THÈSE

Présentée à l'UFR des Sciences de la Vie et de la Terre de l'Université Louis Pasteur à Strasbourg pour obtenir le grade de

DOCTEUR DE L'UNIVERSITÉ STRASBOURG I UNIVERSITÉ LOUIS PASTEUR

Discipline : Aspects moléculaires et cellulaires de la Biologie Option : Biologie Structurale

> par Yves NOMINÉ

Caractérisation biophysique de la qualité des protéines de fusion : application à l'étude structurale et fonctionnelle de l'oncoprotéine virale E6.

soutenue le 16 décembre 2002 devant la Commission d'Examen :

Jean CAVARELLI	Rapporteur
Analisa PASTORE	Rapporteur
François PENIN	Rapporteur
Bernard ROQUES	Examinateur
Bruno KIEFFER	Co-directeur de thèse
Gilles TRAVE	Directeur de thèse

A mes amis

A mon frère

A mes parents

A ma nouvelle famille ...

« Quand les grimpeurs observent de loin une montagne, tout est obstacle ; c'est en avançant qu'ils trouvent des passages. Mais ils n'avanceraient point s'ils n'espéraient point.] ... [Essayer avec l'idée que la route est barrée, ce n'est pas essayer. Décider d'avance que les choses feront obstacle au vouloir, ce n'est pas vouloir. Aussi voit-on des inventeurs, explorateurs et autres réformateurs qui ne croient pas à ce barrage imaginaire que ferait la montagne de loin. Et cette vertu d'essayer aussitôt et devant soi, est bien l'espérance. »

Alain « Les passions et la Sagesse ». Pg 187.

Je dédie ce manuscrit et ce travail à Jean-François Lefèvre. Jean-François m'aura exprimé immédiatement sa confiance en m'accueillant dans son laboratoire. Il m'aura soutenu et encouragé jusqu'à la fin. Je le laisse juge de la progression de ce travail, et ose espérer qu'il ne le regrette pas.

Je voudrais exprimer toute ma gratitude à mon directeur de thèse, Gilles Travé, ainsi qu'à mon codirecteur, Bruno Kieffer, qui ont su tous deux me faire partager leurs larges connaissances et leur passion pour la recherche. Je suis très reconnaissant à Gilles de m'avoir intégré dès le début au projet E6. Je lui dois aujourd'hui entre autre toutes mes connaissances des techniques de paillasse et de biochimie, et bien plus encore. J'ai par ailleurs beaucoup apprécié qu'il me laisse libre de m'engager dans de nouvelles directions de recherche. Je n'oublierai jamais sa gentillesse, sa patience et sa personnalité emplie d'humanité. Bruno, pour sa part, m'a offert la possibilité de travailler dans une équipe à la pointe de la technologie RMN, en me faisant partager sa vision de la Science « avec les mains » (sans arrière pensée aucune). Durant toutes ces années, son investissement pour moi aura finalement porté ses fruits. Il a su aussi donner la cohésion dont le groupe avait besoin dans les moments difficiles. Je lui suis par ailleurs reconnaissant d'avoir cru en mes capacités d'enseignant. Merci à eux deux.

Je remercie Madame A. Pastore et Messieurs J. Cavarellí, F. Penín, B. Roques de m'avoir fait l'honneur de juger ce travail de thèse.

Je remercie très chaleureusement Tutik Ristriani. Evidemment je pourrai dire que sans elle, ce travail n'aurait pas acquis sa dimension biologique. Mais surtout notre collaboration ne s'est pas arrêtée là : un grand merci pour son amitié et sa bonne humeur de tous les jours. Les critiques justifiées, l'enthousiasme et la richesse des discussions avec Etienne Weiss m'ont aidés à persévérer dans ce sujet difficile. J'ai trouvé en Claude Kédinger une personne toujours à l'écoute, prête à m'appuyer lorsque cela s'avérait nécessaire. Je remercie Danièle Altschuh et Marc Van-Regenmortel de m'avoir laissé accéder au BIAcore 2000, ainsi que Laurence Choulier pour son aide précieuse. J'ai côtoyé pendant mon doctorat deux équipes de recherche : l'immunotechnologie et la RMN. Toutes deux sont indissociables de ce travail de recherche. J'admire, en Claude Ling et en Georges Schwalbach, leur patience, leur disponibilité et leurs conseils de tous les instants. Je n'oublierai jamais les discussions aussi diverses que variées avec Emeric Wasielewski, ni la rigueur sans faille -qualité parmi tant d'autres plus que nécessaire dans un tel sujet- de Sebastian Charbonnier. J'exprime toute ma sympathie à Andrew Atkinson, Marie-Aude Coutouly, Annick Dejaeggere, Baudouin Dillman, Virginie Gervais, Florence Gaudin, Esther Kellenberger, Virginie Lafont, Finton Sirokin (en RMN), ainsi que Géraldine Brandner, Dominique Desplancq, François Déryckère, Georges Orfanoudakis, Magali Lagrange, Murielle Masson, Philip Robinson, Annie-Paule Sibler (en Immuno) et salue la contribution de chacun dans la vie d'un groupe.

Mon travail de thèse m'aura amené à rencontrer, à l'ESBS, bien d'autres personnes à chaque étage : je tiens à les remercier pour ces conseils de tous les jours. La liste est bien trop longue pour être mentionnée ici. Je ferai juste une exception envers Jean-Baptiste pour ces quelques escapades neigeuses, rares moments d'évasion totale. C'est en effet dans les Vosges que j'ai trouvé, trouve et trouverai un terrain de détente, de ressources et de défoulements.

Enfin, je voudrais finir en exprimant mon affection à ma famille : mes parents, mon frère, et plus particulièrement à toi, Cathy, pour ton soutien de tous les instants et, ... pour notre fils Pierrot que tu nous as donné.

ABREVIATIONS

ADN	Acide DésoxyriboNucléïque
ARN	Acide RiboNucléique
ANS	8-Anilino-1-Naphtalène Sulfonate
BSA	Albumine de Sérum de Bovin
DC	Dichroïsme Circulaire
DLS	« Dynamic Light Scattering »
DO	Densité Optique
DTT	DithioTréiTol
EDC	N-ethyl-N'-(dimethylaminopropyl)
	carbodiimide hydrochloride
	Ethylàna Diamina Tatraga actia A aid
LDIA	
FCR	Force Centrifuge Relative
FCR FTIR	Force Centrifuge Relative Infra Rouge par Transformée de
FCR FTIR	Force Centrifuge Relative Infra Rouge par Transformée de Fourier
FCR FTIR GST	Force Centrifuge Relative Infra Rouge par Transformée de Fourier Glutathion S-Transférase
FCR FTIR GST HSQC	Force Centrifuge Relative Infra Rouge par Transformée de Fourier Glutathion S-Transférase « Heteronuclear Simple Quantum
FCR FTIR GST HSQC	Force Centrifuge Relative Infra Rouge par Transformée de Fourier Glutathion S-Transférase « Heteronuclear Simple Quantum Correlation »
FCR FTIR GST HSQC IPTG	Force Centrifuge Relative Infra Rouge par Transformée de Fourier Glutathion S-Transférase « Heteronuclear Simple Quantum Correlation » IsoPropyl-β-D-
FCR FTIR GST HSQC IPTG	Force Centrifuge Relative Infra Rouge par Transformée de Fourier Glutathion S-Transférase « Heteronuclear Simple Quantum Correlation » IsoPropyl-β-D- ThioGalactopyranoside

LB « Luria Broth »

- MBP « Maltose Binding Protein »
- NHS N-hydroxysuccinimide
- Nm Nanomètre (10^{-9} m)
- NOESY Nuclear Overhauser SpectroscopY
 - PAGE « PolyAcrylamide Gel Electrophoresis »
 - PM PhotoMultiplicateur
 - RMN Résonance Magnétique Nucléaire
 - rmsd « root mean square deviation »
 - RPE Résonance Paramagnétique Electronique
 - rpm rotation par minute
 - RU « Response Unit »
 - SDS Sodium Dodécyl Sulfate
 - SLS « Static Light Scattering »
 - SPR Résonance Plasmonique de Surface s seconde
- TOCSY « Total Correlation SpectroscopY »
 - Tris Tris-hydroxyméthylamino-méthane

TABLEAU DES ACIDES AMINÉS

A Ala	Alanine	M Met	Méthionine
C Cys	Cystéine	N Asn	Asparagine
D Asp	Acide Aspartique	P Pro	Proline
E Glu	Acide Glutamique	Q Gln	Glutamine
F Phe	Phenylalanine	R Arg	Arginine
G Gly	Glycine	S Ser	Serine
H His	Histidine	T Thr	Thréonine
I Ile	Isoleucine	V Val	Valine
K Lys	Lysine	W Trp	Tryptophane
L Leu	Leucine	Y Tyr	Tyrosine

PLAN

AVANT PROPOS	1
A. INTRODUCTION	5
CHAPITRE I : Repliements et stabilité des protéines	7
A.I/ Le repliement d'une protéine	7
A.II/ Les différents états d'une protéine exprimée <i>in vivo</i>	8
A.II.1/ L'expression de protéines recombinantes	8
A.II.1.a/ Corps d'inclusion et dégradation	9
◆ La composition des corps d'inclusion	10
♦ La réversibilité des corps d'inclusion	10
A.I.1.b/ Agrégation et protéines repliées correctement	10
 Comment remédier à l'agrégation et donc à la formation des corps d'inclusion 	11
A.I.2/ L'agréga <i>tion</i> in vivo dans une cellule eucaryote	11
A.III/ Les différ <i>ents</i> états d'une protéine en solution	
A.III.1/L'agrégation in vitro.	
A.III.2/ Etats microscopiques et macroscopiques	13
Protéine correctement repliée	13
♦ Protéine mal repliée	14
A IV/ La stabilité d'une protéine	15
A IV 1/ La stabilité thermodynamique d'une protéine native	15
A.IV.1.a/ La stabilité thermodynamique selon la relation de Gibbs	
A.IV.1.b/ La stabilité thermodynamique d'un point de vue biologique	16
A.IV.1.c/ La détermination de la stabilité thermodynamique	18
♦ Stabilité face à une dénaturation thermique	18
 Stabilité face à un agent dénaturant 	18
A.IV.2/ Autres notions incluses dans le terme de stabilité	
CHAPITRE II : Techniques biophysiques	
A V/ Principes généraux de spectroscopies	23
A V 1/L'interaction lumière / matière	23 23
A.V.1.a/ Absorption de lumière	
◆ Absorption de photon et phénomène de fluorescence	
◆ Absorption de photon par une molécule	
A.V.1.b/ Le phénomène de diffusion de lumière	
◆ Traitement physique de la diffusion de lumière	27

• Diffusion diffraction et milieu isotrope	27
A.V.1.c/ Absorption. diffusion et fluorescence.	
A.V.2/ Les montages utilisés en spectrométrie	
A.VI/ Aspects expérimentaux de spectroscopies appliquées à l'étude de protéines	
A.VI.1/L'absorption de lumière ou la mesure de densité optique	
A.VI.1.a/ Absorption de la référence.	
A.VI.1.b/ Absorption par la molécule d'intérêt.	

◆ Intérêt des mesures d'absorption en biologie	
A.VI.2/ Spectroscopie de fluorescence	34
◆ Intégrité du repliement d'une protéine	
• Mesure de dénaturation par fluorescence	
♦ Intensité de fluorescence et concentration	
A.VI.3/ Le Dichroïsme Circulaire (DC)	37
A.VI.4/ La résonance Magnétique Nucléaire (RMN)	40
A.VI.5/ Les spectroscopies infrarouge et Raman	44
A.VI.5.a/ La spectroscopie infrarouge	44
A.VI.1.b/ La spectroscopie Raman	46
A.VI.6/ La diffusion de lumière	48
A.VI.6.a/ Diffusion Statique de Lumière	48
• Diffusion de lumière et tailles des particules en solution	
 Masse, rayon de giration et diffusion de lumière 	
♦ Masse moyenne pondérée	
 Précautions quant à l'utilisation de la diffusion de lumière. 	
♦ Apports de la SLS en mesures directes	51
• Exemple : mesure de l'agrégation dans le temps	51
A.VI.6.b/ La diffusion dynamique de lumière	
A.VI.7/ La Résonance Plasmonique de Surface (technologie BIAcore)	53
♦ Principes des mesures en temps réel	53
♦ Immobilisation du ligand à la surface	56
◆ Analyse des courbes d'association et de dissociation	56
♦ Les principaux modèles et leur spécificité	
♦ Limite du transport de masse	
♦ Analyses quantitative et qualitative par SPR	58
A.VII/ Tableau récapitulatif des techniques utilisées	59

CHAPITRE III : E6, les virus à papillomes humains	61
A.VIII/ Les virus à papillome humains (HPV) et les cancers	61
♦ Les souches bénignes et malignes de HPV	61
♦ Génome et expression des différentes protéines de HPV	62
◆ Spécificité des HPV dans les cancers des épithéliums	62
A.IX/ Rôle joué par E6 et E7 dans la réplication des virus dits à « haut risque »	63
A.X/ Etat des connaissances sur la protéine E6 du virus HPV 16	65
A.X.1/ Séquence de E6	65
A.X.2/ E6 dans la cellule	67
A.X.3/ La difficulté d'obtenir des données biologiques pour E6	68

B. MATERIEL ET METHODES

(n
6	ч
	1

B.I/ Techniques générales d'expression et de purification de protéines	71
B.I.1/ Milieux de culture	71
♦ Milieu de culture LB (Luria Broth)	71
• Milieu minimum M9 ^{15}N	71
B.I.2/ Expressions bactériennes	71
♦ Souches bactériennes	71
♦ Transformation par électroporation	

♦ Les conditions de préculture	72
• Les conditions de culture	72
B.I.3/ Purifications de protéines	73
♦ Lyse des bactéries	73
 Purification de la protéine fusionnée à MBP : la colonne amylose 	73
 Séparation des molécules agrégées de celles monomériques : l'ultracentrifugation 	74
Coupure de la fusion ; séparation de MBP et de la protéine E6, ou d'un de ses domaines	74
B.I.4/ Concentrations de protéines	74
B.I.5/ Méthode de séparation de la protéine agrégée de celle monodiperse	75
B.II/ Techniques générales de spectroscopies (autres que la RMN)	77
B.II.1/ Spectrofluorimètre	77
♦ Mesures de fluorescence	78
♦ Mesures de diffusion de lumière	78
 Mesures du rapport d'agrégation 	78
B.II.2/ Dichroïsme circulaire	79
B.II.3/ Dénaturation de protéines	80
♦ Aspects généraux	80
♦ Cas de la protéine en quantite suffisante	81
◆ Cas de la protéine en quantité limitée	81
◆ Traitement des courbes de dénaturation	82
B.III/ Techniques générales de RMN	82
B.III.1/ Acquisition des spectres RMN	82
B.III.2/ Calculs de structure	83
◆ Collecte des contraintes de distances	83
◆ Procédure itérative sur le SA	85
◆ Un affinement des structures basé sur le protocole SA	87
B.IV/ Techniques liées à l'instrumentation BIAcore.	87
B.IV.1/ Préparation des réactifs	87
◆ Le ligand (molécule accrochée à la surface)	87
◆ L'analyte (molécule en solution)	88
B.IV.2/ Immobilisation des ligands à la surface	88
B.IV.3/ Conditions de régénérations	89
B.IV.4/ Acquisition des données expérimentales	90
B.IV.5/ Traitement des données expérimentales	90
B.V/ Ajustement non linéaire de courbes experimentales sous MatLab	91
♦ Ajustement non linéaire	91
◆ Etude numérique de l'interaction E6 / ADN cruciforme	91

C. RESULTATS ET DISCUSSIONS	 93

PARTIE I : Caractérisation biophysique de l'oncoprotéine E6 et de ses domaines	95
C.I.1/ Publication 1 : Formation of soluble inclusion bodies by HPV16 E6 oncoprotein fused to MBP	97
C.I.2/ Publication 2 : A strategy for optimising the monodispersity of fusion proteins: application to purification	
of HVP16 E6 oncorprotein	109
C.I.3/ Publication 3 : Biophysical characterisation of E6 C-terminal DNA-binding domain	119
Evolution de la qualité du domaine E6-C 4C/4S au cours des purifications	135

PARTIE II : Etude structurale du domaine C-terminal de E6	137
C.II.1/ Publication 4 : 1H, 15N resonances assignment and secondary structure of the C-terminal domain of E6 C.II.2/ Calculs de la structure du domaine C-terminal de E6	139 145
PARTIE III : Identification et caractérisation de l'interaction entre E6 et l'ADN	153
C.III.1/ Publication 5 : HPV oncoprotein E6 is a structure-dependant DNA-binding protein that recognizes four-way DNA junction	155
C.III.2/ Publication 6 : Specific recognition of four-way DNA junctions by the C-terminal zinc-binding domain of EC C.III.3/ Publication 7 : Protein mutagenesis with monodispersity-based quality probing: selective inactivation of p5	5.171 3
degradation and DNA binding properties of HPV E6 oncoprotein	183
C.III.4/ Etude de l'interaction E6 / ADN cruciforme par BIAcore	195
C.III.4.a/ Conception de l'expérience	197
C.III.4.b/ Détermination de la concentration active de protéine E6	197
C.III.4.c/ Analyse de la phase de dissociation	198
C.III.4.d/ Analyse globale des phases d'association et de dissociation	198
C.III.4.e/ Effet des conditions experimentales	200
D. CONCLUSIONS ET PERSPECTIVES	. 203
E. ANNEXES	. 209

F. BIBLIOGRAPHIE	 211
<u>F. BIBLIOGRAPHIE</u>	 211

AVANT-PROPOS

Repliement et stabilité seront des fils conducteurs retrouvés tout au long de ce manuscrit. Ces termes, très usités en biologie, ont des sens très variés que nous tenterons de détailler. Ces principes, nécessaires pour une meilleure compréhension des résultats décrits dans ce manuscrit, seront donc détaillés dans le premier chapitre de l'introduction bibliographique dans le cas de protéines globulaires, exception faite des protéines imbriquées dans de grands complexes protéiques (actine, tubuline, facteurs de transcription, ...). Celles-ci constituent en effet un cas particulier de protéines pouvant s'agréger même sous leur forme native. Le second chapitre reprendra les bases des techniques spectroscopiques utilisées dans ce travail, et introduira le principe de Résonance Plasmonique de Surface. Adaptée à la biologie depuis 1992, cette technologie permet d'entreprendre l'étude de cinétiques d'interactions en temps réel sans marquage des molécules. Le but de ce chapitre n'est pas d'établir une liste exhaustive des techniques utilisées dans l'étude des protéines, mais plutôt d'amener le lecteur à considérer ces techniques biophysiques dans l'étude de la qualité des objets biologiques étudiés.

