTGKRNGA--SDDMLVALVMAVYLSSLPPTSDAFSSLPAQ------
BHV-5            YLFAQLQNLTESTAG---PGEARAFSGKRRGA--ADDLMVALVMAVFVGSLPPTDSAFAPLAARPPAD--
HHV-3            YLCKQLDNLTEVIKGGMGNLDTKTYTGKGTTGTMSDDLMVALIMSVYIGSSCIPDSVFMPIK--------
MDV-2            YLTKQLTNISEVVTG---PTCTRTFSGKKGG---NDDTVVALTMAVYISAH-IPDMAFAPIRV-------
ALPHA            YL--QφδNα-E-φ-σ-----δφπ-ρσGK--σ---σDDφφVAαφMταραττ---φ--φFφσα---------
AD169            YLLEQIQNLHRVTLA---EGTTARYSAKRQN-RISDDLIIAVIMATYLCDDIHAIRFRVS----------
CCMV             YLLEQLQNLHRVTLA---EGGAGSYSAKRHQ-RFSDDLIIAVIMATYLCDEMHAIHFRVS----------
RhCMV            YLLEQIQNLHRITLA---EGNT-TYSAKRHH-KYSDDLVVAVIMATYLCDDIYTARFKAS----------
TuHV-1           YLLEQLKNIQRITVN---EYVT--YNAKKHN--CADDLAIALIMAVYLCLDERGYHFKPI----------
HHV6-B           YALEQIRNIHQISVN---NYIT--YSAKKQA--CSDDLIIAIIMAIYVCSGNSSASFREI----------
HHV-7            YVIEQIKNLHQININ---EHVT--YNAKKQT--CSDDLLISIIMAIYMCHEGKQTSFKEI----------
BÊTA             YααEQαδNαδδασασ---δ--φ--YσAKπδ---φτDDLααταIMAφYφC---δ-φδFπ------------
PL_HV3           YLIEQIRSIKCIPLK---DGGH-TYHAKAKN--MSDDLLVAVVMAMHLSNSNSLPFKALHNKRFF-----
PL_HV1           YLIEQIRAIKCITLK---DGGH-TYHAKTKN--MSDDLLVAVVMSSHLANSSTLPFKALHNKRFF-----
HHV-8            YLVEQVRAIKCVPLR---DGGQ-SYSAKQKH--MSDDLLVAVVMAHFMATDDRHMYKPISPQ--------
HHV-4            YLISQIRAIKPIPLR---DGTY-TYTGKQRN--LSDDVLVALVMAHFLATTQKHTFKKVH----------
GAMMA            YLαδQαRτIKσαφLπ---DGσδ-σYδτK-πδ--φSDDαLVAαVMτ-ρφτσσδδφφρKφαδ----------
ET               Yα--Qφδσα-x-x-x------x--ρxτK------σDDφφαταφMτ-ρφτ-------φx------------
PSI-PRED         hhhhhhhcceeeecc---ccceeeeeccccc-ccchhhhhhhhhhhhhcccccccceeecc----------
```

**Alignment data :**

Alignment length: 770
Residues conserved for 90 % or more (upper-case letters): 131 is 17.01 %
Residues conserved for 50 % and less than 90 % (lower-case letters): 153 is 19.87 %
Residues conserved less than 50 % (white space): 451 is 58.57 %
IV conserved positions (!): 15 is 1.95 %
LM conserved positions ($): 6 is 0.78 %
FY conserved positions (%): 6 is 0.78 %
NDQEBZ conserved positions (#): 8 is 1.04 %
HSV-1: NP_044616 (735 residues) | AD169: CAA35363 ( 674 residues) | PL_HV3: AAO12336 ( 684 residues)
HSV-2: NP_044484 (734 residues) |CCMV: AAM00728 ( 672 residues) | PL_HV1: AAM22132 ( 683 residues)
CeHV-2: YP_164457 (735 residues) | RhCMV: AAT07420 ( 671 residues) | HHV-4: CAA24834 ( 690 residues)
CeHV-1: AAP41433 (738 residues) | TuHV-1: AAK57138 ( 673 residues) | HHV-8: AAB62677 ( 687 residues)
EHV-1: YP_053090 (734 residues) | HHV6-B: AAD49667 ( 666 residues) |
BHV-5: NP_954931 (737 residues) | HHV-7: YP_073802 ( 663 residues)
HHV-3: NP_040165 (747 residues)
MDV2: BAA82909 (737 residues)

## B.2. ANNEXE B2: Alignement de séquence de pUL56 et 22 de ses homologues herpesviraux

```
                                        10        20        30        40        50        60        70
                                         |         |         |         |         |         |         |
HSV-1       -----------------MAAPVSEPTVARQKLLALLGQVQTYVFQIELLRRCDPHIGR--GKLPQLKLNALQVRALRRRLRPGLEAQAG-AFLTPLSV
HSV-2       -------------MAAAPPAAVSEPTAARQKLLALLGQVQTYVFQLELLRRCDPQIGL--GKLAQLKLNALQVRVLRRHLRPGLEAQAA-AFLTPLSV
CeHV-1      -----MASGPTAAASPTDGDAVSEPTAARQKLLALFGQVQTYVFQLELLRRCDPRIGR--GKLAALKLNALQVRAVRRRLRPGLAAQAR-AFLTPLSV
CeHV-2      MSAPSGVAAASAPASAADDAAVSEPTAARQKLLALFGQVQTYVFQLELLRRCDPRIGR--GKLAALKLNALQVRAVRRRLRPGLAAQAR-AFLTPLSV
EHV-1       -------------MEQDDAPAAMGSAQARQRLLAIFGQVQAYIFQVEMLKRCDPLALL--PLVGSLKLNALTIRMLRRKLGGALIEQAQ-HQQTPLAC
MDV-2       -------MLGMSHNRLQSQGTEHDK-FATQKLFAIWGQIQSYLFQVELLKRCDPTVGV--RMINRLKLNVLMIYYLEKKMVPALKEQRE-MNLTPLTY
GaHV3       ----------MPHGNAQALDVERDSIFASQKLLAVWGQLQSYLFQVELLKRCDPRVGV--RMIKPLKLNVLMVRYLENLMAPGLTEQLE-RNISPLSY
MEHV-1      ----------MEHGPHQPSELEYEGSIARQKLLAVWGQIQSYLFQIEILKRCDPVVGM--RMINRLKLNVLMIRYLEKKLLPGWKEQED-ISITPITY
HHV-3       -----------------------MELDINRTLLVLLGQVYTYIFQVELLRRCDPRVAC--RFLYRLAANCLTVRYLLKLFLRGFNTQLK-FGNTPTVC
SUID        --------------------------MAERRLVAVLGQVQTYVFQLEMLKRCDPAVVR--ELAPRVKLNALMCRYLARRL------PLE-AQTTPLTC
GaHV1       -----------------------MAEETQLRLMALAGQLQTLLFQIDILKRCDPEILLQKRLRSRIKHNALMALYIHSLLSNELSMQVN-YRLTLQTY
GaHV2       -----------------------MAEETQLRLMALAGQLQTLLFQIDILKRCDPEILLQKRLRSRIKHNALMALYIHSLLSNELSMQVN-YRLTLQTY
ALPHA       ---------------------------φδ-πLφAαφGQαδσφαFQανφLπRCDP-α-----φ---αφφNσL-φ-φα-δ-φ-------------σα-σφ
AD169       -----------------------MEMNLLQKLCVVCSKCNEYAMELECLKYCDPNVLL--AESTPFKRNAAAIVYLYRKIYPEVVAQNR-TQSSLLTL
CCMV        -------------------------MNLLQKLCVVCSKCNEYAMELECLKYCDPNVLL--AECTPFKRNAAAIVYLYRKIYPEVVAQNR-TQSSLLTL
RhCMV       -----------------------MAMNTLQKLCVVCSKCNECAMDVECLKYCDPNIVS--MDSTAFRRNGVMVIHLYRTLYPALVSQNA-VQTSVLTL
MCMV        -----------------------MAMNTLQKLCVVCSKCNECAMDVECLKYCDPNIVS--MDSTAFRRNGVMVIHLYRTLYPALVSQNA-VQTSVLTL
RCMV        -------------------------MNTLQKLCVVCSKCNECAMDAECLKYCDPSIVL--MDSTAFQRNGATVIHLYRTLYPALVAQNT-AQTSVLSL
Tupaii      -------------------------MNTLQRLCVVCSKCNECAMELECLKYCDPAIVL--VDSAPFKRNALTVVHLYRRLCPALAEQNARYQASLITL
HHV-7       -------------------------MNSLQSLCVLCSRLSECALELECLRFCDPVTLI--PDMTNFRKNGVVIIHLFKTLFAELCHQNF-NCASPVTI
BÊTA        -------------------------MN-LQπLCVαCSπφδEφAφναECLπρCDPσφα---φν-σσFδπNταφααρLρπδαφα-ασ-QN---δσSαασα
HHV-8       ---------------------------MAKELAAVYADVSALAMDLCLLSYADPATL----DTKSLALTTGKFQSLHGTLLPLLRRQNA-HE SGLSL
HHV-4_1     ---------------------------MSGLLAAAYSQVYALAVELSVCTRLDPRSL----DVAAVVRNAGLLAELEAILLPRLRRQND-RA SALSL
HHV-4_2     ---------------------------MSGLLAAAYSQVYALAVELSVCARLDPRSL----DVAAVVRNAGLLAELEAILLPRLRRQND-RA SALSL
GAMMA       ---------------------------Mτ--LAAαYτδVδALAφνLδαφσδαDP-σL----Dσφταα-δσGφφ-δLδτφLLP-LRRQNσ-π-CSτLSL
ET          ----------------------------x--LxσαφτδφδxφαφδαδφφxxφDP-x-----x---φx-δσxφ--xφ-x-φ-------------σ----
PSI-PRED    ----------------------cchhhhhhhhhhhhhhhhhhhhhhhhcccchhhhh--hhhhhhhhhhhhhhhhhhhhhhhhhhhh-hcccchhhh
```

```
                 80        90        100                 110         120        130       140       150
                  |         |         |                   |           |          |         |         |
HSV-1       TLELLLEYAWREGERLLGSLETF---ATAGD------VAAFFTETMGLARP--CPYHQRVRLDTYG--GTVHMELCFLHDVENFLKQLNYCHLITPSR
HSV-2       TLELLLEYAWREGERLLGHLETF---ATTGD------VSAFFTETMGLARP--CPYHQQIRLQTYG--GDVRMELCFLHDVENFLKQLNYCHLITPPS
CeHV-1      TLELLLEYAQREGERLLGCLEAFGEGGEEGD------VAAFFVESMGLARP--CPYHQPIRLETYG--GCVDMELCFLHDVENFLKQLNYCHLITPPR
CeHV-2      TLELLLEYAQREGERLLGCLEAFGEGAGAGD------VAAFFAESMGLARP--CPYHQRIRLETYG--GGVDTELCFLHDVENFLKQLNYCHLVTPPR
EHV-1       ALTMALEYAEVEGERVLRAVDDVNLAGPEG----------FFRATMRLDEP--CEYHVRVHLDTYG--GPIDAEVQFLHDAENFLKQLNYCHLITGFE
MDV-2       GLWLALRRAKLEGELLLDALCEF---KDGGN------LRDFFRKSMSMCGD--CPYHSTVELDTYG--GKVSTEIKFLHDVENVLKQLNYCHLILKAD
GaHV3       GLWLALRRARLDGEVLLRALCEF---TDTGD------VHGFFIKSMTISGD--CPYHEVVELDTYG--GSVRTEIKFLHDVENILKQLNYCHLIVKAG
MEHV-1      GVWSAIRRARLEGEVLLSALCDY---VVEGN------TRNFFVKSMSLQGV--CPYHVVAELDTYG--GPVTTEIKFLHDVENVLKQLNYCHLIINSR
HHV-3       ALHWALCYVKGEGERLFELLQHFKTRFVYGETKDSNCIKDYFVSAFNL-KT--CQYHHELSLTTYG--GYVSSEIQFLHDIENFLKQLNYCYIITSSR
SUID        ALRLALAYARAEGDRVLGAL------AAAGDD-----AEAYFERTMG--GA--CRFHARVALDTYG--GRVETELQFLHDAENLLKQLNYCHLITPHA
GaHV1       CVWLWLRRACSEIAALCNEIDKYSVLKDR---------KCFFEATLGYGKGFICPLHAEVQLQLYGMSSIISKELVLINDVENFMKQLNYCHCVVSCK
GaHV2       CVWLWLRRACSEIAALCNEIDKYSVLKDR---------KCFFEATLGYGKGFICPLHAEVQLQLYGMSSIISKELVLINDVENFMKQLNYCHCVVSCK
ALPHA       σαδ-φα-δαδ-ν---αφ--α--------σ-----------ρF--σφ-------C-φH--α-LδφYG--τ-α--Eα-φαδDαENφφKQLNYCρφαφ---
AD169       YLEMLLKALHEDTALLDRALMAYSRQPDR---------AAFYRTVLRLDR---CDRHHTVEL-QFTDNVRFSVSLATLNDIERFLCKMNYVYGILAPE
CCMV        YLEMLLKALHEDTALLDRALTAYSRHPDR---------AAFYRTVLRLDH---CDRHHTVEL-RFTDSVRFSVSLATLNDIERFLCKMNYVYGILAPE
RhCMV       YMEMLLQGLYDTMREIDMALTDFGTHRDR---------QRYYRRVLKLDS---CNRHESITI-TFAPELALTIDLATLNDVERLLCKINCVYGAVDAS
MCMV        YMEMLLQGLYDTMREIDMALTDFGTHRDR---------QRYYRRVLKLDS---CNRHESITI-TFAPELALTIDLATLNDVERLLCKINCVYGAVDAS
RCMV        YLEMLLAGLYRTMRELDTALTDFARHRDR---------ERYYRRILNLDS---CP--ESVTL-VFAPGLHLTIELATLNDIERLLCKINCVYGALDRT
Tupaii      YLEMLLRCLYEDVLLVDEALGEFERHGDR---------QRYYRRVLRLDR---CACHDTLEV-TFTERIRLTVDVATLNEVERLLCKINCVYGVLEPT
HHV-7       YLQILLKAMYNQVLLLDASIHQFLLDNDK---------QKYFENIFQLNE---CKQHLKLDL-ALNNYLTFSVDISTINDIEKLLCKMNCIFGLISPL
BÊTA        YφδφLL-σφρδδφ--αD-τα--ρ--δ-Dπ-----------ρρδδαφδLδδ---C-----αδα--φσ--α-φσαδατTLNναEπφLCKφNφαρGαα---
HHV-8       ELEHFWKTWLMLWPRWECALAENCLQKSI---------------FPSCIWTQHATSNRSVRFNFYG-NWALELKLSLINDVEIFFKRLSSVFYCIGSG
HHV-4_1     ELVHLLENSREA----SAALLAPGRKGTR---------------VPPL-----RTPSVAYSVEFYG-GHKVDVSL LINDIEILMKRINSVFYCMSHT
HHV-4_2     ELVHLLENSREA----SAALLAPGRKGTR---------------VPPL-----RTPSVAYSVEFYG-GHKVDVSL LINDIEILMKRINSVFYCMSHT
GAMMA       EL-Hφφδδδ--α----δτALα-στ-δ-σ----------------φPσφ------Tσσ-σφδφδFYG--ρφανα-LτLINDαEIφφKRασSVFYCφτ-σ
ET          -l--ll----------x--α-----------------------xφ-------------φ-φ--φσ--x-φ---α-φαδναExφφφδφσxxρxφφ---
PSI-PRED    hhhhhhhhhhhhhhhhhhhhhhhcccccch---------hhhhhhhhhccc---ccccceeeee-eeecceeeeeeeeeehhhhhhhhhhhcceeeeecch
```

