



HAL
open science

Contribution au projet de génomique structurale du virus d'Epstein-Barr : l'Uracile-ADN Glycosylase et les enzymes du métabolisme des acides nucléiques

Thibault Geoui

► To cite this version:

Thibault Geoui. Contribution au projet de génomique structurale du virus d'Epstein-Barr : l'Uracile-ADN Glycosylase et les enzymes du métabolisme des acides nucléiques. Biochimie [q-bio.BM]. Université Joseph-Fourier - Grenoble I, 2006. Français. NNT: . tel-00160082

HAL Id: tel-00160082

<https://theses.hal.science/tel-00160082>

Submitted on 4 Jul 2007

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.

THESE - Présentée par :

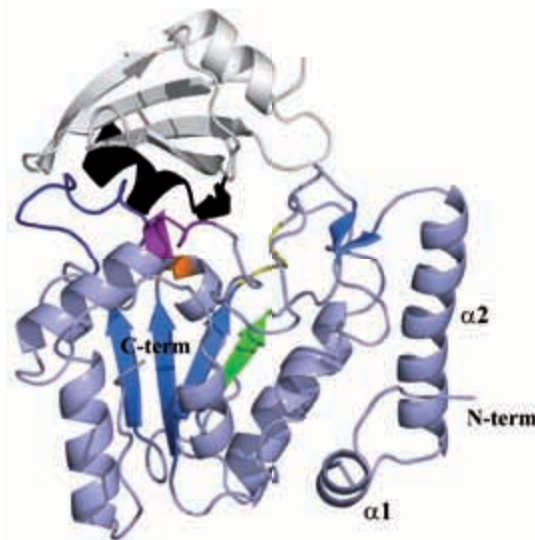
Thibault Géoui

Pour obtenir le titre de :

Docteur de l'Université Joseph Fourier - Grenoble 1

Discipline : Biologie Structurale & Nanobiologie

Contribution au projet de génomique structurale du
virus d'Epstein-Barr :
l'Uracile-ADN Glycosylase et les enzymes du
métabolisme des acides nucléiques



Soutenance publique le lundi 18 décembre 2006 devant le jury composé de :

Président : - Pr. Alain Favier
Rapporteurs : - Dr. Jonathan M. Grimes
- Dr. Evelyne Manet
Examineurs : - Dr. Patrice Morand

Thèse préparée au sein de l'Institut de Virologie Moléculaire et Structurale
sous la direction du Pr. Wilhelm Pascal Burmeister.

En couverture :

Structure cristallographique de l'Uracile-ADN Glycosylase d'EBV en complexe avec l'Ugi, sa protéine inhibitrice.

Remerciements ...

Il y a quatre ans, alors que j'étais encore étudiant en Biochimie à Lyon, j'envoyai un mail au **Dr. Steven Cusack**, directeur de l'EMBL-Grenoble ... c'était une bouteille à la mer qui contenait une description précise de l'image que je me faisais du monde de la recherche et du désir irréprensible d'en faire partie.

Peu de temps après je recevais un appel du **Pr. Rob Ruigrok**, alors chef de groupe à l'EMBL pour un entretien. C'est de cette façon que j'ai rencontré le **Pr. Wim Burmeister** qui deviendrait mon superviseur durant mon DEA et ma Thèse.

L'histoire n'est pas aussi simple, et pour être tout à fait honnête, elle débute bien plus tôt, alors que j'étais encore un jeune lycéen en 1^{ère}, dans ma Haute-Savoie natale. C'est à cette époque que mon père me présenta le **Dr. Joseph Curran**, chez qui je passerai quelques jours à l'université de Genève ... ce fut le premier à m'initier avec patience et passion à la recherche. Là où les histoires se croisent et les coïncidences sont heureuses, ce fut le jour où, de passage à Grenoble pour l'entretien mentionné quelques lignes plus haut, je me rendis compte que le Pr. Ruigrok et le Dr. Curran avaient collaboré quelques années plus tôt ... était-ce un hasard ou alors ce qu'on appelle communément le destin ? Je ne me prononcerai pas ... Le Dr. Joseph Curran me reçut une nouvelle fois chez lui et après une longue discussion je pris la décision de travailler avec le Pr. Wim Burmeister.

Je tiens à remercier le Dr. Joseph Curran, de m'avoir guidé durant mes années à l'université, de m'avoir reçu chez lui et de m'avoir montré le côté le plus humain et désintéressé de la recherche.

Merci au Dr. Steven Cusack pour avoir fait circuler mon message dans son institut et plus tard m'avoir permis de travailler pendant près de deux ans et demi dans son laboratoire.

Merci au Pr. Ruigrok, de m'avoir accueilli au sein de l'institut qu'il dirige : l'IVMS.

Merci au Pr. Wim Burmeister de m'avoir accueilli dans son équipe, de m'avoir accordé sa confiance et d'avoir supporté mon mauvais caractère et mon goût prononcé pour la contradiction.

Je tiens à remercier le **Pr. Alain Favier** qui a présidé mon jury de Thèse.

Les **Dr. Evelyne Manet** et **Dr. Jonathan M. Grimes** qui furent les rapporteurs de mon travail. Enfin, le **Dr. Patrice Morand**, que j'ai côtoyé pendant quatre ans au sein de l'IVMS. J'ai souhaité qu'il fasse partie de mon jury car, c'est non seulement un expert reconnu du virus d'Epstein-Barr, c'est aussi un des rares médecins à comprendre quelque chose à la biologie structurale. C'est surtout une personne que j'ai beaucoup appréciée et ce fut un plaisir de discuter avec lui de sujets scientifiques et surtout de tout le reste ...

Quand j'ai commencé mon DEA, l'équipe du Pr. Burmeister était composé ... de lui-même et du **Dr. Marlyse Buisson**. C'était une petite équipe ... mais quelle équipe ! En effet, si il y a une personne que je me dois de remercier, c'est bien le Dr. Marlyse Buisson. Marlyse, je te dois tout ou presque, sans toi, cette thèse n'existerait certainement pas, et d'ailleurs, la majorité des résultats obtenus dans l'équipe ne seraient pas là. Ton importance ne s'arrête pas seulement à ces détails matériels, ce fut un réel plaisir non seulement de travailler avec toi (c'est une évidence) mais surtout d'apprendre à te connaître ... ce qui, à mon avis nécessite bien plus longtemps que la durée de cette thèse. Compte sur moi pour ne pas t'oublier, bien plus qu'une collègue, je te considère comme une amie.

Plus tard, le **Dr. Nicolas Tarbouriech** arriva dans l'équipe, très rapidement il devint incontournable, pour ne pas dire un pilier du laboratoire. Merci à toi d'avoir fait preuve de patience aussi bien dans le laboratoire quand une expérience ne se déroulait pas de la façon dont je l'avais prévue (souvent ...) que devant un ordinateur, alors que les cartes de densité ne me semblaient pas encore très parlantes.

Je tiens également à remercier le **Dr. Lucie Rivail** et **Lucy (Nancy) Freeman**, pour les nombreux bons moments que nous avons passé ensemble ... la liste serait trop longue et les détails n'ont certainement pas leur place dans un document aussi officiel, ainsi je n'en dirais pas plus.

Les expériences de cristallographie menées par notre équipe, n'auraient certainement jamais abouti sans le support, à différentes époques et sur différents projets, des **Dr. Hassan Belrhali** et **Dr. David Flot**.

L'EMBL et plus tard, l'IVMS furent des lieux de rencontre où de nombreuses amitiés naquirent et j'en suis sûr perdureront après mon départ :

Je pense au **Dr. Carlos Fernandez-Tornero** ; il m'est difficile de résumer en quelques phrases tout ce qui nous lie, à commencer par notre amour du cinéma qui aboutira peut-être, un jour (je l'espère) sur une réalisation commune. Bien au-delà de ça, Carlos, tu es et tu resteras un ami et j'espère (j'en suis sûr !) que ni le temps et ni la distance n'affecteront cette relation.

La liste des personnes auxquelles je pense est très longue et j'espère ne pas en oublier :

Le **Dr. Aurélie Albertini** avec qui j'ai commencé et avec laquelle je finis ma thèse...

Céline ma « colocataire de bureau », assez bavarde surtout avec son ordinateur et toujours à la pointe de la mode et de la technologie.

Les anciens avec qui j'ai commencé : **Sarah, Hatice, Manos, Vassily, Guy.**

Et ceux et celles qui sont arrivés par la suite : **Francine, Majida, Jeanne, Sophie** (Miss Gizmodo), **Cedric, Estella, Cyril.**

La vie au laboratoire n'aurait certainement pas été ce qu'elle fut sans le personnel qui aura toujours été présent, efficace et amical :

Virginie, Benjamin, Fred, Aymeric, Rémi, Mary-Jane, Annie, Françoise, Jocelyne ...

Heureusement la « période Grenobloise » ne se limitait pas au laboratoire et mes quatre années furent riches en expériences et en rencontres :

A l'université entre les cours et les associations :

JB dès mon DEA me permit de débiter ma vie associative au sein de ³⁸Globule.

Je tiens également à remercier le CIES de Grenoble, notamment **Michelle Vuillet** et **Régine Herbelles.**

Merci au **Pr. Michel Robert-Nicoud**, qui fut non seulement un de mes professeurs à l'université Joseph Fourier, le directeur de l'Ecole Doctorale de chimie et Science du Vivant et également un ami avec lequel j'ai passé de bons moments.

Dès le départ, Grenoble fut une ville particulièrement accueillante, et rapidement je rencontrais de vrais amis, qui le jour de ma soutenance me le rendraient bien :

Pauline avec laquelle je suis arrivé de Lyon à Grenoble ; ne t'en fais pas la distance n'est pas un frein à l'amitié !

David (mon pote de DEA et mon colloque pour quelques temps à qui je dois beaucoup !), **Greg, Julien, Nathalia, Roya, Carole** et **Fabien** avec lesquels je ne compte plus les soirées, les voyages ... j'espère que ça va continuer ainsi !

Je tiens à remercier **ma famille** :

Ma tante **Marie-Claire** qui accepta avec patience de corriger l'orthographe perfectible de ma Thèse.

Mes grands mères, **Alice** et **Blanche.**

Et enfin, mes parents **Joseph** et **Marinette**, sans qui (c'est une évidence) je ne serais pas là ... et je ne serais pas qui je suis. Ces quelques mots représentent si peu de chose pour leur exprimer toute ma gratitude.

Enfin, **Mila** avec qui je me prépare pour une nouvelle vie ...

Table des matières

I	Introduction Générale	11
1	Burkitt et la découverte d'EBV	13
1.1	L'enfance et les études de Denis Burkitt	13
1.2	L'Afrique et les réalisations de Denis Burkitt	16
2	Le virus d'Epstein-Barr	23
2.1	Aspect médical de l'étude d'ebv	23
2.1.1	Identification du virus	23
2.1.2	EBV et pathologies Humaines	25
2.2	Biologie d'EBV	31
2.2.1	EBV est un herpes-virus	31
2.2.2	Structure du virion	32
2.2.3	Réplication des herpes-virus	33
3	Contexte de l'étude et but du projet	45
II	Génomique Structurale	47
4	Sélections des cibles	49
4.1	Biologie structurale et conception rationnelle de médicaments	49
4.1.1	Le projet SPINE	52
4.2	Acides nucléiques : métabolisme/réparation/réplication	52

4.2.1	Une régulation complexe	52
4.2.2	Principe de la chimiothérapie anti-cancéreuse et anti-virale	55
4.2.3	L'inhibition de la voie de synthèse du dTMP est un "classique" de la chimiothérapie	58
5	Matériel et Méthodes	61
5.1	Annotation du génome	61
5.1.1	Conception d'une base de données des protéines d'EBV	61
5.2	Clonage et expression de la TK et de la RNR	63
6	Article : Génomique structurale d'EBV	65
6.1	Résumé	65
6.2	Article 1	66
7	Résultats supplémentaires	77
7.1	Recherche d'une Thymidylate Synthase (TS) virale	77
7.2	Protéines virales : <i>E.coli</i> n'est pas le système d'expression idéal	78
III	Structure de l'Uracile-ADN Glycosylase d'EBV	81
8	Introduction : Mécanismes de réparation de l'ADN	83
8.1	L'organisme est le sujet d'agressions constantes	83
8.2	L'organisme a développé des mécanismes de défenses	84
8.3	Hiérarchisation de la réponse à un dommage	86
8.4	Mécanisme d'Excision et de Réparation des Bases (BER)	87
8.4.1	Les ADN Glycosylases, une super-famille avec une grande diversité de substrats	89
8.4.2	L'Uracile-ADN Glycosylase (UNG) catalyse l'élimination de l'uracile	94
8.5	BKRF3 est une Uracile-ADN Glycosylase codée par EBV	97
8.5.1	Pourquoi EBV exprime-t-il des enzymes présente chez l'hôte?	97
8.5.2	L'Uracile-ADN Glycosylase d'EBV est une cible potentielle pour le développement de nouveaux médicaments	98

9	Matériel et Méthodes	99
9.1	Conception de la construction UNG Δ 24	99
9.1.1	Protéolyse limitée et séquençage N-terminal	99
9.2	Caractérisation de l'échantillon	100
9.2.1	Diffusion dynamique de lumière (DLS)	100
9.2.2	Dichroïsme circulaire	102
9.3	Principe du test enzymatique	106
9.4	Résolution de Structure par Remplacement Moléculaire	106
9.4.1	Principe	106
9.4.2	Les étapes dans la résolution d'une structure	108
10	Article : Structure de l'UNG d'EBV	115
10.1	Résumé	115
10.2	Article 2	116
11	Résultats supplémentaires	133
11.1	Définition et obtention du domaine UNG Δ 24	133
11.2	Vérification du repliement de la protéine par CD	135
11.3	Influence du β -mercaptoéthanol et de l'Ugi sur l'homogénéité	135
11.4	Information structurale supplémentaire	136
11.4.1	Diagramme de Ramachandran	136
11.4.2	Empilement cristallin	138
IV	Etude structurale préliminaire de l'Alcaline Exonucléase (BGLF5) du virus d'Epstein-Barr	143
12	Introduction : BGLF5, un défi technique	145
12.1	L'Alcaline Exonucléase d'EBV	145
12.2	Une étude structurale compliquée	147

13 Matériel et Méthodes	151
13.1 Méthodes Alternatives de Cristallogénèse	151
13.1.1 Diffusion de vapeur	151
13.1.2 Croissance cristalline en capillaire	152
13.2 Cristallographie en “Conditions Difficiles”	153
13.2.1 La ligne de lumière micro-focus	154
13.2.2 Mélange de jeux de données	157
14 Résultats et perspectives sur l’Alcaline Exonucléase	159
14.1 Propriétés des cristaux d’Alcaline Exonucléase	159
14.1.1 Les cristaux de l’Alcaline Exonucléase	159
14.2 Analyse structurale préliminaire	160
14.2.1 Microscopie électronique	160
14.2.2 Jeu de donné natif	161
14.2.3 Self-Rotation Function	161
14.3 Conclusion et perspectives	161
V Conclusion et Perspectives	167
15 Conclusions et perspectives	169
15.1 Génomique structurale d’EBV	169
15.1.1 Conclusions sur le travail mené	169
15.1.2 Evolution de la génomique structurale d’EBV	170
15.2 L’Uracile-ADN Glycosylase : conclusions et perspectives	171
15.2.1 Perspectives Fonctionnelles	172
15.2.2 Perspectives Structurales	173
15.3 Résolution de la structure de l’Alcaline Exonucléase	174
15.4 Conclusion sur l’étude structurale des protéines d’EBV	175

VI Annexes **177****16 Informations complémentaires** **179**

16.0.1 Homologie / identité /similarité 179

16.0.2 Base de données généraliste 179

16.0.3 Base de données de protéines virale 180

16.0.4 Base de données de protéines impliquées dans la réparation de l'ADN 180

16.0.5 Prédiction de structure secondaire 180

16.0.6 Symétrie cristalline 181

Abréviations :

5-FU	=	5-Fluorodeoxyuridine
AA	=	Acide-Aminé
Å	=	Angstrom
BER	=	Base Excision Repair
BL	=	Lymphome de Burkitt
CD	=	Cluster de Différentiation (immunologie)
CD	=	Dichroïsme Circulaire (physique)
CMH	=	Complexe Majeur d'Histocompatibilité
CTL	=	Lymphocyte T cytotoxique
DHF, THF	=	Di, Tetra HydroFolate
DHFR	=	DiHydroFolate Réductase
DLS	=	Diffusion Dynamique de Lumière
dTMP	=	désoxyribo-Thymidine Monophosphate
dUMP	=	désoxyribo-uridine Monophosphate
EBV	=	Virus d'Epstein-Barr
GP	=	Glycoprotéine
HCMV	=	Cytomégalovirus Humain
HHV	=	Herpes Virus Humain
HIV	=	Virus de l'Immunodéficience Humaine
HSV	=	Herpes Simplex Virus
KSHV	=	Kaposi Sarcoma Herpes Virus
LB	=	Lymphocyte B
LCL	=	Lymphocyte infecté de façon Latente
MNI	=	Mononucléose Infectieuse
MTX	=	Méthotrexate
NPC	=	Carcinome du Nasopharynx
OEM	=	Onde ElectroMagnétique
RNR	=	Ribonucléotide Réductase
SIDA	=	Syndrome d'Immunodéficience Acquise
TF	=	Transformée de Fourier
TK	=	Thymidine Kinase
TS	=	Thymidylate Synthase
UNG	=	Uracile-ADN Glycosylase
VZV	=	Varicella Zoster Virus

Première partie

Introduction Générale

Chapitre 1

Denis Parsons Burkitt et la découverte du virus d'Epstein-Barr¹

1.1 L'enfance et les études de Denis Burkitt

ENNISKILLEN, district de Fermanagh, le 28 février 1911. C'est dans cette petite bourgade d'Irlande du Nord que les époux Burkitt attendaient la naissance de leur premier enfant : Denis Parsons Burkitt. Denis Burkitt fut scolarisé à l'école protestante de *Portora Royal*. A la même époque, Samuel Beckett était également élève dans cette institution, se sont-ils rencontrés ? Les archives n'en font pas mention. Mais lorsque l'on consulte ces mêmes archives avec attention, il est un fait que l'on ne peut nier, et qui se reflète dans l'absence du nom de Denis Burkitt sur la page retraçant l'histoire, "glorieuse" de cette institution nord irlandaise, Denis Burkitt était humble, son humilité était telle qu'elle le poussait à "accuser" le sort ou la conjonction heureuse de certains événements pour justifier des grandes découvertes dont il fut crédité, et qu'aucun ne se risquerait de lui discuter.

Cette personnalité effacée, ne facilita certainement pas l'enfance de Denis Burkitt, l'aîné d'une famille de deux enfants. Le début de sa scolarité ne fut pas une réussite. Denis Burkitt est décrit comme un étudiant médiocre, son frère, Robin, fait la fierté de sa famille, il surpasse Denis que ce soit dans le cadre des activités sportives aussi bien que par sa popularité (Nelson, 1995). Si les biographes

¹Les éléments biographiques de cette partie proviennent des références suivantes : Nelson C & Temple N (1994), Nelson (1995), Lee (1996), Jay (1998).

conservent un élément de cette période de la vie de Denis Burkitt, c'est certainement cet accident fâcheux, ce petit drame qui aura des répercussions sur sa vie entière : alors qu'il est en route pour l'école, le jeune garçon se trouve pris dans une bagarre, il en sort avec de graves blessures à un œil ... dont il perdra l'usage assez rapidement. Il a alors 11 ans.

A l'âge où Denis Burkitt doit choisir ce qu'il fera de son avenir ... et bien justement ... il ne sait pas ce qu'il veut faire. Son père James Burkitt, est ingénieur en travaux publics dont la carrière est brillante. A cette époque il est aussi connu pour ses travaux en ornithologie. James Burkitt fut en effet parmi les premiers à marquer, d'une bague métallique, les oiseaux, afin de définir leur territoire. Personne ne sait si Denis Burkitt souffrait d'un complexe d'infériorité résultant d'une hérédité aussi bien que d'une fraternité qui le dépassaient, en apparence, en tout point. C'est ainsi que, sur les conseils de son père il entra, en 1929 à l'âge de 18 ans, au *Trinity College* de Dublin pour entamer des études d'ingénieur en travaux publics. Mais Denis Burkitt est un étudiant sans passions et encore une fois, il se révèle un élève médiocre. C'est un fait, son père, aussi bien que ses professeurs en sont conscients. Plus tard, il retrouvera, parmi les documents conservés par son père, une lettre envoyée par son tuteur de l'époque, Luce A.A (Luce, 1929), dans laquelle celui-ci met en garde James Burkitt : “[Votre fils] ne réussira certainement jamais à obtenir le moindre diplôme [de notre] Université ... et, visiblement ... il va falloir penser à sa réorientation”. C'est à cette époque que se produit une rencontre qui fera date dans la vie de Burkitt.

Il est facile de juger à posteriori si un événement est important ou s'il ne l'est pas, mais dès à présent, nous pouvons noter un élément du caractère de Denis Burkitt qui sera certainement la clef de son succès : son *opportunité* et sa *clairvoyance* face aux événements qui seront décisifs aussi bien dans sa vie personnelle que professionnelle. Pasteur ne disait-il pas que “*le hasard ne favorise que les esprits préparés*” ?

Le hasard, que certains seraient tentés d'appeler destin ... mais je ne me risquerai pas sur ce terrain sachant que ce n'est pas l'objet de ce travail ... et bien le hasard se manifesta à Denis Burkitt un an après son entrée au *Trinity College*. Un étudiant plus âgé que lui, le convia à une réunion d'une communauté évangélique. Denis Burkitt est très croyant, et cette rencontre est providentielle. Peu de



FIG. 1.1: Olive et Denis Burkitt.

temps après, il entrera dans cette communauté connue sous le nom de *No 40*, en raison du numéro de la salle où ses membres avaient l'habitude de se réunir. Denis Burkitt décide de *“vouer sa vie à Jésus Christ ... ce n'était que peu de temps avant que des prières incessantes à la recherche de mon chemin ne me conduisent à la conviction que je serais amené à devenir Docteur”* (Burkitt, 1994).

Alors étudiant peut concerné par le destin auquel sa famille le destinait, il change radicalement d'orientation et aborde sa carrière médicale, en 1930, avec *“un enthousiasme sans limite [...] mon Q.I., que je n'ai jamais considéré comme élevé, n'a pas changé, à l'opposé de mon attitude et de ma confiance”* (Burkitt, 1994).

En 1935, il obtient, après de nombreux prix pour excellence académique, son titre de *MB* de l'Université de Dublin. J.N.P Moore, un de ses camarades de classe se rappelle cette période :

“Il aurait été normal de lui proposer un des postes convoités de praticien et de chirurgien au centre hospitalo-universitaire. Le poste fut offert à un élève moins bien classé mais plus populaire, un international de rugby qui permit à l'hôpital de remporter la coupe. Burkitt était très déçu particulièrement à cette période de sa carrière, mais il accepta cette situation avec beaucoup de charisme et de dignité.”

(Coakley, 1992). En 1938, alors jeune chirurgien diplômé du *Royal College of Surgeon* d'Edinburgh, Burkitt ne sait quel sens donner à sa vie, il décide de partir pendant 5 mois pour exercer la chirurgie sur un bateau en partance pour la Chine. A la suite de ce voyage, il décida qu'il consacrerait alors sa vie aux moins favorisés, mais il doit acquérir plus d'expériences. Il travaille comme chirurgien au *Plymouth Hospital*, où il rencontrera Olive Mary Rogers (Figure 1.1), alors étudiante infirmière, qui



FIG. 1.2: *Microfilaria* - Agent responsable de l'hydrocèle.

trois ans plus tard allait devenir sa femme (Nelson, 1995).

1.2 L'Afrique et les réalisations de Denis Burkitt

Il se porte volontaire pour le *service médical colonial d'Afrique de l'Ouest*. Mais qui voudrait d'un chirurgien borgne ? Sa demande se voit rejetée, il n'est convié à aucune audition. C'est à cette époque que la seconde guerre mondiale éclate. Le besoin en personnel est tel, que toute compétence est bonne à exploiter, et, les autorités chargées du recrutement qui, peu de temps avant ne voulait pas de Denis Burkitt, furent moins "regardantes" en cette période de crise, sur le profil des candidats recrutés, ils se contentèrent donc d'un chirurgien borgne. Burkitt sera chirurgien pour l'armée pendant 5 ans. C'est durant cette période, faste, qu'il épouse Olive Rogers et découvre enfin l'Afrique. "*J'avais le sentiment que c'était à cet endroit que j'étais appelé*" (Burkitt, 1994), dit Burkitt des années après. A la fin de la guerre il obtient un poste en Ouganda.

En 1946, Denis Burkitt est nommé Officier Médical dans le district de Lango. "*Quand je suis enfin arrivé en Afrique, Dieu, dans sa grande bonté, me permit de voir avec mon seul et unique œil, ce que mes prédécesseurs n'avaient pu voir avec leur deux yeux*" (Jay, 1998). L'activité de Denis Burkitt est débordante, on le comprend aisément, il est responsable de la santé de 170.000 personnes réparties sur 4.400 Km². Dès cette époque, il entreprend ses premières études épidémiologiques. Burkitt opère

jusqu'à 6 à 8 cas d'hydrocèle² par semaine. Burkitt se rend compte que les cas d'hydrocèle sont plus répandus à l'Est avec 30% des individus de sexe masculin atteints, contre seulement 1% à l'Ouest. Il identifia le *microfilaria* comme agent potentiellement responsable de la maladie (Figure 1.2 page 16, Burkitt (1951), Burkitt (1947)).

Après 18 mois passés dans son district quelque peu éloigné, Burkitt est nommé enseignant en chirurgie au *Makerere College Medical School* et également chirurgien à l'hôpital de Mulago dans le district de Kampala.

En 1957, Hugh Trowell, un collègue de Burkitt lui adresse un patient. C'était un jeune garçon de 5 ans, "*Son visage était énorme et présentait d'étranges lésions sur les parties supérieures et inférieures de la mâchoire. Je n'ai jamais rien vu de tel*". Plus tard, on lui adressa un enfant avec des lésions similaires mais qui présentait en plus des grosseurs au niveau de l'abdomen. "*Une curiosité ne se produit qu'une fois, deux cas similaires indiquent plus qu'une curiosité*" (Jay, 1998). Les deux premiers cas que Burkitt observa à Mulago ne furent pas diagnostiqués, mais par la suite, le troisième patient qu'il examina à Jinga le fut au regard "*de photographies des cas précédents et représentait une entité clinique définie*" (Burkitt (1958b), Figure 1.3 page 18).

L'observation de tumeurs au niveau de la mâchoire chez des enfants africains n'était pas une nouveauté, Albert Cook, un missionnaire présent en Afrique au début du xx^e siècle, avait répertorié et dessiné précisément de telles tumeurs dans ses carnets. Burkitt partit à la recherche de preuves dans des rapports d'autopsie et découvrit que les enfants présentant ces tumeurs massives au niveau de la mâchoire étaient également atteints de tumeurs disséminées à d'autres endroits du corps.

Burkitt présenta les résultats d'une étude qu'il effectua sur 41 enfants, au cours d'une conférence auprès de l'Association Est Africaine des Chirurgiens (Burkitt, 1958b). Burkitt écrit des rapports précis et tenta de classer ses observations concernant ses patients, mais il se trouva dans l'obligation d'avouer qu'il n'avait "*aucune idée de ce qui peut en être la cause*" (Burkitt, 1957). Il conclut sa conférence ainsi :

" J'ai limité mes remarques à la classification de faits observés et à la présentation d'un problème auquel je n'apporte pas de solutions. Ma contribution au soulagement de ces enfants pourrait appa-

²C'est un gonflement des testicules qui se remplissent d'une grande quantité de liquide, cette maladie est souvent d'origine parasitaire.



FIG. 1.3: Jeune fille de 5 ans atteinte par un lymphome de Burkitt.

raître comme faible voire nulle. La contribution la plus significative proviendra de ceux qui sauront interpréter et donner une signification à ces faits, et qui un jour, espérons, découvriront et peut-être élimineront la cause du développement de cette tumeur” (Burkitt, 1958b).

Burkitt ne le savait pas encore, mais, non seulement il allait contribuer à l’identification de la cause de cette tumeur mais en plus, il allait permettre son traitement.

En 1958, les travaux de Burkitt furent publiés dans le *British Journal of Surgery* (Burkitt, 1958a), mais il n’eurent que peu d’écho.

Puisque personne ne semblait s’intéresser au travail conduit par “*un petit chirurgien de brousse*”, Burkitt décida de mener l’enquête lui-même. Deux bourses gouvernementales de 10 £ et 15 £ lui permirent d’imprimer et d’envoyer des questionnaires à de nombreux hôpitaux et missions africaines. A partir de 1960, il avait réuni des informations concernant 120 cas, provenant de l’Afrique entière. Les réponses lui permirent de tracer la carte de ce qu’il appellera la “*ceinture des lymphomes*” (“*The*



FIG. 1.4: La *ceinture des lymphomes*. Le parcours grisé représente le safari de 10.000 miles que Burkitt effectuera pour collecter ses données (Glemser, 1971).

lymphoma belt”, Figure 1.4, Glemser (1971)).

La ceinture des lymphomes couvrait une bande allant de 10° au Nord jusqu’à 10° au Sud de l’équateur, avec une queue parcourant une partie de la côte Est (Jay, 1998). En 1959, O’Connor et Davies identifient la tumeur comme étant d’origine lymphatique (O’Conor & Davies (1960), Burkitt & O’Conor (1961)). Les articles de Burkitt et de ses collègues décrivaient un syndrome clinique qui liait une tumeur située dans une grande variété de sites anatomiques : mâchoires, reins, surrénales, foie, thyroïde, cœur, estomac et ovaires - avec la même apparence histologique, la même incidence en fonction de l’âge et la même distribution géographique. Ce qui était précédemment un “*fourre-tout*” de cas cliniques était en fait le cancer le plus commun chez les enfants en Ouganda. Les premiers pas vers une thérapie étaient entrepris. La ceinture des lymphomes suggérait un facteur étiologique - mais quel était-il ? Davies, le collègue de Burkitt, suggéra alors un virus (Burkitt, 1961).

Le 22 mars 1961, Denis Burkitt, à l’occasion d’un séjour à Londres, donna un séminaire à l’hôpital de Middlesex intitulé “The commonest children’s cancer in tropical Africa. A hitherto unrecognized syndrome”.

Tony Epstein (Figure 1.5 page 20), un virologue qui faisait partie des auditeurs, s’intéressait au lien entre virus et cancer. Peu de temps après, Burkitt accepta de lui faire parvenir des échantillons



FIG. 1.5: M.A. Epstein, B.G. Achong et Y.M. Barr en 1964, (Robertson, 2005).

congelés de tumeurs.

Alors que Denis Burkitt était en plein safari, Epstein et ses collègues, Bert Achong et Yvonne Barr, une jeune étudiante du Trinity College, ne chômaient pas. En 1964, dans le numéro du 28 mars de la revue *The Lancet*, ils allaient publier leur découverte concernant l'observation de particules virales à partir de cellules de lymphome mises en culture (Epstein *et al.*, 1964).

Alors que la tumeur identifiée par Burkitt allait être nommée lymphome de Burkitt, le virus que l'équipe anglaise venait de mettre en évidence prendrait le nom de virus d'Epstein-Barr. Pourquoi Bert Achong a-t-il été évincé? Cela restera un mystère. Dans ses lettres avec McDonnell, Burkitt écrira :

“[Le virus aurait dû être nommé] Epstein-Achong et non Epstein-Barr puisque Yvonne Barr [...] ne travaillait que de façon temporaire dans le laboratoire alors que Bert Achong [...] allait travailler avec Tony (Epstein) pendant presque 20 ans (Burkitt, 1992).

La façon dont le virus a été identifié sera discutée dans la suite de ce travail. Pour le moment, retrouvons Denis Burkitt qui décidément ne s'est pas beaucoup reposé durant sa “période africaine”.

Burkitt allait définir les contours de la ceinture des lymphomes au cours d'un safari de plus de 10.000 miles qui se sera une véritable “biopsie géographique” (Burkitt, 1964). *“ C'est certainement le safari au travers de l'Afrique le plus sûr qui n'ait jamais été. Nous sommes trois médecins, chacun avec notre stock de médicaments voyageant d'un hôpital à un autre.”* (Nelson C & Temple N (1994),

Figure 1.4 page 19).

Burkitt reçut un accueil favorable de la part de ses collègues médecins de campagne *“Ils comprenaient que je pouvais me mettre à leur place, ils étaient heureux de partager leurs expériences, et parfois, je pouvais les aider à résoudre leurs propres problèmes ... cette approche personnelle était basique mais fondamentale pour ma collecte d’informations épidémiologiques”* (Burkitt, 1993).

La description précise de l’incidence et l’étiologie probable du lymphome était d’un intérêt scientifique certain mais qu’en était-il des patients ? Burkitt était consterné par le taux de mortalité élevé résultant de la maladie. Dès 1961, Burkitt fit part d’une rémission rapide mais néanmoins temporaire d’un patient traité par radiothérapie en Inde (Burkitt & O’Conor, 1961), la radiothérapie n’étant pas disponible en Ouganda. Il écrivit alors aux laboratoires pharmaceutiques qui produisaient des agents cytotoxiques, pour leurs expliquer que l’accès aux radiothérapies était compliqué en Afrique et que ce terrain leur permettrait de tester leurs chimiothérapies sur des tumeurs jusqu’alors non traitées. Les laboratoires lui envoyèrent gratuitement des médicaments . Burkitt obtint de très bons résultats avec le méthotrexate, la cyclophosphamide et la vincristine. Le lymphome de Burkitt fut une des premières tumeurs traitées par chimiothérapie uniquement (Burkitt *et al.*, 1965).

Burkitt démissionna de son poste de chirurgien en 1964 mais resta à Kampala en tant que personnel délocalisé du Medical Research Council (MRC). En janvier 1966, à l’occasion du congrès de l’association internationale de lutte contre le cancer, il présenta 12 patients guéris avec les photographies de leurs tumeurs avant thérapies. Sir Alexander Haddow, le président de cette association déclara, à l’occasion de la conférence de clôture de la conférence, que c’était la chose la plus remarquable qu’il ait jamais vue. Depuis cette époque, les caractéristiques du lymphome de Burkitt à ses différentes étapes ainsi que la chimiothérapie ont considérablement progressé. Une étude en 1991 montrait que le taux de survie est de 75% trois ans après le traitement (Schwenn *et al.*, 1991). La ceinture des lymphomes, le virus et les moustiques, en revanche, n’ont pas été éradiqués.

Burkitt quitta l’Afrique en 1966 et débuta une nouvelle carrière au MRC de Londres. Suite à sa rencontre avec Peter Cleave, il s’intéressa à l’influence de l’alimentation sur la santé. Il montra qu’un régime alimentaire riche en fibres, tel que celui pratiqué en Afrique, permet de prévenir certaines maladies très fréquentes en Occident et pratiquement absentes en Afrique³ (Burkitt, 1971). Burkitt

³Obésité, diabète, certain cancers, maladies cardio-vasculaire ...

recevra de nombreux prix pour récompenser l'ensemble de sa carrière. Il mourut à l'âge de 82 ans, le 23 mars 1993 .

40 ans de recherches ...

La quête qui mena Burkitt de la découverte d'un lien entre différentes formes d'un même cancer et l'identification par Epstein et ses collègues du virus responsable de ce cancer est une épopée fascinante, mais ce n'est que le début de plus de 40 ans de recherches.

Chapitre 2

Le virus d'Epstein-Barr

2.1 Aspect médical de l'étude d'ebv

LE but avoué par l'auteur en incluant cette partie, n'est certes pas de faire une revue exhaustive de l'"état de l'art" en terme de biologie d'EBV. Je me contenterai d'aborder brièvement les concepts nécessaires à la mise en perspective du travail effectué au cours de cette thèse. L'aspect très fondamental, physique, parfois même éloigné de la réalité médicale que devrait sous-tendre tout travail lié de près ou de loin à l'étude d'un agent pathogène humain, doit être replacé dans son contexte biologique. Pourquoi avons-nous abordé l'étude d'EBV de cette façon ? Quelle sera la conséquence des réponses que nous apportons ? Ce dernier propos, aussi déterministe qu'il soit, laisse sa part d'aléatoire à toute découverte, mise à la disposition de la communauté scientifique. Personne ne peut affirmer ce qu'il en adviendra, néanmoins, et je tiens une nouvelle fois à insister sur ce point, la cristallographie ... méthode d'étude du vivant, sur laquelle je reviendrai en détail par la suite, n'est qu'une technique qui doit être "asservie" à la discipline à l'intérieur de laquelle elle est utilisée (la Biologie) et qui doit apporter de nouvelles réponses et susciter d'autres interrogations.

2.1.1 Identification du virus

Suite à la conférence de Burkitt au *Middlesex Hospital*, Anthony Epstein était convaincu que le lymphome de Burkitt était d'origine virale. Quelques jours après cette conférence, Anthony Epstein

parvint à un accord avec Burkitt, ce dernier lui enverrait des échantillons de biopsie de patients atteints de lymphome de Burkitt.

Les deux premières années de travail furent désespérantes, aucune des méthodes standards de l'époque ne permirent d'identifier quoi que ce soit. Le sort changea le cours des événements, le jour où Epstein et ses collègues se rendirent compte qu'ils pouvaient cultiver des cellules issues de patients atteints par le lymphome de Burkitt *in-vitro*, c'est-à-dire dans un environnement dépourvu de système immunitaire. La conséquence directe pour le virus était de se multiplier sans subir la "répression" du système immunitaire (Epstein *et al.* (1964), Epstein *et al.* (1965)). L'obtention d'une culture de cellules de lymphome était un véritable défi à cette époque puisque personne n'avait réussi un tel tour de force (Epstein, 2005). Néanmoins, le 3 décembre 1963, un échantillon de biopsie de lymphome de Burkitt arriva de Kampala à l'aéroport de Manchester ... il fallait compter avec la météo anglaise : le brouillard Londonien , ne permit pas au prélèvement d'arriver dans les délais impartis à l'endroit convenu. Quand Epstein récupéra le prélèvement, celui-ci était trouble, comme contaminé par des bactéries. Après observations au microscope, il se rendit compte que la turbidité n'était pas due aux bactéries : les cellules s'étaient détachées de la tumeur et se trouvaient en solution. L'obtention de culture de cellules n'était que la première étape de la longue histoire qui allait mener à l'identification du virus. A cette époque, la démonstration de la présence d'un virus nécessitait la mise en évidence d'anticorps dirigés contre ce dernier ou l'observation de ses effets sur un animal de laboratoire ou sur une culture de cellules. L'utilisation de la microscopie était loin d'être une méthode admise, les images obtenues en microscopie électronique étant considérées comme des artefacts dus aux contraintes expérimentales (Epstein, 2005).

Epstein observa les premiers échantillons en microscopie électronique le 24 février 1964 et nota immédiatement des particules qui ne pouvaient être que virales et dont la forme caractéristique lui rappelait la famille des herpes virus (Epstein (1962a), Epstein (1962b)). Afin de confirmer la véracité de ses résultats, Epstein prit contact avec Werner et Gertrude Henle qu'il avait rencontrés quelques années plus tôt à Philadelphie alors qu'ils travaillaient sur le virus de la grippe. Le couple Henle accepta de collaborer avec Epstein alors que d'autres groupes de virologistes anglais refusèrent la collaboration, jugeant l'approche d'Epstein trop peu "canonique". En 1965, les deux groupes publièrent le résultat d'un travail commun confirmant que le virus étudié était bien un herpes virus humain



FIG. 2.1: Gertrude et Werner Henle ont montré que la Mononucléose-Infectieuse est causée par EBV.

(Epstein *et al.*, 1965). Le virus ne fut désigné par le nom que nous connaissons (Virus d'Epstein-Barr (EBV)) qu'après un certain nombre d'années et sa caractérisation par de nombreux groupes (Henle *et al.*, 1968), jusqu'alors on le désignait comme "*herpes-like*" ou "*herpes-type*".¹

2.1.2 EBV et pathologies Humaines

Les travaux de Gertrude et Werner Henle ... les conséquences de la primo-infection

En 1963, le couple Henle (Figure 2.1) fut contacté par C.Everett Koop, un chirurgien de retour d'une conférence en Afrique. Il connaissait l'intérêt du couple pour les virus entraînant une infection persistante², il leur dit : "*Si vous voulez vraiment travailler sur les tumeurs causées [probablement] par des virus vous devriez travailler sur le lymphome de Burkitt*". la collaboration proposée par Epstein en 1964 fut une coïncidence heureuse ...

Dès 1966, les Henle avaient développé un test d'immunofluorescence pour détecter des antigènes d'EBV en culture cellulaire (Henle & Henle, 1966). Ce test permit de montrer que les patients atteints d'un lymphome de Burkitt (Henle & Henle, 1966) ou d'un carcinome du nasopharynx (NPC, Old *et al.* (1966)) possèdent un titre élevé d'antigènes viraux. Les Henle démontrèrent également que

¹Cette partie est inspirée de Epstein (2005).

²eg. : Virus de l'hépatite.

l'infection par EBV est ubiquitaire et que le virus est très certainement largement répandu dans la population (Henle *et al.*, 1969). En 1968, les Henle démontrèrent qu'EBV est l'agent responsable de la Mononucléose-Infectieuse (MNI) (Henle *et al.*, 1968) : “*La réponse à notre recherche [d'une maladie causée par EBV] apparut de façon inattendue, au beau milieu de notre laboratoire, alors que l'une de nos techniciennes était en arrêt maladie durant le mois d'août 1967. Elle était séronégative [pour EBV] et c'était une donneuse fréquente [plus ou moins volontaire] de leucocytes. [Son sang] était utilisé pour des études de transmission / transformation conduites par Volker Diehl, un chercheur post-doctorant allemand. Notre technicienne reprit le travail 5 jours plus tard. [Mais visiblement elle n'allait pas bien. Elle souffrait d'une mononucléose-infectieuse (MNI)]. Après de nombreux tests nous avons découvert qu'elle avait développé des anticorps anti-EBV! Ses leucocytes, qui n'avaient jamais survécu en culture, formaient une lignée lymphoblastoïde stable! C'était la preuve qu'EBV était la cause de la MNI*”.³

Transmission

Avant de définir de quelle façon EBV coexiste avec son hôte, il est essentiel de savoir comment survient l'infection ? Les maladies avec lesquelles EBV est associé sont relativement bien connues du grand public, nous reviendrons par la suite sur ce détail, mais dès maintenant arrêtons-nous sur la *Mononucléose-Infectieuse* ... connue sous le nom de *Maladie du Baiser*, maladie qui, nous l'avons vu, fut rapidement identifiée comme une réaction à l'infection tardive par EBV. Il est généralement admis que EBV se transmet par la salive. Certaines études montrent que le virus peut être transmis *in-utero* de la mère à l'enfant (Meyohas *et al.*, 1996). Il a été montré que la transmissions sexuelle est une possibilité marginale par rapport à la transmission salivaire (Thomas *et al.*, 2006).

Pathologies associées à EBV

Un grand nombre de pathologies sont causées ou associées à EBV. Ces maladies résultent de la particularité de ce virus qui, une fois à l'intérieur de l'organisme, persiste tout au long de la vie. Plus de 90% de la population est infectée, mais, en général, l'infection latente reste asymptomatique. Néanmoins, dans certains contextes physiologiques, le sujet infecté va développer des complications.

³Cette partie est inspirée de zur Hausen (2005).

Parmi ces pathologies, il convient de distinguer les maladies du sujet *immuno-compétent* de celles du patient *immuno-déprimé*. Nous allons faire une brève revue des maladies le plus souvent associées à EBV sachant que, pour la plupart, elles sont la conséquence du cycle latent d'EBV sur lequel nous reviendrons dans la partie consacrée au cycle de vie du virus.

La primo-infection est, en général, asymptomatique et a lieu durant la petite enfance, cependant, si le premier contact avec le virus survient chez l'adolescent ou le jeune adulte, le sujet déclenchera, dans 50% des cas, une **Mononucléose-Infectieuse** (MNI). La MNI est caractérisée par de la fièvre, une pharyngite, des polyadénopathies et à une splénomégalie (Nicolas & Seigneurin, 2002). Les patients atteints de MNI souffrent d'une grande fatigue, les symptômes pouvant durer plus de 6 mois.

EBV peut infecter de façon latente et stimuler la prolifération de lymphocytes B, ainsi, le virus est associé à de nombreux cancers (Tableau 2.1 page 30).

Historiquement, le **lymphome de Burkitt** (BL) fut le premier cancer associé à EBV. C'est une tumeur maligne que l'on trouve dans le monde entier mais dont l'incidence est extrêmement élevée en Afrique et en Nouvelle-Guinée, là où le paludisme est particulièrement répandu (Rasti *et al.*, 2005). Dans ce contexte, l'agent responsable du paludisme pourrait stimuler en permanence les lymphocytes B entraînant leur prolifération (Young & Rickinson, 2004). D'autres auteurs supposent que le système est encore plus complexe et qu'il impliquerait d'autres agents environnementaux tels que des plantes⁴ dont les extraits stimulent le développement de tumeurs (van den Bosch, 2004). La répartition géographique de ces plantes étant similaire à celle de la ceinture des lymphomes.

Les patients souffrant d'un BL développent des tumeurs au niveau des mâchoires et des ganglions lymphatiques. Cette tumeur est mortelle si elle n'est pas traitée. Les cellules cancéreuses sont caractérisées par des mutations très spécifiques et caractéristiques des BL tant endémiques que sporadiques. On retrouve toujours l'une des trois translocations chromosomiques suivante : soit dans 80% des tumeurs, une translocation entre le chromosome 8 (à proximité du locus *myc*) et le chromosome 14 (au niveau du gène des chaînes lourdes des Ig) ; soit entre le chromosome 8 et l'un des loci des chaînes légères des Ig sur les chromosomes 2 ou 22 (dans 20% des cas). Ces mutations affectent des gènes codant pour des protéines impliquées dans le contrôle de la prolifération cellulaire telle que : *c-myc* (Kelly *et al.*, 2006), et le virus exprime des protéines (Eg. : EBNA-1) inhibant les protéines cellulaires de contrôle

⁴*Euphorbia tirucalli*.

de la prolifération telle que p53 (Kennedy *et al.*, 2003). De même, des marqueurs antigéniques importants, pour l'identification et l'élimination des cellules anormales, par les lymphocytes T cytotoxiques (CTL), sont dérégulés dans les cellules cancéreuses du lymphome de Burkitt, ce qui rend le contrôle par le système immunitaire peu efficace (Staege *et al.* (2002), Khanna & Burrows (2000)).

La **maladie de Hodgkin** est une autre forme de tumeur associée dans 40% des cas à EBV. C'est un syndrome prolifératif affectant les lymphocytes-B (Andersson, 2006). La structure des nœuds lymphatiques est perturbée et on note la présence de cellules malignes dites de Hodgkin/Reed Sternberg (HRS) (Murray & Young, 2005). Cette maladie se déclenche le plus souvent chez le jeune adulte et présente un second pic d'incidence autour de 45 ans (Nicolas & Seigneurin, 2002). Sa répartition est mondiale avec une incidence plus forte dans les pays développés. On suppose qu'EBV est un facteur nécessaire mais pas suffisant au déclenchement de la maladie, d'autres facteurs doivent entrer en jeu tels que l'origine géographique, le sexe, l'ethnie, l'âge (Murray & Young, 2005)...

Bien que le tropisme d'EBV soit principalement dirigé vers les lymphocytes B, certains cancers des lymphocytes T, dénommés "**lymphome des cellules-T**", sont associés à EBV. Néanmoins les recherches ne sont pas très avancées dans ce domaine (Dupuis *et al.*, 2006).

Le **carcinome du nasopharynx** (NPC), est un carcinome peu ou pas différencié associé à EBV dans 100% des cas. Il atteint l'adulte âgé de 20 à 50 ans. Sa prévalence est particulièrement élevée en Asie du Sud-Est, chez les Esquimaux et en Afrique du Nord et de l'Est (Wei & Sham, 2005). Les données actuelles suggèrent une combinaison de facteurs environnementaux et génétiques dans le déclenchement de la maladie. Le type de CMH⁵ (Hildesheim *et al.*, 2002) mais aussi l'alimentation jouent un rôle (Gallicchio *et al.*, 2006). Les populations asiatiques émigrées aux Etats-Unis, dont les habitudes alimentaires se sont adaptées aux coutumes locales, présentent une incidence moins forte de NPC (Buell (1974), Dickson & Flores (1985)). Les études ne disent pas si ces populations développent de nouvelles pathologies, causées non plus par EBV mais davantage, comme le suggérait Denis Burkitt, par la richesse de la nourriture Nord-Américaine ... Le NPC est un carcinome qui résulte de la transformation des cellules épithéliales du nasopharynx. Comme la plupart des cancers associés à EBV, les événements clés dans le déclenchement de la maladie ne sont pas ou peu connus. Il semblerait qu'une infection latente par EBV soit suffisante (Strauss & Strauss, 2002).

⁵Les **Complexe Majeure d'Histocompatibilité** sont des protéines clés dans l'immunité cellulaire.

Les **personnes dont le système immunitaire est compromis** (HIV⁺, transplantés avec traitement immuno-suppresseur) présentent un risque particulièrement élevé de développer un syndrome prolifératif associé à EBV (Tableau 2.1 page 30). La raison principale serait l'incapacité des lymphocytes T cytotoxiques (LTC) à contrôler la prolifération des lymphocytes B infectés de façon latente par EBV (Thorley-Lawson, 2005).

TAB. 2.1: Tumeurs associées à EBV, adapté de Strauss & Strauss (2002) et Cohen (2005)

Tumeur	Sous-type	Période de latence	Prévalence d'EBV (%)	Antigène détecté	Type de latence	Distribution mondiale
Lymphome de Burkitt	Endémique	3-8 ans	100		I	Infecte en générale les enfants de moins de 15 ans dans les régions où le paludisme est endémique
	Sporadique	3-8 ans	15-85	EBNA1	?	
Carcinome du nasopharynx	Associé au SIDA	3-8 ans post-VIH	30-40		I/II	
	N/A	> 30 ans	100	EBNA1, LMP1, LMP2	II	Majoritairement en Asie du Sud-EST
Maladie de Hodgkin	Type cellulaire variable	> 30 ans	80-90	EBNA1, LMP1, LMP2	II	Répartition mondiale, majoritaire dans les pays développés
	Sclérose Nodulaire	> 10 ans	30		?	
Lymphome des cellules-T	MNI* fatale	< 6 mois	? 100	?	?	
	Nasale AILD pleomorphique**	> 30 ans	100	EBNA1, LMP1, LMP2	II II/III	
Lymphome immunoblastique	MNI* fatale	< 6 mois	100	EBNA1, 2, 3A, 3B, 3C	?	
	Associé à la transplantation (précoce)	< 6 mois après la transplantation	100	EBNA1, 2, 3A, 3B, 3C	III	
	Associé à la transplantation (tardif)	> 1 an après la transplantation	100	LMP1, 2	III	
	Associé au SIDA	5-10 ans post VIH	70-80		?	

* MNI = Mononucléose-Infectieuse, ** AILD = Angioimmunoblastique-lymphoœdénopathie-like

TAB. 2.2: Caractéristiques biologiques des trois familles d'Herpes-virus, adapté de Viruses and Human Disease (Strauss & Strauss, 2002)

Caractéristiques	Alpha herpes-virus	Beta herpes-virus	Gamma herpes-virus
Gamme d'hôtes infectés	Variable, large	Restreint	limité à la famille de l'hôte naturel
Cycle de vie	court	long	relativement long
Infection en culture de cellule	Se répand rapidement infecte de nombreux types cellulaires	Progresse doucement	Infecte en priorité les cellules lymphoblastoïdes
Cytotoxicité	La plupart des cellules sont détruites	Cytomégalie	Infection lytique possible (cellule épithéliale, fibroblaste)
Latence	Ganglion sensoriel	Différents types cellulaires glande de sécrétion, reins ...	Spécifique des lymphocytes B ou T
Genre / représentant Humain	<i>Herpes simplex</i> HHV-1 (HSV-1) HHV-2 (HSV-2) <i>Varicellovirus</i> HHV-3 (VZV)	<i>Cytomegalovirus</i> HHV-5 (CMV) <i>Roseolovirus</i> HHV-6 HHV-7	<i>Lymphocryptovirus</i> HHV-4 (EBV) <i>Rhadinovirus</i> HHV-8 (KSHV)

2.2 Biologie d'EBV

2.2.1 EBV est un herpes-virus

Les **herpes-virus** constituent une grande famille de plus de 100 virus qui infectent les vertébrés, en général les mammifères et les oiseaux mais également les amphibiens, des reptiles et des poissons (Strauss & Strauss, 2002). Ce sont d'anciens virus qui ont co-évolué et qui se sont particulièrement bien adaptés à leurs hôtes. Chaque herpes-virus n'infecte qu'une espèce de vertébré dans la nature. Les herpes-virus ont été divisés en trois familles : alpha, beta et gamma, en fonction de leurs propriétés biologiques ainsi que des informations disponibles relatives à leurs séquences d'ADN (Tableau 2.2).

Les *alpha herpes-virus* infectent un grand nombre d'animaux et un grand nombre de cellules (Tableau 2.2). En culture, ils se propagent rapidement et détruisent efficacement les cellules infectées. Chez leurs hôtes naturels, ils établissent une infection latente des neurones sensoriels et l'infection lytique a lieu, en général au niveau des cellules de l'épiderme. Chez l'Homme, cette famille est représentée par **herpes simplex** et par **Varicella zoster** (Strauss & Strauss, 2002).

Les *beta herpes-virus* infectent un nombre restreint d'hôtes et leur cycle d'infection en culture est long

(Tableau 2.2 page 31). Les cellules infectées grossissent⁶. Chez son hôte naturel, le virus persiste de façon latente, notamment, dans les glandes de sécrétion et dans les reins. Le **cytomégalo**virus est un des représentant humain de cette famille (Strauss & Strauss, 2002).

Les *gamma herpes-virus* sont les herpes-virus qui possèdent le nombre d'hôtes le plus restreint, ils n'infectent que la famille ou l'ordre auquel l'hôte naturel appartient (Tableau 2.2 page 31). Ils se répliquent dans les cellules lymphoblastoïdes et certains peuvent infecter de façon lytique les cellules épithéliales et les fibroblastes. Ils sont spécifiques des lymphocytes B et T et l'infection est en générale latente. **HHV-8** appartient au genre *Rhadinovirus*, alors que **HHV-4**, l'autre nom d'**EBV**, appartient au genre *Lymphocryptovirus* (Strauss & Strauss, 2002).

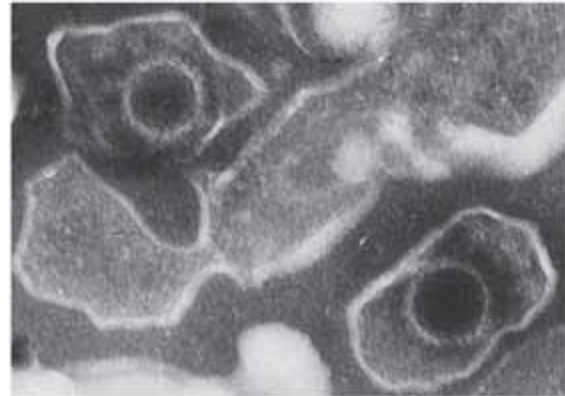
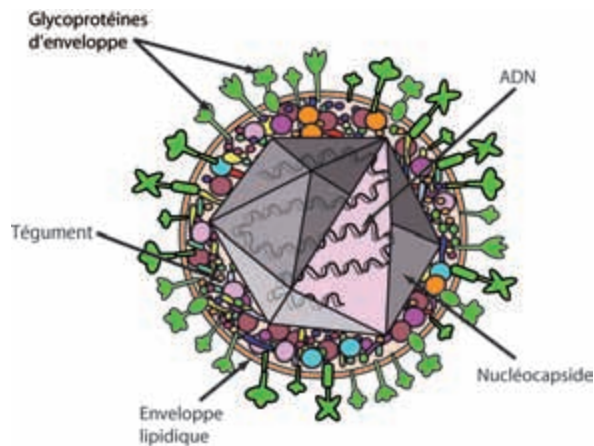
Les 8 herpes-virus connus chez l'Homme sont également désignés par l'appellation *HHV*, pour "*Human Herpes Virus*" suivie du numéro 1 à 8. En général, les maladies auxquelles ils sont associés ne constituent pas une menace pour la survie de l'individu infecté, à l'exception des patients dont le système immunitaire est affaibli, soit par une maladie du type *HIV*, ou, suite à un traitement immunosuppresseur⁷. Nous reviendrons par la suite au cas particulier d'EBV, dont l'association à de nombreux cancers est la conséquence d'interactions complexes entre le virus, son hôte ainsi que des facteurs environnementaux (Gallicchio *et al.* (2006), Hildesheim *et al.* (2002), van den Bosch (2004)).

2.2.2 Structure du virion

Les herpes-virus, sont des virus enveloppés. Ils possèdent une *capside icosaédrique* d'une centaine de nanomètres de diamètre (150 nm pour EBV Figure 2.3(b) page 34). Cette capsid est entourée d'un ensemble de protéines regroupées sous l'appellation de *tégument*. Le tégument est entouré par une membrane constituée d'une *bicouche lipidique* dans laquelle sont enchâssées des *glycoprotéines* (Figure 2.2(a) page 33). Les événements permettant la formation d'un virus mature ne sont pas entièrement connus, la plupart des travaux concernent herpes-simplex chez qui, il semblerait que, la nucléocapsid s'assemblerait à l'intérieur du noyau, les protéines constituant le tégument interagissant avec la nucléocapsid en partie dans le noyau et au moment du bourgeonnement au travers de la membrane nucléaire et avant l'entrée dans le golgi ; ces différentes étapes permettent au virus d'acquérir sa bi-

⁶On parle de cytomégalie.

⁷Chez les patients greffés ayant subi une greffe d'organe, par exemple.



(a) Représentation schématique d'un herpes-virus, adapté de Flint & al (2004)

(b) Image en microscopie électronique d'EBV (Altmann & Hammerschmidt, 2005)

FIG. 2.2: Morphologie et organisation structurale des herpes-virus

couche lipidique et les protéines membranaires qui la composent (Mettenleiter (2006) , Mettenleiter *et al.* (2006)).

2.2.3 Réplication des herpes-virus

Tropisme cellulaire d'EBV

Le virus d'Epstein-Barr infecte préférentiellement les lymphocytes B. Il se fixe sur la membrane cellulaire par l'interaction entre la glycoprotéine d'enveloppe virale *gp350/250* et le *CD21*, qui est le récepteur du complément. Le *CD21* est également présent sur les membranes de certains lymphocytes T et de certaines cellules épithéliales mais ceci n'est pas la règle et, jusqu'à récemment, l'infection des cellules épithéliales restait un mystère (Nicolas & Seigneurin, 2002). Récemment, les équipes de Rickinson et Delecluse ont montré que les lymphocytes B au repos n'internalisent pas systématiquement le virus, EBV reste fixé à la surface cellulaire et il est transféré aux cellules épithéliales par formation d'un complexe ("Synapse virale") entre lymphocytes B et cellules épithéliales. EBV a mis au point un moyen d'accès à différents compartiments de l'organisme : les lymphocytes B et les cellules épithéliales de l'épithélium pharyngé (Figure 2.4 page 35, Shannon-Lowe *et al.* (2006)). Des résultats récents de l'équipe de Hutt-Fletcher suggèrent que l'infection des cellules épithéliales pourrait procéder d'un mécanisme n'utilisant pas la fixation sur les *gp350/250* et le *CD21* mais d'autres protéines membranaires

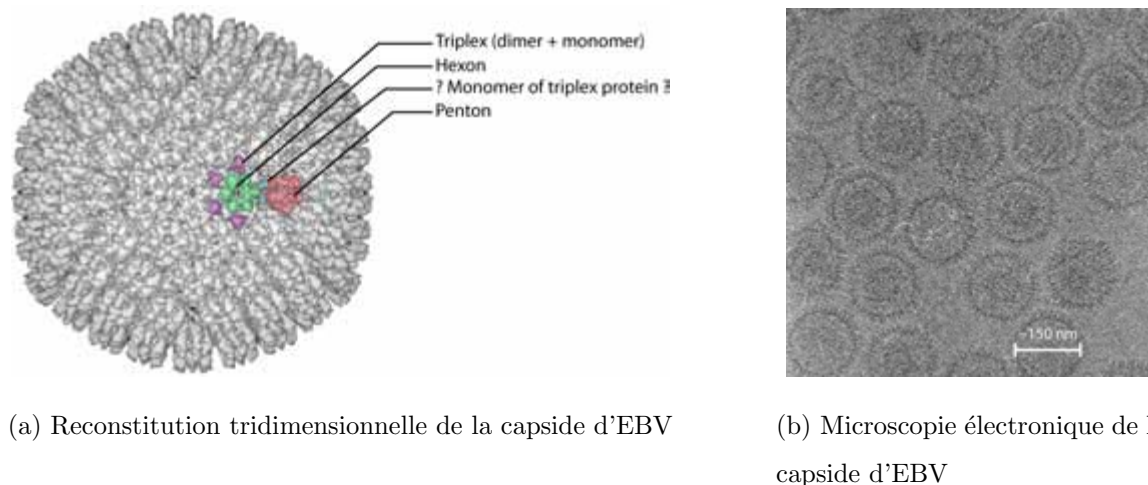


FIG. 2.3: Capside du Virus d'Epstein-Barr (Travail réalisé par : Raphaele Germe, Josette Guimet, Guy Schoehn)

telles que la protéine gH (Wu *et al.*, 2005). Ces résultats laissent à penser que l'infection de différents types cellulaires par EBV utilise un grand nombre de voies dont la plupart ne sont pas encore connues (Turk *et al.*, 2006).