Lorsque j'ai rejoint le laboratoire de RMN en 1998, Gilles Travé travaillait à l'établissement d'un protocole d'expression et de purification de l'oncoprotéine E6 issue du papillomavirus humain de type 16. Le troisième chapitre introduira le contexte général de cette oncoprotéine impliquée dans le cancer du col de l'utérus. La séquence de 158 résidus de cette protéine est connue depuis 1985 (Schwarz, 1985). Pourtant, aucune étude structurale n'était disponible. Cet état de fait était lié à l'impossibilité d'exprimer et de purifier la protéine seule. Afin de contourner ce problème, la majorité des études biologiques utilisaient la protéine E6 fusionnée à une porteuse du type MBP (Maltose Binding Protein) ou GST (Glutathion-S-Transférase), puis sur-exprimée dans des systèmes bactériens. Ces systèmes avaient en effet été décrits comme un moyen efficace pour solubiliser des protéines récalcitrantes (Kapust & Waugh, 1999) à la manière d'une chaperonine (Fox, 2001). Dans le cas de l'oncoprotéine E6, ces systèmes permettaient effectivement d'augmenter les quantités de protéines produites. Pourtant, il était toujours impossible d'isoler la protéine E6.

Dans ce contexte, l'étude structurale de E6 représentait donc un objectif ambitieux : d'une part, celui d'obtenir la protéine pure sous une forme biologiquement active permettant sa caractérisation biophysique ; d'autre part, celle d'obtenir un échantillon RMN permettant de déterminer une structure en solution. Les étapes majeures de ce processus sont la détermination des conditions d'expression et de purification permettant d'aboutir à une concentration de protéine de 1mM, mais aussi celle d'un marquage isotopique.

Les résultats des premiers tests d'expression des protéines de fusion MBP-E6 nous ont conduits à étudier de façon systématique le comportement de ces molécules afin de mieux comprendre les mécanismes moléculaires responsables des comportements « anormaux » observés : la précipitation quasi-instantannée de E6 après coupure protéolytique spécifique de la fusion. L'utilisation systématique des techniques biophysiques à chaque étape du protocole d'expression ou de purification nous a permis de caractériser l'état de la protéine en solution, en particulier son repliement et la taille des particules en solution. La meilleure connaissance de ces fusions nous a été utile pour concevoir des méthodes de séparation des deux types de fusions coexistants en solution (fusions monodisperses et corps d'inclusion solubles). Ce travail a permis d'isoler la protéine E6 sauvage de HPV16 en solution. Pourtant, cette molécule ainsi purifiée est sensible à des processus d'oxydation, conduisant à sa précipitation. La mutation en sérine de 6 résidus cystéines non conservés dans les alignements de séquence de E6 de différentes souches virales (appelé E6 6C/6S) a permis d'augmenter sensiblement la resistance de la protéine face à l'oxydation. Parallèlement à cette étape déterminante, un contrôle strict des conditions d'oxydation nous a permis d'obtenir la protéine E6 recombinante sous une forme repliée, isolée, thermodynamiquement stable, active et monomérique en solution. Ces résultats seront détaillés dans les publications 1 et 2 incluses à ce manuscrit.

Au cours des essais de purification de la protéine E6 6C/6S, nous avons pu identifier 2 domaines stables de taille équivalente. La manipulation en solution du domaine C-terminal, plus aisée que celle de la protéine entière, nous a permis d'amener ce domaine à une concentration de l'ordre du milli-molaire, conduisant à un échantillon RMN. La publication 3 présentera ces résultats. Obtenus courant janvier 2001, les premiers spectres homonucléaires bidimensionnels du domaine C-terminal de E6 (E6-C 4C/4S₈₇₋₁₅₈) à 1 mM (pH 6.8 et 5°C) ont clairement montré que le domaine est replié et que l'étude de sa structure tri-dimensionnelle était envisageable. Cependant, ce résultat n'a pu être obtenu qu'avec un échantillon marqué uniformément à l'azote 15. Cette difficulté etait

essentiellement liée à un taux important de recouvrement des fréquences, mais aussi et surtout à la présence d'équilibre conformationnels conduisant à des élargissements de raies. Les résultats de l'étude RMN seront présentés dans la seconde partie des résultats. La structure tri-dimensionnelle ayant été obtenue au cours des dernières semaines de mon travail, ces résultats n'ont pas fait l'objet d'une analyse détailllée.

L'oncoprotéine E6 était suspectée d'interagir directement avec l'ADN. Cette idée se basait sur le fait que : i/ E6 active la transcription du gêne de la sous-unité catalytique de la télomérase (Gewin & Galloway, 2001), ii/ E6 est une protéine très basique contenant deux domaines de liaison au zinc (Cole & Danos, 1987), iii/ E6 active ou réprime la transcription de plusieurs promoteurs cellulaires et viraux (Sedman, 1991). Au laboratoire, l'obtention de la protéine sous une forme soluble et monomérique a permis de mettre en évidence une interaction spécifique entre la protéine E6 et un ADN particulier, l'ADN cruciforme. Par ailleurs, nous avons montré que le domaine C-terminal de E6 correspond à la région de E6 nécessaire et suffisante à cette interaction. Ma contribution à cette partie a consisté en la mise au point du protocole de purification de la protéine et de ses domaines (mutés ou non), la vérification de la qualité des molécules utilisées pour ces études d'interaction, et l'élaboration de modèles mathématiques de l'interaction. J'ai également étudié la cinétique de cette interaction par BIAcore. Les résultats relatifs à cette interaction seront présentés dans la troisième partie des résultats.

Ce travail a été réalisé au sein de l'unité Biotechnologie des Interactions Macromoléculaires (UMR 7100 CNRS-ULP, ESBS) dans l'équipe structure-fonction et immunologie de l'oncoprotéine E6, en étroite collaboration avec le groupe de Résonance Magnétique Nucléaire (UMR 7104 CNRS-ULP-INSERM, ESBS).

INTRODUCTION

6

CHAPITRE I : REPLIEMENTS ET STABILITE DES PROTEINES

A.I/ Le repliement d'une protéine

Les protéines interviennent dans tous les processus du vivant, du décodage du matériel génétique au niveau cellulaire à la transmission d'information entre organes chez les êtres complexes. Les milliers de protéines ont chacune des fonctions et des spécificités très précises, bien que la chaîne carbonée principale ne puisse adopter qu'un nombre restreint de structures secondaires. En 1960, J. Kendrew qui a déterminé par cristallographie la première structure d'une protéine (la myoglobine) estime que « les caractéristiques les plus remarquables de la protéine sont probablement sa complexité et son absence de symétrie » (Kendrew, 1960). En fait, cette spécificité est due à l'arrangement dans l'espace des différentes structures secondaires entre elles, mais aussi grâce à la diversité des chaînes latérales des résidus impliqués dans ces structures secondaires. Ceci permet, en rapprochant des régions parfois très éloignées dans la séquence primaire, d'obtenir des cœurs hydrophobes, des régions chargées positivement (ou négativement) ..., conférant ainsi aux protéines leurs propriétés caractéristiques et spécifiques à chacune d'elle.

Cet arrangement, appelé *conformation* de la molécule, dépend de plusieurs interactions : dipolaire, ionique, hydrophobe et des liaisons hydrogène. Pour une molécule comportant jusqu'à plusieurs milliers ou dizaines de milliers d'atomes, cette conformation n'est évidemment pas unique. Cependant, les lois de la thermodynamique, ainsi que la cinétique de repliement, impliquent que seule une conformation unique sera favorisée par rapport à toutes les autres.

Les protéines de grande taille s'organisent en domaines structuraux, capables d'adopter une structure correcte indépendamment du reste de la molécule (Janin & Chothia, 1985). Pour qu'un domaine structural se replie de manière autonome, il est nécessaire que sa séquence comporte un minimum d'une cinquantaine de résidus (Doolittle, 1995). En dessous de cette taille, le gain en énergie libre apporté par les interactions favorables à l'état replié (liaisons hydrogènes et ioniques, interactions hydrophobes et de van der Waals) reste inférieur à l'entropie conformationnelle de l'état non replié.

Les ponts disulfure intramoléculaires ou la coordination d'un ou plusieurs ions zinc constituent un moyen séléctionné par la Nature pour stabiliser des petites protéines capables, malgré leur petite taille (entre 10 et 70 acides aminés), d'adopter des structures tridimensionnelles stables et biologiquement actives.

A.II/ Les différents états d'une protéine exprimée in vivo

A.II.1/ L'expression de protéines recombinantes



Fig 1 : les différentes voies possibles lors de l'expression hétérologue de protéines. Les disques ombrés correspondent aux états finaux possibles après expression [adapté de (Carrio & Villaverde, 2002)].

La figure 1 résume les différentes voies possibles lors de l'expression de protéines hétérologues. Après expression dans une bactérie, la chaîne peptidique passe par différents états intermédiaires nécessaires pour atteindre son état natif. A ce stade, plusieurs voies sont ouvertes, chacune d'elle étant caractérisée par la vitesse de formation de l'état final. Les processus de repliement conduisant au repliement natif de la protéine sont caractérisés par une vitesse k_{rep} . Pourtant les protéines non repliées ou mal repliées sont fréquentes *in vivo*. Deux voies s'offrent alors : celle de la dégradation dans le cytoplasme par les protéases –définie par k_{deg} – ou bien celle de la formation d'agrégats –définie par k_{inc} , celle-ci étant privilégiée si la vitesse de dégradation k_{deg} est trop lente comparée à celle de la formation des corps d'inclusion k_{inc} . Betton *et al*. ont mis en évidence cet équilibre entre agrégation et dégradation dans le cas de mutants MBP ayant perdu leur faculté de repliement. Ils ont également noté que cet équilibre se déplaçait vers une plus grande agrégation lorsque le taux de protéines exprimées augmente (Betton, 1998).

Comment définit-on un agrégat ? Un *agrégat* caractérise une agglomération de protéines entre elles. Cependant, contrairement au précipitat, un agrégat peut rester soluble. Pourquoi utiliser ce terme plutôt que celui de multimère ? En fait, l'idée d'agrégat sous-entend, contrairement au cas d'un multimère, une hétérogénéité de taille et une variabilité en fonction du temps. Ceci se traduit généralement par un état polydisperse. A l'inverse, une protéine qui ne s'agrège pas, est donc *monodisperse*, ou tout du moins davantage monodisperse que l'agrégat. En effet, une protéine en solution, composée du nombre minimal d'oligomères définissant sa conformation native –monomère, ..., hexamère ou d'un ordre encore plus élevé– est considérée comme étant dans un état d'énergie le plus favorisé (Garel, 1992).

A.II.1.a/ Corps d'inclusion et dégradation

Les chaînes peptidiques non repliées correctement peuvent se rassembler entre elles autour d'un point de nucléation du fait des interactions hydrophobes (London, 1974). Ce point peut continuer à grossir avec l'apport de protéines mal repliées, jusqu'à former des agrégats contenant plusieurs centaines, voire plusieurs milliers de protéines. Lorsque ces agrégats sont localisés au sein d'une cellule, on les appelle des corps d'inclusion (Marston, 1986). Ceux-ci semblent davantage résistants à la digestion cellulaire que ne le serait une protéine mal repliée libre dans le cytoplasme, ce qui explique d'ailleurs leur accumulation dans les cellules. Ils tendent à être insolubles et métaboliquement stables dans le temps (Enfors, 1992) (Corchero, 1997).

• *la composition des corps d'inclusion*

Ces agrégats sont soit amorphes, soit structurés comme les amyloïdes –généralement sous forme de brin β (Dobson, 1999). De formes très variables, ces particules seraient fortement hydratées (Carrio & Villaverde, 2002). Bien que la formation des agrégats soit souvent considérée comme une précipitation désordonnée et non spécifique des chaînes polypeptidiques, deux équipes au moins (Rajan, 2001 ; Speed, 1996) ont mis en évidence une réelle spécificité de cette agrégation. D'après ces auteurs, les protéines mal voire non repliées ne s'agrègent qu'entre elles, expliquant ainsi la relative pureté des corps d'inclusion. Cette propriété peut être utilisée comme première étape de purification (purification de E7 HPV16 (Suttnar, 1994) ; de p53 (Bell, 2002)), à condition que l'on dispose ensuite d'un protocole efficace de repliement *in vitro*.

• La réversibilité des corps d'inclusion

Le caractère solide de ces corps d'inclusion laisse penser qu'ils persistent dans la cellule aussi longtemps que la cellule vit. Depuis quelques années, cette vision semble évoluer : l'agrégation des protéines *via* les corps d'inclusion pourrait être réversible. Par exemple, les corps d'inclusion de la β -galactosidase sont partiellement protéolysés dans la bactérie *E. coli* lorsque la synthèse de protéines est interrompue (Corchero, 1997).

A.II.1.b/ Agrégation et protéines repliées correctement

Le fait qu'un agrégat de protéines ne s'accumule pas dans une cellule non stressée, et ce en dépit d'une production continue, traduit l'existence d'une machinerie cellulaire de « contrôle qualité » (Wickner, 1999). Les chaperonines constituent l'un de ces éléments, dans le sens où elles n'interagissent qu'avec les protéines mal repliées. Agissant comme un catalyseur, elles facilitent le repliement en protégeant temporairement du milieu cytoplasmique des résidus hydrophobes exposés. Ces chaperonines ne peuvent cependant aider plus qu'une quantité fixée de protéines au repliement (Agashe & Hartl, 2000).

Dans le cas d'une sur-expression recombinante, la quantité de protéines natives reflète une compétition cinétique entre repliement correct et agrégation (Yon, 1996 ; Garel, 1992). Des situations de stress –choc en température, sur-expression d'une protéine– peuvent déséquilibrer cette compétition, et conduire à une agrégation. Ainsi, lors d'une période de stress que constitue la

sur-expression, l'action renforcée de ces chaperonines ne permet pas toujours de « traiter » le surplus de protéines produites dans un temps donné.

• Comment remédier à l'agrégation et donc à la formation des corps d'inclusion

Dans un premier temps, il est nécessaire d'éviter les conditions favorables aux corps d'inclusion comme l'acidification des milieux de culture ou l'environnement oxydo-réducteur dans la bactérie (Yon, 1996). Par ailleurs, l'idée est de favoriser l'état replié plutôt que celui de protéines mal ou non repliées, (i) soit en augmentant la vitesse k_{rep} de formation de protéines correctement repliées, (ii) soit en diminuant la vitesse k_{mal} de formation de protéines mal ou non repliées (Fig. 1).

- i/ L'ajout d'une chaperonine, elle aussi sous le contrôle d'un promoteur inductible, minimise généralement la formation des corps d'inclusion (Georgiou & Valax, 1996), même si des résultats parfois contradictoires ont déjà été rapportés (Thomas & Baneyx, 1996). Cette approche est d'autant plus intéressante que la protéine d'intérêt est induite par un promoteur fort.
- ii/ Au contraire, l'utilisation d'un promoteur faible et / ou d'une température basse (de l'ordre de 20° à 30°C, voire même moins) réduit la formation d'agrégat. Le but est ici de ralentir la production de protéines, en supposant que la bactérie puisse alors replier correctement la faible quantité synthétisée.

Enfin, il semblerait que, même lorsque ces corps d'inclusion se sont déjà formés, l'arrêt de la synthèse de protéine par l'ajout de chloramphénicol dans le milieu de culture, permette à la bactérie *via* les chaperonines, de défaire ces agrégats et de replier ensuite correctement la protéine [(Carrio & Villaverde, 2001] (Fig. 1).

A.II.2/ L'agrégation in vivo dans une cellule eucaryote

Dans les cellules de patients atteints de maladies dégénératives (maladies d'Alzheimer, de Parkinson), il existe un lien étroit entre l'existence de dépôts amyloïdes en quantité anormalement élevée et un défaut de repliement (Dobson, 1999 ; Tran & Miller, 1999). La question reste cependant posée de savoir si ces agrégats sont la cause ou la conséquence de ces maladies. Ces agrégats solides pourraient en effet se former suite à une mutation dans une protéine provoquant une perte de fonctionnalité, ou bien, comme dans le cas de l'encéphalite bovine spongiforme

(maladie associée au prion) ou de la maladie d'Alzheimer, suite à un processus encore mal connu conduisant à une transition conformationnelle en brins β [(Nandi, 1996; Soto, 1995)]. L'équipe du Professeur Nandi PK. avait par ailleurs suggéré que ces amyloïdes n'étaient en fait constitués que d'une seule protéine dans les maladies d'Alzheimer, de Huntington, ou dans le cas du prion : il y aurait donc, au moins dans certains cas, une spécificité d'interaction comme celle retrouvée dans les corps d'inclusion.

Les agrésomes constituent un autre type d'agrégat cellulaire, à rapprocher de ces amyloïdes. Ces derniers, connus depuis de nombreuses années, apparaissent sans aucune action directe sur la cellule ; leur expansion est capable de percer la membrane cellulaire. Au contraire, les agrésomes ont été mis plus récemment en évidence suite à des expériences de surexpression de protéines hétérologues dans des cellules eucaryotes. Ils sont l'une des conséquences de l'agrégation *in vivo*, et correspondraient donc à un équivalent du corps d'inclusion dans une cellule eucaryote. Kopito *et al.* estiment que ces agrésomes existent également dans une cellule eucaryote qui ne soit pas en situation d'expression hétérologue. Ils apparaîtraient lorsque la quantité produite de protéines sensibles à l'agrégation dépasse la capacité cellulaire d'élimination de ces agrégats (Kopito, 2000), traduisant encore ici cette idée d'équilibre.

A.III/ Les différents états d'une protéine en solution.

A.III.1/ L'agrégation in vitro

Plusieurs équipes de chercheurs ont mis en évidence le caractère agrégatif de certaines protéines. Par exemple, Bier & Nord démontrent en 1949 la dépendance au cours du temps, en température et en concentration, de l'agrégation de l'ovalbumine [Doty & Edsall, 1951]. Halwer *et al.* rapportent en 1951 un phénomène d'agrégation de la BSA : ils observent une augmentation de la masse des particules au cours du temps (à l'échelle de plusieurs mois), et remarquent qu'ils peuvent revenir à des valeurs de masse initiale en réalisant une ultracentrifugation préalablement à la mesure [Doty & Edsall, 1951]. Ces chercheurs soulignent justement l'importance de la préparation des solutions pour éviter ces phénomènes d'agrégation, ou tout du moins l'utilisation de techniques telles la diffusion de lumière pour écarter les solutions présentant une agrégation. Ainsi, le problème d'agrégation décrit lors de l'expression de protéines ne semble pas se limiter au cas *in vivo*. Au cours de ma thèse, j'ai été confronté à des problèmes d'agrégation, *in vitro* et *in vivo*. Il apparaît qu'il faut distinguer les agrégats produits *in vivo* au sein de la bactérie que l'on retrouve ensuite pendant la purification, des agrégats formés lors du processus de purification. Bien que les mécanismes de formation puissent être parfois similaires –mauvais repliement d'une protéine–, ce deuxième type d'agrégats n'exclut pas que la protéine ait pu être, pendant un temps, correctement repliée et monodisperse.