```
                 160       170       180       190       200       210       220       230       240       250
                  |         |         |         |         |         |         |         |         |         |
HSV-1            GATAALERVREFMVGAVGSGLIVPPELSDPSHPCAVCFEELCVTANQGATIARRLADRICNHVTQQAQVRLDANELRRYLPHAAGLSDADRARALSVLD
HSV-2            GATAALERVREFMVAAVGSGLIVPPELSDPSHPCAVCFEELCVTANQGATIARRLADRICNHVTQQAQVRLDANELRRYLPHAAGLSDAARARALCVLD
CeHV-1           GAIAALERVRAFLVGAVGSGLVVPPEISDPSHPCAVCFEELCVTANQGATIARRLADRICNHVTQQAQVRLDDNELRRYLPHASGFSEAARARALRVLD
CeHV-2           GAVAALERVRDFLVGAVGSGLIVPPEISDPSHPCAVCFEELCVTANQGATVARRLADRICNHVTQQAQVRLDDNELRRYLPHAAGFSEAARARALRVLD
EHV-1            AGLDALESVARFLTRTVGSGIVVPPELCDPTHPCSVCFEELCVTANQGEAVHRRLLECTCDHITRQMAVRVANIDIARHLPHALSVASERRAAAEAALR
MDV-2            TVENFMVSLDNYLLKTLGSGSVVPPELYDPSQPCSVCFEELCVTANSGDSTHKRIVRKICDHITKQINIRVNSDDMVTHLPHATYVPDDKRTTAQTALD
GaHV3            TAENFMLALDRYLIETLGAGSVVSPELCDPSQPCSVCFEELCVTANSGDSAHKRIIGKICNHVTKQIVLRVHPDDMVAHIPHAACLSDEKRSVVQAALS
MEHV-1           SVGRFMASIDKYLLKTLGSGSIVPPELYDPSQPCSVCFEELCVTANSGDSAHKRIVGKMCDHVTKQMILRVDPDDMVAHLPHATYILPDKLAVATTALN
HHV-3            EALNTLETVTRFMTDTIGSGLIPPVELFDPAHPCAICFEELCITANQGETLHRRLLGCICDHVTKQVRVNVDVDDIIRCLPYIPDVPDIKRQSAVEALR
SUID             VD---LSAVDEFLARTIGGGLVVPPELYDPAQPCAVCFEELCVTANQGEATHRRLLGCVCDHLTRQLAVRVDPEDVAKNLPHVHGLDEARRGRALAALA
GaHV1            NAIEALNTIDEFLVLSSGGSLVAAPDTYDHGQPCYVCLEELSVTANQGETIYRRLGYKICDHLTRQFPVNVSTDDLKRHLPFLRDVDPEKLKNVLS---
GaHV2            NAIEALNTIDEFLVLSSGGSLVAAPDTYDHGQPCYVCLEELSVTANQGETIYRRLGYKICDHLTRQFPVNVSTDDLKRHLPFLRDVDPEKLKNVLS---
ALPHA            -σ---φ—α---ρφφ-σ-Gττ-αασανφ-DφσδPC-αCφEELσαTANδG-σαφπRα---φCDHαTδQφ-αδα-σ-νφ---αPρα--φσ------α-----
AD169            AGLEV AQLLELLRRLCGISPVARQEVYVEGTTCAQCYEELTIIPNQGRSLNKRLQGLLCNHIAVHRPS-SQSDVNIQTVEQDLLDLTTRIPHLAGVLS
CCMV             AGLEV TQLLALLRRLCGISPVARQEAYVEGTTCAQCYEELTIIPNQGRSLNKRLQGLLCNHIAVHRPS-SQSDVNIQTVEQDLLELTERIPRLAGVLS
RhCMV            QGVAV RRLLSLLARLCDICPVAGPEIYRETVTCFQCYEELMAVPNQGRSINRRMQGLLCDHITIKKVL-VQLDMDAQAVEQDMGDIAIRAPSVKGIIR
MCMV             QGVAV RRLLSLLARLCDICPVAGPEIYRETVTCFQCYEELMAVPNQGRSINRRMQGLLCDHITIKKVL-VQLDMDAQTVEQDMGDIAIRAPSVKGIIR
RCMV             QGVEV RRLLSLLGALCDTSPVAGPETYRETSTCYQCYEELMAVPNQGRSINRRIEGLLCDHVTVKKVL-VQLDTDAQTVEQDLGEVAIRAPRVKSIIR
Tupaii           RGLEL RRLLSLMGRLCGISPVAAPEAYVENLTCLQCYEELAAVPNQGRSILKRLRGLLCDHLTVRKSL-VQLETGIQTMEQDIVETVGSRPRLSSLLE
HHV-7            DGINA SQIIEFLTILCGVCVVMKPEVFSETTTCLKCYEELSLVPNQGKSIRKRLAGKFCNHLTETHMV-SNLEKNVDIIEKDLDFSTKQYGLVKEYMA
BÊTA             -Gα-αC-δαα-φφ--LCσφσαVφ--Eαρ-Eσ-TCφδCYEEL-ααPNQGπSα-πRφ-GφφCσHασ-δπ----δ-νφσαδφφEδDφ---σ-δ-σ-α--φφ-
HHV-8            SALEGLGEVLRFVGKLRGISPVPGPDLYVSNLPCLECLQEVCLTPNQGTSLQAMLPDTACSHICTPACGEPVRGLFENELKQLGL-----QTPESI---
HHV-4_1          MGLESLERALDLLGRFRGVSPIPDPRLYITSVPCWRCVGELMVLPNHGNPSTAEGTHVSCNHLAVPVNPEPVSGLFENEVRQAGLGHLLEAEEKAR---
HHV-4_2          MGLESLERALDLLGRFRGVSPIPDPRLYITSVPCWRCVGELMVLPNHGNPSTAEGTHVSCNHLAVPVNPEPVSGLFENEVRQAGLGHLLEAEEKAR---
GAMMA            -τLEτL-δαLδφαGπφRGαSPαPσPδLYαδσαPCφδCα-EαφαφPNhGσσ-δA--σδφτCσHατφPασσEPVδGLFENEαπQαGL--------δτ----
ET               -σ---φ-x---φφ--x-σxσ--φ--δφ--xσ-σC-xCφ-Eα-αφσNδG-σ--x-----xCσHα---x-------xx---φxxx----------x-----
PSI-PRED         Hhhhhhhhhhhhhhhcccccccccccccccccchhhhhhhheeccccchhhhhhhcccccccccccccc-cccccchhhhhcchhhhhhcccccccchhh
```

```
                    260              270       280       290       300                     310       320
                     |                |         |         |         |                       |         |
HSV-1       HA-----LARTAGGDGQPH--PSPENDSVRKEADALLEAHDVFQATTPGLYAISELRFWLAS----------GDRAGQTTMDAFASNLTALARRELQQE
HSV-2       QA-----LARTAAGGGARAGPPPADSSSVREEADALLEAHDVFQATTPGLYAISELRFWLAS----------GDRARHSTMDAFADNLNALAQRELQQE
CeHV-1      QA-----LERAAAGEAQGR---APADESARREADALLEAHDVFQATTPGLYAISELRFWLAS----------GDRAHQSTMDAFADNLTELARRELRHE
CeHV-2      QA-----LERTAAGASRER---GPADESARREADALLEAHDVFQPTTPGLYAVSELRFWLAS----------GDRAHQSTMDAFADNLTELARRELRHE
EHV-1       AL-----EARRVQGHNGKS---AGTEDPTQQVASRLLESHHVFKPASRCLYAVSELKFWLAST-K------HGDMGQPRAIDTFTENLETLDKQEKFFH
MDV-2       VI-----QSTMRDTTTEND----SN-ISVSKAAAAALDAHNVFLPASGDLYAISELQFWIASSGR------KLHQPRGNTVESFADNLEALVSKERLFD
GaHV3       LI-----RPTRGDKDPGGS----GDTVSVADAATAVLDAHNVFLPASGNLYAISELQFWIASSER------KPNQPRANTIDSFADNLEDLAFKERLFD
MEHV-1      HI-----NNPNINDTVSTN----VSPDSVSQAAMVALDAHNVFLPASGDLYAISELQFWTASSGR------KASHPQGNTIDSFADNLEQVVAKERLFD
HHV-3       TL-----QTKTVVNPMGAK---NDTFDQTYEIASTMLDSYNVFKPAPRCMYAISELKFWLTSN---------STEGPQRTLDVFVDNLDVLNEHEKHAE
SUID        AVDAAEAREAEAASTAAAGAEAGDAGETARRRADALLDAHDVFRPASRRLYAVSELQFWLAS-----------TNQAVRALDLFTHNLDDLERRERRAE
GaHV1       -------NETRIEAENHVEAGRTLSSSPGESEASKILNKYDVFTEPPGPVYQLSELQYWLASGKRTPSGVESQTTTSQSVLKQLDKDLSELFKRAEDFE
GaHV2       -------NETRIEAENHVEAGRTLSSSPGESEASKILNKYDVFTEPPGPVYQLSELQYWLASGKRTPSGVESQTTTSQSVLKQLDKDLSELFKRAEDFE
ALPHA       ----------------------------σ-δ-A--φLδ-ρδVF--σσ--φY-αSELδρWφσS----------------δφφδ-φσδδLδ-α--δ----δ
AD169       AL-------KSLFSSSSA------YHS-YIQEAEEALREYNLFTDIPERIYSLSDFTYW--------------SRTSEVIVKRVGITIQQLNVYHQLCR
CCMV        AI-------KSLFSSSSA------YHS-YIQEAEEALREYNLFTDIPERIYSLSDFTYW--------------SRTSEVIVKRVGITIQQLNVYHQLCR
RhCMV       AI-------KSLASFSPA------SYA-YINDAEEALRGYNLFSEIPDRIYSLSDYTYW--------------SRTSEVIVKRVGVTIQQLNVYHHLCR
MCMV        SL-------KSLASFSPA------SYA-YINDAEEALRGYNLFSEIPDRIYSLSDYTYW--------------SKTSEAIVRHVGITMRQLNVSHSLWK
RCMV        LL-------KNLASSSAG------SCA-YINDAEEALRGYNLFSEIPERIYSLSDYTYW--------------SKTSEAIVRHVGITMRQLNVSHSLWK
Tupaii      KI-------KGLSSSAAV------SHA-YISEAEDTLRRYNLFTDIPPRIYSLSDFTYW--------------SKTSEVIVRHVGVTMRQLNASHGLWK
HHV-7       KI-------TNIFQQQLY------SKPPHLQEAENTLINFDLFSKIPDTIYSLSEFTYW--------------SKISESVIQKASITLNQLNLCHSLYA
BÊTA        -α-------φσα-δ----------δδ--ραδνAEδφL--ρδLFσδIP-δIYSLSνρTYW--------------SπφSEσααδπαταTφδQLNαδH-Lφ-
HHV-8       ---------PTTPCQSRVR----QDDEIR-QSSLMAVGDHHIFGEVTRSVLEISNLIYWSSGH--------SDATCDGDRDC---SHLASLFTHEADMH
HHV-4_1     ---------PGGPEEGAVP----GPGRPEAEGATRALDTYNVFSTVPPEVAELSELLYWNSGG--------HAIGATGQGEGGGHSRLSALFARERRLA
HHV-4_2     ---------PGGPEEGAVP----GPGRPEAEGATRALDTYNVFSTVPPEVAELSELLYWNSGG--------HAIGATGQGEGGGHSRLSALFARERRLA
GAMMA       ---------PσσPδδτ-V-------σδαδ-δττφ-AασσρδαFτδVσ-δVαEαSδLαYWσSG---------δσασφσGδ-ντ---SπLττLFφπE-δφφ
ET          ----------------------------x-δ-τ--φL--ρδαF--xσ--φφ-αSδφxρW-------------------x-xx---xδφx-α--δ-----
PSI-PRED    hh-------hhhcccccc------cch-hhhhhhhhhhhcccccccchhhhhhhhhhhh--------------cccccchhhhhccchhhhhhhhhhhh
```

```
                330       340       350            360       370       380       390       400       410
                 |         |         |              |         |         |         |         |         |
HSV-1         TAAVAVELALFGRRAE-HFDRAFGSHLAAL--DMVDALIIGGQATSPDDQIEALIRACYDHHLTTPLLRRLVSPEQCDEEALRRVLARMGAGGAADAPK
HSV-2         TAAVAVELALFGRRAE-HFDRAFGGHLAAL--DMVDALIIGGQATSPDDQIEALIRACYDHHLTTPLLRRLVSPEQCDEEALRRVLARLGAGGAT----
CeHV-1        TAVAAVELALFGRRAE-HFERAFGDRLAAL--DMVDALLVGGQATSPEDQIEALIRACYDHHLTTPLLRRLVNPEQSDEEALRRVLARLGGG-------
CeHV-2        TAVAAVELALFGRRAE-HFERAFGERLAAL--DMVDALLVGGQATSPEDQIEALIRACYDHHLTTPLLRRLVSPEQSDEEALRRVLARIGGG-------
EHV-1         LQAATVELALFGRTLD-HFDRLFADQLLGL--DVIDGMLVGSCAVSPDDHIEALIKACYTHHMSAPLLQRLTDPDTSNREALKQLLGRIGVDTD-----
MDV-2         LRTSIVETAVFDRRMD-HFERVFAQEIEHM--NAADRLLLGGRAAAPDDIIEALIKACYDHHMSAPLLKRLLYPDEAAHDALKTVLERVSSHCIGNDIQ
GaHV3         LRTSVVEMTVFGRRMD-HFERVFAQEIENM--NAADRLLLGGRAVAPDDIIEALIRACYDHHMSTPLLKRLLYPEEAAQDALKTTLEQVTSHSTRNGPL
MEHV-1        LRSSVIETALFDKRMN-HFERIFSDEIEQM--NTADRLLLGGRTAAPDDIIEALIKACYDHHLSAPLLKRLLYPDEAAHDALKSTLETLTSRSVR----
HHV-3         LTAVTVELALFGKTPI-HFDRAFSEELGSL--DAIDSILVGNRSSSPDSQIEALIKACYAHHLSSPLMRHISNPSHDNEAALRQLLERVGCED------
SUID          VRAAAVELALFGRRPE-HFDRARAAR--EL--DIIDGLLVGGCAASPDERLEALIRACYDHHMSTPML-RMLDPDRANRDALERLLEGGDDA-------
GaHV1         QECIRLERTLFEKVHA-HFHRVAQAELISPQHSLVDTLITGVLATSPDTEIAALIKACYDHHLSLPLFRRLRNPEKVDTDALEELVEKLRGSD------
GaHV2         QECIRLERTLFEKVHA-HFHRVAQAELISPQHSLVDTLITGVLATSPDTEIAALIKACYDHHLSLPLFRRLRNPEKVDTDALEELVEKLRGSD------
ALPHA         --σ--αE-φαF-π-φ--HFδRα---δ---φ--σφαD-φαφGσ-σστPνδ-α-ALIπACYσHHφσ-Pφφ-πφ-δPδδσσ--ALδδφα----σ--------
AD169         ALMNGISRHLYGEDVE-DIFVLGEKALDG-----EERMFVGSVFAAPNRIIDLITSLSIQAFEDNPVFNKLHESNEM-YTKIKHILEEIRRPLPDGTG-
CCMV          VLMNTISRHLYGEDVE-DIFALGEKVLAG-----EERMFVGSVFAAPSRIIDLITSLSIQAFEDNPVFNKLHESNEM-YTKIKRILEEIRRPLPDGTG-
RhCMV         TLRTELSRYHYGEDLE-DVFTLGEGRFGG-----DERIYVGSIFAAPGKVVDMITSMSIKSFENNPLFNRLHESNEI-YAKIKSLIEEIRGVGDG----
MCMV          TLRTELSRYHYGEDLE-DVFTLGEGRFGG-----DERIYVGSIFAAPGKVVDMITSMSIKSFENNPLFNRLHESNEI-YAKIKSLIEEIRGVGDG----
RCMV          ALRGELSRFLYGEAVE-DIFSLGEGRFAG-----HERMFVGSISAAPGKVVDMITSMSVKSFENNPLFNRLHESNEI-YAKIRSLIEEIRDADDPGSTA
Tupaii        TLRNELGQYLYGDRVE-DLFTLSEAQLGE-----DERLYVGSIYAAPERIVDLMTSLSLQSFENNPVFNKLHENNEI-YSKIRSLLEEIRRPLDG----
HHV-7         DLQNEISKFLYGETIQ-DVFNFNEENVTN-----DDKLYIGSRFISPCRLVDIITNVSIKNLEEDPVFTKLAEEDEI-QTKIKTLLNELEN--------
BÊTA          σL-σ-ατδρφYGν-αδ-DαFσφσE--φσ------δνπφραGS-ρατP-ππαDφφTσφSαδσφEδσPαFσπLφEδνEφ--σKIπ-ααδEαδ---------
HHV-8         KRRVDLAGCLGERGTPKHFFDCFRPDS-------LETLF GGLFSSVEDTIESLQKDCSSAFYQQVNYTTALQKQNEFYVRLSKLLAAGQL--------
HHV-4_1       LVRGACEEALAGARLT-HLFDAVAPGA-------TERLF GGVYSSSGDAVEALKADCAAAFTAHPQYRAILQKRNELYTRLNRAMQRLGR--------
HHV-4_2       LVRRACEEALAGARLT-HLFDAVAPGA-------TERLF GGVYSSSGDAVEALKADCAAAFTAHPQYRAILQKRNELYTRLNRAMQRLGR--------
GAMMA         φ-R-σφ--φLτ---φφ-HφFDφφ-Pσσ-------φEδLFCGGαρSSσ-DφαEσLδ-DCσσAFφ-δαδYδσαLQKδNELYσRLδπαφ-------------
ET            -----φ--φφx---φ--xφxxφ------------xν-φφφGσ-xxτσ-δ-α-xφx-xσxxxφx--αxφ-xφ-δxδδx---xαδ-φφ-------------
PSI-PRED      hhhhhhhhhhcccchhh-hhcccccccccc-----chheeeccccccchhhhhhhhhhhhhhhhhhhhhhhhhchhhh-hhhhhhhhhhcccccccc-
```

```
                                                 420       430       440       450       460       470
                                                  |         |         |         |         |         |
HSV-1        G-------------------------------------------------------GAGPDDDGDRVAVEEGARGLGAPGGGGEDEDRRRGPGGQ
HSV-2        -------------------------------------------------------GGAEEEEPRAAAEEGGRRRGAGTPASEDGERGPEPGAQ
CeHV-1       -------------------------------------------------------RRDAGEGEDGAAGEAGEGAAAPEAGRGDGRE------D
CeHV-2       -------------------------------------------------------RGDAGE--TGEEGEGGAGAAAGEPAR----E------A
EHV-1        -------------------------------------------------------DGAGELGDALDVDLDNLGGAPPVNSTPCGEDALCRTVS
MDV-2        C------------------------------------------------------QDGDGTCGERMNETGHFRTNDSFAMSTT---------SL
GaHV3        C------------------------------------------------------QDGDGAPEDDFSNSEHLRIREGVTPLPD---------VR
MEHV-1       -------------------------------------------------------QDSAPKTSESTGDTTQVTPRDAHRSLSD---------AR
HHV-3        -------------------------------------------------------DLTKEASDSATASECDLNDDSSITFA------------
SUID         -------------------------------------------------------DADGGAAGGADAGDGGVGDEDGPGAPP---------PA
GaHV1        -------------------------------------------------------KETQPVTEGTSTSDLTANLRGTIE-------------R
GaHV2        -------------------------------------------------------KETQPVTEGTSTSDLTANLRGTIE-------------R
ALPHA        ----------------------------------------------------------σ----------------------------------
AD169        --------------------------------------GDGPEGEAIHLRGREAMSGTGTTLMTASNSSNSSTHSQRNNGGGGRARGGGKKVVGGGV
CCMV         --------------------------------------GDGPERGT----SRHPQMHPKTGSRTEEEESMEGTGTYGHNGSS-RFRGRGRRSAAGGA
RhCMV        ------------------------------------------------------PAAGAARSRAEAASGAGAGGEEGAGA------AAGRGNTG
MCMV         ------------------------------------------------------PAAGAAASRAEAASGAGAGGEEGAGA------AAGRGNTG
RCMV         ATGTSGGTSGGTSGGTWGGTSGGISGGTAPAPGTEVPERGRRSSSHGSSTSRPGTSRRAAGGAGRSRRGADTSAEGSGSGAAGARDADSRCGGGGVTR
Tupaii       -------------------------------------------------------AGTRGAAAATAAGARGAAEGTGGAAAAGGAAAAAGEAA
HHV-7         ------------------------------------------------------------------SAHETVPKKYVTHS--------------------
BÊTA         ---------------------------------------------------------------------σ---φ-σ--δ-------------------
HHV-8         ---------------------------------------------------------NLGKCSTESCQSEARRQLVGGG---------------
HHV-4_1       ---------------------------------------------------------GEEEASRESPEVPRPAGAREPG---------------
HHV-4_2       ---------------------------------------------------------GEEEASRESPEVPRPAGAREPG---------------
GAMMA        -------------------------------------------------------------σ--δτSδESσδσ-----α--σG---------------
ET           ---------------------------------------------------------------------------------------------
PSI-PRED     --------------------------------------cccccccchhhcccccccccccccccchhhhhccccchhhhcccccchhcccccccccccc
```