EBV infecte les lymphocytes B de façon latente

In-vitro, EBV a la capacité unique de **transformer** des lymphocytes B quiescents en cellules lymphoblastoïdes infectées de façon latente et permanente (*LCLs* pour Latently infected Lymphoblastoid cell lines). Cet outil est essentiel pour étudier les propriétés transformantes du virus. Les cellules épithéliales constituent un système d'étude plus compliqué puisqu'il est extrêmement difficile d'obtenir un cycle répliatif complet menant au cycle lytique. La transformation des lymphocytes B en LCLs constitue le modèle canonique utilisé pour décrire l'infection par EBV (Young & Rickinson, 2004). L'ADN d'EBV est linéaire à l'intérieur de la capside et se circularise au cours de l'infection. Sa répliation a lieu au niveau du noyau. Les herpes-virus codent pour de nombreuses protéines (environ 86 protéines chez EBV, Tarbouriech *et al.* (2006)). Un grand nombre de ces protéines sont impliquées dans la régulation du métabolisme, la réparation et la répliation des acides nucléiques, nous y reviendrons en détail⁸. Il existe différents promoteurs répartis sur le génome des herpes-virus, ils sont

⁸Eg. : Le cœur répliatif est constitué des protéines suivantes : BALF5, BMRF1, BBLF2, BBLF4, BSLF1 et BBLF2/3. Remarque : dans cette partie et celles qui vont suivre, nous allons nommer de nombreuses protéines exprimées par EBV. La désignation des gènes codant pour ces protéines obéit à une nomenclature particulière qui dérive

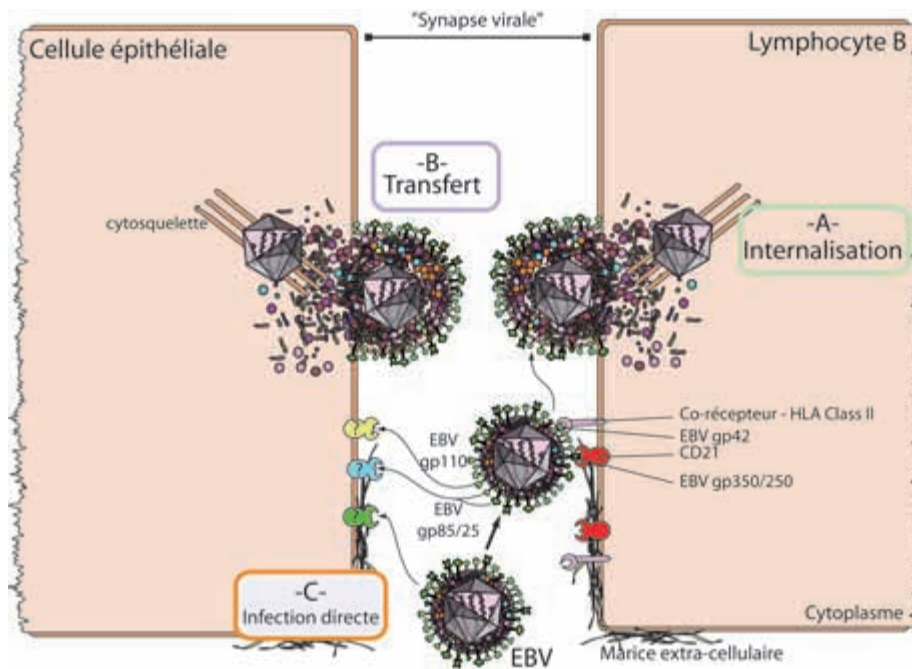


FIG. 2.4: Mécanisme hypothétique de l'infection par EBV de différents types cellulaires. (A) EBV se fixe sur un LB par l'interaction entre la gp350 virale et le récepteur cellulaire CD21 mais également par l'interaction entre la glycoprotéine virale gp42 et le récepteur HLA de classe II. La fusion entre le virus et la cellule est assurée par les protéines gp85/25 et gp110. Le virus est ensuite internalisé. (B) EBV infecte une cellule épithéliale par transfert à partir du LB. Les glycoprotéines protéines gp85/25 permettent la liaison à la cellule épithéliale et, avec la glycoprotéine gp110, elle permettent la fusion à la cellule épithéliale. La protéine gp42 ne semble pas intervenir dans ce processus. (C) EBV infecte directement une cellule épithéliale.

utilisés de façon séquentielle pour temporiser l'expression des gènes au cours du cycle viral (Figure 2.5 page 36, Nicolas & Seigneurin (2002)).

Les protéines, dont la fonction est de réguler l'expression du génome, sont codées par des gènes que l'on appelle "*gènes très précoces*". Ils sont transcrits en l'absence de protéines virales. L'expression de ces gènes permet l'expression des "*gènes précoces*", impliqués dans la réplication de l'ADN. Ces gènes ne sont transcrits que si les protéines codées par les gènes très précoces sont présentes. L'expression

de la façon dont le génome a été étudié. Chez EBV, les gènes sont nommés en fonction du fragment de restriction *Bam*HI auquel ils appartiennent. Les fragments *Bam*HI sont ensuite nommés en fonction de leurs tailles. Par exemple BARF1 est le premier cadre de lecture vers la droite du fragment *Bam*HI A ; BXL2 est le deuxième cadre de lecture vers la gauche du fragment *Bam*HI X (Robertson, 2005).

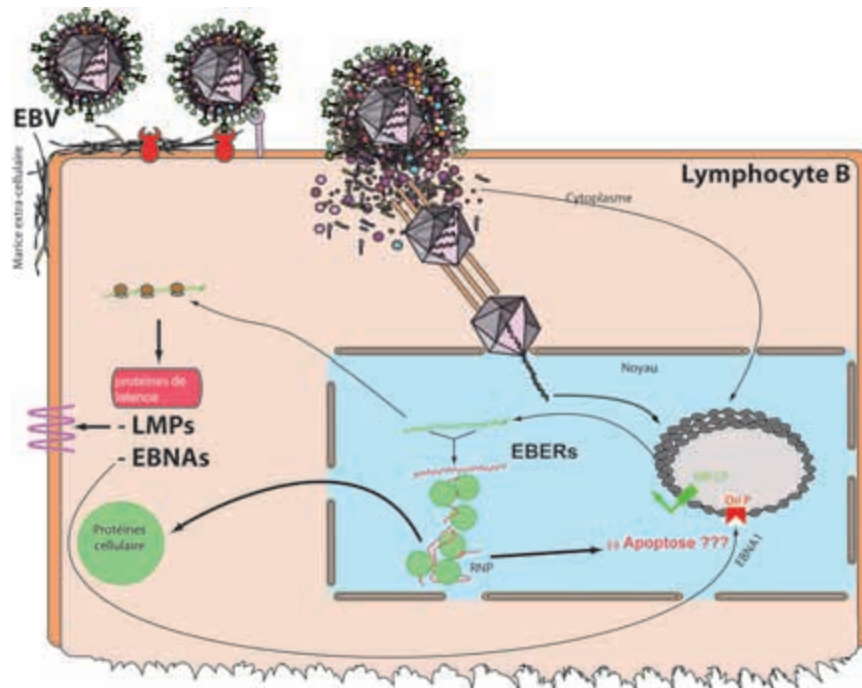


FIG. 2.5: Infection latente des lymphocytes B. EBNA1 interagit avec l'origine de réplication Orip et maintient le génome sous formes épisomale. Les gènes correspondants aux autres protéines EBNA sont transcrits à partir des promoteurs Wp ou Cp. Les EBERS restent majoritairement dans le noyau. Leurs rôles vont de l'inhibition de l'apoptose à la stimulation de la prolifération cellulaire. La protéine membranaire LMP1 est essentielle pour le processus de transformation.

des "gènes tardifs" ne survient qu'une fois l'ADN virale répliqué. Les protéines exprimées permettent l'assemblage des virions (Strauss & Strauss, 2002).

Les ARN_m sont produits par l'ARN polymérase II des cellules hôtes, ils sont polyadénylés et exportés vers le cytoplasme pour être traduits, ceci implique un mécanisme spécifique utilisant la protéine EB-2 (Hiriart *et al.*, 2003). La plupart des transcrits ne codent qu'une protéine et la plupart ne sont pas épissés. Il a été montré qu'un grand nombre de protéines exprimées ne sont pas nécessaires à la survie en culture cellulaire, elles fournissent au virus les outils lui permettant de tromper le système immunitaire de l'hôte ou de réguler l'infection *in-vivo* (Reddy *et al.*, 1998).

Contrôle de l'expression génique au cours de l'infection latente

Les lymphocytes B infectés par EBV comportent de multiples copies de son génome. Ce génome exprime un nombre limité de produits (protéines et ARN) dont le but est de maintenir **la latence**

TAB. 2.3: Séquence d'expression des gènes d'EBV en fonction du type de latence, adapté de Cohen (2005)

Latence	EBNA-1	EBNA-2	LMP-1	LMP-2	EBER	BARTs
I	+	-	-	-	+	+
II	+	-	+	+	+	+
III	+	+	+	+	+	+
0	-	-	-	-	+	?

(Figure 2.5). Les produits du génome d'EBV exprimés au cours de la latence sont les suivants (Nicolas & Seigneurin (2002), Young & Rickinson (2004)) :

- 6 Antigènes Nucléaires⁹ (**EBNAs** 1, 2, 3A, 3B, 3C, LP). Ces protéines permettent le maintien de l'épisome viral au cours des divisions cellulaires (EBNA1, Kieff & Rickinson (2001), Figure 2.5 page 36) ainsi que la croissance et la prolifération des cellules infectées (EBNA2 en complexe avec EBNA-LP, Mannick *et al.* (1991), Sinclair *et al.* (1994)) en même temps que la dérégulation du cycle cellulaire (EBNA3A et EBNA3C, Tomkinson *et al.* (1993)).
- 3 protéines membranaires de latence¹⁰ (**LMPs** 1, 2A, 2B). LMP1 est la principale protéine oncogène d'EBV. Elle entraîne, entre autre, une augmentation de la synthèse de répresseur de l'apoptose. LMP1 ressemble au CD40¹¹ et peut se substituer à ce dernier¹², cette protéine peut fournir les signaux nécessaires à la prolifération et à la différenciation des lymphocytes B *in-vivo* (Hatzivassiliou *et al.*, 2006).
- ARN non polyadénylé, non codant¹³ (**EBER** 1, 2). Ces ARNs ne sont pas indispensables à la transformation des lymphocytes B naïfs (Nanbo *et al.*, 2002). Une autre "piste" est fournie par l'étude d'HSV-1 chez qui le *Transcrit Associé à la Latence* (LAT), seul gène exprimé au cours de l'infection latente des neurones, confère une résistance à l'apoptose aux neurones infectés en diminuant la quantité de TGF- β disponible (Gupta *et al.*, 2006).

⁹EBNA : EB-Virus Nuclear Antigen ; LP : Leader Protein

¹⁰LMP : Latent Membrane Protein

¹¹Le CD40 appartient à la famille des Tumor Necrosis Factor (TNFR)

¹²LMP1 peut jouer le rôle d'un récepteur activé de façon constitutive

¹³EBER : EB-virus Encoded RNAs

- **BARTs** : BamHI-A rightward transcripts, famille d'ARNm exprimés au cours de la latence, dont le rôle est incertain .

Toutes les protéines impliquées dans la latence ne sont pas exprimées au même moment, ce qui permet de définir différents types de latences en fonction de leur profil d'expression (Tableau 2.3 page 37, Cohen (2005)). Les étapes de la latence d'EBV sont associées aux différentes maladies que nous avons décrites précédemment (Tableau 2.1 page 30). Les LCLs présentent un grand nombre de marqueurs membranaires caractéristiques de l'activation des lymphocytes B (CD23, CD30, CD39, ...). Ces marqueurs sont en général absents des lymphocytes B quiescents et ne sont exprimés que de façon transitoire, suite à une stimulation antigénique ou mitogénique. L'immortalisation résultant d'une infection par EBV emprunte certainement les voies métaboliques suivies par les lymphocytes B lors de leur activation et de leur prolifération physiologique (Wang *et al.*, 1990).

Le cycle latent d'EBV est le principal responsable des cancers associé à EBV (Tableau 2.1 page 30).

Le cycle de vie d'EBV comporte un cycle lytique : expression des gènes très précoces

L'étude du passage du cycle latent au cycle lytique a toujours été problématique en raison de l'absence d'un modèle cellulaire approprié. Les lymphocytes de la lignée B95-8 sont les plus utilisés. *In-vivo*, on trouve fréquemment du virus dans la salive de sujets sains, suite à une réactivation asymptomatique du virus dans l'épithélium oral ou, parmi la population de lymphocytes B des amygdales (Walling *et al.*, 2003). De façon occasionnelle, des patients atteints par le SIDA, non traités par des anti-viraux, souffrent de "leucoplasie chevelue de la langue", pathologie au cours de laquelle on trouve une forte infection lytique au niveau de l'épithélium (Walling *et al.*, 2003). Nous avons vu précédemment que les tumeurs associées à EBV sont le résultat de la transformation des cellules par le virus latent, néanmoins, certains auteurs ont postulé qu'une réactivation préalable du cycle lytique est nécessaire à la transformation (Menet *et al.*, 1999). *In-vivo*, la réactivation du cycle lytique est mal connue, on suppose qu'elle résulte de l'activation des lymphocytes B en réponse à un antigène (Antigen-mediated activation of B cells). *In-vitro*, différents facteurs ont la capacité de réactiver le cycle lytique d'EBV.

- Le "cross-linking" des immunoglobulines de surface (Tovey *et al.* (1978), Takada (1984)).
- La voie du TGF- β a la capacité de réactiver le cycle lytique d'EBV dans les cellules de lymphome

de Burkitt (Adler *et al.*, 2002). De même, l'interaction entre les lymphocytes T CD4⁺ et les lymphocytes B infectés par EBV réactive le cycle lytique du virus (Fu & Cannon, 2000).

- Au cours de la leucoplasie chevelue de la langue, la réactivation du virus est induite par la différenciation des cellules épithéliales infectées par EBV (Young *et al.*, 1991).
- Des produits chimiques, tels que le TPA¹⁴ ont la capacité de réactiver EBV (Fresen *et al.*, 1978).

Le cycle lytique peut être divisé en trois étapes. Ici encore, l'expression séquentielle est de rigueur. Pendant longtemps, le facteur déclenchant le basculement du cycle latent vers le cycle lytique était inconnu. Vers le milieu des années 80, les produits des gènes *BZLF1* et *BRLF1* furent identifiés en tant qu'élément essentiel permettant le déclenchement de ce processus (Countryman & Miller (1985), Chevallier-Greco *et al.* (1986), Countryman *et al.* (1987), Hardwick *et al.* (1988)). Ces protéines sont qualifiées de **très précoces**. *BZLF1* code pour la protéine *ZEBRA*¹⁵ (Z EB Replication Activator, Figure 2.6, Petosa *et al.* (2006)) , *BRLF1* code pour la protéine Rta (R transactivator, Miller & El-Guindy (2002)). Quel que soit le stimulus qui entraîne la différenciation, *BZLF1* et *BRLF1* sont exprimées (Figure 2.7 page 41, Feng *et al.* (2004)), ces protéines induisent la transactivation des gènes précoces (Ragoczy & Miller, 1999). Il a été montré que la transfection de lymphocytes en culture, infectés par EBV, avec un vecteur codant pour la protéine *ZEBRA* était un élément nécessaire et suffisant au déclenchement du cycle lytique (Countryman & Miller (1985), Chevallier-Greco *et al.* (1986)).

L'activation des gènes précoces par *ZEBRA* et *RTA* permet la production des **protéines précoces**¹⁶. Les gènes codant pour ces protéines sont transcrits avant la réplication lytique du génome. Ces gènes ne sont pas transcrits quand il y a une inhibition de la synthèse protéique (à l'opposé des gènes de *BZLF1* et de *BRLF1*). Les **gènes tardifs** codent pour les composants structuraux du virus tels que la capside. Leur expression est interrompue par l'inhibition de la réplication virale. Dans la plupart des cas, la cellule infectée ne survit pas au cycle lytique (Figure 2.7 page 41, Kim & YI (2003), Inman *et al.* (2001), Gutiérrez *et al.* (1996)).

¹⁴TPA : 12-O-tetradecanoyl-phorbol-13-acetate.

¹⁵Autrement appelée : EB1, Zta ou Z.

¹⁶Eg. : L'ADN polymérase virale est une protéine précoce.



FIG. 2.6: Structure du domaine de liaison à l'ADN de la protéine ZEBRA liée à un promoteur de gènes lytique d'EBV (Petosa *et al.*, 2006).

Cycle lytique d'EBV : expression des gènes précoces

Les protéines codées par les gènes précoces couvrent une large gamme de fonctions. La régulation de la réplication est assurée par BMRF1, c'est une protéine associée à l'ADN polymérase virale (DNA polymerase processivity factor, Tsurumi *et al.* (1993)). Si l'on supprime le gène codant pour cette protéine, la réplication lytique ne peut avoir lieu ; néanmoins, le phénotype sauvage peut être restauré par trans-complémentation en fournissant BMRF1 au virus (Neuhierl & Delecluse, 2006). Certaines protéines permettent le transport des ARNm et assurent leur stabilité dans le cytoplasme. Beaucoup de gènes du cycle lytique d'EBV n'ont pas d'introns. La protéine EB2¹⁷ (BMLF1) est une protéine de liaison à l'ARN qui augmente la stabilité des transcrits sans introns et permet leur transport vers le cytoplasme (Boyer *et al.* (2002), Batisse *et al.* (2005)). BHRF1 et BARF1 permettent respectivement l'inhibition de l'apoptose¹⁸ (Henderson *et al.*, 1993) et l'évasion du système immunitaire¹⁹ (Strockbine *et al.*, 1998) (Figure 2.8 page 42, Tarbouriech *et al.* (2006)).

Les protéines codées par les gènes lytiques précoces permettent également la production d'une kinase virale : BGLF4²⁰ (Gershburg *et al.*, 2004). Cette protéine est l'homologue de UL-97 que l'on trouve chez HCMV (Michel & Mertens, 2004). Elle entraîne la phosphorylation de protéines d'EBV

¹⁷EB2 est également appelée SM ou Mta.

¹⁸BHRF1 a des effets similaires à l'anti-apoptotique Bcl-2.

¹⁹BARF1 aurait la capacité de lier le CSF-1 (Colony Stimulating Factor) qui est une cytokine stimulant la différenciation des macrophages. Ceci permettrait à EBV, selon Strockbine *et al.*, de moduler la réponse immunitaire.

²⁰Autrement appelée EBV-PK (PK = Protéine Kinase).

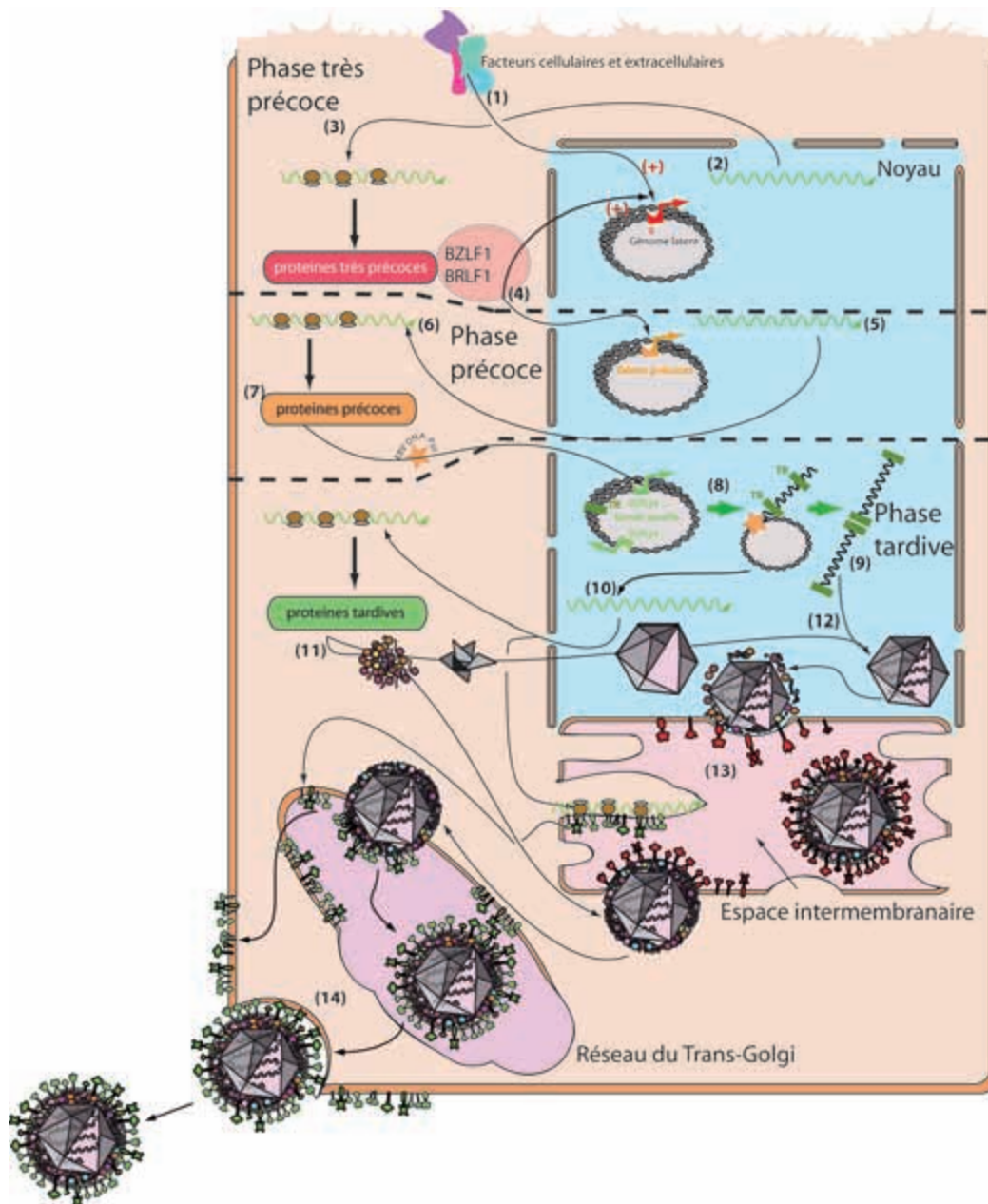


FIG. 2.7: Cycle lytique d'EBV. (1) Le cycle lytique est activé par des facteurs cellulaires ou extracellulaires qui interagissent avec le promoteur des gènes très précoces. (2) L'ARNm des gènes très précoces est exporté vers le cytoplasme (3) où il est traduit. (4) Les protéines BZLF1 et BRLF1 vont auto-activer leur promoteur et interagir en trans avec le promoteur des gènes précoces. (5) L'ARNm des gènes précoces est exporté et (6) traduit en protéines précoces, (7) notamment une ADN polymérase virale. (8) Le génome est répliqué à partir des origines de replication "OriL" ou "OriLyt". (9) Les segments de génomes sont coupés au niveau des répétitions terminales ("TR") (Israel & Kenney, 2005). (10, 11) Une fois la répllication de l'ADN effectué, les protéines tardives sont exprimées. (12) A partir de ce moment, le génome est encapsidé et le virion suit les voies d'exocytose, il utilise, vraisemblablement, le réseau du Trans-Golgi (Turcotte *et al.*, 2005) (13). Le passage au travers de la membrane cellulaire fournit l'enveloppe lipidique du virus sur laquelle sont enchâssées des glycoprotéines (14). Adapté de Flint & al (2004)

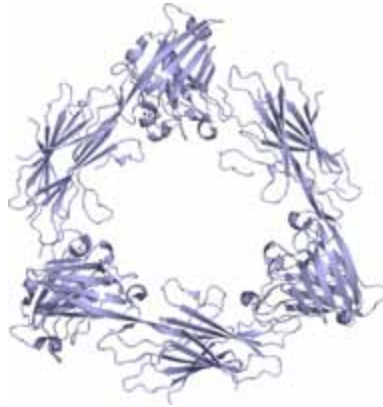


FIG. 2.8: Structure cristallographique de BARF1. Cette protéine permet au virus d'échapper au système immunitaire.

(BMRF1 et EBNA-LP, Chen *et al.* (2000)) aussi bien que de protéines cellulaires (Facteur 1 Δ , Kato *et al.* (2001)). En revanche, on ne sait pas quelles sont les conséquences de ces phosphorylations.

D'autres protéines interviennent dans la réplication du virus et interagissent avec l'origine de réplication lytique (OriLyt, Figure 2.7 page 41) (Fixman *et al.*, 1995). Le "cœur" du système de réplication virale comprend : BALF5 (ADN polymérase), BMRF1 (le facteur de processivité de l'ADN polymérase), BSLF1 (une primase), BBLF2/3 (protéines associées à la primase), BALF2 (une protéine de liaison à l'ADN simple brin) et ZEBRA (un facteur de transcription, Schepers *et al.* (1993)) (Liao *et al.*, 2001). BGLF5 (DNase ou Alkaline Exonucléase) pourrait également participer à la machinerie de réplication, son rôle n'est pas documenté chez EBV mais, son homologue chez HSV intervient dans les phénomènes de recombinaison (Reuven *et al.* (2003), Reuven *et al.* (2004)).

D'autres protéines précoces sont impliquées dans le métabolisme des acides nucléiques. Ce sont BXLF1 (une thymidine kinase, Littler *et al.* (1986)), BaRF1 et BORF2 (les petites et grandes sous unités de la Ribonucléotide réductase, Kieff & Rickinson (2001)) et BLLF3 (dUTPase, Tarbouriech *et al.* (2005), Fleischmann *et al.* (2002) Figure 2.9 page 43). On suppose que ces protéines permettent d'assurer la production de substrat nucléotidique dans les cellules quiescentes. BKRF3 est une Uracile-ADN Glycosylase ... cette protéine étant le sujet central de ce travail, nous y reviendrons plus en détails.

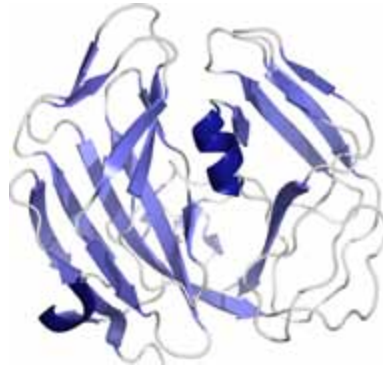


FIG. 2.9: Structure cristallographique de BLLF3. C'est une dUTPase, elle permet au virus de maintenir un ratio dUTP/dUMP le plus bas possible, ce qui minimise l'incorporation d'uracile dans le génome.

Cycle lytique d'EBV : expression des gènes tardifs

Dans un souci d'exhaustivité, il convient de citer les protéines exprimées au dernier moment, une fois que la replication de l'ADN a eu lieu. Ces protéines dites "tardives" permettent la formation de la capsid virale (BNRF1, BFRF2/3, BLRF1/2/3, BcLF1, Johannsen *et al.* (2004)) des glycoprotéines de membrane (BLLF1 : gp350/220, BXLF2 : gp85, BZLF2 : gp42, BKRF2 : gp25) qui déterminent le tropisme cellulaire du virus (Guerreiro-Cacais *et al.*, 2004). La phase tardive permet également la production de BCLF1 ou vIL10, c'est une protéine homologue de IL10, elle permet au virus de se protéger contre la réponse immunitaire de l'hôte (Hsu *et al.*, 1990). Pour finir, de nombreuses protéines parmi les 86 protéines potentiellement codées par le génome d'EBV, devraient être impliquées dans la maturation du virion et son bourgeonnement au travers de la membrane cellulaire, mais très peu d'éléments sont connus, on suppose que le schéma suivi est équivalent à celui des α -herpesvirus (Turcotte *et al.*, 2005).

Chapitre 3

Contexte de l'étude et but du projet

LORSQUE j'ai commencé ce travail en DEA, le groupe de "*génomique structurale du virus d'Epstein-Barr*" dirigé par le Professeur **Wilhelm Pascal Burmeister** venait de débiter son activité. L'équipe était amenée à faire partie de l'"*Institut de Virologie Moléculaire et Structurale*" (IVMS). Cet institut regroupe différentes équipes et surtout différentes compétences allant de la biologie moléculaire à la cristallographie en passant par la virologie médicale. Cet environnement permet d'envisager une approche globale de la virologie, aussi bien fondamentale qu'appliquée. Le but de l'équipe étant, à terme, de conduire à l'élaboration rationnelle de médicaments pour lutter contre les pathologies associées à l'infection par EBV.

Qu'appelle-t-on conception rationnelle d'un médicament ? Quel est le rôle de la biologie structurale dans ce contexte ? Les exemples concrets abordés dans cette thèse permettent d'appréhender l'intérêt mais aussi les difficultés d'une telle approche.

- Dans un premier temps nous verrons de quelle façon la parallélisation du traitement d'un grand nombre de cibles permet de rationaliser l'approche de la biologie structurale. Nous verrons quels ont été les développements techniques ainsi que les limitations associées à ce type d'approche.
- L'étude structurale et fonctionnelle de l'Uracile-ADN Glycosylase occupe une place centrale dans ma thèse. Nous verrons comment ce projet s'intègre à la thématique du groupe, les problématiques spécifiques auxquelles il répond et les questions nouvelles qu'il soulève.
- Les complications techniques sont inhérentes à la biologie structurale. L'étude structurale préliminaire de l'Alcaline Exonucléase nous permettra d'aborder un certain nombre de problèmes

récurrents, par exemple, la difficulté d'obtention de cristaux de tailles et de qualité suffisantes pour entreprendre une étude par cristallographie aux rayons-X. Nous aborderons les alternatives expérimentales permettant de résoudre ces problèmes.

- Dans une dernière partie nous verrons quelles perspectives nouvelles ce travail apporte à la recherche sur EBV et quels sont les développements ultérieurs qu'il suggère.

Deuxième partie

Génomique structurale du virus d'Epstein-Barr

Chapitre 4

Introduction : Sélection de cibles et Génomique structurale

4.1 Biologie structurale et conception rationnelle de médicaments

UNE définition simple de la biologie structurale pourrait être la suivante : “C’est l’étude de la forme des molécules biologiques”. Une telle définition laisse la porte ouverte aux interprétations les plus diverses et nécessite quelques précisions. A quand remonte la première expérience de biologie structurale ? Avec la définition précédente, nous pourrions la dater à l’invention du microscope photonique par **Anton Van Leeuwenhoek** (Figure 4.1 page 50), c’est-à-dire au milieu du XVII^e siècle. Van Leuvenhoek fit les premières observations microscopiques de cellules. Il fut le premier à décrire ce qu’on ne peut voir à l’œil nu. En quelques siècles, la “seule” différence tient à l’échelle d’observation. Nous avons amélioré les techniques pour observer des objets de plus en plus petits. L’utilisation des rayons-X nous permet d’envisager l’étude de la matière au niveau atomique, moyennant un artefact expérimental : la préparation de cristaux. Les difficultés inhérentes à cette technique ainsi que la technique elle-même seront discutées dans la partie consacrée à BGLF5¹ (page 145).

La connaissance de la structure tridimensionnelle d’une protéine, au niveau atomique permet de

¹BGLF5 sera appelée Alkaline Exonucléase ou DNase, ces appellations étant équivalentes.



FIG. 4.1: Anton Van Leeuwenhoek (1632-1723), est considéré comme le père de la microscopie.

corrélér sa forme à sa fonction. Dans le cas d'une enzyme, l'identification précise du site actif présente un intérêt évident si l'on veut modérer son activité de façon rationnelle. Historiquement, la découverte des médicaments reposait sur l'utilisation empirique et systématique de composés naturels plus ou moins modifiés. L'identification de la pénicilline par Alexander Fleming (Figure 4.2 page 51) est un exemple célèbre de l'utilisation d'un produit naturel pour le traitement d'infection bactérienne (Fleming, 2001).

L'industrie pharmaceutique a évolué vers de nouvelles techniques avec notamment le développement de banques de composés permettant un criblage haut débit, pour découvrir un/des inhibiteur(s) potentiel(s) d'une cible donnée (Bajorath, 2002). La biologie structurale permet d'adopter une nouvelle approche, dite *rationnelle* de la conception de médicaments : la structure d'une cible importante pour le cycle de vie d'un virus, par exemple, est déterminée et son site actif est précisément identifié. Il est possible de concevoir *in-silico* un inhibiteur spécifique pour ce site actif. Dans certains cas, un inhibiteur ou le substrat naturel est connu, l'utilisation des informations structurales permet de modifier ces composés, dans le but d'en faire des médicaments efficaces. L'inhibition de la protéase du virus du SIDA (Figure 4.3 page 52) est l'exemple classique du succès de cette approche (Miller *et al.* (1989), Whittle & Blundell (1994), Blundell *et al.* (1990)).

La génomique structurale est née au milieu des années 1990 aux Etats-Unis et au Japon avec

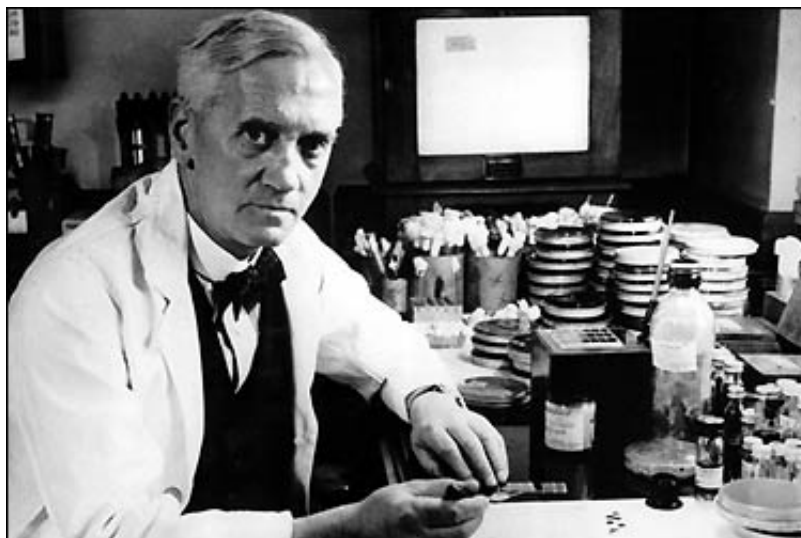


FIG. 4.2: Sir Alexander Fleming, prix Nobel de Physiologie et de Médecine en 1945 pour la découverte de la pénicilline.

la fondation de l'**Organisation Internationale de Génomique Structurale** (ISGO). A la suite des campagnes de séquençage à haut débit de génomes entiers, l'idée fut émise qu'il serait possible d'appliquer des méthodes similaires pour obtenir les structures tridimensionnelles de toutes les protéines d'un organisme (son protéome). Cette approche systématique permettrait d'approfondir nos connaissances relatives aux différents repliements adoptés par les protéines. (Stuart *et al.*, 2006).

L'équipe "*génomique structurale du virus d'Epstein-Barr*" propose d'apporter une connaissance approfondie de la structure des protéines du virus d'Epstein-Barr. Par analogie avec ce qui a été fait sur le virus du SIDA, la première cible fut la protéase d'EBV (Figure 4.4 page 53, Buisson *et al.* (2002)). Le résultat de cette étude démontra que cette protéine, essentielle à la réplication du virus, n'est pas une bonne cible pour la conception rationnelle de médicaments. Son site actif, contrairement au site actif de la protéase du virus du SIDA (Figure 4.3 page 52), est trop en surface, l'affinité des inhibiteurs ne serait pas suffisante. Néanmoins, l'obtention de la structure permet d'adopter une autre approche, ciblant l'inhibition de la dimérisation de la protéine.

Les enzymes, dont le rôle dans le cycle de vie d'un virus est essentiel, constituent une cible de choix pour la conception d'inhibiteurs, ainsi, il fut décidé de poursuivre l'étude par la résolution, systématique, d'un grand nombre d'enzymes d'EBV. Ce projet, qui se poursuit jusqu'à aujourd'hui, est effectué dans le cadre du partenariat Européen SPINE.

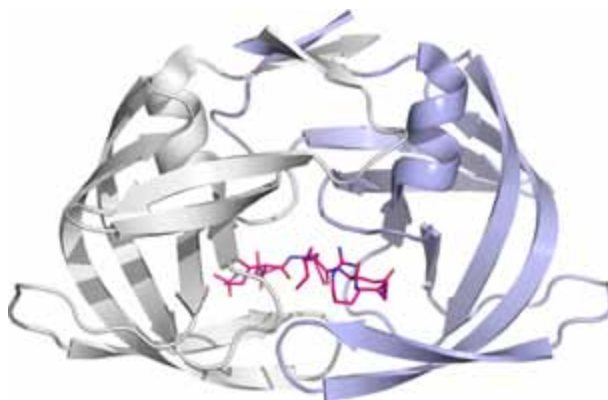


FIG. 4.3: La structure de la protéase du virus du SIDA en complexe avec un inhibiteur est l'exemple classique de l'approche rationnelle de la conception de médicaments (Miller *et al.*, 1989).

4.1.1 Le projet SPINE

Le projet SPINE ou "*Structural Proteomics IN Europe*" est une initiative européenne permettant le rapprochement d'une vingtaine de laboratoires de biologie structurale pour développer de nouvelles méthodes et technologies, afin d'accroître la détermination de nouvelles structures de protéines. Le projet débuta le 1^{er} octobre 2002 pour une durée de 3 ans. Les différentes équipes travaillent sur des thématiques en relation avec la santé humaine, allant des structures de protéines virale, bactérienne ou en lien avec le système immunitaire et le cancer. Rétrospectivement, le projet SPINE aura surtout permis un développement technique en permettant d'automatiser un grand nombre de tâches répétitives (robot de clonage, robot de cristallisation, changeur d'échantillons). C'est dans ce contexte que débuta le travail sur les protéines d'EBV.

4.2 Métabolisme des acides nucléiques / réparation / répliation : des cibles de choix

4.2.1 Une régulation complexe

Le génome du virus d'Epstein-Barr code pour 3 enzymes impliquées dans le métabolisme des acides nucléiques (page 40). Outre les enzymes spécifiques d'EBV telles que l'ADN polymérase virale et ses

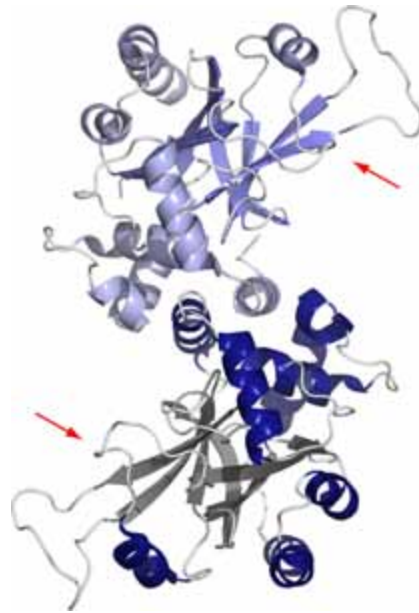


FIG. 4.4: Structure du dimère de la protéase d'EBV. Le site actif (flèche rouge) en surface ne permet pas la conception facile d'inhibiteurs (Buisson *et al.*, 2002).

cofacteurs, essentiels à la replication de l'ADN viral, EBV code pour d'autres enzymes qui sont déjà présentes dans la cellule qu'il infecte. La nécessité pour le virus de coder de telles enzymes viendrait de leur absence ou de leur faible expression chez l'hôte (Chen *et al.*, 2002).

Le métabolisme des acides nucléiques est composé d'un ensemble de voies interconnectées (Figure 4.5 page 54) dont le but est la production des (désoxy/ribo)-nucléotides triphosphates entrant dans la composition des acides nucléiques : ADN (si le sucre est un désoxyribose) ou ARN (si le sucre est un ribose).

Sans entrer dans les détails, ces voies de synthèse sont hautement régulées et possèdent des étapes de contrôle appelées "committed step" que l'on pourrait traduire par "étape d'engagement". Ces étapes sont cruciales et déterminent si une voie métabolique peut continuer jusqu'à la synthèse du produit final ou si elle doit s'arrêter (en raison d'une carence par exemple). Parmi les enzymes sur lesquels nous avons travaillé, plusieurs font partie de la voie métabolique des desoxyribonucléotide², plus particulièrement des pyrimidines (Figure 4.6 page 56) et de la voie de synthèse qui mène à la formation

²BaRF1, BORF2 : Ribonucléotide Réductase, RNR.

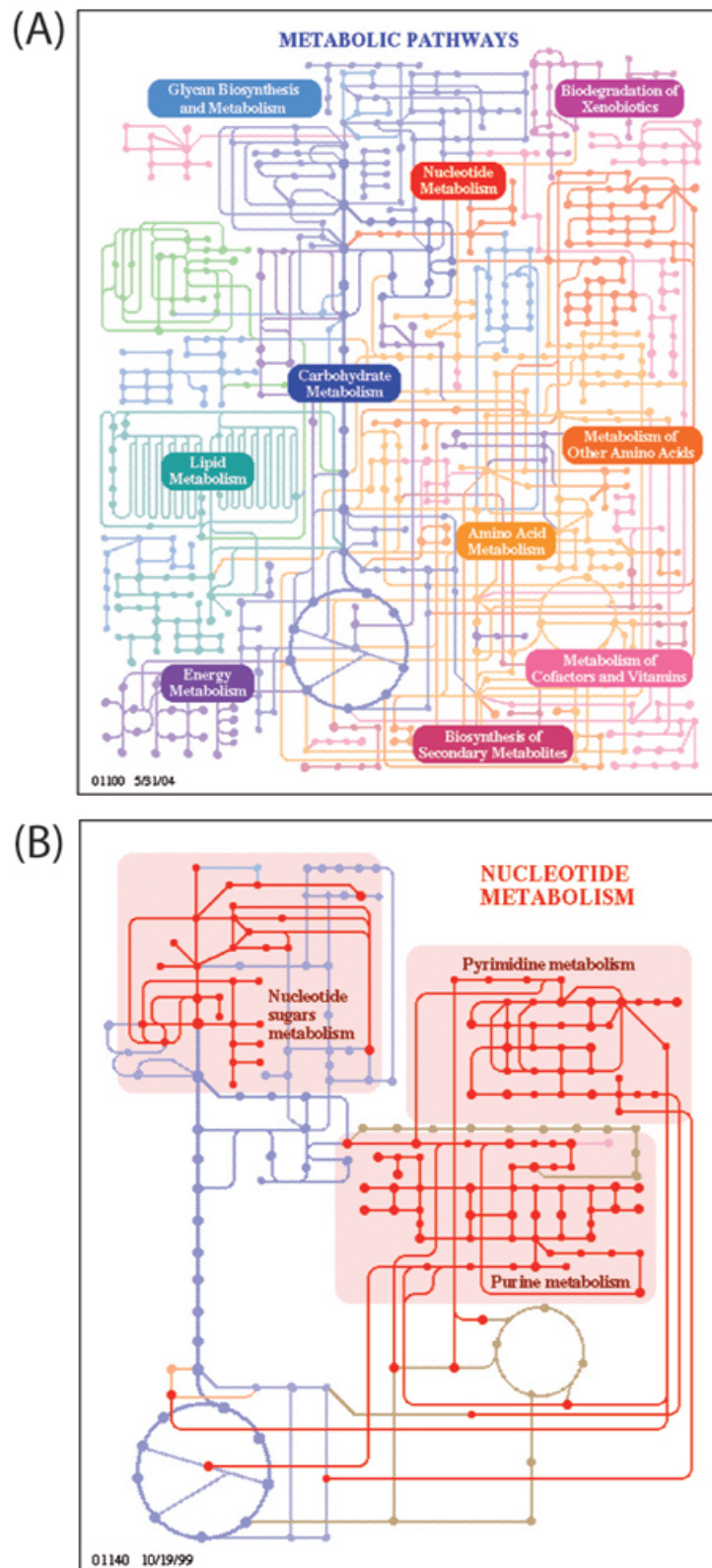


FIG. 4.5: (A) Vue d'ensemble du métabolisme cellulaire. (B) Le métabolisme des acides nucléiques est divisé en trois parties : la synthèse du sucre, la synthèse des purines et la synthèse des pyrimidines (Kanehisa *et al.*, 2006).

du désoxythymidylate³ (dTMP, Figure 4.7 page 57).

La **Ribonucléotide Réductase** (RNR) occupe une position centrale puisqu'elle est responsable de la formation des désoxyribonucléotides à partir des ribonucléotides (Figure 4.9 page 58). Elle est essentielle pour la synthèse de l'ADN (Figure 4.6 page 56). La RNR est une enzyme qui comporte 2 sous unités : une petite sous-unité BaRF1 et une grande BORF2. Les voies de synthèse des désoxyribonucléotides pyrimidiques sont hautement connectées et il existe plusieurs "routes" menant à la synthèse de chaque désoxyribonucléotide, excepté pour le **dTMP**. En effet, la synthèse *de-novo* du dTMP n'est catalysée que par la **thymidylate synthase** (TS). Cette enzyme utilise comme substrat le dUMP et le tétrahydrofolate (THF). La réaction catalysée par la TS produira, en plus du dTMP, du dihydrofolate (DHF), lui même régénéré en THF par la DiHydroFolate Reductase (DHFR). L'inhibition⁴ de la TS ou de la DHFR bloque la synthèse *de-novo* de dTMP, ce qui conduit à la mort cellulaire (Van Triest *et al.*, 2000). Il existe néanmoins une voie de recyclage (*salvage pathway*), permettant la synthèse de dTMP, qui n'utilise pas le dUMP comme substrat. Cette réaction fait intervenir la **thymidine kinase** (Figure 4.8) qui catalyse la phosphorylation de désoxy-thymidine en désoxy-thymidine monophosphate⁵. L'inhibition de cette enzyme pourrait contribuer au dysfonctionnement de la voie de synthèse du dTMP, elle constitue par conséquent une cible intéressante.

4.2.2 Principe de la chimiothérapie anti-cancéreuse et anti-virale

Lorsque l'on veut inhiber la croissance d'une cellule cancéreuse ou infectée par un virus, plusieurs stratégies sont envisageables. Ce type de cellule prolifère plus vite ; c'est d'ailleurs le principal problème associé à la tumorigénèse, c'est-à-dire : une prolifération rapide et non contrôlée, qui ne tient pas compte des signaux environnementaux (Eg. : inhibition de contact). Historiquement, l'inhibition de la croissance tumoral repose sur le principe suivant : la cellule cancéreuse croît plus vite que les cellules saines qui l'environnent. Par conséquent, l'administration d'un produit toxique, conduit plus les cellules tumorales à la mort cellulaire que les cellules "normales" dont le métabolisme est moins

³Différentes enzymes sont concernées : BLLF3 la dUTPase, BKRF3 l'Uracile-ADN Glycosylase à laquelle la partie centrale de cette thèse est consacrée (voir page 83), BXLF1 la Thymidine Kinase (TK).

⁴Par le methotrexate ou le 5FdUMP.

⁵La TK est l'enzyme responsable de la phosphorylation de l'acyclovir chez HSV (Reardon & Spector, 1989).

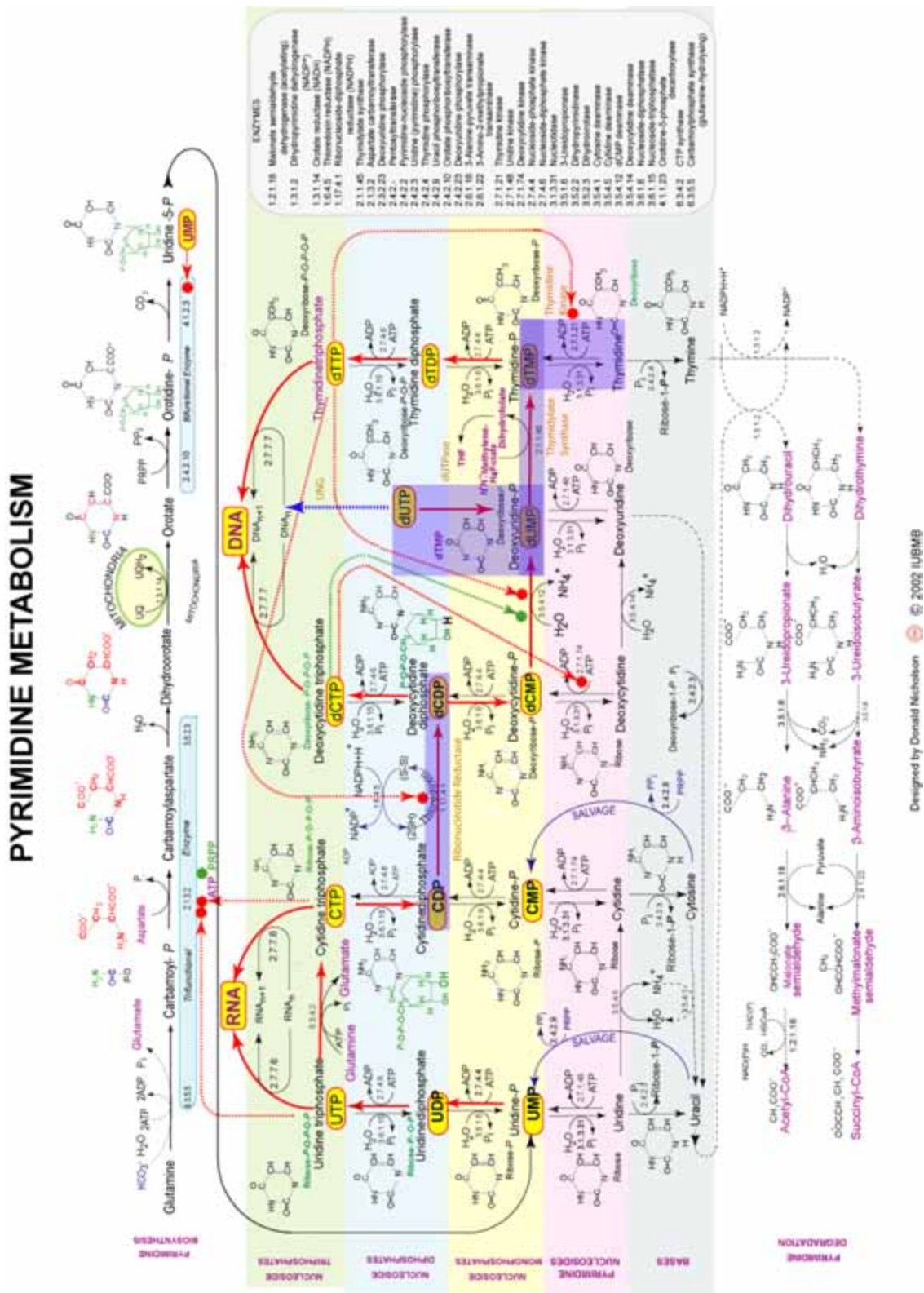


FIG. 4.6: Métabolisme des pyrimidines. Les enzymes sur lesquels nous avons travaillé sont indiqués en marron, les voies de synthèse concernées sont surlignées en bleu. La voie de synthèse *de-novo* du dTMP est unique. L'UNG permet d'exciser l'uracile incorporé dans l'ADN.

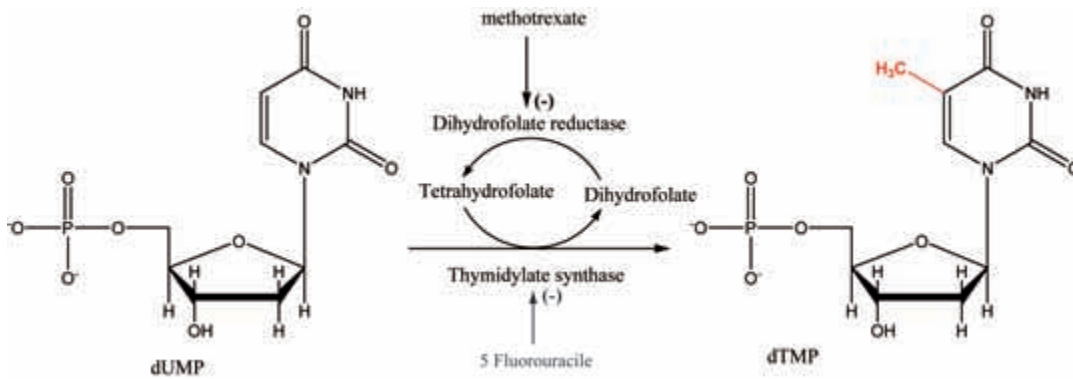


FIG. 4.7: La seule voie de synthèse *de-novo* du dTMP suppose la méthylation du dUMP. Cette voie métabolique est une cible classique des chimiothérapies contre le cancer.

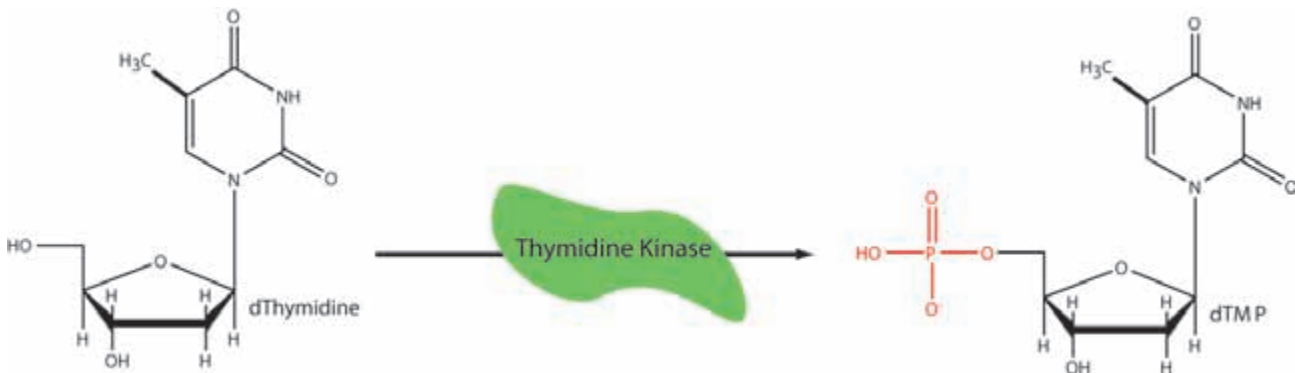


FIG. 4.8: La Thymidine Kinase permet la phosphorylation de désoxy-thymidine en désoxy-thymidine-phosphate.

actif. Pour inhiber la croissance des cellules infectées par un virus, il existe d'autres stratégies. Un virus exprime des enzymes dont la spécificité pour certains métabolites est différente des enzymes cellulaires. Afin d'inhiber le développement des cellules infectées, on administrera un métabolite qui n'est toxique qu'après transformation par une enzyme spécifique du virus. Seules les cellules infectées seront affectées par la présence de ce métabolite. C'est le cas de l'**Acyclovir** qui nécessite une modification par la Thymidine Kinase virale (Furman *et al.*, 1980) pour être activé et inhiber l'ADN polymérase virale (Figure 4.10 page 58, Gnann *et al.* (1983), Reardon & Spector (1989)).

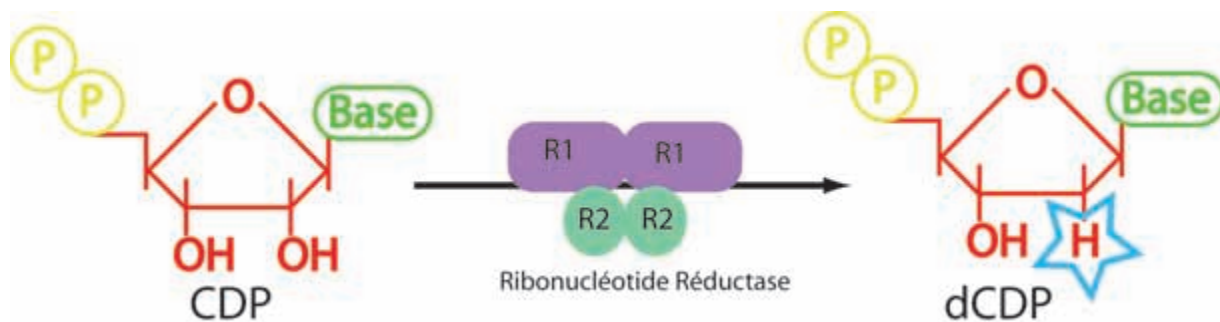


FIG. 4.9: La Ribonucléotide Réductase est la seule enzyme responsable de la formation de désoxyribonucléotide à partir de ribonucléotide.

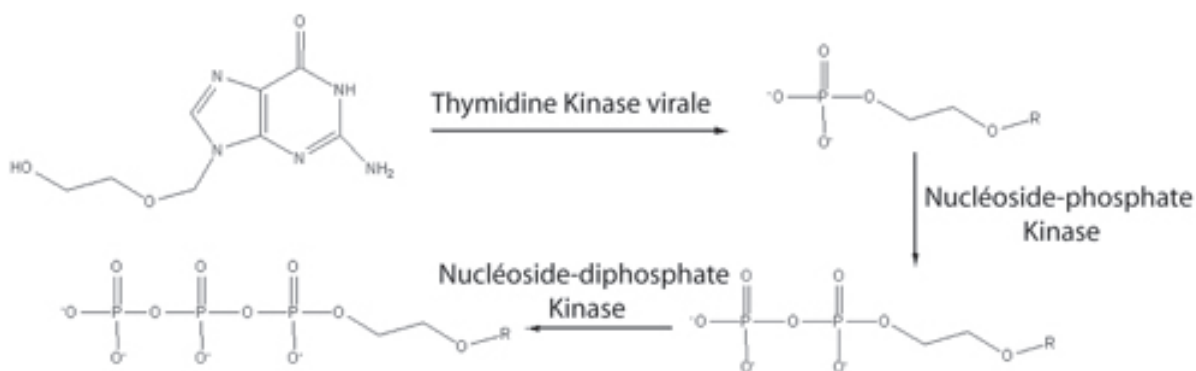


FIG. 4.10: L'Acyclovir est un anti-viral utilisé pour lutter contre l'infection par HSV1. Il inhibe l'ADN polymérase viral après activation par la TK viral.

4.2.3 L'inhibition de la voie de synthèse du dTMP est un "classique" de la chimiothérapie

Comme nous pouvons le voir dans les figures 4.5 page 54 et 4.6 page 56, le métabolisme procède par un jeu d'interconnexions qui permet de compenser la carence d'un métabolite par un autre. Une chimiothérapie qui viserait une voie métabolique complexe, permettant l'obtention d'un produit de différente manière, serait inefficace en raison de la compensation d'une voie par une autre. Or, la voie de synthèse du dTMP ne peut être compensée puisqu'elle est unique. Le 5-fluorouracile⁶ (5-FU, Figure 4.11(a) page 59) entre en compétition avec le dUMP et forme un complexe stable avec la TS.

⁶La forme active de ce médicament est le 5FdUMP ou 5-Fluoro-2'-deoxy-5'-monophosphateuridine, mais en générale on parle de 5-FU.

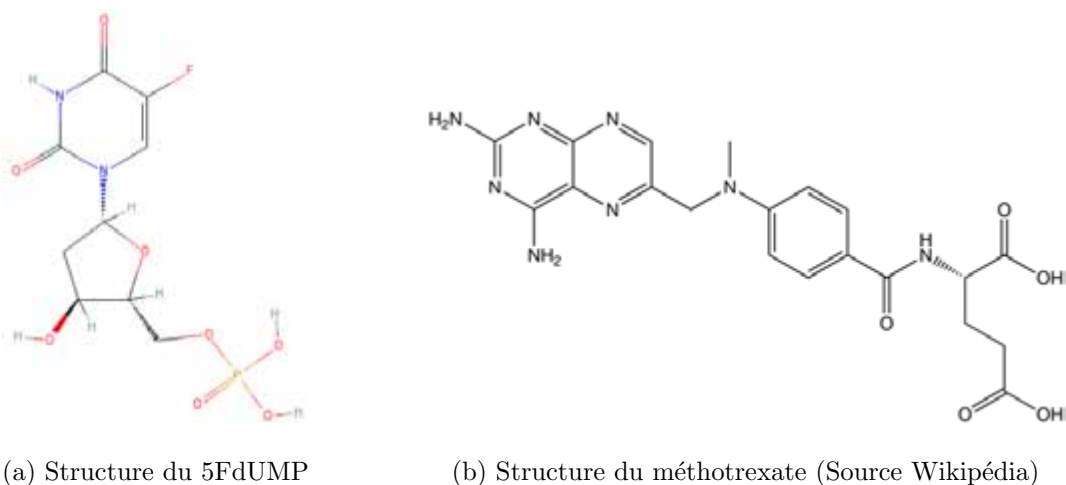


FIG. 4.11: Le 5-Fluorouracile et le méthotrexate sont des composés utilisés depuis de nombreuses années en chimiothérapie.

Le 5-FU est le médicament le plus utilisé dans le traitement des cancers du colon (Van Triest *et al.*, 2000). Le méthotrexate⁷ (Figure 4.11(b) page 59) s'attaque à la même voie métabolique mais inhibe la DHFR qui est l'enzyme responsable de la catalyse de la réaction $DHF \rightarrow THF$, le THF étant un cofacteur indispensable à la réaction $dUMP \rightarrow dTMP$ (Figure 4.7 page 57). L'inhibition de la synthèse de dTMP provoque non seulement un déséquilibre des pools de désoxyribonucléotides, mais aussi une cascade d'événements qui conduisent à un phénomène connu sous le nom de **“Thymine-less death”** ou mort consécutive à l'absence de thymine (Mennigmann & Szybalski, 1962). L'accumulation de dUMP, suite à l'inhibition de la TS, conduit à une accumulation de dUTP. Le ratio dUTP/dUMP est maintenu bas par la dUTPase (Figure 2.9, page 43). Lorsque le pool de dUMP est trop élevé, la concentration en dUTP augmente également, la dUTPase est dépassée et ne peut convertir le dUTP en dUMP suffisamment vite. La grande concentration en dUTP provoque son incorporation massive à l'intérieur de l'ADN. L'uracile n'étant pas une base “normale” de l'ADN, il existe des mécanismes de réparation qui permettent de l'éliminer et de la remplacer par la base canonique (ie. : la thymine). Lorsque le pool de dUTP est déséquilibré au point de provoquer une incorporation trop importante d'uracile, on n'assiste plus à une réparation du génome mais à sa fragmentation, donc à une mort cellulaire certaine (Van Triest *et al.*, 2000).

⁷La formule systématique du méthotrexate est : (S)-2-(4-(((2,4-diaminopteridin-6-yl) methyl)methylamino)benzamido) pentanedioic acid.

Conception d'inhibiteurs d'enzymes virales

Pendant des années, l'industrie pharmaceutique s'est surtout concentrée sur le développement d'inhibiteurs nucléotidiques visant les ADN polymérases virales (Coen & Schaffer, 2003). Ces médicaments agissent en tant que terminateurs de chaîne, ils bloquent la synthèse d'ADN des virus à l'intérieur des cellules infectées (Reardon & Spector, 1989). L'émergence de résistances conduit à la recherche de nouvelles cibles . Nous venons de voir que la voie de synthèse du dTMP comporte des enzymes dont l'inhibition conduit à la mort cellulaire et qui constitueraient, de fait, de bonnes cibles (Coen & Schaffer, 2003).

Chapitre 5

Matériel et Méthodes

5.1 Annotation du génome d'EBV : recherche d'une Thymidylate Synthase

5.1.1 Conception d'une base de données des protéines d'EBV

LE rôle de toutes les protéines codées par le génome d'EBV n'est pas connu. Si la distance évolutive n'est pas trop grande, il est possible d'assigner une fonction à ces protéines inconnues en se fondant sur leur ressemblance avec des protéines connues dont la fonction est correctement documentée. Cette approche repose sur l'utilisation d'outils bioinformatiques tels que les alignements de séquences et la recherche de motifs. Des bases de données¹ spécialisées permettent de faciliter ces tâches. Néanmoins, la création d'une base de données spécifiques, ne contenant que les protéines d'EBV permet d'accélérer la recherche de protéines homologues, à condition que l'on ait une idée de ce que l'on recherche.

Les séquences correspondant aux protéines codées par les cadres de lecture ouverts d'EBV sont disponibles sur la base de donnée VIDA (Albà *et al.*, 2001). Il est possible de récupérer l'ensemble de ces séquences au format FASTA afin de créer une base de données locale. La création d'une base de données locale suppose un formatage des séquences FASTA, pour cela il faut utiliser la commande “for-

¹VIDA pour les virus et SwissProt pour toutes les protéines, pour plus de détails, voir Annexe 179.

matdb”² du programme BLAST. Une fois la base de données construite, il est possible de confronter la séquence d’une protéine d’intérêt avec la base de données.

Recherche de protéines homologues : BLAST

Le programme BLASTp permet de rechercher une séquence dans une base de données de protéines³. BLAST⁴ est un algorithme heuristique qui utilise des approximations pour réduire le temps de calcul. Le but de BLAST n’est pas de produire un alignement exact de toutes les séquences, mais d’aller chercher les séquences homologues au plus vite, dans une base de données qui en contient un grand nombre. Un score et une “E-value”⁵ sont attribués à chaque résultat obtenu. Le score est la valeur donnée à la comparaison de 2 séquences en utilisant une matrice⁶. Plus le score est élevé plus les séquences sont similaires. La “E-value” ou “Expected Value” est un paramètre statistique qui décrit la probabilité qu’un alignement soit dû au hasard. La E-value diminue de façon exponentielle avec le score “S” attribué à l’alignement de deux séquences. En fait la E-value décrit le bruit de fond aléatoire qui existe quand on cherche à aligner des séquences. Par exemple, une E-value de 1 que l’on attribuerait à une touche peut être interprétée comme : “dans une base de données d’une taille donnée, on est certain de trouver une touche de même score uniquement par chance ; par conséquent, plus la E-value est proche de zéro plus la touche est significative”.