A.III.2/ Etats microscopiques et macroscopiques

Nous allons désormais nous concentrer sur les deux états extrêmes d'une protéine unique : repliée ou mal repliée. Le cas de l'état totalement déplié, en conformation « random » ne sera pas discuté ici puisque, au cours de la synthèse d'une protéine, cet état transite très rapidement vers un état intermédiaire de repliement. Les deux états replié et mal replié, que j'appellerai « *microscopiques* », induisent différents états de protéines en solution : précipitat, agrégat soluble ou oligomère minimal. Je nommerai ces états des états « *macroscopiques* » (Fig. 2).

♦ protéine correctement repliée

Une protéine repliée reste théoriquement monodisperse en solution, tout du moins pour les protéines globulaires (flèche épaisse). De manière plus exceptionnelle (flèche fine), certaines protéines bien que repliées au départ, peuvent précipiter, à la suite par exemple, de la formation de ponts disulfures inter-moléculaires entre des cystéines non impliquées dans des interactions intra-moléculaires. De manière encore plus rare (flèche discontinue), des protéines repliées peuvent s'agréger[#]. Certaines conservent d'ailleurs leur repliement dans certaines conditions de tampon : le sulfate d'ammonium par la désolvatation qu'il produit sur les protéines, ou bien la cristallogénèse. Nous reviendrons plus loin sur cet état.

[#] C'est effectivement un événement rare si l'on exclut de ces protéines celles appartenant à des complexes biologiques de grandes tailles et qui, à l'état « naturel » sont toujours protégées du milieu cytoplasmique.



Fig. 2 : les différents états d'une protéine, microscopique ou macroscopique, et les liens entre chacun. L'épaisseur des flèches est fonction de la fréquence, pour une protéine dans un état microscopique donné, d'être dans tel état « macroscopique ».

♦ protéine mal repliée

Il est généralement admis qu'une protéine précipite lorsqu'elle est mal repliée (flèche épaisse) du fait de l'exposition au solvant de résidus hydrophobes. De manière très rare, certaines protéines restent monodisperses bien que mal repliées (flèche discontinue) : c'est le cas pour des molécules telles que le lysozyme, l'apomyoglobine ou la α -lactalbumine dont la transition repliée / mal repliée est réversible. Ces molécules, qui possèdent une activité une fois repliée, ont d'ailleurs été extensivement utilisées comme modèle pour l'étude du repliement des protéines *in vitro* (Privalov, 1996). A l'inverse, une protéine mal repliée peut rester en solution, mais sous la forme d'un agrégat comme dans le cas décrit précédemment de l'ovalbumine (flèche fine).

Dans cette description, on distingue donc deux cas conduisant à la précipitation : soit la protéine ne s'est jamais repliée correctement ; soit la protéine une fois repliée, tend à s'agréger. Ces repliements incomplets, ou mauvais repliements, peuvent par exemple exposer au solvant, pendant un laps de temps trop long, des résidus hydrophobes : ces résidus non polaires vont se regrouper, et donc former des agrégats. De plus, la réactivité des cystéines est un autre exemple d'agrégation ou de précipitation pouvant intervenir même lorsque la protéine est correctement repliée (cas de la BSA

(Maruyama, 2001) ou de manière plus générale (Costantino, 1994)), en particulier dans le cas des cystéines non impliquées dans la formation de ponts disulfure intra-residus, ou dans une liaison au zinc.

A.IV/ La stabilité d'une protéine

A.IV.1/ La stabilité thermodynamique d'une protéine native

A.IV.1.a/ La stabilité thermodynamique selon la relation de Gibbs

Ce terme de stabilité d'un état natif (N) par rapport à l'état dénaturé (D) évoque bien souvent la stabilité thermodynamique, c'est-à-dire les contributions de l'enthalpie ($\Delta H = H_N - H_D$) et de l'entropie ($\Delta S = S_N - S_D$) à l'énergie libre ($\Delta G = G_N - G_D$), selon la relation de Gibbs (à température constante) :

$$\Delta G = \Delta H - T \Delta S \tag{1}$$

La forme native (N) d'une protéine sera dite plus stable que la forme dépliée (D) si l'énergie libre associée à N est inférieure à celle associée à D, autrement dit si $\Delta G < 0$.

L'état correctement replié d'une chaîne peptidique résulte principalement de l'équilibre entre deux forces opposées (Baldwin, 1986) :

- celle associée aux interactions hydrophobes. Afin de minimiser l'énergie de solvatation, la chaîne va adopter une conformation tendant à minimiser les contacts entre l'eau et les groupes hydrophobes, et en parallèle à favoriser ceux avec les groupements hydrophiles. Le repliement à l'intérieur de la molécule des chaînes latérales aliphatiques et aromatiques s'accompagne d'interactions de van der Waals, contribuant elles aussi à la stabilité de l'édifice. De manière générale, toutes les interactions apportent une contribution <u>enthalpique</u> à l'état replié.
- celle associée à la perte <u>d'entropie</u> de conformation. La chaîne peptidique repliée en une conformation unique a perdu la majeure partie de ses degrés de libertés. Il ne subsiste que ceux correspondant aux mouvements internes de rotation et de vibration autour de la conformation moyenne. La relation de Gibbs implique que cette variation d'entropie négative est associée à

une variation positive de l'énergie libre. Cette unique contribution est donc défavorable à l'existence de la forme compacte.

Les interactions électrostatiques (interactions de charges, liaisons hydrogènes), quant à elles, ne jouent qu'un rôle secondaire dans la stabilisation de la structure de la protéine native, comparativement aux interactions hydrophobes. Par contre, leur rôle devient beaucoup plus important lors du processus de repliement en permettant de se rapprocher dans l'espace des charges de signes opposés (Privalov, 1992).

Pour des protéines, le ΔG observé d'une forme repliée est petit par rapport aux deux valeurs des énergies antagonistes. Par exemple dans le cas du lysozyme, ΔG vaut 60 kJ à 25°C (soit 298 K) alors que ΔH et T ΔS valent respectivement 236 kJ/mol et 175 kJ/mol (Baldwin, 1986).

Plus généralement, la valeur de l'énergie de Gibbs pour la stabilisation d'une protéine globulaire est de l'ordre de 50 kJ, soit une énergie moyenne par résidu de l'ordre de 0,5 kJ (dans le cas d'une protéine globulaire de 100 résidus). Cette valeur représente $1/5^{\text{ème}}$ de l'énergie thermique RT à température ambiante (la constante des gaz parfaits R = 8,31 J/mol/Kelvin). La stabilité d'une protéine native ne se limite donc pas aux contributions individuelles de chacun des résidus, mais tient compte de la coopérativité efficace de ces interactions. Cette coopérativité est d'ailleurs augmentée par le caractère dual de l'interaction hydrophobe : attractive à courte distance, répulsive à longue distance.

A.IV.1.b/ La stabilité thermodynamique d'un point de vue biologique

Les structures secondaires, tertiaires et quaternaires des protéines sont le résultat d'un savant équilibre entre des forces attractives (interactions électrostatiques, hydrophobes, liaisons hydrogènes), et répulsives (électrostatiques, encombrements stériques, ...). Les liaisons fortes entre résidus éventuellement éloignés dans la séquence primaire[#] apportent également leur contribution à cet édifice moléculaire. La stabilité conformationnelle d'une protéine dépend donc directement de sa séquence primaire.

[#] ces liaisons fortes peuvent être par exemple des ponts disulfures existant entre les groupes S-H de deux cystéines, ou des liaisons à un ion (Fe²⁺, Ca²⁺, Zn²⁺, ...).

Dans cette séquence, la mutation d'un résidu peut avoir différentes repercussions :

- Provoquer la perte du repliement. Ce cas est aisément détectable puisqu'il entraine souvent la précipitation de la protéine. Cet état solide conduit d'ailleurs à la non validité du cadre de la thermodynamique.
- Provoquer la perte d'une fonction biologique. Ce cas peut être détecté lorsque cette fonction est connue. Malheureusement celle-ci ne l'est pas toujours, en particulier lorsque l'on travaille sur un domaine.
- Provoquer une modification de la stabilité thermodynamique sans repercussion ni sur le repliement, ni sur une fonction connue. Pour s'en assurer, il est donc nécessaire d'étudier sa stabilité thermodynamique par rapport à celle de la protéine sauvage.

Sans réaliser de mesure de stabilité, il est difficile de prédire l'effet d'une mutation. Quelques règles assez générales ressortent des nombreuses études de mutagénèse dirigée. Ainsi, la mutation d'un résidu impliqué dans le cœur de la protéine peut avoir de graves répercussions sur cette stabilité, une unique mutation pouvant aller jusqu'à diminuer la stabilité de près de 40 kJ/mol (Creighton, 1993). En comparant cette valeur à la différence d'énergie libre d'une protéine, une unique mutation peut donc dans certains cas conduire à la perte du repliement. Par ailleurs, le degré d'implication d'un résidu dans la stabilité d'une structure est souvent corrélé au nombre d'interactions auxquelles cet acide aminé participe, ce nombre étant généralement proportionnel à la longueur de sa chaîne latérale. La proline constitue un cas particulier parmi les acides aminés. De par son cycle carboné, ce résidu, bien qu'il établisse peu d'interactions hydrophobes ou de charges avec ses voisins, impose une contrainte géométrique particulière à la chaîne peptidique, capable à elle seule de rompre la symétrie d'une hélice.

Et pourtant, ces règles générales ne sont de loin pas toujours respectées. Par exemple, les séquences de protéines appartenant à une même famille de protéines, ont pu changer au cours de l'évolution avec cependant des effets mineurs sur la conformation des protéines de cette famille. Dans les cas extrêmes, la divergence de séquences primaires de protéines d'une même famille peut conduire à ne conserver que 20% d'identité, tout en conservant pourtant un repliement similaire (Chothia & Lesk, 1986). Ceci justifie l'utilisation de la mutagénèse dirigée en biochimie.

Rappelons cependant la non-relation de cause à effet entre une protéine formant des corps d'inclusion solides et un quelconque défaut de stabilité de cette protéine. Altamirano *et al.* montrent

dans un premier temps que que l'un des mutants de IGPS (enzyme impliquée dans les voies de synthèse du tryptophane) est produit systématiquement sous forme de corps d'inclusion. En utilisant une méthode originale de renaturation sur colonne, ils parviennent à obtenir une protéine repliée, et qui présente une stabilité thermodynamiquement comparable à d'autres protéines (Altamirano, 2000).

A.IV.1.c/ La détermination de la stabilité thermodynamique

♦ stabilité face à une dénaturation thermique

Sur une protéine, l'augmentation de température provoque d'abord une augmentation de flexibilité des régions les moins contraintes –principalement les boucles–, puis passé un certain seuil de température, la destruction des structures secondaires type hélice α ou feuillet β . Ce seuil est d'autant plus élevé que le nombre d'interactions attractives est important. Ce nombre important d'interactions attractives contribue à l'augmentation de l'enthalpie Δ H, tandis que l'augmentation de flexibilité contribue à une diminution de l'entropie Δ S, et donc à l'augmentation de -T Δ S. Passée une certaine température, la variation d'énergie libre Δ G résultante deviendra positive : la forme native (N) n'est plus favorisée par rapport à la forme dépliée (D).

Je ne détaillerai cependant pas davantage ce type de dénaturation car, hormis quelques rares exceptions, elle conduit à l'agrégation ou à la précipitation de la protéine. Ce cas s'éloignant d'un système à deux états, il n'autorise plus à utiliser les lois de la thermodynamique classique.

• stabilité face à un agent dénaturant

La dénaturation peut être induite par l'augmentation de concentration d'un agent dénaturant type urée ou chlorure de guanidium.

$$\begin{array}{ccc} O & O \\ || \\ H_2N--C--NH_2 & H_2N--C==NH_2^+ \\ ur\acute{e} & ion \ guanidium \end{array}$$

Ces additifs interagissent préférentiellement avec les surfaces de protéines. Ils ont un effet dénaturant puisque les protéines ont une plus grande surface accessible à ces agents lorsqu'elles sont dépliées. Cependant, dans ce cas, ce dépliement n'implique pas une agrégation car ces agents

protègent justement la chaîne peptidique du solvant. Les transitions de dépliements peuvent être suivies avec toute méthode sensible à la conformation des protéines. Les plus couramment utilisées sont la fluorescence et le dichroïsme circulaire (Cf. chapitre II de cette introduction).

Ces agents chaotropiques ont eux-mêmes un effet sur les propriétés spectrales des états replié et déplié qui peut être non négligeable. Il en résulte dans ce cas une pente non nulle dans les parties pré- et post-transitions des courbes de dénaturation qui doit être préalablement déterminée par ajustement linéaire (Creighton, 1993) (Fig. 3 A/).



Fig. 3 : mesures à l'équilibre de transitions de dénaturation de la phosphoglycérate kinase. A/ dénaturation au chlorure de guanidium suivie par fluorescence (cercle plein) et DC (cercle vide). La fraction de protéine dépliée est tracée en B/ [d'après (Nojima, 1977)].

En supposant ensuite que les valeurs extrêmes recouvrent les états totalement repliés ou dépliés, les valeurs sont ensuite converties en pourcentage : de 0% pour l'état natif à 100% pour l'état déplié[#]. Dans le cas de la dénaturation de la phosphoglycérate kinase, les techniques de fluorescence et de DC donnent deux courbes superposées (Fig. 3 B/), en accord avec une transition à deux états. Ce type de transition est d'ailleurs confirmé par l'idée de coopérativité détaillée précédemment. Cette coopérativité implique en effet une cassure très rapide lors du processus de débobinage de la protéine : en d'autres termes il n'existe qu'un très faible nombre d'états intermédiaires (Privalov, 1979).

[#] pour les aspects pratiques, voir Pace, C., Shirley, B. & JA., T. (1994) Measuring the Conformational Stability of a Protein. In T. Creighton (ed.) *Protein Structure: A Practical Approach*. IRL Press, Oxford, Vol. 174, pp. 311-329..

Ces courbes de dénaturation apportent une première information sur l'énergie libre de la protéine (Creighton, 1993). En effet, la constante à l'équilibre entre les états natif (N) et déplié (D) peut être mesurée directement à partir de la fraction moyenne dépliée (α) dans la région de transition.

L'écart d'énergie libre ΔG de N par rapport à D est donné par :

$$\Delta G = G_N - G_D = -RT \ln(K_{eq}) \quad \text{où} \qquad K_{eq} = [N]/[D] = (1 - \alpha)/\alpha. \tag{2}$$

D'un autre côté, ce même ΔG dépend linéairement de la concentration en dénaturant :

$$\Delta G = \Delta G_{H20} + m.[dénaturant]$$
(3)

Le paramètre *m* reflète la dépendance pour l'énergie libre de la concentration en agent dénaturant. Typiquement, *m* vaut 1 ou 3 kcal/mol pour l'urée ou le chlorure de guanidium. A la demidénaturation, K_{eq} est égal à 1, et donc ΔG est nul. Bien qu'il n'existe aucune base théorique pour cela, la concentration d'agent chaotropique à la demi-dénaturation est généralement extrapolée linéairement, permettant alors de déterminer sa valeur ΔG_{H20} en absence de dénaturant.

Cette valeur de l'énergie libre ΔG_{H20} renseigne sur la stabilité thermodynamique de la protéine. Qui plus est, dans un contexte de stabilité thermodynamique, c'est-à-dire lorsque l'on considère que la mutation n'affecte pas le repliement de la structure, il est possible de juger du gain en énergie libre (c'est-à-dire du gain en stabilité) apporté par une mutation par rapport à la protéine sauvage (Fig. 4).





Une autre application consiste à déterminer par cette même technique le ΔG_{H20} de la protéine seule, et le ΔG_{H20} de la protéine avec son ligand.

A.IV.2/ Autres notions incluses dans le terme de stabilité

On parle également de stabilité, ou instabilité, lorsque la mesure de l'activité d'une protéine diminue au cours d'un test biochimique. Cette stabilité est pourtant différente de la stabilité thermodynamique décrite précédemment qui ne tient pas compte des étapes intermédiaires de repliement. Par exemple, l'augmentation de la stabilité thermodynamique par mutation n'explique pas nécessairement une diminution de l'agrégation. En effet, le gain en stabilité apporté par une mutation est déterminé par rapport au repliement correct de la protéine, tandis que cette même mutation peut perturber le processus de repliement, et donc conduire éventuellement à une augmentation de l'agrégation. En d'autres termes une mutation, déterminée pour augmenter la stabilité d'une protéine, pourrait dans le même temps augmenter la barrière cinétique entre les états déplié et replié, et ainsi empêcher ce repliement.

En fait, l'instabilité observée lors d'un test biologique est un terme mal défini qui peut cacher différentes causes possibles. Citons quelques exemples parmi ceux les plus fréquents :

- i/ lorsque l'analyse sur gel SDS d'une protéine ne laisse pas apparaître la bande attendue, la protéine est dite instable. Cette instabilité n'est en fait que le résultat d'une protéolyse (pendant l'expression ou pendant le processus de purification) ou d'autres modifications post-transcriptionnelles, non nécessairement corrélée à une instabilité thermodynamique.
- ii/ la protéine peut également être considérée comme instable lorsqu'on la retrouve dans des corps d'inclusion. Là aussi, comme décrit précédemment, ces corps d'inclusion peuvent résulter d'un défaut de repliement, ou d'un déséquilibre entre repliement correct et agrégation. La notion de stabilité thermodynamique ne peut non plus s'appliquer ici puisqu'elle sous-entend la stabilité d'un repliement donné, alors que cet état n'a pas été atteint.
- iii/ la protéine est dite aussi instable lorsqu'elle est précipitée après purification. Celle-ci peut intervenir suite à une incompatibilité entre le tampon et une protéine –même correctement repliée–, suite à la formation de ponts dissulfure intermoléculaires, au rapprochement de zones hydrophobes exposées au solvant alors même que la protéine est repliée, ou encore suite à un repliement incomplet comme état métastable entraînant la précipitation lente de la protéine.