```
                480       490       500       510       520       530       540       550       560       570
                 |         |         |         |         |         |         |         |         |         |
HSV-1        GPE-----------------------TWGDIATQAAADVRERRRLYADRLTKRSLASLGRCVREQRGELEKMLRVSVHGEVLPATFAAVANGFAARA
HSV-2        GPE-----------------------SWGDIATRAAADVRERRRLYADRLTKRSLASLGRCVREQRGELEKMLRVSVHGEVLPATFAAVANGFAARA
CeHV-1       GPE-----------------------SWADVAGRAAADVRERRRLYADRLTKRSLASLGRCVREQRGELEKMLRVSVHGEVLPATFAAVSNGFAARV
CeHV-2       GPE-----------------------SWADVAGRAAADVRERRRLYADRLTKRSLASLGRCVREQRGELEKMLRVSVHGEVLPATFAAVSNGFAARV
EHV-1        EER-----------------------PWDKLLERATADASQRRRMYAERLSKRSIASLGRCVREQRRELEKTLRVNVYGEVLLHTYVSSYNGFCARR
MDV-2        GHD-----------------------EWLEMVKSASSDVARRRKMYAERLTKKSLASLDKCITEQRHELEKMLRVNVYGEVLIDSYTALFNGFRSRK
GaHV3        EHY-----------------------DWLELVRSASADVARRRKMYAERLSKRSLASLDKCITEQRKELEKMLRVNVYGEILIDSYTTLFNGFRSRN
MEHV-1       GDV-----------------------EWLELVRAASADAARRRKMYAERLTKRSLASLDKCITEQRKELEKILRVNVYGEVLIDSYVAIFNGFRSRR
HHV-3        -VH-----------------------GWENLLSKAKIDAAERKRVYLEHLSKRSLTSLGRCIREQRQELEKTLRVNVYGEALLQTFVSMQNGFGARN
SUID         DAV-----------------------AWADLPAAALRDAERRRRLYADRLSRRSAASLAQCVREQRRELEKTLRVNVYGDALLHTYVAVAAGFRARR
GaHV1        DED-----------------------SIEVLLKKAERDLDVRQKNYVEKLSARSFSNLDRCIKIQRAELEKLMRVNIYGETLPVMYVTLKNAFLARR
GaHV2        DED-----------------------SIEVLLKKAERDLDVRQKNYVEKLSARSFSNLDRCIKIQRAELEKLMRVNIYGETLPVMYVTLKNAFLARR
ALPHA        ---------------------------φ--φα--A--Dα--Rδπ-YαvπLσ-πSφσσLσδCαδ-QR-ELEKφφRVδαρGvφLα--ρφσ--στF-τR-
AD169        NGQDGDGSENGLRVRNCDEHEALDLVDARSRIHNVTREVNVRKRAYLQKVSEVGYGKVIRCIKTQERLTSKLIDVNLVGPLCLDFISKLMNGFLYRS
CCMV         -------ATLSQAPRGEEEHEALDPLDARSRVHNVTREVNVRKRAYLQKVSEVGYGKVIRCIKAQERLTTKLIDVNLVGPVCLDFLSKMMNGFIYRS
RhCMV        GDE----GAGTTTA----MSSALECGDPLLRVHDVNKEVNVRKRAYLKKVSEMGYNKVMACIRNQEHLVTKLVNVNLVGTVCLEAVSKIMNGFLSRQ
MCMV         GDE----GAGTTTA----MSSALECGDPLLRVHDVNKEVNVRKRAYLKKVSEMGYNKVMACIRNQEHLVTKLVNVNLVGTVCLEAVSKIMNGFLSRQ
RCMV         GDPPEAAAASSTTAPARGMNFALECGDPLLRTHDVNKEVNVRKRAYLKKVSEVGYNKVMACIKSQEHLVTKLVNVNLVGTVCLEALSKVMNGFLSRQ
Tupaii       GGP-------------------FQCGDPAARAHDVVREVHVRKKAYLQKVSELGYNRVMQCIKGQEKLIKKLVNVNLLGTVCFEVLAKVINGFIRRR
HHV-7        ----------------------------MTQDHNLQQEIHIRKKAYYQKISESGYSKVMLCIKEQEALINKLMNINILGNHIFESLSKMMNAFANRQ
BÊTA         ------------------------------δ-Hδα-δEαδαRKπAYφδKαSE-GYσπVφ-CIπ-QE-LφδKLφδαNααG-φφφν-ατKφφNτFαδRδ
HHV-4_1      -------------------------------PSGALSDALKRKEQYLRQVATEGLAKLQSCLAQQSETLTETLCLRVWGDVVYWELARMRNHFLYRR
HHV-4_2      -------------------------------PSGALSDALKRKEQYLRQVATEGLAKLQSCLAQQSETLTETLCLRVWGDVVYWELARMRNHFLYRR
HHV-8        ------------------------------KPEEVLRDAKHRQELYLQKVARDGFKKLSDCIRHQGHILSQTLGLRLWGSVIYNEASALQNHFLHRA
GAMMA        ------------------------------Pδ-αLδDAφπRδE-YLδδVA-νGφφKLδδCα-δQτδφLδδTLτLRαWGδVαYδEατ-φδNHFLρR-
ET           -----------------------------------α--να--Rδδ-Yφδδασ--τφxxαx-Cα--Qx-xφδδφφxαδαφG-φφφ--φx---σxF-xR-
PSI-PRED     cccccccccccccccccccchhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhh
```

```
                          580       590       600       610       620       630       640       650
                           |         |         |         |         |         |         |         |
HSV-1        RFCALT-AG----AGTVIDNRSAPGVFDAHRFMRASLLRHQVDPALLPSITHRFFELVNGPLFDHSTHSFAQPPNTALYYSVENVGLLPHLKEELA
HSV-2        RFCALT-AG----AGTVIDNRAAPGVFDAHRFMRASLLRHQVDPALLPSITHRFFELVNGPLFDHSTHSFAQPPNTALYYSVENVGLLPHLKEELA
CeHV-1       RFCDLT-AA----AGTVVDNRAAPGVFDAHRFMRASLLRHHVDPALLPSITHRFFELVNGPLFDHATHSLAQPPNTALYYSVENVGLLPHLKEELA
CeHV-2       RFRDLT-AA----AGTVVDNRAAPGVFDAHRFMRASLLRHHVDPALLPSITHRFFELVNGPLFDHATHSLAQPPNTALYYSVENVGLLPHLKEELA
EHV-1        GFCAAV-SR----AGTIIDNRSSTSAFDSHQFMKAALLRHPIDQSLMPSITHKFFELINGPVFDNAGHNFAQPPNTALYYSVENVGLLPHLKEELA
MDV-2        RLLEAVKNC----CANIIDNRNSDDAFDAHRFMQTSLLKHRIDPAMLPSLTHKFFQLVNGPMFSHDRHRFAQPSNTALYFSVENVGLLPHLKEEMA
GaHV3        RLLEAIRKR----CTNIIDNRITADAFDAHRFMQTSLLKHRIDPAMLPSLTHKFFRLLNGPMFNHDVHRFAQPSNTALYFSVENVGLLPHLKEEIA
MEHV-1       GLLEAVQTR----FANVIDNRRGDEAFDAYRFMQTALLKHKIDPAMLPSLTHKFFQLVNGPLFNHDVHRFAQPSNTALYFSVENVGLLPHLKEEMA
HHV-3        VFLAKV-SQ----AGCIIDNRIQEAAFDAHRFIRNTLVRHTVDAAMLPALTHKFFELVNGPLFNHDEHRFAQPPNTALFFTVENVGLFPHLKEELA
SUID         AFCEAA-AR----AGTVVDER-ETGCFDAHSFMKATVQRHPVDAALLPAVTRKFFELVNGPLFAHDTHAFAQPPNTALYFAVENVGLLPHLKEELA
GaHV1        AFLDIVCAEESLHSKILRIDQNDGRAYDWHQYIRSSITRYQIDQTMLPRLTNRFFEILSGPLFKHHKERFPQPPNTSIYFSVENVGLLPHLKEELA
GaHV2        AFLDIVCAEESLHSKILRIDQNDGRAYDWHQYIRSSITRYQIDQTMLPRLTNRFFEILSGPLFKHHKERFPQPPNTSIYFSVENVGLLPHLKEELA
ALPHA        -φ--φφ----------α--δδ----σρD-ρδρφδσσα-πρ-αD-σφφP-αTδπFFδαασGPφF-δ--δ-φσQPσNTταρρσVENVGLLPHLKEEφA
AD169        QYHQ---------DQD-VVDVGDQFTYDEHLYVVNNLIHKSLPVESLPLLGQQIYELCNGPLFTHCTDRYPLSHNVDMAYACDNAGVLPHVKDDLV
CCMV         RYGQ---------DRE-MLDVGQSLTYDEHLYVVNNLIHKSLPVESLPLLGQQIYELCNGPLFTHCTDRYPLSHNVDMAYACDNAGVLPHAKDDLV
RhCMV        RSIT---------EAETYPDVAESLGYDEHLYVINNLVHKRLPSELLPQLGQQIYRFINGPMFTHYLDRHPLPYNVNMAYACDNAGILPHVKEDLV
MCMV         RSIT---------EAETYPDVAESLGYDEHLYVINNLVHKRLPSELLPQLGQQIYRFINGPMFTHYLDRHPLPYNVNMAYACDNAGILPHVKEDLV
RCMV         RATK---------EARALPDVGESLAYDEHLYVVNNLVHKRLPVELLPHLGQQVYRFINGPLFTRYLDRYPLPYNVNMAYACDNGGILPHVKEDLV
Tupaii       AYLE---------RVDGVVDVDRLLQYDDHLYVVNNLVHRRLPAESLPALGQEFYRFVNGPVFQHHADRYPLPYNIDMAYACDNAGMLPHLKEDLV
HHV-7        LQSL---------ENFS----ADPFTYDDHLYIKNNLLSKKLPQELLPNLSQEMYRLLTGPLSNYHTASFPLSSNISMAYACDVADFLPHMKEDLA
BÊTA         -------------δ-------σ--φ-YDνHLYαφNNLαδπδLP-E-LP-LτQδφYδφφσGPφ-δδδφσδρPLσδNασMAYACDστσφLPHφKνDLσ
HHV-4_1      AFVS---------GPWEDRRAGEGAAFENSKYIKTHLFTQTLSSEHLHALTHSLYTFITGPLAEESG-LFPPPSNVALARCCDAAGTLPHQKAFLT
HHV-4_2      AFVS---------GPWEDRRAGEGAAFENSKYIKTHLFTQTLSSEHLHALTHSLYTFITGPLAEESG-LFPPPSNVALARCCDAAGTLPHQKAFLT
HHV-8        QFIS---------LPWQDLTVDCPTRFENSKYIKNSLYCQRLGREHVEILTLEFYKLITGPLSKRHT-LFPSPPNVTLAQCFEAAGMLPHQKMMVS
GAMMA        -FvS----------PWδD-δασδσφ-FENSKYIKσδLρδQδLτ-EHαδαLTφδφYδφITGPLτδδδσ-LFPσPσNVσLAδCφνAAGφLPHQKφφασ
ET           --------------------------ρν-xxρφxσxα-δδ-ασ-x-φx-ασxδφρδφφσGPφ--δ----φσxσxNφσφφxσφνσσσφLPHxKxxφσ
PSI-PRED     hhhh---------cce-eeccccchhhhhhhhhhhhhhcccccchhhhhhhhhhhhhhcccccccccccccccccchhhhhhhhhcccchhhhhhh
```

```
            660       670       680       690       700       710       720       730       740       750
             |         |         |         |         |         |         |         |         |         |
HSV-1       RFIMGAGGSGADWAVSEFQRFYCFDGISGITPTQRAAWRYIRELIIATTLFASVYRCGELELRRPDCSRPTSEGRYRYP----------PGVYLTY
HSV-2       RFIMGAGGSGADWAVSEFQKFYCFDGVSGITPTQRAAWRYIRELIIATTLFASVYRCGELELRRPDCSRPTSEGLYRYP----------PGVYLTY
CeHV-1      RFIMGAGGAGTEWAVSEFQKFYSFDGISGITPTQRAAWRYIRELIIATTLFASVYRCGEVELRRPDYGRPTPEGLYRYP----------PGVYLTY
CeHV-2      RFIMGAGGAGTEWAVSEFQKFYSFDGISGITPTQRAAWRYIRELIIATTLFASVYRCGEVELRRPDYGRPTPEGLYRYP----------PGVYLTY
EHV-1       RFMITA--AKGDWSISEFQRFYCFEGVTGVTATQRLAWKYIGELILAAAVFSSVFHCGEVRLLRADRTYPDSSGAQRCV----------SGIYITY
MDV-2       RFMFHSS-RKTDWTVSKFRGFYDFSTIDNVTAAHRMAWKYIKELIFATALFSSVFKCGELHICRADSLQINSNGDYVWK----------NGIYITY
GaHV3       RFMIHSC-SSSDWIVSKFRGFYDFDGIDSITAAHRMAWKYIKEVILATTLFSSAFSCGELHLRRADCLPLNVNVEYIWE----------DGIYVTY
MEHV-1      RFMLQCN-NTSSWIISKFRGFYDFGGVDNITTAHRMAWKYIRELILATALFASVFKCGDLQVCRADGLQLTLGGEYIWE----------DGVYLTY
HHV-3       KFMGGV--VGSNWLLSPFRGFYCFSGVEGVTFAQRLAWKYIRELVFATTLFTSVFHCGEVRLCRVDRLGKDPRGCTSQPKGIGSSHGPLDGIYLTY
SUID        RFMV-----ARDWCVSEFRGFYRFQ-TAGVTATQRQAWRYIRELVLAVAVFRSVFHCGDVEVLRADRFAGR------------------DGLYLTY
GaHV1       NFTKRH--RSCDWMTSEFRRFYNFAGISNSTAIQRAACGYIREALFATTAFEKIFHSGNTKLLRADCVEFDLDGPILE-----------NGLYLTF
GaHV2       NFTKRH—-RSCDWMTSEFRRFYNFAGISNSTAIQRAACGYIREALFATTAFEKIFHSGNTKLLRADCVEFDLDGPILE-----------NGLYLTF
ALPHA       δFφ------σ-δW-φS-Fδ-FYδF--φ-σσTφφδR-Aφ-YI-EαaφAφφαF-δαρδδGδφδα-RαD----δ------------------σGαYαTρ
AD169       KCAEGTV-YPSEWMVVKYMGFFNFSDCQDLNVLQKEMWMHVRELVLSVALYNETFGKQLSIACLRDELHPDRD-VILTYNKEWPLLLRHEGSLYKS
CCMV        KCAEGTV-YPSEWMVVKYMSFFNFSECQDLNVLQKEMWMHVRELVLSVALYNETFGKQLSIACLRDELNPDRD-LILTYNKEWPLLLRHEGSLYKS
RhCMV       RCADGTV-VPSDWMTVGYMGFFRFADIRELNDLQKMVWAHIRELVLSVALYNETFGKQLALWRVEDG-DEIGDGIILTYNPESPLILRRGDRSYRS
MCMV        RCADGTV-VPSDWMTVGYMGFFRFADIRELNDLQKMVWAHIRELVLSVALYNETFGKQLALWRVEDG-DEIGDGIILTYNPESPLILRRGDRSYRS
RCMV        RCADGTA-EPGDWMVAGYQGFFSFVDVRDVNDLQKAVWAHVRELVLSVALYNETFGKQLDVWRVEDG-DAAGDGIVLTYNREAPLLLRYAGRRFRS
Tupaii      RLAEGTV-APSEWMVAPYRRFFDLGAALDLNELQKGFWAHVREIVFSVALYNEAFGKELRLCRADEPVDEERERLVVTYNVDGPLFLYAGGGVWKS
HHV-7       KCVEGTI-YPENWMLCTYNKFFNFDGLHNINDMQRQMWNFIRELVLSVALYNDVFGKQLSIVKFGEE-TETVEKILLTFDSGSPLLFKRGTTTTKF
BÊTA        πφαvGTα--P-δWMφφ-Y--FFδφσ-φ-δαN-φQπ-φW-ραRELVφSVALYNvφFGKδL—αφδφ-v--δ---v-αaαTρδ---PLφφδ--σ--φπ-
HHV-4_1     SLIWPGI-EPSDWIETSFNSFYSVPG-GSLASSQQILCRALREAVLTVSLYNKTWGRSLILRRADAVSPGQALPPD--------------GLYLTY
HHV-4_2     SLIWPGI-EPSDWIETSFNSFYSVPG-GSLASSQQILCRALREAVLTVSLYNKTWGRSLILRRADAVSPGQALPPD--------------GLYLTY
HHV-8       EMIWPSI-EPKDWIEPNFNQFYSFEN-QDINHLQKRAWEYIRELVLSVSLYNRTWERELKILLTPQGSPGFEEPKPAGLT---------TGLYLTF
GAMMA       δφIWPσI-EPδDWIEφσFNδFYSφ-σ--σασδ-Qδ-αφδφαREαVLσVSLYNπTW-RδL-α--φσ-σSPG---Pφσ--------------GLYLTρ
ET          δφφ------σ-δW-xx-ρ--Fρδφ----xxσ--δδ-φφ-φα-Eαaφσφxαρ-δφρ-δxx-x--x---------------------------σ--φδ-
PSI-PRED    hhhhccc-ccccceeeccccccccccccchhhhhhhhhhhhhhhhhhhhhhhhhhhcccceeeeccccccccccc-eeeeeccccceeeeeccccccc
```