²*formatdb -iX.fasta -pT -oT -nY* ; X : nom du fichier contenant les séquences au format fasta, -pT : dits au programme que l’on utilise des séquences de protéines, -oT : permet d’indexer les séquences, -nY = permet de donner un nom à la base de données créée.

³La commande utilisée est : *blastall -pblastp -dY -iZ -eN -xxx.html -T* ; Y : nom de la base de données, Z : fichier correspondant à la séquence protéique d’intérêt au format fasta, N : valeur limite de la *E.value*, xxx : nom du fichier de sortie, -T : permet d’obtenir un résultat au format HTML.

⁴Basic Local Alignment Search Tool.

⁵E-value signifie Expected value ou valeur attendue.

⁶la matrice PAM ou Point Accepted Mutation, permet la comparaison de 2 séquences, et attribue une pénalité pour le remplacement d’un acide aminé par un autre.

5.2 Clonage et expression de la TK et de la RNR

Parmi les cibles sélectionnées comme candidates pour l'étude structurale de protéine d'EBV, un certain nombre d'enzymes du cycle lytique d'EBV ont été choisies. Nous reviendrons en détail sur les résultats obtenus pour l'Uracile-ADN Glycosylase et pour l'Alcaline Exonucléase. Néanmoins, nous avons travaillé sur d'autres enzymes pour lesquelles nous avons eu plus de difficultés et pour lesquelles les approches employées se sont révélées peu fructueuses : La **Thymidine Kinase** (BXLF1) dont le numéro d'accèsion Swissprot est le P03177, est une enzyme de 67.2 KDa qui comporte 607 acides-aminés. Le protocole de clonage et d'expression standard de cette protéine est décrit dans Tarbouriech *et al.* (2006). Nous avons également travaillé sur La **Ribonucléotide Réductase** (RNR). La RNR est une enzyme qui comporte 2 sous unités. Une petite sous-unité dont le numéro d'accèsion Swissprot est le P03175, cette enzyme est codée par le cadre de lecture ouvert BaRF1, elle comporte 302 acides-aminés et sa masse est de 34.3 KDa. Une grande sous-unité dont le numéro d'accèsion Swissprot est le P03190, elle est codée par le cadre de lecture ouvert BORF2. Cette enzyme comporte 826 acides-aminés et sa masse est de 93 KDa. En raison de la complexité de cette cible, différentes constructions ont été envisagées : une native et une tronquée des 95 premiers acides-aminés que le serveur FoldIndex[©] prédit comme étant désordonnée (Figure 5.1 page 64, Prilusky *et al.* (2005)). Deux protocoles de clonage sont utilisés. Le protocole standard décrit par Tarbouriech *et al.* (2006), un autre protocole utilise le vecteur pETDuet-1 (Novagen). Ce vecteur permet le clonage et la co-expression de 2 protéines, ce qui, en théorie, permet d'améliorer la solubilité des protéines qui ne fonctionnent pas seules et que l'on trouvera systématiquement associées à d'autres protéines *in-vivo*, ce qui est le cas de la RNR.

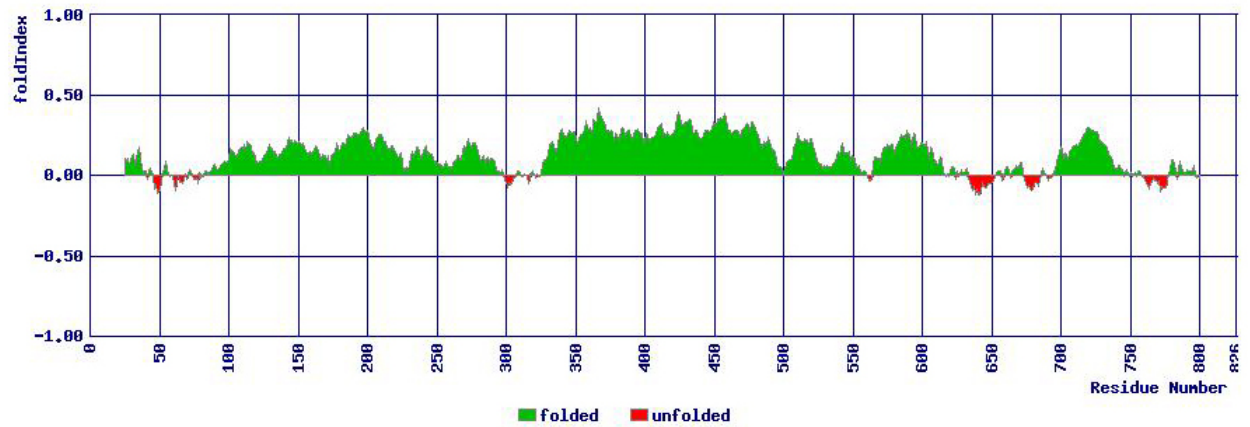


FIG. 5.1: Le serveur Folindex[©] nous permet de définir des zones désordonnées (rouge). La RNR Δ 95 est tronquée de 95 acide-aminés dans la région N-terminale, cette région est prédite comme désordonnée.

Chapitre 6

Article : Génomique structurale d'EBV

6.1 Résumé

LE virus d'Epstein-Barr est un herpesvirus. Il est responsable de maladies telles que la mononucléose infectieuse, de nombreux carcinomes et il est associé à des syndromes immunoprolifératifs. Son génome code pour 86 protéines, ce qui représente autant de cibles potentielles pour une approche de génomique structurale. Après une vérification de l'annotation du génome, 23 cadres de lecture ouverts ont été sélectionnés et exprimés chez *E.coli*. Les premières cibles furent sélectionnées en se fondant sur la connaissance d'une activité enzymatique ; par la suite, la liste fut complétée par d'autres protéines en fonction d'une série de critères, notamment leur haut degré de structure secondaire. De façon inattendue, le principal problème rencontré lors de ce projet fut la faible expression ainsi que l'insolubilité d'une grande partie des cibles sélectionnées. Ce problème ne put être résolu ni par la modification des constructions, ni par le changement de la température d'expression ou de la souche bactérienne, ou l'utilisation de techniques de renaturation. Parmi les 8 protéines solubles, 5 furent cristallisées à l'aide de nano-gouttes faites par un robot et un criblage de solutions. A l'issue de cela, 4 structures furent résolues. La clef de la réussite de ce projet fut un traitement individuel de chaque cible plutôt que l'utilisation de protocoles standards. Néanmoins, le criblage miniaturisé à haut débit des conditions de cristallisation, fut un élément essentiel pour le traitement de ces cibles difficiles.

6.2 Article 1

Structural genomics of the Epstein–Barr virus

Nicolas Tarbouriech,^{a,b} Marlyse Buisson,^{b,c} Thibault Géoui,^{a,b} Susan Daenke,^d Stephen Cusack^a and Wim Pascal Burmeister^{b,e*}

^aEMBL–Grenoble Outstation, BP 181, F-38042 Grenoble CEDEX 9, France, ^bInstitut de Virologie Moléculaire et Structurale, FRE 2854 CNRS–UJF, BP 181, F-38042 Grenoble CEDEX 9, France, ^cLaboratoire de Virologie, CHU Michallon, BP 217, F-38043 Grenoble CEDEX 9, France, ^dWellcome Trust Centre for Human Genetics, University of Oxford, Oxford OX3 7BN, England, and ^eInstitut Universitaire de France, 103 Boulevard Saint-Michel, F-75005 Paris, France

Correspondence e-mail: wpb@embl-grenoble.fr

Received 15 November 2005

Accepted 1 August 2006

Epstein–Barr virus is a herpesvirus that causes infectious mononucleosis, carcinomas and immunoproliferative disease. Its genome encodes 86 proteins, which provided targets for a structural genomics project. After updating the annotation of the genome, 23 open reading frames were chosen for expression in *Escherichia coli*, initially selecting for those with known enzyme activity and then supplementing this set based on a series of predicted properties, in particular secondary structure. The major obstacle turned out to be poor expression and low solubility. Surprisingly, this could not be overcome by modifications of the constructs, changes of expression temperature or strain or renaturation. Of the eight soluble proteins, five were crystallized using robotic nanolitre-drop crystallization trials, which led to four solved structures. Although these results depended on individual treatment rather than standardized protocols, a high-throughput miniaturized crystallization screening protocol was a key component of success with these difficult proteins.

1. Introduction

Human herpesviruses comprise three subfamilies: (i) α -herpesviruses [herpes simplex viruses (HSV) 1 and 2 and varicella zoster virus (VZV)], (ii) β -herpesviruses [cytomegalovirus (CMV) and human herpesvirus (HHV) 6 and 7] and (iii) γ -herpesviruses, comprising the Kaposi's sarcoma-associated herpesvirus (KSHV or HHV8) and Epstein–Barr virus (EBV or HHV4). The last infects the vast majority of the world's human population, establishing and maintaining a lifelong persistence in the infected host.

Primary infection typically occurs in childhood and is frequently asymptomatic. In contrast, a delayed primary infection in adolescents or young adults results in infectious mononucleosis (IM) in approximately half of cases, with symptoms including fever, pharyngitis, lymphadenopathy and splenomegaly. IM is a self-limiting lymphoproliferative disorder characterized by an expansion of EBV-infected B-lymphocytes associated with viral lytic replication in the oropharynx, controlled by a vigorous CD8⁺ cytotoxic T-cell immune response. The majority of cases of acute IM recover, but serious complications can occasionally lead to death. EBV is associated with a number of cancers in the immunocompetent host (Rickinson & Kieff, 1996), in particular Burkitt's lymphoma and nasopharyngeal carcinoma, which are endemic in African and Asian populations (Raab-Traub, 2005). Furthermore, EBV can lead to immunoproliferative disease in immunosuppressed patients, notably those infected with HIV (Rickinson & Kieff, 1996). Currently licensed anti-herpesvirus drugs (acyclovir and related compounds) directed against viral DNA synthesis (Coen & Schaffer, 2003) show little effect against EBV.

Table 1
Proteins of EBV.

Accession, SwissProt, TrEMBL or PIR (Protein Information Resource) accession number. NCBI, GI numbers assigned by NCBI. Function, information on name, synonyms and function of the protein. ORF EBV, name of the EBV open reading frame. F, classification based on the function into C, capsid; M, membrane (glyco)protein; N, nucleotide metabolism; L, latency; P, packaging; R, replication; S, transcription factors, transactivators, signalling; T, tegument. S, an × means translated from spliced messenger RNA. Occ., occurrence in herpesvirus subfamilies, no entry for proteins present only in EBV and very closely related viruses such as rhesus lymphocryptovirus (LCV). Homologue, name of the homologue in HSV, if existing, otherwise of human CMV. N, number of constructs used in the project. St, current status. Proteins with enzymatic activity are shown in bold. For space reasons, only a limited number of references to original work are given; otherwise, review articles are cited. The annotation extensively used the *BLAST* program at NCBI (Altschul *et al.*, 1997).

Accession	NCBI	Function	ORF EBV	F	S	Occ.	Homologue	Reference	N	St	Comment
P03229	gi:140202	Bcl-2 homologue, negative regulator of anti-apoptosis protein BHRF1	BALF1	S		γ		Marshall <i>et al.</i> (1999), Bellows <i>et al.</i> (2002), Cabras <i>et al.</i> (2005)	1	†	Toxic
P03227	gi:118744	Single-stranded DNA-binding protein, part of replication fork/ machinery	BALF2	R		$\alpha\beta\gamma$	UL29	Decaussin <i>et al.</i> (1995), Robertson <i>et al.</i> (1996)			
P25939	gi:124087	Terminase large subunit/ATPase	BALF3	P		$\alpha\beta\gamma$	UL28	Alba (2002), Hwang & Bogner (2002), Savva <i>et al.</i> (2004)			
P03188	gi:138191	Membrane glycoprotein B (gB, gp110) precursor, fusion and co-receptor binding	BALF4	M		$\alpha\beta\gamma$	UL27	Gong <i>et al.</i> (1987), Spear & Longnecker (2003)			
P03198 Q8AZJ4	gi:118858 gi:23893662	DNA polymerase BARF0, nuclear/perinuclear in epithelial cells, start at amino acid 298 of SWISS-PROT, homologue in rhesus LCV	BALF5 BARF0	R U	×	$\alpha\beta\gamma$	UL30	Kieff (1996) Hitt <i>et al.</i> (1989), Gilligan <i>et al.</i> (1991), Fries <i>et al.</i> (1997)	12	‡	Insoluble
P03175	gi:132620	Ribonucleoside reductase, small 38 kDa subunit	BaRF1	N		$\alpha\gamma$	UL40	Kieff (1996)	1	†	No expression
P03228	gi:115128	CD80 homologue (p33), oncogene, soluble glycoprotein, 2 Ig domains, binds CSF-1	BARF1	S				Wei & Ooka (1989), Stroockbine <i>et al.</i> (1998), Tarbouriech <i>et al.</i> (2006)	1	§	PDB 2ch8
P03216	gi:136813	Myristoylated phosphoprotein in tegument (MyrP)	BBLF1	T		γ		Johannsen <i>et al.</i> (2004)			
Not in DB	Not in DB	Primase-associated factor, spliced, full sequence not in DB, part of helicase-primase complex	BBLF2/3	R	×	$\alpha\beta\gamma$	UL8	Yokoyama <i>et al.</i> (1999)			
P03214	gi:122807	Helicase, part of helicase-primase complex	BBLF4	R		$\alpha\beta\gamma$	UL5	Yokoyama <i>et al.</i> (1999)	1	†	No expression
P03213	gi:136792	Portal protein UL6 homologue, function by homology	BBRF1	P		$\alpha\beta\gamma$	UL6	Newcomb <i>et al.</i> (2001)			
P29882	gi:267198	Unknown function	BBRF2	U		$\alpha\beta\gamma$	UL7	Johannsen <i>et al.</i> (2004)			
P03215	gi:140278	Glycoprotein M, part of gN-gM complex involved in envelope-tegument interaction	BBRF3	M		$\alpha\beta\gamma$		Lake & Hutt-Fletcher (2000), Johannsen <i>et al.</i> (2004)			
P03226	gi:137569	Major capsid protein, MCP, VP5, 155K	BcLF1	C		$\alpha\beta\gamma$		Kieff (1996)			
P03180	gi:114886	Viral interleukin-10 homologue precursor, vIL-10	BCRF1	S		$\beta\gamma$	hcmvUL84	Hsu <i>et al.</i> (1990)			PDB 1vlk, 1y6m
P25215 Q8AZK7	gi:136975 gi:7441833	Unknown function EBNA-Lp (EBNA-5) nuclear phosphoprotein, highly spliced, 12 exons, proline-rich, enhances EBNA-2 transactivation	BcRF1 BWRF1	U L	×	$\beta\gamma$ γ		Alba (2002) Kieff (1996), Waltzer <i>et al.</i> (1996)			
P25214	gi:139189	Minor capsid protein (mCP), triplex protein HSV-1 VP23 homologue	BDLF1	C		$\alpha\beta\gamma$	UL18	Trus <i>et al.</i> (2001)	1	¶	Purified
P03225	gi:136827	Tegument protein, transmembrane protein?	BDLF2	T		γ		Johannsen <i>et al.</i> (2004)	5	‡	Insoluble
P03224	gi:138369	Envelope glycoprotein gp150 (gp117)	BDLF3	M		γ		Borza & Hutt-Fletcher (1998), Johannsen <i>et al.</i> (2004)			
P03223	gi:136983	Protein (gp115)	BDLF4	U		$\beta\gamma$	hcmvUL92	Alba (2002), Boeckmann <i>et al.</i> (2003)			
Q66541	gi:1334910	Assemblin without protease domain coded by 2nd reading frame	BdRF1	C		$\alpha\beta\gamma$	UL26.5				
P03203	gi:119113	EBNA-3B nuclear protein (EBNA-4)	BERF2a/ BERF2b	L	×			Kieff (1996), Young & Rickinson (2004)			

research papers

Table 1 (continued)

Accession	NCBI	Function	ORF EBV	F	S	Occ.	Homologue	Reference	N	St	Comment
P03204	gi:119114	EBNA-3C nuclear protein (EBNA-6, EBNA-4B) (–) effect on transactivation, EBNA-2 and cell cycle, essential for immortalization	BERF3/ BERF4	L	×			Kieff (1996), Young & Rickinson (2004)			
P03184	gi:136878	Role in DNA packaging, cytosolic zinc-binding protein, cysteine-rich	BFLF1	P		$\alpha\beta\gamma$	UL32	Chang <i>et al.</i> (1996)			
P03183	gi:136873	Nuclear membrane phosphoprotein, part of intracellular virions, egress protein, complex with BFRF1	BFLF2	M		$\alpha\beta\gamma$	UL31	Lake & Hutt-Fletcher (2004), Gonnella <i>et al.</i> (2005)	2	‡	Insoluble
P03185	gi:136886	Nuclear membrane protein p38, transmembrane with large cytoplasm domain, complex with BFLF2	BFRF1	M		$\beta\gamma$	hcmvUL50	Lake & Hutt-Fletcher (2004)			
P14347	gi:140660	Unknown function, not included in virions	BFRF2	U		$\beta\gamma$	hcmvUL49	Bellows <i>et al.</i> (2002), Johannsen <i>et al.</i> (2004)			
P14348	gi:139195	Smallest capsid protein (sCP) on outer capsid surface, HSV1-VP26 homologue, KSHV ORF 65 homologue	BFRF3	C		$\alpha\beta\gamma$	UL35	Nealon <i>et al.</i> (2001), Bowman <i>et al.</i> (2003), Johannsen <i>et al.</i> (2004)			
AAA45868	gi:330334	DNA-cleavage and packaging protein, part of the DNA-packaging machinery (BFRF0.5, HS4BAM)	BFRF4	P		$\alpha\beta\gamma$	UL33	Alba (2002), Beard & Baines (2004)			
P03222	gi:136833	Tegument protein (gp115)	BGLF1	T		γ		Johannsen <i>et al.</i> (2004)	1	‡	Insoluble
P03221	gi:114952	Tegument protein, MyrPBP	BGLF2	T		$\alpha\beta\gamma$	UL16	Boeckmann <i>et al.</i> (2003), Johannsen <i>et al.</i> (2004)			
P03220	gi:136988	Homologue to HSV-1 tegument protein, but not included in virion, gp118	BGLF3	U		$\alpha\beta\gamma$	UL14	Johannsen <i>et al.</i> (2004)			
P13288	gi:125627	Ser/Thr kinase, phosphorylation of nucleoside analogues, tegument protein	BGLF4	N		$\alpha\beta\gamma$	UL13	Smith & Smith (1989), Marschall <i>et al.</i> (2002), Johannsen <i>et al.</i> (2004)	4	‡	Insoluble
P03217	gi:119691	Alkaline exonuclease, involved together with BALF2 in DNA recombination	BGLF5	R		$\alpha\beta\gamma$	UL12	Cheng <i>et al.</i> (1980), Reuven <i>et al.</i> (2004)			
P03219	gi:23893636	DNA-packaging protein, terminase small subunit	BGRF1/ BDRF1	P	×	$\alpha\beta\gamma$	UL15	Alba (2002)			
P03181	gi:140778	Proline-rich protein LF3, unknown function, tandem repeats, <i>NotI</i> repeat (125 bp), homologue in rhesus LCV	BHLF1	U				Laux <i>et al.</i> (1985), Rivailler <i>et al.</i> (2002), Farrell (2005)			
P03182	gi:119102	Anti-apoptotic factor bcl-2 homologue, early antigen protein R (EA-R), nuclear antigen	BHRF1	S		γ		Huang <i>et al.</i> (2003)	2	††	PDB 1q59 (NMR)
P03208	gi:138777	G-protein coupled receptor (G-PCR), 7 transmembrane helices, 6 glycosylation sites, 2 disulfide bridges gp64	BILF1	M		γ		Hutt-Fletcher (2005), Paulsen <i>et al.</i> (2005)			
P03218	gi:138183	Membrane glycoprotein gp55/80, Ig-like, gp78	BILF2	M				Mackett <i>et al.</i> (1990), Boeckmann <i>et al.</i> (2003), Johannsen <i>et al.</i> (2004)			
P03211	gi:119110	Latent nuclear protein EBNA-1, assures EBV episome maintenance replication, Gly-rich domain, essential for immortalization	BKRF1	L				Bochkarev <i>et al.</i> (1996)			PDB 1b3t, 1vhi
P03212	gi:140976	Glycoprotein L precursor, gp25, in gL–gH complex involved in viral fusion together with gB	BKRF2	M		$\alpha\beta\gamma$	UL1	Hutt-Fletcher (2005), Spear & Longnecker (2003)			
P12888	gi:137034	Uracil-DNA glycosylase	BKRF3	N		$\alpha\beta\gamma$	UL2	Winters & Williams (1993)	10	§	Publication in preparation
P30117	gi:267499	Tegument phosphoprotein	BKRF4	T		γ		Johannsen <i>et al.</i> (2004)			
P03200	gi:138368	Envelope glycoprotein gp350 (gp340) initial cell binding through complement receptor 2 (CR2, CD21)	BLLF1	M		$\alpha\gamma$		Spear & Longnecker (2003), Hutt-Fletcher (2005)			

Table 1 (continued)

Accession	NCBI	Function	ORF EBV	F	S	Occ.	Homologue	Reference	N	St	Comment
AAA45880	gi:330361	Envelope glycoprotein gp220, obtained through in-frame splicing from BLLF1	BLLF1b	M	×	$\alpha\gamma$		Beisel <i>et al.</i> (1985)			
P03199	gi:140999	Unknown function	BLLF2	U		γ					
P03195	gi:118952	dUTP pyrophosphatase, dUTPase	BLLF3	N		$\alpha\gamma$	UL50	Sommer <i>et al.</i> (1996), Tarbouriech <i>et al.</i> (2005)	2	§	PDB 2bsy, 2bt1
P03196	gi:141001	Membrane glycoprotein gN, part of the gM–gN complex, part of the envelope–tegument interaction	BLRF1	M		$\beta\gamma$	hcmvUL73	Lake & Hutt-Fletcher (2004)			
P03197	gi:141002	Tegument protein	BLRF2	T		γ		Johannsen <i>et al.</i> (2004)		‡‡	Soluble
P12977	gi:119112	EBNA-3A nuclear protein (EBNA-3), (–) effect on transactivator EBNA-2 and cell cycle, essential for immortalization	BLRF3/ BERF1	L	×			Kieff (1996), Waltzer <i>et al.</i> (1996), Young & Rickinson (2004)	1	†	No expression
P03191	gi:119098	Processivity factor, sliding clamp, early antigen protein D (EA-D, polymerase accessory protein)	BMRF1	R		$\alpha\beta\gamma$	UL42	Tsurumi <i>et al.</i> (1993)			
P03192	gi:141066	Receptor for cellular integrins, needed for infection of epithelial cells, 10 TM helices, RGD motif	BMRF2	M		γ		Modrow <i>et al.</i> (1992), Tugizov <i>et al.</i> (2003)			
P03230	gi:126373	Latent membrane protein 1 (LMP-1), interferes with signalling, TRAF-binding through CTAR1 and 2, essential for immortalization	BNLF1	L	×	γ		Young & Rickinson (2004)			
B28918	gi:23893667	Potential membrane protein	BNLF2a	U				Rivailler <i>et al.</i> (2002), Farrell (2005)			
A28918	gi:7460898	Potential gp141	BNLF2b	U				Rivailler <i>et al.</i> (2002), Farrell (2005)			
P03179	gi:139165	Major tegument protein (MTP), viral surface protein involved in B lymphocyte binding	BNRF1	T		γ		Johannsen <i>et al.</i> (2004), Lopez <i>et al.</i> (2005)			
P03189	gi:136893	Large tegument protein-binding protein (LTPBP)	BOLF1	T		$\alpha\beta\gamma$	UL37	Johannsen <i>et al.</i> (2004)			
P03187	gi:139172	Minor capsid protein-binding protein (mCP-BP), Triplex protein HSV-1 VP19C homologue	BORF1	C		γ	UL38	Johannsen <i>et al.</i> (2004)			
P03190	gi:132602	Ribonucleoside reductase, large 140 kDa subunit	BORF2	N		$\alpha\beta\gamma$	UL39	Kieff (1996)	2	†	No expression
P03186	gi:135574	Large tegument protein (LTP)	BPLF1	T		$\alpha\beta\gamma$	UL36	Johannsen <i>et al.</i> (2004)			
P03209	gi:115130	Transcription activator, R transactivator, Rta, dimeric (TAF50)	BRLF1	S		γ		Hardwick <i>et al.</i> (1988), Gruffat <i>et al.</i> (1990)	3	‡	Insoluble
Not in DB	Not in DB	Spliced BRLF1-BZLF1 protein, repressor of BZLF1, RAZ	BRLF1/ BZLF1	S	×			Manet <i>et al.</i> (1989), Furnari <i>et al.</i> (1994), Segouffin <i>et al.</i> (1996)			
Not in DB	Not in DB	Hypothetical BRLF1-BZLF1 splice variant	BRLF1/ BZLF1b	S	×			Farrell (2005)			
P03207	gi:141395	Enhancement of the induction of the lytic cycle	BRRF1	S		γ		Segouffin-Cariou <i>et al.</i> (2000), Hong <i>et al.</i> (2004)	1	†	No expression
P03210	gi:141396	Tegument protein, unknown function	BRRF2	T		γ		Johannsen <i>et al.</i> (2004)			
P03193	gi:136937	Primase, subunit of the helicase–primase complex, part of the replication machinery	BSLF1	R		$\alpha\beta\gamma$	UL52	Yokoyama <i>et al.</i> (1999)		†	No expression
Q04360	gi:1708410	mRNA-export factor (EB2, Mta,SM), mRNA splicing, interaction with human Spen proteins, IE63 (ICP27) homologue	BSLF2/ BMLF1	S	×	$\alpha\beta\gamma$	UL54	Boeckmann <i>et al.</i> (2003), Hiriart <i>et al.</i> (2005), Swaminathan (2005)	6	¶	In crystallization
P03194	gi:141432	Palmitoylated tegument protein (PalmP)	BSRF1	T		$\alpha\beta\gamma$	UL51	Johannsen <i>et al.</i> (2004)			
P30119	gi:267575	Capsid maturation, capsid-associated	BTRF1	C		$\alpha\gamma$	UL21	Wagenaar <i>et al.</i> (2001)			
P03233	gi:136861	Portal plug (EC-RF2), capsid-associated tegument protein, seals DNA inside capsid	BVRF1	C		$\alpha\beta\gamma$	UL25	Sheaffer <i>et al.</i> (2001), Johannsen <i>et al.</i> (2004)			

research papers

Table 1 (continued)

Accession	NCBI	Function	ORF EBV	F	S	Occ.	Homologue	Reference	N	St	Comment
P03234	gi:139231	Protease/assemblin	BVRF2	C		$\alpha\beta\gamma$	UL26	Buisson <i>et al.</i> (2002)	2	†	PDB 1o6e
P03177	gi:1170666	Thymidine kinase	BXLF1	N		$\alpha\gamma$	UL32	Littler <i>et al.</i> (1986)	3	‡	Insoluble
P03231	gi:138312	Glycoprotein gp85, gH, part of gHgLgp42, fusion	BXLF2	M		$\alpha\beta\gamma$	UL22	Spear & Longnecker (2003), Hutt-Fletcher (2005)			
P03232	gi:136852	Nucleoprotein	BXRF1	U		$\alpha\beta\gamma$	UL24	Pearson & Coen (2002)	2	‡	Insoluble
P12978	gi:119111	EBNA-2 nuclear protein, transcription factor, interacts with RBPJ-K, essential for immortalization	BYRF1	L	×			Henkel <i>et al.</i> (1994), Waltzer <i>et al.</i> (1994), Young & Rickinson (2004)			
P03206	gi:115196	Trans-activator ZEBRA, origin binding protein (EB1, Zta), bZip similar to CCAAT/enhancer binding protein α	BZLF1	S	×	γ		Chevallier-Greco <i>et al.</i> (1986), Rooney <i>et al.</i> (1988), Giot <i>et al.</i> (1991), Petosa <i>et al.</i> (2006)			PDB 2c9n, 2c9l
QQBE27	gi:73982	Function unknown, BZLF1 splice variant	BZLF1b	U	×			Countryman & Miller (1985), Farrell (2005)			
P03205	gi:141578	gp42, MHC class II binding protein, part of gHgLgp42 complex	BZLF2	M		γ		Mullen <i>et al.</i> (2002), Rensing <i>et al.</i> (2003)			PDB 1kg0
P03235	gi:140616	EC-RF4 (ECRF4) protein	ECRF4	U				Rivailler <i>et al.</i> (2002), Rensing <i>et al.</i> (2003)			
Q8AZJ5	gi:23893655	Protein LF1, contains a dUTPase like domain, γ -herpes ORF10 family	LF1	U		γ		Rivailler <i>et al.</i> (2002), Davison & Stow (2005), Farrell (2005)			
Q99306	gi:23893654	Protein LF2, contains a dUTPase like domain, γ -herpes ORF11 family	LF2	U		γ		Rivailler <i>et al.</i> (2002), Davison & Stow (2005), Farrell (2005)			
Not in DB	Not in DB	RK-BARF0, interaction with notch	RK-BARF0	S	×			Fries <i>et al.</i> (1997), Kusano & Raab-Traub (2001)			
P13285	gi:126379	LMP-2A, interference with protein kinase signalling, gene terminal protein, essential for immortalization		L	×			Kieff (1996), Young & Rickinson (2004)			
Q8AZK9	gi:23893578	LMP-2B, negative regulator of LMP-2A		L	×			Kieff (1996), Young & Rickinson (2004)			

† Cloned. ‡ Expressed. § Structure solved. ¶ Purified protein. †† Crystals. ††† Soluble protein.

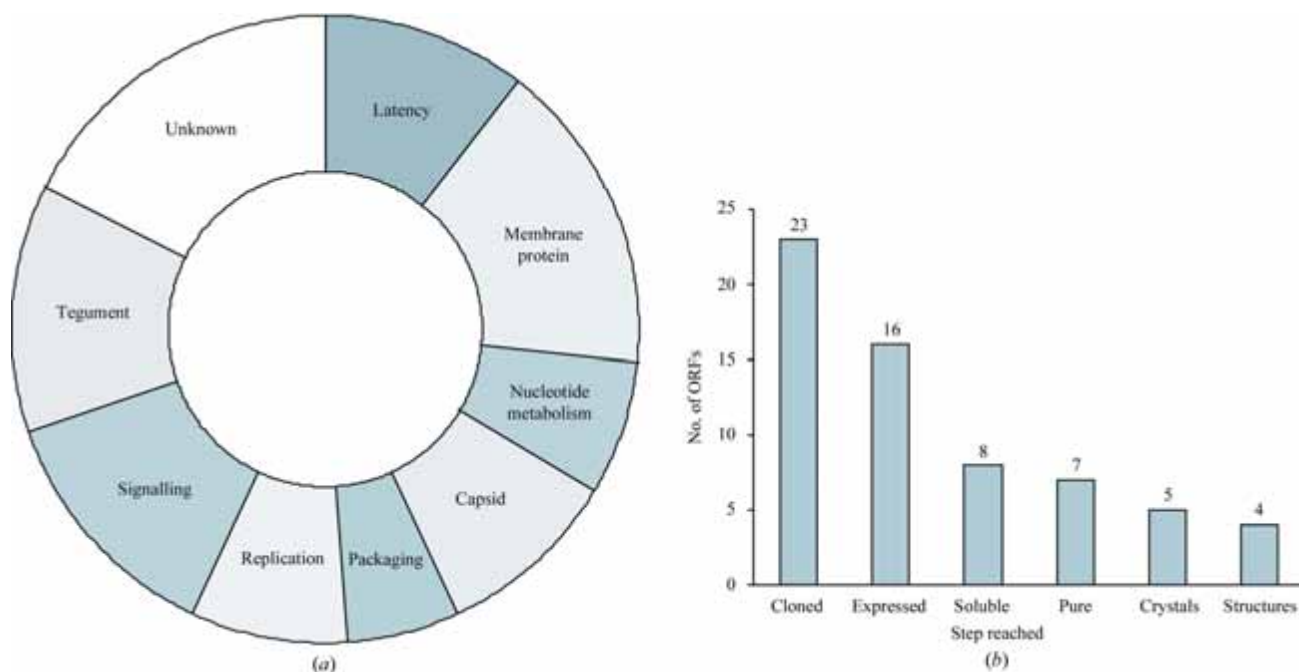


Figure 1
 (a) Classification of the EBV proteins according to function. (b) Outcome for the proteins entering into the structural genomics project.

Table 2
Crystallographic results.

Protein reference	EBV ORF, PDB code	Construct†	Inhibitor	Protein buffer	Precipitant	Space group	Unit-cell parameters (Å)	Resolution (Å), method‡	Structure
Protease§	BVRF2, 1o6e	GSHMAS + (2–235) + RS	Diisopropyl-fluorophosphate	100 mM NaCl, 20 mM Tris–HCl pH 7.5, 1 mM EDTA, 10 mM β-mercaptoethanol	1.4 M sodium formate, 100 mM sodium acetate pH 4.6, 5 mM EDTA	<i>P</i> 3 ₁ 21, twinned	<i>a</i> = <i>b</i> = 52.8, <i>c</i> = 330.5	2.3, MR	Fig. 2(a)
dUTPase¶	BLLF3, 2bsy	GAMGSGIP + (1–278)	dUMP	250 mM NaCl, 20 mM imidazole, 10 mM MgCl ₂ , 20 mM HEPES pH 7.5	20% PEG 3350, 200 mM LiSO ₄ , 100 mM Tris–HCl pH 8.5	<i>P</i> 2 ₁ 2 ₁	<i>a</i> = 56.7, <i>b</i> = 55.8, <i>c</i> = 81.1	1.5, SAD	Fig. 2(b)
	2bt1	GAMGSGIP + (1–278)	α,β-Imino-dUTP	250 mM NaCl, 20 mM imidazole, 10 mM MgCl ₂ , 20 mM HEPES pH 7.5	3 M NaCl, 100 mM Bis-Tris–HCl pH 6.5	<i>P</i> 622	<i>a</i> = 146.6, <i>c</i> = 77.1	2.7, MR	
BARF1††	BARF1, 2ch8	(21–221)	—	100 mM NaCl, 20 mM HEPES–NaOH pH 7.5	1 M (NH ₄) ₂ SO ₄ , 2% PEG 3350, 100 mM Bis-Tris–HCl pH 6.0	<i>H</i> 3, twinned	<i>a</i> = <i>b</i> = 179.3, <i>c</i> = 95.7	2.3, SAD	Fig. 2(c)
Uracil DNA glycosylase‡‡	BKRF3, N/A	GAM + (25–255)	UGI of bacteriophage, PBS-2	100 mM NaCl, 20 mM Tris pH 7.5, 10 mM DTT	20% PEG 3350, 50 mM NH ₄ Cl	<i>C</i> 222 ₁	<i>a</i> = 62.8, <i>b</i> = 83.5, <i>c</i> = 273.4	2.3, MR	Fig. 2(d)

† Extra residues arising from cloning are given; the residues of the viral ORF are indicated in parentheses. ‡ Method used for structure determination. MR, molecular replacement; SAD, single-wavelength anomalous dispersion on a heavy-atom derivative. § Buisson *et al.* (2002). ¶ Tarbouriech *et al.* (2005). †† Tarbouriech *et al.* (2006). ‡‡ Unpublished.

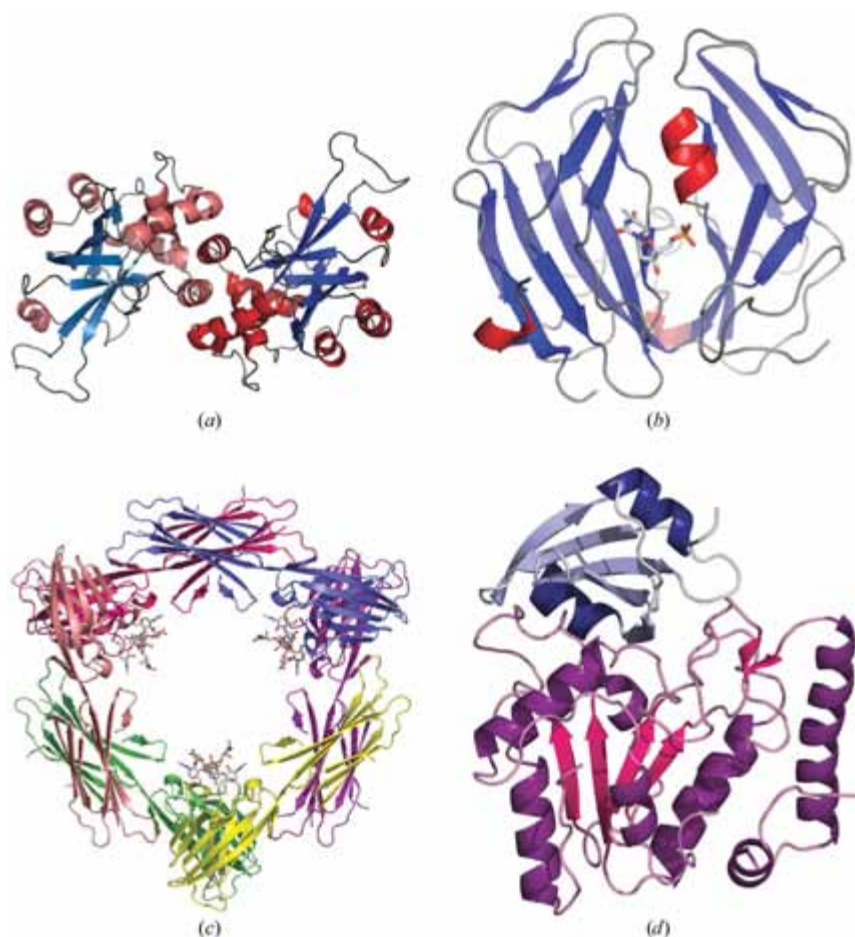


Figure 2
Protein structures (see Table 2).

EBV is composed of an inner capsid that contains the viral double-stranded DNA genome, surrounded by a membrane carrying various surface glycoproteins. Tegument fills the space between the capsid and the membrane. During the latent stage of infection in B-lymphocytes a very limited set of proteins is expressed. The viral DNA forms a circular episome which is associated with the cellular chromosomes and is replicated by the cellular machinery during cell division. After activation, the infection can switch to the lytic cycle, leading to the expression of the full set of viral proteins and production of viral particles. This complex lifestyle utilizes about 86 predicted proteins (Table 1), meaning that EBV has one of the largest genomes of human viruses. The principal viral functions are receptor binding and cell entry, maintenance of latency, nucleotide metabolism, DNA replication and packaging and capsid assembly (Fig. 1*a*, Table 1). EBV also codes for a number of immune-modulators. Some little-studied proteins shuttle viral particles from the nucleus, the site of viral replication, to the extracellular space and a number of proteins still have no assigned function. With the aim of obtaining insight into

the protein functions and in order to identify new drug targets, SPINE (Structural Genomics In Europe) included the structural proteomics of herpesviruses in workpackage 9 (human pathogen targets; see Fogg *et al.*, 2006) and here we report our contribution to this, namely the analysis of a cohort of 23 EBV proteins.

2. Project design, methods and results

2.1. Target annotation

The project included a major continued effort in protein annotation since the information available in databases [principally SWISS-PROT (Boeckmann *et al.*, 2003) and VIDA (Alba, 2002)] was rather incomplete, in particular for spliced reading frames, or no longer up to date. Our annotation is given in Table 1 with the results on the SPINE targets, together with as much bibliographic information as possible. We identified 86 proteins encoded by the EBV genome. The existence of a few of these remains questionable, owing to alternative splicing. The function of 15 proteins is unknown and could not be inferred from sequence homology or bibliographic information (Table 1, Fig. 1*a*). In general, little is known about the role of the tegument proteins, even though they have been recently localized unambiguously in the virus particle (Johannsen *et al.*, 2004).

2.2. Target selection

As one aim of the project was to obtain structures of potential new drug targets, we first targeted proteins with known enzymatic activity (11 ORFs; Table 1). Next, proteins were ranked according to several predicted properties. Firstly, they were given priority if they had a high predicted secondary structure by the NSP@ server (Deleage *et al.*, 1997), small size and a high stability index according to the ExpASY *ProtParam* tool (Gasteiger *et al.*, 2005). Known membrane proteins, surface glycoproteins and proteins involved in the packaging mechanism were omitted in order to avoid redundancy with other teams of the SPINE project. Furthermore, we selected against components of known multi-protein assemblies and eliminated proteins containing transmembrane domains using the *DAS* software (Cserzo *et al.*, 1997) and the TMHMM server (Krogh *et al.*, 2001) available from the ExpASY web site.

2.3. Cloning and protein production

We opted for a small-scale parallel approach using simple restriction-based cloning into a vector containing a tobacco etch virus protease (TEV) cleavable N-terminal His₆ tag, allowing the targets to be closely followed through purification and crystallization.

2.3.1. Cloning and expression tests. The selected genes were cloned by PCR amplification of EBV DNA extracted from the B95-8 cell line using primers introducing restriction sites at the 5' and 3' ends of the gene and ligated into the pPROEX-HTb plasmid (Invitrogen) using standard methods. The PCR products were cloned between *Nco*I or *Bam*HI sites as a first choice, *Eco*RI as a second choice and *Hind*III or

*Xho*I sites. The ligated products were directly transformed into *Escherichia coli* BL21(DE3) GOLD cells (Invitrogen), which were used for both DNA preparation for sequencing and small-scale expression tests. DNA preparation was performed either manually or automatically on the RoBioMol platform at the IBS (Grenoble). Small-scale expression tests used 1 ml LB media inoculated with single colonies. Protein production was induced with 0.5 mM isopropyl β -D-thiogalactoside and continued for 3–5 h at 310 and 303 K and overnight at 296 and 289 K. Cells were lysed with BugBuster (Novagen). Protein solubility was checked on SDS-PAGE by loading both the cell extract and the soluble fraction after centrifugation at 18 000g for 20 min. If soluble protein was not detected, the *E. coli* strains Rosetta, Origami, BL21 (DE3) STAR (Invitrogen), C41 and C43 (Avidis) were tested with overnight induction at 289 K.

2.3.2. Protein expression and purification. Proteins were produced using either classical LB or an auto-inducible medium (Studier, 2005). Cells were lysed by sonication and cell debris was removed by centrifugation at 30 000g for 30 min. The supernatant was loaded onto an Ni-NTA (Qiagen) column equilibrated with 20 mM Tris-HCl pH 7.5, 100 mM NaCl and 20 mM imidazole, washed using the same buffer containing 50 mM imidazole and eluted at an imidazole concentration of 500 mM. After buffer exchange back to the loading buffer, the protein was incubated overnight at room temperature with a ratio of 1/100 of recombinant His-tagged TEV protease. This was loaded again on an Ni-NTA column and the eluate of this column was concentrated by ultra-filtration and loaded onto a Superdex S75 or S200 gel-filtration column (GE/Amersham), depending on the protein size.

2.3.3. Refolding. When good expression levels of insoluble protein were obtained, refolding was attempted. Following large-scale production with induction at 310 K for 4–5 h, the protein was purified from inclusion bodies using buffers supplemented with 8 M urea. After purification and concentration to 5 mg ml⁻¹, a 20-fold dilution in refolding buffers was followed by 24 h incubation at 277 K. Refolding buffers varied in salt concentration (0 or 500 mM NaCl), pH (Bis-Tris-HCl pH 5, Tris-HCl pH 7 or Tris-HCl pH 9) or divalent cation contents (10 mM EDTA or 5 mM CaCl₂/5 mM MgCl₂), leading to 12 different basic conditions. Samples were centrifuged for 15 min at 16 000g and supernatants were assayed for soluble protein either by ammonium sulfate precipitation and SDS-PAGE or by concentration followed by gel filtration.

2.4. Crystallization

Proteins were analyzed by dynamic light scattering (Protein Solutions) prior to crystallization. Crystallization screening was carried out at the High Throughput Crystallization Laboratory of the EMBL Grenoble Outstation (HTX Lab). Typically, 576 conditions were tested per sample using a PixSys4200 robot (Cartesian) and the vapour-diffusion method in CrystalQuick (Greiner Bio-One) 96-well sitting-drop crystallization plates with square wells. Drops contained 100 nl protein solution and 100 nl buffer solution. Crystal

Screen, Crystal Screen II, PEG/Ion Screen, Crystal Screen Lite, Natrix, Membfac, Grid Screens and Index Screen (Hampton Research) were used as well as Clear Strategy Screens (Molecular Dimensions). Crystallization plates were stored and automatically imaged by a CrystalMation robot (RoboDesign) including a RoboIncubator and a Minstrel III module. Successful crystallizations were reproduced and refined manually using 1 + 1 μ l hanging drops.

3. Discussion

A significant bottleneck in the structure-determination pipeline for EBV proteins was obtaining levels of protein expression (16/23) and soluble protein sufficient for crystallization (7/23; Fig. 1*b*, Table 1), although the success rate at crystallization was unexpectedly high (5/7). Surprisingly, changing the bacterial strain or expression temperature did not increase soluble expression levels compared with our standard protocol using BL21 cells at 303 K. A bioinformatics analysis using secondary-structure prediction (Deleage *et al.*, 1997) and *ClustalW*-based alignments (Thompson *et al.*, 1994) only rarely suggested obvious truncations. Perhaps as a consequence of this, modification of the constructs by N-terminal and C-terminal truncations, although attempted for the majority of the studied reading frames (Table 1), was successful in only one case, uracil-DNA glycosylase (UNG), where deletion of the N-terminal 24 residues increased expression levels and led to diffraction-grade crystals. The deleted residues may contain a nuclear localization signal based on sequence identity with human UNG2 (Otterlei *et al.*, 1998). Seven soluble proteins were expressed in *E. coli*: the EBV protease domain, dUTPase, uracil-DNA glycosidase, BHRF1, BLRF2, BDLF1 and a fragment of BMLF1 (EB2), but the last three proteins were unstable after purification. In the case of the dUTPase, the low solubility of the protein necessitated intensive optimization of purification and crystallization conditions (Tarbouriech *et al.*, 2005). Work on the EBV protease domain predated the SPINE project (Buisson *et al.*, 2002). Structural determination of BHRF1 was abandoned despite the existence of small crystals when an NMR structure was reported (Huang *et al.*, 2003). BARSF1 was obtained through an external collaboration and expressed in eukaryotic cells (de Turenne-Tessier *et al.*, 2005) before entering our structure-determination pipeline. Protein purification using an N-terminal His₆ tag together with a TEV protease cleavage site, sometimes including size-exclusion chromatography, reliably produced pure protein for crystallization. In line with other unpublished results in SPINE, refolding from inclusion bodies failed to produce soluble protein from any of the 12 cases. However, we subsequently tested expression in insect cells using baculovirus and obtained three soluble proteins from six ORFs. Overall in SPINE the experience has been that viral proteins tend to be more difficult to express in bacterial systems than prokaryotic proteins (*e.g.* 27% of viral proteins were expressed in *E. coli* compared with 33–77% of some bacterial proteins; Fogg *et al.*, 2006). It is clear that eukaryotic expression is a real alternative for difficult viral proteins.

Crystallization screening used 200 nl sitting drops dispensed robotically and achieved a very high success rate; however, for proteins except BARSF1 this required the addition of enzyme inhibitors (Table 2). Crystallographic details for each EBV structure are given in Table 2 and further details on the structure determinations and refinement have been or will be published elsewhere.

The study described here highlights the particular problems associated with the application of pipeline technologies to difficult proteins. In this case, EBV proteins were poorly suited to bacterial expression systems and success was dependent on a much more individual approach to protein production. Although a simple pipeline approach with standard protocols is unlikely to be universally applicable for structural determination, pipeline components can be extremely effective, exemplified here by the high-throughput nanolitre crystallization platform. This major breakthrough in crystallization screening undoubtedly contributed to the high crystallization rates observed with the soluble EBV proteins.

This work was undertaken as part of the European Union Framework Programme 'Quality of Life and Management of Living Resources', Integrated Project SPINE (Structural Proteomics In Europe), contract No. QLG2-CT-2002-00988. We thank Jean-Marie Seigneurin for providing DNA from the B95.8 cell line and support for the project, Florine Dupeux, Benoit Gallet, José-Antonio Marquez, Martin Rower and Thierry Vernet for the operation of high-throughput facilities at the Grenoble Partnership for Structural Biology (PSB), and Lucy Freeman and Lucie Rivail for work on individual proteins. We are grateful to Henri Gruffat and Tadamas Ooka for help with the genome annotation.

References

- Alba, M. M. (2002). *VIDA Database*. http://www.biochem.ucl.ac.uk/bsm/virus_database/.
- Altschul, S. F., Madden, T. L., Schaffer, A. A., Zhang, J., Zhang, Z., Miller, W. & Lipman, D. J. (1997). *Nucleic Acids Res.* **25**, 3389–3402.
- Beard, P. M. & Baines, J. D. (2004). *Virology*, **324**, 475–482.
- Beisel, C., Tanner, J., Matsuo, T., Thorley-Lawson, D., Kezdy, F. & Kieff, E. (1985). *J. Virol.* **54**, 665–674.
- Bellows, D. S., Howell, M., Pearson, C., Hazlewood, S. A. & Hardwick, J. M. (2002). *J. Virol.* **76**, 2469–2479.
- Bochkarev, A., Barwell, J. A., Pfuetzner, R. A., Bochkareva, E., Frappier, L. & Edwards, A. M. (1996). *Cell*, **84**, 791–800.
- Boeckmann, B., Bairoch, A., Apweiler, R., Blatter, M. C., Estreicher, A., Gasteiger, E., Martin, M. J., Michoud, K., O'Donovan, C., Phan, I., Pilbout, S. & Schneider, M. (2003). *Nucleic Acids Res.* **31**, 365–370.
- Borza, C. M. & Hutt-Fletcher, L. M. (1998). *J. Virol.* **72**, 7577–7582.
- Bowman, B. R., Baker, M. L., Rixon, F. J., Chiu, W. & Quijcho, F. A. (2003). *EMBO J.* **22**, 757–765.
- Buisson, M., Hernandez, J. F., Lascoux, D., Schoehn, G., Forest, E., Arlaud, G., Seigneurin, J. M., Ruigrok, R. W. & Burmeister, W. P. (2002). *J. Mol. Biol.* **324**, 89–103.
- Cabras, G., Decaussin, G., Zeng, Y., Djennaoui, D., Melouli, H., Brouilly, P., Bouguermouh, A. M. & Ooka, T. (2005). *J. Clin. Virol.* **34**, 26–34.
- Chang, Y. E., Poon, A. P. & Roizman, B. (1996). *J. Virol.* **70**, 3938–3946.

- Cheng, Y. C., Chen, J. Y., Hoffmann, P. J. & Glaser, R. (1980). *Virology*, **100**, 334–338.
- Chevallier-Greco, A., Manet, E., Chavrier, P., Mosnier, C., Daillie, J. & Sergeant, A. (1986). *EMBO J.* **5**, 3243–3249.
- Coen, D. M. & Schaffer, P. A. (2003). *Nature Rev. Drug Discov.* **2**, 278–288.
- Countryman, J. & Miller, G. (1985). *Proc. Natl Acad. Sci. USA*, **82**, 4085–4089.
- Cserzo, M., Wallin, E., Simon, I., von Heijne, G. & Elofsson, A. (1997). *Protein Eng.* **10**, 673–676.
- Davison, A. J. & Stow, N. D. (2005). *J. Virol.* **79**, 12880–12892.
- Decaussin, G., Leclerc, V. & Ooka, T. (1995). *J. Virol.* **69**, 7309–7314.
- Deleage, G., Blanchet, C. & Geourjon, C. (1997). *Biochimie*, **79**, 681–686.
- Farell, P. J. (2005). *Epstein–Barr Virus*, edited by E. S. Robertson, pp. 263–287. Norwich: Caister Academic Press.
- Fogg, M. J. *et al.* (2006). *Acta Cryst.* **D62**, 1196–1207.
- Fries, K. L., Sculley, T. B., Webster-Cyriaque, J., Rajadurai, P., Sadler, R. H. & Raab-Traub, N. (1997). *J. Virol.* **71**, 2765–2771.
- Furnari, F. B., Zacny, V., Quinlivan, E. B., Kenney, S. & Pagano, J. S. (1994). *J. Virol.* **68**, 1827–1836.
- Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S., Wilkins, M. R., Appel, R. D. & Bairoch, A. (2005). *The Proteomics Protocols Handbook*, edited by J. M. Walker, pp. 571–607. Totowa, NJ, USA: Humana Press.
- Gilligan, K. J., Rajadurai, P., Lin, J. C., Busson, P., Abdel-Hamid, M., Prasad, U., Tursz, T. & Raab-Traub, N. (1991). *J. Virol.* **65**, 6252–6259.
- Giot, J. F., Mikaelian, I., Buisson, M., Manet, E., Joab, I., Nicolas, J. C. & Sergeant, A. (1991). *Nucleic Acids Res.* **19**, 1251–1258.
- Gong, M., Ooka, T., Matsuo, T. & Kieff, E. (1987). *J. Virol.* **61**, 499–508.
- Gonnella, R., Farina, A., Santarelli, R., Raffa, S., Feederle, R., Bei, R., Granato, M., Modesti, A., Frati, L., Delecluse, H. J., Torrisi, M. R., Angeloni, A. & Faggioni, A. (2005). *J. Virol.* **79**, 3713–3727.
- Gruffat, H., Manet, E., Rigolet, A. & Sergeant, A. (1990). *Nucleic Acids Res.* **18**, 6835–6843.
- Hardwick, J. M., Lieberman, P. M. & Hayward, S. D. (1988). *J. Virol.* **62**, 2274–2284.
- Henkel, T., Ling, P. D., Hayward, S. D. & Peterson, M. G. (1994). *Science*, **265**, 92–95.
- Hiriart, E., Gruffat, H., Buisson, M., Mikaelian, I., Keppler, S., Meresse, P., Mercher, T., Bernard, O. A., Sergeant, A. & Manet, E. (2005). *J. Biol. Chem.* **280**, 36935–36945.
- Hitt, M. M., Allday, M. J., Hara, T., Karran, L., Jones, M. D., Busson, P., Tursz, T., Ernberg, I. & Griffin, B. E. (1989). *EMBO J.* **8**, 2639–2651.
- Hong, G. K., Delecluse, H. J., Gruffat, H., Morrison, T. E., Feng, W. H., Sergeant, A. & Kenney, S. C. (2004). *J. Virol.* **78**, 4983–4992.
- Hsu, D. H., de Waal Malefyt, R., Fiorentino, D. F., Dang, M. N., Vieira, P., de Vries, J., Spits, H., Mosmann, T. R. & Moore, K. W. (1990). *Science*, **250**, 830–832.
- Huang, Q., Petros, A. M., Virgin, H. W., Fesik, S. W. & Olejniczak, E. T. (2003). *J. Mol. Biol.* **332**, 1123–1130.
- Hutt-Fletcher, L. M. (2005). *Epstein–Barr Virus*, edited by E. S. Robertson, pp. 359–378. Norwich: Caister Academic Press.
- Hwang, J. S. & Bogner, E. (2002). *J. Biol. Chem.* **277**, 6943–6948.
- Johannsen, E., Luftig, M., Chase, M. R., Weicksel, S., Cahir-McFarland, E., Illanes, D., Sarracino, D. & Kieff, E. (2004). *Proc. Natl Acad. Sci. USA*, **101**, 16286–16291.
- Kieff, E. (1996). *Fields Virology*, edited by B. N. Fields, pp. 2343–2396. Philadelphia: Lippincott–Raven.
- Krogh, A., Larsson, B., von Heijne, G. & Sonnhammer, E. L. (2001). *J. Mol. Biol.* **305**, 567–580.
- Kusano, S. & Raab-Traub, N. (2001). *J. Virol.* **75**, 384–395.
- Lake, C. M. & Hutt-Fletcher, L. M. (2000). *J. Virol.* **74**, 11162–11172.
- Lake, C. M. & Hutt-Fletcher, L. M. (2004). *Virology*, **320**, 99–106.
- Laux, G., Freese, U. K. & Bornkamm, G. W. (1985). *J. Virol.* **56**, 987–995.
- Littler, E., Zeuthen, J., McBride, A. A., Trost Sorensen, E., Powell, K. L., Walsh-Arrand, J. E. & Arrand, J. R. (1986). *EMBO J.* **5**, 1959–1966.
- Lopez, R., Urquiza, M., Patino, H., Suarez, J., Reyes, C., Patarroyo, M. A. & Patarroyo, M. E. (2005). *Biochimie*, **87**, 985–992.
- Mackett, M., Conway, M. J., Arrand, J. R., Haddad, R. S. & Hutt-Fletcher, L. M. (1990). *J. Virol.* **64**, 2545–2552.
- Manet, E., Gruffat, H., Trescol-Biemont, M. C., Moreno, N., Chambard, P., Giot, J. F. & Sergeant, A. (1989). *EMBO J.* **8**, 1819–1826.
- Marschall, M., Stein-Gerlach, M., Freitag, M., Kupfer, R., van den Bogaard, M. & Stamminger, T. (2002). *J. Gen. Virol.* **83**, 1013–1023.
- Marshall, W. L., Yim, C., Gustafson, E., Graf, T., Sage, D. R., Hanify, K., Williams, L., Fingerroth, J. & Finberg, R. W. (1999). *J. Virol.* **73**, 5181–5185.
- Modrow, S., Hoflacher, B. & Wolf, H. (1992). *Arch. Virol.* **127**, 379–386.
- Mullen, M. M., Haan, K. M., Longnecker, R. & Jardetzky, T. S. (2002). *Mol. Cell*, **9**, 375–385.
- Nealon, K., Newcomb, W. W., Pray, T. R., Craik, C. S., Brown, J. C. & Kedes, D. H. (2001). *J. Virol.* **75**, 2866–2878.
- Newcomb, W. W., Juhas, R. M., Thomsen, D. R., Homa, F. L., Burch, A. D., Weller, S. K. & Brown, J. C. (2001). *J. Virol.* **75**, 10923–10932.
- Otterlei, M., Haug, T., Nagelhus, T. A., Slupphaug, G., Lindmo, T. & Krokan, H. E. (1998). *Nucleic Acids Res.* **26**, 4611–4617.
- Paulsen, S. J., Rosenkilde, M. M., Eugen-Olsen, J. & Kledal, T. N. (2005). *J. Virol.* **79**, 536–546.
- Pearson, A. & Coen, D. M. (2002). *J. Virol.* **76**, 10821–10828.
- Petos, C., Morand, P., Baudin, F., Moulin, M., Artero, J. B. & Muller, C. W. (2006). *Mol. Cell*, **21**, 565–572.
- Raab-Traub, N. (2005). *Epstein–Barr Virus*, edited by E. S. Robertson, pp. 71–92. Norwich: Caister Academic Press.
- Ressing, M. E., van Leeuwen, D., Verreck, F. A., Gomez, R., Heemskerck, B., Toebes, M., Mullen, M. M., Jardetzky, T. S., Longnecker, R., Schilham, M. W., Ottenhoff, T. H., Neeffjes, J., Schumacher, T. N., Hutt-Fletcher, L. M. & Wiertz, E. J. (2003). *Proc. Natl Acad. Sci. USA*, **100**, 11583–11588.
- Reuven, N. B., Willcox, S., Griffith, J. D. & Weller, S. K. (2004). *J. Mol. Biol.* **342**, 57–71.
- Rickinson, A. & Kieff, E. (1996). *Fields Virology*, edited by B. N. Fields, Vol. 2, pp. 2397–2446. Philadelphia: Lippincott–Raven.
- Rivailler, P., Jiang, H., Cho, Y. G., Quink, C. & Wang, F. (2002). *J. Virol.* **76**, 421–426.
- Robertson, E. S., Ooka, T. & Kieff, E. D. (1996). *Proc. Natl Acad. Sci. USA*, **93**, 11334–11340.
- Rooney, C., Taylor, N., Countryman, J., Jenson, H., Kolman, J. & Miller, G. (1988). *Proc. Natl Acad. Sci. USA*, **85**, 9801–9805.
- Savva, C. G., Holzenburg, A. & Bogner, E. (2004). *FEBS Lett.* **563**, 135–140.
- Segouffin-Cariou, C., Farjot, G., Sergeant, A. & Gruffat, H. (2000). *J. Gen. Virol.* **81**, 1791–1799.
- Segouffin, C., Gruffat, H. & Sergeant, A. (1996). *J. Gen. Virol.* **77**, 1529–1536.
- Sheaffer, A. K., Newcomb, W. W., Gao, M., Yu, D., Weller, S. K., Brown, J. C. & Tenney, D. J. (2001). *J. Virol.* **75**, 687–698.
- Smith, R. F. & Smith, T. F. (1989). *J. Virol.* **63**, 450–455.
- Sommer, P., Kremmer, E., Bier, S., Konig, S., Zalud, P., Zeppezauer, M., Jones, J. F., Mueller-Lantzsch, N. & Grasser, F. A. (1996). *J. Gen. Virol.* **77**, 2795–2805.
- Spear, P. G. & Longnecker, R. (2003). *J. Virol.* **77**, 10179–10185.
- Strockbine, L. D., Cohen, J. I., Farrah, T., Lyman, S. D., Wagener, F., DuBose, R. F., Armitage, R. J. & Spriggs, M. K. (1998). *J. Virol.* **72**, 4015–4021.
- Studier, F. W. (2005). *Protein Expr. Purif.* **41**, 207–234.

- Swaminathan, S. (2005). *Epstein-Barr Virus*, edited by E. S. Robertson, pp. 631–649. Norwich: Caister Academic Press.
- Tarbouriech, N., Buisson, M., Seigneurin, J. M., Cusack, S. & Burmeister, W. P. (2005). *Structure*, **13**, 1299–1310.
- Tarbouriech, N., Ruggiero, F., de Turenne-Tessier, M., Ooka, T. & Burmeister, W. P. (2006). *J. Mol. Biol.* **359**, 667–678.
- Thompson, J. D., Higgins, D. G. & Gibson, T. J. (1994). *Nucleic Acids Res.* **22**, 4673–4680.
- Trus, B. L., Heymann, J. B., Nealon, K., Cheng, N., Newcomb, W. W., Brown, J. C., Kedes, D. H. & Steven, A. C. (2001). *J. Virol.* **75**, 2879–2890.
- Tsurumi, T., Kobayashi, A., Tamai, K., Daikoku, T., Kurachi, R. & Nishiyama, Y. (1993). *J. Virol.* **67**, 4651–4658.
- Tugizov, S. M., Berline, J. W. & Palefsky, J. M. (2003). *Nature Med.* **9**, 307–314.
- Turenne-Tessier, M. de, Jolicœur, P., Middeldorp, J. M. & Ooka, T. (2005). *Virus Res.* **109**, 9–18.
- Wagenaar, F., Pol, J. M., de Wind, N. & Kimman, T. G. (2001). *Vet. Res.* **32**, 47–54.
- Waltzer, L., Logeat, F., Brou, C., Israel, A., Sergeant, A. & Manet, E. (1994). *EMBO J.* **13**, 5633–5638.
- Waltzer, L., Perricaudet, M., Sergeant, A. & Manet, E. (1996). *J. Virol.* **70**, 5909–5915.
- Wei, M. X. & Ooka, T. (1989). *EMBO J.* **8**, 2897–2903.
- Winters, T. A. & Williams, M. V. (1993). *Virology*, **195**, 315–326.
- Yokoyama, N., Fujii, K., Hirata, M., Tamai, K., Kiyono, T., Kuzushima, K., Nishiyama, Y., Fujita, M. & Tsurumi, T. (1999). *J. Gen. Virol.* **80**, 2879–2887.
- Young, L. S. & Rickinson, A. B. (2004). *Nature Rev. Cancer*, **4**, 757–768.