Cette liste n'est évidemment pas exhaustive, mais donne d'autres notions incluses dans ce terme polysémique de stabilité. Mon travail de thèse m'a amené à rencontrer ces autres notions de stabilité bien plus que celle de stabilité thermodynamique. Ces aspects demandent en fait d'autres outils

d'analyse que ceux rencontrés classiquement en biochimie, en particulier des outils axés sur la biophysique. Ce sont ces outils que j'introduis dans le chapitre suivant.

CHAPITRE II : TECHNIQUES BIOPHYSIQUES

A.V/ Principes généraux de spectroscopies.

Une expérience classique de spectroscopie est conceptuellement extrêmement simple. Une radiation électromagnétique d'une longueur d'onde donnée est dirigée vers l'échantillon. Les propriétés de la radiation sortante sont analysées. La propriété la plus souvent mesurée est la fraction de lumière absorbée ou dissipée par l'échantillon (spectroscopies d'absorption, de diffusion). Il est également possible d'analyser la lumière réémise par l'échantillon à une longueur différente de celle utilisée pour l'excitation (spectroscopies de fluorescence, de phosphorescence, diffusion Raman, ou encore diffusion inélastique). Hormis l'intensité et la longueur d'onde, des techniques plus complexes analysent également la polarisation de la radiation émise par l'échantillon (dichroïsme circulaire, anisotropie de fluorescence, ...).

Chaque spectroscopie utilise une plage de longueur d'onde bien définie (Fig 5), fonction des propriétés de la molécule que l'on étudie : des ondes radio pour la RMN ou la RPE, aux rayons X pour la cristallographie en passant par l'infrarouge pour la technique FTIR, l'ultraviolet proche et/ou lointain pour le dichroïsme circulaire ou la fluorescence.

A.V.1/L'interaction lumière / matière

La spectrométrie se base sur l'interaction des ondes électromagnétiques avec la matière (atomes, molécules, solution, ...). La lumière est constituée d'un champ électromagnétique oscillant très rapidement. La matière, quant à elle, possède des propriétés électriques et magnétiques liées aux distributions de charges et de spins. L'onde, en interagissant avec cette matière, modifie sa distribution de charges et de spins. Les techniques spectroscopiques consistent à étudier, après excitation, le retour à l'équilibre de cette distribution.





24

En fonction des propriétés mises en jeu, la lumière est considérée soit comme une particule d'énergie E –un photon–, soit comme une onde de longueur d'onde λ . Suivant le type de spectrométrie, on évoque davantage le côté corpusculaire (fluorescence, absorption, ...) ou le côté ondulatoire (RMN, RPE, ...). La relation liant énergie et longueur d'onde est :

$$E = h \cdot \frac{c}{\lambda} \tag{4}$$

où h est la constante de Planck et c la vitesse de la lumière.

Une molécule en solution, irradiée par une onde, peut soit absorber, soit diffuser la lumière incidente. Nous allons désormais distinguer ces deux types de phénomènes.

A.V.1.a/ Absorption de lumière

Un système, s'il est excité par une radiation d'énergie suffisante, peut passer d'un état fondamental à un état excité en absorbant un photon, et donc son énergie : c'est *l'absorption* de lumière (Fig. 6). Il s'agit d'un phénomène très rapide, de l'ordre de quelques femtosecondes (fs.) (1fs. = 10^{-15} s.).



Fig. 6 : représentation schématique des états d'énergie d'une particule. Les flèches en trait épais continus représentent les transitions électroniques entre deux états associées à un photon, celles en traits épais discontinues les transitions électroniques entre deux états sans émission de photons, celles en trait fins des transitions par conversion interne sans émission de photon au sein d'un même état. Enfin, celles en pointillées notées **r** et **v** correspondent respectivement aux transitions rotationnelles et vibrationnelles. Les niveaux d'énergies ne sont pas à l'échelle.

◆ Absorption de photon et phénomène de fluorescence

La lumière peut être absorbée par interaction directe avec une particule, par exemple un électron d'une orbitale atomique, et l'amener ainsi dans un état excité. Les règles de mécanique quantique n'autorisent l'absorption d'un photon que pour certaines valeurs d'énergie, au moins supérieures à e1-e0 (e1 et e0 sur la figure 6). L'électron revient ensuite dans le sous-état excité de plus faible énergie e1. Puis, le retour à l'état fondamental peut alors suivre différents procédés : soit l'électron cède son surplus d'énergie par collision ou par dissipation de chaleur au milieu environnant, soit le
système réémet un photon. Ce phénomène est appelé *fluorescence* ou *phosphorescence* suivant que la durée de vie de l'état excité est de l'ordre de la nanoseconde (ns), ou de la milliseconde (ms). Les systèmes pouvant fluorescer sont principalement les électrons des cycles aromatiques. Tout chromophore fluorescent peut être caractérisé par son *rendement quantique*, défini comme le rapport du nombre de photons émis par le nombre de photons absorbés. Par ailleurs, il est intéressant de noter que l'énergie du photon réémis est toujours plus basse –et donc sa longueur d'onde plus grande– que celle du photon excitateur.

Ces principes sont généralisables aux particules autres que l'électron : les nucléons (neutron et proton), ou même les particules constituants ces nucléons (quarks, ...). La différence réside dans les énergies mises en jeu : plus l'espace occupé par la particule diminue, plus l'énergie pour l'exciter augmente. Ainsi, l'énergie nécessaire pour exciter un proton ou un neutron est bien supérieure à celle nécessaire pour exciter un électron. Ce domaine d'énergie, non représenté sur la figure 5 se situe au-delà des rayons X à des fréquences supérieures à 10^{20} Hz, ou des longueurs d'onde inférieures à 0,01 nm.

• Absorption de photon par une molécule

Un électron associé à une orbitale moléculaire peut lui aussi absorber un photon. Là également, seuls les photons ayant une certaine valeur d'énergie peuvent être absorbés. Cependant ces électrons étant délocalisés, l'énergie d'un photon absorbé est plus faible que dans le cas de l'absorption d'un photon par un électron d'une orbitale atomique. Les énergies de liaison ou de rotation, correspondant à des sous-états de l'état fondamental ou états vibrationnels, s'en trouvent donc affectées (Fig. 6).

A.V.1.b/ Le phénomène de diffusion de lumière

Dans les phénomènes d'interaction lumière / matière, nous avons pour l'instant évoqué le cas où le photon est absorbé par la matière. Lorsque le photon n'a pas l'énergie suffisante pour être absorbé par les électrons de la molécule présente en solution, le photon est alors diffusé par la matière. Ces phénomènes sont regroupés sous le terme de *diffusion de lumière*. On distinguera deux cas de diffusion :

- le photon diffusé a « rebondi » sur la matière sans perte d'énergie. La longueur d'onde du photon diffusé est donc identique à celle du faisceau incident. On parle alors de diffusion élastique, ou *diffusion de Rayleigh élastique*.
- un couplage entre le photon et le mouvement subi par l'électron conduit un petit nombre de photons à échanger une partie de leur énergie avec des électrons de la molécule. Cette énergie est beaucoup plus faible que celle correspondant aux transitions électroniques. Du fait de la relation (1) reliant l'énergie à l'inverse de la longueur d'onde, la longueur d'onde du photon diffusé sera donc plus grande que celle du faisceau incident. Si les mouvements sont d'origine brownienne, on parle de *diffusion de Rayleigh inélastique*. S'ils correspondent à des vibrations intramoléculaires, on parle de *diffusion Raman*.

• Traitement physique de la diffusion de lumière

Le traitement quantitatif de la diffusion de lumière commence avec les travaux de Lord Rayleigh en 1871. La lumière est constituée de la variation périodique des champs électrique et magnétique exercant une force sur les électrons des molécules de la solution. Cette force engendre un déplacement de l'électron à la même fréquence que la lumière excitatrice, la position moyenne de l'électron coïncidant avec sa position à l'équilibre. Or les équations de Maxwell stipulent que toute charge soumise à une accélération va émettre une onde électromagnétique –c'est-à-dire de la lumière– de même fréquence que l'onde excitatrice dans une direction perpendiculaire à celle de l'accélération : ce principe physique trouve une application dans les rayonnements synchrotrons. Dans notre cas, cette nouvelle radiation correspond à ce que l'on appelle la *lumière diffusée*. Par contre, bien que la fréquence soit identique, sa longueur d'onde pourra différer selon le milieu traversé. Ce principe est par exemple à l'origine du phénomène de réfraction (apparence bleue du ciel en journée ou rouge en soirée, impression de cassure d'une cuillère à l'interface entre l'air et l'eau, …).

Contrairement à l'absorption, le photon n'est pas absorbé par la molécule lors des phénomènes de diffusion de lumière. Arrivé selon la direction du faisceau incident, ce photon repartira dans une direction qui peut être toute autre : ceci induit une diminution de l'intensité du faisceau en sortie de l'échantillon, mais surtout –phénomène plus facile à observer– une apparition de lumière dans les directions autres que celle du faisceau incident.

♦ Diffusion, diffraction et milieu isotrope

Dans le cas où la matière est ordonnée dans l'espace, tel dans un cristal de protéine la lumière semble n'être diffusée que selon certaines directions privilégiées. Le terme de *diffraction* remplace

alors celui de diffusion, bien que le principe physique soit identique à la base. Dans ce cas, les ondes diffusées -initialement de manière isotrope- interfèrent de manière soit constructive, soit destructive, laissant apparaître une distribution anisotrope de taches et de trous de lumière. Les particules diffusantes s'apparentent alors à des sources de lumières dépendantes les unes des autres. Au contraire, dans une solution ou dans un gaz, chaque particule diffusante occupe une position et une orientation aléatoires. La lumière diffusée par l'ensemble des particules l'est donc de manière *isotrope*, c'est-à-dire avec la même intensité dans toutes les directions de l'espace. Ces sources de lumière sont donc considérées comme des sources indépendantes.

A.V.I.c/ Absorption, diffusion et fluorescence

La figure 7 résume les différences entre absorption, diffusion et fluorescence, tels que décrits dans les paragraphes précédents. Dans un souci de clarté, dans le cas de la fluorescence (Fig. 7c), la radiation de même longueur d'onde λ que l'onde incidente, existante en sortie de l'échantillon, n'a pas été représentée. Cette radiation ne joue cependant aucun rôle dans les phénomènes de fluorescence.



Figure 7 : vision schématique des différences entre l'absorption (a), la diffusion (b) et la fluorescence (c). Le carré grisé représente l'échantillon.

Dans une expérience d'absorption ou de fluorescence, les longueurs d'onde incidentes doivent être absorbées par la matière. Au contraire, la diffusion de lumière exige de se placer à des longueurs d'onde pour laquelle la matière est « transparente ».

A.V.2/ Les montages utilisés en spectrométrie



Figure 8 : représentation schématique des trajets optiques (flèches) pour différentes géométries de montages rencontrés en spectrophotométrie. Le carré grisé représente l'échantillon.

Il existe quatre possibilités de détection, résumées dans la figure 8 : observations (a) à 180° dit montage linéaire, (b) à 90° appelé montage en L ou perpendiculaire, (c) à 45°, et (d) à 0° ou montage en rétrodiffusion. Toute spectrométrie impliquant une analyse de la lumière en sortie de l'échantillon, se base sur l'un de ces quatre montages. Chaque spectromètre possède une source et un détecteur qui lui sont propres, appropriés aux longueurs d'onde de travail.

Le montage linéaire est utilisé pour comparer les intensités lumineuses incidente et transmise : il est donc idéal pour mesurer une absorption. Ceci implique que la solution induise des modifications non négligeables, sans pour autant atténuer totalement le faisceau sortant. Au contraire, le montage en L est préféré pour analyser la lumière réémise : il est donc idéal pour effectuer des mesures directes de diffusion, ou de fluorescence. Le signal en sortie du détecteur étant généralement amplifié par un photomultiplicateur, l'avantage de ce montage réside donc dans sa très grande sensibilité, le faisceau incident ne perturbant pas les mesures. Le montage à 45° n'est utilisé que lorsque la solution absorbe trop de lumière. Ce montage est rarement rencontré en biologie. Enfin, le montage en retrodiffusion s'utilise dans des cas très particuliers comme la Résonance Magnétique Nucléaire, la mesure de fluorescence sur des plaques ELISA, ... : la source d'ondes et le détecteur étant confondus, il faut développer une technique de détection propre à chaque spectroscopie. Par exemple, la RMN utilise une unique bobine pour l'excitation et pour l'acquisition, mais à des instants différents.

Reprenons maintenant quelques techniques spectroscopiques plus en détail.

A.VI/ Aspects expérimentaux de techniques spectroscopiques appliquées à l'étude de protéines.

A.VI.1/ L'absorption de lumière ou la mesure de densité optique

A.VI.1.a/ Absorption de la référence.

Toute molécule absorbe de la lumière, mais avec une efficacité variable d'un type de molécule à l'autre. Ainsi, il est envisageable de mesurer aussi bien l'absorption d'une solution aqueuse que celle d'une protéine. Ceci traduit la relativité de la mesure. La mesure d'absorption de lumière par une solution acqueuse implique l'utilisation de l'eau comme référence, tandis que la mesure d'absorption par une protéine implique celle du tampon dans lequel elle est dissoute. Dans le cas d'une protéine, cette référence tiendra compte de tous les éléments l'entourant (solvant, ions), y compris également des paramètres tels que le pH ou la température qui peuvent avoir un effet sur l'absorption.

Un agent dénaturant tel que urée ou chlorure de guanidium est parfois utilisé à des concentrations atteignant plusieurs molaires, valeurs bien supérieures à celle de la protéine. Une attention toute particulière doit être apportée à leur pureté. En effet, à de telles concentrations, une impureté même à l'état de trace aura une concentration de même ordre de grandeur que celle de la protéine. Or ces impuretés peuvent absorber la lumière aux longueurs d'onde de travail, et donc perturber sensiblement le spectre de la protéine que l'on étudie.

A.VI.1.b/ Absorption par la molécule d'intérêt.

La loi de Beer-Lambert permet de relier directement l'absorbance $A(\lambda)$ et la concentration d'une molécule :

$$A(\lambda) = \log\left(\frac{I_0}{I}\right) = \mathcal{E}(\lambda).l.c$$
(5)

 I_0 et I sont les intensités de la lumière traversant respectivement la référence et l'échantillon, ε (M⁻¹.cm⁻¹) le coefficient d'extinction molaire, l la longueur du trajet optique (en cm) et c la concentration (en M). Remarquez que ε , et donc l'absorbance A, dépendent de la longueur d'onde. Pratiquement, le coefficient d'extinction molaire correspond à la densité optique (DO) d'une

solution à 1 molaire. Cette relation dénote le caractère linéaire de l'absorption, à condition de rester dans des gammes de concentrations pas trop élevées, typiquement telle que la DO reste inférieure à 1. Les grandes variations de concentrations limitent la sensibilité de l'appareil, car celui-ci doit être conçu pour couvrir une grande plage d'intensités (de très fortes à très faibles). Au delà d'une DO de 1, l'imprécision de la mesure ne permet plus d'appliquer cette loi.

Même si les substances biologiques absorbent un grand nombre de fréquences, seules certaines domineront le spectre d'absorption. Parmi ces objets, nous retrouvons celles ayant les cycles aromatiques (Fig. 9A) et la liaison peptidique dans le cas des protéines, ainsi que les bases dans le cas des acides nucléiques (Tableau 1). L'absorption d'une molécule résulte de la combinaison linéaire de l'absorption de chacun des chromophores, pondérée par leur nombre.

Groupements	Absort	bance	Fluorescence			
Groupements	$\epsilon (M^{-1}.cm^{-1})$	$\lambda_{max}(nm)$	φ	$\lambda_{max}(nm)$		
Tryptophane (W)	5600	280	0,20	348		
Tyrosine (Y)	1400	275	0,15	303		
Phénylalane (F)	200	257	0,04	282		
Adénosine (A)	13400	260	2,6 10 ⁻⁴	321		
Cytosine (C)	6100	267	0,8 10 ⁻⁴	313		
Guanine (G)	8100	275	3,0 10 ⁻⁴	329		
Uracile (U)	9500	260	0,4 10 ⁻⁴	308		
liaison peptidique	4-8000	≈190				
flavine FAD	11000	450	0,03	530		
NADH	6200	339	0,02	470		
DANSYL	3400	330	0,10	510		
ANS	6800	374	0,98	454		
bromure d'éthidium	3800	515	1,00	600		

Tableau 1 : caractéristiques d'absorption et de fluorescence de quelques composés classiquementrencontrés en biologie (acides aminés, bases, quelques chromophores intrinsèques et extrinsèques).Les valeurs des longueurs d'onde indiquées (absorption et fluorescence) sont sensibles à
l'environnement et sont donc données à titre indicatif (adapté de [Eftink, 1991]).



Fig. 9 : spectres d'absorption (A) et de fluorescence (B) des résidus tryptophane (Trp), tyrosine (Tyr) et phenylalanine (Phe). Les mesures d'absorption sont réalisées à pH 7.0 à 25°C avec des concentrations de 1 mM (Phe) ou 0,1 mM (Tyr et Trp). L'excitation pour les mesures de fluorescence est réglée à : 257 nm (Phe à 100 μM), 274 nm (Tyr à 6 μM) et 278 nm (Trp à 1 μM).

♦ Intérêt des mesures d'absorption en biologie

L'utilisation la plus fréquente de la loi de Beer-Lambert consiste, à partir d'une mesure de DO à 280 nm, à déterminer la concentration d'une protéine en solution connaissant son coefficient d'extinction molaire (calculé à partir de la séquence primaire et du tableau 1) (Gill & von Hippel, 1989). A 280 nm, l'environnement peut cependant influer de manière limitée sur ce coefficient d'extinction : la concentration mesurée sera donc réelle dans une limite de 5 à 10%.

Un spectre d'absorption entre 220 et 320 nm permet également de détecter la présence de contaminants, et plus particulièrement la présence d'ADN ou ARN dans une solution de protéine (ou inversement), les uns absorbant principalement vers 260 nm, les autres vers 280. Cette mesure est donc complémentaire d'un gel de protéines.

Un spectrophotomètre est aussi utilisé pour mesurer une concentration de bactéries en solution : cette mesure utilise cependant un autre principe physique qui est celui de la turbidimétrie. En se plaçant à une longueur d'onde où les molécules absorbent le moins –soit donc au delà de 550 nm–, la diminution de lumière dans la direction du faisceau incident est provoquée par la diffusion et l'absorption de lumière par les bactéries. Cette diminution est également traduite en terme de DO.