```
                                                  760         770         780        790        800        810
                                                   |           |           |          |          |          |
HSV-1        DSDCPLVAIVESAPDGCIG----PRSVVVYDRDVFSILYSVLQHLAPRLPDGGHDGPP-------------------------------------
HSV-2        NSDCPLVAIVESGPDGCIG----PRSVVVYDRDVFSILYSVLQHLAPRLAGGGSDAPP-------------------------------------
CeHV-1       DSDCPLVAIVECEPDGGIG----PRSVVVYDRDVFSILYSVLQHLAPRLAAGGPDQPPP-------------------------------------
CeHV-2       DSDCPLVAIVECEPDGGIG----PRSVVVYDRDVFSILYSVLQHLAPKLAAGGPESTPA-------------------------------------
EHV-1        EASCPLVAVLSAAPHGAIG----AETVVIYDSDVFSLLYAVLQQLAPGSGAN-------------------------------------------
MDV-2        ETEYPLIMILGSESSTSET----QNMTAIIDTDVFSLLYSILQYMAPVTADQVRVEQITNSHAPI------------------------------
GaHV3        ETECPLITVLGAESD---------NVTVILDADVFSLLYSILQYMAPAAAIRYVRNEQTMTTPPLDQCDIFRTAARLSTLSMRLVKINP-------
MEHV-1       ETDCPLVALIGCRAYLTHN----QSPTVLVDTDVFSLLYSILQFMAPATADQLRSDRFSNSHTHN------------------------------
HHV-3        EETCPLVAIIQSGETG-ID----QNTVVIYDSDVFSLLYTLMQRLAPDSTDPAFS----------------------------------------
SUID         EASCPLVAVFGAGPGG-IG----PGTTAVLASDVFGLLHTTLQLRGAPSR---------------------------------------------
GaHV1        EEEHPLVAVWGVDDRGKLG----PASTIIVEKDLYAVLYAILHRREH-----------------------------------------------
GaHV2        EEEHPLVAVWGVDDRGKLG----PASTIIVEKDLYAVLYAILHRREH-----------------------------------------------
ALPHA        δ-δφPLαφαφ-σ--------------σααy--DαρταLyσφφδ-------------------------------------------------
AD169        ------------------------------KDLYLLLYRHLSRPDE--SGDVPTAPVA--KPSTLTAAAAVSGVFREPDRLWLPSPYPSSSTAGV
CCMV         ------------------------------KDLYLLLYRHLSRPDE--SGDAPTAGAVFREPNPQQQAWAPPPVPPGALLSSASSSSTSSLCSTA
RhCMV        ------------------------------RDLYLLLYKHLSVDSETLAD---AGSRASVADLCQVERPGP-----IAEQRS-STQNVKKKRKRM
MCMV         ------------------------------RDLYLLLYKHLSVDSETLAD---AGSRASVADLCQVERPGP-----IAEQRS-STQNVKKKRKRM
RCMV         ------------------------------KDLYLLLYKHLALCSDEASGHGVGGPPGSVRLVAPSPRAGPPPADATANRRDPSPSGGRKRRKKT
Tupaii       ------------------------------KDLYLLLYQHLN------------GAAAPPPPAAPSPPPAEPATTAGASRKRPAVESPGV-----
HHV-7         -------------------------------NDLYSLLYFDLKTQCD-------PVQISQTKQVSHIPAPNLLDLCRQNENSIPECFYNF------
BÊTA         ------------------------------δDLY-LLY—δL---σν--------σ----------------α------------σ----------
HHV-4_1       DSDRPLILLY--------------KGRGWVFKDLYALLYLHLQMRDDSA-----------------------------------------------
HHV-4_2       DSDRPLILLY--------------KGRGWVFKDLYALLYLHLQMRDDSA-----------------------------------------------
HHV-8         ETSAPLVLVD--------------KKYGWIFKDLYALLYHHLQLSNHNDSQV-------------------------------------------
GAMMA        νσδ-PLαLαδ--------------K-rGWαFKDLYALLYφHLQφδσδσσ-----------------------------------------------
ET           -------------------------------Dαρ-αLρ-xL-------------------------------------------------------
PSI-PRED     ------------------------------hhhhhhhhhhhhcccc—cccccccccc--cccccccccccccccccccccccccccccccccccc
```

```
                 820        830        840        850
                  |          |          |          |
HSV-1          ------------------------------------
HSV-2          ------------------------------------
CeHV-1         ------------------------------------
CeHV-2         ------------------------------------
EHV-1          ------------------------------------
MDV-2          ------------------------------------
GaHV3          ------------------------------------
MEHV-1         ------------------------------------
HHV-3          ------------------------------------
SUID           ------------------------------------
GaHV1          ------------------------------------
GaHV2          ------------------------------------

AD169          SRRVRATRKRPRRASSLLDLARDEHGIQDLVPGSLR--
CCMV           RRSRVISSRKRTRNAASLLDLARIDHVGQDLVPGSLR-
RhCMV          SLLELVRDVDGAGGDDLVPPCLYK--------------
MCMV           SLLELVRDVDGAGGDDLVPPCLYK--------------
RCMV           SLVDLVRDADRLADGEVVPDSMRVHPEPAVFRPGGGDR
Tupaii         -LLDLVRDADR--ESSLVPDCLLYDP------------
HHV-7           -------------------------------------
HHV-8           -------------------------------------

HHV-4_1         -------------------------------------
HHV-4_2         -------------------------------------
PSI-PRED       cccccccccccccchhhhhhhhhhhcccccccccccccc-
```

**Alignment data :**

Alignment length: 983
Residues conserved for 90 % or more (upper-case letters): 44 is 4.48 %
Residues conserved for 50 % and less than 90 % (lower-case letters): 191 is 19.43 %
Residues conserved less than 50 % (white space): 724 is 73.65 %
HSV-1: CAA32321 ( 785 residues) | AD169: X17403 ( 850 residues) | HHV-4_1: YP_401715( 685 residues)
HSV-2: CAB06753 ( 785 residues) | CCMV: AAM00906 ( 839 residues) | HHV-4_2: YP001129509( 685 residues)
CeHV-1: BAC58068 ( 785 residues) | RhCMV: CAA47415 ( 798 residues) |
CeHV-2: AAU84532 (784 residues) | MCMV: CAA47415 ( 798 residues).|
EHV-1: YP_053077 (775 residues) | RCMV: AAC56431 ( 893 residues). |
MDV-2: BAA82984 (793 residues) | Tupaii: AAK57105 (782 residues). |
GaHV-3: BAB16538 ( 811 residues) | HHV-8: AAC57084 (695 residues). |
MeHV-1: AAG30068 (787 residues) |
HHV-3: CAA27913(770 residues) |
SUID: YP_068331 (724 residues) |
GaHV-1: YP_182357 (764 residues) |
GaHV-2: AAF66763 (764 residues) |

## B.3. ANNEXE B3: Alignement de séquence de pUL89 et gp17 (bactériophage T4).

```
                  10        20        30        40        50        60        70        80        90        100
                   |         |         |         |         |         |         |         |         |         |
PUL89        MLRGDSAAKIQERYAELQKRKSHPTSCISTAFTNVATLCRKRYQMMHPELGLAHSCNEAFLPLMAFCGRHRDYNSPEESQRELLFHERLKSALDKLTFRP
gp17         ---------MEQPINVLN--DFHPLNEAGKILIKHPSLAERKDEDGIHWIKSQWDGKWYPEKFSDYLRLHKIVKIPNNSDKPELFQTYKD----------
                      :::    *:  . ** .  ..: : .:*..:::    :    .:    : :   *: : *::*::  **:   .
                     110       120       130       140       150       160       170       180       190       200
                      |         |         |         |         |         |         |         |         |         |
PUL89        CSEEQRASYQKLDALTELYRDPQFQQINNFMTDFKKWLDGGFSTAVEGDAKAIRLEPFQKNLLIHVIFFIAVTKIPVLANRVLQYLIHAFQIDFLSQTSI
gp17         -KNNKRSRYMGLPNLKRAN--IKTQWTREMVEEWKKCRDD------------------------IVYFAETYCAITHIDYGVIKVQLRDYQRDMLKIMSS
              .:::*: *  *  *..    : *  .::: ::**  *.                         * * :   *.:   *::  :: :* *:*.  *
                     210       220       230       240       250           260       270       280       290
                      |         |         |         |         |             |         |         |         |
PUL89        DIFKQKATVFLVPRRHGKTWFIIPIISFLLKHMIGISIGYVAHQKHVSQFVLK--EVEFRCRHTFARDYVVENKDNVISIDHRGAKSTALFASCYNTNSI
gp17         ----KRMTVCNLSRQLGKTTVVAIFLAHFVCFNKDKAVGILAHKGSMSAEVLDRTKQAIELLPDFLQPGIVEWNKGSIELDN----GSSIGAYASSPDAV
                 :: **  :.*: *** .:  :::.:: .  . ::* :**:  :*  **. :  :.    *:  :** :.. *.:*:    .::: * . ..:::
                300      310         320       330       340       350      360       370       380       390
                 |        |           |         |         |         |        |         |         |         |
PUL89        RGQNFHLLLVDEAHFIK--KEAFNTILGFLAQNTTKIIFISSTNTTSDSTCFLTRLNNAPFDMLNVVSYVCEEHLHSFTEKGDATACPCYRLHKPTFISL
gp17         RGNSFAMIYIDECAFIPNFHDSWLAIQPVISSGRRSKIIITTTPNGLNHFYDIWTAAVEGKSGFEPYTAIWNSVKERLYNDEDIFDDGWQWSIQTINGSS
             **:.* :: :**. **   :::: :*  .::..  . *:*::* . :   :        . :: : : :.  .: :. *          :.   *
                400       410       420       430       440       450       460       470       480       490
                 |         |         |         |         |         |         |         |         |         |
PUL89        NSQVRKTANMFMPGAFMDEIIGGTNKISQNTVLITDQSREEFDILRYSTLNTNAYDYFGKTLYVYLDPAFTTNRKASGTGVAAVGAYRHQFLIYGLEHFF
gp17         LAQFRQEHTAAFEGTSGTLISG--MKLAVMDFIEVTPDDHGFHQFKKPEPDRKYIATLDCSEGRGQDYHALHIIDVTDDVWEQVGVLHSN----TISHLI
              :*.*:  . : *:    * *   *::   .: .  . . *. :: . : :    :.:     *       ..:.     **.: :     :.*::
```

```
                          510       520       530       540       550       560       570       580       590
                           |         |         |         |         |         |         |         |         |
PUL89        LRDLSESSEVAIAECAAHMIISVLSLHPYLDELRIAVEGNTNQAAAVRIACLIRQSVQSSTLIRVLFYHTPDQNHIEQPFYLMGRDKALAVEQFISRFNS
gp17         LPDIVMRYLVEYNECPVYIELNSTGVS-VAKSLYMDLEYEG-----VICDSYTDLGMKQTKRTKAVGCSTLKDLIEKDKLIIHHR---ATIQEFRTFSEK
             * *:     *   **..:: :.  .:    ..*: :*:      *   .    .::.:.  :.:   * .:   ::: :  *    ::::*:  :.

            600       610       620       630           640       650       660       670
             |         |         |         |             |         |         |         |
PUL89        GYIKASQELVSYTIKLSHDPIEYLLEQIQNLHRVTLAEG-TTARYSAKRQNRISDDLIIAVIMATYLCDDIHAIRFRVS--
gp17         GVSWAAEEGYHDDLVMSLVIFGWLSTQSKFIDYADKDDMRLASEVFSKELQDMSDDYAPVIFVDSVHSAEYVPVSHGMSMV
               *    *::*     : :*  : :*  *: :. .  :   ::.  :*.: :***   .
```

Alignment length: 681

Strongly similar (:): 155 is 22.76% |Different: 360 is 52.86% |gp17(610 residues).
Identity (*): 89 is 13.07% |Weakly similar (.): 77 is 11.31% |AD169(674 residues)

## B.4. ANNEXE B4: Alignement de séquence de pUL89(161-570) et du domaine hélicase de la famille des snf2/rad54 de *Sulfolobus solfataricus* (PDB 1Z63)

```
                    170       180       190       200       210       220       230       240       250
                     |         |         |         |         |         |         |         |         |
PUL89       NLLIHVIFFIAVTKIPVLANRVLQYLIHAFQIDFLSQTSIDIFKQKATVFLVPRRHGKTWFIIPIISFLLKHMIGISIGYVAHQKHVSQF
1z63xx0     ---ELLDRLRGNKSFQLLEPYNIKANLRPYQIKGFSWMRFMNKLGFGICLADDMGLGKTLQTIAVFSDAKKENELTPSLVICPLSVLKNW
                : : . ..: :*    ::  ::.:**. :*  :      . :      ***   *.::*   *.    .   :.  . :.::
                    260       270       280       290       300       310       320       330       340
                     |         |         |         |         |         |         |         |         |
PUL89       VLKEVEFRCRHTFARDYVVENKDNVISIDHRGAKSTALFASCYNTNSIRGQNFHLLLVDEAHFIKKEAFNTILGFLAQNTTKIIFISSTN
1z63        EEELSKFAPHLRFA---VFHEDRSKIKLEDYDIILTTYAVLLRDTR-LKEVEWKYIVIDEAQNIKNPQTKIFKAVKELKSKYRIALTGTP
             :  :* :  **   *..:. . *.::..    *:  .   :*. ::  ::: :::***: **:  : : ..   ::.  * ::.*
                    350       360       370       380       390       400       410       420        430
                     |         |         |         |         |         |         |         |          |
PUL89       TTSDSTCFLTRLNNAPFDMLNVVSYVCEEHLHSFTEKGDATACPCYRLHKPTFISLNSQVRKTANMFMPGAFMDEIIGGTNKISQNTVLI
1z63        IENKVDDLWSIMTFLNPGLLGSYS-EFKSKFATPIKKGDNMAKEELKAIISPFILRRTKYDKAIINDLPDKIETNVY--CNLTPEQAAMY
              ..  : : :.    .:*.  *   :.:::  :***  *   :   ..**  .::  *:    :*.:  ::    *  .::::.:
                    440       450       460       470       480       490       500       510       520
                     |         |         |         |         |         |         |         |         |
PUL89       TDQSREEFDILRYSTLNTNAYDYFGKTLYVYLDPAFTTNRKASGTGVAAVGAYRHQFLIYGLEHFFLRDLSESSEVAIAECAAHMIISVL
1z63        KAEVENLFNNIDSVTG----IKRKGMILSTLLKLKQIVDHPALLKGGEQS--VRRSGKMIRTMEIIEEALDEGDKIAIFTQFVDMG----
            .: .: *::   *      .  *  * . *.    .:: *  .*       *:. :    .:: . *.*..::**    ..*
                   530       540       550       560       570
                    |         |         |         |         |
PUL89       SLHPYLDELRIAVEGNTNQAAAVRIACLIRQSVQSSTLIRVLFYHTPDQN
1z63        ------KIIRNIIEKELNTEVPFLYGELSKKERDDRECSHAVILFDII--
                  . :*  :*: *  ...  . * ::. :.    :.:: .
```

**Alignment data :** Alignment length: 410

Identity (*): 52 is 12.68 % | Weakly similar (.): 61 is 14.88 % | Sequence 0001: 1z63 ( 382 residues).Strongly similar (:): 84 is 20.49 % |

Different: 213 is 51.95 % | Sequence 0002: PUL89 ( 410 residues).

## B.5. ANNEXE B5: Sous domaines conservés à travers quelques phosphotransférases herpesvirales comparés avec celles de Cdk2 et PKA.

| | I | II | III | IV-V | VI | VII | VIII | IX |
|---|---|---|---|---|---|---|---|---|
| **UL13_HSV-1** | GGSGGYGEVQ | AVK | KEW | FTELARAVVFLNTTCGISHL | DIKCAN | DFSL | TAL | DLYALG |
| **UL13_HSV-2** | GGSGGYGEVQ | AVK | KEW | FTDLARAVVFLNTRCGISHL | DIKCAN | DFSL | AAL | DLYALG |
| **ORF47_VZV** | AGRGTYGRVH | AVK | MDS | FLDLAQALTFLNRTCGLTHL | DVKCGN | DYSL | RDA | DLYALG |
| **BGLF4_EBV** | LGRGSYGAVY | TVK | TEL | FQGLKDAVYFLNRHCGLFHS | DISPSN | DYGT | FSI | DLQSLG |
| **UL97_HCMV** | LGQGSFGEVW | VVK | SET | FCTLADAIKFLNHQCRVCHF | DITPMN | DYSL | REC | ELSALG |
| **pR97_RCMV** | IGKGSFGQVW | ALK | DEA | FTGLADGIRFLNLQCGVAHF | DVTPMN | DYSL | TDL | DLCALG |
| **M97_MCMV** | IGKGSFGQVW | ALK | NET | FSGIADALRFLNLRCGVGHF | DVMPMN | DFSL | TDI | DLCALG |
| **U69_HHV_6A** | LGVGAYGKVF | AIK | DES | FCKLADAVRFLNLKCRINHF | DISPMN | DYSL | CDM | DLCAFA |
| **U69_HHV_6B** | LGVGAYGKVF | AIK | DES | FCKLADAVRFLNLKCRINHF | DISPMN | DYSL | CDM | DLCAFA |
| **Cdk2** | IGEGTYGVVY | ALK | REI | YLFQLLQGLAFCHSHRVLHR | DLKPQN | DFGL | PEI | DIWSLG |
| **PKA** | LGTGSFGRVM | AMK | NEK | YAAQIVLTFEYLHSLDLIYR | DLKPEN | DFGF | PEI | DWWALG |

| | I | II | III | IV-V | VI | VII | VIII | IX | taille |
|---|---|---|---|---|---|---|---|---|---|
| **UL13_HSV-1** | 157-166 | 174-176 | 181-183 | 257-276 | 277-282 | 301-304 | 335-337 | 382-387 | 518AA |
| **UL13_HSV-2** | 157-166 | 174-176 | 181-183 | 257-276 | 277-282 | 301-304 | 335-337 | 382-387 | 518AA |
| **ORF47_VZV** | 138-147 | 155-157 | 159-161 | 237-256 | 257-262 | 282-285 | 315-317 | 362-367 | 510AA |
| **BGLF4_EBV** | 84-93 | 100-102 | 108-110 | 175-194 | 195-200 | 219-222 | 254-256 | 297-302 | 429AA |
| **UL97_HCMV** | 337-346 | 353-355 | 361-363 | 436-455 | 456-461 | 481-484 | 506-508 | 542-547 | 707AA |
| **pR97_RCMV** | 242-251 | 259-261 | 268-260 | 341-360 | 361-365 | 386-389 | 423-425 | 457-462 | 612AA |
| **M97_MCMV** | 271-280 | 288-290 | 298-300 | 371-390 | 391-395 | 416-419 | 452-454 | 488-493 | 643AA |
| **U69_HHV_6A** | 201-210 | 216-218 | 223-225 | 293-312 | 313-318 | 335-338 | 370-372 | 407-411 | 562AA |
| **U69_HHV_6B** | 202-211 | 217-219 | 224-226 | 294-313 | 314-319 | 336-339 | 371-373 | 408-412 | 563AA |
| **Cdk2** | 10-19 | 31-33 | 50-52 | 107-126 | 127-132 | 145-148 | 171-173 | 185-190 | 298AA |
| **PKA** | 50-59 | 71-73 | 91-93 | 147-166 | 167-172 | 185-188 | 208-210 | 221-226 | 351AA |

## B.6. ANNEXE B6: Alignement de séquences entre pUL27 et ses homologues issus de bêta herpesvirus