Chapitre 7

Résultats supplémentaires

7.1 Recherche d'une Thymidylate Synthase (TS) virale

TROIS protéines intervenant dans le métabolisme du dTMP ont été découvertes dans le génome d'EBV (Tarbouriech *et al.*, 2006). Or, il reste environ 15 protéines dont la fonction est inconnue. De plus, les protéines virales peuvent comporter plusieurs fonctions à l'intérieur d'une même protéine. Aucune Thymidylate Synthase n'a été décrite chez EBV. Pour rechercher la présence de cette enzyme, plusieurs stratégies ont été envisagées :

- Confronter la séquence d'une TS connue (Eg. : Celle de KSHV, un autre γ -herpesvirus) à la base de données de protéine virale.
- Confronter la signature¹ des TS à la base de données.

La référence swissprot (Bairoch *et al.*, 2004) de la séquence de la TS de KSHV est **P90463**.

C'est une protéine de 38.5 KDa et de 337 AA. Le serveur **Interproscan** (Mulder *et al.*, 2005) accessible à partir du site **Expasy** (Gasteiger *et al.*, 2003) permet de faire une recherche de motif. Le motif TS correspondant à la signature du site actif, dont le code d'accès **Prosite** (Hulo *et al.*, 2006) est **PS00091**, comporte entre 27 et 34 acides-aminés (Figure 7.1 page 78), il est retrouvé dans 171 séquences.

La recherche d'une protéine assurant le rôle d'une Thymidylate Synthase ne fut pas fructueuse. La

¹La signature d'une protéine est la séquence consensus qui correspond aux résidus conservés que l'on retrouve dans la famille de protéines exprimant une fonction donnée.

R-x(2)-[LIVMT]-x(2,3)-[FWY]-[QNYDI]-x(8,13)-[LVES]-x-P-G-[HAVMLC]-x(3)-[QMTLHD]-[FYWL]-x(0,1)-[LV

FIG. 7.1: Signature du site actif de la Thymidylate Synthase. “x” correspond à n’importe quel acide-aminé. Le nombre entre parenthèse correspond à la répétition d’un acide-aminé. Une lettre entre crochet signifie qu’un acide-aminé de cette série est possible à cet endroit de la séquence.

meilleur E-value obtenue est de **1.9**, elle correspond à la séquence d’une sous-unité de l’hélicase-primase. Bien entendu, un tel résultat ne correspond à rien de pertinent. Il représente l’alignement de 9 acides-aminés non consécutifs, et résulte vraisemblablement du hasard.

7.2 Expression de protéines virales : *E.coli* n’est pas le système d’expression idéal

In-vivo, les enzymes du virus d’Epstein-Barr sont exprimées soit dans des lymphocytes B soit dans certaines cellules épithéliales (voir Tropisme page 33), or, le système d’expression le plus simple et le plus largement utilisé, dans le cadre d’étude structurale, reste *E.coli*. Une étude structurale par cristallographie aux rayons-X nécessite la croissance de cristaux de protéines², or l’obtention de cristaux repose sur une méthode empirique de criblage d’un grand nombre de conditions. Ceci nécessite l’utilisation d’une grande quantité de protéine. Un minimum d’une centaine de microlitres à une concentration d’environ 10 mg.ml⁻¹ est une bonne base de départ. *E.coli* permet, en général, la production d’une grande quantité de protéines pour de faibles coûts, néanmoins, certaines protéines ne s’expriment pas ou très peu, ou alors, leur solubilité est extrêmement faible. Ce fut le cas de la TK et de la RNR. Différentes souches d’*E.coli* furent testées (voir Tarbouriech *et al.* (2006)) aucune d’entre elles n’améliora la production des enzymes. La solution de ce problème serait l’utilisation de système d’expression plus proche de l’hôte naturel du virus. L’utilisation d’un système d’expression eucaryote a permis d’obtenir de bons résultats sur d’autres protéines d’EBV. L’**Alcaline Exonucléase** sur laquelle nous reviendrons au chapitre “*Etude structurale préliminaire de l’Exonucléase du virus d’Epstein-Barr*” page 145, a été exprimée dans le système baculovirus/cellules d’insectes. La

²Pour plus de détail sur les techniques permettant de faire pousser des cristaux, se référer à la partie sur la cristallogénèse de l’Alcaline Exonucléase, page 151.

structure de la protéine oncogène **BARF1** (Figure 2.8, page 42) a été obtenue à partir d'une protéine exprimée en cellule de mammifère infecté par un adénovirus recombinant (Tarbouriech *et al.*, 2006). L'expression en cellule eucaryote est un système beaucoup plus proche de l'hôte naturel d'EBV et a permis de résoudre les problèmes d'expression et de solubilité rencontrés chez *E. coli*. La limite de ces systèmes d'expression tient à leur coût et à la faible quantité de protéines qu'ils permettent de produire. Néanmoins, la robotisation des étapes de criblage permet de travailler avec moins de matériel et permet d'envisager le travail sur de telles cibles.

Troisième partie

Structure cristallographique de l'Uracile-ADN Glycosylase d'EBV : une enzyme de réparation de l'ADN

Chapitre 8

Introduction : Mécanismes de réparation de l'ADN

8.1 L'organisme est le sujet d'agressions constantes

LE maintien de l'intégrité du matériel génétique est un élément essentiel qui contribue à la survie et au contrôle de l'évolution de tout organisme. L'organisme est amené à lutter en permanence contre des facteurs environnementaux (radiations ionisantes tels que les UV produits par le soleil, toxines, agents mutagènes pouvant s'intercaler dans l'ADN) mais également intrinsèques (Eg. : La production de variété réactive de l'oxygène est hautement mutagène) pouvant endommager le matériel génétique (Figure 8.1 page 85). Il existe plusieurs types de lésions caractéristiques du matériel génétique qui affecte la structure de l'ADN à différents niveaux (Figure 8.1 page 85). Les rayons UV de la lumière ainsi que certains produits chimiques¹ entraînent des lésions au niveau des nucléotides telles que la formation de dimère de pyrimidine (Mouret *et al.*, 2006). Des rayons ionisants tels que les rayons-X, mais aussi des produits utilisés au cours des chimiothérapies (comme le cis-platine) entraînent la rupture des brins d'ADN (Khanna & Jackson, 2001). Le métabolisme normal de la cellule n'est pas exempt d'erreurs et, au cours de la réplication il peut y avoir insertion, délétion ou un appariement avec un mauvais nucléotide (Jiricny, 2006). Nous nous intéresserons à un autre type de

¹Notamment les polycycliques aromatiques contenus dans la fumée de cigarettes.

dommage, celui causé aux bases. Les rayons-X, les formes réactives de l'oxygène, des agents alkylants mais aussi des réactions chimiques spontanées sont la cause de 5 types de lésions des bases (Huffman *et al.*, 2005) :

1. Dommage oxydatif (Eg. : 8-oxo-7,8-dihydroguanine (8-oxoG)).
2. Désamination (Eg. : On estime que la désamination de la cytosine en uracile survient entre 100 et 500 fois par cellule humaine et par jour (Lindahl & Nyberg (1974), Figure 8.2 page 86).
3. Incorporation d'un nucléotide à la place d'un autre (incorporation de dUMP à la place du dTMP).
4. Hydrolyse des bases qui entraîne une dépurination ou une dépyrimidination (environ 10.000/jour, Lindahl & Barnes (2000)).
5. Alkylation (En générale c'est une méthylation, l'exemple le plus classique étant la formation de 7-methylguanine).

8.2 L'organisme a développé des mécanismes de défenses

Tous les organismes ont développé des mécanismes de réparation des erreurs engendrées par ces différentes agressions. Le génome humain code pour près de 150 enzymes de réparation de l'ADN². Loin de moi l'idée d'opter pour une approche déterministe de la nature et de l'évolution, néanmoins, une telle "armada" et même, dans certains cas, une telle redondance³ dans l'activité des enzymes de réparation de l'ADN, suggèrent une nécessité absolue de protection et plus encore de réparation du matériel génétique. La complexification d'un organisme va de paire avec une régulation stricte de sa replication. Une estimation de la quantité de mutation dans le génome des mammifères suggère un taux d'erreurs de 2.2×10^{-9} paire de base / an (Kumar & Subramanian, 2002). De nombreuses études portent sur le développement d'organismes dont les enzymes de réparation auraient été mutées ou supprimées : plus l'organisme est complexe, plus les conséquences pour son développement sont importantes. Des

²Une base de données est maintenue au laboratoire "Cancer Research UK" par l'équipe de Tomas Lindahl, voir annexe page 180, Wood *et al.* (2005), Wood *et al.* (2001).

³Il existe deux types d'Uracile-ADN Glycosylase dans les cellules Humaines : l'une est nucléaire (UNG2), l'autre est mitochondriale (UNG1).

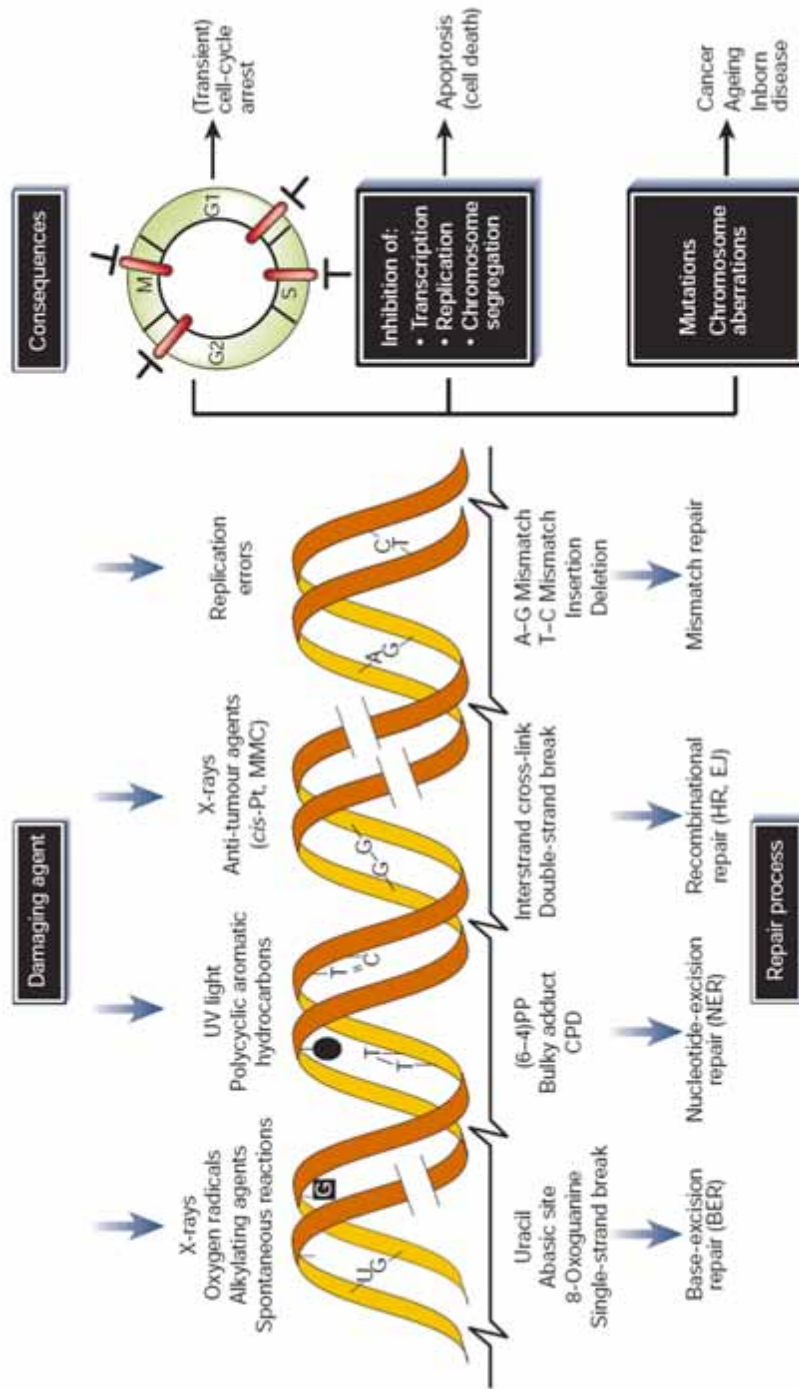


FIG. 8.1: Agressions, conséquences et réparation du génome, (Hoeijmakers, 2001).

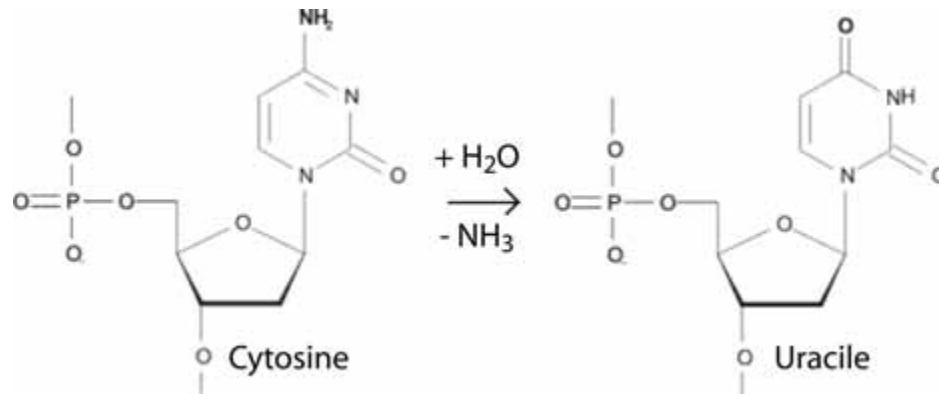


FIG. 8.2: La désamination de la cytosine en uracile peut provoquer des mutations de l'ADN.

souris présentant de telles mutations montrent un panel complet de problèmes de développement au cours de l'embryogenèse, des disfonctions spécifiques de certains tissus, une hyper-sensibilité vis-à-vis de certains agents mutagènes, un vieillissement prématuré, une instabilité génétique et enfin, un taux de cancer très élevé (Wilson & Thompson, 1997). Il en va de même chez *Saccharomyces Pombe*, chez qui le contrôle de l'intégrité du matériel génétique par les enzymes de réparation de l'ADN influence la progression du cycle cellulaire (Tinkelenberg *et al.*, 2002). Nous allons voir quels sont les ouvriers qui veillent à l'entretien de notre patrimoine génétique.

8.3 Hiérarchisation de la réponse à un dommage

Puisqu'il existe une grande variété de dommages au niveau du matériel génétique, l'organisme possède une grande variété de mécanismes de réparation. La réparation des erreurs est organisée de façon hiérarchique, de telle sorte que l'ADN endommagé est pris en charge, dans un premier temps, par des enzymes spécifiques d'une lésion. Ensuite, des étapes "générales" de réparation prennent en charge la suite des événements (Figure 8.3 page 88) en commençant par l'hydrolyse du squelette phosphate de l'ADN⁴. Ensuite, l'ADN polymérase- β (Pelletier *et al.*, 1996) comble le trou laissé par la/les bases éliminées, enfin une ADN Ligase catalyse la réaction phosphodiester entre les extrémités 5' et 3' de l'ADN. En fonction du type de lésion, une partie plus ou moins grande de l'ADN devra être éliminée

⁴Certaines enzymes dites bi-fonctionnelles possèdent à la fois l'activité ADN-Glycosylase et AP-Endonucléase.

pour être remplacée. Une grande quantité de données structurales sont disponibles et permettent une meilleure compréhension des mécanismes mis en jeu (Figure 8.4 page 89, Parikh *et al.* (1997)). Dans le cadre de ce travail, nous nous focaliserons sur l'excision et la réparation des bases⁵ (“**Base Excision Repair**” ou BER) et plus particulièrement de l'une d'entre elles : **l'uracile**.

8.4 Mécanisme d'Excision et de Réparation des Bases (BER)

Une base endommagée est la lésion de l'ADN la plus commune⁶. La première étape de réparation de l'ADN suppose la reconnaissance spécifique et l'élimination de la base endommagée. C'est une étape d'engagement (committed step) ce qui signifie que, une fois entamée, ce processus doit se conduire jusqu'à son terme. En effet, toutes les étapes qui suivent l'élimination de la base inadéquate présente un danger pour la cellule, notamment les sites abasiques qui sont hautement mutagènes (Shikazono *et al.*, 2006). La reconnaissance d'une base inadéquate représente un véritable “challenge” qui s'apparente à rechercher une aiguille (la base inadéquate) dans une botte de foin (les milliards de bases du génome). Pendant de nombreuses années, des recherches ont été menées afin de déterminer le mécanisme par lequel ces enzymes procèdent⁷ (Verdine & Bruner (1997), Chen *et al.* (2002), Zharkov & Grollman (2005)). Certains supposaient que l'enzyme “scannait” et retournait une à une toutes les bases jusqu'à identifier une base mutée (Verdine & Bruner, 1997). D'autres théories supposaient que la liaison de l'enzyme à l'ADN entraînait une baisse de l'énergie de retournement spontanée des bases, une base inadéquate ayant de plus forte chance de se retourner spontanément (Huang & MacKerell (2004), Fuxreiter *et al.* (2002)). En glissant le long de l'ADN, l'enzyme interagira, statistiquement, plus souvent avec les bases inadéquates, ce qui permettra la formation du complexe catalytiquement actif. Récemment, différentes équipes ont publié de nouveaux résultats montrant que la recherche des lésions

⁵En effet, il existe d'autres mécanismes de réparation de l'ADN où, par exemple, un nucléotide entier est excisé, on parle alors de “Nucleotide Excision Repair”.

⁶Il y a plusieurs milliers de bases endommagées par cellule et par jours chez l'Homme (Lindahl, 1993).

⁷C'est une question dont l'intérêt dépasse largement le domaine des spécialistes de la réparation de l'ADN. En effet l'identification d'une séquence particulière, par un facteur de transcription par exemple, pose le même type de problème : comment trouver l'information pertinente au milieu du bruit que représente les bases / séquences sans intérêt pour une enzyme donnée.

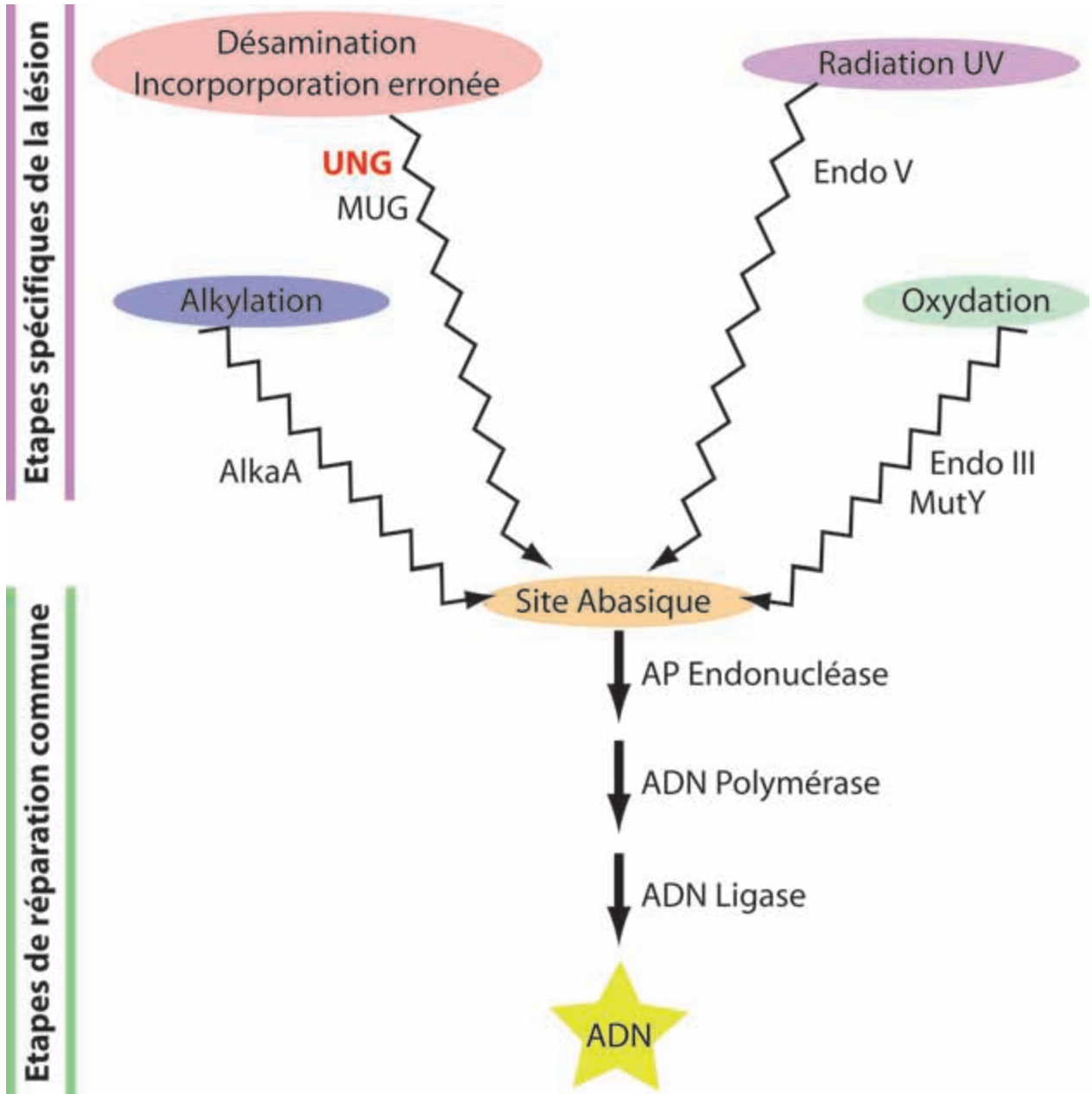


FIG. 8.3: La réparation des lésions du matériel génétique comprend 2 étapes : une étape spécifique de lésion, une étape commune quelle que soit la lésion, adapté de (Mol *et al.*, 1999).

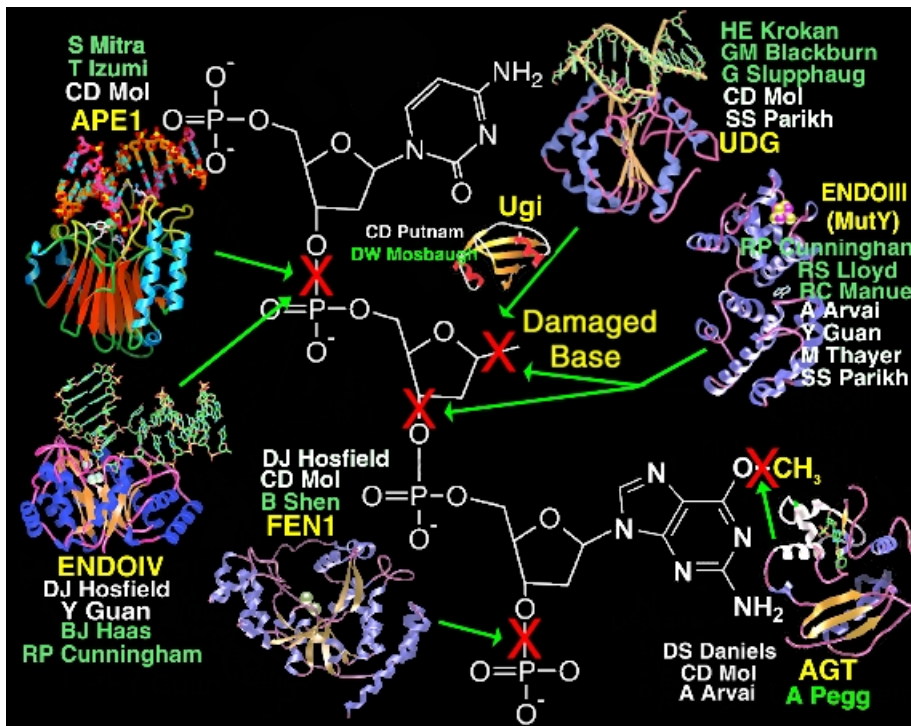


FIG. 8.4: Représentation de la structure tridimensionnelle des protéines intervenant dans la réparation de l'ADN (Parikh *et al.*, 1997)

ne nécessite pas le retournement de toutes les bases de l'ADN. Cao *et al* ont montré que la recherche d'une lésion est un mécanisme passif, les enzymes de réparation utilisent le retournement spontané des bases de l'ADN et les stabilisent dans leur conformation extrahélicale (Cao *et al.*, 2004). Les travaux de Banerjee *et al* montrent que la protéine MutM insère un résidu à l'intérieur de la double hélice, ce qui contraint fortement la paire de bases interrogée, qui reste néanmoins intrahélicale (Banerjee *et al.*, 2006). Enfin, Blainey *et al* ont démontré que la cinétique à laquelle l'enzyme hOgg1 se déplace le long de l'ADN est incompatible avec le retournement systématique de chaque base (Blainey *et al.*, 2006).

8.4.1 Les ADN Glycosylases, une super-famille avec une grande diversité de substrats

La première étape de réparation de l'ADN fait intervenir une enzyme spécifique d'une lésion. Si cette lésion concerne une base, celle-ci devra être éliminée. La précision de cette étape d'engagement

est déterminante pour la survie de la cellule et l'équilibre de l'organisme. Afin d'assurer une spécificité maximale, l'organisme possède tout un panel d'enzymes (Figure 8.5 page 91, Huffman *et al.* (2005)) spécifiques d'un type de dommage et de base (Figure 8.6 page 92, Krokan *et al.* (1997)).

Parmi ces enzymes, il faut distinguer les ADN Glycosylase *mono-fonctionnelles*⁸ dont l'activité enzymatique permet la coupure de la liaison N-Glycosidique entre l'azote de la base et le C1' du désoxyribose. Ces enzymes utilisent une molécule d'eau activée qui effectue une attaque nucléophile sur le C1' du désoxyribose (Dodson *et al.*, 1994). Les enzymes *bi-fonctionnelles* comportent non seulement l'activité N-Glycosylase, mais en plus elles exercent une activité AP-Lyases en éliminant le sucre et le phosphate qui composaient le nucléotide⁹ (Figure 8.7 page 93). Contrairement aux enzymes mono-fonctionnelles, elles n'ont pas recours à une molécule d'eau catalytique mais elle utilise la chaîne latérale d'une lysine (EndoIII, Fromme & Verdine (2003)) ou un résidu proline (EndoVIII, Zharkov *et al.* (2002)) pour exercer leur activité (Dodson *et al.*, 1994).

Un mécanisme “unifié” quel que soit le type d'ADN Glycosylase

Quel que soit le type d'ADN Glycosylase, il existe un certain nombre d'étapes et de concepts communs qui permettent à ces enzymes de reconnaître et d'exciser une base inadéquate.

- Toutes ces enzymes ont la capacité de retourner la base inadéquate dans une poche qui lui est hautement spécifique¹⁰.
- Elles interagissent avec le petit sillon de l'ADN.
- Elles provoquent la courbure de l'ADN.

Hans Krokan fut le premier à démontrer que la poche de reconnaissance de la base est spécifique d'une base donnée (Krokan & Wittwer, 1981). Seule cette base fournit les contacts et possède l'orientation nécessaire au déroulement de la réaction. Il est possible de changer cette spécificité en mutant les résidus environnants (Kavli *et al.*, 1996). Sudip Parikh, dans un article de référence montrant la structure du complexe entre l'UNG humaine et un double brin d'ADN (Parikh *et al.*, 1998), mit en évidence une autre caractéristique commune des ADN Glycosylase : le retournement de la base

⁸Eg. : Uracile-ADN Glycosylase (Kavli *et al.*, 2002).

⁹Eg. : MutM(Fpg) (Fromme & Verdine (2002), Gilboa *et al.* (2002)).

¹⁰On parle de *Base Flipping*.

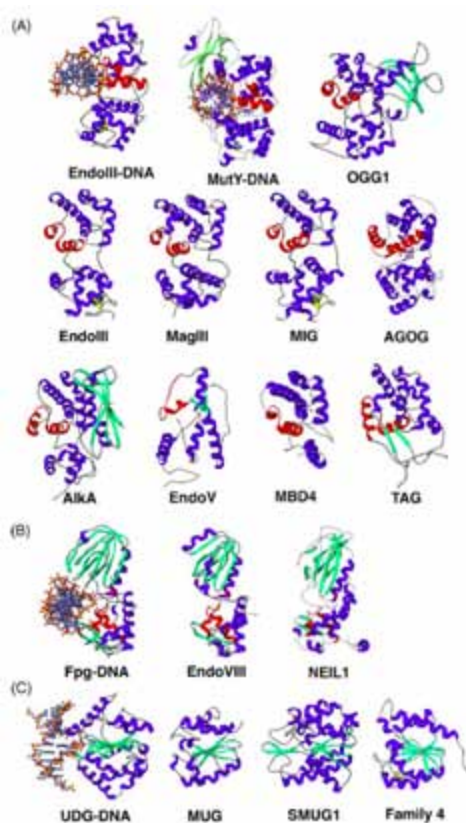


FIG. 8.5: Familles des enzymes de reconnaissance et d'élimination des dommages dans l'ADN . (A) Représentant de la famille des HhH (Helix-hairpin-Helix). Le motif HhH est coloré en rouge, les hélices- α sont colorés en bleu et les feuillets- β en vert. EndoIII (Endonuclease III ou Nth, PDB : 1P59), MutY (adenine DNA glycosylase, reconnaît les paires 8oxoG :A, élimine A, PDB : 1RRQ). Il faut remarquer que EndoIII (PDB : 2ABK), MutY, et MIG (thymine DNA glycosylase, reconnaît les paires G :T et enlève T , PDB : 1KEA) contiennent un cluster Fe-S, alors que AlkA (3-méthyladenine DNA glycosylase II, PDB : 1DIZ) et OGG1 (8-oxoguanine glycosylase, PDB : 1KO9) possèdent des domaines additionnels comportant des feuillets- β . EndoIII, AlkA et Ogg1 reconnaissent et éliminent un large spectre de bases alkylées et oxydées. MAGIII (3mA DNA glycosylase III, élimine le 3,9-diméthyladenine, PDB : 1PU7). TAG (3-méthyladenine glycosylase, élimine 3mA et 3mG, PDB : 1LMZ) et MBD4 (methyl-CpG binding domain protein 4, élimine T et U dans les paires G :T et G :U respectivement, PDB : 1NGN) ne montrent que peu de similarité structurale avec les autres membres de la famille HhH en dehors du motif HhH. AGOG (archaeal GO glycosylase, élimine 8oxoG, PDB 1XQP). EndoV (T4 Endonucléase V, élimine les dimères de pyrimidine, PDB : 1VAS). (B) La famille H2TH (Helix-two-turns Helix) : Fpg (formamidopyrimidine DNA glycosylase) ou MutM, éliminent les purines oxydées (PDB : 1KFV), EndoVIII (Endonuclease VIII ou Nei, élimine les pyrimidines oxydées, PDB : 1K3W) et NEIL1 (Nei-Like, élimine les pyrimidines oxydées, PDB : 1TDH). Le motif H2TH est coloré en rouge, les 6 acides-aminiés N-terminaux sont colorés en pourpre. (C) Les représentants des 4 familles d'Uracile-ADN Glycosylase : UDG (PDB : 1SSP), MUG (PDB : 1MUG), SMUG1 (PDB : 1OE5) et la famille 4 UDG (PDB : 1UIO). Les représentants de cette famille sont décrits en détails dans la suite de ce document. Adapté de Huffman *et al.* (2005).

Alkylated and halogenated bases	Enzyme	Source/gene	Reported DNA substrates	β -Lyase activity
<p>3-Methyladenine (1) 7-Methyladenine (2) 3-Methylguanine (3) R = CH₃: 7-methylguanine (4) R = CH₂CH₂OH: 7-hydroxyethylguanine (5) R = CH₂CH₂Cl: 7-chloroethylguanine (6)</p> <p>O²-Alkylthymine (7) O²-Alkylcytosine (8) 5-Fluorouracil (9)</p>	Uracil-DNA glycosylases (UDGs)	Viral Bacterial (<i>ung</i>) <i>S. cerevisiae</i> (<i>UNG1</i>) Plants Human (<i>UNG</i>)	29 9, 19, 23, 23 29 29 19, 20, 23, 25	No No No No No
	G/T(U)mismatch-DNA glycosylases	<i>M. thermoautotrophicum</i> Insects Human	1, 23, G/G, A/G, T/C, U/C 29 T, 29	No No No
	Alkylbase-DNA glycosylases	<i>E. coli</i> (<i>tag</i>) <i>E. coli</i> (<i>alkA</i>) <i>S. cerevisiae</i> (<i>MAG</i>) <i>S. pombe</i> (<i>magT</i>) <i>A. thaliana</i> (<i>MPG</i>) Rodent/human (<i>MPG</i>)	1, 3 1-8, 24, 25, 30-32 (G, A) 1-6, 30, 31 1 1 1-6, 10, 30, 31	No No No ? ? No
Oxidized and ring-fragmented bases	5-Methylcytosine-DNA glycosylase	Chick Human	5-MeC, T in G/T 5-MeC	No No
<p>8-Oxoguanine (10) 2,5-Amino-5-formamidopyrimidine (11) 4,6-Diamino-5-formamidopyrimidine (12) 2,6-Diamino-4-hydroxy-5-formamidopyrimidine (13)</p> <p>5-Hydroxycytosine (14) 5,6-Dihydrothymine (15) 5-Hydroxy-5,6-dihydrothymine (16) Thymine glycol (17) Uracil glycol (18)</p> <p>Isotaluric acid (19) Alloxan (20) R = H: 5,6-dihydrouracil (21) R = OH: 5-hydroxyuracil (23) R = OH: 5-hydroxy-5,6-dihydrouracil (22) R = CHO: 5-formyluracil (24) R = CH₂OH: 5-hydroxymethyluracil (25)</p> <p>urea (26) methylartronylurea (27) 5-hydroxyhydantoin (28)</p>	Adenine-specific mismatch-DNA glycosylases	<i>E. coli</i> (<i>mutY</i>) Bovine, human (<i>MYH</i>)	A in G/A, C/A, C/A A in G/A, 10/A, C/A	Yes/no Yes?
	DNA glycosylases removing oxidized pyrimidines (EndoIII-like)	<i>E. coli</i> EndoIII (<i>nth</i>) <i>S. cerevisiae</i> (<i>NTG1</i>) <i>S. pombe</i> (<i>nth</i>) Bovine/human EndoIII homologue	14-18, 20-23, 23-28 11, 12, 13, 17 17, 23 17, 23	Yes Yes Yes Yes
	EndoVIII	<i>E. coli</i>	15, 17	Yes
	EndoIX	<i>E. coli</i>	26	?
	Hydroxymethyl-DNA glycosylase	Mouse Bovine	25, 29 25	No Yes?
	Formyluracil-DNA glycosylase	Human	24	?
Deaminated bases	DNA glycosylases removing oxidized purines	<i>E. coli</i> (<i>fpg</i>) <i>S. cerevisiae</i> (<i>OGG1</i>) <i>S. cerevisiae</i> (<i>OGG2</i>) <i>D. melanogaster</i> S3	10-13 10 (opposite G or T) 10 (opposite G or A) 10	Yes Yes Yes Yes
Others	Pyrimidine-dimer-DNA glycosylases	T4 <i>M. luteus</i> (<i>pdg</i>) <i>N. mucosa</i>	33, 12 33 33	Yes Yes Yes

FIG. 8.6: Les groupements chimiques que l'on ne trouve pas dans l'ADN "naturel" sont colorés en rouge. Les substrats sur lesquels les différentes ADN Glycosylases sont actives sont indiqués par un numéro correspondant aux formules chimiques de la partie gauche du tableau. Pour les ADN Glycosidases qui reconnaissent plus d'un substrat, le substrat de prédilection est coloré en rouge, adapté de Krokan *et al.* (1997).

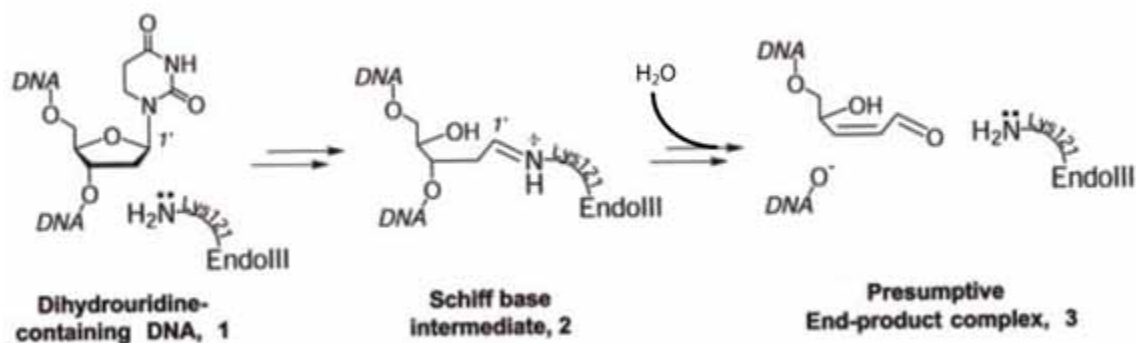


FIG. 8.7: Réaction catalysée par l'enzyme EndoIII. Adapté de Fromme & Verdine (2003)

à l'intérieur de sa poche de reconnaissance s'accompagne d'une torsion autour du phosphate 3', ce phénomène avait été observé pour la première fois sur une autre enzyme : la O(6)-alkylguanine-DNA alkyltransférase (Daniels *et al.*, 2004). Le dernier point commun entre toutes ces ADN Glycosylase, est la nécessité de coordonner la réponse à la détection et à l'élimination de la base inadéquate avec les autres enzymes de réparation de l'ADN qui interviennent au cours des étapes dites "générales" (Figure 8.3 page 88). Un site abasique ou un brin d'ADN rompu est bien plus toxique qu'une mutation non réparée¹¹ (Lindahl, 1993).

Le travail effectué au cours de cette thèse porte sur une enzyme qui reconnaît et excise l'uracile. Cette enzyme appartient à une sous famille des ADN Glycosylase : les Uracile-ADN Glycosylase de famille 1. Nous allons voir quelles sont leurs spécificités.

¹¹Une base inadéquate peut survenir n'importe où dans le génome à des endroits où sa présence peut n'avoir que peu de conséquences.

8.4.2 L'Uracile-ADN Glycosylase (UNG) catalyse l'élimination de l'uracile

Source d'uracile dans l'ADN

L'activité catalysée par l'Uracile-ADN Glycosylase fut observée pour la première fois par Tomas Lindahl en 1974 (Lindahl, 1974). Les Uracile-ADN Glycosylase sont présentes dans tous les organismes, des bactéries à l'Homme et dans certains virus (Huffman *et al.*, 2005). Elles permettent la reconnaissance et l'élimination de l'uracile présent dans l'ADN. L'uracile dans l'ADN provient d'une multitude de sources. Il peut être le résultat du métabolisme normal de la cellule avec l'incorporation de dUMP (Tye *et al.* (1977), Wist *et al.* (1978)) ou la conséquence des multiples agressions que subit l'organisme (Lindahl, 1993). La désamination de la cytosine en uracile est potentiellement mutagénique puisqu'elle peut entraîner des transitions $C : G \rightarrow T : A$ si elle n'est pas réparée avant le cycle de réplication suivant. Outre son rôle dans les mécanismes de réparation de l'ADN, il a été montré que l'UNG2¹² humaine joue un rôle dans le système immunitaire et permet la diversification des anticorps (Petersen-Mahrt *et al.*, 2002).

Les quatre familles d'Uracile-ADN Glycosylase

Il existe quatre familles d'Uracile-ADN Glycosylase qui diffèrent par les détails de leur repliement et par leur spécificité par rapport aux substrats. *La première famille* est représentée par l'Uracile-ADN N-Glycosylase ou UNG, elle est présente dans presque tous les organismes. *La deuxième famille* est représentée par la MUG bactérienne pour **Mismatch-specific UdG** et la TDG ou **Thymine-DNA Glycosylase**, son équivalent eucaryote. *La troisième famille* est représentée par **SMUG1** ou **Single-stranded-selective Monofunctional udG**, cette enzyme possède une large spécificité et reconnaît de nombreux produits oxydés ainsi que l'uracile, néanmoins elle agit surtout sur des substrats double-brin (Masaoka *et al.* (2003), Wibley *et al.* (2003)). Enfin, *la quatrième famille* est la seule à contenir un cluster *Fe-S*, on la trouve chez l'organisme extrémophile *Thermotoga Maritima* (Sandigursky *et al.*, 2001) (Figure 8.5 C page 91).

La séquence des UNG est très variable, néanmoins leur topologie est commune (Figure 8.8 page 95).

¹²C'est l'UNG nucléaire, l'UNG1 humaine est mitochondriale.

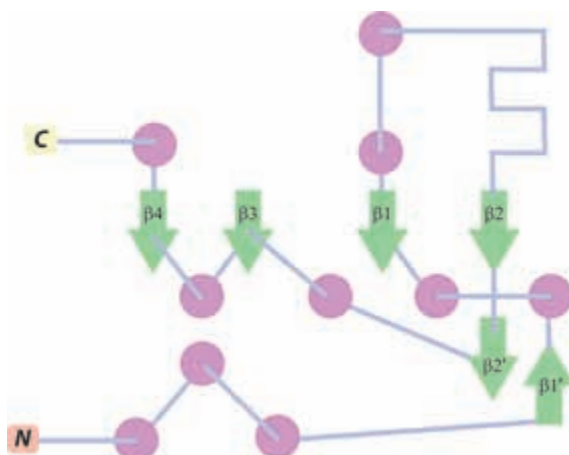


FIG. 8.8: La topologie commune à toutes les UNG comporte un cœur composé de 4 feuillets- β entouré d'hélices- α (Mol *et al.* (1995b), Savva *et al.* (1995)).

Elles possèdent toutes un feuillets- β centrale composé de 4 brins parallèles entourés d'hélices- α . La liaison à l'ADN se fait au niveau d'un sillon, chargé positivement, qui s'étend sur toute une face de l'enzyme du côté c-terminale du feuillet centrale. La poche de liaison à l'uracile se situe au cœur de l'enzyme.

Interaction avec l'ADN et mécanisme catalytique

Malgré une faible identité de séquence¹³, les Uracile-ADN Glycosylase de famille 1 partagent des motifs communs qui permettent de décrire le mécanisme par lequel ces enzymes reconnaissent spécifiquement l'uracile, et ne présentent qu'une activité négligeable vis à vis de la thymine ou de la cytosine (Kavli *et al.* (1996), Caradonna & Cheng (1980), Talpaert-Borlé *et al.* (1982)). Cinq motifs sont décrits de façon systématique dans toutes les structures d'UNG décrite jusqu'à présent¹⁴ et les variations à l'intérieur de ces motifs sont minimales¹⁵.

1. Boucle d'activation de l'eau : 144-GQDPYH-148

¹³Inférieure à 20%.

¹⁴HSV1 (Savva *et al.*, 1995), Humaine (Slupphaug *et al.*, 1996), *E.coli* (Xiao *et al.*, 1999), Gadus Morhua (Leiros *et al.*, 2003), Deinococcus Radiodirans (Leiros *et al.*, 2005).

¹⁵Remarque : la numérotation ainsi que la dénomination des motifs se réfèrent à la structure Humaine publié par Sudip Parikh (Parikh *et al.*, 1998).

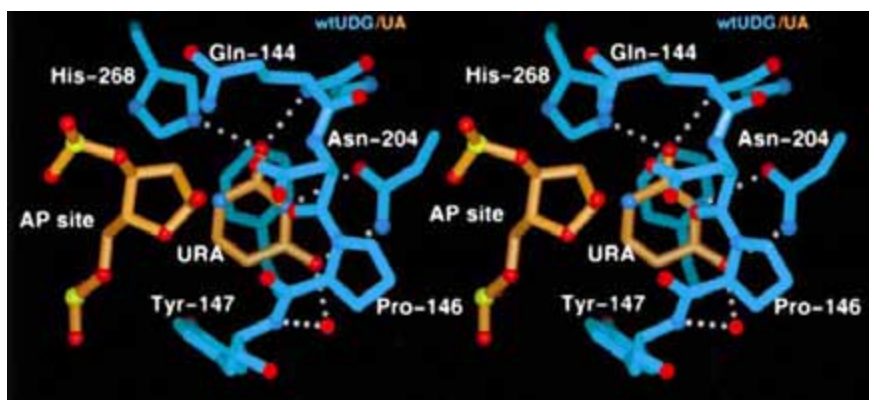


FIG. 8.9: Détail du site actif de l'UNG Humaine contenant de l'uracile. Le groupe carbonyle de l'uracile interagit avec l'amide de la liaison Gly143-Gln144. La chaîne latérale de l'Asn204 interagit avec le N3 et le O4 de l'uracile. La cytosine est exclue car elle possède un N4 en dehors du cycle (Parikh *et al.* (1998), Pearl (2000)).

2. **Boucle Proline**¹⁶ : 165-PPPPS-169

3. **Brin- β de reconnaissance de l'uracile** : 199-GVLLLN-204

4. **Boucle Glycine-Sérine**¹⁷ : 246-GS-247

5. **Boucle Leucine**¹⁸ : 268-HPSPLS-273

L'UNG interagit avec l'ADN grâce à ses motifs 2, 4 et 5. La boucle Proline ainsi que la boucle Glycine-Sérine forme une pince qui comprime le squelette phosphate de l'ADN et lui impose une courbure de 45°, ce qui tend à éjecter la base en direction de la poche de reconnaissance (Parikh *et al.* (1998), Wong *et al.* (2002)). La poche de reconnaissance de l'uracile est hautement spécifique et complémentaire de l'uracile et ne peut s'accommoder d'une autre base (Figure 8.9). La sélection contre la thymine ou la 5-méthyl-cytosine est due à un résidu tyrosine, alors que la sélection contre la cytosine est due au réseau de liaisons hydrogènes (Pearl, 2000).

¹⁶autrement appelée : boucle de compression de l'ADN 5'.

¹⁷autrement appelée : boucle de compression de l'ADN 3'.

¹⁸autrement appelée : boucle d'interaction avec le petit sillon de l'ADN.

8.5 BKRF3 est une Uracile-ADN Glycosylase codée par EBV

Tous les herpesvirus de séquence connue possèdent une Uracile-ADN Glycosylase¹⁹, néanmoins, dans le cas d'EBV, aucune publication ne traite spécifiquement de son rôle et ne la caractérise. Or, l'intérêt de cette protéine dans le cycle de vie des virus est certain.

8.5.1 Pourquoi EBV exprime-t-il des enzymes présente chez l'hôte ?

Vous pouvez désormais vous rendre compte que le génome d'EBV code pour un grand nombre de protéines, certaines sont spécifiques au virus et leur présence ne souffre aucune question. Il est toutefois surprenant de remarquer que certaines enzymes exprimées au cours de la phase lytique précoce ont un équivalent cellulaire. La taille du génome viral est limitée ... dans une certaine mesure, on pourrait les qualifier de "champion de l'organisation". Il est communément admis que les virus ne s'encombrent pas du superflu et préfèrent utiliser à leur profit les protéines cellulaires plutôt que de les coder dans leur génome. Par exemple, le VIH encapsule l'UNG2 cellulaire avant de bourgeonner hors de la cellule qu'il infecte (Chen *et al.* (2004), Priet *et al.* (2005)). A l'opposé, de nombreux virus expriment cette protéine de même que la dUTPase. Ce constat est surprenant, sachant que, dans la plupart des cas, la fonction de ces enzymes ne semble pas "vitale" pour le virus. De plus, elles n'interviennent que dans les premières étapes des mécanismes de réparation de l'ADN ... cette affirmation étant à nuancer dans le cas des herpes-virus dont le grand génome n'a certainement pas livré l'intégralité de ses secrets. Il semble raisonnable de supposer que certaines fonctions nécessaires au maintien de l'intégrité du génome, soient assurées par des protéines dont la fonction n'est pas encore identifiée, ou serait une fonction secondaire d'autres protéines (Eg. : Fonction exonucléase de l'ADN-polymerase). Néanmoins, le virus privé de la fonction assurée par ces enzymes se révèle moins efficace au cours de l'infection. Herpes Simplex Virus 1 (HSV1) dont l'UNG est mutée est 10⁵ fois moins neuro-invasif (Pyles & Thompson, 1994). De même, le cytomegalovirus (HCMV) ou le virus de la Vaccine, dont l'activité UNG est éliminée (Ellison *et al.* (1996), Prichard *et al.* (1996)) se réplique de façon peu efficace. A l'opposé, le virus de la Varicelle (VZV) n'a besoin ni de l'UNG cellulaire ni de l'UNG viral pour se répliquer en culture de fibroblastes (Reddy *et al.*, 1998). L'exemple le plus étonnant est certainement

¹⁹Information provenant de la base de données VIDA(VIDA : Virale DAtabase), voir annexe page 180.

celui de KSHV un γ -herpes virus, dont la protéine LANA (Latency Associated Nuclear Antigen) interagit avec l'UNG2 cellulaire. Si l'UNG2 est absente, alors le virus réplique son ADN de façon moins efficace et produit des particules virales incapables de propager l'infection, or, le génome de KSHV comporte un gène codant pour l'UNG (Verma *et al.*, 2006).

Ce travail ne prétend pas apporter de réponses à ces interrogations, le système d'étude est complexe et nous sommes loin de nous douter de ce qu'il se passe réellement *in-vivo*. L'activité de ces enzymes est-elle nécessaire dans certains types cellulaire ? Qu'advient-il au génome du virus lorsque celui-ci infecte de façon latente une cellule quiescente ? Dans ce cas, l'expression de telles enzymes trouverait sa justification, sachant que l'activité de ces cellules est quasi "végétative". Elle n'exprime qu'un petit nombre de protéines dont l'activité est requise pour leur survies. L'expression des enzymes de réparation de l'ADN n'est pas nécessaire puisque le génome de ces cellules n'est pas répliqué et ne risque pas d'incorporer d'uracile²⁰. Dans ce cas, le virus a besoin de remplir cette fonction de protection de son génome, et d'exprimer lui-même des enzymes de réparation (Chen *et al.*, 2002) .

8.5.2 L'Uracile-ADN Glycosylase d'EBV est une cible potentielle pour le développement de nouveaux médicaments

L'étude du métabolisme conduisant à la synthèse de dTMP a fourni, historiquement, les médicaments anti-cancéreux les plus utilisés²¹. Les thérapies utilisées pour lutter contre les infections virales ont tendance à abuser de l'utilisation d'analogues nucléotidiques, ce qui a pour conséquence la multiplication des résistances. Il est nécessaire de rechercher de nouvelles cibles. La dUTPase (Studebaker *et al.*, 2001) et plus récemment l'UNG (Studebaker *et al.*, 2005) pourraient être de nouvelles cibles de choix.

²⁰Une nuance est à apporter : en effet la désamination de la cytosine peut conduire à la présence d'uracile, mais ces cellules ne se répliquant pas, les conséquences sont certainement moins graves que pour des cellules en division active.

²¹5-FU et methotrexate notamment.

Chapitre 9

Matériel et Méthodes

9.1 Conception de la construction UNG Δ 24

9.1.1 Protéolyse limitée et séquençage N-terminal

Protéolyse limitée

LES enzymes protéolytiques attaquent les parties accessibles de la protéine, elles permettent de détecter des parties flexibles. Nous avons utilisé l'élastase, la chymotrypsine et la trypsine pour digérer l'UNG. L'UNG est utilisée à une concentration de 30 μ M, les différentes enzymes protéolytiques (Sigma) sont utilisées à 0.2 μ M. L'expérience est effectuée dans le tampon¹ de fixation à température ambiante avec agitation sur un plateau basculant. La "cinétique" de coupure de l'UNG a été estimée par des prélèvements de 20 μ l du mélange réactionnel à plusieurs intervalles de temps (1', 2', 5', 10', 15', 30', 45', 60', 2h, 3h, 16h), la réaction est stoppée par l'ajout de Pefabloc-Sc² et l'incubation à 95°C pendant 4min, avec du tampon de chargement³. L'expérience a par la suite été réalisée avec de la trypsine immobilisée sur billes d'agarose⁴ pour vérifier si le niveau de digestion était équivalent en 16h à celui obtenu avec l'enzyme en solution. Le profil de digestion est analysé sur gel SDS-PAGE laemmli 15%.

¹Tris 20 mM pH 7.5 NaCl 100 mM imidazole 20 mM.

²[4-(2-Aminoethyl)benzenesulfonylfluoride.HCl], Roche.

³Tris 400mM pH 6.8, SDS 4%, glycérol 4%, ?- mercaptoéthanol 4%, bleu de bromophénole 1%.

⁴Trypsine-agar, Sigma.

Séquençage N-terminal

Les échantillons d'UNG sont déposés sur un gel SDS-PAGE Laemmli 15% que l'on fait migrer pendant 1h à 150V. Le gel est incubé dans le tampon de transfert⁵ pendant 10min. L'UNGtryps⁶ est transféré pendant 1h30 à 50V à température ambiante vers la membrane de nitrocellulose, que l'on a préalablement immergée dans du méthanol 100% pendant 1 min, puis, dans le tampon de transfert pendant 20 min. Après le transfert, la membrane est rincée à l'eau puis dans du méthanol 100% pendant quelques secondes, elle est ensuite colorée avec 0.1% de bleu de coomassie R-250, 40% de méthanol, et 1% d'acide acétique pendant 10 min. La décoloration s'effectue en quelques minutes dans du méthanol 50%. La membrane est finalement rincée à l'eau. Le séquençage N-term est effectué à l'IBS⁷ selon le principe de la dégradation d'Edman, sur un séquenceur *Applied Biosystems gas-phase model 492*. La chromatographie a été utilisée pour identifier et quantifier le dérivé d'acide aminé éliminé à chaque cycle de séquençage.

9.2 Caractérisation de l'échantillon

9.2.1 Diffusion dynamique de lumière (DLS)

Principe physique de la DLS

Nous avons déterminé l'homogénéité de la solution de protéine au moyen de la DLS, pour ce faire, nous avons utilisé un appareil *Dynapro*TM de *Wyatt Inc.*

Quand un rayon de lumière passe au travers d'une solution de macromolécule ou de colloïdes, les particules composant la solution dispersent la lumière dans toutes les directions. Quand les particules sont très petites par rapport à la longueur d'onde, l'intensité de la lumière diffusée est uniforme dans toutes les directions (diffusion Rayleigh), pour de plus grandes particules ($>250 \text{ \AA}$), l'intensité diffusée dépend de l'angle. Si la lumière utilisée est cohérente et monochromatique (ie. : Celle d'un laser) on peut observer une fluctuation de l'intensité diffusée dépendante du temps. Ces fluctuations

⁵10 mM 3-[cyclohexilamino]-1-propanesulfonic acid ou CAPS, pH 11, méthanol 10%.

⁶UNGtryps = UNG après digestion tryptique.

⁷IBS : Institut de Biologie Structurale.

sont la conséquence des déplacement aléatoires (mouvement Brownien), la distance entre ces particules varie constamment. Ainsi, des interférences constructives et destructives de la lumière diffusée par les particules voisines entraînent une fluctuation de l'intensité reçue au niveau du détecteur, puisque cette variation d'intensité est la conséquence du mouvement des particules, elle contient l'information relative à ce mouvement. Une analyse de la fluctuation de l'intensité en fonction du temps, grâce à une fonction d'autocorrélation, permet de mesurer un coefficient de diffusion D_i (Equation de Stokes-Einstein 9.1) des particules à partir duquel on peut calculer le rayon de Stokes des particules en solution⁸.

$$D_i = \frac{k_B T}{6\pi\eta r_i} \quad (9.1)$$

La fluctuation de l'intensité I par rapport au temps t permet d'abord de déterminer la fonction d'autocorrélation de l'intensité, qui peut être décrite comme la moyenne du produit du signal avec ce même signal décalé dans le temps (Equation 9.2).

$$R_I(\tau) = \int_0^{\infty} I(t)I(t + \tau)\partial t \quad (9.2)$$

Le signal, dans ce cas, est le nombre de photons comptés dans un interval d'échantillonnage. Pour un temps de décalage court, l'autocorrélation est haute, avec le temps, les particules diffusent et la corrélation diminue jusqu'à être nulle. La décroissance exponentielle de la fonction d'autocorrélation est caractérisée par le coefficient de diffusion des particules dans le cas d'un seul type de particule. L'analyse de la fonction d'autocorrélation, en terme de distribution de la taille des particules, est effectuée en ajustant les données numériquement avec plusieurs exponentielles donnant une distribution théorique. Une solution réellement monodisperse donnera une seule décroissance exponentielle à laquelle ajuster les données correspondant à une taille de particule est aisé. En pratique, une solution polydisperse fait que l'on est confronté à une série d'exponentielles et plusieurs schémas existent pour ajuster des données au modèle théorique.

⁸connaissant l'équation de Stokes-Einstein et la viscosité de la solution.

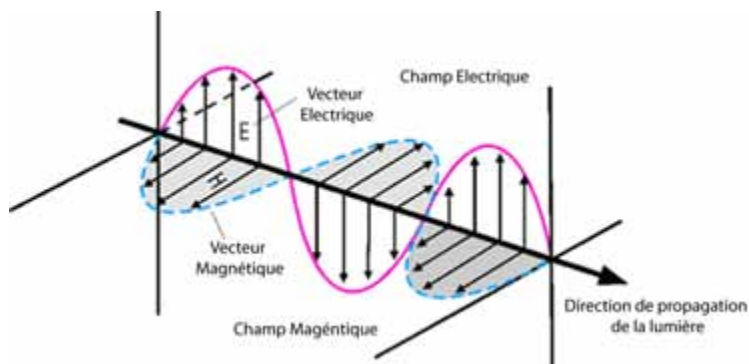


FIG. 9.1: La lumière est une onde électromagnétique (OEM) formée d’une composante électrique et d’une composante magnétique perpendiculaire.

9.2.2 Dichroïsme circulaire

Les ondes électromagnétiques (OEM, Figure 9.1) sont des oscillations périodiques d’un champ électrique perpendiculairement à un champ magnétique, qui évolue dans l’espace et dans le temps, elles se propagent à la vitesse de la lumière dans une direction perpendiculaire au plan dans lequel elles oscillent. Les OEM peuvent interagir de façon constructives ou destructives. Une OEM polarisée circulairement, que l’on peut décrire par l’équation 9.3, peut interagir avec un échantillon de protéine. La structure des acides aminés est asymétrique, ils vont absorber différemment la composante polarisée circulairement à droite et la composante polarisée circulairement à gauche (Figure 9.2 page 103).

$$\vec{E} = \begin{pmatrix} E_0 \cos(\omega t - \kappa z) \\ E_0 \sin(\omega t - \kappa z) \end{pmatrix} = \begin{pmatrix} E_x \\ E_y \end{pmatrix} \quad (9.3)$$

La structure secondaire d’une protéine peut être déterminée par CD dans l’UV lointain (entre 190 et 250 nm), à ces longueurs d’onde, la liaison peptidique est le chromophore, le signal est mesuré quand il provient d’un environnement régulier et replié. Les hélices- α , les feuillets- β et les coudes donnent un spectre de CD caractéristique. La méthode la plus simple pour extraire les informations de structure secondaire à partir d’un spectre de CD est de considérer que le spectre de CD est une combinaison linéaire de spectres de CD de chaque structure secondaire (ex : hélice- α “pure”, brin- β “pure”, etc. Figure 9.3 page 104) rapporté à sont abondance dans le polypeptide. Le principal problème de cette approche est qu’il n’y a pas de spectre CD standard pour de telle structure (“pure”). Pour

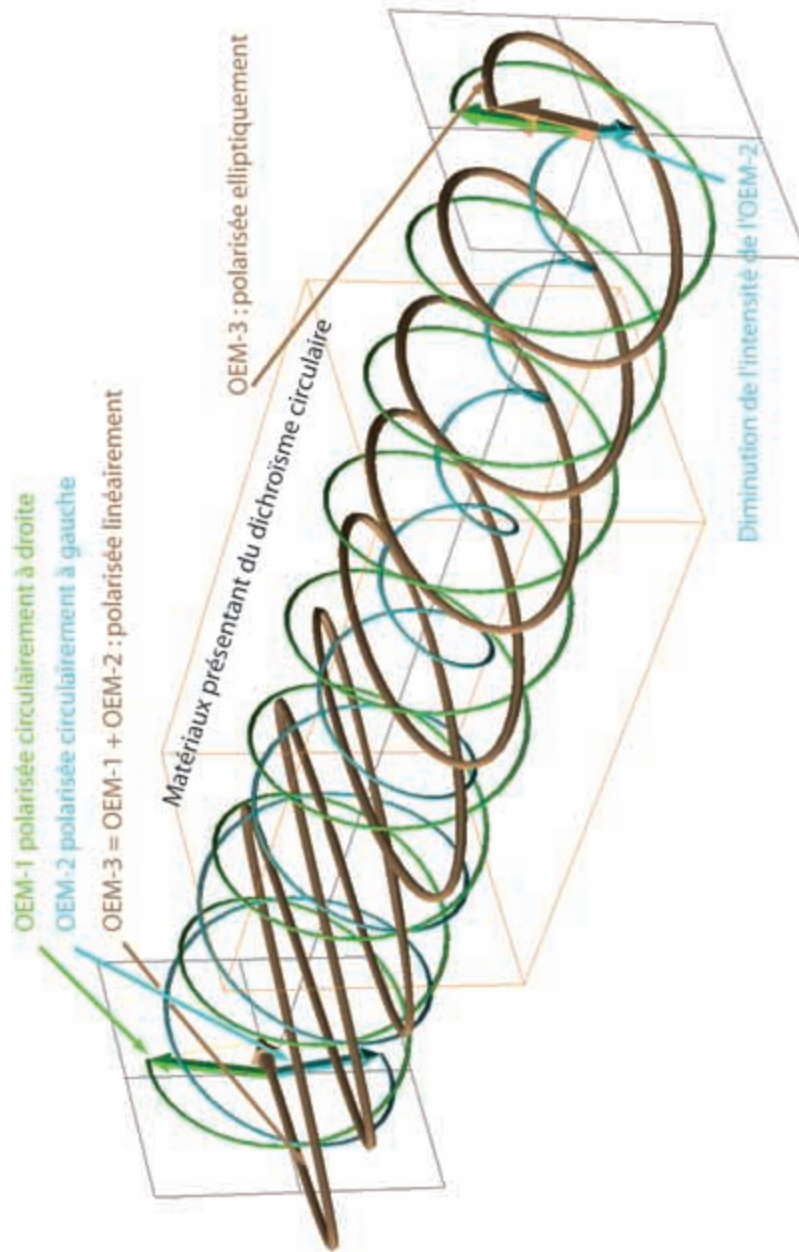


FIG. 9.2: Principe physique du dichroïsme circulaire. Les OEM-1 et 2 sont polarisées circulairement et en sens inverse l'une de l'autre, leur interaction donne une OEM-3 polarisée linéairement. Lorsque les OEM-1 et 2 traversent un matériau qui absorbe en partie l'OEM-2, alors l'OEM-3 est polarisée elliptiquement. (Szilágyi, a)

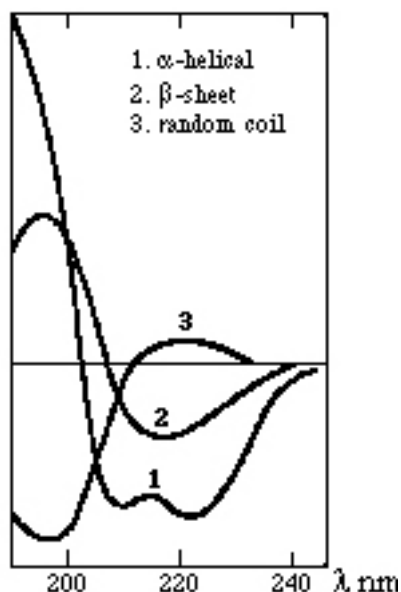


FIG. 9.3: L'analyse des spectres de CD permet de déterminer la composition en structures secondaires d'une protéine : hélice- α , feuillet- β parallèle et anti-parallèle, coude, ... Le dichroïsme circulaire permet d'obtenir une qualité de prédiction (par rapport à la structure déterminée par cristallographie aux rayons-X) d'environ 97% pour les hélice- α , de 75% pour les feuillets- β , et de 50% pour les coudes.

cette raison, des méthodes utilisant les spectres CD de protéines de références contenant des quantités connues de structures secondaires ont été développées pour analyser les spectres CD expérimentaux. Comme "toutes" les techniques spectroscopiques, le signal de CD est une moyenne de la population moléculaire étudiée. Ainsi, le CD permet de déterminer le pourcentage de telle ou telle structure secondaire, mais il ne permet pas de déterminer quel résidu appartient à quelle structure.