Il existe d'autres applications moins courantes à cette technique, comme par exemple :

- 1/ détermination du nombre de tyrosines dans une séquence. Le coefficient d'extinction molaire de la tyrosine est particulièrement sensible à l'état de protonation du groupement OH (pKa \approx 10,9), mais bien moins à son environnement polaire. Ainsi, une titration en pH entre 8 et 14, réalisée sur un spectromètre d'absorption, permet de détecter la présence de tyrosine(s) et peut même aider à remonter jusqu'à leur nombre inclus dans la séquence primaire [d'après (Cantor & Schimmel, 1980).
- 2/ effet d'hypochromicité : la théorie quantique de l'interaction lumière / matière démontre que l'interaction des chromophores entre eux provoque une diminution de l'intensité absorbée, à condition que ces chromophores soient à des distances inférieures à quelques angstrœms et soient distribués régulièrement. Ainsi, un ADN en double hélice absorbe moins de lumière que ce même ADN à l'état dénaturé ou digéré (Fig. 10).

Cette situation se retrouve également dans le cas de la poly-L-Lysine : la conformation de ce peptide, variable en fonction du pH, induit des modifications du coefficient d'extinction molaire entre 180 et 220 nm, région d'absorption de la liaison peptidique (Fig 11). Dans le cas des protéines, il n'a pas pu être établi de spectres d'absorption de références en fonction de la structure secondaire. Par ailleurs, l'acquisition de spectres d'absorption dans cette région (180 – 220 nm) ne sont pas évidents à obtenir : forte absorption du tampon (Cf. \$II.B.4), grande imprécision des spectrophotomètres. Les différences observées ne peuvent donc être reliées directement à une structure secondaire donnée, faisant plus de cette technique un moyen de comparaison entre différents lots de protéines, qu'une mesure absolue.



Fig. 10 : spectres d'absorption pour un ADN de E. coli. à différentes températures. On remarque que le spectre à haute température est similaire au spectre de l'ADN digéré, mais qu'ils diffèrent sensiblement du spectre de l'ADN natif [d'après (Voet, 1963)].

Fig 11 : coefficient d'extinction molaire de la poly-L-Lysine en fonction de la structure secondaire. pH 6.0, 25° C : structure aléatoire ; pH 10.8, 25° C : hélice α ; pH 10.8 52° C : feuillet β [d'après (Cantor & Schimmel, 1980)].

A.VI.2/ Spectroscopie de fluorescence

Le phénomène d'émission de fluorescence est un processus de l'ordre de 10^{-9} ou 10^{-8} s, à comparer à la vitesse relative de l'absorption (10^{-15} s). L'environnement du fluorophore sera donc modifié avant réémission du photon par des mécanismes intervenant sur cette même échelle de temps : collisions avec le solvant, réactions de protonation ou de déprotonation, changements conformationnels locaux, diffusions rotationnelle et translationnelle. L'intérêt de la spectroscopie de fluorescence réside donc tout particulièrement dans sa sensibilité à l'environnement bien meilleure que celle de la spectroscopie d'absorption.

Pour remonter à ces informations, il a été nécessaire de développer un large éventail de techniques : spectre d'excitation, d'émission, ou fonction du temps ; lampe à spectre continu, faisceau polarisé, impulsions laser ; mesures du maximum d'émission, rendement quantique, demi-vie, inhibition (« quenching »), polarisation ... Je ne décrirai dans la suite que la fluorescence à l'état d'équilibre excitée par une lampe à flux continu. Tous les résultats exploités dans ce manuscrit ont en effet été obtenus avec un tel spectrofluorimètre ; ce type d'appareil est d'ailleurs celui le plus couramment rencontré.

Un autre atout de la fluorescence tient dans l'utilisation du montage perpendiculaire. Le faible niveau de lumière parasite provenant du faisceau incident procure une grande sensibilité de détection. Il est donc possible de travailler avec des concentrations de l'ordre du nanomolaire.

Le phénomène de fluorescence se traduit par une réémission de lumière, à des longueurs d'onde toujours plus grande que la longueur d'onde excitatrice. Le tableau 1 donne les caractéristiques principales de fluorescence de quelques composés classiquement rencontrés en biologie. La figure 9B montre les spectres de fluorescence de trois résidus aromatiques : le tryptophane (W ou Trp), la tyrosine (Y ou Tyr) et la phénylalanine (F ou Phe).

Dans une expérience de fluorescence, deux paramètres sont particulièrement intéressants : l'intensité maximale et la longueur d'onde correspondant à ce maximum. Pour peu que le solvant soit polaire, ses molécules se redistribuent autour du fluorophore, impliquant une modification des propriétés électriques avant le retour à l'état fondamental. Cette redistribution tend à abaisser son énergie, et donc à augmenter la longueur d'onde de la lumière réémise. Ainsi, le maximum d'émission du tryptophane dans l'eau, solvant fortement polaire, se situe à 355 nm, tandis qu'il se décale à 325 nm pour ce même tryptophane plongé dans une solution de dioxane, solvant considéré comme apolaire (Gerard, 1975).

L'intensité de fluorescence est, elle aussi, sensible à l'environnement du tryptophane. Un environnement rigide ou non polaire provoque une augmentation du rendement quantique, tandis qu'une exposition au solvant peut entrainer une inhibition de fluorescence par des molécules en solution (en particulier l'oxygène ou les atomes lourds, type césium ou iode). Ainsi, le tryptophane possède un rendement quantique maximal de 0,30 lorsqu'il est enfoui contre 0,15 lorsqu'il est

exposé au solvant (Gerard, 1975). Le rapprochement du tryptophane d'un groupement SH ou d'un groupement carbonyle CO de la liaison peptidique peut induire également une inhibition (Ababou & Bombarda, 2001).

Enfin, il est parfois préférable de s'éloigner de la longueur d'onde d'excitation des résidus Tyr et Phe. On choisira alors une excitation à 295 nm, sélective du Trp (Eftink, 1998).

♦ Intégrité du repliement d'une protéine

La fluorescence renseigne dans une certaine limite sur le repliement d'une protéine (si elle contient au moins un tryptophane). Un maximum de fluorescence décalé vers des basses longueurs d'onde –jusqu'à 308 nm dans le cas de l'azurine (Eftink, 1998)– indique que le tryptophane est protégé du solvant. Ceci peut signifier, soit que la protéine est correctement repliée, soit qu'elle forme des agrégats isolant le tryptophane du solvant. Il en est de même pour le rendement quantique de fluorescence.

• Mesure de dénaturation par fluorescence

On suit le paramètre pour lequel la différence entre l'enfouissement et l'exposition au solvant des résidus aromatiques est la plus grande. Ceci dit, lorsque cela est possible, on suivra l'évolution de l'intensité plutôt que celle du maximum d'émission. En effet, ce maximum est pondéré par les concentrations relatives des états repliés et dépliés (comme pour l'intensité), mais aussi par le rendement quantique de chacun des états ; le cas extrême étant celui où l'état dénaturé ne possède plus de fluorescence (Eftink, 1998). Enfin, lorsque le signal est proche du bruit de fond comme ce qui est souvent à haute concentration d'agent dénaturant, il est plus facile d'en extraire le maximum d'intensité que la longueur d'onde.

♦ Intensité de fluorescence et concentration

On peut se dire que plus la concentration en molécules fluorescentes augmente, plus l'intensité de fluorescence sera grande. En fait, ceci n'est vrai que dans une certaine limite (Fig. 12A). A des concentrations suffisamment faibles, la fluorescence est directement proportionnelle à la concentration de l'échantillon. Par contre, à plus forte concentration, deux phénomènes interviennent en parallèle : la lumière incidente est absorbée ou diffusée dans une trop grande proportion avant d'avoir traversé toute la cuve ; la lumière émise par fluorescence est grandement diffusée dans la cuve. La diminution de l'intensité de fluorescence intervient lorsque la

concentration de protéines dépasse un seuil compris généralement entre 0.1 et 1 mM. Il est cependant bien rare de travailler à de telles concentrations vu la sensibilité des détecteurs.



Fig. 12 : Fluorescence et absorption de lumière. A/ relation entre intensité de fluorescence et concentration. Le rectangle en pointillé délimite la zone de linéarité. B/ Le carré en trait épais représente l'échantillon, I₀ et I_F respectivement les intensites des faisceaux incident et de fluorescence.

Cependant, puisque le rendement de fluorescence dépend dans une large mesure de l'environnement des fluorophores –et plus particulièrement du tryptophane–, la détermination de la concentration d'une protéine doit être réalisée en présence d'un agent dénaturant.

Une expérience très simple permet de vérifier l'absorption de lumière par la solution (Fig. 12B) : par un positionnement correct des fenêtres d'excitation et d'émission, il est possible d'évaluer l'absorption de la lumière incidente (comparaison entre a et b), ou la réabsorption de la lumière de fluorescence (comparaison entre a et c). En pratique, dans les expériences de fluorescence, seule la lumière provenant du centre de la cuve est recueillie.

A.VI.3/ Le Dichroïsme Circulaire (DC)

Le phénomène de dichroïsme circulaire découle de l'activité optique des molécules asymétriques en solution. Des chromophores chiraux –c'est-à-dire qui ne présentent ni centre, ni plan de symétrie– ne vont pas absorber de la même manière une lumière polarisée circulairement gauche ou droite. Un spectre de dichroïsme circulaire est obtenu en exposant un échantillon alternativement à une lumière monochromatique polarisée circulairement gauche puis droite, puis en détectant la différence d'absorption ΔA , et ce à chaque longueur d'onde. Cette différence, typiquement de l'ordre de 0,05% à 0,5% de l'absorption totale, est détectable très précisément. D'après la loi de BeerLambert, la différence d'absorption est directement reliée à la différence $\varepsilon_G - \varepsilon_D^{\#}$, appelée également pouvoir rotatoire. Ce pouvoir rotatoire peut être positif ou négatif (Fig. 13). Ces signes seront d'ailleurs opposés pour deux énantiomères L ou D. Un mélange racémique ne présente donc pas de signal en dichroïsme circulaire.



Fig. 13 : absorbance et dichroïsme circulaire. Le spectre d'absorption A/ contient deux pics d'absorption (a) et (b) optiquement actifs avec des pouvoirs rotatoires de signes opposés, et un pic (c) optiquement inactif.

Afin de comparer des spectres de molécules entre eux, la différence d'absorption est convertie en une unité indépendante de la concentration, l'ellipticité molaire [θ] (en degré.cm²/decimole) :

$$[\theta] = 3300.\Delta A / (d.M) \tag{3}$$

où d représente la longueur (en cm) du trajet optique, et M la concentration (en mol/l). Le facteur 3300 provient des conversions depuis le système international (de radian en degré, et de mole et décimole) multplié par le facteur numérique $\ln 2 / 4$.

En UV lointain, le dichroïsme circulaire d'une protéine tire principalement son origine de la chiralité de la liaison peptidique induite par les éléments de structures secondaires (hélice α , feuillet β , structure aléatoire, boucle, ...). Des spectres de références, dits spectres de Greenfield, ont pu être établis pour chaque type de structure secondaire (Fig 14) (Adler, 1973). Le spectre obtenu pour une protéine correspond à la combinaison linéaire de ces références.

 $^{* \}epsilon_D$ et ϵ_G sont les coefficients d'extinction molaires induits par une lumière polarisée circulairement droite et gauche.





Cette technique permet de travailler avec des solutions de 20 à 200 μ l à des concentrations de l'ordre de 50 à 1000 μ g/ml. Cependant, vu la gamme de longueurs d'onde utilisées, une attention toute particulière doit être portée au choix du tampon qui doit être le plus « transparent » possible. Sans cela, l'absorption serait telle que l'infime différence due à la chiralité des chromophores serait diluée dans le bruit de fond. Ceci constitue l'une des contraintes majeures en terme d'absorption.

nН	lim. sup.	Absorbance d'une solution à 10 mM.				
P	d'absorption	210 nm	190 nm	180 nm		
	170 nm	0	0	0		
	205 nm	0	> 0,5	> 0,5		
	210 nm	0	0,5	> 0,5		
	195 nm	0	0,01	0,15		
	220 nm	0,03	> 0,5	> 0,5		
9,1	200 nm	0	0,09	0,3		
12,0	230 nm	> 0,5	> 2	> 2		
8,0	220 nm	0,02	0,24	> 0,5		
7,5	230 nm	0,37	> 0,5	> 0,5		
	рН 9,1 12,0 8,0 7,5	pH lim. sup. d'absorption 170 nm 205 nm 210 nm 195 nm 220 nm 9,1 200 nm 12,0 230 nm 8,0 220 nm 7,5 230 nm	pH lim. sup. d'absorption Absorbance 170 nm 210 nm 170 nm 0 205 nm 0 210 nm 0 195 nm 0 220 nm 0,03 9,1 200 nm 0 12,0 230 nm > 0,5 8,0 220 nm 0,02 7,5 230 nm 0,37	$\begin{array}{c c c c c c c c c c c c c c c c c c c $		

Tableau 2 : absorption de quelques tampons à trois longueurs d'onde [d'après (Schmidt, 1994)].

Le tableau 2 résume la transparence de différents tampons ou composés à des longueurs d'onde inférieures à 220 nm : des composés ou tampons tels que l'Hepes ou le chlorure de sodium sont

donc à proscrire à ces longueurs d'onde, et ce même à des concentrations de l'ordre de 10 mM. Le DTT, l'histidine ou l'imidazole sont d'autres composés également à proscrire.

Un spectre de DC permet, à partir du spectre de la protéine d'intérêt et d'une base de données de références, de déterminer le pourcentage de structures secondaires existant dans la protéine en solution. Il existe pour cela plusieurs méthodes numériques d'analyse, dont les plus connues sont la régression finie (programme CONTIN) (Provencher & Glockner, 1981) et la décomposition en valeurs singulières par sélection variable (programme VARSELEC) (Manavalan & Johnson, 1987). Toutes ces méthodes donnent des estimations raisonnables des pourcentages de structures secondaires. Une excellente revue de Greenfield N.J. compare les différentes approches en donnant leurs avantages et leurs inconvénients, tout en apportant des renseignements pratiques pour l'obtention de mesures de qualité (Greenfield, 1996).

A noter que la comparaison du spectre de la protéine d'intérêt avec ceux compris dans ces bases de données nécessitent l'utilisation d'une unité commune, indépendante de la concentration de l'échantillon. Ceci implique une connaissance stricte de cette concentration. L'absorption dépendant faiblement de la conformation de la protéine en solution, une détermination précise de la concentration pourra être obtenue par une mesure d'absorption en condition dénaturante (Gill & von Hippel, 1989).

Bien que le Dichroïsme Circulaire rende compte du pourcentage de structures secondaires existant dans la protéine, il ne permet pas de savoir quel résidu y est impliqué, contrairement à la Résonance Magnétique Nucléaire.

A.VI.4/ La résonance Magnétique Nucléaire (RMN)

La Résonance Magnétique Nucléaire est une spectroscopie à part entière, puisqu'elle étudie l'interaction des ondes radiométriques avec les spins nucléaires. La faible sensiblité de détection oblige à travailler avec des concentrations de l'ordre du millimolaire. Je ne rentrerai pas davantage dans les principes de cette technique, mais renverrai plutôt aux excellents ouvrages tels que : (Wüthrich, 1986), .

Les fréquences de résonances des spins étant très sensibles à leur environnement, la spectroscopie RMN permet de différencier la plupart des protons entre eux. Ainsi, dans un spectre, les protons amides se situent généralement entre 7 et 10 ppm[#], les protons amides de bouts de chaînes latérales (Asn et Gln) ainsi que les protons des cycles (Trp, Tyr, Phe, His) entre 5 et 7 ppm, les protons H_{α} entre 4 et 5 ppm, puis les protons aliphatiques en decà de 4 ppm.

Qui plus est, les protons amides d'une protéine dépliée auront un environnement chimique davantage similaire entre eux que s'ils appartenaient à une protéine repliée. La RMN, de part son extrême sensibilité à l'environnement, permet de différencier une protéine dépliée d'une repliée, et ce même à partir d'un spectre à une dimension (spectre 1D) (Fig. 15). En particulier, les CH_3 résonant en deçà de 0.7 - 0.8 ppm sont généralement impliqués dans des interactions hydrophobes. Ceci constitue, avec la dispersion des fréquences de résonance, le critère essentiel du repliement détecté par RMN.



Fig. 15 : spectres RMN 1D à 500 MHz de l'αlactalbumine de porc, soit dans son état natif à pH 5,4, soit dépliée dans 9M d'urée [adapté de (Ptitsyn, 1992)].

Il est cependant impossible, dans le cas d'une protéine, d'attribuer chacun des pics à son proton correspondant sur de tels spectres. Pour y parvenir, il est nécessaire d'enregistrer des spectres à deux, 3, voire 4 dimensions. Ce type de spectres permet de visualiser des corrélations proton-proton (RMN 2D homonucléaire), ou entre proton, azote N¹⁵ et carbone C¹³ (RMN 2D, 3D, ...). Pour ces derniers, l'échantillon devra être enrichi avec ces isotopes pour une meilleure sensibilité de détection.

[#] le ppm, ou Partie Par Million, est l'unité utilisée en RMN : elle correspond à l'écart de fréquence entre la fréquence de résonance du spin observé et celle d'une référence externe. Cette unité est indépendante du spectromètre sur lequel a été enregistré le spectre.

La RMN est la seule spectroscopie permettant d'observer des taches de corrélation. Par ailleurs, ces corrélations peuvent varier en fonction de l'expérience utilisée en RMN : elles pourront traduire soit des interactions scalaires (interactions entre proches voisins liéés par des liaisons covalentes permettant par exemple de suivre l'enchaînement des protons d'un même résidu), soit des interactions dipôle / dipôle d'où l'on extrait des informations de distance. Selon le type d'interactions exploité dans une expérience RMN, on parlera respectivement de spectres TOCSY (« Total Correlation SpectroscopY ») ou NOESY («Nuclear Overhauser Effect SpectroscopY »). La RMN est donc une technique spectroscopique permettant d'identifier séparemment chaque résidu.

Les fréquences de résonance, très sensibles à l'environnement, dépendent aussi des propriétés chimiques de chaque résidu. La figure 16 indique une corrélation manifeste entre déplacement chimique du proton H α et structure secondaire : par exemple dans le cas de feuillets β , les fréquences des protons H α sont systématiquement déplacés vers les hauts ppm.