```
                                                                  10         20          30        40
                                                                   |         |           |         |
HCMV_AD169     ------------------------------------------------MNPVDQPPPPLPTQQPEEQAKEDHD--DGDERLFRDPLTTYE
PoHV_4         ------------------------------------------------MTELQQSQVQTPEPQ-QEHAKDEHGELDDDERVYRDPLTTYE
rh46_RhCMV     ---------------------------------------------------MTANPP--------------------LDDIYSDQHAAFE
Cytomegalo     ------------------------------------------------------------------------------νδαρδD-φσσρE
pR27_MuHV2     ------------------------------------------------------------------------------------------
gPM27_MuHV1    ---------------------------------------------------MADRGPCAEGSGGTVRRRSSPS----ESSEVDDTQFVFE
Muromegalo     ------------------------------------------------------------------------------------------
Gp13_HHV_6B    EQEKMWFGETKCGREEFILLQRNLPAMKDYVAKHCGKRIRVDAFQDFDFSFCSLSDIYYLTGPGILEKITEKDYAIIGTCARSQAEPNCRAA
U5_HHV-7        ELDRLWYGETKCSGEEFVLLQKNIPALKNFVNRQCGQKIRIDAFQNFDLSFCSSNDIHYITGSGILEKILRRKYVVIGTCARCQVEPNCRAV
U5_HHV_6A      ------------------------------------------------------------------------------------------
Roseolo        ------------------------------------------------------------------------------------------
T27_TuHV-1     ------------------------------------------------------------------------------METAPLHARRRVVYE
ET             ------------------------------------------------------------------------------------------
               ------------------------------------------------cccccccccccccccccchhhhhcccc--ccchhhhccccchhh
```

```
                        50        60        70                                80        90
                         |         |         |                                 |         |
HCMV_AD169     YLDDCR-DDEEFCHQFLRAYLTPIRNRQEAVRAGLLCR------------------------TPEDLAAAGGQKKKTPAPKHPK--------
PoHV_4         YLDDCRGDDDEFCRNFIRAYLTPIRNRQEAVRHGLLCR------------------------TPEDLATRHQKPRPAGSGKRSK--------
rh46_RhCMV     FVADCP--DDDFCHNFVRTYLTPVRSLREAALAGLFPVD----------------------SSSSSLTKKPRDTRVVQNAESAKK-------
Cytomegalo     ρασDC---DννFCπNFαRσYLTPαRσ-δEAσ-φGLφσ-------------------------σσ-σLσ----δ-πφσ-σσδδσK--------
pR27_MuHV2     -----------------------------------MDS----------------------WSKRNGKAPREITRFLSVDEPFD--------
gPM27_MuHV1    LVVDET--ADSFCRRFIADHLVQLRNYEQASRFGFVMDA----------------------WSRESGKAT-EIRSFIRLNKPFN--------
Muromegalo     -----------------------------------MDτ----------------------WSπδσGKAσ-EIδ-Fαδαδδ PFσ--------
Gp13_HHV_6B    IVMGSNCHIYIYVENRISKVSKSLRTFIRRGFDELLYKEKYSLDWNDDTLFYISDTETENLNRMLNGELPVLRSKPRHMCVRKD--------
U5_HHV-7       ILLGPNFHIYVYCDNKINKVARSIREFIRRGFEELLYKERYALNWHDDSLIYVSESEAENLNRMLNGESPILRKKPRHMYPRCDRYVKIKVL
U5_HHV_6A      ------------------------------------------------------------------------------------------
Roseolo        ---------------------------RRGFνELLYKEπYτLδWδDDσLφYαSνσEσENLNRMLNGE-PαLRδKPRHMφαRφD--------
T27_TuHV-1     LIRDDP--PTPFGEAFIGAHLQPVEVALDSAPQSLAELR----------------------LRRKRSAARDDGPGCSSARDVVYR-------
ET             ------------------------------------------------------------------------------φ---------------
PSI-PRED       hhcccc-cchhhhhhhhhhhcccccchhhhhhhccccc------------------------cccchhhhcccccccccccccc--------
```

```
                                                                                  100       110       120       130
                                                                                   |         |         |         |
HCMV-AD169     ------------------------------------------------------------HAMVYIRRSCLVHSACATAHGKYDIRGLTLE
PoHV_4         ------------------------------------------------------------HAMVYVRRSCLAHSAEVTAHGKYDIRGLTLE
rh46_RhCMV     ------------------------------------------------------------SPIIYIRRSFLVNSKFATAHGKYNIRGLSLQ
Cytomegalo     ------------------------------------------------------------δαφαYαRR-φLαδSφ-σTAHGKYσIRGLσLδ
pR27_MuHV2     ------------------------------------------------------------HVELKISYLKNARFATANGHYGIIGLALR
GPM27_MuHV1    ------------------------------------------------------------HIELKYSYLKNARLATANGGYGIRGLTLA
Muromegalo     ------------------------------------------------------------HαELKφSYLKNARφATANG-YGI-GLσL-
Gp13_HHV_6B    -------------------RLVKDRSKILFAVRLDEEDSLTVKFITKFLTPVFVGRLPATSRFVVPVSCARLTNGLQGTAAARFGIKGLHPS
U5_HHV-7       LFGILYTILLLMVPWMFFFRLLKNMPSILFAVHSSEISNPLVQSVTKFLHPIIIPNGDTELKYIVPVTESRLINGLQASAAGRFGIKGLRLC
U5_HHV_6A      ------------------------------------------------------------------------------------------
Roseolo        ------------------------------------------------------------σσδ-πραVPVσδτRLφNGLQτσAAτRFGIKGLπασ
T27_TuHV-1     ------------------------------------------------------------PIVVIQF SSYLQNAEFATAAGKYRIRGLLLP
ET             ------------------------------------------------------------------------L-δτ--σσA-τ-ρ-I-GL-α-
PSI-PRED       ------------------------------------------------------------cceeeehhhh???hhhhcccccccccccccc

                        140       150       160       170       180       190       200       210       220
                         |         |         |         |         |         |         |         |         |
HCMV_AD169     SDLAVWAALRGVPLPPDPQHFRWLNAGAFRRLVHEAQYLPEISRAAKRIALAVATGQYVVCTLLDYKTFGTRTHYLRQLCSMTEELYLRLDG
PoHV_4         WDLAVWAALRAIPTMPDPQHFRWNNAGAFRRLVHRARYLPEISRAAKRMALAVATGQYVVCSLLDYKSFGTRTNYLRQLCTLTEELYMRLDG
rh46_RhCMV     TDLAVWGLLRGVPLRPDPGHFHWMNLMQFRRLVAAGPFLPEISKSGKRIALAVATGQYAVSTLLNYKVFGNKTHYLRQLCVITEELFRRLAG
Cytomegalo     δDLAVWτφLRταPφ-PDP-HFπW-Nα--FRRLVφ-σ-ρLPEISπσσKRφALAVATGQYσVτσLLσYKσFGσπTδYLRQLCσφTEElρ-RLσG
pR27_MuHV2     WDTAVWGLLREIECVPYSNPFVFPNCARLSGIIAAMQFRPEANEVVRALCRACVAAQYGVSILFSYDHCAAVPRYLRSLVQELVELNDRIRH
gpM27_MuHV1    WDSAIWGLLREIEFVPFANPFSFPSCDGIAKILFDLEWRPEASDSCKALGRATVAIQYAISLLYGYDRSASMSRYLRRLVNELIDLHDRMRN
Muromegalo     WDσAαWGLLREIEφVPρτNPFσFPσCσ-ατ-Iαφσφδρ RPEAσνσσπALτRAφVAαQYταSαLρτYDπτAτφσRYLR-LVδELανLδDRφRδ
Gp13_HHV_6B    SDCVVWNRLIDYEYETYKYPSTYVRADQIADMVKNLKFMDNFDEKWQCITKLAFIGLYAGASLFNFASKPTLGYWCRYLCEYASMLLFQFES
U5_HHV-7       SDGVIWNRLIDYEYEMFKYPSTFTRADKFLLQLRDLKFMEDFDPKWQCITKLAAVGFYSGASLFNLGAKPGIGYWCRYLCEYLSMLFFKLDG
U5_HHV_6A      -------------------------------MVKDLKFMDNFNEKWQCITKLAFLGLYAGASLFNFTSKPTLGYWCRYLSEYASMLLFQFES
Roseolo        SDτVαWNRLIDYEYEφρKYPSTρφRADδφα--απσLKFMνδFσ-KWQCITKLAφαGφYτGASLFNφστKPσαGYWCRYLτEYαSMLφFδφντ
T27_TuHV-1     WDTAVWGLLRRSALPSRFNPLQRLRAADLCSYLNTVAFRPEASDVARHVARTVAVGQYAVQTVFDYRNPGGLSHYLRFLCESVHALYGRLER
ET             δD-ααWσ-L----φ-----φ-----φ--φ---α--φ-ρ--δφσ---δ-φσ-φφφ---Yσ---αφσφ---σσ-σ-ρφR-Lσ--φ--L--δφ--
```

```
PSI-PRED      hhhhhhhhhhcccccccccchhhhhhhhhhhhhhcccchhhhhhhhhhhhhcccc?eeeee?cccccccchhhhhhhhhhhhhhhhhhcc
                        230       240       250       260       270       280           290       300
                         |         |         |         |         |         |             |         |
HCMV_AD169    T---LCLFLEPEERELIGRCLPAALCRGLPVKYRTHRAAVFFHATFMARAEAALKDLYAAFCECGDG----RDNGGNHDGNHDGNDHSSL--
PoHV_4        T---MCLFLDPEEREVIARCLPQFLCRGLAVRYRSHRAAVFFNATFGARAETALKRLYSAFCECGDW----RDPDVPVEA-AEAVAAASQ--
rh46_RhCMV    T---VCLFLDAGDRDLIQKCLPPHLCKKLPVNYRVQKNAVFFNPRFALRAEAALHTLYARFCTCGDC----HDLFVASSD----TQATP---
Cytomegalo    T---φCLFLνσ-νRναI-πCLP-φLCπ-LσVδρRσδπσAVFFδσδF-αRAEφALπδLYτ-FCδCGD-----πD--σσ-δσ----σ-φσσ---
pR27_MuHV2    LSIVSNLRVEAEDLGVIQGVLPERVAALLGVP---AAHELFLSDEYYDCVEAAVDAMFDRFCECAECGMRRSERIVAVRARALGDARGLTAR
gpM27_MuHV1   MSLMSSLRIPGHDLEIVQRLLPERICQQLGIP---FSFESCLLDSYYEYVEMTMEVIYDQLCYCRECRMRRGERLERLRAQKIGTKY----R
Muromegalo    φSαφSσLRα-τδDL-ααQ-αLPERατ--LGαP---φτρE-φL-DδYYνφVEφφφνσφρDδφC-C-EC-MRRτERα--αRAδ-αGσ-δ----R
Gp13_HHV_6B   K----LKELTKESTRQLGGYNLCHWGQELKDCLE-NKSDVFFRYDFFERIESCLIEHFMLLCGCVEC----RRMFIMYNK------------
U5_HHV-7      K----LKELSKESKQKLGGFSCAFWSESFKTEMQ-NKTESFFRRDFFDRFQLYLLEHFLLFCGCEEC----RYNFFRFKN------------
U5_HHV_6A     K----LKELTKESTRQLGGYNLCHWGQELKDCLE-NKSDVFFRYDFFERIESCLIEHFMLLCGCVEC----RRMFIMYNK------------
Roseolo       K----LKELσKESφδδLGGρσφτρWτδδφKσδφδ-NKσνσFFRδDFFνRφδ-φLαEHFφLφCGC-EC----Rδ-Fφ-ρδδ------------
T27_TuHV-1    IG----DTLDDRARLLIADLLPRAVCRQLGYAYDPDDQTFIFDPQYLHASEQVLQILYDRLCDCEECQ---ERRFTAAAAATSLRRGPK---
ET            φ-------α---δ---α--φ-φ--φτ--φ---φ-------φφ--δρ----δ-φφ--φρ--φC-C-ν------δ------------------
PSI-PRED      c---hh??ccchhhhhhhhhhhhhhhcccccc?ccccccccccchhhhhhhhhhhhhhhhhhcccccc----cccccccccccccccccccc--

                  310       320       330       340       350       360             370       380       390
                   |         |         |         |         |         |               |         |         |
HCMV_AD169    -SPSAVASHHSRLEHAELRLERNRHLGAFHLPAIRHLTAGDVARVQDSVSRDLGFADWS-----QTLIDDYFLLPAGWACANPRRGYAMYLA
PoHV_4        -HAHMLPPPHFHLENAELRLENNRHLGAFHLPAIRHLTVDDISRIRASVSRDLGFAEWS-----QTLVDDYFLLPTGWACANPRRGYVMYLA
rh46_RhCMV    -SEEEPSPPQPRLEDAELTLLHNAHLGKFHLPAIRHLTAAEQLRVQQSVARDLGFTDWS-----HTLVDDYFLLPVGLSCGTPCEGYAMYLM
Cytomegalo    -δ-δ-ασσφδ-πLEδAELδLEδN-HLGφFHLPAIRHLTσσν--Rαδ-SVτRDLGFσνWS-----δTLαDDYFLLPφGφτCτσPδδGYαMYLφ
pR27_MuHV2    RGMLRKKRRNRFFEENDVALPFCPELGRLRLPRIRHLDLVQQELLCRHIGRDLMFDSMFGVCRGMSRMRGVPIPFCSAEKLNLDLAYVLYLA
gpM27_MuHV1   R---RNPDGIRLFEENYVRVAAYPNIGVLRLPKIRHLSDEHQTLVCGHLARDLMYDVHFGSSSPLARVNGIPLPFTAVDCMDIDLGYALYMA
Muromegalo    R---Rδ-δ--RφFEENδV-ασφφPδαG-LRLPπIRHLσ--δQδLαC-HατRDLMρDσφFGστ-σφτRφδGαPαPFφτανφφσαDLτYαLYφA
Gp13_HHV_6B   -----RGRKCDFG--HCVRIQCFPMIGSIRLPAFLHLGEPYSVSLSSLIAKDLGLSMIEG------QIEHSRLPISLQISVTPDKKALLTFL
U5_HHV-7      -----VGTMKKNP--GSVKLHFFPALGKIDLPIFPHLSEKYSN-LSMFVAKDLCLSFIEG------QIEHSRFPISVTVDMGQDKRNLLNIL
U5_HHV_6A     -----RGRKFDFG--HSVRIQCFPMIGSIRLPAFLHLGEPYDVSLSSLIAKDLGLSMIEG------QIELSRLPISLQISVTPDKKALLTFL
Roseolo       ------G-φφδ-σ---τVπαδφFPφαG-IδLPαFαHLτEφYσ--LS-φαAKDLτLSφIEG------QIEφSRφPISαδαδσ--DKπσLLσφL
T27_TuHV-1    ---RRKRSSWAAFERRDVVLEQYRELGVLRLPCLRHLKRDQQLELCQQLRHDLCFRRRQ-----TLTARELLLPIGVPCGHGVSDAYALYLA
```

```
ET            -----------------α-α-----αG-φδLP-φ-HL-------αδ--α-πDL-φ------------φδ---φαφ--------δ---αφ-φφ
PSI-PRED      -cccccccccchhhhhhhhhhhhhcccccchhhhhchhhhhhhhhhhhhcccccchh-----hhhhc?c?cccccccccccccchheehh
                     400       410       420       430       440                 450       460
                      |         |         |         |         |                   |         |
HCMV_AD169    SNAVLALRIIRLLRASIRHEYTACIRMLSGDVQRLIRLFKGEAALLRKGLAQNPVQRR----------ELSRFRKHVHDLKRIRFT------
PoHV_4        SNAVLALRIIRLLHASIRHEQTACVRMLSGDVQRLIRLFKGEAAMLRKGLAQNPVQRR----------ELSRFRKHVHDLKRVRFT------
rh46_RhCMV    SNAVLALKVIRLLYITIRHEHSACVRALCKDVAFLIRMFRYEASIIRKGLAQNATQRM----------ELSRFRRHIHDLKHIHFT------
Cytomegalo    SNAVLALπαIRLLδασIRHEδσACαRφLτ-DV--LIRφFπ-EAτφαRKGLAQNασQR-----------ELSRFRπHαHDLKπαπFT------
pR27_MuHV2    SNSYFLCFLIRTVRDVMRHEERAYGELCLGLARDAARALREDVEARVAAGAVPPRRTYFDFLDP-GLSAEEKDLSIVGALERVTFRGCRVVE
gpM27_MuHV1   NNVYFMLFLIRCVRDVVRSEQREYVDLVTGLVREASVMLRRDVDARIRAGADRPVFVASDPRQD-GISDEERDLSIANALDRLTFSDERPDG
Muromegalo    σNσYFφφFLIRφVRDVφRδEδR-YσνLφφGLαRνAτ-φLRδDVνARα-τGAσRP--φφ-Dφ-δσ-GαSσEEπDLSIασALνRαTFδσδRσσ-
Gp13_HHV_6B   TNIVFIVFVVNTLYRVINAELDIYYDLFTEEVGKLCVTMEEEMKLGRNGCLGDLCYFS--------------SMKQMKEIVRCPGE------
U5_HHV-7      SNIVFLLFIIQTLNSVLFTELKMYYDVYLDELKNLESSMECEMKLGSKGCMNNIVYFN--------------MLKQVKDIVRNPGT------
U5_HHV_6A     TNIVFIVFVVNTLYRVINAELDIYYDLFTEEVGKLCVAMEEEMKLGRNGCLGDLCYFS--------------PMKQMKEIVRCPGE------
Roseolo       σNIVFααFααδTLδδVα-σELδφYYDαρφνEα-δLδσσMEδEMKLGδδGCφσδασYFσ---------------φKQφKνIVRσPGδ------
T27_TuHV-1    SNLVFTLFLAQTLHAIVRIEARAYIDLLSGHADALVGALRRQASYLQHGGLGDGDGNDAGPSGGRSPSDGRKLARLTRRLRGLDFGFGGG--
ET            σN-φφφφ-ααδφαδ--φ--E---φ-δφφ---α--α---φδ-δφ-φ---τ-φ-δ---------------------δ----α------------
PSI-PRED      hhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhhcccccccchh----------hhhhhhhhhhhhhcc?c------