En pratique : 200 μ l de protéine à 0.1 mg/ml sont placés dans une cuve en quartz de 1 mm de trajet optique, l'évaluation de la quantité et de la nature des structures secondaires est effectuée dans l'UV lointain entre 190 et 250nm à l'aide d'un spectromètre JASCO les résultats sont analysés avec le logiciel DICROPROT (Deléage, a).

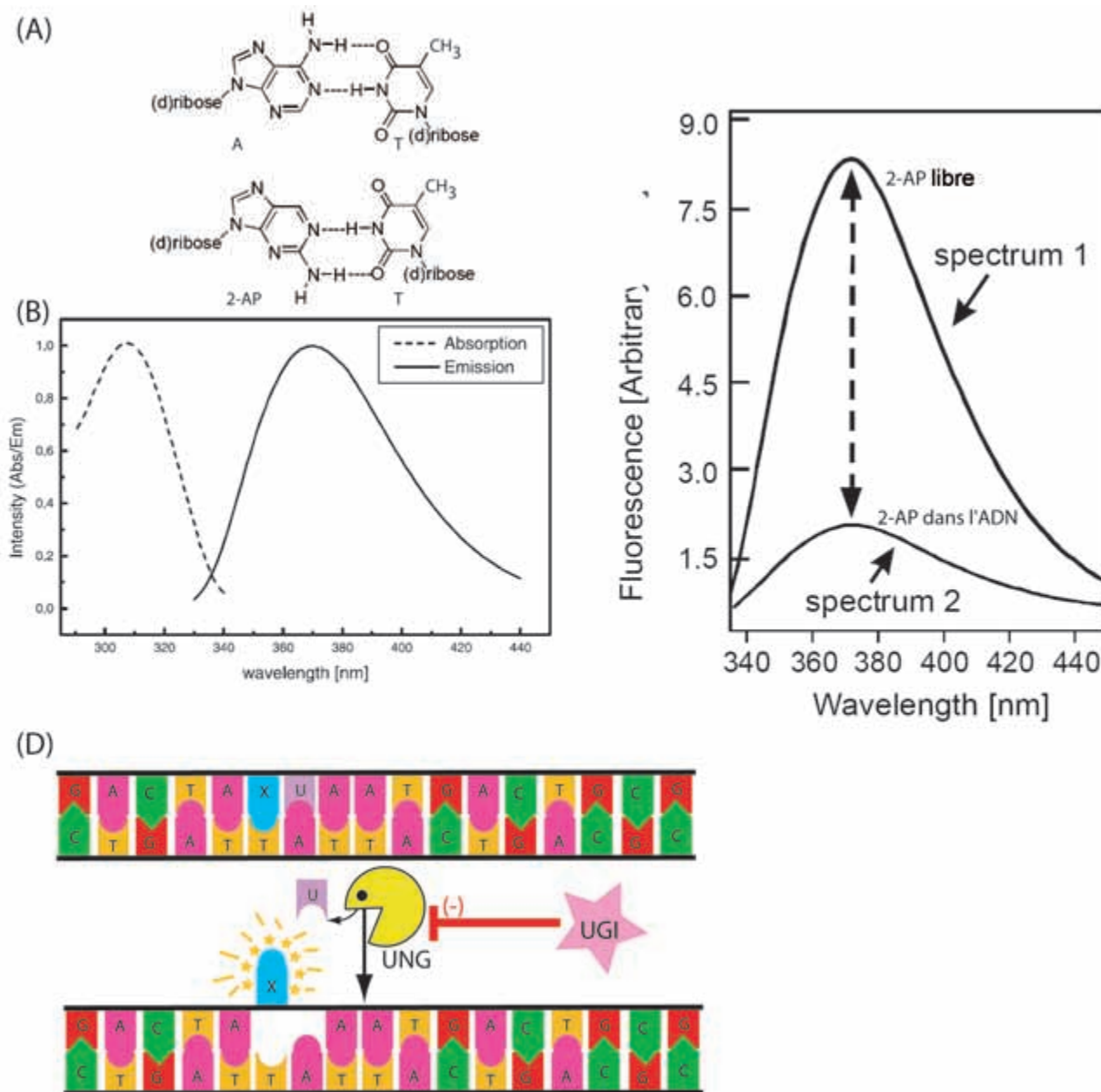


FIG. 9.4: Principe du test enzymatique mis en œuvre. (A) Interaction Watson-Crick entre l'adénine et la thymine (en haut) et entre la 2-amino-purine et la thymine (en bas). La 2-aminopurine diffère de l'adénine par la position du groupement NH_2 . (B) La 2-AP présente un maximum d'absorption à 305 nm et un maximum d'émission à 370 nm. (C) La fluorescence du 2-AP est atténuée dans l'environnement du double brin (Rist, 2003). (D) Test enzymatique : le double brin d'ADN comporte un résidu 2-Amino-Purine (2-AP) à côté d'un uracile. Lorsque l'UNG excise l'uracile, la fluorescence du 2-AP n'est plus atténuée par l'environnement de l'ADN. Adapté de Rist (2003).

9.3 Principe du test enzymatique

Le test enzymatique employé repose sur l'utilisation de 2-Amino Purine (2-AP, Figure 9.4 A page 105). Cette base diffère de l'adénine⁹ par la position du groupement NH₂. Ceci confère une fluorescence au 2-AP (Figure 9.4 B page 105). Or, sa fluorescence est atténuée¹⁰ dans l'environnement du double brin d'ADN et, lorsque le 2-AP est exposé au solvant l'intensité de sa fluorescence augmente d'un facteur 4x (Figure 9.4 C page 105). L'oligomère double brin utilisé pour ce test enzymatique comporte un 2-AP à côté d'un uracile. L'ajout d'UNG entraîne l'excision de l'uracile, le 2-AP se trouve exposé au solvant, sa fluorescence n'est plus atténuée par l'environnement du double brin (Figure 9.4 D page 105)

9.4 Résolution de Structure par Remplacement Moléculaire

9.4.1 Principe

L'utilisation de la cristallographie aux rayons-X donne accès à l'organisation atomique de la matière. Mais, comme son nom le suggère, cette méthode nécessite un artefact expérimental ... en l'espèce : la préparation de cristaux. A l'intérieur d'un cristal les molécules sont réparties de façon régulière dans les 3 dimensions de l'espace. L'interaction entre une OEM dont la longueur d'onde est proche de la dimension de l'objet étudié (ie. : les rayons-X) permet de tirer partie de cet état particulier d'organisation de la matière et d'en déduire la structure de ce qui la compose.

Sans entrer dans les détails, la nature des rayons-X fait qu'ils interagissent avec les électrons qui constituent l'enveloppe des atomes. Lorsqu'un photon d'un faisceau de rayons-X interagit avec un électron, l'électron peut ré-émettre une OEM de façon circulaire (diffusion de Rayleigh). Or, un cristal est composé de couche régulière d'atomes. Tous les électrons vont se comporter comme de petits émetteurs d'OEM. Ces OEM vont interagir de façon constructive ou destructive. Au niveau du cliché de diffraction¹¹ (Figure 9.5 B page 107) on détecte des taches correspondant aux photons diffrac-

⁹L'adénine est une 6-Amino Purine.

¹⁰On parle de *Quenching*.

¹¹Les rayons-X interagissent avec le cristal, ils sont diffractés et entrent en collision avec un détecteur au niveau duquel une image correspondant à l'interaction entre les rayons-X et le cristal est enregistré : c'est le cliché de diffraction.

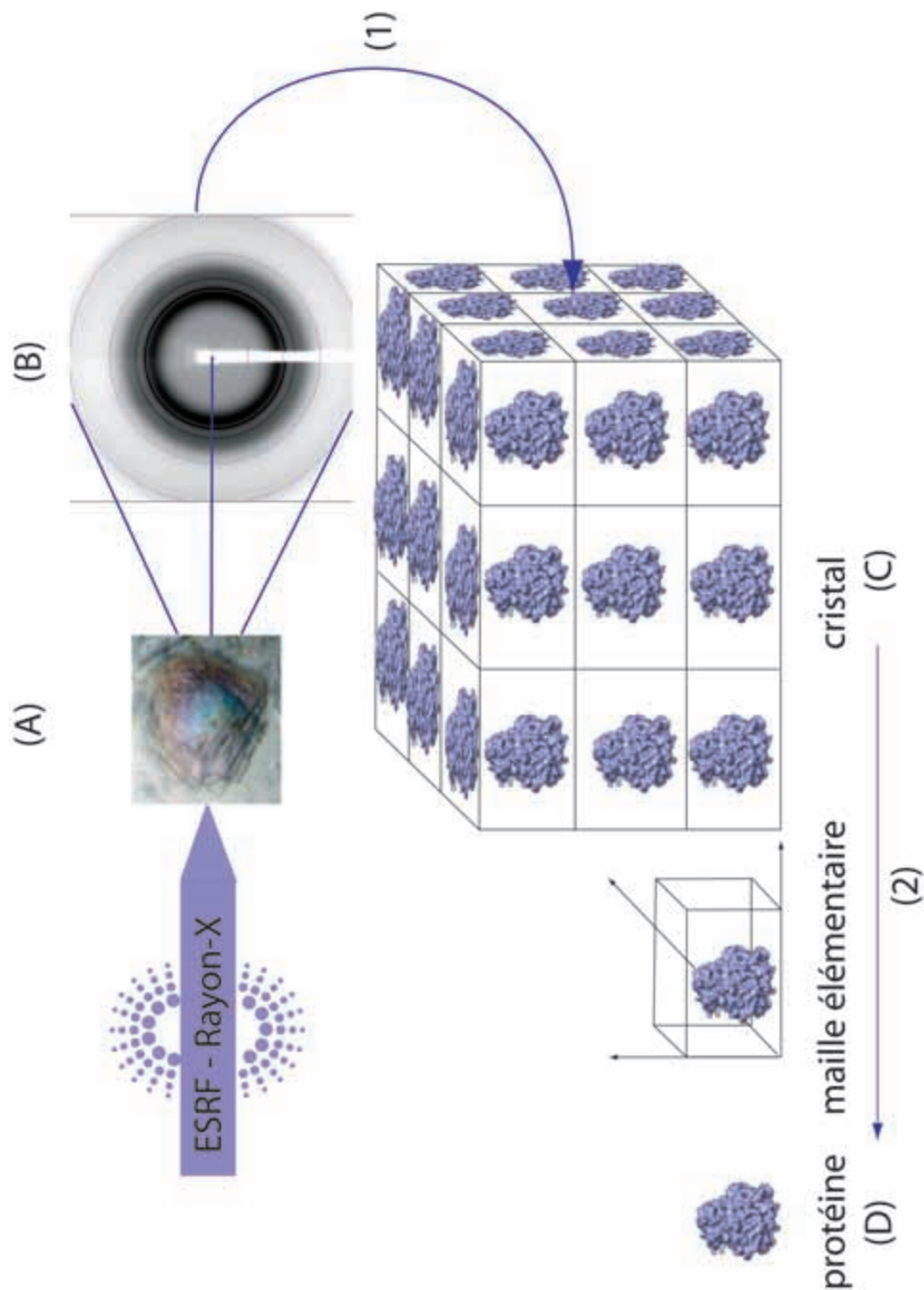


FIG. 9.5: Principe de la cristallographie aux rayons-X. Un cristal (A) est l'organisation régulière d'une protéine dans l'espace (C). L'utilisation des rayons-X permet d'obtenir un cliché de diffraction (B) à partir duquel on peut remonter à la structure du cristal, c'est-à-dire à son organisation (1). La problématique que l'on cherche à résoudre est la suivante : comment le cliché de diffraction obtenu nous permet d'expliquer l'organisation du cristal et la structure de la protéine qui le compose (2, D).

tés. L'analyse de la **position** des taches nous renseigne sur l'**organisation** du cristal (Figure 9.5 C page 107), l'analyse de l'**intensité** des taches nous renseigne sur la **composition** du cristal (Figure 9.5 D page 107). Pour résumer simplement, le travail du cristallographe consiste à utiliser les données mesurées sur le détecteur pour obtenir la structure tridimensionnelle de la protéine étudiée (Figure 9.5 B, D page 107).

9.4.2 Les étapes dans la résolution d'une structure

Afin de comprendre de quelle façon on résout une structure à l'aide de la cristallographie aux rayons-X, il convient de s'interroger sur la nature de l'information contenue dans le cliché de diffraction. En fonction des informations dont on dispose sur la protéine qui compose le cristal, la méthode de résolution de la structure ne sera pas la même.

Diffraction des rayons-X par un cristal

On observe de la diffraction quand les OEM interagissent avec le cristal. Tout le problème consiste à comprendre comment et dans quelles circonstances les OEM sont diffusées en phase. C'est dans ce cas que l'on observe des taches de diffraction au niveau du détecteur. Les taches de diffraction sont souvent appelées réflexions, car le cristal peut être considéré comme une boîte composée d'un grand nombre de miroirs qui reflètent les rayons-X. Ces miroirs sont appelés "Plan réticulaire" (Figure 9.6 page 109). Si l'on considère le cristal comme un réseau dont les nœuds sont A, B et C, un plan de ce réseau peut être défini par l'équation 9.4 :

$$h.u + k.v + l.w = 1 \quad (9.4)$$

Il existe une infinité de plans réticulaires parallèles à ce plan. Cette infinité de plans contient l'ensemble des nœuds du réseau. On désigne cette famille par **(hkl)**. Les trois entiers h, k et l sont les **indices de Miller** de la famille de plans réticulaires (Rousseau, a). Les rayons-X vont interagir avec ces plans. Pour qu'une tache se forme au niveau du détecteur, il faut que cette interaction soit constructive. La loi de Bragg (Equation 9.5) démontre que cette interaction est constructive si la différence de marche¹²

¹²Dans la loi de Bragg, d_{hkl} est la distance entre les plans réticulaires décrits par les indices h, k, l.

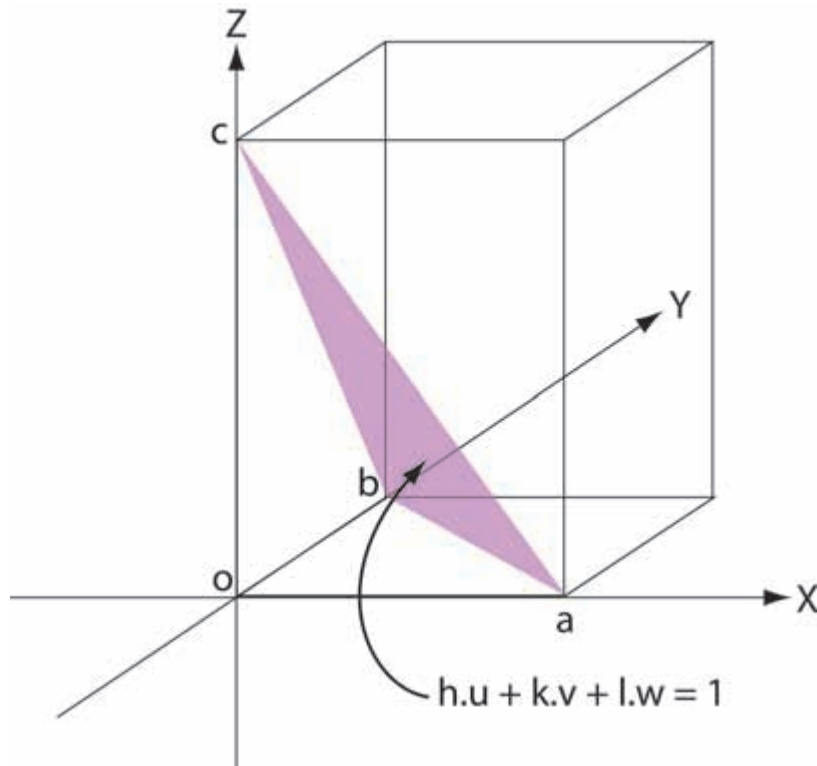


FIG. 9.6: Un cristal est composé d'une infinité de plan réticulaires qui se comportent comme des miroirs vis à vis des rayons-X.

entre les rayons incidents et les rayons diffractés par les électrons est égale à un nombre entier “ n ” de fois la longueur d’onde “ λ ”.

$$n \cdot \lambda = 2d_{hkl} \sin \theta \quad (9.5)$$

Lorsque que l’on étudie un cristal, la répétition des plans qui le composent dans les 3 dimensions de l’espace peut être nommée **réseau direct**, auquel on peut associer un **réseau réciproque** formé par les taches diffractées sur le détecteur. Lors d’une collecte de données, on fait tourner le cristal (ie. : le réseau direct) par rapport à la source de rayons-X, le réseau réciproque qui lui est associé tourne également. **La sphère d’Ewald** est une construction mathématique qui permet de connaître dans quelles conditions il y a diffraction (Figure 9.7 page 110). C’est une sphère de rayon $1/\lambda$ (λ est la longueur d’onde du rayon incident), centrée en O sur le cristal. Il y a diffraction quand un nœud du réseau réciproque se trouve sur le périmètre de la sphère d’Ewald.

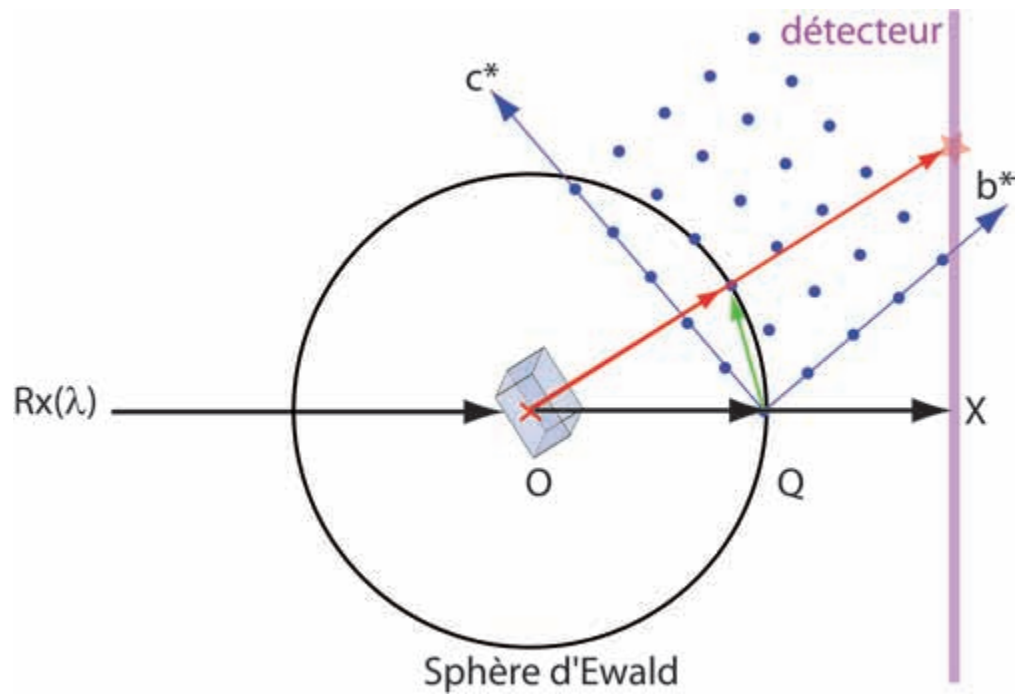


FIG. 9.7: La sphère d'Ewald. Le faisceau de rayons-X est le long de l'axe X. Le point où les rayons-X sortent de la sphère d'Ewald définit l'origine du réseau réciproque (Q). Le cristal définit l'origine de la sphère (O). a^* est orienté autour de l'axe de rotation. Le plan du réseau réciproque $h = 1$ est représenté. Chaque tache de diffraction sur le détecteur possède un vecteur de diffraction correspondant dans l'espace réciproque. Un de ces vecteurs est représenté par une ligne verte entre l'origine de l'espace réciproque (Q) et la surface de la sphère d'Ewald. (Leslie, 2006).

Le cliché de diffraction de la méthode de rotation

Avant de résoudre une structure, il faut collecter des données. C'est une étape cruciale au cours de laquelle il faut accumuler suffisamment d'informations, c'est-à-dire, de clichés de diffraction (cliché de rotation, Figure 9.5 B page 107) à différents angles. Un cristal est un objet symétrique, on utilise la symétrie pour avoir accès à toute l'information qu'il contient. Par exemple, les cristaux de l'UNG d'EBV appartiennent au groupe d'espace $C222_1$ (Figure 9.5 A page 107 et Annexe, Figure 16.1 page 182). C'est un système **orthorhombique**, par conséquent, il suffit de collecter 90° pour avoir accès à la totalité de l'information contenue dans le cristal.

De nombreuses statistiques permettent de caractériser la qualité des données cristallographiques. Lors de la collecte des données, la plupart des taches enregistrées sur le détecteur ont des taches équivalentes reliées par symétrie ou collectées sur un autre cristal. On mesure la déviation par rapport à un idéal théorique en utilisant les statistiques R_{sym} ou R_{merge} ¹³ (Weiss, 2001). L'équation 9.6 nous donne une définition de la statistique R_{sym} , où i indique une mesure individuelle.

$$R_{sym} = \frac{\sum_N [\sum_n (I_i(h) - \bar{I}(h))/n]}{\sum_N \bar{I}(h)} \quad (9.6)$$

Pour faire "simple", on appelle **paire de Friedel** des réflexions de Bragg reliées par une inversion à partir de l'origine¹⁴. La **loi de Friedel** mentionne que des paires de Friedel ont la même amplitude mais une phase opposée. Au niveau du détecteur, on n'enregistre que l'intensité des taches¹⁵, ainsi, chaque tache a une tache équivalente de même intensité. La statistique R_{sym} mesure la somme des différences de toutes les mesures par rapport à la valeur moyenne des mesures, le tout divisé par la somme de toutes les mesures. Si R_{sym} est bas, cela signifie que les valeurs ont été mesurées correctement.

¹³Ces statistiques sont pratiquement interchangeables. R_{sym} était utilisé à l'époque où les réflexions n'étaient mesurées qu'une seule fois ou si l'on ne considère que les réflexions provenant d'un seul cristal, la redondance dans la mesure des réflexions, introduites par les méthodes modernes de collectes de données, et l'utilisation de plusieurs jeux de données provenant de différents cristaux nous poussent à utiliser la statistique R_{merge} . Mais ces statistiques décrivent "grosso-modo" la même chose.

¹⁴ie. : centro-symétrie.

¹⁵On n'a pas accès à la phase ... c'est un problème que nous traitons par la suite.

La théorie de Fourier

On appelle **facteur de structure** (F_{hkl}) la transformée de Fourier de la densité électronique $\rho(x, y, z)$ dans la maille (Equation 9.7).

$$F_{hkl} = \int_0^x \int_0^y \int_0^z \rho(x, y, z) e^{(2\pi i)(hx+ky+lz)} dx dy dz \quad (9.7)$$

Le cliché de diffraction est relié à l'objet diffractant (ie. : la densité électronique) par une opération mathématique : **La Transformée de Fourier** (TF). Or, nous nous intéressons à l'objet diffractant, donc, pour l'obtenir, il "suffit" d'effectuer l'opération inverse, c'est-à-dire une **Transformée de Fourier inverse** (TF^{-1}) ... Mais il y a un problème ... afin de calculer la TF^{-1} nous avons besoin, à la fois de l'amplitude et de la phase des ondes diffractées, or le cliché de diffraction ne permet de mesurer uniquement l'amplitude des ondes diffractées¹⁶. Toute la problématique du cristallographe sera d'accéder à la phase de l'onde diffractée.

Déterminer les phases

Pour déterminer les phases de façon expérimentale, il faut user d'un subterfuge. Si l'on perturbe la structure par l'addition d'atomes lourds alors, le cliché de diffraction sera légèrement différent. Ce sont les électrons qui entourent les atomes qui diffractent, or les atomes lourds¹⁷ possèdent beaucoup plus d'électrons que les atomes que l'on trouve en générale dans une protéine¹⁸. Il est donc possible de déduire la position des atomes lourds dans la maille et, d'une façon indirecte, la valeur des phases. Cette technique est appelée : **Remplacement Isomorphe**.

A la place de l'atome lourd, on peut utiliser des atomes possédant une **diffusion anormale** qui permettent d'observer des différences au niveau des paires de Friedel donc de les localiser.

¹⁶On mesure le nombre de photons au niveau de chaque tache du cliché de diffraction, ce qui correspond à l'intensité de l'onde diffractée. L'intensité de l'onde diffractée est proportionnelle au carré de son amplitude.

¹⁷Plomb, Mercure, Platine ...

¹⁸Carbone, azote, oxygène ... hydrogène ...

Deviner les phases à partir d'un modèle

Si la protéine étudiée a déjà été cristallisée, ou, si une structure voisine existe¹⁹, alors on peut utiliser ce modèle et le placer dans le cristal pour calculer des phases que l'on obtiendrait ... ensuite il faut déterminer de quelle façon ce modèle permet d'expliquer ce que l'on observe. Si les 2 structures sont assez proches, cette technique dite de “**Remplacement Moléculaire**”, permet de résoudre la structure de la protéine sans avoir à passer par l'étape de détermination des phases expérimentales. C'est la technique que nous avons utilisée (voir Article 2, page 116).

Nous avons utilisé le programme **Molrep** développé par Alexei Vagin (Vagin & Teplyakov, 1997). Pour comprendre le fonctionnement de ce programme, il faut définir la **fonction de Patterson**. La fonction de Patterson est la TF des intensités des facteurs de structure. On peut calculer une fonction de Patterson à partir d'un modèle atomique et le comparer à la Patterson déterminée expérimentalement. Quand le modèle est dans la bonne orientation, les 2 Patterson montrent une corrélation due aux contributions intra-moléculaires. Le remplacement moléculaire est un problème à 6 dimensions :

- 3 paramètres définissent l'orientation (α, β, γ) .
- 3 paramètres définissent la position (x, y, z) .

La carte de Patterson est une carte représentant des vecteurs que l'on peut diviser en 2 catégories :

- les vecteurs intramoléculaires qui dépendent seulement de l'orientation de la molécule.
- les vecteurs intermoléculaires qui dépendent de l'orientation et de la position de la molécule.

Le programme va procéder en 2 étapes :

- le programme calcule une **fonction de rotation** qui calcule une corrélation due aux vecteurs intramoléculaires, par rapport aux différentes orientations du modèle décrits par les 3 angles d'Euler. (α, β, γ) .
- le programme utilise un certain nombre des meilleurs pics de la fonction de rotation et calcule la corrélation pour toutes les **translations** (x, y, z) dans la maille²⁰.

Le maximum de corrélation détermine la bonne position du modèle dans la maille élémentaire. Si plusieurs copies indépendantes sont présentes, un certain nombre des meilleures solutions sont combinées pour déterminer un modèle du contenu de l'unité asymétrique. Finalement, il y a un affinement

¹⁹La même protéine chez une autre espèce, par exemple.

²⁰Fonction de translation.

de la position de la molécule traitée comme un corps rigide pour minimiser l'écart entre les $|F_{obs}|$ et $|F_{calc}|$.

Construction du modèle et affinement

Une fois le remplacement moléculaire effectué, il faut substituer les AA du modèle par ceux de la protéine étudiée et essayer d'ajuster le modèle à la densité électronique. On considère que le modèle est fini quand les statistiques R_{cryst} et R_{free} sont satisfaisantes. La statistique R_{cryst} est définie par l'équation 9.8 page 114, c'est la somme des différences entre les amplitudes observées et calculées divisée par la somme des amplitudes observées. Pour une structure aux alentours de 2 Å, on considère que R_{cryst} doit être environ 20% en fin d'affinement.

$$R_{cryst} = \frac{\sum_{hkl} ||F_{obs}| - |F_{calc}||}{\sum_{hkl} |F_{obs}|} \quad (9.8)$$

Si l'on considère uniquement la diminution de la statistique R_{cryst} comme révélateur de l'amélioration du modèle, il est possible d'affiner la construction de manière artificielle, pour éviter cela, Axel Brünger proposa à partir de 1993 d'introduire une autre statistique : R_{free} . L'équation 9.9 décrit la statistique R_{free} .

$$R_{free} = \frac{\sum_{test.set} ||F_{obs}| - |F_{calc}||}{\sum_{test.set} |F_{obs}|} \quad (9.9)$$

R_{free} est une statistique non biaisée : un groupe de réflexions est choisi de manière aléatoire et exclue du processus d'affinement avant que l'affinement ne commence²¹.

²¹Un grand nombre des schéma et explications de cette partie proviennent du site internet de "Structural Medicine" de l'Université de Cambridge (Read, a).

Chapitre 10

Article : Structure de l'UNG d'EBV

10.1 Résumé

LE virus d'Epstein-Barr (EBV) est un γ -herpesvirus Humain. Son génome comporte 86 cadres de lecture ouvert parmi lesquels 2 protéines luttent contre l'incorporation d'uracile dans l'ADN, ce sont l'Uracile-triphosphate hydrolase (dUTPase) et l'Uracile-ADN Glycosylase (UNG). L'UNG permet l'élimination de l'uracile présent de manière erronée, suite à la désamination d'une cytosine ou à l'incorporation accidentelle de dUMP. L'enzyme d'EBV appartient à la première famille d'UNGs. L'utilisation des rayons-X nous a permis de résoudre sa structure tridimensionnelle à une résolution de 2.3 Å, en complexe avec la protéine inhibitrice de l'Uracile-ADN Glycosylase (Ugi) que l'on trouve chez le bactériophage PBS-2. La structure de l'UNG d'EBV, codée par le cadre de lecture BKRF3 démontre l'excellente conservation de la structure des UNG de la première famille. 4 des 5 motifs catalytiques sont identiques, néanmoins, le cinquième motif, la "boucle leucine" présente une insertion de 7 résidus. Malgré cette différence, les constantes catalytiques de l'UNG d'EBV sont similaires à celles des autres UNGs. La modélisation du complexe de l'UNG d'EBV avec de l'ADN montre que la longue "boucle leucine" interagit tout de même avec l'ADN et devrait remplir son rôle de déformation de l'ADN d'une façon différente des autres UNGs dont la structure est déjà connue. Bien que l'enzyme exprimée par EBV soit distante d'un point de vue évolutif par rapport à l'enzyme codée par *Bacillus Subtilis*, l'hôte naturel qu'infecte le bactériophage PBS-2, l'Ugi inhibe l'UNG d'EBV avec une constante de 8 nM. Cette excellente affinité résulte de la spécificité de l'Ugi pour 4 motifs du site actif de l'UNG ainsi que

pour un coude- β structurant le site actif.

10.2 Article 2

JMBAvailable online at www.sciencedirect.com ScienceDirect

New Insights on the Role of the γ -Herpesvirus Uracil-DNA Glycosylase Leucine Loop Revealed by the Structure of the Epstein-Barr Virus Enzyme in Complex with an Inhibitor Protein

Thibault Géoui¹, Marlyse Buisson^{1,2}, Nicolas Tarbouriech¹
and Wim Pascal Burmeister^{1,3*}

¹Institut de Virologie
Moléculaire et Structurale
FRE 2854 CNRS-UJF, BP 181
F-38042 Grenoble cedex 9
France

²Laboratoire de Virologie
CHU Michallon, BP 217
F-38043 Grenoble cedex 9
France

³Institut Universitaire de
France, 103 bd Saint-Michel
F-75005 Paris, France

Epstein-Barr virus (EBV) is a human gamma-herpesvirus. Within its 86 open reading frame containing genome, two enzymes avoiding uracil incorporation into DNA can be found: uracil triphosphate hydrolase and uracil-DNA glycosylase (UNG). The latter one excises uracil bases that are due to cytosine deamination or uracil misincorporation from double-stranded DNA substrates. The EBV enzyme belongs to family 1 UNGs. We solved the three-dimensional structure of EBV UNG in complex with the uracil-DNA glycosylase inhibitor protein (Ugi) from bacteriophage PBS-2 at a resolution of 2.3 Å by X-ray crystallography. The structure of EBV UNG encoded by the BKRF3 reading frame shows the excellent global structural conservation within the solved examples of family 1 enzymes. Four out of the five catalytic motifs are completely conserved, whereas the fifth one, the leucine loop, carries a seven residue insertion. Despite this insertion, catalytic constants of EBV UNG are similar to those of other UNGs. Modelling of the EBV UNG–DNA complex shows that the longer leucine loop still contacts DNA and is likely to fulfil its role of DNA binding and deformation differently than the enzymes with previously solved structures. We could show that despite the evolutionary distance of EBV UNG from the natural host protein, bacteriophage Ugi binds with an inhibitory constant of 8 nM to UNG. This is due to an excellent specificity of Ugi for conserved elements of UNG, four of them corresponding to catalytic motifs and a fifth one corresponding to an important beta-turn structuring the catalytic site.

© 2006 Elsevier Ltd. All rights reserved.

Keywords: Epstein-Barr virus; DNA repair; uracil-DNA glycosylase; lytic protein; crystal structure

*Corresponding author

Introduction

Epstein-Barr virus (EBV) is a human γ -herpesvirus carrying a double-stranded DNA genome coding for 86 expressed open reading frames.^{1,2} EBV primo-infection can lead to infectious mononucleosis but it often goes unnoticed, in particular when it occurs during childhood. After the acute phase of the infection, EBV persists lifelong in B

lymphocytes so that more than 90% of the world population is infected.³ EBV infects B lymphocytes and epithelial cells. *In vitro*, EBV has the ability to immortalize B lymphocytes, *in vivo* it can lead to an immunoproliferative syndrome,⁴ a life threatening complication for patients whose immune system is compromised, as for transplanted or AIDS patients. EBV is associated with certain cancers, in particular with endemic nasopharyngeal cancer in Asia and Burkitt's lymphoma in Africa.⁵

Uracil-DNA glycosylase (UNG) is a DNA repair enzyme and, together with the 5'-triphosphate pyrophosphatase (dUTPase), it keeps the genome free of uracil bases. dUTPase maintains a low dUTP/dTTP ratio, whereas UNG removes uracil from the DNA. Uracil resulting from accidental incorporation of

Abbreviations used: EBV, Epstein-Barr virus; UNG, uracil-DNA glycosylase; Ugi, uracil-DNA glycosylase inhibitor.

E-mail address of the corresponding author:
wpb@embl-grenoble.fr

dUTP by a polymerase is not miscoding in contrast to uracil produced by the deamination of cytosine, the second source of uracil in DNA. Such sponta-

neous deamination events occur at a rate of 500 times per day in proliferating cells.⁶ The deamination leads to a C:G→T:A transition mutation if it is

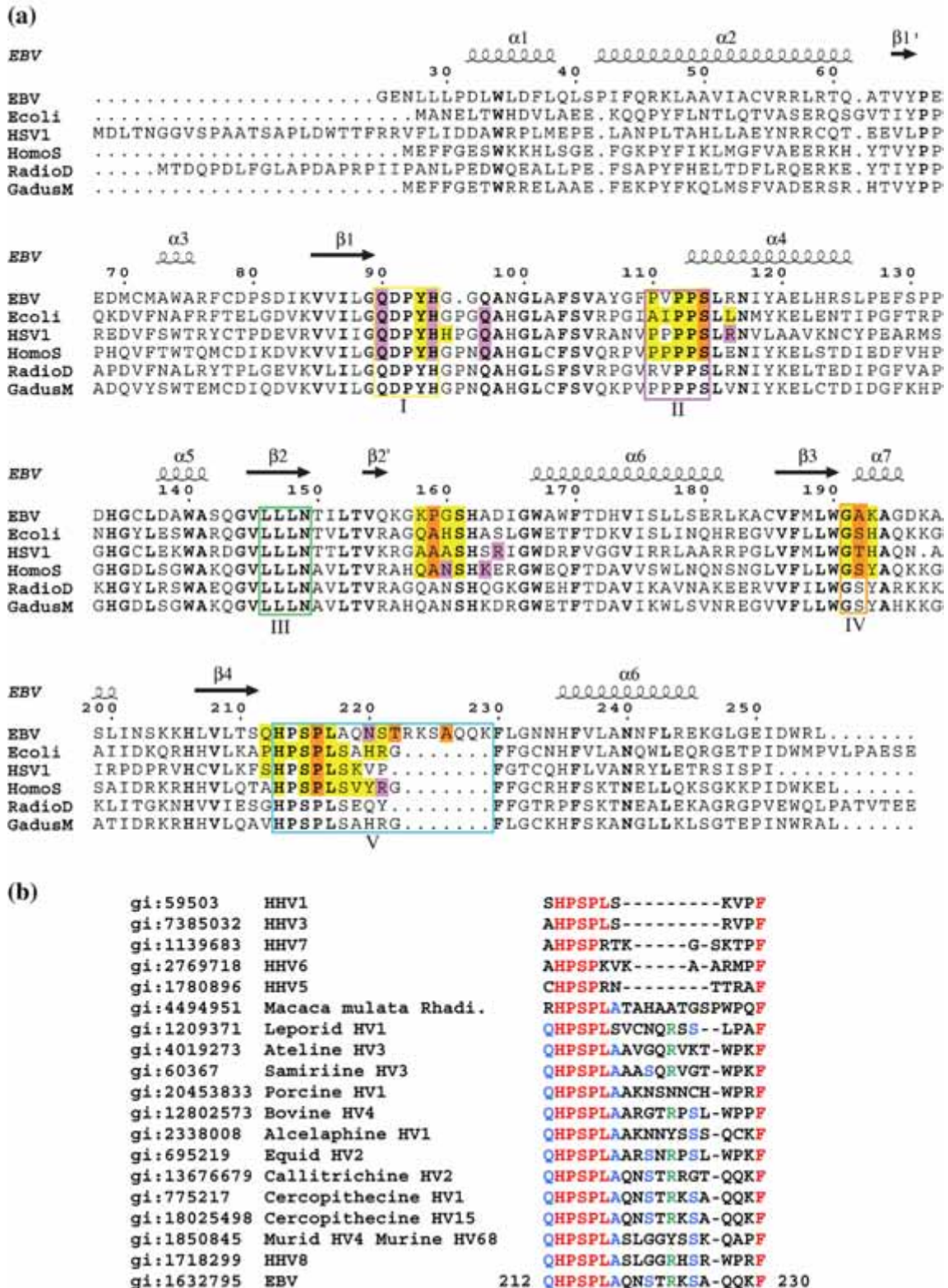


Figure 1 (legend on opposite page)

not repaired before the next round of replication. Thus DNA repair mechanisms have been developed that clean the genome from uracil in order to maintain its integrity. Base excision repair (BER) activity releasing uracil was first identified in *Escherichia coli*,⁷ meanwhile it has been shown that almost all organisms have enzymes with UNG activity.⁸

UNG is encoded and expressed by DNA viruses such as herpesvirus and poxvirus, whereas the cellular enzyme is encapsulated by some retroviruses such as HIV-1.⁹ It is surprising that viruses use their own UNG regarding that, in most of the cases, (i) it is only the first step in the DNA repair pathway, hence leaving it unclear why the virus provides only this enzyme and (ii) it is often dispensable for viral replication depending on the model that has been used. For example it has been shown that in a culture of neuronal cells, herpes simplex virus 1 (HSV-1) deficient for UNG is 10⁵ times less neuro-invasive.¹⁰ Cytomegalovirus (CMV) and vaccinia virus deleted in the UNG gene were impaired when growing in fibroblasts,^{11,12} whereas it could be shown that varicella zoster virus (VZV) needs neither cellular nor viral UNG activity in fibroblasts.¹³ A partial explanation of the contradictory results may be that the virus only needs its UNG when growing in quiescent cells where DNA repair enzymes are not, or only to a lesser extent, produced.¹⁴

UNGS are classified into four families based on phylogenies and substrate specificity.^{8,15} They all share a common core structure: EBV UNG belongs to the first family, which acts on double-stranded DNA containing uracil. Their structures are known for a number of organisms: HSV-1¹⁶ *Homo sapiens*,¹⁷ *Escherichia coli*,¹⁸ arctic cod *Gadus morhua*¹⁹ and *Deinococcus radiodurans*.²⁰ Family 2 is specific of G:U/T mismatches, its structure is represented by the bacterial mismatch-specific uracil-DNA glycosylase (MUG).²¹ Family 3 contains enzymes with a broader specificity for oxidized pyrimidines, its structure is represented by the (so-called) single-stranded monofunctional uracil-DNA glycosylase (SMUG1).²² Finally, family 4 corresponds to archeal UNGs²³ with a characteristic iron-sulfur cluster and is represented by the UNG of *Thermus thermophilus*.²⁴

Within family 1 enzymes, structures are very similar and they present five conserved sequence motifs (Figure 1(a)): the so-called 4-Pro loop and the Gly-Ser loop which together form the proline-pinch interacting with the phosphate backbone at the and

the side of the DNA strand carrying the base being questioned by UNG. Together with the leucine loop, these elements are supposed to locally compress the B-DNA and lower the activation energy in order to flip out the base.²⁵ In addition the leucine loop helps to widen the minor groove by water-mediated interactions with bases adjacent to the flipped base. Furthermore the leucine loop carries a residue that intercalates into the DNA at the position of the flipped-out base. It has been described as a “door-stop” preventing the flipped-out nucleotide from going back to its original position.²⁵ The identity of the flipped-out base is verified by a residue of the “uracil-recognition β 2-strand” and other residues from the leucine loop and the “water activating loop”. This loop provides the residues that position and activate a catalytic water molecule for a nucleophilic attack on the C1' atom of the ribose and hydrolysis of the N-glycosylic bond.^{16,25}

Structures have also been solved in complex with the uracil-DNA glycosylase inhibitory protein (Ugi),^{18,26-27} which is produced by *Bacillus subtilis* bacteriophages PBS-1 and PBS-2.²⁹ These phages use uracil instead of thymine in their DNA genomes,³⁰ which are protected from cellular UNG by the expression of Ugi. Although this protein is primarily intended to inhibit bacterial UNG it has been shown to inhibit UNGs from various organisms³¹ with the exception of vaccinia virus UNG.¹²

Here we report the 2.3 Å crystal structure of a complex between EBV UNG and Ugi. We determined kinetic constants for its activity of uracil excision from a fluorogenic double-stranded DNA substrate containing an U-A base-pair and kinetic constants of the Ugi inhibition.

Results

Overall structure

Only the UNG with a 24 residue N-terminal deletion (UNG Δ 24) complexed to Ugi yielded crystals, although numerous crystallisation conditions were screened using full-length UNG and UNG Δ 24 alone or in complex with Ugi. Furthermore we attempted in vain the crystallization of complexes with heteroduplex DNAs containing a uracil base in order to take advantage of the relatively high affinity of UNG for abasic sites created after the excision of uracil.³²

Figure 1. Sequence alignment. (a) Sequences from EBV, *E. coli* (pdb 1LQG), HSV-1 (pdb 1UDI), human (pdb 1UGH), *G. morhua* (pdb 1OKB) and *D. radiodurans* (pdb 2BOO) have been used. Strictly conserved residues are printed in bold. Residues forming the five motifs involved in DNA binding and enzymatic activity are boxed: the water activating loop (I, yellow), the 4-Pro loop (II, pink), the uracil recognition β 2-strand (III, green), the GS-loop (IV, orange) and the leucine loop (V, blue). Residues interacting with Ugi are shaded in yellow for contacts, in orange for main chain hydrogen bonds and in pink for side-chain hydrogen bonds (obtained from DIMPLOT). The secondary structure elements of EBV UNG are indicated above the alignment. (b) Alignment of the sequences of the leucine loop within the herpesviruses. Residues conserved within family 1 UNGs are printed in red, the residues with an important role in the loop conformation in EBV and their equivalents within the γ -herpesvirus family are printed in blue and Arg223 and possible equivalents in green.

Table 1. Crystallographic data collection and refinement statistics

<i>Data collection</i>	
Resolution (Å)	67–2.30 (2.42–2.30)
Number of reflections	169,145 (23,926) ^a
Unique reflections	29,123 (4095)
Multiplicity	5.8 (5.8)
Completeness (%)	92 (91)
R_{sym}	0.101 (0.348)
Wilson plot B factor (Å ²)	33
<i>Refinement</i>	
Resolution (Å)	50–2.30 (2.36–2.30)
R_{free}	0.253 (0.281)
R_{cryst}	0.196 (0.234)
rms deviation from ideal	
Bond lengths (Å)	0.004
Bond angles (°)	0.83
Average atomic B factor (Å ²)	24.2

^a Values in parentheses concern the highest resolution bin.

Crystals of the UNG Δ 24–Ugi complex belong to space group C222 with cell parameters of $a=62.4$ Å, $b=82.9$ Å, $c=269.1$ Å. The model contains two UNG–Ugi complexes in the asymmetric unit with a total of 623 amino acid residues, 389 water molecules and two urea molecules. The urea molecules have been identified from the electron density and the hydrogen bonding pattern, but could alternatively be interpreted as a guanidinium molecule and probably result from a contamination of one of the chemicals. The initial molecular replacement solution using the *E.coli* UNG–Ugi complex³³ (pdb 1LQM) with a correlation of 0.29 and a R_{cryst} of 0.53 using data between 67 and 3 Å resolution has been rebuilt and refined at 2.3 Å resolution (see Table 1) to an R_{cryst} of 19.8% ($R_{\text{free}}=25.7\%$). Two residues lie in generously allowed regions of the Ramachandran plot and the two residues Phe103 of the active site in disallowed regions as observed in the other UNG structures. Due to a more dense packing inside the crystal, one UNG–Ugi complex (chains A and B in the pdb entry, average $B=16$ Å²) is much better ordered than its counterpart (chains C and D in the pdb entry, average $B=34$ Å²) and has been used for analysis and structure comparison. Molecules can be superposed onto their counterpart with an rms deviation of 0.4 Å for C α atoms. EBV UNG (Figure 2(a)) has an α/β -fold consisting of a central β -sheet with four parallel strands surrounded by eight α -helices and two additional small β -strands ($\beta 1'$, $\beta 2'$) as observed in the known UNG structures. The superposition of the EBV UNG structure with human²⁶ (pdb 1UGH),

*E. coli*³³ (1LQG) and HSV-1 (pdb 1UDI)²⁷ UNG in complex with Ugi shows that about 205 C α atoms of the UNG domain superpose with an overall rms deviation of about 1.0 Å (Figure 2(b)). The major deviations concern helices $\alpha 1$ and $\alpha 2$ at the N terminus (Figure 2(a) and (b)). In the same superposition, Ugi C α atoms superpose with an rms deviation between 0.5 Å and 1 Å showing a complete conservation of the position of Ugi relative to UNG in the different complex structures.

The main difference of EBV UNG structure lies in an insertion of seven residues (residues 223–229) into the leucine loop (Figures 1 and 2(b)). The leucine loop consists of a conserved N-terminal HPSPL motif and a basic hydrophilic loop, which ends before a strictly conserved phenylalanine residue. The strictly conserved doorstep residue Leu217 protrudes from the loop. The structure of the leucine loop is stabilized by a network of interactions around Gln212 involving on one side the hydroxyl group of Ser225 and the carbonyl oxygen of Gln227 and on the other side the hydroxyl group of Ser221 and the carbonyl oxygen of Ala218 (Figure 3(a)).

UNG–Ugi interaction

The interaction of UNG with Ugi can be divided into four conserved regions (Figure 1, Table 2), which correspond approximately to the areas observed to bind DNA²⁵ except for the uracil-recognition $\beta 2$ -strand, which is not involved in Ugi binding. A fifth area of interaction involves residues around Pro159 of EBV UNG. Interactions with Ugi involve a contact surface of 1180 Å², which correspond to 13% of the UNG and 28% of the Ugi solvent accessible surface.

The leucine loop region has a contact surface of 560 Å² with Ugi. This interaction accounts for about 47% of the contact surface whereas the interaction with the other motifs that form the DNA binding groove create the remaining 53% or 620 Å² of the contact surface (Table 2). Almost half of the contact surface is contributed by polar atoms. With almost 150 Å² of mainly hydrophobic contact surface, the doorstep residue Leu217 accounts for 13% of the total interacting surface. Leu217 is nested in a hydrophobic cavity formed by the side-chains of eight residues of Ugi (Met24, Val29, Val32, Ile33, Val43, Met56, Leu58 and Val71, Figure 2(d)).

At the N-terminal side of the loop, the carbonyl oxygen of Pro216 makes a hydrogen bond to the Ugi Gln73 side-chain and hydrophobic interactions

Figure 2. Structure of the EBV UNG–Ugi complex. (a) Secondary structure, UNG is shown in light blue, its β -strands in blue. Motifs involved in the catalytic activity of UNG are coloured as in Figure 1(a). Ugi is shown in white, its first β -strand and second α -helix in black. (b) Stereoview of the structural alignment of UNGs in complex with Ugi. EBV UNG (white) is superimposed with human UNG (1UGH, green), *E.coli* UNG (pdb 1LQG, red) and HSV-1 UNG (pdb 1UDI, yellow). (c) Charge complementarity of UNG (left) and Ugi (right) coloured according to the surface potential for the individual molecules ranging from +15 kT/e (blue) to –15 kT/e (red). The C α backbone of the partner molecule is shown in white. (d) Stereoview of the binding of Leu217 (red, with solvent accessible surface in mesh representation) in the hydrophobic pocket (light green) formed by Ugi (gray). A modelled phenylalanine residue is shown in yellow together with its surface. The backbone of the leucine loop is shown in blue.

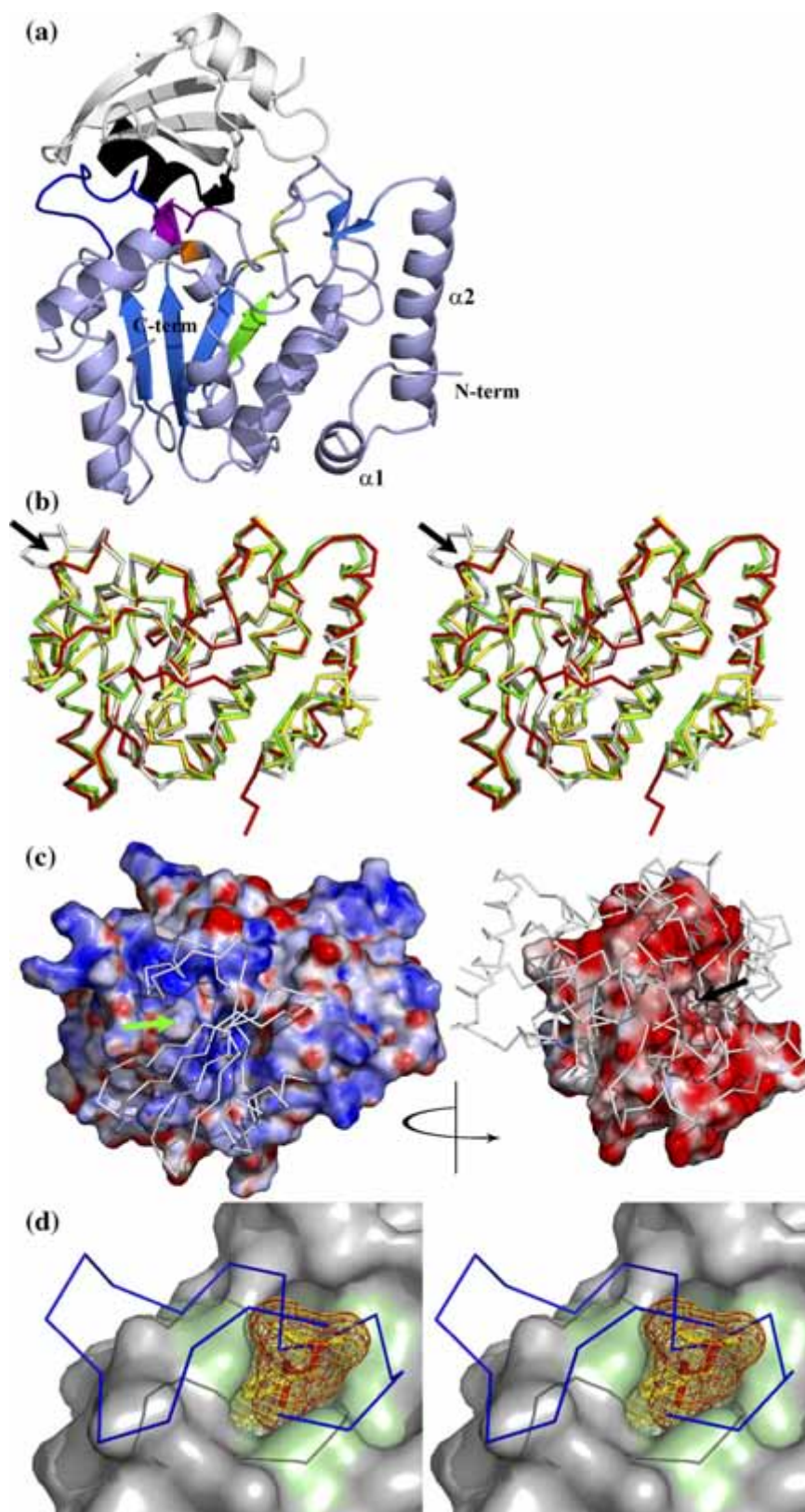


Figure 2 (legend on opposite page)

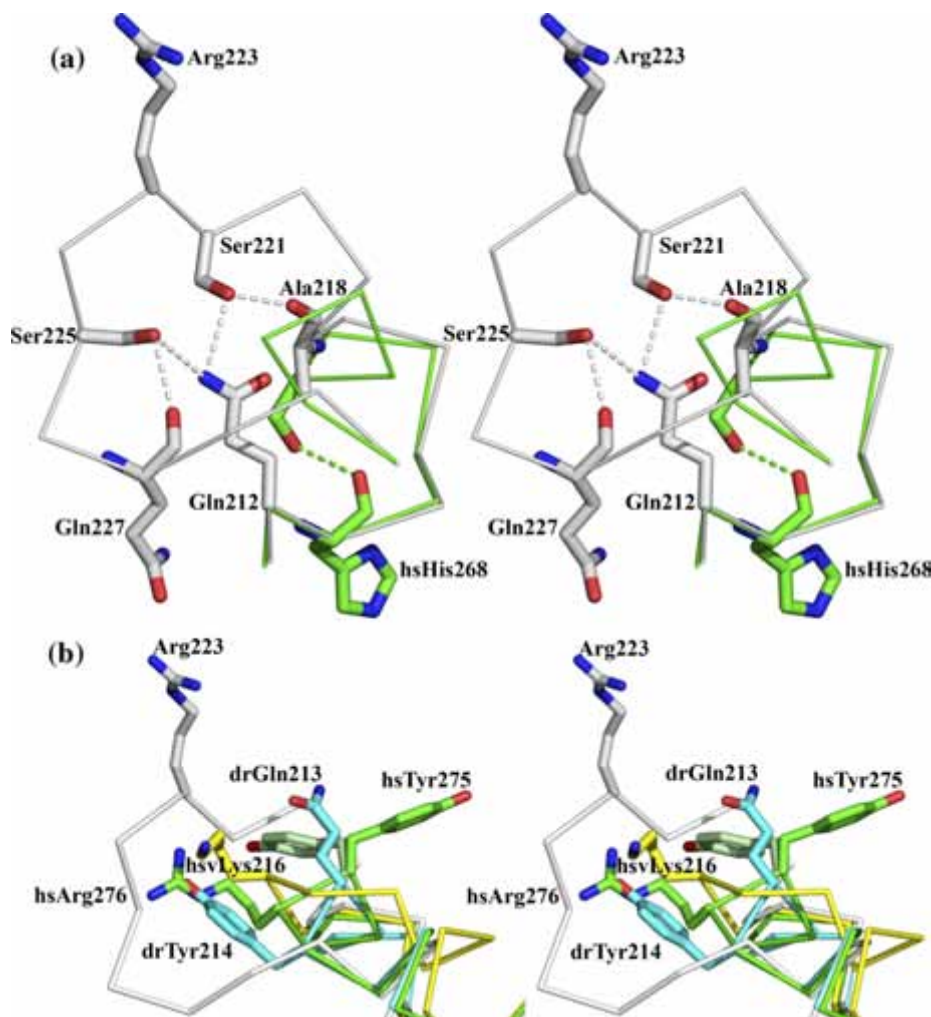


Figure 3. The leucine loop. Stereoviews. (a) The C α backbone is shown, residues involved in the stabilisation of the EBV loop (white) and the human loop (green) are shown as sticks. Hydrogen bonds are shown as dotted lines. For clarity, hsSer273 is not labelled. (b) The leucine loop in UNG–Ugi complexes from EBV (white), *H. sapiens* (green) and HSV-1 (yellow) and from the free *D. radiodurans* protein (cyan). The side-chains of the *H. sapiens* enzyme contacting the minor groove are shown as sticks and labelled (hs), together with side-chains of the *D. radiodurans* (dr) and HSV-1 (hsv) enzymes that may play similar roles. hsTyr275 is present as a different rotamer (light green) in the human UNG–DNA complex.

involve conserved His213 and Ser215 in addition to the major contact with Leu217 (Table 2). The seven residue insertion into the EBV UNG leucine loop extends the interactions with Ugi compared to other UNGs (Figures 1 and 3(b), Table 2). The side-chain of UNG Asn220 interacts with the side-chain of Ugi Gln73 as well as the main chain of Gly77. At the C-terminal end of the leucine loop insertion, the main chain of Thr222 and Ala226 interact with Glu31 of Ugi.

The electrostatic surface of UNG presents a positively charged patch along the DNA binding region as well as on the leucine loop (Figure 2(c)). Ugi is negatively charged, with a very acidic pI of 4.2.³⁴ Nevertheless, no ionic interactions are present in the EBV UNG–Ugi complex as observed for the other complex structures analyzed so far where only the *E. coli* complex²⁸ (pdb 1UUG) features

one ionic interaction between His134 and Ugi Asp61.

Enzymatic activity

We determined the catalytic constants of the enzyme using an assay based on the fluorescence increase of a 2-aminopurine residue adjacent to the cleaved uracil base both incorporated into a synthetic double strand oligonucleotide (Figure 4(a) and (b)) as described by Bellamy and Baldwin.³⁵ In this way the substrate uracil bases are present in only one type of context in contrast to assays that are based on the uracil liberation from uracil containing DNA prepared by nick translation.³⁶ Using a substrate containing an adenine base facing uracil, we determined a K_m of 1.5 μM and 4.4 μM for UNGwt and UNG Δ 24. The catalytic constant k_{cat} is

Table 2. Interactions in the EBV UNG–Ugi complex

UNG	Motif	Ugi
<i>Hydrogen bond</i>		
Gln90 N¹², O^{1a}	I	Leu23 N, O
His94 N¹²	I	Ser21 O ⁷
<i>Ser114 N, O⁷</i>	II	Glu20 OE1, OE2
<i>Pro159 O</i>		Tyr65 O ¹
<i>Ala192 N</i>	IV	Glu28 O ²
Pro216 O	V	Gln73 N ¹²
<i>Asn220 O^{δ1}</i>	V	Gln73 N ¹²
<i>Asn220 N^{δ2}, O^{δ1}</i>	V	Gly77 O
<i>Thr222 N</i>	V	Glu31 O
<i>Ala226 N</i>	V	Glu31 O ²
<i>van der Waals contact</i>		
<i>Pro110</i>	II	Gln19
Pro112	II	Glu20
Pro113	II	Gln19, Glu20, Thr45
<i>Lys158</i>		Tyr65
<i>Pro159</i>		Ser21, His44, Ala62, Tyr65
Gly191	IV	Glu28
<i>Ala192</i>	IV	Glu28
<i>Lys193</i>		Leu23
<i>Gln212</i>		Glu31
His213	V	Ile22
Ser215	V	Met24
Leu217	V	Val32, Ile33, Met56
<i>Gln219</i>	V	Gln73
<i>Asn220</i>	V	Asn79
<i>Ser221</i>	V	Glu31
<i>Ser225</i>	V	Glu31

^aStrictly conserved residues of UNG are printed in bold, those whose main chain atoms are involved in hydrogen bonds in italics. Motifs are numbered as in Figure 1(a). Contacts have been analyzed with DIMPLLOT.⁵⁰

3.9 s⁻¹ and 1.7 s⁻¹ for UNGwt and UNGΔ24, respectively. Ugi acts as a competitive inhibitor with a K_i of 10 nM for UNGwt and 8 nM for UNGΔ24 (Figure 4(c)).

Discussion

Overall structure

The 29 amino acid residue N-terminal extension of EBV UNG preceding the sequence of the UNG domain may carry a nuclear localisation signal (sequence KRKQ) based on a sequence homology between EBV UNG and human UNG2.³⁷ As the deletion of the 24 N-terminal residues reduces the catalytic activity of the enzyme in the Δ24 mutant this extension may also interact with DNA.

The comparison of EBV UNG domain with the known family 1 UNG structures shows a conservation of all the structural motifs and an overall rms deviation of equivalent C^α atoms in the range of 1 Å when around 205 C^α atoms of the UNG domain are superposed. Uracil recognition and catalytic machinery are particularly conserved. The uracil recognition site is composed of Asn149 of the uracil recognition β2-strand, the backbone carbonyl of Gln90, Tyr93 which packs against the edge of the uracil ring, both belonging to the water activating

loop, His213 belonging to the leucine loop and Phe103 which stacks onto the uracil base. The residues and their positions are conserved in the other known UNG structures such as the human one (Table 3). Catalysis involves a reactive water molecule, which is observed in the EBV enzyme structure. It is coordinated with three conserved residues of the water activating loop: Asp91, the carbonyl oxygen of Pro92 and His94. The putative catalytic water molecule of the EBV UNG structure is suitably placed for a nucleophilic attack on the C1' atom of the ribose allowing hydrolysis of the β-glycosylic linkage with inversion of configuration. In a complex between a human Leu272Ala mutant and DNA²⁵ (pdb 2SSP) a similar water molecule is located next to C1' of the ribose of the abasic site and opposite to the O1' atom, which is in a β-anomeric conformation in this complex.