		6	5	p.p.m. 4		3
Ala	A	1020				
Cys	С	1 1555	1	1		Ì
Asp	D	1				1
Glu	E	in the second	unun a	000000		1
Phe	F		20200	53		1
Gly	G			SCORE STORE		
His	н		SAMA SA	1		
lle	I	1	10000	23 2323	E	i
Lys	к		100000			1
Leu	L		00000	1.000		1
Met	м		10,9252 07			1
Asn	N	1				Ì
Pro	Ρ		10000000	55.555 55 1 555		1
Gln	Q		820000	2272273 2000		1
Arg	R	1	IN NO			
Ser	S		2222223			-
Thr	т		2000000	225223 200000000000000000000000000000000	8	-
Val	v	1	1			
Trp	W	1	10000000			1
Tyr	Y	1	200000.92			1

Fig. 16 : distributions des déplacements chimiques des protons alpha pour chacun des 20 acides aminés communs, pour une hélice alpha (rectangles gris foncés), et un feuillet beta (rectangles gris clairs). Les barres noires verticales indiquent la valeur médiane [d'après (Wishart, 1991)].

La méthode de Wishart, basée sur la comparaison du déplacement chimique du proton H α pour chaque résidu avec celui de ce même résidu en structure aléatoire (Wishart, 1992), s'est révélée une manière efficace et rapide pour repérer les éléments de structures secondaires d'une protéine.

La détermination d'une structure par RMN consiste donc à : i/ identifier sur les spectres chacun des protons –et éventuellement azote et carbone dans le cas d'un échantillon marqué– d'un nombre maximum de résidus de la séquence primaire, ii/ identifier et accumuler les informations de distances entre protons (dans le cas de la RMN 2D homonucléaire), iii/ générer des structures calculées respectant l'ensemble des informations de distances. D'autres informations peuvent également être receuillies par RMN afin de faciliter cette détermination (vitesse d'échange de protons, dynamique, constante de couplage scalaire, ...).

A première vue, il existe deux inconvénients majeurs à cette technique :

- i/ elle nécessite des concentrations élevées de protéines, de l'ordre du millimolaire. Cette contrainte est en partie résolue avec l'utilisation de cryosondes : l'augmentation du rapport signal / bruit permet d'utiliser des échantillons à des concentrations de l'ordre de la centaine de μM.
- ii/ la qualité du signal mesuré en RMN se dégrade avec la taille de la molécule. En effet, la largeur à mi-hauteur des raies dans un spectre RMN est inversement proportionnelle à la vitesse de relaxation transversale, elle-même inversement proportionnelle au temps de corrélation globale de la protéine (Dayie, 1996). Ce temps caractéristique augmentant avec la taille des molécules en solution, il existe une limite de taille au delà de laquelle la largeur des pics devient telle que le pic disparaît dans le bruit du spectre. Les progrès récents, ainsi que le marquage isotopique, ont repoussé la limite théorique de taille des molécules observées par RMN au delà de 100 kDa (Wider & Wuthrich, 1999). Mais, la possibilité de résoudre une structure 3D reste conditionnée par l'observation de NOEs, et donc de la résolution des superpositions dans le spectre. En pratique, on ne sait pas calculer une structure si la taille est supérieure à 30 kDa. Cependant, pour des molécules de taille résonnable, les difficultés majeures qui se présentent maintenant dans une étude structurale se situent : en aval avec l'exploitation manuelle des données enregistrées, et en amont avec l'obtention d'échantillons enrichis (azote 15, carbone 13 et deutérium) sans caractère agrégatif. En effet, l'agglomération des molécules entre elles augmentent le temps de corrélation globale de la particule, jusqu'à éteindre tout signal.

Malgré ces désavantages, la Résonance Magnétique Nucléaire est devenue, avec la possibilité de déterminer des structures de macromolécules biologiques à une résolution atomique, un outil de choix de la biologie structurale. La connaissance de cette structure peut être d'une grande aide dans la compréhension de la fonction d'une protéine, ses mécanismes d'interactions (Wagner, 1997). Elle peut par exemple aider à mettre en évidence le rôle prépondérant de certains acides aminés

dans la structure d'une protéine ; aucune mutation ne pourra affecter ces résidus sans risquer de perdre ce repliement. Par contre, cette analyse ne permet pas toujours de détecter les résidus clefs dans les processus de repliement d'une protéine.

A.VI.5/ Les spectroscopies infrarouge et Raman

A.VI.5.a/ La spectroscopie infrarouge

L'absorption de la lumière dans le domaine de l'infrarouge met en jeu des transitions entre les niveaux de vibration des molécules. Ces niveaux correspondent à des mouvements coordonnés des noyaux, principalement en phase ou en anti-phase, ces mouvements provoquant l'élongation ou la rétraction des liaisons. Chaque liaison pouvant apporter sa contribution, le spectre infrarouge des molécules est très riche. Cependant, il est beaucoup moins exploité que les autres spectres d'absorption pour plusieurs raisons. Tout d'abord, les coefficients d'extinction molaires ɛ dans l'infrarouge, avec des valeurs inférieures à 1000 M⁻¹.cm⁻¹, sont environ 10 fois plus faibles que ceux dans l'ultraviolet. Les échantillons doivent donc être relativement concentrés, de l'ordre du milligramme / ml dans des volumes de quelques dizaines de µl. Par ailleurs, l'eau absorbe fortement dans les zones intéressantes du spectre. Les mesures se font donc soit sur des poudres ou des films, soit en solution aqueuse, mais en les limitant alors uniquement à quelques fenêtres où l'absorption du solvant reste faible. Avec les appareils de nouvelles générations, ce problème est partiellement résolu grâce à l'emploi de l'interférométrie de Michelson et la Transformée de Fourier Rapide. L'augmentation du rapport signal / bruit permet de supprimer une partie du signal de l'eau. L'autre possibilité est de remplacer les molécules de H₂O par du D₂O : l'inconvénient étant un risque de dénaturation pendant la substitution -non négligeable- ou une deutération partielle de la molécule.

Krimm *et al.* ont édité une revue très complète sur ces techniques spectroscopiques (Krimm & Bandekar, 1986), dont je donne ci-dessous un très bref aperçu. Les spectres IR présentent 4 bandes distinctes, renfermant chacune des informations très spécifiques (Fig. 17) :



Fig. 17 : les quatre bandes principales d'absorption en spectroscopie IR de protéines. L'unité utilisée [le nombre d'onde (cm⁻¹) est égal à l'inverse de la longueur d'onde] a été choisie par les premiers spectroscopistes infrarouge. (source : http://www.imbjena.de/ImgLibDoc/ftir/IMAG E_FTIR.html)

- <u>La bande amide A</u> : elle est constituée à près de 95% par la vibration d'élongation de la liaison
 N-H. Ce mode de vibration ne dépend pas de la conformation de la chaîne carbonée principale,
 mais est très sensible à la longueur de la liaison hydrogène.
- <u>La bande amide I</u>: c'est la plus intense des bandes d'absorption pour les protéines. Elle est gouvernée principalement par la vibration d'élongation des liaisons C=O (75 à 85%) et C-N. Sa fréquence est retrouvée entre 1600 et 1700 cm⁻¹. La position de la résonance dépend de la conformation de la chaîne principale et de l'organisation des liaisons hydrogène.
- <u>La bande amide II</u> : elle se retrouve dans la région de 1510 à 1580 cm⁻¹, et découle en grande partie de la torsion du vecteur N-H, des vibrations d'élongation des liaisons C-N et C-C.
- La bande amide III : c'est la plus complexe des quatre. Elle dépend de la nature des chaînes latérales et des liaisons hydrogène : elle est donc d'une utilisation très limitée pour l'obtention d'informations structurales.





Tout comme le DC, elle ne renseigne pas sur l'implication de tel résidu dans tel type de structure secondaire, mais permet de remonter jusqu'aux pourcentages de structures secondaires par déconvolution (Fig. 18). En effet, les pics de références ont pu être déterminés pour chaque type de structure secondaire, principalement à l'aide de peptides modèles tels la poly-L-Lysine.

La figure 19 montre les différences très visibles observables sur des spectres IR du lysozyme natif et dénaturé.



(source : http://www.imb-jena.de/ImgLibDoc/ftir/IMAGE_FTIR.html)

Par ailleurs, contrairement à beaucoup d'autres spectroscopies, elle offre l'avantage de travailler également avec des poudres, autrement dit sur des protéines précipitées. Son application peut se révéler très utile, comme par exemple dans l'étude des corps d'inclusion d'une sous unité d'une kinase (Luo, 1994) : ses structures secondaires étaient majoritairement constituées de brins β .

A.VI.5.b/ La spectroscopie Raman

La spectroscopie Raman se base, comme l'IR, sur les transitions entre niveaux de vibrations de liaisons dans une molécule. Son principe repose sur la diffusion inélastique de la lumière. Lors de la diffusion, le photon transfère une partie de son énergie à un mode de vibration de la molécule, et ressort donc de l'échantillon avec une longueur d'onde plus grande. Cet effet Raman est cependant très rare, et n'intervient que pour un photon parmi 10⁷ du faisceau incident : il est donc nécessaire d'utiliser des sources de lumière de très grande luminescence telles que le laser. Les angles de détection couramment utilisés sont de 90 et 180°. Enfin, il est important pour une meilleure

efficacité de la diffusion Raman, d'utiliser des longueurs d'onde ne provoquant ni absorption ni fluorescence des molécules.

La longueurs d'onde du pic de diffusion Raman dépend de la longueur d'onde d'excitation. En effet, l'énergie absorbée ΔE par une liaison donnée est toujours identique. La relation (4) implique :

$$\Delta E = h.c. \left(\frac{1}{\lambda_{exc}} - \frac{1}{\lambda_{Raman}} \right)$$
(6)

Par exemple, le pic de diffusion Raman des molécules d'eau est centré sur 309 nm si l'excitation est à 280 nm ; ou sur 397 nm si l'excitation est à 350 nm.

Le spectre Raman est obtenu en tracant l'intensité diffusée en fonction de ce décalage Raman ΔE . Les pics occupent des positions similaires à celles relevées sur un spectroscope IR, mais leurs intensités peuvent varier. Les différences majeures entre ces deux techniques sont détaillées cidessous :

- 1/ Bien que les fréquences d'absorption, pour une liaison donnée, soient les mêmes en Raman et en IR, les intensités des pics diffèrent. Par exemple, la vibration d'élongation de liaisons ioniques telles que O-H ou N-H est forte en IR, mais faible en Raman, et inversement pour des liaisons covalentes telles que C=C ou S-S.
- 2/ La spectroscopie Raman semble idéale pour les études de molécules biologiques puisque la diffusion Raman du solvant est faible, tandis que les fréquences infrarouge sont fortement absorbées par l'eau. Cependant, la spectroscopie IR est souvent nécessaire parce que beaucoup de vibrations ont une intensité très faible en Raman.
- 3/ L'utilisation d'un laser dans la spectroscopie Raman expose l'échantillon à des échauffements locaux et/ou à de la photo-décomposition.
- 4/ Les volumes utilisés en Raman, de l'ordre de 1 à 10 μl, sont inférieurs à ceux utilisés en IR (au mieux de quelques dizaines de μl).

Des études Raman peuvent se limiter à une comparaison de spectres en présence ou non d'un composé. Par exemple, les modifications structurales d'une protéine liant le zinc peuvent être suivies avec l'ajout de zinc ou d'un complexant du zinc.

L'estimation du pourcentage des structures secondaires est également réalisable avec une bonne efficacité pour les structures du type hélice et feuillet. Cependant, cette information structurale est assez complexe à obtenir : un décalage du spectre de 1 cm⁻¹ peut conduire à un décalage de 3% de

l'estimation de structures secondaires (Williams, 1983) : une référence externe sûre est donc requise.

Lorsque ces contraintes sont connues, l'utilisation de la spectroscopie Raman peut apporter des informations intéressantes. Elle a permis par exemple de mettre en évidence prédominance d'une structure en feuillet β dans le cas de la β -lactamase sous forme corps d'inclusion. (Przybycien, 1994).

A.VI.6/ La diffusion de lumière

Il existe principalement trois types de diffusion de lumière, apportant chacune une information particulière pouvant se révéler intéressante en biologie. La diffusion dynamique de lumière (DLS) renseigne sur le rayon hydrodynamique (rayon de Stokes) des particules en solution. La diffusion statique de lumière aux petits angles permet de remonter jusqu'au rayon de giration ainsi qu'à la masse moyenne pondérée <M> de ces particules. Cette masse moyenne <M> est également déterminable à partir d'une expérience de diffusion statique de lumière aux grands angles.

A.VI.6.a/ Diffusion Statique de Lumière (SLS pour « Static Light Scattering »)

• Diffusion de lumière et tailles des particules en solution

On sépare généralement les cas où les particules sont de taille très petite (facteur 1/10 à 1/20), comparable, ou très grande, par rapport à la longueur d'onde λ du faisceau incident. Nous nous limiterons au cas où les particules sont de petite taille face à λ . En effet, comme nous le verrons par la suite, des particules dont les tailles sont supérieures à quelques centaines de nanomètres (longueur d'onde typique des faisceaux incidents) ne restent pas longtemps en solution, mais tendent à sédimenter. Quant aux particules de taille intermédiaire (de l'ordre de grandeur de la longueur d'onde), il faut tenir compte de phénomènes d'interférences dont l'exploitation permet d'ailleurs d'obtenir des informations sur le rayon de giration et la forme de la particule. Ces aspects théoriques ont été développés par Debye en 1915 (Debye, 1915).

♦ Masse, rayon de giration et diffusion de lumière

L'intensité de lumière diffusée I(θ) est donnée par la relation de Zimm (Wen, 1996) :

K.c / I(
$$\theta$$
) = [1 + (16 π^2 / 3 λ^2).g²>.sin²(θ /2) + ϵ] / M + 2.A₂.c (7)

où
$$K = [4\pi^2 n^2 (dn/dc)^2] / (\lambda^4.Na)$$
 (8)

- avec c : concentration de l'échantillon en mg/ml,
 - λ : longueur d'onde dans le vide,
 - M : masse moyenne pondérée (nous revenons plus en détail sur cette masse plus loin),
 - R_g : rayon de giration,
 - A₂ : second coefficient de virial,
 - ϵ : reste du développement limité,
 - n : indice de réfraction,
 - dn/dc : incrément de l'indice de réfraction,
 - Na : nombre d'Avogadro.

Ainsi, en tracant K.c / I(θ) en fonction de sin²(θ /2) pour une concentration c donnée, l'extrapolation de l'ordonnée à l'origine est égale à 1/M et la pente à < R_g^2 >/M. A noter que la connaissance du rayon de giration peut aider à remonter à la forme de la molécule. Le paramètre K peut être déterminé à partir d'une solution de calibration du type toluène (Kunitani, 1997) ou BSA (Bauer, 2000).

Pour des particules en solution dont la masse est égale à environ 5.10^7 Da (Wen, 1996), le rayon de giration vaut approximativement 15 nm. Cette valeur permet de négliger le terme contenant $\langle R_g^2 \rangle$ dans l'expression (7). Par ailleurs, aux faibles concentrations, le coefficient du virial est lui aussi négligeable. La relation de Zimm se résume alors à (Wyatt, 1993) :

$$I(\theta) = K.c.M \tag{9}$$

et ce, quel que soit l'angle θ . Ainsi, plus la solution contient des macromolécules de grosses tailles (dans les limites déjà discutées par rapport à la longueur d'onde utilisée), plus la quantité de lumière diffusée sera importante. Cette quantité de lumière peut être mesurée directement avec un spectrofluorimètre à 90° du faisceau incident : c'est la *mesure directe de diffusion* (Kunitani, 1997). Ce type de montage permet de travailler avec des concentrations de l'ordre de quelques dizaines de nanomolaires. Si la solution contient des macromolécules de masses distinctes, la masse M déterminée par cette technique correspond à la moyenne des masses de chaque particule, pondérée par leur concentration relative (Doty & Edsall, 1951).

♦ Masse moyenne pondérée ([Doty & Edsall, 1951])

La masse moyenne pondérée des particules en solution est donnée par la relation (10) :

$$\langle M \rangle = \frac{\sum_{i} n_i . M_i^2}{\sum_{i} n_i . M_i}$$
(10)

M_i et n_i correspondent respectivement à la masse et au nombre de particules i en solution.

Ainsi, la masse moyenne pondérée sera identique, que nous ayons un mélange de 9900 particules de masse M = 10 et 1 particule de masse M' = 1000, ou bien un mélange de 3000 particules de masse M = 10 et 30 particules de masse M' = 30. Des multimères de 20 protéines chacun donneront là aussi la même masse moyenne pondérée. D'après la relation (9), pour peu que la concentration de protéines totale soit identique dans ces trois cas, le signal diffusé sera le même.

♦ *Précautions quant à l'utilisation de la diffusion de lumière.*

Une conséquence liée à la diffusion de lumière est une extinction de lumière dans la direction parallèle au faisceau incident. La mesure de cette diminution est appelée la *turbidimétrie*, et se traduit, de manière analogue à la loi de Beer-Lambert, par le facteur de turbidimétrie λ :

$$I = I_0.e^{-\lambda.l} \tag{11}$$

Préalablement à toute mesure de diffusion, il est nécessaire de s'assurer que toute diminution de lumière transmise est due uniquement à la diffusion et non à l'absorption, sans quoi l'intensité diffusée $I(\theta)$ sera sur-évaluée.

Dans le cas des mesures d'absorption ou de fluorescence, nous avions insisté sur le fait qu'il est nécessaire d'utiliser une solution de référence identique en composition à celle contenant la molécule d'intérêt. Ceci est encore plus important pour des mesures de SLS car le tampon peut avoir un effet au moins du même ordre de grandeur que celui mesuré pour la molécule d'intérêt. Une attention particulière doit être apportée à l'élimination de toute poussière en solution qui risquerait de fausser totalement la masse moyenne pondérée. Ceci est réalisé en utilisant par exemple des tampons fraîchement filtrés à base d'eau bidistillée, dé-ionisée, et éventuellement ultracentrifugés. Enfin, les fentes de sortie doivent être les plus fines possibles pour éviter toute contamination par rediffusion de lumière diffusée.

♦ Apports de la SLS en mesures directes

La quantité de lumière diffusée par une solution est proportionnelle à la fois à la concentration en protéine, et à la masse moyenne pondérée des particules. Cette technique peut être utilisée, dans un premier temps, pour comparer différentes solutions de molécules entres elles, en précisant si l'une d'elle contient des particules agrégées ou non. Suivant la précision apportée aux mesures (élimination de poussières) afin de diminuer le bruit de la mesure, l'agrégat se définira alors soit comme un multimère de quelques molécules chacun, soit comme un multimère de plusieurs centaines de molécules.

Si maintenant l'appareil a été calibré avec une solution de masse, d'indice de réfraction et de concentration connus, et après avoir éliminé les molécules agrégées voire précipitées, il est alors possible de différencier un dimère d'un monomère, et ce même à des concentrations de l'ordre du micromolaire [(Tanford, 1967), (Bauer, 2000)].