                                   470       480       490       500       510                       520
                                    |         |         |         |         |                         |
HCMV_AD169    --------------------EDTFVETFCDFLELVQRIPDYRSVSLRIKRELLCLHVFKLRRGCRAPPT----------------PETA---
PoHV_4        --------------------VDGFVESFCDFLDVIHRIPDYRCISLRIKRELLCLHTFKLRRVYRPPPSSLLAGSHHNNSVVLPPPTTTNAS
rh46_RhCMV    --------------------VSNFLETFCDFIKIVNKIPDYQCISLTIKRELVLLHLLRLRRVYECPSA-----------------ETR---
Cytomegalo    ---------------------σσFαEσFCDFαδααδπIPDYδταSLδIKRELαφLHφφπLRRσφδσPσσ-----------------δT----
pR27_MuHV2    RRDPDGTSVD--------ERVGCFRQDAAVVLGAAAAYHVPRALREHRAREELLAHTLGIRRLYDERVLAEFHG------------------
gpM27_MuHV1   YEDDDNEALERMFLEQRYERMGGFSRDAAKFMDLLDAFRVPKALRENRAREELLLHTFFIKRMYDERPLAGFYG------------------
Muromegalo    δδDσDσδταν--------ERφGτFδδDAAφφφσαασAρπVPπALREδRAREELLαHTφ-IπRφYDERαLA-FρG------------------
Gp13_HHV_6B   --------------------KSQFILKCWEALRIGFSVPAYKDYDETRFMEMFFLHHLHIKRFHEHNDRDLVSS------------------
U5_HHV-7      --------------------SSNFIFNCLEVIKMSFEIPYYKNYDETNFMESFYLHHLYIQR-QPAKHTDLVAA------------------
U5_HHV_6A     --------------------KSQFILKCWEALRIGFSVPAYKDYDETPFMEMFFMHHLHIKRFHEDNDRDLVSC------------------
Roseolo       --------------------δSδFIφδCφEααπφτFδαPφYKσYDET-FME-FρφHHLρIδR-δ--δδδDLVττ------------------
```

```
T27_TuHV-1     ----SG------------S-GTAFFTAFCDFLEIVEKLPDYARLSEFQVREELLVHHFYLAQVYADPSPETQAG------------------
ET             ---------------------σ-F---φφ-φφ-φ---φ--φ--φδ---φ-E-φφφHφφ-α-δ------------------------------
PSI-PRED       --------------------hhhhhhhhhhhhhhhhhccccchhhhhhhhhhhhhhhhhhhhhhcccccc----------------chhh---
                                     530       540       550               560       570       580       590
                                      |         |         |                 |         |         |         |
HCMV_AD169     ---------------RVQRLLWHSLRHGDAPQDRTRLPQFSSALSDAELSN-------HAN C   APLELGPAVVAAPGPSV Y A IQ F
PoHV_4         SPAPSSPHPDPQLLLRAERLAWHYLRHGDVPQDRSRLPQFSSALSDAELSN-------NAN C   APVEFAG---FAGGPSV Y A IQ F
rh46_RhCMV     ---------------KGEKLAWYYFRHGDVTLDPGLLMTFDSLLADCELSN-------NAN C   APFEIPT----SATVTV YQG IE L
Cytomegalo     ---------------πσδπLαWρδφRHGDσφ-D-σ-LφδFσSαLτDτELSN-------δAN C   APφEφσσ----σσσασV Yδτ Iδ F
pR27_MuHV2     ----------------EKLVPYHLCVGGHWRRDGWAVQARNVTLLDEEIVG--V G--DDMFAG VDFYE FEEVEVI EIPV  RTFMMEL
gpM27_MuHV1    ----------------DRLIPYYMFVGGYRRRDGAVIRMGAVTLSDTEIV GHW A--DDLVGM L YYDTFEYVDML NISVN QTLMMEL
Muromegalo     -----------------πLαPYρφφVGGρδRRDGφααδφ-σVTL-DδEIV---φ τ--DDφφτ- αδρYνδFEδVνφα δIσVδ δTφMMEL
Gp13_HHV_6B    ----------------DNLIPGFFIVNTHEENFLQRLQRVVLPVAEDYLTNT CINGTMAFFFSGL YFGSGNH GVQISPE DV AIGY L
U5_HHV-7       ----------------NNLAPGFFIVNAKEKSFIDVLERSIVNIEAEYLSNT NINGAMALFFSGL YFGNFGNGNFQTSPE DV AVGY L
U5_HHV_6A      ----------------DNLIPGFFIVNTDGENFLQRLQRVVLPVVEDYLTNT CINGTMAFFFSGL YFGSGNH GFQISPE DV AIAY L
Roseolo        ----------------σNLαPGFFIVNσδ-δσFαδ-LδRσα-σα--νYLσNT σINGσMAφFFSGL YFGσ-σδ-σφQφSPE DV AατY L
T27_TuHV-1     ----------------ERLVSYYVQNGDACCQDAARLRALVVDLSD ELS -------AP- AEPSWAWTIEP----TEIVFPP I EVT F
ET             -----------------δ-α--ρ----σ--------α------α---δασ---------σ-------φφ-----------α-φ-------δφ
PSI-PRED       ---------------hhhhhhhhhhccccccccccchhhhhhhhhhhhhhhh-------hhhhhhcccccccccceeeccccceceehhhhhhh

                        600
                         |
HCMV_AD169     ERLHVRRFRPHEVGGHAT-------------------------------------------------------------------------
PoHV_4         DRLYVRRFRPHEVGGHAT-------------------------------------------------------------------------
rh46_RhCMV     HRLYVRRFRPHEVGGCAI-------------------------------------------------------------------------
Cytomegalo     -RLρVRRFRPHEVGGδAφ-------------------------------------------------------------------------
pR27_MuHV2     QRMYVSKSELAAYADADVEIEEAADEDDGREVEFGPAGRDRREEGIRLEDYDDSDLGLDADHDPMFGIVDGDDDDDDGDDDDGGDDAMSGDE
gpM27_MuHV1    DRLYVPKRELVELVISD-------------------------------------------------KDSSGSDEETEDADSG----HSEP
Muromegalo     δRφYVσKδELα-φσ-τD-------------------------------------------------φDτσστDννσνDσDτG----δτν-
Gp13_HHV_6B    GSLDVLRDDYKYYEYAP-------------------------------------------------DCAGELNGHDGDE----------
U5_HHV-7       GGLDKIQNDLCYFANVET-------------------------------------------------LACVGVDASDGNE----------
U5_HHV_6A      GSLDVLRDDYKYYEYTPP-------------------------------------------------DCPGELNGHGGDE----------
```

```
Roseolo        GτLDφαδδDφφYρ-δφ-----------------------------------------------σσ--αστδσGσE----------
T27_TuHV-1     EDLYVRRYWSGEVGGWMP-----------------------------------------------HGEYTIEDEDRRVQHALEETARA
ET             --φδφ-δ-δ-------------------------------------------------------------------------------
PSI-PRED       hhhhhhcccccccccccc-----------------------------------------------------------------------
```

```
HCMV_AD169     ----------------------
PoHV_4         ----------------------
rh46_RhCMV     ----------------------
Cytomegalo     ----------------------
pR27_MuHV2     EDDEDADEGPRYVVERLEGYRT
gpM27_MuHV1    SDDGGESDGSEFVVERV-----
Muromegalo     -DD-σ-σνGσδρVVERα-----
Gp13_HHV_6B    ----------------------
U5_HHV-7       ----------------------
U5_HHV_6A      ----------------------
Roseolo        ----------------------
T27_TuHV-1     AYPWWHDHSAYAASGRRL----
ET             ----------------------
```

**Alignment data :**

| | | |
|---|---|---|
| Alignment length: 1122 | gpM27_MuHV1: YP_214040 (682 residues) | Gp13_HHV_6B: NP_050187 (903 residues). |
| Identity (*): 18 is 1.60 % | pR27_MuHV2: NP_064132 (671 residues). | U5_HHV_6A: NP_042896 (444 residues). |
| Strongly similar (:): 47 is 4.19 % | HCMV_AD169: NP_039961 (608 residues). | U5_HHV-7: P52522 (865 residues). |
| Weakly similar (.): 10 is 0.89 % | PoHV_4: NP_612671 (640 residues). | T27_TuHV-1: NP_116370 (636 residues). |
| Different: 1047 is 93.32 % | rh46_RhCMV: YP_068140 (579 residues). | |

# C. *ANNEXE C: Fichiers Utilisés Pour Le Calcul Du Modèle Du Sous-domaine Hélicase De PuL89.*

## C.1. Annexe C1: Fichier de contraintes (.rstrnt)

```
!BIOSYM        restraint      1
!
!

#NOE_distance
 1:ASP_309:OD1          2:OHH_1:HH1      2.000   3.000   3.000   100.00  100.00  1000.000  0.00     #Asp309/H2O
1:ASP_309:OD2  2:OHH_1:HH2      2.000   3.000   3.000   100.00  100.00  1000.000  0.00
 1:GLU_310:OE1 2:OHH_2:HH1      2.000   3.000   3.000   100.00  100.00  1000.000  0.00    #Glu310/H2O
2:OHH_2:HH2    2:PHOS_1:P1      2.000   3.500   3.000   100.00  100.00  1000.000  0.00      #H2O/γ PO4
2:MAG_1:MG1   2:OHH_1:OH       2.000   3.100   3.000   100.00  100.00  1000.000  0.00
2:MAG_1:MG1   1:THR_219:OG1  2.000   3.100   3.000   100.00  100.00  1000.000  0.00
2:MAG_1:MG1   2:PHOS_1C:O3    2.000   3.100   3.000   100.00  100.00  1000.000  0.00  # Chélation Mg2+
2:MAG_1:MG1   2:PHOS_1:O2      2.000   3.100   3.000   100.00  100.00  1000.000  0.00
 1:LYS_218:NZ    2:PHOS_1:O3      2.000   3.500   3.000   100.00  100.00  1000.000  0.00 #Lys218/ β et γPO4
1:LYS_218:NZ    2:PHOS_1C:O2    2.000   3.500   3.000   100.00  100.00  1000.000  0.00
 1:ARG_214:CZ   2:PHOS_1:P1      2.000   5.000   3.000   100.00  100.00  1000.000  0.00   #Arg253/ γPO4
 1:LYS_253:NZ    2:A_1B:P            2.000   3.500   3.000   100.00  100.00  1000.000  0.00  #Lys253/ α PO4
 1:GLN_191:O     2:A_1B:N6           2.000   3.500   3.000   100.00  100.00  1000.000  0.00 #Gln191/ N6
#mixing_times
-9.990000E+02
```

## C.2. Annexe C2: Fichier entrée pour minimsation sous contraintes (.inp)

```
!   INPUT FILE FOR DISCOVER GENERATED BY INSIGHT
!
!    fichier
      ifile = 1
      lastj = 100
      jfile = 1
      overlap = 0.01
      cutoff =  12.000000
      cutdis =  10.000000
      swtdis =  2.0
      FKCHIR = 10000000000.000
      IGRPCK = 0
      DEMAX = .10000E+10
      NSAVE = 100
!     BOXLIM  =     90
!     MIPBC   =     0
      Demax            = 10000000.000000
      begin simulation
!    *    add-automatic bond torsion valence out-of-plane
!     reduce
     set dielectric  = 4*r
       fixed atom list generation
    *    add main  atoms
    *   molecule 1 residues LEUN 138 to ILE 175
    *    add main  atoms
    *   molecule 1 residues ILE 192 to VAL 209
    *    add main  atoms
    *   molecule 1 residues VAL A44 to VAL A57
    *    add main  atoms
    *   molecule 1 residues LYS  218 to SERC 364
      constrain using "helicatpwmg3.rstrnt"
     scale nothing
     fkchir            =     10.000000
     scale nonbond term    by     1.000000
     scale coulombic       by     1.000000
     scale 1-4             by     1.000000
      cscale             =     1.000000
     dscale             =     1.000000
     fkchir    =  fkchir   *      1.000000
     scale_noe          =     0.000000
      scale out-of-plane    by     1.000000
     scale bond            by     1.000000
     scale theta           by     1.000000
     scale phi             by     1.000000
!
     minimize using steepest descents for 100 cycles
    *     with no morse functions and no cross terms
    *     until the maximum derivative is less than 0.050000
     minimize using conjugate gradients for 500 cycles
    *     with no morse functions and no cross terms
    *     until the maximum derivative is less than 0.050000
      initialize dynamics
    *   at 300 K
    *   for 5000 steps of 1.0 fs
    *   with no cross terms and no morse functions
    *   write history file every 5000 steps
```

```
    *   write averages every 5000 steps
lop1     resume dynamics
    *   at 300 K for 5000 steps
    *   with no cross terms and no morse functions
      minimize using steepest descents for 100 cycles
    *     with no morse functions and no cross terms
    *     until the maximum derivative is less than 0.10000
      minimize using conjugate gradients for 500 cycles
    *     with no morse functions and no cross terms
*     until the maximum derivative is less than 0.10000
  print noe_dist  violations exceeding  0.100000
  print dihedral  violations exceeding  1.000000
  print chirality violations
    archive as file number jfile "helicatpwmg3.arc"
    jfile = jfile + 1
    if jfile .le. lastj then lop1

end
```

# D. ANNEXE D: Modélisation Sous Contraintes RMN Et Tables De Déplacements Chimiques

## D.1. Annexe D1: Création du fichier de contraintes NOEs (.tbl)

En combinant les fichiers de volumes (vol) et de peak picking (.txt) par le biais du fichier .com qui va convertir au format XPLOR les données obtenues sous FELIX, le fichier de contraintes NOEs va être génér (.tbl).

pUL89DPC-293K.vol:

```
vol     4
   1    0.158E+09    0.148E+09    0.153E+08    0.943E+07    ┐
   2    0.419E+08    0.381E+08    0.339E+07    0.171E+07    |
   3    0.247E+08    0.264E+08    0.279E+07    0.188E+07    |
   4    0.246E+08    0.255E+08    0.284E+07    0.187E+07    |Volumes mesurés pour chaque pic
   5    0.109E+09    0.135E+09    0.198E+08    0.167E+08    |
[…]
 442    0.585E+08    0.502E+08    0.489E+07    0.266E+07    |
 443    0.314E+08    0.239E+08    0.205E+07    0.107E+07    ┘
 444    0.200E+03    0.150E+03    0.100E+03    0.500E+02    #  Temps de mélange
  ↑
 numéro du pic
```

pul89DPC-293K.txt:

```
   1  314.150    2.987    0  18hn        1226.358    8.658    0  18ha*
   2  402.095    2.329    0  19hn        1226.133    7.536    0  18ha*
   3  368.211    2.389    0  16hn        1456.117    5.696    0  16hb2
   4  368.148    2.834    0  16hn        1411.993    5.210    0  16hb1
   5 1456.376    3.826    0  16hb2       1412.193    5.590    0  16hb1
[…]
```

```
  442   530.558    2.361    0  11hd*          1140.701    4.249    0  15ha
  443   530.672    2.518    0  11hd*          1806.398    4.129    0  8hg2#
   ↑        ↑          ↑          ↑              ↑          ↑          ↑
 numéro  position   position    proton         position   position    proton
 du pic   D1 (points)  D1(ppm)   D1           D2 (points)  D2(ppm)   D2
```

pul89DPC-293K.com:

```
# 1. Buildup file(s) format (Felix/Xeasy/Xeasn)
#     Felix assignment-filename volume-filename (both in Felix 2.3 format)
#     Xeasy atomlist-filename sequence-filename
#     Xeasn assignment-filename sequence-filename
#     Examples
#     Felix assignment.file volume.file
Felix pul89DPC-293K.txt pul89DPC-293K.vol    ←invocation des fichiers .txt et .vol
#
# 2. Calibration Peak ids (all on one line)
#264 210 252
190 100 163                                  ←numéros des pics de calibration (corrélation Hε/Hδ de Phe590, Phe594 et Tyr598)
#241 225 244 284 209 232
#
# 3. Calibration distance {for the average of above peaks}
#   Power model
#     power is used as dist = (caldis/volume)^(1/power)
#     for single mixing time (e.g. as in H2O spectra)
#   Distance cutoff for all protons except the given list
#   List of protons for which the HIGHER cutoff applies
```

```
#    HIGHER distance cutoff for (ONLY) the give list of protons
#2.20 6 6.0 H5# H7# 5.5
2.46 6 6.0 hb# hg2# hd1# 5.5                    ←distance imposée aux pics de calibration
#
# 4. -/+ Delta for distance (dminus dplus)
#     defined as either % of distance
#                     or absolute
#     Examples:
#     10.0%           ! if - is the same as +, specified as % of actual dis
#     10.0% 20.0%  ! if - is different from +, specfied as % of actual dis
#      15.0% 0.25    ! - is %, + is absolute
#     0.20 0.50       ! if - is different from +
#      0.50            ! if - is the same as +
20.0% 20.0%                                     ←intervalle d'incertitude sur la mesure des distance
#
# 5. Ignore data point (vol at high mix time) < (vol at lower mix time)
No
#Yes
#
# 6. Output format (X-PLOR/DGPlay filename)
#     DGPlay filename
X-PLOR pul89DPC-293K.tbl                        ←nom du fichier de contraintes en sortie au format XPLOR (.tbl)
#
# 7. Supress output for following peaks
#     Syntax:
#      SKIP PEAKS peak-id peak-id ...
#      SKIP ATOM atomname res-numbers ...
```