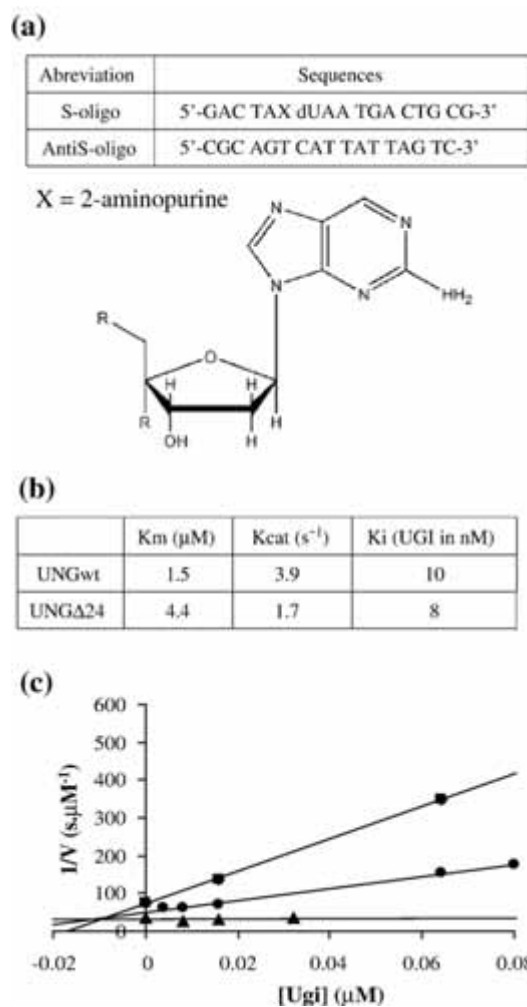


Figure 4. Enzyme kinetics. (a) Oligonucleotides and fluorescent label used in the enzyme assay. (b) Overview of the kinetic constants. (c) Dixon plot of the UNGΔ24 inhibition by Ugi, with the concentration of fluorescent DNA substrate ranging from 1 μM (▲) over 2 μM (●) to 6 μM (■).

Table 3. Interactions of human UNG with DNA and in a modelled EBV UNG–DNA complex

DNA base	Motif	Human UNG	EBV UNG (modelled DNA complex ^a)	Interaction type
<i>Uracil containing strand</i>				
Thy4	V	Pro271	Pro216	VdW ^b
Ura5 PO ₄	II	Pro167	Pro112	vdW
Ura5 PO ₄	II	Pro168	Pro113	vdW
Ura5 PO ₄	II	Ser169 N ^{o1}	Ser114 N ^{o1}	H-bond
Ura5 PO ₄	II	Ser169 O ^y	Ser114 O ^y	H-bond
Ade6 PO ₄	V	Ser270 O ^y	Ser215 O ^y	H-bond
Ade6 N3	V	Leu272 O	Leu217 O	H ₂ O mediated
Ade6 N3	V	Arg276 N ^e		H ₂ O mediated
Ade6 N3	V		Ser221 O ^y	H ₂ O mediated ^c
Thy7 PO ₄	V	Ser273 O ^y		H-bond
Thy7 PO ₄	IV	Gly246	Gly191	vdW
Thy7 PO ₄	IV	Ser247 N	Ala192 N	H-bond
Thy7 PO ₄	V	His268 N	His213 N	H-bond
Thy7 backbone	V	Arg276		vdW
Thy7 backbone	V		Gln212	vdW?
Cyt8 PO ₄	V		Ala226 O	H-bond?
Cyt8 PO ₄	V		Gln227 O ^{e1} , N ^{e2}	H-bond?
<i>Complementary strand</i>				
Thy26 PO ₄	V		Arg223	Ionic interaction?
Ade27 N3	V	Leu272 O	Leu217 O	H ₂ O mediated
Ade27 N3	V	Tyr275 O ^{o1}		H ₂ O mediated
Ade27 N3	V		Asn220	H ₂ O mediated?
Ade27 backbone	V		Thr222	?
Ade27 PO ₄	V		Arg223	Ionic interaction?
Ade28 N3	V	Pro271 O	Pro216	H ₂ O mediated
Ade28 ribose	V	Tyr275		vdW
Ade28 ribose	V		Gln219	vdW?
Ade28 ribose	V		Asn220	vdW?
Cyt29 PO ₄	V		Gln219	H-bond?
Gua31 PO ₄	V		Arg116	Ionic interaction?
<i>Uracil base</i>				
Uracil		Phe158	Phe103	Stacking
Uracil N1	V	His268 N ^{e2}	His213 N ^{e2}	H-bond
Uracil O2	I	Gln144 N	Gln90 N	H-bond
Uracil O2	V	His268 N ^{e2}	His213 N ^{e2}	H-bond
Uracil N3	III	Asn204 O ^{o1}	Asn149 O ^{o1}	H-bond
Uracil O4	III	Asn204 N ^{o2}	Asn149 N ^{o2}	H-bond
Uracil C5, C6	I	Tyr147	Tyr93	vdW
<i>Catalysis</i>				
Catalytic H ₂ O	I	Pro146 O	Pro92 O	H-bond
Catalytic H ₂ O	I	His148 N ^{o1}	His94 N ^{o2}	H-bond
Catalytic H ₂ O	I	Asp145 O ^{o1}	Asp91 O ^{o1}	H-bond

^a The model of the EBV UNG–DNA complex has been obtained as described in the text. Conserved interactions are printed in bold.

^b vdW: van der Waals contact. Motifs are numbered as in Figure 1(a).

^c Interactions only predicted from the modelling are marked with a question mark.

Some conservative changes affect the proline pinch, which are discussed below in the context of enzyme–DNA contacts. EBV UNG differs due to the insertion of seven amino acid residues into the leucine loop. Whereas in the other known structures, such as the one of the human enzyme, this loop is stabilized by a hydrogen bond between the carbonyl oxygen of His268 on the N-terminal side of the doorstop leucine residue and the hydroxyl group of Ser273 located at the C-terminal side (Figure 3(a)), the serine residue is changed to alanine leading to a loss of this particular hydrogen bond. The larger loop in EBV UNG is stabilized by a network of interactions around Gln212 bridging as before residues located N-terminally and C-terminally of the doorstop residue. The stabilization due to this hydrogen bond network leads to a well defined structure for

this loop reflected by an identical conformation for the two molecules in the asymmetric unit, although, in the second monomer, residues Thr222 and Arg223 lack most of their electron density.

The leucine loop has several roles: (i) to carry the leucine residue that intercalates into the DNA when the uracil base is flipped out, (ii) to provide additional interactions with the backbone to compress the helix in order to favour bending and base flipping and (iii) to widen the minor groove next to the uracil base in order to facilitate base flipping²⁵ and to increase the interactions with double-stranded DNA.³⁸ Whereas the first two functions are located on the part of the loop that is structurally conserved (sequence HPSPL in EBV), the last function is located on the more variable C-terminal side of the loop. Here the enzymes from HSV-1, *D. radiodurans* and

EBV (or γ -herpesviruses in general) adopt obviously different solutions (Figures 1(b) and 3(b)). For human UNG the importance of residues located in this part of the loop had been shown by the drastic reduction of the catalytic efficiency for the excision of uracil bases in the context of double-stranded DNA when Arg276 is mutated³⁸ while activity against single-stranded DNA was less affected.

A comparison of the sequences of the herpesvirus UNG leucine loops (Figure 1(b)) shows, that γ -herpesviruses share this loop structure (Figure 3(b)). In β -herpesviruses arginine or lysine residues replace the doorstep leucine residue and the residues that are important for the stabilisation of the leucine loop either in the HSV-1 or in the EBV enzyme are not conserved. In the γ -herpesvirus family, two members have leucine loops of differing lengths; the one for *Macaca mulata* rhadinovirus, which is larger and probably organized around an arginine residue replacing the conserved glutamine residue 212 and the loop of leporid herpesvirus 1, which is shorter, but still conserves most of the important residues. The other γ -herpesviruses largely conserve Ser221 and Ser225, which hydrogen bond to Gln212. Ser221 may be replaced by glycine residues with the loss of a hydrogen bond or by asparagine residues, which could play a similar role as serine in the formation of hydrogen bonds to Gln212. Ser225 is sometimes replaced by glycine, leading again to a loss of a hydrogen bond or more rarely by cysteine or lysine. Ala218 is also strictly conserved, its role as a small residue oriented towards the inside of the loop probably does not allow the accommodation of a bigger side-chain. Arg223 is the only other residue of the insertion that shows a wide conservation, which may indicate a functional role in the interaction with DNA.

Despite the different structure of the leucine loop, the catalytic efficiency of the EBV enzyme appears very similar to the one of herpes simplex virus. A K_m of 1.5 μM for the full length EBV enzyme and 4.4 μM observed for the truncated form has been observed, which compares to 1.6 μM for the HSV-1 enzyme obtained using the same assay.³⁵ At 37 °C, the catalytic constant k_{cat} is 3.9 s^{-1} and 1.7 s^{-1} for UNGwt and UNG Δ 24, respectively, compared to the 2.1 s^{-1} found for HSV-1 at 25 °C. Still herpesvirus UNG has a lower activity than human UNG, where in the same assay $K_m=0.12 \mu\text{M}$ and $k_{\text{cat}}=7.4 \text{s}^{-1}$ have been found.³² The additional interactions of the leucine loop in the EBV enzyme do not seem to fully compensate for the lost interactions compared to the human enzyme (Tyr275 and Arg276), as discussed below in the section about the interactions with DNA, and do not restore catalytic efficiency completely. Other subtle, so far not identified differences between herpesvirus and human UNG may also be responsible for the difference in catalytic constants.

UNG-Ugi interaction

We observed a $K_i=8 \text{ nM}$ of Ugi for wild-type EBV UNG. But as numerical values for the inhibition

constant of Ugi have so far only been published for the *E. coli* enzyme where they vary from an essential irreversible binding³⁹ over an IC_{50} around 34 pM ²⁶ to a K_i of 140 nM ⁴⁰ it is impossible to put the value into context and to correlate it with the tri-dimensional structure of the complexes. The contact surface of 1180 \AA^2 between UNG and Ugi is larger than the one for *E. coli* and HSV-1 (1030 \AA^2) and human (1140 \AA^2) UNGs, principally due to an increased buried surface involving the leucine loop, which carries an insertion in the EBV enzyme. This surface increases from 410 \AA^2 for HSV-1, 460 \AA^2 for *E. coli*, 470 \AA^2 for the human enzyme to 560 \AA^2 for the EBV enzyme.

Residues of Ugi interacting with UNG (Figure 1(a), residues printed on a shaded background) can be classified in three categories: (i) residues interacting with a conserved β -turn of UNG situated around Pro159. This turn is part of the environment of the catalytic residue Asp91, which is positioned by a hydrogen bond to the amide nitrogen of Gly160 and is partially shielded from the solvent by the C^β atom of Pro159. The turn is positioned by a hydrogen bond between the hydroxyl group of Ser161 and the amide nitrogen atom of Gln155. Ugi contacts the conserved elements of this turn by a hydrogen bond from the hydroxyl group of Tyr65 to the carbonyl oxygen atom of Pro159 and van der Waals interactions. Although not part of the classical catalytic motifs of UNG, this turn carrying the sequence motif [P/A]xSH is sufficiently conserved to be targeted by Ugi.

(ii) Residues forming the hydrophobic pocket binding the doorstep residue of UNG (Figure 2(d)). The doorstep residue in the natural partner of Ugi, UNG of *B. subtilis* is a phenylalanine residue, which is replaced by leucine in most of the UNG sequences; some sequences present an arginine residue (vaccinia virus, human cytomegalovirus); these are most likely not inhibited by Ugi as shown for vaccinia UNG.¹² The interaction of the doorstep residue with the eight hydrophobic side-chains of residues Met24, Val29, Val32, Ile33, Val43, Met56, Leu58 and Val71 of Ugi described by Putnam *et al.*²⁸ is completely conserved (Figure 2(d)). The binding pocket of Ugi is deeper than what would be needed for the interaction with leucine and could obviously accommodate a phenylalanine residue (Figure 2(d)). The interaction with the doorstep leucine residue buries about 140 \AA^2 of hydrophobic surface for the leucine residue and 220 \AA^2 at the level of Ugi, reflecting the large size of its pocket. Consequently, not all of the hydrophobic residues of Ugi are in direct van der Waals contact with Leu217 (Table 2).

(iii) Residues mimicking the phosphate backbone. These are the main chain amide groups of Gln19, Leu23 and the side-chain oxygen atoms of Glu20, Ser21 and Gln28 (pdb entry 1UUG).²⁸ These residues are involved in conserved interactions with EBV UNG (Table 2) with the exception of Gln19.

Despite the unique character of the insertion into the leucine loop in EBV UNG, it interacts with

additional residues from Ugi, which are Glu31, Val32, Gln73 and Gly77. Potentially, these interactions involving Asn220, Ser221, Thr222 and Ala226

on the UNG side (Table 2) increase the affinity of the inhibitor and could explain its excellent inhibitory effect. The ability of Ugi to bind EBV UNG with high

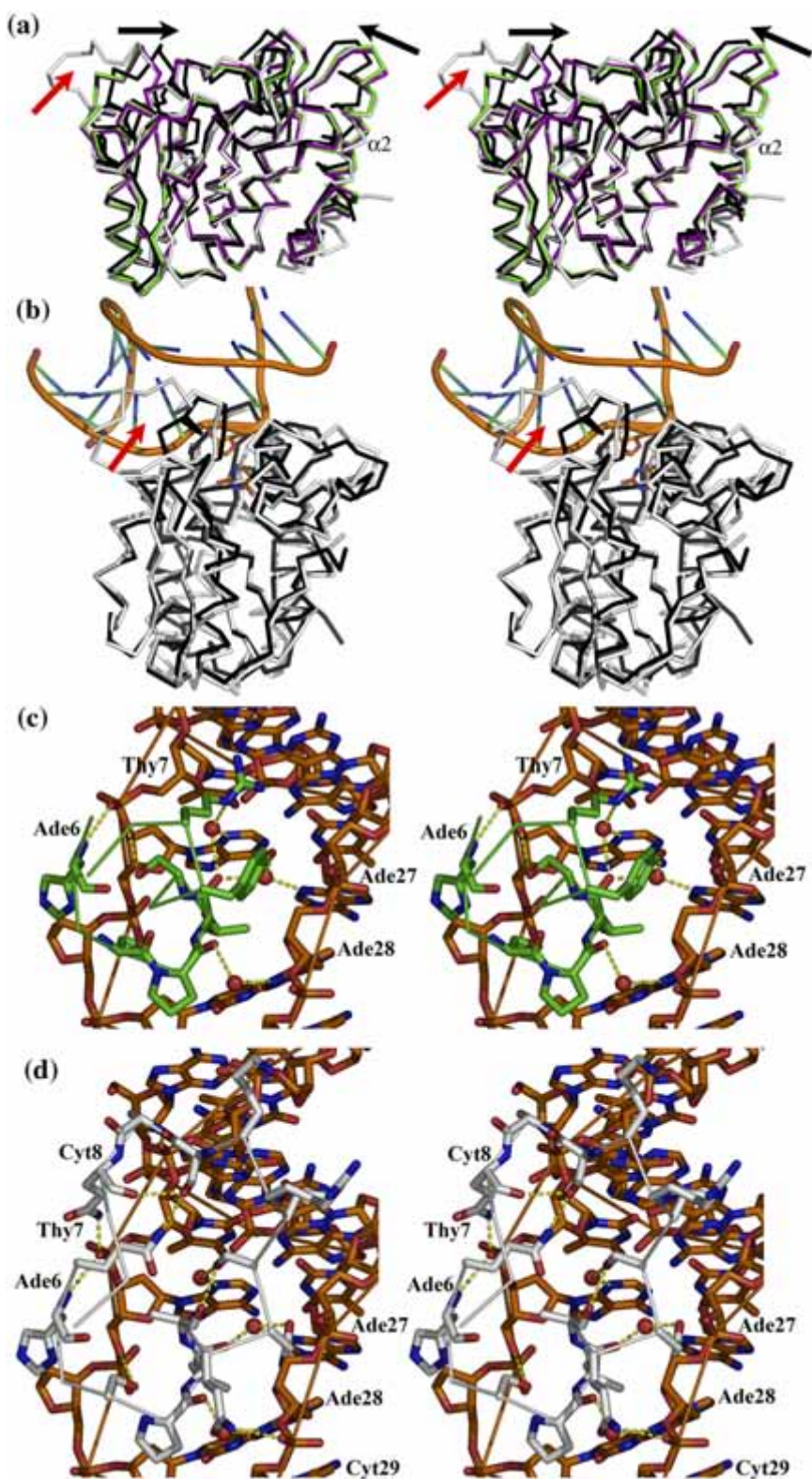


Figure 5 (legend on opposite page)

affinity despite the large evolutionary distance of the enzymes is surprising and shows the excellent adaptation of Ugi to the recognition of the conserved features of UNG. Compared to the natural *B. subtilis* host of bacteriophage PBS-2, the *H. sapiens* enzyme for example has a sequence identity of about 50% within the UNG domain,⁴¹ whereas for EBV UNG the identity drops to 42%. The enzymes from mycoplasma have similar sequence identities with the *B. subtilis* enzyme to herpesvirus UNGs whereas for poxviruses the sequence identity drops to 20 to 30%. Still the active sites of UNG show a remarkable conservation through evolution.

Model of the EBV interaction with DNA

Superposition of the free human UNG¹⁷ (pdb 1AKZ), the human UNG-Ugi complex²⁶ (pdb 1UGH) and the human UNG-DNA complex²⁵ (pdb 1SSP) shows a progressive closing of the enzyme structure. The difference between the unbound open state, and the Ugi bound intermediate state is very small, with an rms deviation of 0.4 Å for the 223 superimposed C α atoms, and is mainly characterized by a slight inward movement of the α 2-helix (Figure 5(a)). Upon binding of DNA, the whole structure appears to close itself on the substrate. The rms difference for C α positions between free UNG or UNG bound to Ugi on one hand and UNG bound to DNA on the other hand is about 1.5 Å. The structure of the EBV UNG-Ugi complex corresponds to the Ugi-bound intermediate state.

It is possible to model the interaction of the EBV enzyme with DNA based on the superposition of EBV UNG onto the human complex structure (Figure 5(b)). Concerning the DNA-binding motifs, it appears that in the 4-Pro loop the first two proline residues play a conformational role and are actually not in contact with DNA in the complex structure. For EBV UNG, the conformation of the loop is conserved but the second proline residue is changed to valine. The serine residue of the GS-loop is changed to an alanine residue, which would lead to the loss of a hydrogen bond with the DNA phosphate backbone. The uracil recognition β 2-strand and the water activating loop are fully conserved.

Furthermore it appears that despite the seven residue insertion, the leucine loop of EBV UNG is positioned over the DNA minor groove with only

one minor clash for Gln219 (Figure 5(b)). Due to the internal movements within UNG when it binds to DNA, the superposition of the structure around the leucine loop was not very good. It could be improved by a least-squares fit of atom positions of UNG of the conserved base of the loop to the corresponding ones in the human complex structure. The model has been refined further by re-orientations of side-chains before it was used for analysis (Figure 5(d), Table 3). In this way any clashes in the modelled complex could be avoided. While sequence and structure of the N-terminal part of this loop (HPSPL) are conserved, the C-terminal part is completely different for EBV UNG. The two prominent residues Tyr275 and Arg276 of the human sequence, which are described as key residues as they interact with DNA bases in the minor groove²⁵ (Figure 5(c)), no longer have structural equivalents. The importance of Arg276 for the enzyme activity towards uracil in double-stranded DNA has been shown by mutagenesis.³⁸ Whereas the human enzyme carries a tyrosine and an arginine residue, the tyrosine residue is replaced by a histidine in the *E. coli* and *G. morhuua* enzymes, which conserve the same conformation of the loop (not shown). For *D. radiodurans* and HSV-1, the conformation of the loop differs due to the deletion of a residue (Figures 1(a) and 3(b)). It is likely, that in these structures other residues play the role of Tyr275 and Arg276: for *D. radiodurans* the residues Tyr-Arg are probably swapped to Gln-Tyr maintaining the same function, whereas in HSV-1 UNG Lys216 probably replaces Arg276 of the human enzyme. All these residues are likely to interact directly or indirectly through water molecules with the bases of the minor groove. In EBV UNG the indirect interactions with the bases could involve Asn220 and Ser221 and additional interactions with the DNA backbone could use Gln212 and two stretches of residues: Gln219, Asn220, Ser221, Thr222 and Ala226 and Gln227 (Figure 5(d), Table 3). The low conservation of these residues within the γ -herpesviruses (Figure 1(b)) suggests that the interactions may involve peptide backbone atoms interacting through hydrogen bonds and van der Waals contacts with the phosphate backbone or ribose groups. The conservation of Arg223 within the γ -herpesviruses suggests a role for this residue. But in order to establish a contact with the DNA backbone the residue would not only have to change its orientation but also the conformation of the

Figure 5. Modelled EBV UNG-DNA interaction. Stereoviews. (a) Superposition of EBV UNG (white) with human UNG: free (violet, pdb 1AKZ), bound to Ugi (green, pdb 1UGH) and bound to DNA (black, pdb 1SSP). The progressive closing of the human UNG structure upon complex formation is indicated by black arrows; a red arrow points to the leucine loop. (b) Superposition of EBV UNG (white) with human UNG bound to DNA (black). The alignment is based on the equivalent C α atom positions. The DNA backbone from the human UNG-DNA complex is shown in orange together with the flipped-out ribose of the abasic site and the released uracil in the active site. A red arrow points to the leucine loop. (c) Interactions of the human UNG leucine loop (green carbon atoms) with DNA (orange carbon atoms). Interacting residues are shown as sticks, hydrogen bonds as yellow dotted lines. Water molecules mediating indirect contacts with the DNA are shown as spheres. (d) Similar view as in (c) of a modelled interaction of the EBV UNG leucine loop (white carbon atoms) with DNA (orange carbon atoms, bases are numbered as in (c)). The loop has been positioned onto the DNA by an alignment of conserved atom positions at its base and side-chain rotamers have been adjusted.

extremity of the leucine loop between Ser221 and Ser225. In conclusion it is likely that instead of the insertion of a tyrosine and an arginine residue into the minor groove next to the uracil base as observed for human UNG, EBV UNG rather inserts part of its leucine loop in order to splay it apart. It may not only use indirect interactions with the bases but also interactions with the backbone.

Conclusion

In general, EBV UNG confirms the extraordinary structural conservation of family 1 uracil DNA glycosylases. Still one of the motifs of the UNG active site, the leucine loop is highly variable between species and seems to adopt different solutions to fulfil its proposed role of minor groove binding and widening. In γ -herpesvirus UNG, the leucine loop carries a large insertion and appears to insert itself into the minor groove leading to kinetic constants similar to enzymes with a short leucine loop. The strong inhibition of EBV UNG by Ugi despite the evolutionary distance underlines the excellent specificity of Ugi for the interaction with conserved structural elements of UNG. In addition to the four classical motifs to which Ugi binds we could show that a fifth interaction corresponds to a conserved β -turn motif forming part of the active site.

Material and Methods

Cloning, protein expression and purification

The coding sequences for the wild-type EBV UNG (UNGwt) and a construct carrying an N-terminal 24 residue deletion (UNG Δ 24) were amplified by PCR from DNA extracted from B95-8 cells, ligated into the NcoI and the XhoI sites of the *PProEx HTb* vector (Invitrogen) and transformed into the *E.coli* BL21 (DE3) strain. Bacteria were grown at 37 °C in LB medium to an A_{600nm} of 0.6 before induction with 0.5 mM IPTG. The protein was overexpressed at 30 °C and cells are harvested 5 to 7 h post-induction by centrifugation. Bacteria were resuspended in 10 ml/g of binding buffer (20 mM Tris-HCl (pH 7.5), 100 mM NaCl, 20 mM imidazole) and lysed by sonication. His-tagged UNG was purified on a nickel column (Ni-NTA, Qiagen) and eluted with buffer containing 500 mM imidazole. The buffer was changed to binding buffer using a 10DG desalting column (Biorad) and incubated at room temperature overnight with 1:100 Tobacco Etch Virus protease (TEV) in order to remove the hexahistidine tag. The protein was passed again over a nickel affinity column and the flowthrough was concentrated to 0.5 ml before further purification over a Superdex-75 column (GE Healthcare) in 100 mM NaCl, 20 mM Tris-HCl (pH 7.5), 10 mM DTT. UNG was concentrated using Amicon Ultra-15 Centrifugal Filter Units (Millipore). The cleaved protein carries three extra N-terminal residues from the vector (Gly-Ala-Met). The final yield of purified protein was 1–2 mg of UNGwt and 7–10 mg of UNG Δ 24 per litre of cell culture.

Purification of Ugi and of the UNG–Ugi complex

Ugi was cloned in a *pRSETB* vector⁴² and was produced at 37 °C as described above. For the production of the complex bacteria expressing UNG Δ 24 and bacteria expressing Ugi were mixed prior to sonication and the complex was purified using nickel affinity chromatography and TEV cleavage as described above. The complex was purified further on a monoQ HR 10/10 column (GE Healthcare) using a NaCl gradient from 150 mM to 205 mM, 20 mM Tris-HCl (pH 8.0), 10 mM DTT and a Superdex-75 column (GE Healthcare) using 100 mM NaCl, 20 mM Tris-HCl (pH 7.5) and 10 mM DTT. For the preparation of pure Ugi, the complex was heated for 7 min at 90 °C. After centrifugation, the supernatant containing pure Ugi was harvested. Its purity was checked on 15% (w/v) Tris-tricine SDS-PAGE gels.

Glycosylase activity assay

A 17-mer heteroduplex DNA³⁵ (Figure 4(a)), containing a 2-aminopurine base next to a uracil residue, was used in a fluorescent assay. 2-Aminopurine fluorescence is quenched in the context of the DNA double strand and increases after the release of the neighbouring uracil base. All the enzymatic reactions were performed in a fluorimeter (Photon Technology International) using a 3 mm path length quartz cell with an excitation wavelength of 310 nm whereas emission was monitored at 375 nm. Lyophilized oligonucleotides (MWG-Biotech) were annealed in TN buffer (20 mM Tris-HCl (pH 8.0), 60 mM NaCl, 0.1 mM EDTA) by heating to 90 °C and overnight cooling.⁴³ Concentrations in the range from 0 μ M to 40 μ M based on an $\epsilon_{280} = 55,710 \text{ M}^{-1}\text{cm}^{-1}$ for UNGwt and $\epsilon_{280} = 44,710 \text{ M}^{-1}\text{cm}^{-1}$ for UNG Δ 24 were used. The kinetic parameters of UNGwt and UNG Δ 24, and the inhibition constant of Ugi have been determined at 37 °C in 20 mM Tris-HCl (pH 8.0), 100 mM NaCl, 10 mM DTT, 2.5 mM MgCl, 0.5 mg/ml bovine serum albumin (Pierce). Reactions were initiated by adding UNGwt or UNG Δ 24 to a final concentration of 25 nM. The inhibition constant (*K*) of Ugi was determined using Ugi concentrations from 0 nM to 100 nM based on a $\epsilon_{280} = 8480 \text{ M}^{-1}\text{cm}^{-1}$. The steady state velocity ($\mu\text{M s}^{-1}$) was calculated using the initial linear slope of the reaction curve normalized according to the substrate concentration using the following equation:

$$[P](t) = \frac{A_0 - A(t)}{A_0 - A_\infty} [S]$$
, where $[P](t)$ is the concentration of the product, $[S]$ the initial concentration of the substrate, $A(t)$ the fluorescence signal at a given time t , A_0 the initial signal and A_∞ the signal after completion of the reaction. Kinetic parameters were determined using Lineweaver-Burk plots.⁴⁴

Crystallization and data collection

Crystals of the UNG Δ 24–Ugi complex at a concentration of 30 to 50 mg/ml were grown at 20 °C in drops composed of 1 μ l protein solution and 1 μ l reservoir solution using the vapour diffusion method with a reservoir solution of 20% (w/v) PEG 3350 (Hampton Research) and 0.05 M NH_4Cl . Needle crystals grew in a week and were used for seeding of a fresh drop, yielding plate crystals that could be used for data collection. Crystals were dipped into paraffin oil (Panjelly kit, Jena Biosciences) and frozen directly at 100 K in the nitrogen gas stream (Oxford Cryosystems). Data were collected at

the European Synchrotron Radiation Facility (ESRF) on beamline ID23-1 at a $\lambda=0.98$ Å (see Table 1) using a MarMosaic 225 detector (marusa).

Molecular replacement and refinement

Data were integrated with MOSFLM⁴⁵ and further processed with the CCP4 package.⁴⁶ The structure was solved by molecular replacement using MOLREP.⁴⁷ Cycles of model building and refinement were performed with COOT⁴⁸ and REFMAC.⁴⁹ Final refinement included TLS tensors for each of the four polypeptide chains of the model. Statistics and model composition are shown in Table 1.

Structure analysis

Sequences were analyzed on the NPS@ server using BLAST,⁵⁰ aligned with EBV UNG using ClustalW⁵¹ and visualized with Esript.⁵² Interacting residues were identified with DIMPLOT.⁵³ Accessible surfaces and contact surfaces were calculated using the protein-protein interaction server⁵⁴ and the program AREAMOL of the CCP4 suite. Electrostatic surface potentials were calculated using the Adaptive Poisson-Boltzmann Solver⁵⁵ (APBS) together with the PyMol⁵⁶ APBS tools. Structures were superposed using LSQMAN.⁵⁷ The “Fit pairs” option of PyMol has been used for local alignments.

Protein Data Bank accession codes

Coordinates and structure factors have been deposited in the RCSB Protein Data Bank and are available under entry number 2J8X.

Acknowledgements

This work was undertaken as part of the European Union Framework Programme “Quality of Life and Management of Living Resources”, Integrated Project SPINE (Structural Proteomics IN Europe), contract number QL2-CT-2002-00988. We thank Stephen Cusack and the EMBL Grenoble outstation for hosting the group. We are grateful to Laurence Pearl and Renos Savva for the Ugi plasmid, to Lucy Freeman for corrections of the manuscript and to Didier Nurrisso from ESRF for beamtime and support.

References

- de Jesus, O., Smith, P. R., Spender, L. C., Elgueta Karstegl, C., Niller, H. H., Huang, D. & Farrell, P. J. (2003). Updated Epstein-Barr virus (EBV) DNA sequence and analysis of a promoter for the BART (CST, BARF0) RNAs of EBV. *J. Gen. Virol.* **84**, 1443–1450.
- Tarbouriech, N., Buisson, M., Géoui, T., Daenke, S., Cusack, S. & Burmeister, W. P. (2006). Structural genomics of the Epstein-Barr virus. *Acta Crystallog. sect. D*, **62**, 1276–1285.
- Henle, G., Henle, W., Clifford, P., Diehl, V., Kafuko, G. W., Kirya, B. G. *et al.* (1969). Antibodies to Epstein-Barr virus in Burkitt’s lymphoma and control groups. *J. Natl Cancer Inst.* **43**, 1147–1157.
- Ahmed, N. & Heslop, H. E. (2006). Viral lymphomagenesis. *Curr. Opin. Hematol.* **13**, 254–259.
- Gandhi, M. K. (2006). Epstein-Barr virus-associated lymphomas. *Expert Rev. Anti Infect. Ther.* **4**, 77–89.
- An, Q., Robins, P., Lindahl, T. & Barnes, D. E. (2005). C→T mutagenesis and gamma-radiation sensitivity due to deficiency in the Smug1 and Ung DNA glycosylases. *EMBO J.* **24**, 2205–2213.
- Lindahl, T. (1974). An N-glycosidase from *Escherichia coli* that releases free uracil from DNA containing deaminated cytosine residues. *Proc. Natl Acad. Sci. USA*, **71**, 3649–3653.
- Huffman, J. L., Sundheim, O. & Tainer, J. A. (2005). DNA base damage recognition and removal: new twists and grooves. *Mutat. Res.* **577**, 55–76.
- Priet, S., Gros, N., Navarro, J. M., Boretto, J., Canard, B., Querat, G. & Sire, J. (2005). HIV-1-associated uracil DNA glycosylase activity controls dUTP misincorporation in viral DNA and is essential to the HIV-1 life cycle. *Mol. Cell*, **17**, 479–490.
- Pyles, R. B. & Thompson, R. L. (1994). Evidence that the herpes simplex virus type 1 uracil DNA glycosylase is required for efficient viral replication and latency in the murine nervous system. *J. Virol.* **68**, 4963–4972.
- Prichard, M. N., Duke, G. M. & Mocarski, E. S. (1996). Human cytomegalovirus uracil DNA glycosylase is required for the normal temporal regulation of both DNA synthesis and viral replication. *J. Virol.* **70**, 3018–3025.
- Ellison, K. S., Peng, W. & McFadden, G. (1996). Mutations in active-site residues of the uracil-DNA glycosylase encoded by vaccinia virus are incompatible with virus viability. *J. Virol.* **70**, 7965–7973.
- Reddy, S. M., Williams, M. & Cohen, J. I. (1998). Expression of a uracil DNA glycosylase (UNG) inhibitor in mammalian cells: varicella-zoster virus can replicate in vitro in the absence of detectable UNG activity. *Virology*, **251**, 393–401.
- Chen, R., Wang, H. & Mansky, L. M. (2002). Roles of uracil-DNA glycosylase and dUTPase in virus replication. *J. Gen. Virol.* **83**, 2339–2345.
- Krokan, H. E., Standal, R. & Slupphaug, G. (1997). DNA glycosylases in the base excision repair of DNA. *Biochem. J.* **325**, 1–16.
- Savva, R., McAuley-Hecht, K., Brown, T. & Pearl, L. (1995). The structural basis of specific base-excision repair by uracil-DNA glycosylase. *Nature*, **373**, 487–493.
- Mol, C. D., Arvai, A. S., Slupphaug, G., Kavli, B., Alseth, I., Krokan, H. E. & Tainer, J. A. (1995). Crystal structure and mutational analysis of human uracil-DNA glycosylase: structural basis for specificity and catalysis. *Cell*, **80**, 478–869.
- Ravishankar, R., Bidya Sagar, M., Roy, S., Purnapatre, K., Handa, P., Varshney, U. & Vijayan, M. (1998). X-ray analysis of a complex of *Escherichia coli* uracil DNA glycosylase (EcUDG) with a proteinaceous inhibitor. The structure elucidation of a prokaryotic UDG. *Nucl. Acids Res.* **26**, 4880–4887.
- Leiros, I., Moe, E., Lanes, O., Smalas, A. O. & Willassen, N. P. (2003). The structure of uracil-DNA glycosylase from Atlantic cod (*Gadus morhua*) reveals cold-adaptation features. *Acta Crystallog. sect. D*, **59**, 1357–1365.

20. Leiros, I., Moe, E., Smalas, A. O. & McSweeney, S. (2005). Structure of the uracil-DNA N-glycosylase (UNG) from *Deinococcus radiodurans*. *Acta Crystallog. sect. D*, **61**, 1049–1056.
21. Barrett, T. E., Savva, R., Panayotou, G., Barlow, T., Brown, T., Jiricny, J. & Pearl, L. H. (1998). Crystal structure of a G:T/U mismatch-specific DNA glycosylase: mismatch recognition by complementary-strand interactions. *Cell*, **92**, 117–129.
22. Wibley, J. E., Waters, T. R., Haushalter, K., Verdine, G. L. & Pearl, L. H. (2003). Structure and specificity of the vertebrate anti-mutator uracil-DNA glycosylase SMUG1. *Mol. Cell*, **11**, 1647–1659.
23. Chung, J. H., Im, E. K., Park, H. Y., Kwon, J. H., Lee, S., Oh, J. *et al.* (2003). A novel uracil-DNA glycosylase family related to the helix-hairpin-helix DNA glycosylase superfamily. *Nucl. Acids Res.* **31**, 2045–2055.
24. Hōseki, J., Okamoto, A., Masui, R., Shibata, T., Inoue, Y., Yokoyama, S. & Kuramitsu, S. (2003). Crystal structure of a family 4 uracil-DNA glycosylase from *Thermus thermophilus* HB8. *J. Mol. Biol.* **333**, 515–526.
25. Prikh, S. S., Mol, C. D., Slupphaug, G., Bharati, S., Krokan, H. E. & Tainer, J. A. (1998). Base excision repair initiation revealed by crystal structures and binding kinetics of human uracil-DNA glycosylase with DNA. *EMBO J.* **17**, 5214–5226.
26. Mol, C. D., Arvai, A. S., Sanderson, R. J., Slupphaug, G., Kavli, B., Krokan, H. E. *et al.* (1995). Crystal structure of human uracil-DNA glycosylase in complex with a protein inhibitor: protein mimicry of DNA. *Cell*, **82**, 701–708.
27. Savva, R. & Pearl, L. H. (1995). Nucleotide mimicry in the crystal structure of the uracil-DNA glycosylase-uracil glycosylase inhibitor protein complex. *Nature Struct. Biol.* **2**, 752–757.
28. Putnam, C. D., Shroyer, M. J., Lundquist, A. J., Mol, C. D., Arvai, A. S., Mosbaugh, D. W. & Tainer, J. A. (1999). Protein mimicry of DNA from crystal structures of the uracil-DNA glycosylase inhibitor protein and its complex with *Escherichia coli* uracil-DNA glycosylase. *J. Mol. Biol.* **287**, 331–346.
29. Friedberg, E. C., Ganesan, A. K. & Minton, K. (1975). N-Glycosidase activity in extracts of *Bacillus subtilis* and its inhibition after infection with bacteriophage PBS2. *J. Virol.* **16**, 315–321.
30. Takahashi, I. & Marmur, J. (1963). Replacement of thymidylic acid by deoxyuridylic acid in the deoxyribonucleic acid of a transducing phage for *Bacillus subtilis*. *Nature*, **197**, 794–795.
31. Karran, P., Cone, R. & Friedberg, E. C. (1981). Specificity of the bacteriophage PBS2 induced inhibitor of uracil-DNA glycosylase. *Biochemistry*, **20**, 6092–6096.
32. Krusong, K., Carpenter, E. P., Bellamy, S. R., Savva, R. & Baldwin, G. S. (2006). A comparative study of uracil-DNA glycosylases from human and herpes simplex virus type 1. *J. Biol. Chem.* **281**, 4983–4992.
33. Saikrishnan, K., Bidya Sagar, M., Ravishankar, R., Roy, S., Purnapatre, K., Handa, P. *et al.* (2002). Domain closure and action of uracil DNA glycosylase (UDG): structures of new crystal forms containing the *Escherichia coli* enzyme and a comparative study of the known structures involving UDG. *Acta Crystallog. sect. D*, **58**, 1269–1276.
34. Bennett, S. E. & Mosbaugh, D. W. (1992). Characterization of the *Escherichia coli* uracil-DNA glycosylase-inhibitor protein complex. *J. Biol. Chem.* **267**, 22512–22521.
35. Bellamy, S. R. & Baldwin, G. S. (2001). A kinetic analysis of substrate recognition by uracil-DNA glycosylase from herpes simplex virus type 1. *Nucl. Acids Res.* **29**, 3857–3863.
36. Krokan, H. & Wittwer, C. U. (1981). Uracil DNA-glycosylase from HeLa cells: general properties, substrate specificity and effect of uracil analogs. *Nucl. Acids Res.* **9**, 2599–2613.
37. Otterlei, M., Haug, T., Nagelhus, T. A., Slupphaug, G., Lindmo, T. & Krokan, H. E. (1998). Nuclear and mitochondrial splice forms of human uracil-DNA glycosylase contain a complex nuclear localisation signal and a strong classical mitochondrial localisation signal, respectively. *Nucl. Acids Res.* **26**, 4611–4617.
38. Chen, C. Y., Mosbaugh, D. W. & Bennett, S. E. (2004). Mutational analysis of arginine 276 in the leucine-loop of human uracil-DNA glycosylase. *J. Biol. Chem.* **279**, 48177–48188.
39. Bennett, S. E., Schimerlik, M. I. & Mosbaugh, D. W. (1993). Kinetics of the uracil-DNA glycosylase/inhibitor protein association. Ung interaction with Ugi, nucleic acids, and uracil compounds. *J. Biol. Chem.* **268**, 26879–26885.
40. Wang, Z. G., Smith, D. G. & Mosbaugh, D. W. (1991). Overproduction and characterization of the uracil-DNA glycosylase inhibitor of bacteriophage PBS2. *Gene*, **99**, 31–37.
41. Marchler-Bauer, A., Anderson, J. B., Cherukuri, P. F., DeWeese-Scott, C., Geer, L. Y., Gwadz, M. *et al.* (2005). CDD: a Conserved Domain Database for protein classification. *Nucl. Acids Res.* **33**, D192–D196.
42. Savva, R. & Pearl, L. H. (1995). Cloning and expression of the uracil-DNA glycosylase inhibitor (UGI) from bacteriophage PBS-1 and crystallization of a uracil-DNA glycosylase-UGI complex. *Proteins: Struct. Funct. Genet.* **22**, 287–289.
43. Stivers, J. T. (1998). 2-Aminopurine fluorescence studies of base stacking interactions at abasic sites in DNA: metal-ion and base sequence effects. *Nucl. Acids Res.* **26**, 3837–3844.
44. Lineweaver, H. & Burk, D. (1934). The determination of enzyme dissociation constants. *J. Am. Chem. Soc.* **56**, 658–666.
45. Leslie, A. G. (2006). The integration of macromolecular diffraction data. *Acta Crystallog. sect. D*, **62**, 48–57.
46. Collaborative Computational Project Number 4. (1994). The CCP4 suite: programs for protein crystallography. *Acta Crystallog. sect. D*, **50**, 760–763.
47. Vagin, A. & Teplyakov, A. (2000). An approach to multi-copy search in molecular replacement. *Acta Crystallog. sect. D*, **56**, 1622–1624.
48. Emsley, P. & Cowtan, K. (2004). Coot: model-building tools for molecular graphics. *Acta Crystallog. sect. D*, **60**, 2126–2132.
49. Murshudov, G. N., Vagin, A. A. & Dodson, E. J. (1997). Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallog. sect. D*, **53**, 240–255.
50. Altschul, S. F., Madden, T. L., A., S.A., Zhang, J., Zhang, Z., Miller, W. & Lipman, D. J. (1997). Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucl. Acids Res.* **25**, 3389–3402.
51. Thompson, J. D., Higgins, D. G. & Gibson, T. J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucl. Acids Res.* **22**, 4673–4680.
52. Gouet, P., Courcelle, E., Stuart, D. I. & F., M. (1999).

- ESPrnt: analysis of multiple sequence alignments in PostScript. *Bioinformatics*, **15**, 305–308.
53. Wallace, A. C., Laskowski, R. A. & Thornton, J. M. (1995). LIGPLOT: a program to generate schematic diagrams of protein-ligand interactions. *Protein Eng.* **8**, 127–134.
54. Jones, S. & Thornton, J. M. (1996). Principles of protein-protein interactions. *Proc. Natl Acad. Sci. USA*, **93**, 13–20.
55. Baker, N. A., Sept, D., Joseph, S., Holst, M. J. & McCammon, J. A. (2001). Electrostatics of nanosystems: application to microtubules and the ribosome. *Proc. Natl Acad. Sci. USA*, **98**, 10037–10041.
56. DeLano, W. L. (2005). The case for open-source software in drug discovery. *Drug Discov. Today*, **10**, 213–217.
57. Kleywegt, G. J. & Jones, T. A. (1994). A super position. *ESF/CCP4 Newsl.* **31**, 9–14.

Edited by K. Morikawa

(Received 12 September 2006; received in revised form 31 October 2006; accepted 2 November 2006)

Chapitre 11

Résultats supplémentaires

11.1 Définition et obtention du domaine UNG Δ 24

Alignement de séquence

L'alignement de la séquence complète de l'UNG d'EBV avec les séquences des UNG provenant de la Protein Data Bank (PDB) montre que des extensions de la partie N-terminal de la séquence, par rapport aux domaines conservés de l'UNG, ne sont jamais présentes dans les structures (voir Figure 11.1 page 134). La signature de l'UNG, dont le code d'accèsion Prosite est PS00130 (Figure 11.2 page 135), ainsi que le motif UNG pfam03167 défini sur le site NCBI, permettent de définir le domaine UNG qui débute au niveau du premier tryptophane conservé (W34) et s'étend jusqu'à la fin de la séquence.

Protéolyse limitée et séquençage N-terminal

Les parties flexibles d'une protéine sont en général plus exposées vis à vis des enzymes protéolytiques. En 3h la trypsine permet d'isoler un domaine stable de l'UNG (Figure 11.3 page 135), le séquençage N-terminal nous permet de confirmer qu'une vingtaine de résidus du côté N-terminal sont désordonnés (Figure 11.3 page 135). Ce résultat était prévisible compte tenu de l'alignement entre la séquence de l'UNG d'EBV "sauvage" et des séquences d'UNG d'autres espèces présentes dans la PDB (Figure 11.1 page 134).

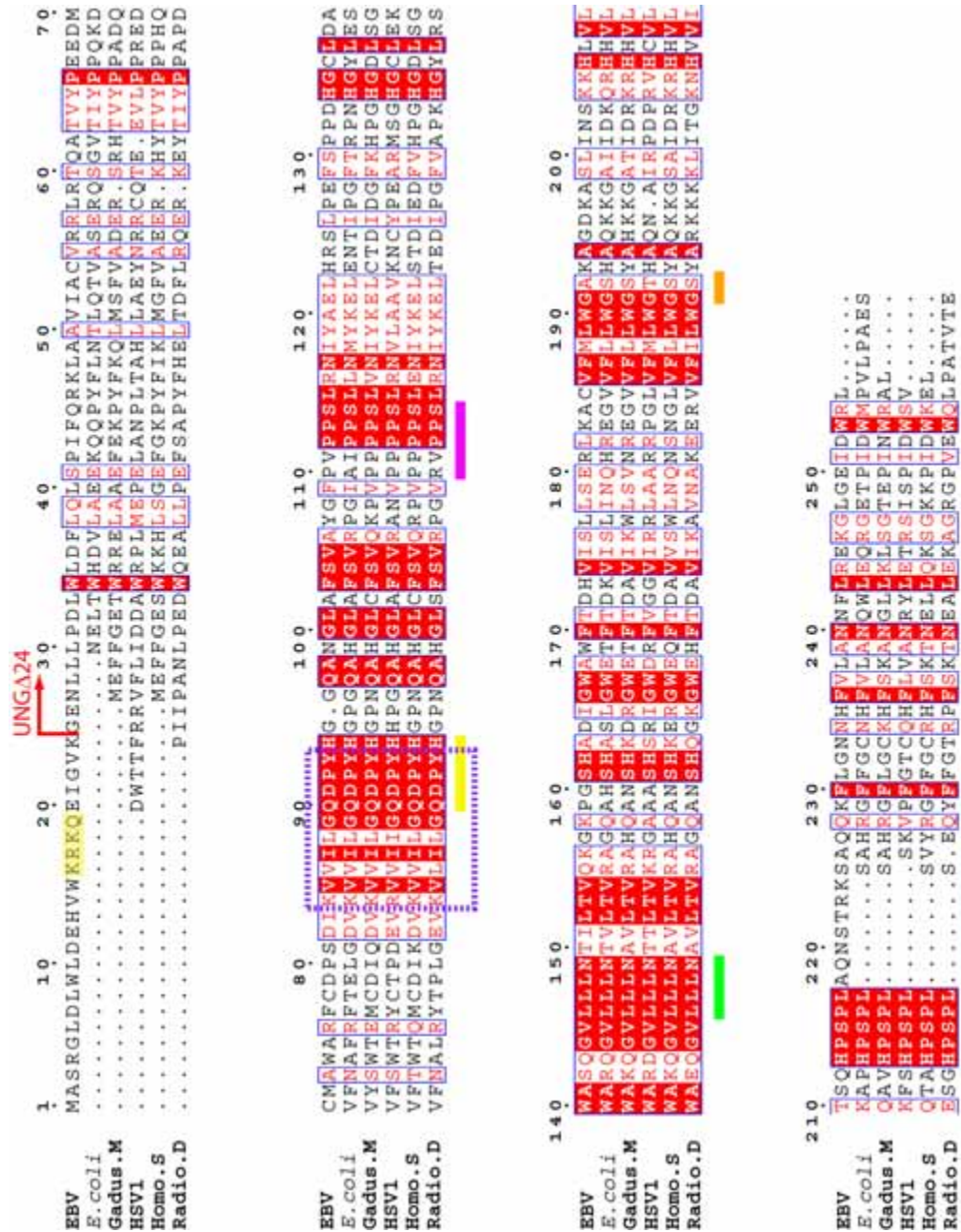


FIG. 11.1: Alignement de l'UNG d'EBV dans sa forme naturelle avec les UNG cristallisées d'autres espèces. La séquence "KRKQ" surlignée en jaune pourrait correspondre à un signal de localisation nucléaire (Otterlei *et al.*, 1998). La flèche indique le début de la construction UNG Δ 24. Les domaines responsables de la liaison à l'ADN et formant le site actif sont soulignés : la boucle d'activation de l'eau en jaune; la boucle 4-proline en rose; le brin- β 2 de reconnaissance de l'uracile en vert; la boucle GS en orange; la boucle leucine en bleu. Le cadre violet en pointillé indique la signature de l'UNG telle qu'elle est définie sur Prosite (Hulo *et al.*, 2006) par le code PS00130. L'étoile verte au dessus du résidu conservé (Trp34 chez EBV) indique le début du domaine UNG tel qu'il est défini par le code pfam03167 sur le site NCBI, ce domaine s'étend jusqu'à la fin de la séquence.

[KR] - [LIVA] - [LIVC] - [LIVM] - x - G - [QI] - D - P - Y

FIG. 11.2: Signature de l'UNG telle qu'elle est définie par Prosite (Hulo *et al.*, 2006).

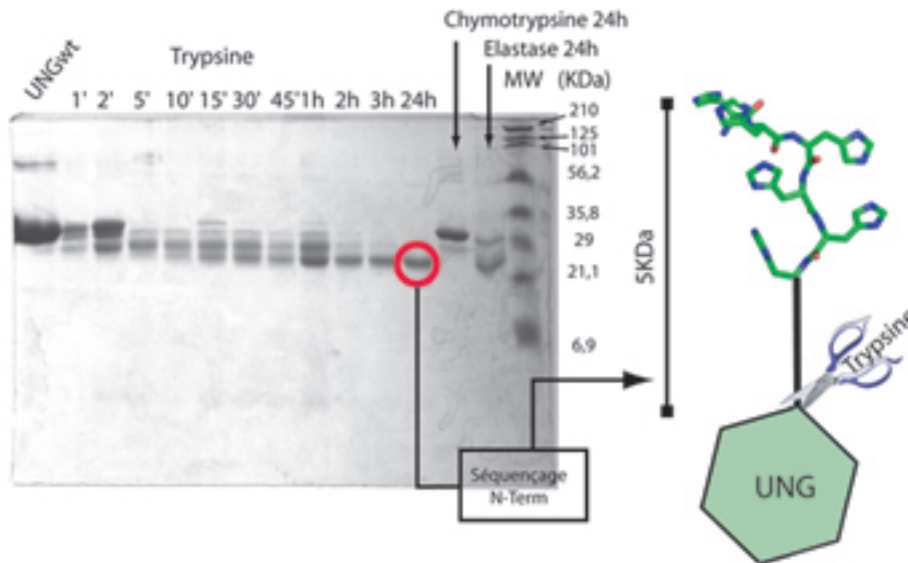


FIG. 11.3: La protéolyse limitée de l'UNGwt suivie d'un séquençage N-terminale permet de définir le domaine stable UNG Δ 24.

11.2 Vérification du repliement de la protéine par CD

Le spectre de dichroïsme circulaire nous permet de vérifier que l'UNG d'EBV est bien repliée dans nos conditions expérimentales. Nous avons pu évaluer la proportion de structure secondaire à 43% d'hélice- α et 12% de feuillet- β ce qui est en accord avec les résultats trouvés à partir de la structure cristallographique (Figure 11.4 page 136).

11.3 Influence du β -mercaptoéthanol et de l'Ugi sur l'homogénéité

La préparation d'un échantillon en vue de l'obtention de cristaux nécessite non seulement un échantillon pure, mais bien plus encore, un échantillon homogène. Si plusieurs états d'une même

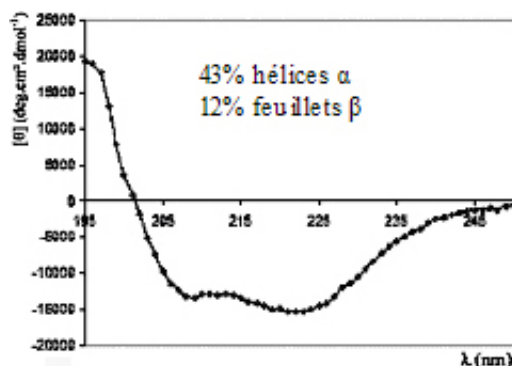


FIG. 11.4: Analyse du contenu en structure secondaire de l'UNG d'EBV par dichroïsme circulaire dans l'UV lointain.

protéine coexistent en solution, les chances d'obtenir des cristaux sont limitées. La diffusion dynamique de lumière nous a permis de tester différentes situations expérimentales et d'évaluer leur influence sur l'homogénéité de l'échantillon (Figure 11.5 page 137). Il ressort de cette étude que la présence de DTT¹ et la formation du complexe avec l'Ugi améliorent de façon substantielle l'homogénéité de la solution². En effet, l'UNG Δ 24 comporte 5 résidus cystéine. Les chaînes latérales de 2 d'entre eux sont orientées en direction de la surface de la protéine (Figure 11.6 page 138). Le DTT rend l'environnement réducteur et permet d'éviter la formation de pont disulfure entre des monomères voisins. La formation d'un empilement cristallin correct dépend également de la rigidité de la structure, une structure trop flexible n'adoptera pas un empilement régulier. Le complexe avec l'Ugi permet de la stabiliser (Figure 11.7 page 139) et favorise un empilement cristallin correct.

11.4 Information structurale supplémentaire

11.4.1 Diagramme de Ramachandran

Le diagramme de Ramachandran (Figure 11.8 page 140) nous montre que la plupart des résidus sont dans les régions les plus favorables (92%) ou assez favorables (7%). Un seul résidu, la Phe103, est situé dans une région non autorisée (soit 0.5% des résidus). La position de ce résidu est hautement

¹DTT : dithiothreitol.

²On observe une polydispersité réduite des formes individuelles et des agrégats.

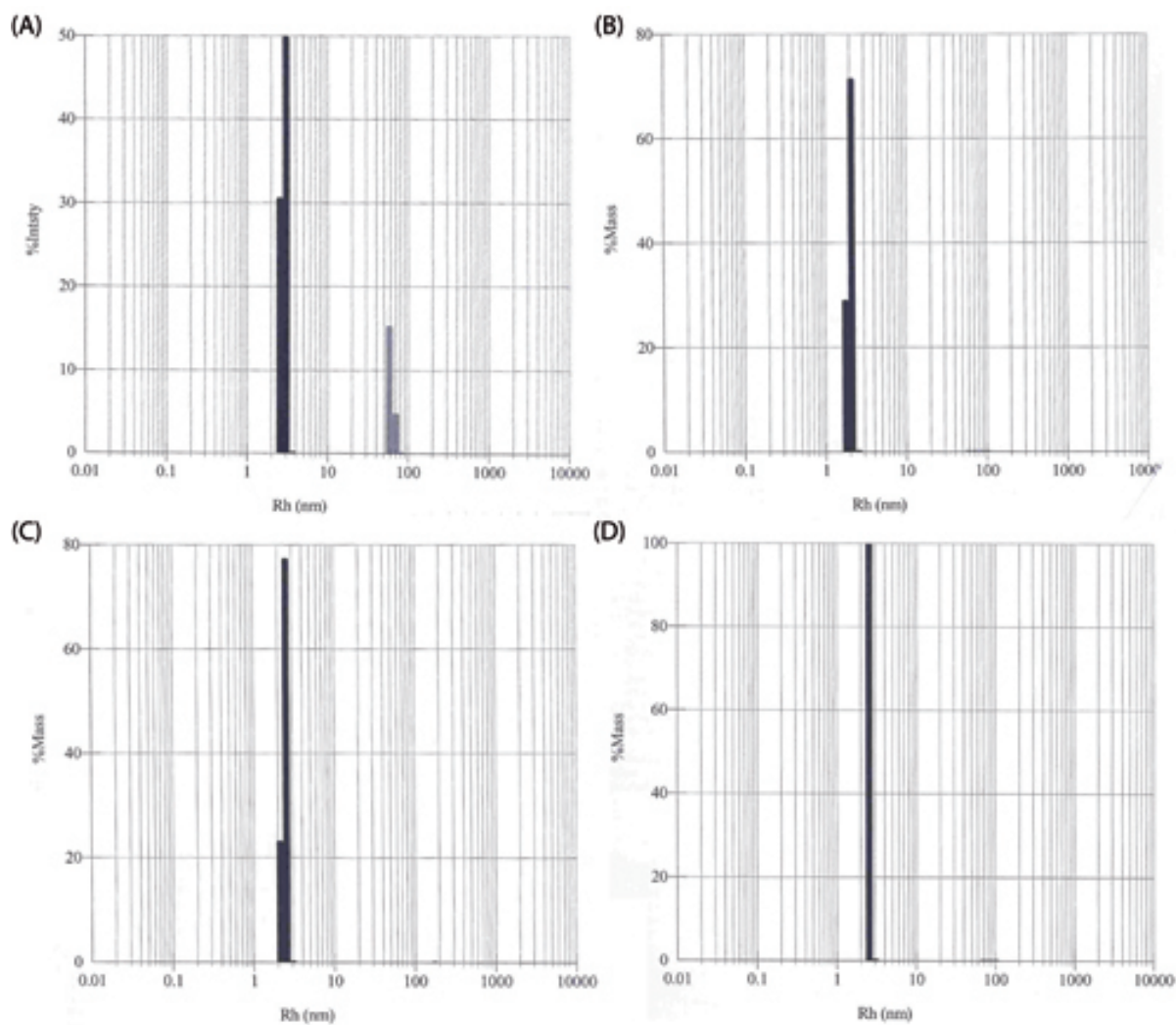


FIG. 11.5: Influence de la nature du tampon et de la présence d'Ugi sur l'homogénéité de l'échantillon. (A) UNG Δ 24 seule. (B) UNG Δ 24 + 10 mM DTT. (C) UNG Δ 24 + Ugi. (D) UNG Δ 24 + 10 mM DTT + Ugi.

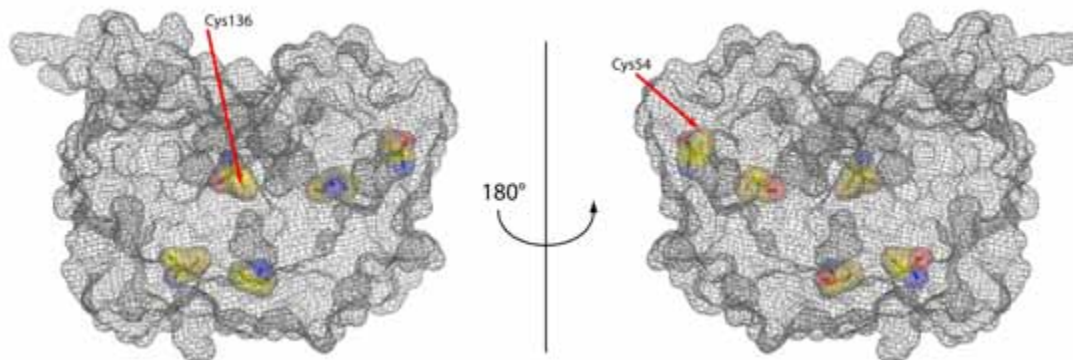


FIG. 11.6: L'UNG Δ 24 d'EBV comporte 5 résidus cystéine. La chaîne latérale des cystéines 54 et 136 est orientée vers le solvant.

conservé dans toutes les structures d'UNG déterminées jusqu'à maintenant. Il interagit avec l'uracile et le maintient dans une position favorable pour effectuer la catalyse de la rupture de la liaison N-glycosidique (Parikh *et al.*, 1998). Sa position atypique n'est donc pas le résultat d'une erreur lors de la construction du modèle, elle est liée à une nécessité fonctionnelle (Shaw *et al.*, 2004).

11.4.2 Empilement cristallin

L'arrangement du dimère AB est plus compacte que le dimère CD, la Figure 11.9 page 141 montre qu'il y a moins de solvant entre les dimères AB qu'entre les dimères CD. Cet empilement plus serré permet de limiter les mouvements des parties flexibles des protéines, ce qui conduit à une meilleure définition des différents domaines. Cette observation est confirmée par l'analyse du facteur de température globale des différentes chaînes qui est plus bas dans la chaîne A que dans la chaîne C (Figure 11.7, page 139).

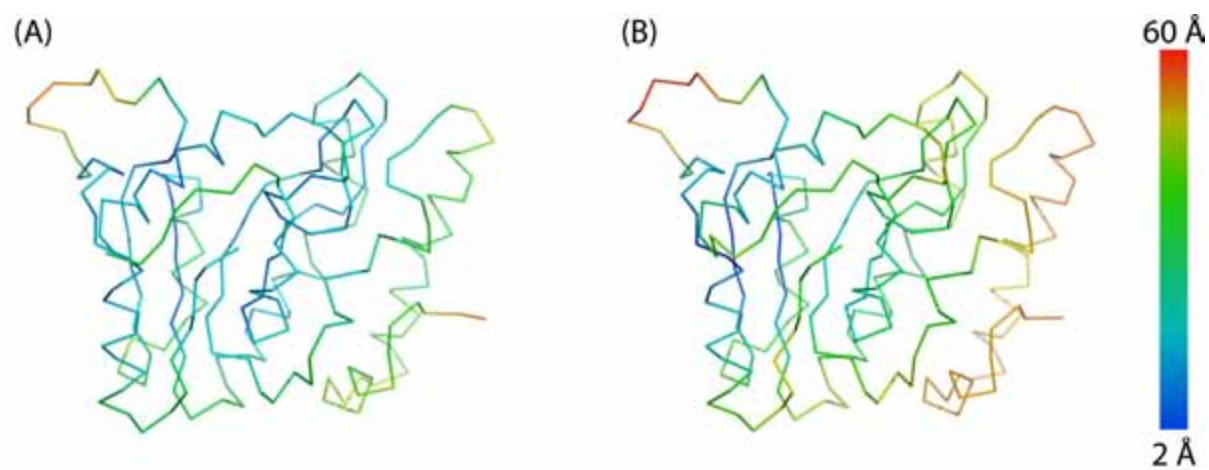


FIG. 11.7: Représentation “ribbon” de la structure de l’UNG d’EBV. Le facteur de température B de la boucle leucine de la chaîne A est de 25.7 \AA^2 contre 38.5 \AA^2 pour la chaîne C.