• exemple : mesure de l'agrégation dans le temps

Une protéine ayant tendance à perdre sa structuration au cours du temps risque donc d'exposer ses résidus hydrophobes au solvant. Ce phénomène entraine souvent l'agrégation ou la précipitation de la molécule. Une mesure de diffusion de lumière sur une échelle de temps adaptée permet donc d'appréhender ce phénomène (Fig. 20).



Fig. 20 : suivi de l'agrégation par diffusion de lumière au cours du temps pour la protéine sauvage α-synucléine et deux mutants [d'après (Li, 2001)].

A.VI.6.b/ La diffusion dynamique de lumière (DLS pour « Dynamic Light Scattering »)

La SLS est très sensible à la présence, même en très faible quantité, d'agrégats. Mais elle ne permet pas de rendre compte de l'homogénéité d'une solution car elle ne constitue qu'une mesure de la masse moyenne de toutes les populations présentes en solution. Au contraire, la diffusion dynamique de lumière (DLS) apporte des informations sur l'homogénéité d'une solution (présence ou non de sous-populations de taille différentes).

Des particules en solution suivent un mouvement brownien : une particule diffusera donc d'autant plus vite que sa taille est petite. Un faisceau laser est focalisé dans un volume très petit de l'ordre de quelques femtolitres. Les particules à l'intérieur de ce volume vont diffuser de manière inélastique la lumière dans toutes les directions et créer, par effet Doppler et à cause de la fluctuation du nombre de molécules dans le volume irradié, une distribution de longueurs d'onde. Les écarts de ces longueurs d'onde par rapport à celle du faisceau incident sont très faibles, rendant obligatoire l'utilisation d'un laser, source extrêmement monochromatique. Les concentrations requises pour ces expériences avoisinent la centaine de micromolaire afin d'avoir en permanence quelques particules au sein de ce volume.

L'enregistrement au cours du temps de l'intensité diffusée apparaît donc comme fluctuant de manière aléatoire. Pour extraire l'information de vitesse de diffusion contenue dans cette fluctuation, le traitement consiste à calculer une fonction d'autocorrélation $C(\tau)$ (Fig. 21). En effet :

$$C(\tau) = A.e^{-2\Gamma\tau} + B$$
⁽⁹⁾

avec : A et B des constantes liées à l'appareil, nécessitant une calibration préalable $\Gamma = q^2 \cdot D$

où q représente le vecteur de diffusion, et D le coefficient de diffusion.

$$q = [4\pi .n.sin(\theta/2)] / \lambda$$
 et $D = (k_BT) / (3\pi.\eta.d)$ (approximation sphérique)

n est l'indice de réfraction de la solution, θ est l'angle de diffusion, λ la longueur d'onde du faisceau laser, k_B la constante de Boltzmann, T la température en Kelvin, η la viscosité du liquide et d le diamètre de la particule (ou diamètre hydrodynamique).



Fig. 21 : intensité mesurée au cours du temps et fonction d'autocorrélation. Le signal fluctuant au cours du temps engendré par une population homogène en taille conduira à une décroissance monoexponentielle de la courbe d'autocorrélation.

Chaque population de particules homogène en taille produira sa propre courbe d'autocorrélation, c'est-à-dire une simple décroissance exponentielle. Un simple ajustement linéaire du logarithme népérien de $C(\tau)$ en fonction du temps de corrélation τ permet de remonter jusqu'au facteur Γ , et donc au rayon hydrodynamique de chaque population.

A.VI.7/ La Résonance Plasmonique de Surface (technologie BIAcore)

La reconnaissance moléculaire est au cœur des processus biologiques. La connaissance de ces processus passe par une description quantitative des interactions entre molécules. La Résonance Plasmonique de Surface permet de suivre en temps réel de telles interactions.

• Principes des mesures en temps réel

Lorsqu'un faisceau lumineux monochromatique illumine une surface, une partie est réfléchie tandis qu'une autre est transmise. Il existe un angle au delà duquel toute la lumière est intégralement réfléchie. Le système de détection optique développé par BIAcore (Biological Interactions Analysis) est calibré pour que le faisceau lumineux rencontre l'interface dans ces conditions, dites de réflexion interne totale. Une fine couche d'or, métal riche en électrons libres, permet sous ces conditions, d'induire le phénomène de Résonance Plasmonique de Surface (SPR) : les photons incidents entrent en résonance avec les électrons du métal, conduisant à un cône d'ombre, très fin, dans le faisceau réfléchi. L'angle de ce cône d'ombre varie avec l'indice de réfraction du milieu

constitué par la couche de 300 nm à l'interface avec l'or. Cet indice est proportionnel aux changements de masse locale résultant de la fixation de molécules sur la surface (Stenberg, 1991). L'enregistrement de la variation de l'angle permet donc de suivre en temps réel l'association ou la dissociation de molécules. Le signal de résonance est quantifié en unité de résonance ou RU. Une variation de 1000 RU correspond à une déviation de l'angle de 0,1° et à une fixation de 1 ng de protéine par mm² de surface. La figure 22 présente le principe de détection.

Dans une expérience de BIAcore, la première étape consiste en l'accrochage d'une molécule –le ligand– sur la surface. Cette surface est la plupart du temps constituée d'une couche de Dextran, permettant de mimer, en première approximation, une solution à l'interface. La seconde étape consiste à injecter une solution contenant la molécule –l'analyte– que l'on suppose interagir spécifiquement avec le ligand. Si l'interaction existe, elle entraînera une modification du signal. La figure 23 expose un cycle comportant l'accrochage, le décrochage de molécules à la surface, puis la régénération de cette surface. Si la régénération n'a pas altéré la fixation du ligand sur la surface, cette dernière peut être utilisée pendant près d'une centaine de cycles.

Même si le fonctionnement du BIAcore repose sur une méthode optique, il ne rentre pas à proprement parler dans les techniques spectroscopiques, puisque les molécules analysées ne sont pas excitées directement par le faisceau lumineux, contrairement à la fluorescence, au dichroïsme, ...



Fig. 22 : Détection des interactions moléculaires par résonance plasmonique de surface. La SPR induit une chute de l'intensité à un angle précis. Ce cône d'ombre dans le faisceau réfléchi varie avec l'indice de réfraction (et donc la masse) à proximité de la surface. Cette variation d'angle est transformée en un signal de résonance variant avec le temps : c'est le sensorgramme.



Fig 23 : L'enregistrement au cours du temps du signal de résonance. La ligne de base correspond au ligand déjà greffé à la matrice de Dextran. Lorsque l'analyte est injecté, la formation de complexes se traduit par une augmentation du signal de résonance. L'association fait place à la dissociation lorsque l'analyte est supprimé du tampon injecté. La phase de régénération consiste à supprimer de la surface toute molécule d'analyte, sans affecter pour autant la quantité de ligands fixés.

• Immobilisation du ligand à la surface

Les surfaces les plus couramment utilisées (« sensor chips ») possèdent des chaînes de Dextran d'une centaine de nanomètres de long servant à capturer le ligand et à le retenir sur la surface durant toute la mesure. La fixation du ligand sur le Dextran peut être covalente (immobilisation chimique), quasi-covalente (interaction avidine-biotine) ou réversible (interaction NTA-His-tag, interaction anticorps-antigène). Il est essentiel que la molécule de ligand ne se détache pas de façon détectable de la surface au cours de la mesure afin de ne pas compliquer les cinétiques d'interaction ligand / analyte. Il semblerait donc idéal de fixer le ligand de manière covalente ou quasi-covalente. Ces deux méthodes présentent cependant quelques inconvénients : i/ la surface ne peut pas être recyclée pour fixer un autre ligand ; ii/ les modifications chimiques nécessaires à la fixation peuvent affecter le site de liaison à l'analyte, voire le repliement de la protéine.

Les anticorps anti-GST (Gluthatione-S-transférase) constituent un exemple de capture de ligand permettant de contourner ce problème. Fixés covalemment à la surface, ils capteront toute protéine ou peptide fusionnée à la GST. Ce système est avantageux à plus d'un titre : i/ le ligand proprement dit reste libre pour interagir avec l'analyte. ii/ La sélection pour leur faible constante de dissociation des anticorps anti-GST monoclonaux utilisés assure une ligne de base stable durant l'expérience (généralement d'une durée inférieure à 24 heures). iii/ De telles surfaces sont recyclables par un flux rapide de tampon acide (pH 2) : ce tampon n'affecte pas en effet le repliement de l'anticorps, mais permet le décrochage des fusions de la surface (Quinn & O'Kennedy, 2001).

Toutes les fois où cela sera possible, on préferera choisir comme ligand le plus petit des deux partenaires. De cette manière, le signal observé lors de l'interaction ligand / analyte aura une plus grande amplitude.



♦ Analyse des courbes d'association et de dissociation

Fig. 24 : interaction de la novobiocine et d'un mutant de l'ADN gyrase suivie par SPR. Les enregistrements ont été obtenus avec des concentrations croissantes de novobiocine : 20 (courbe la plus basse), 40, 60, 80 et 100 nM (courbe la plus haute) [d'après (Kampranis, 1999)]. Une étude cinétique par BIAcore consiste, après accrochage du ligand sur la surface, à envoyer l'analyte à différentes concentrations. La figure 24 montre un exemple de sensorgrammes classiquement obtenus en BIAcore.

Dans le cas d'une interaction de Langmuir (A + B <==> AB), les réponses mesurées pour les phases d'association et de dissociation sont de la forme :

$$R_{ass} = \{(k_{on}.[C].R_{max}) / (k_{on}.[C] + k_{off})\} . [1 - exp\{-(k_{on}.[C]+k_{off}).t\}]$$

 $R_{dis} = R_0 \cdot exp\{-k_{off} \cdot t\}$

avec t le temps, R_{max} le signal maximal en RU et [C] la concentration de l'analyte.

L'ajustement linéaire des courbes dR_{dis} / dt en fonction de R_{dis} permet de déterminer la constante cinétique de dissociation k_{off} . Connaissant cette valeur et avec le même type d'opérations sur le signal dans la phase d'association, on calcule la constante cinétique d'association k_{on} .

♦ Les principaux modèles et leur spécificité

Modèle	Réactions				
1:1 (Langmuir) binding	$A + B \iff AB$				
analyte bivalent	$A + B \iff AB \text{ et } AB + B \iff AB^2$				
analyte hétérogène	$A_1 + B \iff A_1B \text{ et } A_2 + B \iff A_2B$				
ligand hétérogène	$A + B_1 \iff AB_1 \text{ et } A' + B_2 \iff A'B_2$				
dimérisation de l'analyte	$A + A \iff A^2 \text{et } A^2 + B \iff A^2 B$				
changement conformationel	$A + B \iff AB \iff AB \iff AB^*$				

Tableau 4 : résumé de quelques modèles rencontrés dans l'analyse de courbesd'interactions obtenues par Résonance Plasmonique de Surface. A et B représententrespectivement l'analyte et le ligand

Les interactions étudiées par SPR ne suivent pas toujours le modèle de Langmuir. D'autres modèles ont été développés, afin de répondre aux différents systèmes rencontrés (Tableau 4). Il est à noter toutefois que l'utilisation de ces modèles pour décrire ces résultats requière une précision accrue des données. Il faut être en effet certain que le modèle utilisé correspond effectivement au signal observé. Morton *et al.* ont décrit quelques pièges à éviter lors de l'acquisition des données (Morton & Myszka, 1998).

Le modèle de l'analyte bivalent correspond à un anticorps (l'analyte) voyant successivement un premier, puis un second antigène (le ligand). Les modèles hétérogènes rendent compte de l'existence simultanée de deux populations de molécules.

• Limite du transport de masse

Il existe des conditions dans lesquelles la vitesse de transport de l'analyte depuis le flux vers le ligand est plus lente que celle de la liaison avec ce ligand. Cet état, appelé situation de transport de masse, est amplifié par une quantité importante de ligands accrochés sur la surface, ou par une concentration trop faible d'analyte dans le flux. Il est donc préférable de limiter la quantité de ligands accroché à la surface lorsque l'on envisage une étude cinétique. Le coefficient de transport de masse dépend, lui, du coefficient de diffusion de l'analyte en solution, du flux de tampon et des dimensions de la cellule. En condition de transport de masse, la courbe d'association aux temps courts, est linéaire. Pour une même quantité de ligands fixés à la surface, l'étude des pentes des phases d'association mesurées pour des flux de tampon allant de 2 à 100 μ l/mn, permet de remonter jusqu'à la concentration active d'analyte (Christensen, 1997).

Cette mesure dépend de l'interaction étudiée. Une même protéine peut donc avoir des concentrations de protéines actives variant d'un système à l'autre.

♦ Analyses quantitative et qualitative par SPR

L'éventail des applications liées à la détection par Résonance Plasmonique de Surface d'interactions en temps réel est très vaste. Qualitativement, cette technologie permet de rechercher des ligands dans des extraits bruts, de cartographier les résidus primordiaux pour l'association ou la dissociation, de vérifier l'intégrité du repliement d'une protéine non purifiée. Quantitativement, elle informe sur les cinétiques d'interaction, l'affinité (K_D), les mécanismes et la stœchiométrie de l'interaction, la concentration active d'analytes, et permet même de remonter jusqu'aux paramètres thermodynamiques de l'interaction (énergie libre Δ G, enthalpie Δ H et entropie Δ S). D. Myzska, donne une vue générale bibliographique de toutes ces applications (Myszka, 1999).

A.VII/ Tableau récapitulatif des techniques utilisées

Déterminer la technique la mieux appropriée peut souvent paraître délicat. Pour cela, il est en fait important de se souvenir que la technique à utiliser dépend de la question que l'on se pose. Le tableau 5 donne pour différentes techniques les contraintes expérimentales et les informations que l'on peut espérer en déduire.

		informations obtenues sur				paramètres requis				
TECHNIQUES		structure tertiaire	structure secondaie	identification d'interaction	détermination stabilité	agrégation	pureté	homogénéité d'agrégation	difficulté pour utilisateur	concentration
BIAcore	extrait brut	-	-	+++	++	++	+	+	++	+
BIAcore	cinétiques	-	-	+++	++	++	+++	+++	++	+
ultracentrifugation		-	-	+	++	+++	+	+	+	++
spectro. de masse		-	-	+++	-	+	+	-	++	+
gel de protéine	dénaturant	-	-	-	-	-	+	-	+	+
gel de retard	natif	++	-	+++	-	++	++	+++	++	+
dichroïsme circulaire	lointain	-	++	+++	++	-	+++	++	++	++
dichroïsme circulaire	proche	++	-	+++	+	-	+++	++	++	++
fluorescence		-	-	++	++	-	+++	++	++	+
absorption		-	+	+	+	-	+++	++	+	++
diffusion de lumière	statique	-	-	++	+++	++	+++	+++	+	+
diffusion de lumière	dynamique	-	-	++	+++	+++	++	+	++	++
diffraction rayons X		+++	+++	+++	-	-	+++	+++	+++	+++
RMN		+++	+++	+++	++	+	+++	++	+++	+++

Tableau 5 : résumé des apports et des contraintes de quelques techniques expérimentales.Pour les colonnes "Information", le nombre de « + » est proportionnel à la puissante de latechnique ; au contraire, un tiret signifie que la technique n'est pas adaptée pour obtenir untetelle information. Dans les colonnes "paramètres", le nombre de « + » est proportionnel àl'importance requise de la contrainte.

CHAPITRE III : E6, LES VIRUS À PAPILLOMES HUMAINS ET LE CANCER DU COL DE L'UTERUS.

A.VIII/ les virus à papillome humains (HPV) et les cancers

Les Virus à Papillomes Humains (HPVs) sont des virus à double brin d'ADN, d'environ 7500 paires de bases, qui infectent les cellules des tissus épithéliaux. A ce jour, plus de 100 souches distinctes ont été identifiées. Dans les pays industrialisés, entre 10 et 20% des hommes et des femmes âgés de 15 à 49 ans sont infectés par au moins un type de papillomavirus. Dans les pays en voie de développement, ce taux est encore plus élevé, pouvant aller jusqu'à plus de 40% (Orth, 1997). Tous ces virus ont un même but : celui de se reproduire. Après infection, le virus utilise la machinerie de production protéique de la cellule-hôte à ses fins, principalement à l'aide de deux protéines précoces virales agissant en synergie, E6 et E7. Ce détournement cellulaire n'aura cependant pas les mêmes conséquences pathologiques suivant la souche virale. On distingue généralement deux classes de souches : les bénignes, dites à « bas-risque », et les malignes dites à « haut-risque ».

♦ les souches bénignes et malignes de HPV

Les souches bénignes représentent environ les deux tiers de ces souches, et leur infection provoque des condylomes à l'origine de la formation des verrues (expansion des cellules épithéliales de la peau). Par exemple les souches 6 et 11 à l'origine de verrues génitales.

Les souches malignes (HPV 16, 18 et dans une moindre mesure 31, 33 et 51 avec des variations géographiques), spécifiques des épithéliums des organes génitaux, sont transmises par voie sexuelle. Le virus y reste latent dans près de 90% des cas, ou bien s'y réplique. Ces infections sont la plupart du temps non persistantes, car les mécanismes de défense de l'organisme permettent de détecter et d'éliminer les particules virales produites. Cependant, certaines infections peuvent devenir persistantes, et les séquences de E6 et E7 –les deux protéines virales principales pour le virus– sont alors systématiquement intégrées au génome de la cellule infectée (Schwarz, 1985). Elles sont ainsi capables de générer la transformation et l'immortalisation des cellules infectées [(Münger, 1989), (Hawley-Nelson, 1989), (zur Hausen, 1991)]. Après une longue période de latence, ces cellules infectées présentent un risque : celui de passer de l'état de lésions cervicales
bénignes à des carcinomes cervicaux (cancers) [(zur Hausen, 2000), (Mathevet, 2001)]. Ainsi, en Europe, sur un échantillon de 1000 femmes, 200 sont contaminées par le HPV, et seule une développera un cancer. Le but principal du virus est donc bien de se répliquer, et non pas d'induire un cancer.

• génome et expression des différentes protéines de HPV

Il existe deux groupes de protéines de HPV : précoce (E1, E2, E4, E5, E6 et E7) et tardive (L1, et L2) (Fig. 25). Les protéines tardives sont des protéines structurales qui forment la capside du virus : elles sont donc exprimées plus tardivement dans le cycle de multiplication du virus.



Fig. 25 : organisation du génome de HPV 16. Sont localisées sur ce génome les régions : précoce encodant les protéines E1, E2, E4, E5, E6 et E7 ; tardive pour les protéines L1 et L2 (E pour early ; L pour Late) [adapté de (Dell & Gaston, 2001)].