```
#      SKIP ATOM atomname # (for all NOEs involving this atom, all residues)
#      SKIP INTRARES atom1 atom2 res-numbers ... (for intra-residue NOEs)
#      SKIP INTRARES atom1 atom2 # (wildcard for all intra-residue NOEs)
#      SKIP INTERSEG segment1 segment2    (implies all inter-segment NOEs)
#     You may write any number of such Skip lines,
#     provided each line begins with the keyword Skip.
#Skip intrares H5 H6 #
#Skip intrares H2' H2'' #
#Skip intrares H5 H5# #
#Skip atom H4' #
#Skip atom H5'* #
#Skip peaks 264 210 252 #
#Skip interseg DNA1 DRG1
#
# 8. Output distance table arranged sorted by peakid or resid.
#    Syntax:
#     SortBy PeakID
#     SortBy ResID
#SortBy PeakID
SortBy ResID                          ←tri des contraintes par nom du résidu en D1 dans .tbl
#
# 9. GnuPlot output (Yes/No)
#    PostScript output (Yes/No)
#    Peak-ids (optional)
#    No No
#    Yes Yes          (defaults to interactive)
#    Yes Yes all
```

```
#    Yes Yes interactive
#    Yes Yes 7 14 93
Yes Yes                                  ←tracé des courbes de build-up pour chaque pic
#
#10. Xeasy Buildup files (mixing-time filename) {only for Xeasy or Xeasn input}
#      In decreasing order of mixing time.
160.0 bu160.peaks
120.0 bu120.peaks
 80.0 bu80.peaks
 40.0 bu40.peaks
#
# End of file
```

pul89DPC-293K.tbl :

```
set echo=false end
                                        distance mesurée
                                              |      intervalles
                                              |       +    -      n° pic
                                              |       |    |      |   qualité
                                              |       |    |      |   du pic
                                              |       |    |      |     |     build-up
                                              |       |    |      |     |     |   remarque (Bad)
                                              ↓       ↓    ↓      ↓     ↓     ↓       ↓
assi (resi   2 and name hn   and segi PU89)                                                  ←proton en D1
    (resi   1 and name ha*  and segi PU89)  2.97  0.59  0.59 !   21 -v-  4.23                 ←proton en D2
assi (resi   2 and name hd*  and segi PU89)
    (resi   2 and name ha   and segi PU89)  4.79  0.96  0.96 !   33 -v-  3.63
assi (resi   2 and name he   and segi PU89)
    (resi   2 and name hb*  and segi PU89)  3.75  0.75  0.75 !   31 -v-  4.30
!assi (resi   2 and name hn   and segi PU89)
!    (resi   2 and name hd*  and segi PU89)  6.00  1.20  1.20 !   22 -v-  3.56  6.40
assi (resi   2 and name hn   and segi PU89)
    (resi   2 and name ha   and segi PU89)  3.56  0.71  0.71 !   19 -v-  4.47
    (resi   2 and name hd*  and segi PU89)  3.31  0.66  0.66 !  381 -v-  4.49
assi (resi   2 and name hb2  and segi PU89)
    (resi   2 and name ha   and segi PU89)  3.21  0.64  0.64 !  382 -v-  4.24
assi (resi   2 and name hg*  and segi PU89)
    (resi   2 and name hb1  and segi PU89)  2.48  0.50  0.50 !   39 -v-  4.20
!assi (resi   2 and name he   and segi PU89)
!    (resi   2 and name ha   and segi PU89)  6.00  1.20  1.20 !  326 Bad 100.00 60.00
```

## D.2. Annexe D2: pUL89(580-600)

| | | H2O | 30% TFE | 30% TFE | 120eq. DPC |
|---|---|---|---|---|---|
| | | pH 3,5 | pH 6,8 | pH 3,3 | pH 3,7 |
| $G^{580}$ | NH | - | - | - | - |
| | α | 3.81,3.81 | - | 3.92,3.92 | 3.87,3.87 |
| $R^{581}$ | NH | 8.71 | - | 8.88 | 8.87 |
| | α/β | 4.29/ 1.78,1.71 | 4.27/ 1.51,1.51 | 4.31/ 1.85,1.85 | 4.33/ 1.83,1.79 |
| | γ/δ | 1.58,1.58/ 3.16,3.16 | 1.89,1.73/ 3.24,3.24 | 1.85,1.85/ 3.22,3.22 | 1.63,1.63/ 3.16,3.16 |
| | ε | 7.20 | - | 7.27 | 7.48 |
| $D^{582}$ | NH | 8.69 | 8.65 | 8.60 | 8.60 |
| | α/β | 4.59/ 2.81,2.71 | 4.45/ 2.79,2.68 | 4.61/ 2.95,2.84 | 4.59/ 2.89,2.77 |
| $K^{583}$ | NH | 8.51 | 8.15 | 8.37 | 8.92 |
| | α/β | 4.19/ 1.79,1.71 | 4.19/ 1.92,1.92 | 4.14/ 1.91,1.91 | 4.03/ 1.93,1.70 |
| | γ/δ | 1.40,1.40/ 1.64,1.64 | 1.53,1.46/ 1.73,1.73 | 1.53,1.47/ 1.72,1.72 | 1.53,1.42/ 1.70,1.70 |
| | ε/ξ | 2.95,2.95/ 7.56 | 2.99,2.99 | 2.96,2.96/ - | 3.00,2.97/ - |
| $A^{584}$ | NH | 8.34 | 8.03 | 7.99 | 8.44 |
| | α/β | 4.21/ 1.33 | 4.15/ 1.51 | 4.11/ 1.48 | 3.95/ 1.46 |
| $L^{585}$ | NH | 8.18 | 8.16 | 7.75 | 7.83 |
| | α/β | 4.24/ 1.59,1.59 | 4.22/ 1.80,1.74 | 4.18/ 1.71,1.71 | 4.08/ 1.66,1.66 |
| | γ/δ | 1.53/ 0.89,0.83 | 1.68/ 0.98,0.94 | 1.56/ 1.00,094 | 1.60/ 0.96,0.90 |
| $A^{586}$ | NH | 8.32 | 7.90 | 7.81 | 7.83 |
| | α/β | 4.25/ 1.33 | 4.23/ 1.59 | 4.19/ 1.56 | 4.19/ 1.52 |
| $V^{587}$ | NH | 8.16 | 8.19 | 8.15 | 8.05 |

| | | | | | |
|---|---|---|---|---|---|
| | α/β | 4.00/ 2.01 | 3.76/ 2.25 | 3.76/ 2.25 | 3.69/ 2.30 |
| | γ | 0.90,0.90 | 1.08,1.01 | 1.08,1.02 | 1.08,0.99 |
| | NH | 8.45 | 8.38 | 8.39 | 8.45 |
| $E^{588}$ | α/β | 4.22/ 1.93,1.93 | 4.00/ 2.22,2.15 | 4.01/ 2.32,2.25 | 3.86/ 2.33,2.26 |
| | γ | 2.36,2.36 | 2.48,2.28 | 2.67,2.46 | 2.61,2.61 |
| | NH | 8.43 | 8.39 | 8.56 | 8.70 |
| $Q^{589}$ | α/β | 4.19/ 1.92,1.87 | 4.07/ 2.32,2.17 | 4.01/ 2.32,2.14 | 4.03/ 2.26,2.05 |
| | γ/ε | 2.23,2.18/ 7.54,6.91 | 2.58,2.43/ 7.51,6.84 | 2.62,2.41/7.31,6.78 | 2.58,2.42/ 7.53,6.88 |
| | NH | 8.38 | 8.27 | 8.22 | 8.05 |
| $F^{590}$ | α/β | 4.56/ 3.07,2.99 | 4.23/ 3.42,3.29 | 4.25/ 3.43,3.31 | 4.25/ 3.33,3.33 |
| | δ/ε/ξ | 7.18,7.18/ 7.27,7.27/- | 7.22,7.22/ 7.15,7.15/7.22 | 7.23,7.23/ 7.18,7.18/- | 7.23,7.23/ 7.16,7.16/- |
| | NH | 8.15 | 8.88 | 8.92 | 8.45 |
| $I^{591}$ | α/β | 4.05/ 1.74 | 3.67/ 2.01 | 3.68/ 2.00 | 3.69/ 2.06 |
| | γ/δ | 1.42,1.42,1.11/ 0.79 | 1.93,1.32,1.00/0.90 | 1.93,1.36,1.01/ 0.90 | 1.92,1.34,0.99/ 0.88 |
| | NH | 8.40 | 8.21 | 8.22 | 8.16 |
| $S^{592}$ | α/β | 4.32/ 3.84,3.77 | 4.27/ 4.05,4.05 | 4.26/ 4.05,4.05 | 4.21/ 4.02,4.02 |
| | NH | 8.36 | 7.91 | 7.83 | 7.60 |
| | α/β | 4.15/ 1.57,1.57 | 4.10/ 1.84,1.74 | 4.09/ 1.73,1.36 | 4.09/ 1.65,1.65 |
| $R^{593}$ | γ/δ | 1.34,1.34/ 3.02,3.02 | 1.70,1.50/ 3.05,3.05 | 1.62,1.47/3.05,3.05 | 1.56,1.44/ 3.01,3.01 |
| | ε | 7.10 | - | 7.13 | 7.24 |
| $F^{594}$ | NH | 8.25 | 8.45 | 8.37 | 7.89 |

| | | | | | |
|---|---|---|---|---|---|
| | α/β | 4.54/ 3.09,2.91 | 4.26/ 3.09,2.74 | 4.28/ 3.10,2.73 | 4.25/ 3.09,2.83 |
| | δ/ε/ξ | 7.17,7.17/ 7.27,7.27/ - | 7.14,7.14/ 7.24,7.24/ 7.24 | 7.13,7.13/ 7.23,7.23/ - | 7.10,7.10/ 7.03,7.03/7.07 |
| | NH | 8.33 | 8.42 | 8.37 | 8.03 |
| $N^{595}$ | α/β | 4.64/ 2.77,2.66 | 4.67/ 2.90,2.90 | 4.68/ 2.94,2.85 | 4.59/ 2.93,2.70 |
| | δ | 7.65,6.93 | 7.62,6.75 | 7.58,6.73 | 7.69,6.93 |
| $S^{596}$ | NH | 8.37 | 8.03 | 8.05 | 8.30 |
| | α/β | 4.29/ 3.88,3.82 | 4.40/ 3.95,3.95 | 4.38/ 3.98,3.98 | 4.23/ 3.88,3.88 |
| $G^{597}$ | NH | 8.44 | 8.06 | 8.12 | 8.29 |
| | α | 3.86,3.86 | 3.94,3.88 | 3.90,3.90 | 3.80,3.80 |
| | NH | 7.98 | 7.88 | 7.86 | 7.87 |
| $Y^{598}$ | α/β | 4.47/ 2.92,2.92 | 4.53/ 3.09,2.89 | 4.50/ 3.07,2.90 | 4.36/ 2.97,2.84 |
| | δ/ε | 7.01,7.01/ 6.74,6.74 | 7.10,7.10/ 6.83,6.83 | 7.10,7.10/ 6.83,6.83 | 7.06,7.06/ 6.79,6.79/ - |
| | NH | 8.05 | 7.86 | 7.79 | 7.90 |
| $I^{599}$ | α/β | 4.01/ 1.71 | 4.16/ 1.86 | 4.13/ 1.85 | 4.13/ 1.98 |
| | γ/δ | 1.38,1.38,1.07/ 0.82 | 1.45,1.14,0.93/ 0.86 | 1.46,1.14,0.94/ 0.87 | 1.50,1.18,0.91/ 0.87 |
| | NH | 8.27 | 7.78 | 8.16 | 8.10 |
| | α/β | 4.11/ 1.79,1.79 | 4.20/ 1.84,1.84 | 4.35/ 1.92,1.80 | 4.18/ 1.82,1.82 |
| $K^{600}$ | γ/δ | 1.42,1.42/ 1.64,1.64 | 1.40,1.40/ 1.75,1.75 | 1.45,1.45/ 1.70,1.70 | 1.38,1.38/ 1.73,1.64 |
| | ε/ξ | 2.94,2.94/ - | 3.01,3.01/ - | 3.00,3.00/ - | 2.95,2.95/ 7.58 |

## D.3. ANNEXE D3: RMN de pUL89-XI tables de déplacements chimiques

283 K, pH 3,2, 50% H2O/ 50% acétonitrile

| | NH | Hα | Hβ | Autres |
|---|---|---|---|---|
| Asp568 | - | 4,13 | 2,81, 2,76 | |
| Gln569 | 8,69 | 4,17 | 1,98, 1,87 | 2,24(γ*) |
| Asn570 | 8,16 | 4,51 | 2,6, 2,6 | 7,55(δ21), 6,73(δ22) |
| His571 | | 8,02 | 4,59 | 3,19, 2,99 7,18(δ2), 8,49(ε1) |
| Ile572 | | 7,96 | 4,02 | 1,75 1,08(γ12 & 13 & 2), 0,79(δ1) |
| Glu573 | 8,17 | 4,26 | 1,97, 1,84 | 2,28(γ*) |
| Gln574 | 8,01 | 4,47 | 1,97, 1,82 | 2,25(γ*), 7,32(ε21), 6,62(ε22) |
| Pro575 | | - | 4,21 | 2,07, 1,51 1,78(γ2), 1,98(γ3), 3,62(δ2), 3,54(δ3) |
| Phe576 | 7,9 | 4,34 | 2,97, 2,97 | 7,08(δ*), 7,24(ε*) |
| Tyr577 | | 7,52 | 4,28 | 2,93, 2,93 6,98(δ*), 6,71(ε*) |
| Leu578 | 7,66 | 4,08 | 1,53, 1,44 | 1,42(γ), 0,79(δ1 & 2) |
| Met579 | 7,82 | 4,21 | 1,97, 1,97 | 2,53(γ2), 2,44(γ3) |
| Gly580 | 8,06 | 3,84, 3,79 | | |
| Arg581 | 8,08 | 3,99 | 1,76, 1,71 | 1,59(γ2), 1,47(γ3), 3,07(δ*), 7,14(ε) |
| Asp582 | 8,29 | 4,34 | 2,68, 2,68 | |
| Lys583 | 7,94 | 3,99 | 1,76, 1,76 | 1,4(γ2), 1,31(γ3), 1,56(δ*), 2,83(ε*), 7,42(ζ) |
| Ala584 | | 7,92 | 4,00 | 1,36 |
| Leu585 | 7,98 | 4,01 | 1,69, 1,56 | 0,87(δ1), 0,82(δ2) |

| | | | | | |
|---|---|---|---|---|---|
| Ala586 | | 7,77 | 4,08 | 1,44 | |
| Val587 | | 8,07 | 3,68 | 2,15 | 0,99(γ1), 0,92(γ2) |
| Glu588 | 8,17 | 3,94 | | 2,14, 2,14 | 2,54(γ2), 2,40(γ3) |
| Gln589 | 8,43 | 3,95 | | 2,15, 1,99 | 2,45(γ2), 2,29(γ3), 7,08(ε21), 6,76(ε22) |
| Phe590 | 8,12 | 4,15 | | 3,33, 3,19 | 7,17(δ*), 7,18(ε*) |
| Ile591 | | 8,6 | 3,59 | 1,91 | 1,84(γ12), 1,23(γ13), 0,89(γ2), 0,80(δ1) |
| Ser592 | | 8,13 | 4,19 | 3,92, 3,89 | |
| Arg593 | 7,87 | 3,96 | | 1,68, 1,62 | 1,41(γ*), 3(δ*), 7,12(ε) |
| Phe594 | 8,23 | 4,08 | | 3,03, 2,64 | 7,1(δ*), 7,13(ε*) |
| Asn595 | 8,38 | 4,48 | | 2,8, 2,75 | 7,28(δ21), 6,51(δ22) |
| Ser596 | | 8,15 | 4,12 | 3,9, 3,87 | |
| Gly597 | 8,12 | 3,76, 3,76 | | | |
| Tyr598 | | 8,03 | 4,09 | 2,98, 2,78 | 6,92(δ*), 6,66(ε*) |
| Ile599 | | 8,08 | 3,58 | 1,86 | 1,16(γ12 & 13 & 2), 0,81(δ1) |
| Lys600 | 7,93 | - | | 1,76, 1,76 | 1,41(γ2) 1,32(γ3), 1,57(δ*), 2,82(ε*) |
| Ala601 | | 8,07 | 4,01 | 1,31 | |
| Ser602 | | 8,01 | 3,99 | 3,85, 3,66 | |
| Gln603 | 8,10 | 3,92 | | 2,18, 2,18 | 2,47(γ2), 2,26(γ3), 7,10(ε21), 6,51(ε22) |
| Glu604 | 8,13 | 3,99 | | 2,18, 2,12 | 2,49(γ2), 2,36(γ3) |
| Leu605 | 7,99 | 4,16 | | 1,87, 1,68 | 1,75(γ), 0,84(δ1 & δ2) |
| Val606 | | 8,12 | 3,65 | 2,11 | 1,01(γ1), 0,88(γ2) |
| Ser607 | | 8,05 | 4,03 | 3,89,3,83 | |
| Tyr608 | | 8,25 | 4,17 | 3,10, 3,10 | 7,01(δ*), 6,68(ε*) |
| Thr609 | | 8,11 | 3,72 | 4,28 | 1,16(γ2) |
| Ile610 | | 8,21 | 3,66 | 1,85 | 1,05((γ12 & 13), 0,77(γ2), 0,75(δ1) |
| Lys611 | 7,87 | 3,93 | | 1,78, 1,78 | 1,45(γ*), 1,55(δ*), 2,84(ε*) |
| Leu612 | 7,97 | 3,97 | | 1,77, 1,47 | 1,43(γ), 0,68(δ1 & δ2) |

| | | | | | |
|---|---|---|---|---|---|
| Ser613 | | 7,64 | 4,14 | 3,80, 3,74 | |
| His614 | | 7,78 | 4,57 | 3,29, 3,05 | 7,23(δ2), 8,48(ε1) |
| Asp615 | 7,98 | 4,88 | | 2,85, 2,71 | |
| Pro616 | | - | 4,32 | 2,29, 1,97 | 2(γ2), 1,9(γ3), 3,76(δ*) |
| Ile617 | | 7,86 | 3,84 | 1,95 | 1,51(γ12), 1,24(γ13), 0,84(γ2), 0,83(δ1) |
| Glu618 | 7,75 | 3,91 | | 1,97, 1,97 | 2,35(γ2), 2,32(γ3) |
| Tyr619 | | 7,83 | 4,08 | 3, 3 | 6,98(δ*), 6,66(ε*) |
| Leu620 | 7,79 | 3,98 | | 1,73, 1,66 | 1,66(γ), 0,90(δ1), 0,87(δ2) |
| Leu621 | 8,21 | 3,90 | | 1,77, 1,77 | 1,42(γ), 0,81(δ1), 0,78(δ2) |
| Glu622 | 8,08 | 3,9 | | 2,06, 1,99 | 2,52(γ2), 2,3(γ3), |
| Gln623 | 7,95 | 3,96 | | 1,98, 1,98 | 2,10(γ2), 2,06(γ3),6,53(ε21), 6,45(ε22) |
| Ile624 | | 8,15 | 3,59 | 1,86 | 1,70(γ12&13), 1,05(γ2), 0,79(δ1) |
| Gln625 | 8,09 | 3,96 | | 2,06, 1,98 | 2,38(γ2), 2,27(γ3), 7,10(ε21), 6,59(ε22) |
| Asn626 | 8,14 | 4,41 | | 2,75, 2,72 | 7,4(δ21), 6,73(δ22) |
| Leu627 | 7,96 | 4,05 | | 1,68, 1,54 | 1,54(γ), 0,83(δ1), 0,80(δ2) |
| His628 | | 8,06 | 4,41 | 3,29, 3,16 | 7,23(δ2), 8,43(ε1) |
| Arg629 | 7,89 | 4,03 | | 1,85, 1,67 | 1,52(γ*), 3,13(δ*), 7,2(ε) |
| Val630 | | 7,94 | 3,9 | 2,09 | 0,94(γ1), 0,87(γ2) |
| Thr631 | | 7,81 | 4,13 | 4,1 | 1,12(γ2) |
| Leu632 | 7,82 | 4,14 | | 1,59, 1,46 | 1,46(γ), 0,77(δ1 & δ2) |
| Ala633 | | 7,78 | 4,16 | 1,32 | |
| Glu634 | 7,85 | 4,25 | | 2,06, 1,90 | 2,41(γ2), 2,35(γ3) |
| Gly635 | 7,76 | 3,98, 3,84 | | | |

## D.4. ANNEXE D4: pUL56(506-548)

Les déplacements chimiques indiqués ont été obtenu pour la forme réduite du pont disulfure. En rouge et entre parenthèses, sont indiqués les écarts relatifs des déplacements chimiques obtenus avec la forme oxydée (δ réduit – δ oxydé) exprimés en ppm 10 E-02. Un tiret est indiqué pour les protons non attribués.

303 K, pH 6,1, 70% H2O/ 30%TFE

| | **NH** | **Hα** | **Hβ** | **Autres** |
|---|---|---|---|---|
| Arg506 | - | - | - | - |
| Ile507 | | - | - | - |
| His508 | | - | 4,77 | 3,24, 3,17 | 7,21(δ2), 8,18(ε1) |
| Asn509 | 8,29 | 4,81 | 3,00, 2,83 | 7,69(δ21), 6,91(δ22) |
| Val510 | | 8,29 | 4,09 | 2,24 | 1,07(γ1), 1,04(γ2) |
| Thr511 | | 7,98 | 4,34 | 4,13 | 1,34(γ2) |
| Arg512 | 8,03 | 4,24 | 1,99, 1,99 | 1,80(γ2), 1,71(γ3),3,29(δ*), 7,35(ε) |
| Glu513 | 8,21 | 4,15 | 2,21, 2,21 | 2,43(γ*) |
| Val514 | | 8,31 | 3,82 | 2,23 | 1,14(γ1), 1,04(γ2) |
| Asn515 | 8,05 | 4,61 | 3,05, 2,88 | 7,59(δ21), 6,89(δ22) |
| Val516 | | 8,30 | 3,81 | 2,30 | 1,17(γ1), 1,07(γ2) |
| Arg517 | 8,07 | 4,13 | 2,06, 1,97 | 1,76(γ*), 3,27(δ1),3,21(δ2),-(ε) |
| Lys518 | 8,40 | 4,16 | 2,04, 2,04 | 1,51(γ*), 1,81(δ*),3,02(ε*), -(ζ,-) |
| Arg519 | 7,95 | 4,07 | 1,91, 1,91 | 1,69(γ*), 3,27(δ*),7,31(ε,-) |
| Ala520 | | 8,33 | 4,24 | 1,62 | |
| Tyr521 | | 8,26 | 4,32 | 3,28, 3,28 | 7,14(δ*), 6,82(ε*) |
| Leu522 | 8,44 | 4,06 | 2,08, 2,02 | 1,64(γ), 1,01(δ1,**2**),1,01(δ2,**3**) |
| Gln523 | 8,37 | 4,08 | 2,30, 2,30 | 2,56(γ1), 2,46(γ2), 7,64(ε21**,-4**),6,68(ε22) |
| Lys524 | 7,94 | 4,24 | 2,04, 2,04 | 1,61(γ*), 1,80(δ1), 1,80(δ2), 3,02(ε*), -(ζ) |

| | | | | | |
|---|---|---|---|---|---|
| Val525 | | 8,39 | 3,71(**4**) | 2,09 | 0,89(γ1,**-3**), 0,84(γ2,**3**) |
| Ser526 | | 8,07 | 4,30 | 4,15, 4,15 | |
| Glu527 | 8,03 | 4,19 | 2,36, 2,25 | 2,64(γ*) | |
| Val528 | | 8,15 | 3,99(**9**) | 2,30 | 1,16(γ1,**-3**), 1,05(γ2) |
| Gly529 | 8,31 | 3,99, 3,93 | | | |
| Tyr530 | | 8,31 | 4,44 | 3,18, 3,18 | 7,18(δ*), 6,87(ε*) |
| Gly531 | 8,33 | 4,00, 3,84 | | | |
| Lys532 | 7,94 | 4,15 | 2,08, 2,02 | 1,51(γ*), 1,79(δ1), 1,69(δ2), 3,03(ε*),-(ζ) | |
| Val533 | | 7,78 | 3,71(**9**) | 2,31 | 1,13(γ1,**2**), 1,03(γ2,**3**) |
| Ile534 | | 8,23 | 3,81(**4**) | 1,88 | 1,56(γ12,**5**), 1,27(γ13), |
| | | | | | 0,95(γ2,**2**), 0,85(δ1,**5**) |
| Arg535 | 7,95 | 4,23 | 2,04, 2,04 | 1,62(γ*), 3,27(δ*), 7,14(ε,-) | |
| Cys536 | 8,16 | 4,24 | 3,30, 3,12 | | |
| Ile537 | | 8,53 | 3,82 | 2,04 | 1,86(γ12), 1,22(γ13),1,00(γ2),0,90(δ1) |
| Lys538 | 8,75 | 4,24 | 2,06, 2,06 | 1,66(γ*), 1,76(δ*),2,93(ε*), 7,31(ζ) | |
| Thr539 | | 8,14 | 4,51(**8**) | 4,08(**8**) | 1,35(γ2,**2**) |
| Gln540 | 8,23 | 4,14 | 2,31, 2,31 | 2,57(γ2), 2,47(γ3), | |
| | | | | 7,15(ε21,**-11**),6,56(ε22, **-11**) | |
| Glu541 | 8,51 | 4,09 | 2,30, 2,22 | 2,56(γ*) | |
| Arg542 | 7,98 | 4,26 | 2,11, 2,11 | 1,85(γ*), 3,32(δ*),7,36(ε) | |
| Leu543 | 8,26 | 4,33 | 1,94, 1,90 | 1,71(γ), 0,98(δ1),0,98(δ2) | |
| Thr544 | | 7,98 | 4,45 | 4,34 | 1,36(γ2) |
| Ser545 | | 7,88 | 4,50 | 4,07, 4,07 | |
| Lys546 | 8,05 | 4,45 | 2,04, 2,04 | 1,94(γ*), 1,56(δ*),3,10(ε*), -(ζ) | |
| Leu547 | 8,09 | 4,47 | 1,77, 1,77 | 1,71(γ), 1,01(δ1),0,96(δ2) | |
| Ile548 | | 7,41(**6**)4,18 | | 1,92 | 1,53(γ12), 1,23(γ13), |
| | | | | | -(γ2,*(0,97)*),-(δ1,(*0,94*)) |

# E. ANNEXE E : PUL89-endo Par MODELLER

## E.1. Annexe E1: Fichiers d'alignement de séquence pour le calcul des formes libre et lié au duplexe(.ali).

### *E.1.a. Forme libre*