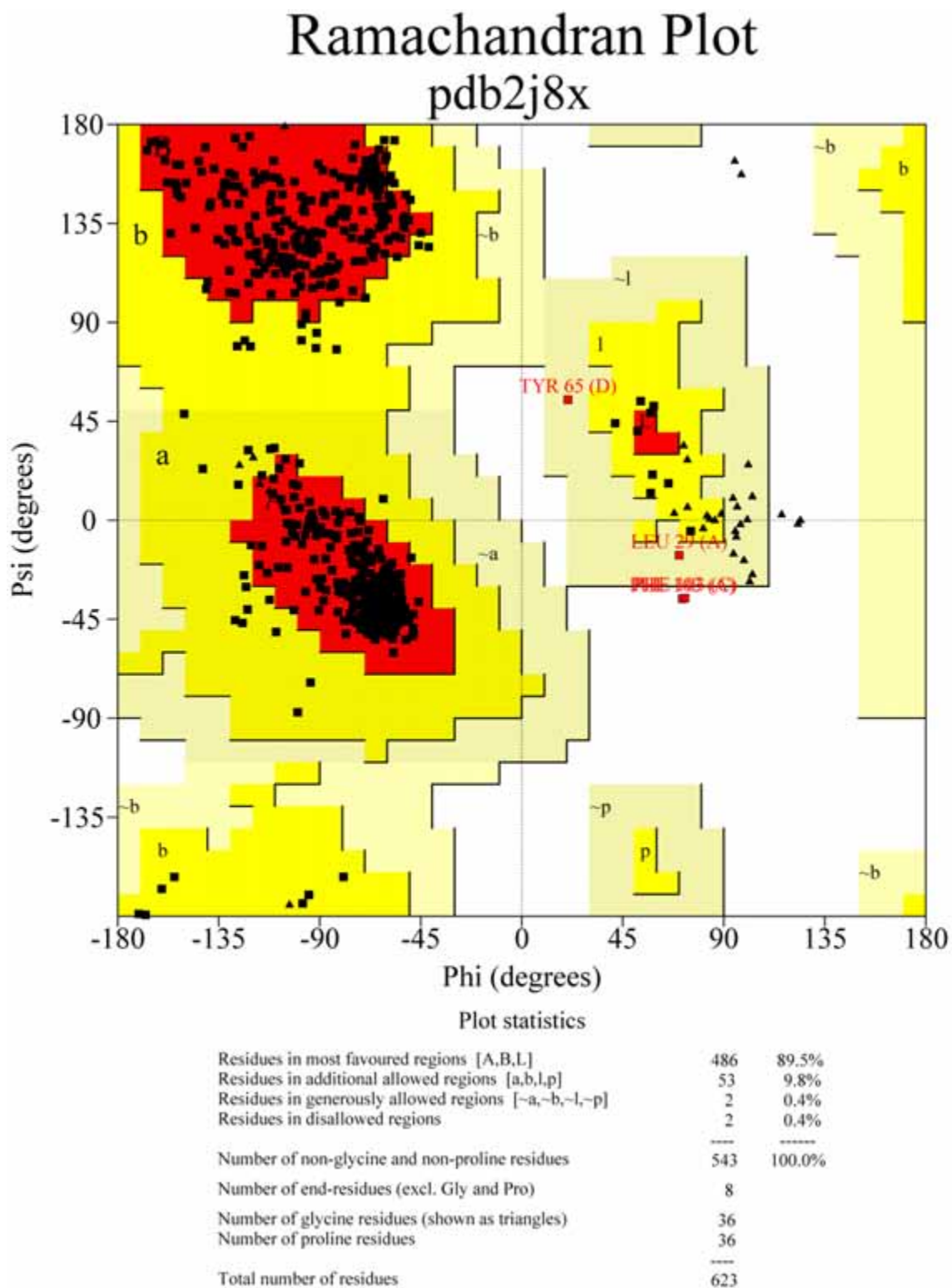


FIG. 11.8: Diagramme de Ramachandran du complexe Uracile-ADN Glycosylase-Ugi.

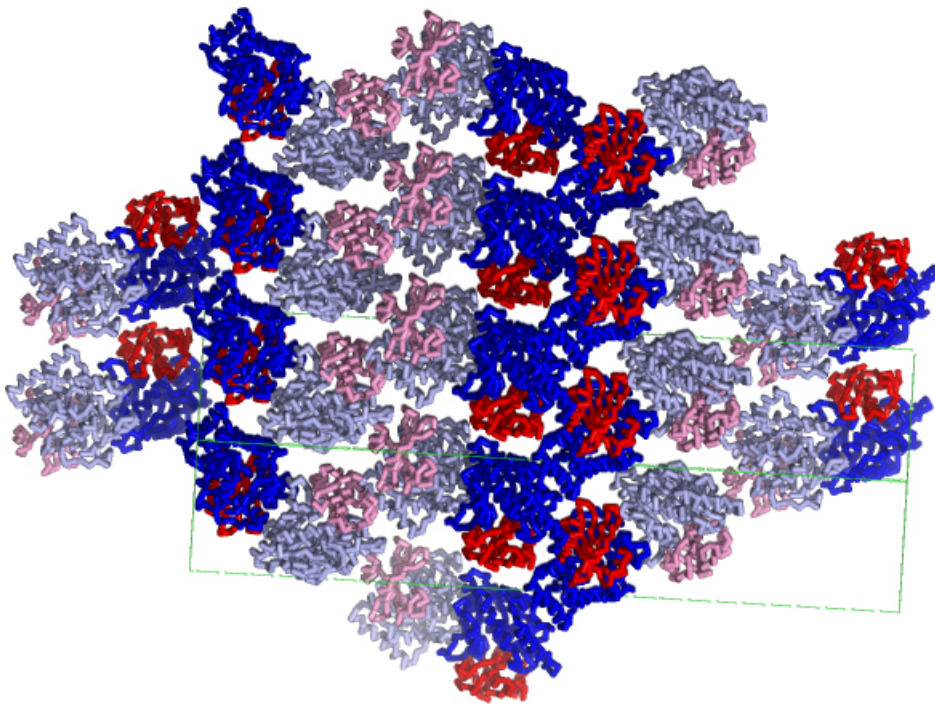


FIG. 11.9: Organisation du cristal du complexe UNG Δ 24-Ugi. La maille est représentée par une boîte verte. Le dimère AB (bleu-rouge) est soumis à un empilement plus compact que le dimère CD (bleu ciel-rose).

Quatrième partie

Etude structurale préliminaire de l'Alcaline Exonucléase (BGLF5) du virus d'Epstein-Barr

Chapitre 12

Introduction : BGLF5, un défi technique

DANS cette partie nous utiliserons l'exemple du travail sur l'Alcaline Exonucléase d'EBV pour illustrer les difficultés techniques majeures que l'on peut rencontrer au cours d'un projet de détermination de structures par cristallographie. Les évolutions techniques, notamment instrumentales, nous ont permis de porter cette étude à un niveau auquel nous n'aurions pas pu l'envisager il y a 4 ans. Ce travail a été initié et conduit en collaboration avec le Dr. Marlyse Buisson, chercheur au sein de l'équipe du Pr. Wim Burmeister.

12.1 L'Alcaline Exonucléase d'EBV

Rôle de l'Alcaline Exonucléase d'EBV

Le propos de cette partie n'est pas de décrire en détail les propriétés de l'Alcaline Exonucléase d'EBV¹ (comme nous avons pu le faire pour l'UNG), mais de l'utiliser davantage comme exemple d'un travail qui n'aurait pu être réalisé sans certain développement technologique. On connaît peu de chose sur l'Alcaline Exonucléase d'EBV, dont l'activité fut identifiée pour la première fois en culture de cellule par Cheng *et al* (Cheng *et al.*, 1980). Une DNase, comme son nom l'indique, est une enzyme qui à la capacité de couper l'ADN au niveau des liaisons phosphodiester entre 2 désoxyriboses, on parle alors d'endonucléase ; si elle coupe l'ADN à une extrémité 5' ou 3', on parle alors d'exonucléase

¹Son numéro d'accès Swissprot est le P03217.

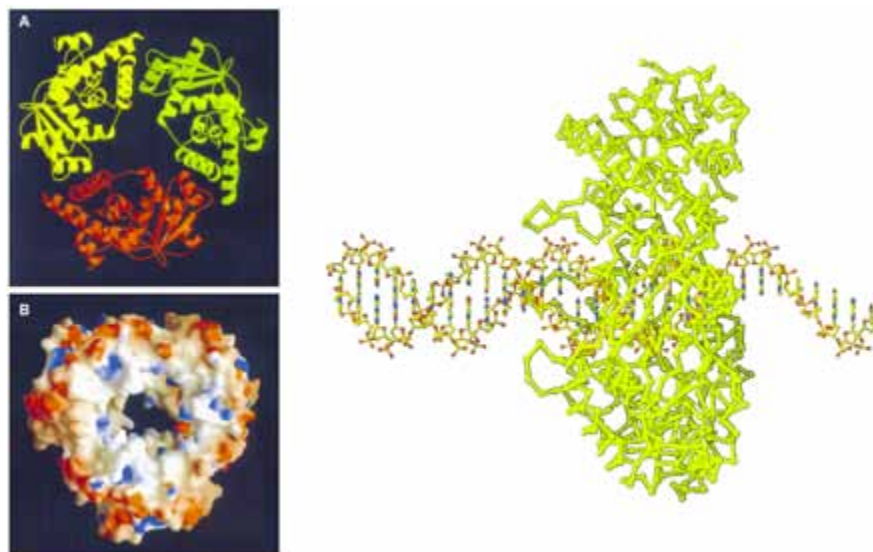


FIG. 12.1: Représentation de la structure tridimensionnelle de l'exonucléase du phage lambda (Kovall & Matthews, 1997)

(Figure 8.4, page 89). On ne sait pas quel est le rôle exact de l'Alcaline Exonucléase d'EBV, néanmoins elle est homologue à la protéine UL12 d'HSV-1 qui joue un rôle essentiel dans le cycle réplcatif de ce virus (Goldstein & Weller, 1998). Cette protéine jouerait également un rôle dans la recombinaison homologue de l'ADN virale (Reuven *et al.*, 2004). Sur la base de l'homologie de ces protéines et si l'on assume une fonction similaire, l'Alcaline Exonucléase d'EBV constitue une cible potentielle pour le développement d'antiviraux (Wathen, 2002).

Des données structurales limitées

Il n'existe que peu de données structurales concernant les Alcaline Exonucléases, seule la structure de la λ -exonucléase du bactériophage- λ^2 est connue (Kovall & Matthews, 1997). Cette protéine de 226 AA est trimérique, sa forme active est toroïdale. La cavité centrale permettrait l'accès à l'ADN (Figure 12.1). Avec 470 AA, la protéine d'EBV est 2 fois plus grande, ce qui suppose une organisation différente. L'alignement de séquences (Figure 12.2 page 147) à l'aide du programme ClustalW (Thompson *et al.*, 1994) ne permet pas de faire une comparaison pertinente des 2 protéines. Avec 10%

²Son numéro d'accès Swissprot est le P03697.

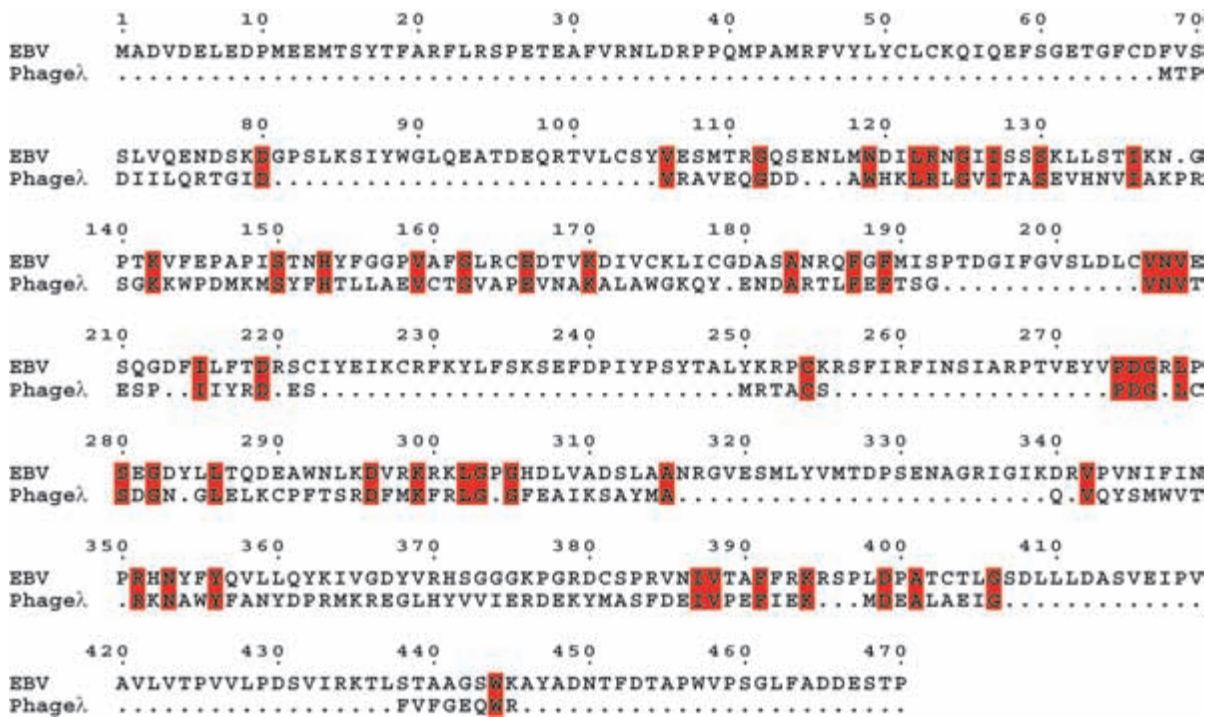


FIG. 12.2: L'alignement des séquences de l'exonucléase du phage lambda (Kovall & Matthews, 1997) et de l'Alcaline Exonucléase d'EBV ne permet pas d'identifier de domaines communs. Les résidus identiques sont surlignés en rouge.

de résidus identiques et 20% de résidus similaires, il est impossible, dans ce cas, d'identifier des zones similaires au moyen de la comparaison de séquences.

Néanmoins, la comparaison de la prédiction de structure secondaire³ de l'Alcaline Exonucléase d'EBV avec la structure secondaire de la protéine du phage- λ permet d'identifier des similarités entre la partie N-terminale de la protéine d'EBV et la λ -exonucléase du phage (Figure 12.3 page 148). La détermination de sa structure tridimensionnelle nous permettra de comprendre l'évolution et la complexification de cette protéine et, éventuellement, de proposer un modèle pour sa fonction.

12.2 Une étude structurale compliquée

Nous avons entrepris une étude structurale de l'Alcaline Exonucléase d'EBV. Le clonage, la purification et la production des premiers cristaux ont été conduits par le Dr. Marlyse Buisson. Les premiers cristaux obtenus étaient des aiguilles (Figure 12.4(a) page 149). En raison de leur forme et

³A l'aide du serveur NPS@ (Combet *et al.*, 2000).

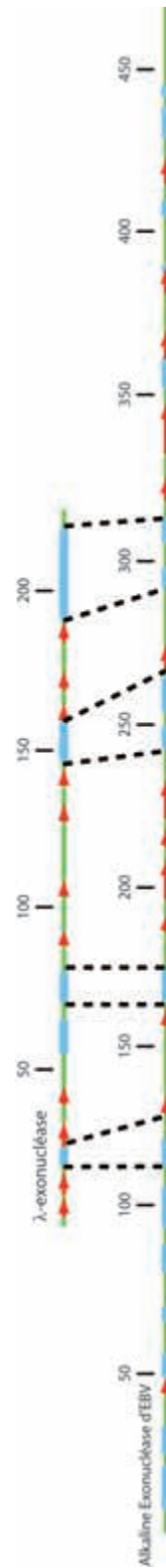
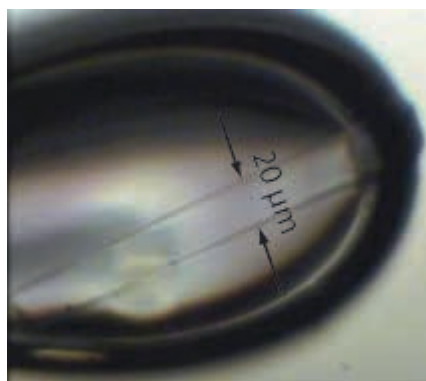


FIG. 12.3: La comparaison des structures secondaires de la λ -exonucléase et de l'Alcaline Exonucléase permet d'identifier des similarités du côté N-terminale de la protéine d'EBV. Les hélices- α sont en bleu, les feuillets- β sont en rouge, les coudes sont en vert.



(a) Les cristaux obtenus en diffusion de vapeur sont de fines aiguilles



(b) Après affinement des conditions, nous obtînmes de fines plaques.

FIG. 12.4: La diffusion de vapeur ne permet pas d'obtenir des cristaux de taille suffisante pour une bonne collecte de données.

de leur faible épaisseur, ce type de cristal ne permet pas d'analyse structurale. Après quelque temps, et en jouant sur les conditions de cristallisation, des cristaux en forme de plaque furent obtenus (Figure 12.4(b)). Ces cristaux sont assez gros pour être montés dans une boucle et exposés devant un faisceau de rayons-X, néanmoins, leur faible volume ne permet pas d'obtenir de cliché de diffraction à haute résolution, et surtout, ils se dégradent très rapidement. Ce projet pose 2 problèmes techniques majeurs :

1. Comment travailler avec des cristaux de très faible épaisseur⁴ ?
2. Comment obtenir de plus gros cristaux ?

La première question trouve sa réponse dans le développement de nouvelles lignes de lumière dites "micro-focus". Ces lignes permettent l'obtention d'un faisceau de rayons-X stable d'un diamètre de l'ordre de quelques μm . La deuxième question nécessite l'utilisation de méthodes alternatives de crystallogénèse.

⁴de l'ordre de quelques μm .

Chapitre 13

Matériel et Méthodes

13.1 Méthodes Alternatives de Cristallogénèse

13.1.1 Diffusion de vapeur

LORS du criblage des conditions de cristallogénèse, la diffusion de vapeur reste la méthode de choix. Cette méthode repose sur le principe suivant : la protéine d'intérêt à la concentration "X" est mélangée à une solution de précipitant à la concentration "Y", la goutte formée par ce mélange est disposée dans une enceinte scellée à proximité d'un réservoir contenant la solution de précipitant à la concentration "Y" (Figure 13.1 page 152).

Au moment où l'on mélange protéine et précipitant, leur concentration respective est divisée par 2¹. Au cours du temps, l'eau quitte le mélange protéine - précipitant, sous forme de vapeur, pour aller vers le réservoir, puis, la concentration en protéine et en précipitant tend vers la concentration initiale. L'augmentation respective de la concentration de la protéine et du précipitant entraîne une diminution de la solubilité de la protéine qui peut éventuellement se traduire par la formation de cristaux (Figure 13.2 page 153).

Il existe différentes façons pour changer la forme ou la taille des cristaux de protéine sur lesquels ont travaillé :

- En utilisant une construction de la protéine différente on peut faciliter l'empilement cristallin

¹Respectivement $\frac{X}{2}$ et $\frac{Y}{2}$.

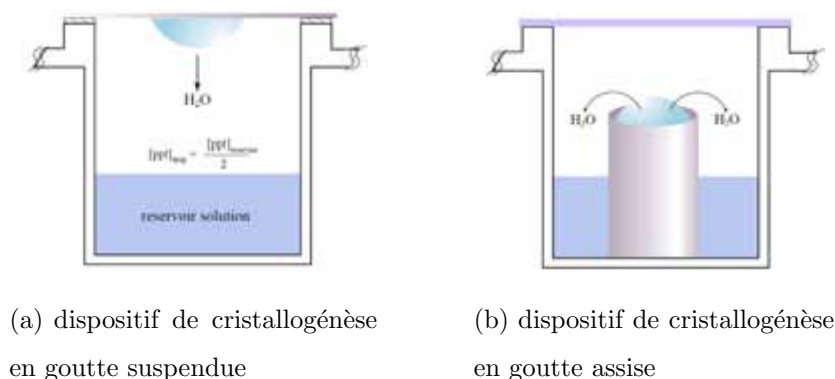


FIG. 13.1: La diffusion de vapeur est la technique de cristallogénèse la plus répandue, source Hampton-research

en supprimant des zones flexibles².

- Le changement des conditions de cristallogénèse ou l’ajout d’additif à la solution de cristallisation est la méthode généralement adoptée.
- Il est possible de changer les conditions physiques de la croissance cristalline, en jouant sur des paramètres comme la température ou les méthodes de diffusion.

Après avoir essayé un grand nombre de conditions de cristallisation et d’additifs et avant de tronquer l’enzyme, comme ce fut le cas pour l’UNG (voir page 99), nous avons décidé de jouer sur les conditions physiques d’obtention des cristaux en adoptant une méthode alternative de croissance cristalline.

13.1.2 Croissance cristalline en capillaire

Nous avons utilisé un nouveau procédé de cristallogénèse développé par l’équipe de Juan Ma. Garcia-Ruiz (Garcia-Ruiz *et al.*, 2002) et commercialisé par Hampton-Research et Trianatech sous le nom de **The Granada Crystallization Box** (Figure 13.3(b) page 154). L’utilisation de ce dispositif repose sur le principe de la **contre-diffusion** (Figure 13.3(a) page 154). La protéine est introduite dans un capillaire dont on bouche l’extrémité supérieure par de la graisse. Le capillaire est introduit dans une boîte dont le fond comprend une couche d’agarose³ d’environ 1.5 cm d’épaisseur. La boîte est ensuite remplie avec la solution de précipitant qui diffuse au travers de l’agarose et remonte à l’intérieur

²Cela suppose le clonage d’un nouveau fragment, de plus on perd l’information structurale correspondant aux zones tronquées.

³La concentration d’agarose est variable, dans le cadre de ce travail nous avons utilisé de l’agarose à 1% ou 2% dissout dans de l’eau. Ces différentes concentrations n’ont pas eu de répercutions au niveau des résultats.

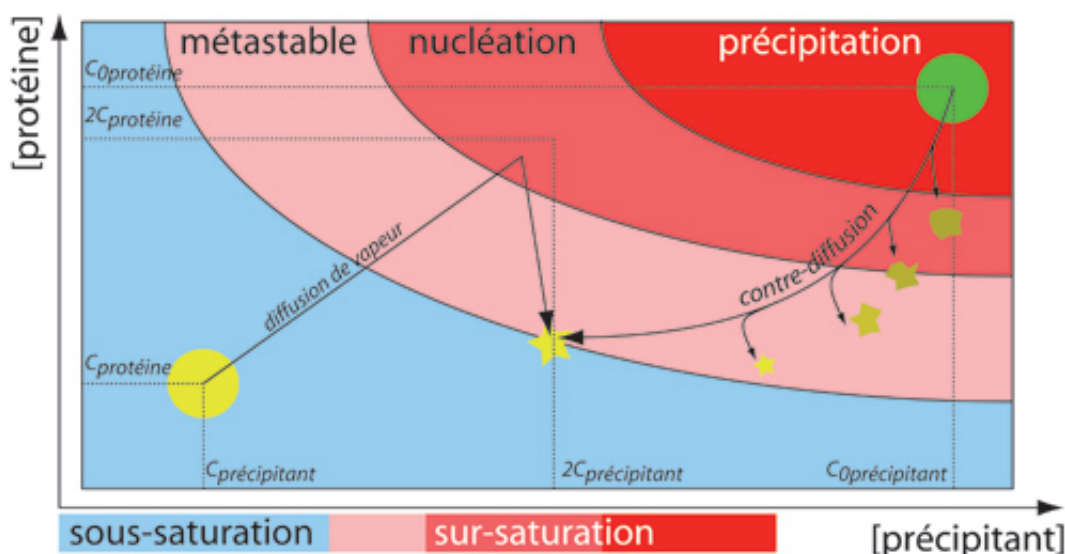


FIG. 13.2: Diagramme de phase. La contre diffusion permet d’explorer le diagramme de manière plus exhaustive (García-Ruiz *et al.*, 2002).

du capillaire (Figure 13.3(b) page 154). L’utilisation de cette méthode apporte de nombreux avantages par rapport à des techniques plus traditionnelles :

- La méthode de contre diffusion permet de “scanner” un grand nombre de conditions à l’intérieur d’un seul capillaire (différentes cinétiques, différents rapports protéine/précipitant, ...).
- L’utilisation de capillaire de petit volume permet de limiter la quantité de protéine utilisée pour chaque expérience à quelque μl .
- Le transport des cristaux est plus sûr qu’avec d’autres dispositifs expérimentaux (Eg. : Goutte suspendue).

13.2 Cristallographie en “Conditions Difficiles”

Le développement technique des infrastructures utilisées dans le domaine de la cristallographie apporte une solution nouvelle à des projets qui, il y a quelques années auraient été qualifiés de “problématiques”. Désormais l’utilisation de petit cristaux à faible durée de vie est possible. Nous allons voir quelles sont les évolution qui ont rendu ces projets possibles, et quels sont les problèmes qui persistent.

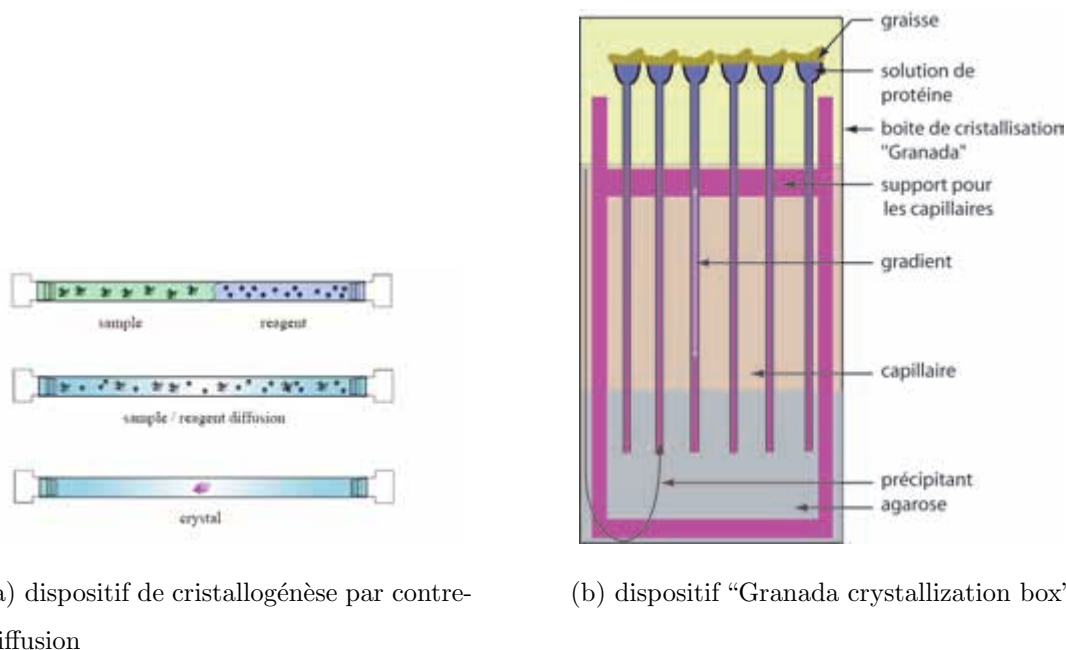


FIG. 13.3: La technique de contre-diffusion nous a permis d'augmenter la taille des cristaux par rapport à ceux obtenus en diffusion de vapeur.

13.2.1 La ligne de lumière micro-focus

Le travail sur des cristaux de petite taille pose de nombreux problèmes techniques, outre la faible durée de vie des cristaux⁴ les principaux défis restent :

- Obtenir un faisceau de rayons-X de faible dimension⁵.
- Le faisceau doit être stable dans le temps.
- La possibilité de passer un grand nombre d'échantillons en un minimum de temps.

C'est à ces caractéristiques que la ligne de lumière micro-focus ID23eh2 de l'ESRF⁶ tente de répondre.

Id23eh2 permet un alignement précis du faisceau avec le cristal

La ligne Id23eh2 comporte un certain nombre de dispositifs permettant un alignement très précis du faisceau de rayons-X avec le cristal. Avant toute expérience, il est possible de vérifier et d'ajuster la position exacte du faisceau de rayons-X à l'aide d'un cristal fluorescent (Figure 13.4 page 155) que l'on

⁴La faible durée de vie des cristaux est la conséquence d'un flux de rayons-X élevé.

⁵I.e. : de dimension plus faible que le cristal étudié.

⁶ESRF : European Synchrotron Radiation Facility.

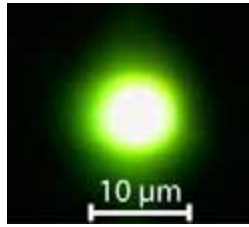
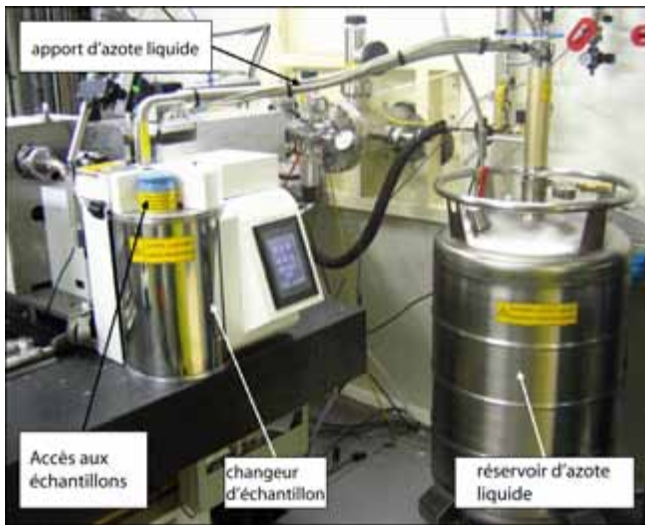
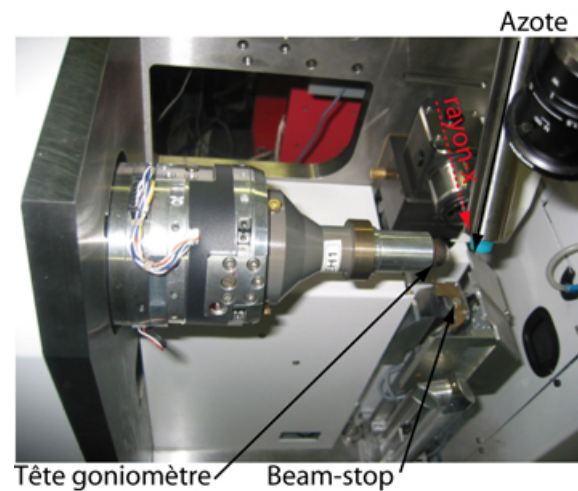


FIG. 13.4: Détection de la position et de la forme du faisceau de rayons-X à l'aide d'un cristal fluorescent.



(a) Vue d'ensemble de la ligne de lumière avec changeur d'échantillon



(b) Zoom sur la tête goniomètre où l'on monte le cristal

FIG. 13.5: La ligne de lumière Id23eh2 possède l'ensemble des équipements permettant le travail sur des cristaux “problématiques”.

peut aligner automatiquement alors même que la boucle comportant le cristal sur lequel l'expérience va être conduite est positionnée sur la tête goniométrique (Figure 13.5(b)). Une fois l'alignement réalisé, le cristal de protéine peut être introduit dans le faisceau de rayons-X. Une interface informatique permet de centrer le cristal grâce à système à 3 clics de souris, le cas échéant, une interface “expert” donne accès à un système encore plus précis (Figure 13.6 page 156). Ce très petit faisceau d'environ $10 \mu\text{m} \times 10 \mu\text{m}$ permet :

- De viser un endroit précis d'un cristal dans le cas d'un cristal de qualité hétérogène.
- De viser un cristal donné dans le cas où la boucle contient plusieurs cristaux.
- De se déplacer le long du cristal dans le cas où la zone irradiée se dégrade.

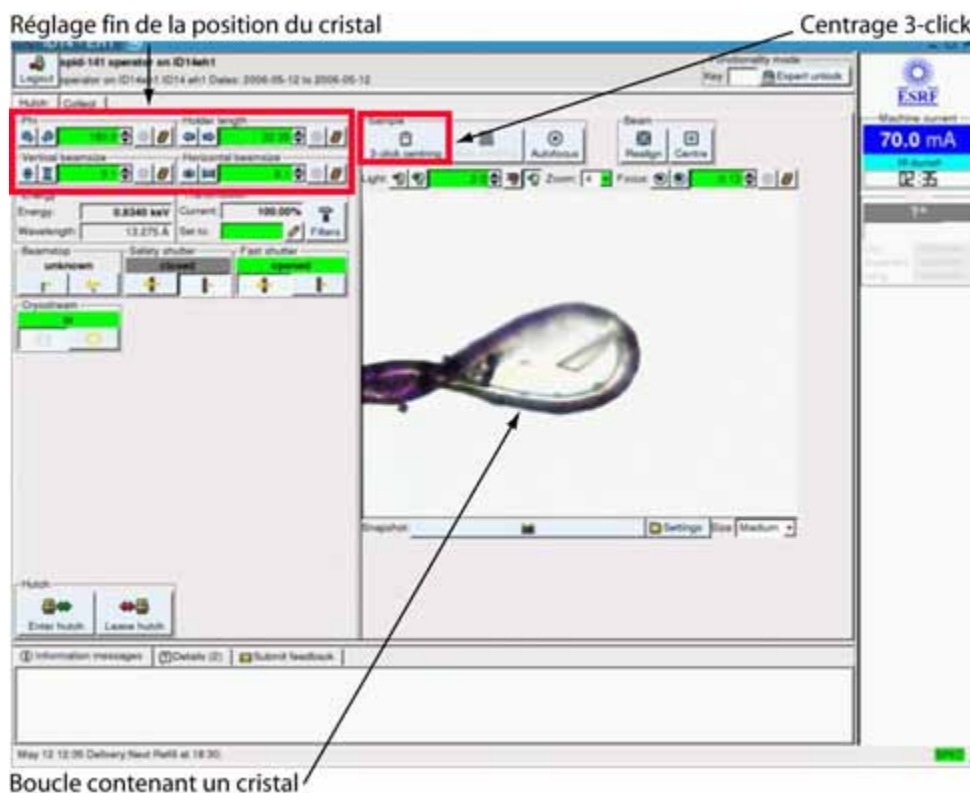


FIG. 13.6: Interface de centrage de l'échantillon MxCube

Le “changeur automatique d'échantillon”

La ligne Id23eh2 est équipée d'un système automatisé de changement d'échantillon : le “**sample changeur**” (Figure 13.5(a) page 155). Ce système permet le chargement de 5 paniers⁷ comportant chacun 10 fioles⁸ chargés de cristaux. Le tout peut être piloté de la salle de contrôle sans avoir besoin de se déplacer au niveau de la cabine expérimentale⁹. Outre le gain de temps résultant de l'absence de manipulation entre le montage et le démontage de chaque cristal, l'utilisation du changeur automatique d'échantillon évite les entrées et sorties répétées à l'intérieur de la cabine expérimentale, ce qui contribue à une plus grande stabilité du faisceau de rayons-X.

⁷Basket.

⁸Vial.

⁹experimental hutch.

13.2.2 Mélange de jeux de données

La collecte de données sur différentes parties d’un même cristal et, à fortiori, sur différents cristaux pose le problème du mélange des jeux de données pour obtenir un jeu complet. En effet, la collecte de données en se déplaçant le long d’un cristal suppose que ce cristal soit isomorphe sur toute sa longueur, ce qui n’est pas forcément le cas, de même, la collecte de données sur plusieurs cristaux suppose également qu’ils soient isomorphes. Avant de mélanger les données, la mise à l’échelle s’effectue avec le programme “Scala” (CCP4, 1994) qui permet de mesurer la statistique R_{merge} (Equation 13.1) correspondant à la qualité de la collecte. Cette statistique doit être la plus basse possible et rester si possible en dessous de 10%.

$$R_{merge} = \frac{\sum_{hkl} \sum_i |I_i(hkl) - \overline{I(hkl)}|}{\sum_{hkl} \sum_i I_i(hkl)} \quad (13.1)$$

Chapitre 14

Résultats et perspectives sur l'Alcaline Exonucléase

14.1 Propriétés des cristaux d'Alcaline Exonucléase

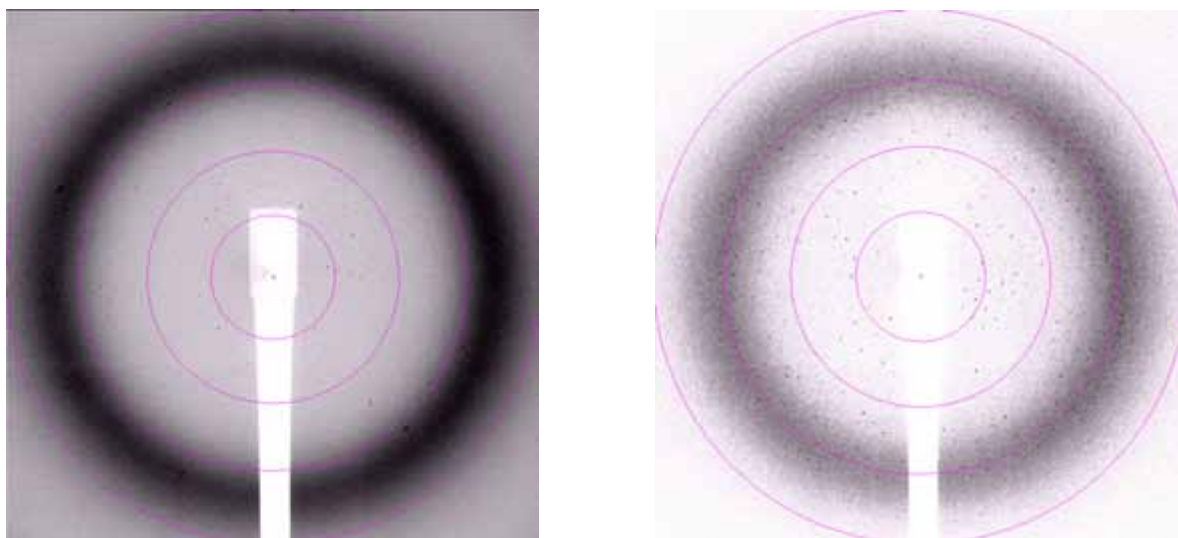
14.1.1 Les cristaux de l'Alcaline Exonucléase

Cristaux obtenus en diffusion de vapeur

LES premiers cristaux obtenus étaient des aiguilles sur lesquels aucune collecte n'était possible (Figure 12.4(a) page 149). L'affinement des conditions de cristallisation a permis l'obtention de plaques dans les conditions suivantes : 10 mM DTT 10 mM MgCl₂ 0.1 M Hépès pH 7,0, 8% PEG 400 (Figure 12.4(b) page 149). Ces cristaux étaient trop fins pour permettre une collecte de données. Ils permettaient la collecte d'une dizaine d'images à une résolution maximale de 10 - 15 Å (Figure 14.1(a) page 160). Ni le changement des conditions de cristallisation, ni l'ajout d'additifs, n'ont permis d'obtenir de plus gros cristaux.

Cristaux obtenus par la méthode de contre-diffusion

Afin de faire grossir les cristaux nous avons opté pour la méthode de contre-diffusion en utilisant les "Granada Crystallisation Box" de Hampton-Research (Figure 13.3(b) page 154). Ce système nous



(a) Un cristal d'Alcaline Exonucléase de petite taille ne permet pas de dépasser une dizaine d'Angström de résolution.

(b) Les cristaux obtenus en capillaire permettent d'obtenir des clichés d'une résolution supérieure à 3\AA .

FIG. 14.1: L'alignement du faisceau joue un rôle crucial pour une collecte de données complète. Les cercles violet indiquent la résolution, du centre vers l'extérieur : 11.2\AA , 5.6\AA , 3.7\AA , 2.8\AA .

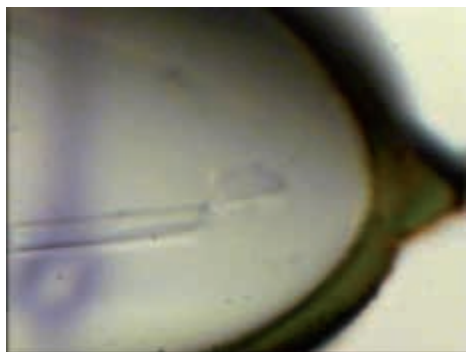
a permis d'obtenir des cristaux de plus grande taille (Figure 14.2(b) page 161), leur finesse reste problématique puisque ils se dégradent après un faible nombre d'images, on parle de "radiation damage"¹. Néanmoins, leur grande taille permet de se déplacer et de collecter à différents endroits (Figure 14.2(b) page 161) jusqu'à une résolution supérieure à 3\AA (Figure 14.1(b)).

14.2 Analyse structurale préliminaire

14.2.1 Microscopie électronique

La microscopie électronique nous permet de vérifier la qualité de la préparation de l'échantillon. L'échantillon est homogène et semble indiquer que le complexe adopte une organisation en anneau

¹Les rayons-X endommagent la protéine. Après un certain temps d'exposition, certaines liaisons sont hydrolysées (Eg. : pont dissulfure) et certains groupements sont endommagés (Eg. : décarboxylation de certains aspartates et de certaines glutamines, perte du groupement hydroxyle de certaines tyrosines ainsi que du groupement méthylthiol de certaines méthionines) (Burmeister, 2000).



(a) Cristal mal aligné avec le faisceau de rayon-X. La trace correspondant à l'impact des rayons-X dans le tampon est située sous le cristal



(b) Un cristal bien aligné avec le faisceau de rayon-X permet une collecte plus complète.

FIG. 14.2: L'alignement du faisceau joue un rôle crucial pour une collecte de données complète.

(Figure 14.3 page 163).

14.2.2 Jeu de donné natif

Nous avons obtenus un jeu de données natif dont les caractéristiques sont résumées dans le Tableau 14.1 page 165.

14.2.3 Self-Rotation Function

La fonction de self rotation nous permet de confirmer l'hypothèse formulée à partir des clichés de microscopie électronique, l'axe d'ordre 3 est aligné avec l'axe b du cristal. Ce résultat serait compatible avec une structure trimérique (Figure 14.4 page 178).

14.3 Conclusion et perspectives

Obtention des phases expérimentales

Le travail sur l'Alcaline Exonucléase n'en est qu'à ses début. En raison de la fragilité des cristaux, il fut difficile d'obtenir un jeu natif complet. Le phasage nécessite la production d'un jeu de données

d'une grande homogénéité. L'idéal étant de pouvoir tout collecter sur un même cristal afin de rester isomorphe. Ceci nécessite la production de cristaux de grande taille et l'utilisation de la ligne de lumière id23eh2.

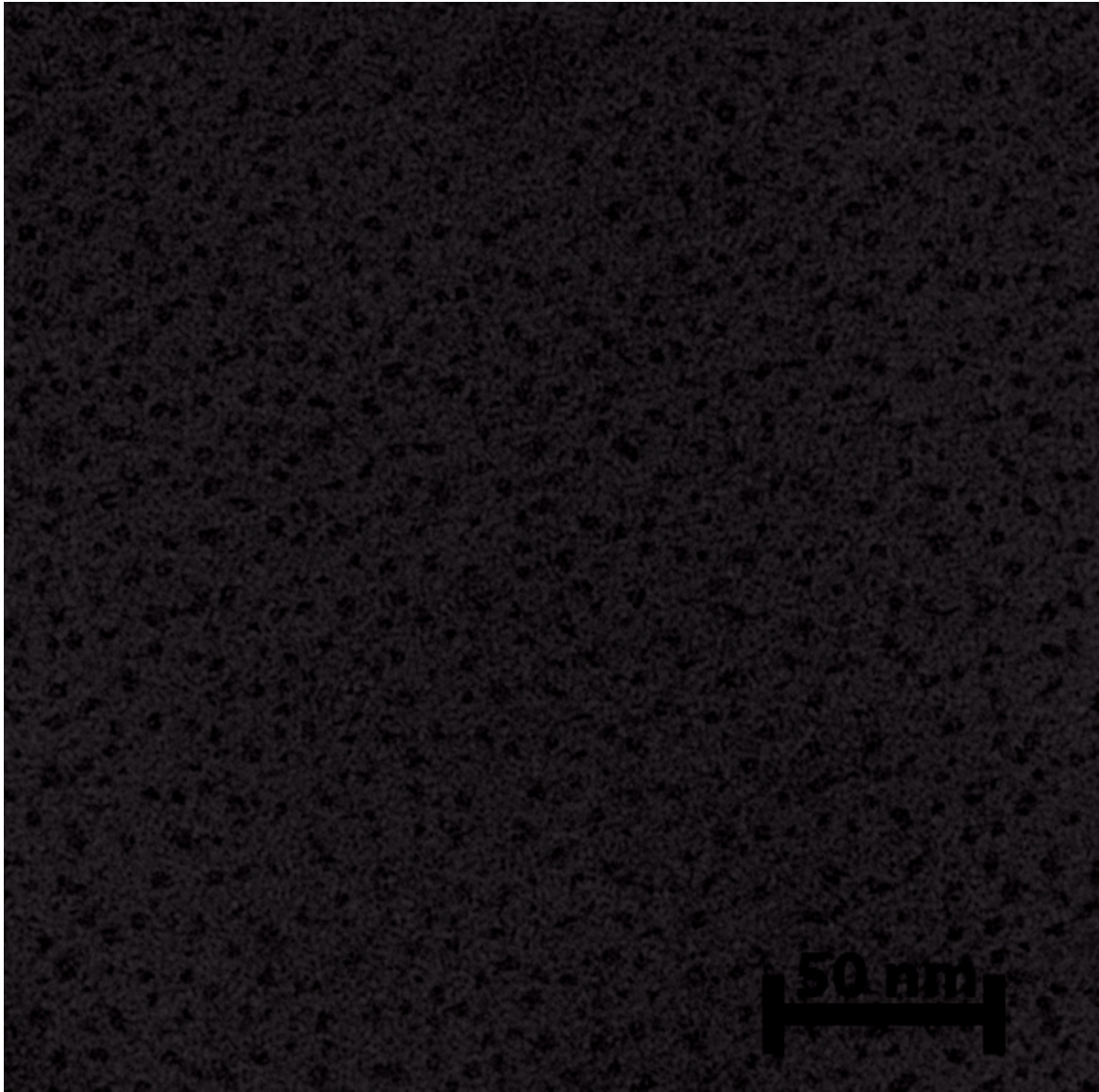


FIG. 14.3: La microscopie de l'Alcaline Exonucléase d'EBV suggère une forme pyramidale telle que celle observée pour l'exonucléase du phage lambda (source : Guy Schhoen).

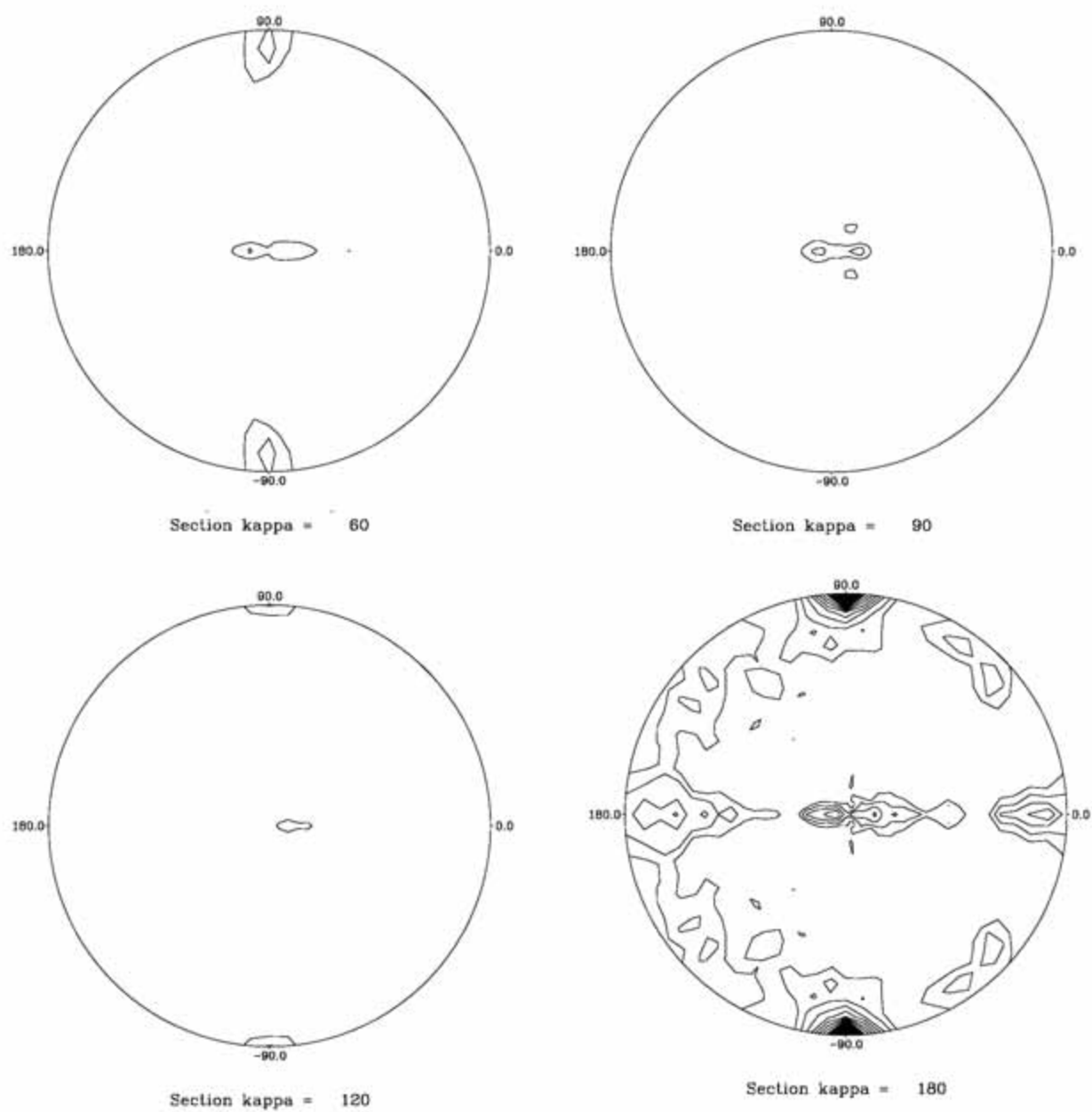


FIG. 14.4: La fonction de self rotation obtenue via le programme polarrfn, suggère une organisation dimérique. Le contour du diagramme est à partir de 2 sigma, tous les 1 sigma. Le seul pic significatif est pour $\kappa = 180^\circ$. La résolution va de 8 à 3.5 . Le rayon de la sphere d'integration est de 30 .

TAB. 14.1: Statistiques associées à la collecte de données natives de l'Alcaline Exonucléase

Jeu de données	
Groupe d'espace	$P2_1$
Maille	$a = 87.5 \text{ \AA}, b = 63.7 \text{ \AA}, c = 113.7 \text{ \AA}; \beta = 93.6^\circ$
Résolution	$3.2 \text{ \AA} - 47.5 \text{ \AA}$
Complétude	97.5 (98.7)
Multiplicité	2 (1.9)
R_{merge}	0.123 (0.402)
B-facteur du Wilson Plot	58.24 \AA^2

Cinquième partie

Conclusion et Perspectives

Chapitre 15

Conclusions et perspectives sur la contribution au projet de génomique structurale

15.1 Génomique structurale d'EBV

CETTE thèse est la conclusion de plus de 4 années de travail au sein de l'équipe "Génomique structurale d'EBV". Durant cette période, nous avons pu développer une méthodologie nous permettant de travailler sur plusieurs de cibles en parallèle. Dans cette dernière partie, je reviendrai brièvement sur les développements, les résultats et les perspectives engendrés par ce travail.

15.1.1 Conclusions sur le travail mené

Le travail mené par l'équipe de génomique structurale d'EBV a permis la résolution de 4 structures¹, d'autres projets sont en cours et devraient aboutir dans l'année à venir. Ce projet fut le premier à aborder systématiquement l'étude du virus d'Epstein-Barr d'un point de vue structurale. Nous avons également montré les limites de l'expression de protéines virales en système bactérien (*E. coli*) et l'intérêt d'utiliser l'expression en cellules d'insectes au moyen de baculovirus recombinants.

¹La protéase, la dUTPase, BARF1 et l'UNG.

Lors de la conférence de l'Association internationale EBV à Regensburg en septembre 2004, la présentation du projet de génomique structurale d'EBV était une première pour cette communauté qui n'était pas habituée à ce genre d'approche. Les résultats obtenus par notre équipe ainsi que la publication de la structure de la protéine ZEBRA par Carlo Petosa (Petosa *et al.*, 2006) ont suscité de plus en plus d'intérêt et les collaborations se multiplient. Chaque projet soulève de nouvelles questions qui pourront être abordées par d'autres équipes pour lesquelles les structures que nous avons publiées permettent d'aborder de vieux problèmes sous un jour nouveau. Dans notre équipe, Lucy Freeman a poursuivi le travail sur la dUTPase² par une caractérisation structurale et fonctionnelle de mutants et par des complexes avec des inhibiteurs. Le contexte scientifique de l'Institut de Virologie Moléculaire et Structurale (IVMS), au sein duquel ces recherches sont menées, permet d'envisager une approche *in-vivo* au travers de collaborations avec d'autres équipes, pour répondre aux questions trop souvent laissées de côté :

- A quel moment du cycle viral telle ou telle protéine est-elle exprimée ?
- Quelle est l'importance de telle ou telle protéine pour la propagation du virus ?
- Quelle est l'importance de telle ou telle protéine pour le maintien du virus sous sa forme latente ?
- Quels sont les partenaires cellulaires des protéines d'EBV ?

Bien sûr de nombreuses équipes se sont déjà penchées sur ces questions, mais notre approche, utilisant l'information tridimensionnelle, est nouvelle. De plus, les cibles auxquelles nous nous intéressons ne sont pas les cibles "historiques" ayant focalisé l'intérêt de la communauté des "EBVistes" durant près de 40 ans³. Cette proposition n'est certainement pas un jugement de valeur, elle doit être considérée comme un regard décalé et complémentaire d'une même problématique.

15.1.2 Evolution de la génomique structurale d'EBV

L'équipe fait partie du projet européen SPINE II qui fait suite à SPINE. Les thématiques sont désormais orientées vers la résolution de **complexes** de protéines. Ces complexes sont d'origines vi-

²Le travail sur la dUTPase a été initié par les Drs Marlyse Buisson et Nicolas Tarbouriech (Tarbouriech *et al.*, 2005).

³Entre autre et dans le désordre : LMP1, EBNA5 ...

rales⁴ ou entre une protéine virale et une protéine humaine⁵. A l'heure actuelle, il existe peu d'études décrivant de manière exhaustive "l'interactome" d'EBV. Cette étude est en cours, et les premiers travaux sont sur le point d'être publiés. Ce type de travail permettra d'apporter une information supplémentaire en décrivant de nouvelles interactions qui resteront à confirmer et à analyser par d'autres méthodes. A plus long terme, ce travail permettra d'identifier de nouvelles cibles thérapeutiques.

15.2 L'Uracile-ADN Glycosylase : conclusions et perspectives

Les de structures d'Uracile-ADN Glycosylase de 3 espèces étaient déjà connues avant que nous entreprenions ce travail : HSV1 (Savva *et al.*, 1995), Humaine (Mol *et al.*, 1995b), *E.coli* (Ravishankar *et al.*, 1998), et les structures de *Gadus Morhua* (Leiros *et al.*, 2003) et de *Deinococcus Radiodurans* (Leiros *et al.*, 2005) ont été publiées durant la thèse. Quelle est notre contribution à ce domaine qui semblait déjà bien pourvu avant que nous ne nous y intéressions ? Sans mentionner le fait qu'il s'agit de la structure d'une protéine d'EBV, ce qui, en tant que telle, est porteur d'intérêt en raison de la rareté des structures disponibles pour ce virus, c'est la première structure publiée qui possède une large insertion de 7 acides-aminés au niveau de la "boucle-leucine" un motif du site actif. L'intérêt de cette structure n'est pas anecdotique puisque cette insertion est une règle générale chez les γ -herpesvirus. En pratique le fonctionnement de l'enzyme ne se trouve pas altéré et les constantes cinétiques restent similaires à ce qui a déjà été observé (Bellamy & Baldwin (2001), Krusong *et al.* (2006)). Néanmoins la stratégie évolutive adoptée par l'enzyme démontre la flexibilité de ce système et le grand nombre de solutions adoptées par la nature pour répondre à une même question. Ce travail propose un modèle de fonctionnement qu'il faudra tester par la suite avec de nouvelles expériences.

⁴Le complexe hélicase-primase ...

⁵Le complexe BARF1-CSF1 est en cours, des de complexes entre la protéine ZEBRA et des partenaires cellulaires sont en projet.

15.2.1 Perspectives Fonctionnelles

Localisation cellulaire et cinétique d'expression

La structure de l'UNG d'EBV publiée est tronquée de ses 24 premiers acides-aminés, or la partie N-terminale de l'enzyme comporte une séquence "**KRKQ**" (Figure 11.1 page 134) homologue au signal de localisation nucléaire observé chez UNG2 l'enzyme nucléaire Humaine (Otterlei *et al.*, 1998). Nous avons fait produire des anticorps monoclonaux dirigés contre l'UNG d'EBV, et il serait intéressant de localiser la répartition cellulaire de cette enzyme en fonction de la délétion ou non de sa partie N-terminale. Dans le même ordre d'idée, l'utilisation de ces anticorps nous permettra de déterminer à quel moment du cycle de vie du virus l'enzyme est exprimée.

Mutagenèse

Le modèle d'interaction avec l'ADN nous a permis de mettre en évidence un certain nombre de résidus qui pourraient potentiellement interagir avec l'ADN et jouer un rôle dans le mécanisme catalytique de l'enzyme. Il convient désormais de tester ces hypothèses, un certain nombre de résidus pourront être mutés pour évaluer leurs rôles et déterminer si ils sont essentiels ou non. Tous ces résidus appartiennent au motif "**V**", c'est à dire à la boucle-leucine, ce sont : Gln212, Gln219, Asn220, Ser221, Thr222, Arg223, et la Gln227. La **Gln212** semble jouer un rôle particulièrement important pour structurer le réseau de liaisons hydrogènes de la boucle-leucine. Sa mutation contre un acide aminé aliphatique ou à chaîne plus courte, ne disposant pas des groupements intervenant dans la formation de liaisons hydrogènes permettrait d'évaluer le rôle de ce résidu sur la flexibilité de cette boucle. L'**Arg223** est situé au "sommet" de la boucle-leucine, les UNG d'autres espèces possèdent également des résidus à chaîne longue à une position équivalente. L'Arg276 de l'UNG Humaine semble jouer un rôle critique au niveau des interactions avec l'ADN et pour la discrimination du substrat simple et double brin (Chen *et al.*, 2005). La mutation de l'Arg223 d'EBV, en un résidu à chaîne plus courte, permettrait de mettre en évidence le rôle de ce résidu. Pour finir, la chaîne latérale du résidu **Asn220** joue un rôle dans la formation du complexe avec l'Ugi et pourrait également interagir avec le squelette phosphate de l'ADN. Que se passerait-il si ce résidu était muté contre un autre ne possédant pas le même type de chaîne latérale ?

Test enzymatique

Nous avons déterminé les constantes catalytiques de l'UNG d'EBV. Nous avons obtenu des K_m et K_{cat} de $1.5 \mu\text{M}$ et 3.9 s^{-1} pour l'UNGwt⁶ et $4.4 \mu\text{M}$ et 1.7 s^{-1} pour l'UNG Δ 24. Ces résultats sont du même ordre que ceux obtenus chez HSV1 ($K_m = 1.6 \mu\text{M}$, $K_{cat} = 2.1 \text{ s}^{-1}$, Bellamy & Baldwin (2001)), mais l'enzyme virale est légèrement moins active que son équivalent humaine ($K_m = 0.12 \mu\text{M}$, $K_{cat} = 7.4 \text{ s}^{-1}$, Krusong *et al.* (2006)). Le test enzymatique employé a été développé par Stuart Bellamy et ses collaborateurs (Bellamy & Baldwin, 2001). La mutagenèse des résidus mise en évidence dans le modèle entre l'UNG d'EBV et l'ADN et la détermination des constantes catalytiques des mutants nous permettront de d'évaluer le rôle de ces résidus, particulièrement ceux de la "boucle leucine", au niveau de laquelle plusieurs résidus seraient susceptibles de jouer un rôle important dans le mécanisme catalytique (Eg. : Arg223, Gln212, Asn220).

Inhibition par l'Ugi

La mise en perspective de notre travail, concernant la détermination des constantes d'inhibition de l'UNG par l'Ugi, est difficile en raison du manque de données permettant la comparaison. Samuel Bennet et ses collaborateurs décrivent une interaction irréversible entre l'UNG d'*E.coli* et l'Ugi (Bennett *et al.*, 1993). Clifford Mol et ses collaborateurs décrivent un IC_{50} de 12 à 36 pM⁷ (Mol *et al.*, 1995a). Zhigang Wang et ses collaborateurs, ont déterminé un K_i de 140 nM pour l'UNG d'*E.coli* (Wang *et al.*, 1991). Il faudra attendre que d'autres équipes répètent ces expériences pour disposer d'assez de données pour tirer des conclusions sur les spécificités de l'enzyme d'EBV par rapport aux autres espèces.

15.2.2 Perspectives Structurales

Un grand nombre d'hypothèses pourront être vérifiées grâce au test enzymatique : des variations des constantes cinétiques indiquant l'importance de tel ou tel résidu ; néanmoins la cristallisation et la résolution de la structure des formes mutantes nous apporteraient davantage d'informations.

⁶wt = wilde type ou naturelle.

⁷respectivement pour l'UNG Humaine et pour l'UNG d'*E.coli*.

Complexes avec de l'ADN

Les hypothèses formulées quant à la fonction de certains résidus reposent, dans la plupart des cas, sur l'analyse d'un modèle entre l'UNG d'EBV et un fragment d'ADN provenant d'une structure d'UNG Humaine (Parikh *et al.*, 1998). Seule la cristallisation de l'UNG d'EBV avec un oligomère d'ADN équivalent nous permettra de tester la véracité de ces hypothèses. Au début de ce projet, nous avons essayé de cristalliser l'UNG avec un fragment d'ADN (voir Parikh *et al.* (1998)), nous n'avons jamais obtenu de cristaux. Des tests de retard sur gel n'ont pas permis de montrer si l'enzyme liait l'oligomère d'ADN. Il faudrait tester différentes séquences d'ADN et évaluer l'influence de certains paramètres tels que la taille de l'oligomère et la nature de la base qui fait face à l'uracile.

Interaction avec d'autres protéines

De quelle façon la cascade de réactions, suivant l'excision de l'uracile, se poursuit-elle? Cette question est essentielle. L'UNG est certainement amené à interagir avec d'autres protéines, qu'elles soient cellulaires ou virales. Le travail publié récemment par Daikoku *et al.* (Daikoku *et al.*, 2006) montre qu'un grand nombre de protéines sont recrutées au niveau des sites de réplication de l'ADN d'EBV. La disponibilité d'anticorps monoclonaux dirigés contre l'UNG d'EBV nous permettra d'effectuer des expériences de co-immunoprécipitation et d'identifier des partenaires potentielles. Le clonage de ces protéines et leur production nous permettront de former des complexes et, éventuellement, d'obtenir les bases structurales de l'interaction entre l'UNG et d'autres protéines avec lesquelles elle interagit.

15.3 Résolution de la structure de l'Alcaline Exonucléase

La résolution de la structure de l'Alcaline Exonucléase est un défi en raison de la finesse des cristaux. Très peu de données sont disponibles concernant la structure de ce type d'enzyme, excepté la structure de l'enzyme du phage lambda (Kovall & Matthews, 1997). Le projet est déjà bien avancé, avec à notre disposition un jeu natif à environ 3 Å. Il nous reste à collecter des données sur des dérivés lourds ou sur une protéine contenant de la sélénométhionine pour déterminer les phases expérimentales, et construire le modèle. La structure de la protéine seule nous permettra de comprendre l'évolution et la complexification entre le phage- λ et les herpesvirus. Le travail n'en sera qu'à ses débuts, en effet, dans

l'article publié par Kovall *et al* en 1997 (Kovall & Matthews, 1997), un modèle de liaison à l'ADN est proposé, ce modèle n'a pas été confirmé expérimentalement : la structure d'un complexe avec l'ADN serait la bienvenue. Cela pose un problème majeur car l'Alcaline Exonucléase est censé couper l'ADN, le substrat utilisé devra être modifié pour que l'enzyme puisse lier l'ADN mais pas l'hydrolyser.