```
>P1;UL89
sequence:UL89:: :: :: : :
KTLYVYLDPAFTTNRKASGTGVAAVGAY------RHQFLIYGL--EHFFLRDLSESSEVAIAECAAHMIISVLS-
LHPYLDELRIAVEGN-----------------TNQAAAVRIACLIRQSVQSSTLIRVLFYHTPDQNHIEQPFYLM
GRDKALAVEQFISRFNSGYIKASQELVSYTIKL------------------------
..ww
------------
------------*

>P1;2ehg
structure:2ehg: 1:A: 149: A:: : :
-MIIGYFDGLCEPKNPGGIATFGFVIYL-----DNRKIEGYGLA-EKPFSIN--STNNVAEYSGLICLMETML-R
LGIS--SPIIKGDSQLVIKQMNGEYKVKA----KRIIPLYEKAIELKKKLN------ATLIWVP-----------
-REENKEADRLSRVAYELVRRGKLRDIGCIILT------------------------
----
------------
------------*
>P1;1zbi2
structure:1zbi2: 64: A: 232: A:: : :
IWESLSVDVGSQGNP---GIVEYKGVDTKTGEVLFEREPIPIG------------TNNMGEFLAIVHGLRYLKER
NSR----KPIYSNSQTAIKWVKDKKAKSTLVRNEETALIWKLVDEAEEWLNT-----------------------
--------------------------------HTYETPILKWQTDKWGEIKADYGRK
..ww .......................*
```

*E.1.b.* *Forme liée*

```
>P1;UL89
sequence:UL89:: :: :: : :
KTLYVYLDPAFTTNRKASGTGVAAVGAY------RHQFLIYGL--EHFFLRDLSESSEVAIAECAAHMIISVLS-
LHPYLDELRIAVEGN-----------------TNQAAAVRIACLIRQSVQSSTLIRVLFYHTPDQNHIEQPFYLM
GRDKALAVEQFISRFNSGYIKASQELVSYTIKL------------------------
..ww
.......................*

>P1;2ehg
structure:2ehg: 1:A: 149: A:: : :
-MIIGYFDGLCEPKNPGGIATFGFVIYL-----DNRKIEGYGLA-EKPFSIN--STNNVAEYSGLICLMETML-R
LGIS--SPIIKGDSQLVIKQMNGEYKVKA----KRIIPLYEKAIELKKKLN------ATLIWVP-----------
-REENKEADRLSRVAYELVRRGKLRDIGCIILT------------------------
----------------------------*
>P1;1zbi2
structure:1zbi2: 64: A: 232: A:: : :
IWESLSVDVGSQGNP---GIVEYKGVDTKTGEVLFEREPIPIG------------TNNMGEFLAIVHGLRYLKER
NSR----KPIYSNSQTAIKWVKDKKAKSTLVRNEETALIWKLVDEAEEWLNT-----------------------
--------------------------------HTYETPILKWQTDKWGEIKADYGRK
..ww.......................*
```

## E.2. Annexe E2: Fichier de contraintes (.py)

```
# Addition of restraints to the default ones
from modeller import *
from modeller.automodel import *    # Load the automodel class
log.verbose()
env = environ()
# Give less weight to all soft-sphere restraints:
env.schedule_scale = physical.values(default=1.0, soft_sphere=0.7)
env.io.atom_files_directory = ['.', '../atom_files']
# Read in HETATM records from templates PDBs
env.io.hetatm = True                                #champs de force des atomes lourds
env.io.water = True                                 #champs de force de l'eau
class MyModel(allhmodel):
    def special_restraints(self, aln):
        rsr = self.restraints
        at = self.atoms
        rsr.add(secondary_structure.alpha(self.residue_range('38:', '44:')))     ┐ hélices PSI-PRED
        rsr.add(secondary_structure.alpha(self.residue_range('84:', '103:')))    ┘
        rsr.add(secondary_structure.alpha(self.residue_range('120:', '126:'))) ┐ hélices
        rsr.add(secondary_structure.alpha(self.residue_range('128:', '139:')))  | contraintes RMN
        rsr.add(secondary_structure.alpha(self.residue_range('141:', '155:'))) ┘
        rsr.add(secondary_structure.strand(self.residue_range('19:', '28:')))    ┐
        rsr.add(secondary_structure.strand(self.residue_range('31:', '36:')))     | brins PSI-PRED
        rsr.add(secondary_structure.strand(self.residue_range('73:', '79:')))     |
        rsr.add(secondary_structure.strand(self.residue_range('105:', '111:'))) |
```

```
    rsr.add(secondary_structure.strand(self.residue_range('116:', '118:')))┘
    rsr.add(secondary_structure.sheet(at['O:74'], at['N:108'], ┐
                        sheet_h_bonds=5))                      | appariements
    rsr.add(secondary_structure.sheet(at['O:36'], at['N:22'],  | des brins
                        sheet_h_bonds=-5))                     |
    rsr.add(secondary_structure.sheet(at['O:75'], at['N:5'],   |
                        sheet_h_bonds=8))                      ┘
    rsr.add(forms.gaussian(group=physical.xy_distance,         ┐
                feature=features.distance(at['MG:159'],        |
                                  at['OD2:8']),                |
                mean=2.6, stdev=0.5))                          | contraintes de distances
    rsr.add(forms.gaussian(group=physical.xy_distance,         | Magnesium/ AA acides
                feature=features.distance(at['MG:158'],        |
                                  at['OD1:8']),                |
                mean=2.6, stdev=0.5))                          |
    rsr.add(forms.gaussian(group=physical.xy_distance,         |
                feature=features.distance(at['OE2:133'],       |
                                  at['MG:159']),               |
                mean=2.6, stdev=0.5))                          |
    rsr.add(forms.gaussian(group=physical.xy_distance,         |
                feature=features.distance(at['OE1:79'],        |
                                  at['MG:158']),               |
                mean=2.6, stdev=0.5))                          ┘
a = MyModel(env,
        alnfile = 'MUSTER_secstr.ali',    # fichier d'alignement
        knowns   = ('2ehg','1zbi2'),      # codes des structure-supports dans .ali
        sequence = 'UL89',                # sequence cible dans .ali
```

```
                    assess_methods=(assess.DOPE, assess.GA341)) #estimation de la qualité par MODELLER
a.starting_model= 1                    # premier model
a.ending_model = 100                   # dernier model
                                 # 100 modèles calculés
 a.make()                       # execution du modelling

# Get a list of all successfully built models from a.outputs
ok_models = filter(lambda x: x['failure'] is None, a.outputs)
# Tri des modèle par la function DOPE
key = 'DOPE score'
ok_models.sort(lambda a,b: cmp(a[key], b[key]))

# génère la liste des modèles triés par DOPE score
m = ok_models[0]
print "Top model: %s (DOPE score %.3f)" % (m['name'], m[key])
```

## Etudes structurales de nouvelles cibles thérapeutiques du CMVH révélées par les dérivés benzimidazolés ribonucléosides

**Le Cytomégalovirus Humain est un virus touchant toutes les populations quelque soit leur milieu socioprofessionnel. Inoffensif chez des sujets sains, les dégâts causés par le virus chez l'hôte se révèlent désastreux dès que celui-ci devient immunodéprimé. Ainsi, les porteurs de VIH, les personnes ayant subi une greffe d'organe ou les nourrissons sont particulièrement exposés à CMVH. Les premiers inhibiteurs formulés visant la polymérase virale se sont avérés inefficaces suite à l'observation de multiples mutations de résistance. Une nouvelle classe de molécule, les dérivés benzimidazolés ribonucléosides, visant les étapes de maturation et d'encapsidation, a donc émergé visant les protéines du complexe terminase, pUL89 et pUL56, et la phosphotransférase pUL97 ainsi que sa protéine accessoire, pUL27.Grâce à l'utilisation de la résonance magnétique nucléaire, des spectroscopies UV/Visible et de dichroïsme circulaire, et de la modélisation par homologie, nous avons collecté des informations structurales relatives aux cibles des dérivés ribobenzimidazolés. Nous avons pu proposer des hypothèses sur le mode d'action de ces molécules. Par ailleurs, la découverte d'un domaine capable d'interagir avec l'ADN viral double brin de CMVH a été décrite.**

---

**Structural studies of new therapeutic targets of HCMV revealed by benzimidazoles derivatives ribonucleoside**

**The Human Cytomegalovirus is a virus affecting all populations regardless of their socio-professional environment. Harmless in healthy subjects, the damage caused by the virus in the host proved disastrous when it becomes immunocompromised. Thus, people with HIV, people who have undergone organ transplants or infants are particularly susceptible to HCMV. The first inhibitors formulated for the viral polymerase have been ineffective due to the observation of multiple resistance mutations. A new class of molecule, benzimidazoles ribonucleoside derivatives, for stages of maturation and encapsidation, has emerged for the terminase complex proteins, pUL89 and pUL56, and the phosphotransferase pUL97 and its accessory protein, pUL27. Through the use of nuclear magnetic resonance, UV / Visible spectroscopy and circular dichroism, and homology modelling, we collected structural information on target of derivatives ribobenzimidazoles. We have proposed hypotheses about the mode of action of these molecules. Moreover, the discovery of a domain capable of interacting with the viral double-stranded DNA of HCMV has been described.**

---

**Mots-clefs: Cytomégalovirus Humain, terminase, pUL89, pUL56, dérivés ribobenzimidazolés, RMN, modélisation moléculaire**

**Adresse électronique: anthony.couvreux@wanadoo.fr**