15.4 Conclusion sur l'étude structurale des protéines d'EBV

L'accumulation des données structurales sur les protéines d'EBV et, plus tard, sur les complexes entre les protéines d'EBV et d'autre protéines⁸ nous fournira une base de travail inestimable pour le développement de nouveau antiviraux. Coen et Schaffer estiment le nombre de cibles potentielles, chez les herpesvirus, à plus de 50 (Coen & Schaffer, 2003). Toutes les protéines impliquées dans le cycle virale sont concernées et il serait dommage de ne s'intéresser qu'aux cibles "historiques" telles que la polymérase. Les limites de l'utilisation des inhibiteurs nucléosidiques sont démontrées par l'émergence de résistances (Kleymann, 2003). De nouvelles stratégies sont nécessaires. L'inhibition de l'**interaction** de certaines protéines essentielles au cycle de vie du virus pourrait être une nouvelle voie dans la recherche de thérapies alternatives.

⁸humaine par exemple.

Sixième partie

Annexes

Chapitre 16

Informations complémentaires

16.0.1 Homologie / identité / similarité

IL ne faut pas confondre homologie, identité et similarité. Le taux d'identité de 2 séquences se mesure par le nombre d'acides aminés identiques déterminé après un alignement. Les protéines sont homologues si, et seulement si, elles dérivent d'un ancêtre commun et qu'elles ont divergé à partir de ce dernier. Si 2 séquences partagent plus de 30% d'identité, il y a de très forte chance pour qu'elles soient homologues. En revanche, il se peut que des protéines possédant une très faible identité (< 15%) soient homologues, c'est le cas de protéines ayant divergé très rapidement. Dans cette situation, l'analyse des séquences primaires seules, ne permet pas de démontrer l'homologie. Il faut avoir recours à l'analyse des structures secondaires ou tertiaires si celles-ci sont disponibles, si elles ne le sont pas, il est possible de les prédire avec une bonne précision. La structure des protéines est mieux conservée que leur séquence du fait d'une plus grande pression de sélection sur la structure. On parlera de similarité dans le cas où un acide aminé est substitué à un autre au sein d'un alignement mais, que ce dernier, présente des propriétés similaires (ie. Acide aminé polaire, ou au contraire aliphatique, fonction OH, ...)

16.0.2 Base de données généraliste

Le site Expasy (www.expasy.ch) héberge la base de données Swissprot (Bairoch *et al.*, 2004) : <http://www.expasy.org/sprot/>

16.0.3 Base de données de protéines virale

La base de données VIDA (pour Virus DAtabase) de l'*University College of London* contient une collection d'informations concernant des familles de protéines homologues dérivées des cadres de lecture ouverts de génomes viraux. On peut accéder à la base de données à l'adresse suivante : http://www.biochem.ucl.ac.uk/bsm/virus_database/VIDA3/VIDA.html

16.0.4 Base de données de protéines impliquées dans la réparation de l'ADN

Le laboratoire "*Cancer Research UK*" maintient une base de données des protéines, codées par le génome Humain, qui interviennent dans les mécanismes de réparation du génome. La base de données peut être trouvée à l'adresse suivante : http://www.cgal.icnet.uk/DNA_Repair_Genes.html

16.0.5 Prédiction de structure secondaire

Si la séquence primaire de deux protéines de même fonction peut différer grandement, la structure secondaire est susceptible d'être très similaire. La prédiction de structure secondaire des UNG alignées permet d'obtenir des informations complémentaires sur les différents domaines de cette protéine. Il existe un grand nombre d'algorithmes de prédiction de structure secondaire faisant appel à différents principes, dans cette études, nous avons choisi de combiner des algorithmes récemment développés pour augmenter la qualité de prédiction (supérieure à 70%). GORIV (Garnier, Osguthorpe, Robson quatrième version) recherche quelle influence un acide aminé peut avoir sur la conformation de tous les acides aminés de la séquence en utilisant l'ensemble des protéines de structures 3D connues. Cet algorithme s'appuie sur la théorie de l'information qui consiste à mesurer des paramètres statistiques sur un échantillon de petite taille et de les extrapoler sur un échantillon de plus grande taille sans perte d'informations. Le logiciel SOPM (Self-Optimized Prediction Method) utilise le principe selon lequel des peptides similaires ont tendance à adopter un repliement similaire (méthode des plus proches voisins), la protéine est comparée à une base de données de peptides homologues issus de la PDB, chaque segment de la protéine adopte la conformation du peptide similaire de la base de données. PhD utilise deux réseaux de neurones successifs et des alignements multiples de séquences en entrée, un

premier réseau permet de passer de la séquence à la structure, un deuxième réseau lisse les informations de prédiction de structures, un système décisionnel, à base de vote majoritaire, définit la prédiction finale. PREDATOR est une méthode statistique qui repose sur l'occurrence d'un acide aminé donné dans une structure donnée.

16.0.6 Symétrie cristalline

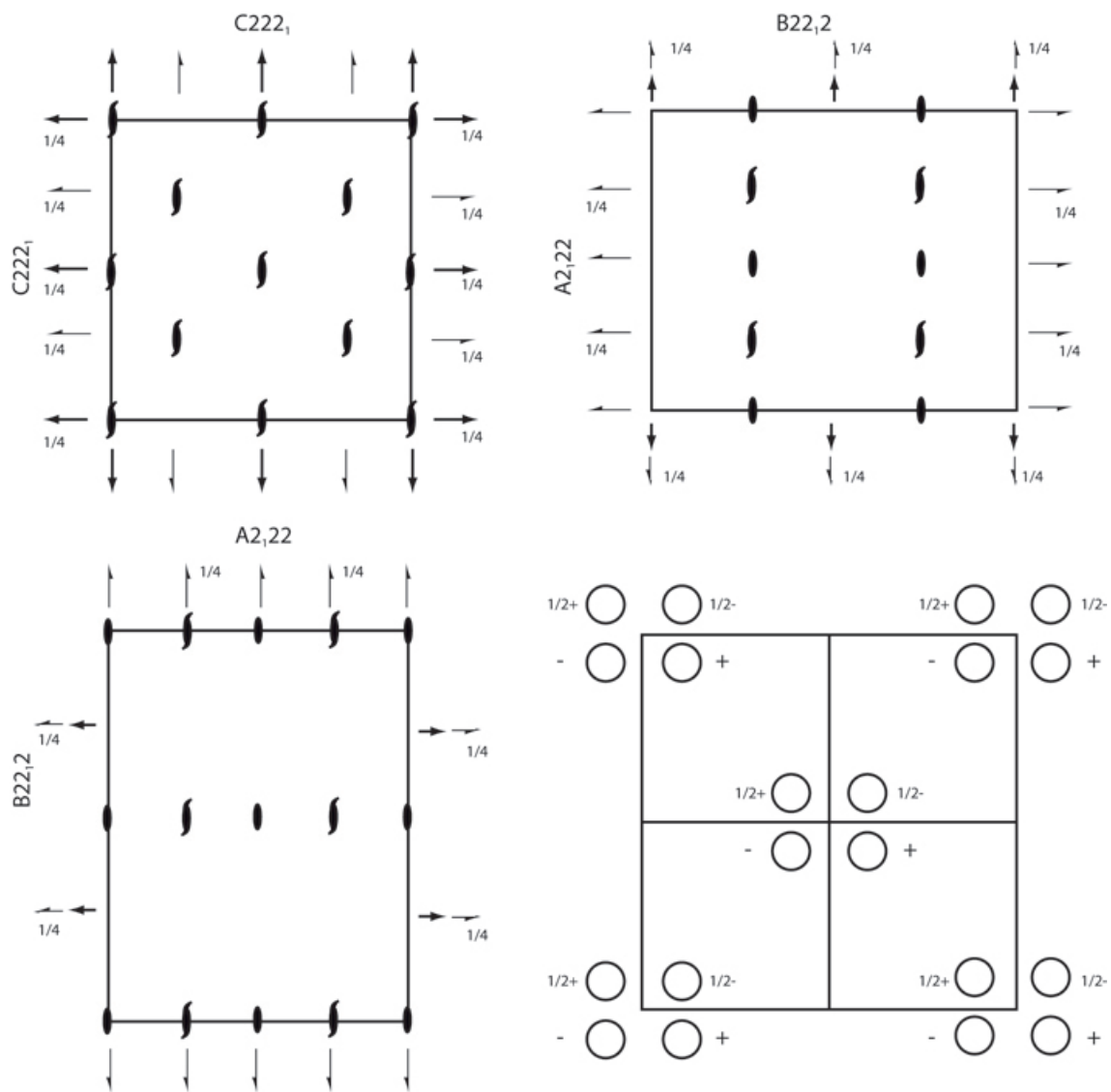


FIG. 16.1: Axes de symétrie et position équivalente à l'intérieur d'un cristal dont le groupe d'espace est $C222_1$.

Bibliographie

- Adler, B., Schaadt, E., Kempkes, B., Zimmer-Strobl, U., Baier, B., & Bornkamm, G. W. (2002). Control of Epstein-Barr virus reactivation by activated CD40 and viral latent membrane protein 1. *Proc Natl Acad Sci U S A*, **99** (1), 437–442.
- Albà, M. M., Lee, D., Pearl, F. M., Shepherd, A. J., Martin, N., Orengo, C. A., & Kellam, P. (2001). Vida : a virus database system for the organization of animal virus genome open reading frames. *Nucleic Acids Res*, **29** (1), 133–136.
- Altmann, M. & Hammerschmidt, W. (2005). Epstein-Barr virus provides a new paradigm : a requirement for the immediate inhibition of apoptosis. *PLoS Biol*, **3** (12).
- Andersson, J. (2006). Epstein-Barr virus and Hodgkin's lymphoma. *Herpes*, **13** (1), 12–16.
- Bairoch, A., Boeckmann, B., Ferro, S., & Gasteiger, E. (2004). Swiss-Prot : juggling between evolution and stability. *Brief Bioinform*, **5** (1), 39–55.
- Bajorath, J. (2002). Integration of virtual and high-throughput screening. *Nat Rev Drug Discov*, **1** (11), 882–894.
- Banerjee, A., Santos, W. L., & Verdine, G. L. (2006). Structure of a DNA glycosylase searching for lesions. *Science*, **311** (5764), 1153–1157.
- Batisse, J., Manet, E., Middeldorp, J., Sergeant, A., & Gruffat, H. (2005). Epstein-Barr virus mRNA export factor EB2 is essential for intranuclear capsid assembly and production of gp350. *J Virol*, **79** (22), 14102–14111.

- Bellamy, S. & Baldwin, G. (2001). A kinetic analysis of substrate recognition by uracil-dna glycosylase from herpes simplex virus type 1. *Nucleic Acids Res*, **29** (18), 3857–3863.
- Bennett, S. E., Schimerlik, M. I., & Mosbaugh, D. W. (1993). Kinetics of the uracil-dna glycosylase/inhibitor protein association. ung interaction with ugi, nucleic acids, and uracil compounds. *J Biol Chem*, **268** (36), 26879–26885.
- Blainey, P. C., van Oijen, A. M., Banerjee, A., Verdine, G. L., & Xie, X. S. (2006). A base-excision dna-repair protein finds intrahelical lesion bases by fast sliding in contact with dna. *Proc Natl Acad Sci U S A*, **103** (15), 5752–5757.
- Blundell, T. L., Lapatto, R., Wilderspin, A. F., Hemmings, A. M., Hobart, P. M., Danley, D. E., & Whittle, P. J. (1990). The 3-d structure of hiv-1 proteinase and the design of antiviral agents for the treatment of aids. *Trends Biochem Sci*, **15** (11), 425–430.
- Boyer, J. L., Swaminathan, S., & Silverstein, S. J. (2002). The epstein-barr virus sm protein is functionally similar to icp27 from herpes simplex virus in viral infections. *J Virol*, **76** (18), 9420–9433.
- Buell, P. (1974). The effect of migration on the risk of nasopharyngeal cancer among chinese. *Cancer Res*, **34** (5), 1189–1191.
- Buisson, M., Hernandez, J. F., Lascoux, D., Schoehn, G., Forest, E., Arlaud, G., Seigneurin, J. M., Ruigrok, R. W., & Burmeister, W. P. (2002). The crystal structure of the epstein-barr virus protease shows rearrangement of the processed c terminus. *J Mol Biol*, **324** (1), 89–103.
- Burkitt (1994). Direction determines destination. *Assoc Old Adelaide Stud J* ., 9–11.
- Burkitt, D. Kunuda, B. (1947). Enormous intra-abdominal hydrocele. *BMJ*, **2**, 175.
- Burkitt, D. (1957). Report on bukuva, 20/8/57. *Burkitt Mss, TCD Archives*, .
- Burkitt, D. (1958a). A sarcoma involving the jaws in african children. *Br J Surg*, **46** (197), 218–223.
- Burkitt, D. (1958b). Sarcoma involving jaws. In : *Presentation to the East African Association of Surgeons, Kampala Burkitt Mss, TCD Archives* .,

- Burkitt, D. (1964). A lymphoma syndrome dependent on environment. part ii. epidemiological factors. In : *Symposium on lymphoreticular tumours in Africa* ,.
- Burkitt, D. (1992). Letter from denis burkitt to d.j. mcconnell, 27 octobre 1992. *Burkitt MSS, TCD Archives*, .
- Burkitt, D. (1993). Unpromising beginnings. *J Ir Coll Phys Surg*, **22**, 36–38.
- Burkitt, D., Hutt, M. S., & Wright, D. H. (1965). The african lymphoma : Preliminary observations on response to therapy. *Cancer*, **18**, 399–410.
- Burkitt, D. & O’Conor, G. T. (1961). Malignant lymphoma in african children. i. a clinical syndrome. *Cancer*, **14**, 258–269.
- Burkitt, D. P. (1951). Primary hydrocele and its treatment. review of two hundred cases. *Lancet*, **1** (25), 1341–1343.
- Burkitt, D. P. (1971). Epidemiology of cancer of the colon and rectum. *Cancer*, **28** (1), 3–13.
- Burkitt, D. Davies, J. (1961). Lymphoma syndrome in uganda and tropical africa. *Med Press*, **245**, 367–369.
- Burmeister, W. P. (2000). Structural changes in a cryo-cooled protein crystal owing to radiation damage. *Acta Crystallogr D Biol Crystallogr*, **56** (Pt 3), 328–341.
- Cao, C., Jiang, Y. L., Stivers, J. T., & Song, F. (2004). Dynamic opening of dna during the enzymatic search for a damaged base. *Nat Struct Mol Biol*, **11** (12), 1230–1236.
- Caradonna, S. J. & Cheng, Y. C. (1980). Uracil dna-glycosylase. purification and properties of this enzyme isolated from blast cells of acute myelocytic leukemia patients. *J Biol Chem*, **255** (6), 2293–2300.
- CCP4 (1994). The ccp4 suite : Programs for protein crystallography. *Acta Cryst.* **D50**, 760–763.

- Chen, C. Y., Mosbaugh, D. W., & Bennett, S. E. (2005). Mutations at arginine 276 transform human uracil-dna glycosylase into a single-stranded dna-specific uracil-dna glycosylase. *DNA Repair (Amst)*, **4** (7), 793–805.
- Chen, L., Haushalter, K. A., Lieber, C. M., & Verdine, G. L. (2002). Direct visualization of a dna glycosylase searching for damage. *Chem Biol*, **9** (3), 345–350.
- Chen, M. R., Chang, S. J., Huang, H., & Chen, J. Y. (2000). A protein kinase activity associated with epstein-barr virus bglf4 phosphorylates the viral early antigen ea-d in vitro. *J Virol*, **74** (7), 3093–3104.
- Chen, R., Le Rouzic, E., Kearney, J. A., Mansky, L. M., & Benichou, S. (2004). Vpr-mediated incorporation of ung2 into hiv-1 particles is required to modulate the virus mutation rate and for replication in macrophages. *J Biol Chem*, **279** (27), 28419–28425.
- Chen, R., Wang, H., & Mansky, L. M. (2002). Roles of uracil-dna glycosylase and dutpase in virus replication. *J Gen Virol*, **83** (Pt 10), 2339–2345.
- Cheng, Y. C., Chen, J. Y., Hoffmann, P. J., & Glaser, R. (1980). Studies on the activity of dnase associated with the replication of the epstein-barr virus. *Virology*, **100** (2), 334–338.
- Chevallier-Greco, A., Manet, E., Chavrier, P., Mosnier, C., Daillie, J., & Sergeant, A. (1986). Both epstein-barr virus (ebv)-encoded trans-acting factors, eb1 and eb2, are required to activate transcription from an ebv early promoter. *EMBO J*, **5** (12), 3243–3249.
- Coakley, D. (1992). *Denis Burkitt - A great twentieth-century Irish doctor*. Town House - Dublin.
- Coen, D. M. & Schaffer, P. A. (2003). Antih herpesvirus drugs : a promising spectrum of new drugs and drug targets. *Nat Rev Drug Discov*, **2** (4), 278–288.
- Cohen, J. (2005). *Clinical Aspects of Epstein-Barr Virus Infection*. Caister Academic Press.
- Combet, C., Blanchet, C., Geourjon, C., & Deléage, G. (2000). Nps@ : network protein sequence analysis. *Trends Biochem Sci*, **25** (3), 147–150.

- Countryman, J., Jenson, H., Seibl, R., Wolf, H., & Miller, G. (1987). Polymorphic proteins encoded within bzlfl1 of defective and standard epstein-barr viruses disrupt latency. *J Virol*, **61** (12), 3672–3679.
- Countryman, J. & Miller, G. (1985). Activation of expression of latent epstein-barr herpesvirus after gene transfer with a small cloned subfragment of heterogeneous viral dna. *Proc Natl Acad Sci U S A*, **82** (12), 4085–4089.
- Daikoku, T., Kudoh, A., Sugaya, Y., Iwahori, S., Shirata, N., Isomura, H., & Tsurumi, T. (2006). Postreplicative mismatch repair factors are recruited to epstein-barr virus replication compartments. *J Biol Chem*, **281** (16), 11422–11430.
- Daniels, D. S., Woo, T. T., Luu, K. X., Noll, D. M., Clarke, N. D., Pegg, A. E., & Tainer, J. A. (2004). Dna binding and nucleotide flipping by the human dna repair protein agt. *Nat Struct Mol Biol*, **11** (8), 714–720.
- Deléage, G. (a). Dichroism of proteins. <http://dicroprot-pbil.ibcp.fr/>.
- Dickson, R. I. & Flores, A. D. (1985). Nasopharyngeal carcinoma : an evaluation of 134 patients treated between 1971-1980. *Laryngoscope*, **95** (3), 276–283.
- Dodson, M. L., Michaels, M. L., & Lloyd, R. S. (1994). Unified catalytic mechanism for dna glycosylases. *J Biol Chem*, **269** (52), 32709–32712.
- Dupuis, J., Emile, J. F., Mounier, N., Gisselbrecht, C., Martin-Garcia, N., Petrella, T., Bouabdallah, R., Berger, F., Delmer, A., Coiffier, B., Reyes, F., & Gaulard, P. (2006). Prognostic significance of epstein-barr virus in nodal peripheral t-cell lymphoma, unspecified : A groupe d' eude des lmpomes de l' adulte (gela) study. *Blood*, .
- Ellison, K. S., Peng, W., & McFadden, G. (1996). Mutations in active-site residues of the uracil-dna glycosylase encoded by vaccinia virus are incompatible with virus viability. *J Virol*, **70** (11), 7965–7973.

- Epstein, M. A. (1962a). Observations on the mode of release of herpes virus from infected hela cells. *J Cell Biol*, **12**, 589–597.
- Epstein, M. A. (1962b). Observations on the fine structure of mature herpes simplex virus and on the composition of its nucleoid. *J Exp Med*, **115**, 1–12.
- Epstein, M. A. (2005). *The origins of EBV Research : Discovery and Characterization of the Virus*. Caister Academic Press.
- Epstein, M. A., Achong, B. G., & Barr, Y. M. (1964). Virus particles in cultured lymphoblasts from burkitt's lymphoma. *Lancet*, **15**, 702–703.
- Epstein, M. A., Henle, G., Achong, B. G., & Barr, Y. M. (1965). Morphological and biological studies on a virus in cultured lymphoblasts from burkitt's lymphoma. *J Exp Med*, **121**, 761–770.
- Feng, W. H., Hong, G., Delecluse, H. J., & Kenney, S. C. (2004). Lytic induction therapy for epstein-barr virus-positive b-cell lymphomas. *J Virol*, **78** (4), 1893–1902.
- Fixman, E. D., Hayward, G. S., & Hayward, S. D. (1995). Replication of epstein-barr virus orilyt : lack of a dedicated virally encoded origin-binding protein and dependence on zta in cotransfection assays. *J Virol*, **69** (5), 2998–3006.
- Fleischmann, J., Kremmer, E., Greenspan, J. S., Grässer, F. A., & Niedobitek, G. (2002). Expression of viral and human dutpase in epstein-barr virus-associated diseases. *J Med Virol*, **68** (4), 568–573.
- Fleming, A. (2001). On the antibacterial action of cultures of a penicillium, with special reference to their use in the isolation of b. influenzae. 1929. *Bull World Health Organ*, **79** (8), 780–790.
- Flint, S. & al (2004). *Principles of Virology, Molecular Biology, Pathogenesis, and Control of Animal Viruses. 2nd Edition*. ASM Press.
- Fresen, K. O., Cho, M. S., & zur Hausen, H. (1978). Recovery of transforming ebv from non-producer cells after superinfection with non-transforming p3hr-1 ebv. *Int J Cancer*, **22** (4), 378–383.
- Fromme, J. C. & Verdine, G. L. (2002). Structural insights into lesion recognition and repair by the bacterial 8-oxoguanine dna glycosylase mutm. *Nat Struct Biol*, **9** (7), 544–552.

- Fromme, J. C. & Verdine, G. L. (2003). Structure of a trapped endonuclease iii-dna covalent intermediate. *EMBO J*, **22** (13), 3461–3471.
- Fu, Z. & Cannon, M. J. (2000). Functional analysis of the cd4(+) t-cell response to epstein-barr virus : T-cell-mediated activation of resting b cells and induction of viral bzlfl1 expression. *J Virol*, **74** (14), 6675–6679.
- Furman, P. A., McGuirt, P. V., Keller, P. M., Fyfe, J. A., & Elion, G. B. (1980). Inhibition by acyclovir of cell growth and dna synthesis of cells biochemically transformed with herpesvirus genetic information. *Virology*, **102** (2), 420–430.
- Fuxreiter, M., Luo, N., Jedlovsky, P., Simon, I., & Osman, R. (2002). Role of base flipping in specific recognition of damaged dna by repair enzymes. *J Mol Biol*, **323** (5), 823–834.
- Galicchio, L., Matanoski, G., Tao, X. G., Chen, L., Lam, T. K., Boyd, K., Robinson, K. A., Balick, L., Mickelson, S., Caulfield, L. E., Herman, J. G., Guallar, E., & Alberg, A. J. (2006). Adulthood consumption of preserved and nonpreserved vegetables and the risk of nasopharyngeal carcinoma : a systematic review. *Int J Cancer*, **119** (5), 1125–1135.
- Garcia-Ruiz, J. M., Gonzalez-Ramirez, L. A., Gavira, J. A., & Otálora, F. (2002). Granada crystallisation box : a new device for protein crystallisation by counter-diffusion techniques. *Acta Crystallogr D Biol Crystallogr*, **58** (Pt 10 Pt 1), 1638–1642.
- Gasteiger, E., Gattiker, A., Hoogland, C., Ivanyi, I., Appel, R. D., & Bairoch, A. (2003). Expasy : The proteomics server for in-depth protein knowledge and analysis. *Nucleic Acids Res*, **31** (13), 3784–3788.
- Gershburg, E., Marschall, M., Hong, K., & Pagano, J. S. (2004). Expression and localization of the epstein-barr virus-encoded protein kinase. *J Virol*, **78** (22), 12140–12146.
- Gilboa, R., Zharkov, D. O., Golan, G., Fernandes, A. S., Gerchman, S. E., Matz, E., Kycia, J. H., Grollman, A. P., & Shoham, G. (2002). Structure of formamidopyrimidine-dna glycosylase covalently complexed to dna. *J Biol Chem*, **277** (22), 19811–19816.

- Glemser, B. (1971). *The Long Safari [Mr. Burkitt and Africa]*. Bodley Head.
- Gnann, J. W., Barton, N. H., & Whitley, R. J. (1983). Acyclovir : mechanism of action, pharmacokinetics, safety and clinical applications. *Pharmacotherapy*, **3** (5), 275–283.
- Goldstein, J. N. & Weller, S. K. (1998). The exonuclease activity of hsv-1 ul12 is required for in vivo function. *Virology*, **244** (2), 442–457.
- Guerreiro-Cacais, A. O., Li, L., Donati, D., Bejarano, M. T., Morgan, A., Masucci, M. G., Hutt-Fletcher, L., & Levitsky, V. (2004). Capacity of epstein-barr virus to infect monocytes and inhibit their development into dendritic cells is affected by the cell type supporting virus replication. *J Gen Virol*, **85** (Pt 10), 2767–2778.
- Gupta, A., Gartner, J. J., Sethupathy, P., Hatzigeorgiou, A. G., & Fraser, N. W. (2006). Anti-apoptotic function of a microrna encoded by the hsv-1 latency-associated transcript. *Nature*, .
- Gutiérrez, M. I., Judde, J. G., Magrath, I. T., & Bhatia, K. G. (1996). Switching viral latency to viral lysis : a novel therapeutic approach for epstein-barr virus-associated neoplasia. *Cancer Res*, **56** (5), 969–972.
- Hardwick, J. M., Lieberman, P. M., & Hayward, S. D. (1988). A new epstein-barr virus transactivator, *r*, induces expression of a cytoplasmic early antigen. *J Virol*, **62** (7), 2274–2284.
- Hatzivassiliou, E. G., Kieff, E., & Mosialos, G. (2006). Constitutive cd40 signaling phenocopies the transforming function of the epstein-barr virus oncoprotein lmp1 in vitro. *Leuk Res*, .
- Henderson, S., Huen, D., Rowe, M., Dawson, C., Johnson, G., & Rickinson, A. (1993). Epstein-barr virus-coded bhrf1 protein, a viral homologue of bcl-2, protects human b cells from programmed cell death. *Proc Natl Acad Sci U S A*, **90** (18), 8479–8483.
- Henle, G. & Henle, W. (1966). Immunofluorescence in cells derived from burkitt's lymphoma. *J Bacteriol*, **91** (3), 1248–1256.

- Henle, G., Henle, W., Clifford, P., Diehl, V., Kafuko, G. W., Kirya, B. G., Klein, G., Morrow, R. H., Munube, G. M., Pike, P., Tukei, P. M., & Ziegler, J. L. (1969). Antibodies to epstein-barr virus in burkitt's lymphoma and control groups. *J Natl Cancer Inst*, **43** (5), 1147–1157.
- Henle, G., Henle, W., & Diehl, V. (1968). Relation of burkitt's tumor-associated herpes-type virus to infectious mononucleosis. *Proc Natl Acad Sci U S A*, **59** (1), 94–101.
- Hildesheim, A., Apple, R. J., Chen, C. J., Wang, S. S., Cheng, Y. J., Klitz, W., Mack, S. J., Chen, I. H., Hsu, M. M., Yang, C. S., Brinton, L. A., Levine, P. H., & Erlich, H. A. (2002). Association of hla class i and ii alleles and extended haplotypes with nasopharyngeal carcinoma in taiwan. *J Natl Cancer Inst*, **94** (23), 1780–1789.
- Hiriart, E., Farjot, G., Gruffat, H., Nguyen, M. V., Sergeant, A., & Manet, E. (2003). A novel nuclear export signal and a ref interaction domain both promote mrna export by the epstein-barr virus eb2 protein. *J Biol Chem*, **278** (1), 335–342.
- Hoeijmakers, J. H. (2001). Genome maintenance mechanisms for preventing cancer. *Nature*, **411** (6835), 366–374.
- Hsu, D. H., de Waal Malefyt, R., Fiorentino, D. F., Dang, M. N., Vieira, P., de Vries, J., Spits, H., Mosmann, T. R., & Moore, K. W. (1990). Expression of interleukin-10 activity by epstein-barr virus protein bcrf1. *Science*, **250** (4982), 830–832.
- Huang, N. & MacKerell, A. D. (2004). Atomistic view of base flipping in dna. *Philos Transact A Math Phys Eng Sci*, **362** (1820), 1439–1460.
- Huffman, J. L., Sundheim, O., & Tainer, J. A. (2005). Dna base damage recognition and removal : new twists and grooves. *Mutat Res*, **577** (1-2), 55–76.
- Hulo, N., Bairoch, A., Bulliard, V., Cerutti, L., De Castro, E., Langendijk-Genevaux, P. S., Pagni, M., & Sigrist, C. J. (2006). The prosite database. *Nucleic Acids Res*, **34** (Database issue), 227–230.
- Inman, G. J., Binné, U. K., Parker, G. A., Farrell, P. J., & Allday, M. J. (2001). Activators of the epstein-barr virus lytic program concomitantly induce apoptosis, but lytic gene expression protects from cell death. *J Virol*, **75** (5), 2400–2410.

- Israel, F. & Kenney, S. (2005). *EBV Lytic Infection*. Caister Academic Press.
- Jay, V. (1998). Extraordinary epidemiological quest of dr. burkitt. *Pediatr Dev Pathol*, **1** (6), 562–564.
- Jiricny, J. (2006). The multifaceted mismatch-repair system. *Nat Rev Mol Cell Biol*, **7** (5), 335–346.
- Johannsen, E., Luftig, M., Chase, M. R., Weicksel, S., Cahir-McFarland, E., Illanes, D., Sarracino, D., & Kieff, E. (2004). Proteins of purified epstein-barr virus. *Proc Natl Acad Sci U S A*, **101** (46), 16286–16291.
- Kanehisa, M., Goto, S., Hattori, M., Aoki-Kinoshita, K. F., Itoh, M., Kawashima, S., Katayama, T., Araki, M., & Hirakawa, M. (2006). From genomics to chemical genomics : new developments in kegg. *Nucleic Acids Res*, **34** (Database issue), 354–357.
- Kato, K., Kawaguchi, Y., Tanaka, M., Igarashi, M., Yokoyama, A., Matsuda, G., Kanamori, M., Nakajima, K., Nishimura, Y., Shimojima, M., Phung, H. T., Takahashi, E., & Hirai, K. (2001). Epstein-barr virus-encoded protein kinase bglf4 mediates hyperphosphorylation of cellular elongation factor 1delta (ef-1delta) : Ef-1delta is universally modified by conserved protein kinases of herpesviruses in mammalian cells. *J Gen Virol*, **82** (Pt 6), 1457–1463.
- Kavli, B., Slupphaug, G., Mol, C. D., Arvai, A. S., Peterson, S. B., Tainer, J. A., & Krokan, H. E. (1996). Excision of cytosine and thymine from dna by mutants of human uracil-dna glycosylase. *EMBO J*, **15** (13), 3442–3447.
- Kavli, B., Sundheim, O., Akbari, M., Otterlei, M., Nilsen, H., Skorpen, F., Aas, P. A., Hagen, L., Krokan, H. E., & Slupphaug, G. (2002). hung2 is the major repair enzyme for removal of uracil from u :a matches, u :g mismatches, and u in single-stranded dna, with hsmug1 as a broad specificity backup. *J Biol Chem*, **277** (42), 39926–39936.
- Kelly, G. L., Milner, A. E., Baldwin, G. S., Bell, A. I., & Rickinson, A. B. (2006). Three restricted forms of epstein-barr virus latency counteracting apoptosis in c-myc-expressing burkitt lymphoma cells. *Proc Natl Acad Sci U S A*, **103** (40), 14935–14940.

- Kennedy, G., Komano, J., & Sugden, B. (2003). Epstein-barr virus provides a survival factor to burkitt's lymphomas. *Proc Natl Acad Sci U S A*, **100** (24), 14269–14274.
- Khanna, K. K. & Jackson, S. P. (2001). Dna double-strand breaks : signaling, repair and the cancer connection. *Nat Genet*, **27** (3), 247–254.
- Khanna, R. & Burrows, S. R. (2000). Role of cytotoxic t lymphocytes in epstein-barr virus-associated diseases. *Annu Rev Microbiol*, **54**, 19–48.
- Kieff, E. & Rickinson, A. B. (2001). *Fields Virology*. Lippincott Williams and Wilkins, Philadelphia.
- Kim, O. & YI, S. J. (2003). Lytic replication of human herpesvirus 8 and induction of apoptosis. *Acta Virol*, **47** (2), 91–95.
- Kleymann, G. (2003). Novel agents and strategies to treat herpes simplex virus infections. *Expert Opin Investig Drugs*, **12** (2), 165–183.
- Kovall, R. & Matthews, B. W. (1997). Toroidal structure of lambda-exonuclease. *Science*, **277** (5333), 1824–1827.
- Krokan, H. & Wittwer, C. U. (1981). Uracil dna-glycosylase from hela cells : general properties, substrate specificity and effect of uracil analogs. *Nucleic Acids Res*, **9** (11), 2599–2613.
- Krokan, H. E., Standal, R., & Slupphaug, G. (1997). Dna glycosylases in the base excision repair of dna. *Biochem J*, **325** (Pt 1), 1–16.
- Krusong, K., Carpenter, E. P., Bellamy, S. R., Savva, R., & Baldwin, G. S. (2006). A comparative study of uracil-dna glycosylases from human and herpes simplex virus type 1. *J Biol Chem*, **281** (8), 4983–4992.
- Kumar, S. & Subramanian, S. (2002). Mutation rates in mammalian genomes. *Proc Natl Acad Sci U S A*, **99** (2), 803–808.
- Lee, T. C. (1996). Seeing the wood for the trees—the early papers of denis burkitt. *J Ir Coll Physicians Surg*, **25** (2), 126–130. Biography.

- Leiros, I., Moe, E., Lanes, O., Smalås, A. O., & Willassen, N. P. (2003). The structure of uracil-dna glycosylase from atlantic cod (*gadus morhua*) reveals cold-adaptation features. *Acta Crystallogr D Biol Crystallogr*, **59** (Pt 8), 1357–1365.
- Leiros, I., Moe, E., Smalås, A. O., & McSweeney, S. (2005). Structure of the uracil-dna n-glycosylase (ung) from *deinococcus radiodurans*. *Acta Crystallogr D Biol Crystallogr*, **61** (Pt 8), 1049–1056.
- Leslie, A. G. (2006). The integration of macromolecular diffraction data. *Acta Crystallogr D Biol Crystallogr*, **62** (Pt 1), 48–57.
- Liao, G., Wu, F. Y., & Hayward, S. D. (2001). Interaction with the epstein-barr virus helicase targets zta to dna replication compartments. *J Virol*, **75** (18), 8792–8802.
- Lindahl, T. (1974). An n-glycosidase from *escherichia coli* that releases free uracil from dna containing deaminated cytosine residues. *Proc Natl Acad Sci U S A*, **71** (9), 3649–3653.
- Lindahl, T. (1993). Instability and decay of the primary structure of dna. *Nature*, **362** (6422), 709–715.
- Lindahl, T. & Barnes, D. E. (2000). Repair of endogenous dna damage. *Cold Spring Harb Symp Quant Biol*, **65**, 127–133.
- Lindahl, T. & Nyberg, B. (1974). Heat-induced deamination of cytosine residues in deoxyribonucleic acid. *Biochemistry*, **13** (16), 3405–3410.
- Littler, E., Zeuthen, J., McBride, A. A., Trost Sorensen, E., Powell, K. L., Walsh-Arrand, J. E., & Arrand, J. R. (1986). Identification of an epstein-barr virus-coded thymidine kinase. *EMBO J*, **5** (8), 1959–1966.
- Luce, A. A. (1929). Letter from luce, a.a. to james burkitt, 11 july 1929. *Burkitt Mss, TCD Archives*, .
- Mannick, J. B., Cohen, J. I., Birkenbach, M., Marchini, A., & Kieff, E. (1991). The epstein-barr virus nuclear protein encoded by the leader of the ebna rnas is important in b-lymphocyte transformation. *J Virol*, **65** (12), 6826–6837.

- Masaoka, A., Matsubara, M., Hasegawa, R., Tanaka, T., Kurisu, S., Terato, H., Ohyama, Y., Karino, N., Matsuda, A., & Ide, H. (2003). Mammalian 5-formyluracil-dna glycosylase. 2. role of smug1 uracil-dna glycosylase in repair of 5-formyluracil and other oxidized and deaminated base lesions. *Biochemistry*, **42** (17), 5003–5012.
- Menet, A., Speth, C., Larcher, C., Prodinger, W. M., Schwendinger, M. G., Chan, P., Jäger, M., Schwarzmann, F., Recheis, H., Fontaine, M., & Dierich, M. P. (1999). Epstein-barr virus infection of human astrocyte cell lines. *J Virol*, **73** (9), 7722–7733.
- Mennigmann, H. D. & Szybalski, W. (1962). Molecular mechanism of thymine-less death. *Biochem Biophys Res Commun*, **9**, 398–404.
- Mettenleiter, T. C. (2006). Intriguing interplay between viral proteins during herpesvirus assembly or : the herpesvirus assembly puzzle. *Vet Microbiol*, **113** (3-4), 163–169.
- Mettenleiter, T. C., Klupp, B. G., & Granzow, H. (2006). Herpesvirus assembly : a tale of two membranes. *Curr Opin Microbiol*, **9** (4), 423–429.
- Meyohas, M. C., Maréchal, V., Desire, N., Bouillie, J., Frottier, J., & Nicolas, J. C. (1996). Study of mother-to-child epstein-barr virus transmission by means of nested pcrs. *J Virol*, **70** (10), 6816–6819.
- Michel, D. & Mertens, T. (2004). The ul97 protein kinase of human cytomegalovirus and homologues in other herpesviruses : impact on virus and host. *Biochim Biophys Acta*, **1697** (1-2), 169–180.
- Miller, I. G. & El-Guindy, A. (2002). Regulation of epstein-barr virus lytic cycle activation in malignant and nonmalignant disease. *J Natl Cancer Inst*, **94** (23), 1733–1735.
- Miller, M., Schneider, J., Sathyanarayana, B. K., Toth, M. V., Marshall, G. R., Clawson, L., Selk, L., Kent, S. B., & Wlodawer, A. (1989). Structure of complex of synthetic hiv-1 protease with a substrate-based inhibitor at 2.3 a resolution. *Science*, **246** (4934), 1149–1152.
- Mol, C. D., Arvai, A. S., Sanderson, R. J., Slupphaug, G., Kavli, B., Krokan, H. E., Mosbaugh, D. W., & Tainer, J. A. (1995a). Crystal structure of human uracil-dna glycosylase in complex with a protein inhibitor : protein mimicry of dna. *Cell*, **82** (5), 701–708.

- Mol, C. D., Arvai, A. S., Slupphaug, G., Kavli, B., Alseth, I., Krokan, H. E., & Tainer, J. A. (1995b). Crystal structure and mutational analysis of human uracil-dna glycosylase : structural basis for specificity and catalysis. *Cell*, **80** (6), 869–878.
- Mol, C. D., Parikh, S. S., Putnam, C. D., Lo, T. P., & Tainer, J. A. (1999). Dna repair mechanisms for the recognition and removal of damaged dna bases. *Annu Rev Biophys Biomol Struct*, **28**, 101–128.
- Mouret, S., Baudouin, C., Charveron, M., Favier, A., Cadet, J., & Douki, T. (2006). Cyclobutane pyrimidine dimers are predominant dna lesions in whole human skin exposed to uva radiation. *Proc Natl Acad Sci U S A*, **103** (37), 13765–13770.
- Mulder, N. J., Apweiler, R., Attwood, T. K., Bairoch, A., Bateman, A., Binns, D., Bradley, P., Bork, P., Bucher, P., Cerutti, L., Copley, R., Courcelle, E., Das, U., Durbin, R., Fleischmann, W., Gough, J., Haft, D., Harte, N., Hulo, N., Kahn, D., Kanapin, A., Krestyaninova, M., Lonsdale, D., Lopez, R., Letunic, I., Madera, M., Maslen, J., McDowall, J., Mitchell, A., Nikolskaya, A. N., Orchard, S., Pagni, M., Ponting, C. P., Quevillon, E., Selengut, J., Sigrist, C. J., Silventoinen, V., Studholme, D. J., Vaughan, R., & Wu, C. H. (2005). Interpro, progress and status in 2005. *Nucleic Acids Res*, **33** (Database issue), 201–205.
- Murray, P. & Young, L. (2005). *Hodgkin's Lymphoma : Molecular Pathogenesis and the Contribution of the Epstein-Barr Virus*. Caister Academic Press.
- Nanbo, A., Inoue, K., Adachi-Takasawa, K., & Takada, K. (2002). Epstein-barr virus rna confers resistance to interferon-alpha-induced apoptosis in burkitt's lymphoma. *EMBO J*, **21** (5), 954–965.
- Nelson, E. R. (1995). Out of africa, major medical discoveries, dr. denis burkitt. *Saturday Evening Post*, **March-April**, 1.
- Nelson C, L. & Temple N, J. (1994). Tribute to denis burkitt. *Journal of Medical Biography*, **2**, 180–183.
- Neuhierl, B. & Delecluse, H. J. (2006). The epstein-barr virus bmrfl gene is essential for lytic virus replication. *J Virol*, **80** (10), 5078–5081.

- Nicolas, J. & Seigneurin, J. (2002). *Virologie Médicale, Herpesviridae : Virus d'Epstein-Barr*. Presses Universitaires de Lyon.
- O'Connor, G. T. & Davies, J. N. (1960). Malignant tumors in african children. with special reference to malignant lymphoma. *J Pediatr*, **56**, 526–535.
- Old, L. J., Boyse, E. A., Oettgen, H. F., Harven, E. D., Geering, G., Williamson, B., & Clifford, P. (1966). Precipitating antibody in human serum to an antigen present in cultured burkitt's lymphoma cells. *Proc Natl Acad Sci U S A*, **56** (6), 1699–1704.
- Otterlei, M., Haug, T., Nagelhus, T. A., Slupphaug, G., Lindmo, T., & Krokan, H. E. (1998). Nuclear and mitochondrial splice forms of human uracil-dna glycosylase contain a complex nuclear localisation signal and a strong classical mitochondrial localisation signal, respectively. *Nucleic Acids Res*, **26** (20), 4611–4617.
- Parikh, S. S., Mol, C. D., Slupphaug, G., Bharati, S., Krokan, H. E., & Tainer, J. A. (1998). Base excision repair initiation revealed by crystal structures and binding kinetics of human uracil-dna glycosylase with dna. *Embo J*, **17** (17), 5214–26. 0261-4189 Journal Article.
- Parikh, S. S., Mol, C. D., & Tainer, J. A. (1997). Base excision repair enzyme family portrait : integrating the structure and chemistry of an entire dna repair pathway. *Structure*, **5** (12), 1543–1550.
- Pearl, L. H. (2000). Structure and function in the uracil-dna glycosylase superfamily. *Mutat Res*, **460** (3-4), 165–181.
- Pelletier, H., Sawaya, M. R., Wolfle, W., Wilson, S. H., & Kraut, J. (1996). Crystal structures of human dna polymerase beta complexed with dna : implications for catalytic mechanism, processivity, and fidelity. *Biochemistry*, **35** (39), 12742–12761.
- Petersen-Mahrt, S. K., Harris, R. S., & Neuberger, M. S. (2002). Aid mutates e. coli suggesting a dna deamination mechanism for antibody diversification. *Nature*, **418** (6893), 99–103.

- Petosa, C., Morand, P., Baudin, F., Moulin, M., Artero, J. B., & Müller, C. W. (2006). Structural basis of lytic cycle activation by the epstein-barr virus zebra protein. *Mol Cell*, **21** (4), 565–572.
- Prichard, M. N., Duke, G. M., & Mocarski, E. S. (1996). Human cytomegalovirus uracil dna glycosylase is required for the normal temporal regulation of both dna synthesis and viral replication. *J Virol*, **70** (5), 3018–3025.
- Priet, S., Gros, N., Navarro, J. M., Boretto, J., Canard, B., Quérat, G., & Sire, J. (2005). Hiv-1-associated uracil dna glycosylase activity controls dutp misincorporation in viral dna and is essential to the hiv-1 life cycle. *Mol Cell*, **17** (4), 479–490.
- Prilusky, J., Felder, C. E., Zeev-Ben-Mordehai, T., Rydberg, E. H., Man, O., Beckmann, J. S., Silman, I., & Sussman, J. L. (2005). Foldindex : a simple tool to predict whether a given protein sequence is intrinsically unfolded. *Bioinformatics*, **21** (16), 3435–3438.
- Pyles, R. B. & Thompson, R. L. (1994). Evidence that the herpes simplex virus type 1 uracil dna glycosylase is required for efficient viral replication and latency in the murine nervous system. *J Virol*, **68** (8), 4963–4972.
- Ragoczy, T. & Miller, G. (1999). Role of the epstein-barr virus rta protein in activation of distinct classes of viral lytic cycle genes. *J Virol*, **73** (12), 9858–9866.
- Rasti, N., Falk, K. I., Donati, D., Gyan, B. A., Goka, B. Q., Troye-Blomberg, M., Akanmori, B. D., Kurtzhals, J. A., Dodoo, D., Consolini, R., Linde, A., Wahlgren, M., & Bejarano, M. T. (2005). Circulating epstein-barr virus in children living in malaria-endemic areas. *Scand J Immunol*, **61** (5), 461–465.
- Ravishankar, R., Bidya Sagar, M., Roy, S., Purnapatre, K., Handa, P., Varshney, U., & Vijayan, M. (1998). X-ray analysis of a complex of escherichia coli uracil dna glycosylase (ecudg) with a proteinaceous inhibitor. the structure elucidation of a prokaryotic udg. *Nucleic Acids Res*, **26** (21), 4880–4887.
- Read, J. (a). The macromolecular crystallography course. <http://www-structmed.cimr.cam.ac.uk/course.html>.

- Reardon, J. E. & Spector, T. (1989). Herpes simplex virus type 1 dna polymerase. mechanism of inhibition by acyclovir triphosphate. *J Biol Chem*, **264** (13), 7405–7411.
- Reddy, S. M., Williams, M., & Cohen, J. I. (1998). Expression of a uracil dna glycosylase (ung) inhibitor in mammalian cells : varicella-zoster virus can replicate in vitro in the absence of detectable ung activity. *Virology*, **251** (2), 393–401.
- Reuven, N. B., Staire, A. E., Myers, R. S., & Weller, S. K. (2003). The herpes simplex virus type 1 alkaline nuclease and single-stranded dna binding protein mediate strand exchange in vitro. *J Virol*, **77** (13), 7425–7433.
- Reuven, N. B., Willcox, S., Griffith, J. D., & Weller, S. K. (2004). Catalysis of strand exchange by the hsv1 ul12 and icp8 proteins : potent icp8 recombinase activity is revealed upon resection of dsdna substrate by nuclease. *J Mol Biol*, **342** (1), 57–71.
- Rist, M. (2003). *2-Aminopurine and Pyrenylethyne-Modified Uridine as Fluorescent Probes for the Investigation of Nucleic Acids and Their Interactions*. PhD thesis Universität München.
- Robertson, E. S., ed (2005). *Epstein-Barr Virus*. Caister Academic Press.
- Rousseau, J. (a). Cours de cristallographie de l'université du maine, le mans. <http://www.univ-lemans.fr/enseignements/physique/02/cristallo/planreti.html>.
- Sandigursky, M., Faje, A., & Franklin, W. A. (2001). Characterization of the full length uracil-dna glycosylase in the extreme thermophile *Thermotoga maritima*. *Mutat Res*, **485** (3), 187–195.
- Savva, R., McAuley-Hecht, K., Brown, T., & Pearl, L. (1995). The structural basis of specific base-excision repair by uracil-dna glycosylase. *Nature*, **373** (6514), 487–493.
- Schepers, A., Pich, D., & Hammerschmidt, W. (1993). A transcription factor with homology to the ap-1 family links rna transcription and dna replication in the lytic cycle of epstein-barr virus. *EMBO J*, **12** (10), 3921–3929.

- Schwenn, M. R., Blattner, S. R., Lynch, E., & Weinstein, H. J. (1991). Hic-com : a 2-month intensive chemotherapy regimen for children with stage iii and iv burkitt's lymphoma and b-cell acute lymphoblastic leukemia. *J Clin Oncol*, **9** (1), 133–138.
- Shannon-Lowe, C. D., Neuhierl, B., Baldwin, G., Rickinson, A. B., & Delecluse, H.-J. (2006). Resting b cells as a transfer vehicle for epstein-barr virus infection of epithelial cells. *Proc Natl Acad Sci U S A*, **103** (18), 7065–7070.
- Shaw, R. W., Feller, J. A., & Bloom, L. B. (2004). Contribution of a conserved phenylalanine residue to the activity of escherichia coli uracil dna glycosylase. *DNA Repair (Amst)*, **3** (10), 1273–1283.
- Shikazono, N., Pearson, C., O'Neill, P., & Thacker, J. (2006). The roles of specific glycosylases in determining the mutagenic consequences of clustered dna base damage. *Nucleic Acids Res*, **34** (13), 3722–3730.
- Sinclair, A. J., Palmero, I., Peters, G., & Farrell, P. J. (1994). Ebna-2 and ebna-lp cooperate to cause g0 to g1 transition during immortalization of resting human b lymphocytes by epstein-barr virus. *EMBO J*, **13** (14), 3321–3328.
- Slupphaug, G., Mol, C., Kavli, B., Arvai, A., Krokan, H., & Tainer, J. (1996). A nucleotide-flipping mechanism from the structure of human uracil-dna glycosylase bound to dna. *Nature*, **384** (6604), 87–92.
- Staege, M. S., Lee, S. P., Frisan, T., Mautner, J., Scholz, S., Pajic, A., Rickinson, A. B., Masucci, M. G., Polack, A., & Bornkamm, G. W. (2002). Myc overexpression imposes a nonimmunogenic phenotype on epstein-barr virus-infected b cells. *Proc Natl Acad Sci U S A*, **99** (7), 4550–4555.
- Strauss, J. H. & Strauss, E. G. (2002). *Viruses and Human Disease*. James H. Strauss and Ellen G. Strauss.
- Strockbine, L. D., Cohen, J. I., Farrah, T., Lyman, S. D., Wagener, F., DuBose, R. F., Armitage, R. J., & Spriggs, M. K. (1998). The epstein-barr virus barf1 gene encodes a novel, soluble colony-stimulating factor-1 receptor. *J Virol*, **72** (5), 4015–4021.

- Stuart, D., Jones, Y., Wilson, K. S., & Daenke, S. (2006). Spine, structural proteomics in europe the best of both worlds. *Acta Crystallogr D Biol Crystallogr*, **62** (Pt 10), perface.
- Studebaker, A. W., Ariza, M. E., & Williams, M. V. (2005). Depletion of uracil-dna glycosylase activity is associated with decreased cell proliferation. *Biochem Biophys Res Commun*, **334** (2), 509–515.
- Studebaker, A. W., Balendiran, G. K., & Williams, M. V. (2001). The herpesvirus encoded dutpase as a potential chemotherapeutic target. *Curr Protein Pept Sci*, **2** (4), 371–379.
- Szilágyi, A. (a). Emanim (a program to visualize electromagnetic waves). EMANIM (a program to visualize electromagnetic waves), <http://www.enzim.hu/szia/cddemo/edemo0.htm>.
- Takada, K. (1984). Cross-linking of cell surface immunoglobulins induces epstein-barr virus in burkitt lymphoma lines. *Int J Cancer*, **33** (1), 27–32.
- Talpaert-Borlé, M., Campagnari, F., & Creissen, D. M. (1982). Properties of purified uracil-dna glycosylase from calf thymus. an in vitro study using synthetic dna-like substrates. *J Biol Chem*, **257** (3), 1208–1214.
- Tarbouriech, N., Buisson, M., Géoui, T., S, D., Cusack, S., & Burmeister, W. P. (2006). Structural genomics of the epstein-barr virus. *Acta Crystallographica Section D*, **62**.
- Tarbouriech, N., Buisson, M., Seigneurin, J. M., Cusack, S., & Burmeister, W. P. (2005). The monomeric dutpase from epstein-barr virus mimics trimeric dutpases. *Structure*, **13** (9), 1299–1310.
- Tarbouriech, N., Ruggiero, F., de Turenne-Tessier, M., Ooka, T., & Burmeister, W. P. (2006). Structure of the epstein-barr virus oncogene barf1. *J Mol Biol*, **359** (3), 667–678.
- Thomas, R., Macsween, K. F., McAulay, K., Clutterbuck, D., Anderson, R., Reid, S., Higgins, C. D., Swerdlow, A. J., Harrison, N., Williams, H., & Crawford, D. H. (2006). Evidence of shared epstein-barr viral isolates between sexual partners, and low level ebv in genital secretions. *J Med Virol*, **78** (9), 1204–1209.

- Thompson, J., Higgins, D., & Gibson, T. (1994). Clustal w : improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res*, **22** (22), 4673–4680.
- Thorley-Lawson, D. A. (2005). Ebv the prototypical human tumor virus—just how bad is it? *J Allergy Clin Immunol*, **116** (2), 251–261.
- Tinkelenberg, B. A., Hansbury, M. J., & Ladner, R. D. (2002). dutpase and uracil-dna glycosylase are central modulators of antifolate toxicity in *saccharomyces cerevisiae*. *Cancer Res*, **62** (17), 4909–4915.
- Tomkinson, B., Robertson, E., & Kieff, E. (1993). Epstein-barr virus nuclear proteins ebna-3a and ebna-3c are essential for b-lymphocyte growth transformation. *J Virol*, **67** (4), 2014–2025.
- Tovey, M. G., Lenoir, G., & Begon-Lours, J. (1978). Activation of latent epstein-barr virus by antibody to human igm. *Nature*, **276** (5685), 270–272.
- Tsurumi, T., Daikoku, T., Kurachi, R., & Nishiyama, Y. (1993). Functional interaction between epstein-barr virus dna polymerase catalytic subunit and its accessory subunit in vitro. *J Virol*, **67** (12), 7648–7653.
- Turcotte, S., Letellier, J., & Lippé, R. (2005). Herpes simplex virus type 1 capsids transit by the trans-golgi network, where viral glycoproteins accumulate independently of capsid egress. *J Virol*, **79** (14), 8847–8860.
- Turk, S. M., Jiang, R., Chesnokova, L. S., & Hutt-Fletcher, L. M. (2006). Antibodies to gp350/220 enhance the ability of epstein-barr virus to infect epithelial cells. *J Virol*, **80** (19), 9628–9633.
- Tye, B. K., Nyman, P. O., Lehman, I. R., Hochhauser, S., & Weiss, B. (1977). Transient accumulation of okazaki fragments as a result of uracil incorporation into nascent dna. *Proc Natl Acad Sci U S A*, **74** (1), 154–157.
- Vagin, A. & Teplyakov, A. (1997). Molrep : an automated program for molecular replacement. *Journal of Applied Crystallography*, **30** (6), 1022–1025.

- van den Bosch, C. A. (2004). Is endemic burkitt's lymphoma an alliance between three infections and a tumour promoter? *Lancet Oncol*, **5** (12), 738–746.
- Van Triest, B., Pinedo, H. M., Giaccone, G., & Peters, G. J. (2000). Downstream molecular determinants of response to 5-fluorouracil and antifolate thymidylate synthase inhibitors. *Ann Oncol*, **11** (4), 385–391.
- Verdine, G. L. & Bruner, S. D. (1997). How do dna repair proteins locate damaged bases in the genome? *Chem Biol*, **4** (5), 329–334.
- Verma, S., Bajaj, B., Ca, i. Q., Si, H., Seelhammer, T., & Robertson, E. (2006). Latency associated nuclear antigen of kshv recruits uracil dna glycosylase 2 at the terminal repeats and is important for latent persistence of the virus. *J Virol*, . J.
- Walling, D. M., Flaitz, C. M., & Nichols, C. M. (2003). Epstein-barr virus replication in oral hairy leukoplakia : response, persistence, and resistance to treatment with valacyclovir. *J Infect Dis*, **188** (6), 883–890.
- Wang, F., Gregory, C., Sample, C., Rowe, M., Liebowitz, D., Murray, R., Rickinson, A., & Kieff, E. (1990). Epstein-barr virus latent membrane protein (lmp1) and nuclear proteins 2 and 3c are effectors of phenotypic changes in b lymphocytes : Ebna-2 and lmp1 cooperatively induce cd23. *J Virol*, **64** (5), 2309–2318.
- Wang, Z. G., Smith, D. G., & Mosbaugh, D. W. (1991). Overproduction and characterization of the uracil-dna glycosylase inhibitor of bacteriophage pbs2. *Gene*, **99** (1), 31–37.
- Wathen, M. W. (2002). Non-nucleoside inhibitors of herpesviruses. *Rev Med Virol*, **12** (3), 167–178.
- Wei, W. I. & Sham, J. S. (2005). Nasopharyngeal carcinoma. *Lancet*, **365** (9476), 2041–2054.
- Weiss, S. (2001). Global indicators of x-ray data quality. *J. Appl. Cryst.* **34**, 130–135.
- Whittle, P. J. & Blundell, T. L. (1994). Protein structure-based drug design. *Annu Rev Biophys Biomol Struct*, **23**, 349–375.

- Wibley, J. E., Waters, T. R., Haushalter, K., Verdine, G. L., & Pearl, L. H. (2003). Structure and specificity of the vertebrate anti-mutator uracil-dna glycosylase smug1. *Mol Cell*, **11** (6), 1647–1659.
- Wilson, D. M. r. & Thompson, L. H. (1997). Life without dna repair. *Proc Natl Acad Sci U S A*, **94** (24), 12754–12757. Comment.
- Wist, E., Unhjem, O., & Krokan, H. (1978). Accumulation of small fragments of dna in isolated hela cell nuclei due to transient incorporation of dump. *Biochim Biophys Acta*, **520** (2), 253–270.
- Wong, I., Lundquist, A. J., Bernards, A. S., & Mosbaugh, D. W. (2002). Presteady-state analysis of a single catalytic turnover by escherichia coli uracil-dna glycosylase reveals a "pinch-pull-push" mechanism. *J Biol Chem*, **277** (22), 19424–19432.
- Wood, R. D., Mitchell, M., & Lindahl, T. (2005). Human dna repair genes, 2005. *Mutat Res*, **577** (1-2), 275–283.
- Wood, R. D., Mitchell, M., Sgouros, J., & Lindahl, T. (2001). Human dna repair genes. *Science*, **291** (5507), 1284–1289.
- Wu, L., Borza, C. M., & Hutt-Fletcher, L. M. (2005). Mutations of epstein-barr virus gh that are differentially able to support fusion with b cells or epithelial cells. *J Virol*, **79** (17), 10923–10930.
- Xiao, G., Tordova, M., Jagadeesh, J., Drohat, A. C., Stivers, J. T., & Gilliland, G. L. (1999). Crystal structure of escherichia coli uracil dna glycosylase and its complexes with uracil and glycerol : structure and glycosylase mechanism revisited. *Proteins*, **35** (1), 13–24.
- Young, L. S., Lau, R., Rowe, M., Niedobitek, G., Packham, G., Shanahan, F., Rowe, D. T., Greenspan, D., Greenspan, J. S., & Rickinson, A. B. (1991). Differentiation-associated expression of the epstein-barr virus bzlfl1 transactivator protein in oral hairy leukoplakia. *J Virol*, **65** (6), 2868–2874.
- Young, L. S. & Rickinson, A. B. (2004). Epstein-barr virus : 40 years on. *Nat Rev Cancer*, **4** (10), 757–768.

- Zharkov, D. O., Golan, G., Gilboa, R., Fernandes, A. S., Gerchman, S. E., Kycia, J. H., Rieger, R. A., Grollman, A. P., & Shoham, G. (2002). Structural analysis of an escherichia coli endonuclease viii covalent reaction intermediate. *EMBO J*, **21** (4), 789–800.
- Zharkov, D. O. & Grollman, A. P. (2005). The dna trackwalkers : principles of lesion search and recognition by dna glycosylases. *Mutat Res*, **577** (1-2), 24–54.
- zur Hausen, H. (2005). *The Early days of Epstein-Barr Virus Research : The Henle Years*. Caister Academic Press.

Résumé:

Le virus d'Epstein-Barr (EBV) est un γ -herpesvirus humain. Il est responsable de maladies telles que la mononucléose-infectieuse et il est associé à de nombreux carcinomes et syndromes immunoprolifératifs. A la différence d'autres herpesvirus tel que herpes-simplex, il n'existe actuellement aucun médicament efficace contre EBV. Le génome d'EBV contient 86 cadres de lecture soit autant de cibles potentielles pour une approche de génomique structurale, permettant le développement rationnel de nouvelles thérapies. La première série de cibles sur laquelle nous avons travaillé code pour 23 protéines. De façon inattendue, le principal problème rencontré lors de ce projet fut la faible expression ainsi que l'insolubilité d'une grande partie des cibles. Parmi les 8 protéines solubles, 5 furent cristallisées à l'issue de quoi, 4 structures furent résolues. La clef de la réussite de ce projet fut un traitement individuel de chaque cible plutôt que l'utilisation de protocoles standards. La partie centrale de ce travail porte sur l'Uracile-ADN Glycosylase (UNG). C'est une enzyme de réparation de l'ADN. Il nous fut impossible d'obtenir des cristaux de la protéine seule, mais, grâce à la formation d'un complexe avec une protéine inhibitrice produite par le phage PBS-2, nous obtînmes des cristaux diffractant à 2.3 Å. La structure de ce complexe nous permis d'expliquer l'organisation de la « boucle leucine », un domaine du site actif, qui comporte une insertion de 7 résidus chez tous les γ -herpesvirus. Malgré cette différence, les constantes catalytiques de l'UNG d'EBV sont proches de celles des autres UNGs ce qui suggère un mécanisme similaire d'interaction avec l'ADN. Le travail sur d'autres cibles et les difficultés qui leur sont inhérentes est également abordé (notamment l'Alcaline Exonucléase).

Summary:

Epstein-Barr Virus (EBV) is a Human γ -herpesvirus. It is the causative agent of diseases such as Infectious Mononucleosis and it is associated with several immunoproliferative syndromes. Unlike other herpesviruses such as herpes-simplex, there is no treatment against EBV. EBV genome features 86 open-reading frames which are potential targets for a structural genomics approach and for the rational design of new drugs. The first round of selected targets encode for 23 proteins. The main problem we encountered was, unexpectedly, the poor expression and solubility of the majority of the selected targets. Among the 8 soluble proteins we obtained, 5 were crystallized and 4 structures were solved. The main success factor was an individual treatment of each target rather than the use of standard protocols. The Uracil-DNA Glycosylase (UNG) is a DNA repair enzyme. It was impossible to obtain crystals of the protein alone, but crystals diffracting to 2.3 Å grew upon the formation of a complex with a phage-encoded inhibitory protein. The analysis of the structure of this complex shows that the « leucin-loop », a major domain of the active site, features a 7 residue insertion which is common to all γ -herpesvirus. Despite this difference, catalytic constants are similar to other organisms UNGs, suggesting a similar interaction with DNA. Current work and problems associated with other targets is also reported (in particular on the Alkaline Exonuclease).