Au contraire, les protéines précoces interviennent dans la régulation du cycle de la cellule hôte, dans le but de détourner la cellule à ses fins. E1 se lie avec l'ADN et serait essentielle pour la réplication du papillomavirus. E2 est connue comme un facteur de transcription régulant le génome viral. La contribution de E5 dans l'infection virale est encore mal comprise. Les deux protéines E6 et E7 induisent la microprolifération cellulaire nécessaire à la réplication de l'ADN viral et à la production de virions. Quant à E4, son rôle dans le cycle viral est encore inconnu à ce jour. D'excellentes revues existent sur ce sujet : [(Dell & Gaston, 2001), (zur Hausen, 1996, 2000)].

• spécificité des HPV dans les cancers des épithéliums

On sait aujourd'hui que le virus HPV est à l'origine de près de 99,7% des cas de cancer invasif du col de l'utérus (Walboomers, 1999), avec en premier lieu la souche de HPV16 (50%), puis HPV18. Ce cancer est, à l'échelle mondiale, le plus fréquent et le plus mortel chez les femmes après le cancer du sein. Chaque année, 470 000 nouveaux cas sont diagnostiqués, dont 80% dans les pays en voie de développement, provoquant plus de 200 000 décès, dont environ 3 000 en France (zur

Hausen, 2000), (Orth, 1997)]. De plus, il a été récemment démontré que les papillomavirus sont impliqués dans 50% des autres cancers génitaux, 50% des cas de cancer du rectum et 20% des cancers des voies aériennes supérieures (larynx et œsophage). Le virus HPV, associé à une forte exposition aux ultraviolets, peut aussi provoquer des cancers de la peau. Au total, environ 15% des cas de cancers au monde seraient donc associés à ces virus (zur Hausen, 1999, 2000).

A.IX/ Rôle joué par E6 et E7 dans la réplication des virus dits à « haut risque ».

Les deux oncoprotéines E6 et E7 induisent en synergie la microprolifération cellulaire nécessaire à la réplication de l'ADN viral et à la production des virions, en agissant sur des points-clés du cycle cellulaire (Kubbutat & Vousden, 1996) (Fig. 26). E7 interagit avec les protéines de la famille "Rb" (apparentées à la protéine du rétinoblastome), dont le fonctionnement normal est de contrôler négativement la progression du cycle. En inhibant ce contrôle, E7 permet de déclencher, dans la cellule infectée, la transition entre la phase G1 (stationnaire) et la phase S (synthèse d'ADN), nécessaire à la duplication de la cellule.



Fig. 26 : modèle de la coopérativité fonctionnelle des oncoprotéines E6 et E7 des HPVs à haut risque pendant la transformation cellulaire [d'après (Ristriani, 2001)].

p53 est une protéine cellulaire appelée « gardienne du génome ». Son rôle principal est de contrôler négativement les événements pouvant aboutir à la propagation d'une altération du matériel génétique. Dans une cellule saine, p53 est exprimée à un faible taux. Au contraire, lorsque des événements (dommages dus aux UV ou à des produits chimiques, etc ...) affectant l'intégrité de l'ADN se produisent, le taux de p53 augmente, induisant alors l'arrêt en phase G1 du cycle cellulaire

pour permettre la réparation des dommages. Lorsque la réparation est impossible, p53 provoque l'apoptose –ou mort cellulaire programmée– de la cellule dont le matériel génétique est endommagé. Dans le cas des cellules transfectées par E7 seule, celles-ci vont alors entrer en apoptose, en réponse à cette stimulation anormale provoquée par E7.

C'est ici que la protéine E6 des souches malignes intervient : elle interagit avec p53 et une autre protéine cellulaire, E6-AP, pour induire la dégradation de p53 (Scheffner, 1990, 1993). La protéine E6-AP est une enzyme du système d'ubiquitination, processus permettant la protéolyse sélective de protéines dont la présence n'est plus requise à certaines étapes du cycle cellulaire. La cible naturelle de E6-AP n'est connue que depuis 1999 : HHR23A (Kumar, 1999). Le papillomavirus a donc détourné E6-AP de sa fonction première. La formation du trimère protéique E6AP-E6-P53 aboutit à la dégradation anormale de p53 par ubiquitination, et par conséquent à une baisse notable du taux de p53 dans les cellules infectées. E6 bloque ainsi la réponse apoptotique. Comme le cycle cellulaire n'est plus arrêté, le virus peut se multiplier. Cette dégradation de p53 par E6 est un événement « secondaire » généré par le virus, mais qui peut se révéler dramatique pour la personne infectée, à savoir la perte de contrôle de l'intégrité du matériel génétique. Dans la plupart des cas, un dommage affectant l'ADN est létal pour la cellule et conduit à son élimination. D'autres, au contraire, n'ont aucun effet. Cependant, parmi ces rares mutations non létales, l'une d'elle peut entraîner l'immortalisation de la cellule infectée. Cette cellule ne connaît donc plus de limites à sa mulplication, donnant naissance à une tumeur. Ce modèle permet ainsi d'expliquer que seule une femme infectée par le HPV sur 200 développera un cancer du col de l'utérus.

Des facteurs tels que le tabac ou les UV ayant des effets mutagènes, augmentent donc réellement le risque de cancer dans des cellules préalablement infectées par les HPVs, puisque p53 n'est plus présent pour jouer son rôle de protecteur du génome. Notons également qu'un défaut de fonctionnement de p53 est retrouvé dans près de la moitié des cancers au monde, signe de l'importance de cette protéine dans les processus de régulation et de contrôle de l'intégrité du génome. D'où aussi son nom de protéine suppresseur de tumeurs.

Très récemment, Jiang & Milner, en utilisant des ARNs silencieux[#] spécifiques des séquences de E6 et/ou E7, ont démontré que l'induction de l'apoptose est plus efficace pour les cellules ciblées par

[#] un ARN silencieux est un petit fragment d'ARN, complémentaire d'une séquence spécifique d'un ARN messager. Ce fragment, hybridé à l'ARNm, bloque la progression du ribosome, et donc la traduction. L'expression de ce gène est donc mis « sous silence ».

l'ARN silencieux de E7, plutôt que par celui de E6 (Jiang & Milner, 2002). Ces résultats semblent donc contredire le modèle classiquement présenté dans la figure 26. Il est intéressant de noter aussi que « l'extinction » de E6 ramène p53 à un taux plus élevé dans la cellule. Les rôles de E6 et de E7 dans la cellule ne semblent donc pas encore réellement élucidés.

A.X/ Etat des connaissances sur la protéine E6 du virus HPV 16.

A.X.1/ Séquence de E6

E6 est une protéine basique de 158 acides aminés (point isoélectrique de 10,1). E6 contenant deux méthionines pouvant être utilisées en codon initiateur, elle peut également être exprimée sous la forme d'un polypeptide de 151 résidus. Les alignements des séquences de E6 d'un grand nombre des souches de HPV –un exemple est représenté figure 27 – suggère la présence de deux motifs de liaisons au zinc (Cole & Danos, 1987). Ces deux motifs sont constitués de 37 résidus chacun et contiennent 4 cystéines distribuées selon le schéma suivant : CxxC-(29x)-CxxC. Plusieurs études ont démontré que E6 fixe l'ion zinc [(Cole & Danos, 1987), (Lipari, 2001)].

La figure 28 est une représentation schématique de E6, dans laquelle on définit les régions de liaison au zinc N- et C-terminales.



Fig. 28 : représentation schématique de la protéine E6 de HPV 16 (numérotation correspondant à la forme de 158 acides aminés).

Е6	
DE	
N-TERMINALE	
RÉGION	

······································	•
MARP VKVCELAHHL NIPIWEVLLP CNFCTGFLTY QELLEFDYKD FNLLW-KDGF VFGCCAACAY RSAYHEFTNYHQE	HPV49–N
MDSTRP LTVQQLSDKL TVPVVDLLLP CRFCSRFLTY LELREFDYKH LQLIWTEEDF VFACCSGCAY ASAQFEIQQFYQL	HPV23-N
MDRPKP QTVRELADTL CIPLVDILLP CRFCNRFLAY IELVAFDLKG LQLIWTEEDF VFACCSSCAY ATAQYEFSKFYEQ	HPV17-N
TPELP TTIKELADLL DIPLVDCLVP CNFCGKFLDF LEVCDFDKKQ LTLIW-KGHF VTACCRSCCA ATAIYEFNEFYQQ	HPV12-N
MARFED-PTRRP YKLPDLCTEL NTSLQDIEIT CVYCKTVLEL TEVFEFAFKD LFVVY-RDSI PHAACHKCID FYSRIRELRHYSD	HPV18-N
MFFPN-PEERP YKLPALCEEV NISIHEIELD CVYCERQLYR CEVYDFIFRD LCVVY-RKGK PLGVCQPCLL FYSKVRQYRRYNQ	HPV34–N
MDRQLFEN-TEERP RTLHQLCEVV NKPLLELQLG CVFCKKALTA SEVYNFAYTD LRVVY-RDGY PYGVCKFCLL FYSKVRKLRYYNC	HPV53–N
MFED-PRERP RTLHELCESL NTTLQNLQVQ CVYCKETLQW ADVYNFAICD LRVVY-RDRS PYAACKRCVI FYSKITEYRRYTC	HPV26–N
MSEENPCP RNIFLLCKQY GLELEDLRLL CVYCRRALSD ADVLAFAIKE LSVVW-RKGF PFGACGKCLI AAGKLRQYRHWHY	HPV27-N
MPMGLHNP TNIMLLCKEI EVDLEDLRIT CIFCKNELTT EELLAIAIKE LQIVW-RDNW PFGVCAPCLA RATKVRELRYWTY	HPV72–N
MSMGAQEP RNILLLCRNC GIFLEDLRLC CIFCTKQLTA AELAAFALRE LYLVW-RAGV PYGACARCLL LQGIVRRLKYWDY	HPV10-N
MASTSASSQP STLYQLCKDF GLTLRNLQIC CIWCKNHLTS AEAYAYHFKD LHVVW-KKGF PYAACAFCLE FYSKVCALRHYDR	HPV32–N
MESKDASTSA TSIDQLCKTF NLSLHTLQIQ CVFCRNALTT AEIYAYAYKN LKVVW-RDNF PFAACACCLE LQGKINQYRHFNY	HPV11-N
MSAR-CGSQA RTLYELCDQC NITLFILQID CVFCKTVLKT AEVLAFAFRE LYVVW-RDDF PHAACPRCLD LHGKVNQYRNFRY	HPV40-N
MFQD-PQERP RKLPQLCTEL QTTIHDIILE CVYCKQQLLR REVYDFAFRD LCIVY-RDGN PYAVCDKCLK FYSKISEYRHYCY	HPV16-N

RÉGION C-TERMINALE DE E6

	סק קר קרצאדנאקד						
•	E	hh.	hR CCL	EK	Fh h.G.CC	•	
HPV16-C	HY <mark>C</mark> YS	LYGTTLEQQY	NKPLCDLLIR	CINCOKPLCP	EEKQRHLDKK QRFHN-IRGR	WTGRCMSCCR	SSRTRRETQL
HPV40-C	NFRYA	AYAPTVEEET	GLTILQVRIR	CCKCHKPLSP	VEKTNHIVKK TQFFK-LKDS	WTGYCLHCWK	KCMEKGQRSETLC
HPV11-C	HFNYA	AYAPTVEEET	NEDILKVLIR	CYLCHKPLCE	IEKLKHILGK ARFIK-LNNQ	WKGRCLHCWT	rCMEDLLP
HPV32-C	HYDRS	AFWHTVEQET	GLLLEEQIIR	CAICQKPLSP	SEKDHHIYNG RHFRF-ILNR	WTGRCTQCRE	
HPV72-C	SYTWY	GYGPTVEQET	GKSLAELYIR	CHACCKPLSC	DEKEYQVQTG IHFHK-ISGL	WTGRCCQCRG	ACTARWQP
HPV10-C	SYDWY	YYVEGVEEET	KQSIYTQLIR	CYMCHKPLVR	EEKDRHRNER RRLHK-ISGY	WRGSCEYCWS	RCTVRIPQ
HPV27-C	SΥHWH−−−−−−−	CYGDTVETET	GIPIPQLFMR	CYICHKPLSW	EEKEALLVGN KRFHN-ISGR	WTGHCMQCGS	TCTAPDPASRTLH
HPV26-C	RYTCS	VYGATLEALT	KKSLCNLLIR	CHRCOMPLGP	EEKQRIVDEK RRFHE-IAGQ	WKGLCTNCWR	PRRQTETQV
HPV34-C	RYNQS	VYGRTLENLT	NKQLCNILIR	CGKC QKPL C P	LEKQRHVDEN KRFHQ-IADQ	WTGRCTQCWR	PSATW
HPV53-C	YYNCS	VYGASLEALT	KKKLSDLSIR	CYRC QHPLTP	EEKQLHCDYK KRFHK-ISHM	WTGSCLTCWR	HTTATESAV
HPV18-C	BUSYH	VYGDTLEKLT	NTGLYNLLIR	CLRCQKPLNP	AEKLRHLNEK RRFHN-IAGH	YRGQCHSCCN	RARQERLQRRRETQV
HPV12-C	FYQQT	VLGRDIELAT	GKSIFDLKIR	CQTCLSFLDT	IEKLDSCGRG LPFHK-VRDR	WKGICRQCKH	LYLNNDR
HPV17-C	EVEQS	VSGRELEEIE	HKPIGEIPIR	CKFCLKKLDL	LEKLDTCYRH QQFHK-VRRN	WKGLCRHCGS	I GI G
HPV23-C	тлочт	VYGREIEQEE	QRPIGQI <mark>C</mark> IR	COYCLKSLDL	IEKLDICSFN QPFHK-VRNH	WKGRCRHCKE	ΤΕ
HPV49-C	VHQEI	VVGIEIEGRA	AANIAEIVVR	CLICLKRLDL	LEKLDICAQH REFHR-VRNR	WKGVCRHCRV	IE

lettre lorsque ce résidu est strictement conservé ; les minuscules, quant à elles, correspondent à : h pour hydrophobe, c pour chargé et alignées entre elles. Les lettres en majuscules, en-dessous de la région N-ter et au-dessus de la région C-ter, reprennent le code à une p pour petit. Les cystèines ont été colorées en rouge pour mettre en évidence leur forte proportion dans la séquence. E6 de HPV16 en Fig. 27: alignement multiple de E6 provenant de différentes souches de HPV. Les régions N- et C-terminales de E6 sont ensuite contient ainsi 15 pour un total de 158 résidus, soit donc près de 10%.

E6, HPV et cancers

A.X.2/ E6 dans la cellule

En plus de son activité de liaison à p53 et E6AP, E6 interagit avec une variété de protéines cellulaires : les co-activateurs transcriptionnels CBP / p300 [(Patel, 1999), (Zimmermann, 1999)] et ADA3 (Kumar, 2002), les facteurs de transcription c-Myc (Gross-Mesilaty, 1998) et IRF3 (Ronco, 1998), la protéine de réplication hMCM7 (Kukimoto, 1998), les protéines de réparation de l'ADN MGMT (Srivenugopal & Ali-Osman, 2002) et XRCC1 (Iftner, 2002), les protéines kinases PKN (Gao, 2000) et Tyk2 (Li, 1999), la protéine activatrice de la GTPase E6TP 1 (Gao, 1999), le récepteur du facteur de nécrose tumorale TNF-R1 (Filippova, 2002), la protéine apoptotique Bak (Thomas & Banks, 1998), le complexe adaptateur de clathrine AP-1 (Tong, 1998), un constituant à l'adhésion cellulaire la paxilline [(Tong & Howley, 1997), (Vande Pol, 1998)], les protéines de liaisons au calcium E6BP (Chen, 1995) et la fibuline-1 (Du, 2002), et enfin plusieurs membres de la famille des protéines "PDZ" comme hDLG (Kiyono, 1997), hScrib (Nakagawa & Huibregtse, 2000), MAGI-1 (Glaunsinger, 2000), MAGI-2 et MAGI-3 (Thomas, 2002), MUPP1 (Lee, 2000).

De plus, E6 induit l'activation ou la répression transcriptionnelle de plusieurs promoteurs cellulaires ou viraux [(Sedman, 1991), (Morosow, 1994), (Dey, 1997), (Ronco, 1998)]. En particulier, E6 active la transcription de la protéine retrotranscriptase hTERT de la télomérase humaine [(Gewin & Galloway, 2001), (Veldman, 2001), (Oh, 2001)].

Les différentes protéines avec lesquelles interagit E6 ont quelques similitudes entre elles. Toutes ont en effet été détectées par la technique du double hybride dans la levure. Par ailleurs, la plupart des régions minimales des protéines reconnaissant E6 (à l'exclusion notable de p53) ont été identifiées. Ainsi, pour les molécules telles que E6-AP, E6-BP, IRF-3 et la paxilline, la région minimale liant E6 se réduit à une famille de peptides contenant le consensus ELLG (Van de Pol, 1998). Ces peptides peuvent adopter une structure en hélice (Chen, 1998). Des domaines minimum de liaison à E6 ont aussi été définis pour les protéines CBP / p300, Bak, E6TP1, AP-1, hMCM7 et Tyk2. Enfin, plusieurs protéines-cibles se positionnent sur la même séquence de E6, une séquence de 5 résidus (RETQL) située à l'extrémité C-terminale. Ceci est notamment le cas pour les protéines hDLG, hScrib et celles de la classe des MAGI [(Gardiol, 1999), (Nakagawa & Huibregtse, 2000)]. Pour ces protéines, la région minimale nécessaire pour lier E6 est également connue. Il s'agit en fait de domaines "PDZ", une famille de modules adapteurs reconnaissant de courtes séquences peptidiques (Cowburn, 1997).

Ces cascades de protéines-cibles sont-elles donc le fruit d'une véritable interaction avec E6, ou bien la résultante d'une similitude de séquences ou d'un consensus en hélice α ? A ce jour, la réponse n'est pas connue.

A.X.3/ La difficulté d'obtenir des données biologiques pour E6

Depuis la mise en évidence en 1985 du caractère oncogène de la protéine E6 (Schwarz, 1985), de nombreuses études ont permis de proposer différentes voies d'actions pour la protéine quant à son rôle dans le processus tumoral. Cependant, il est étonnant de constater que l'importante masse d'observations qui ont été accumulées dans les 15 dernières années par les très nombreux laboratoires s'intéressant à E6 (près de 1500 publications à ce jour) concernent presque exclusivement les effets ex vivo et in vivo de la protéine exprimée dans différents systèmes (principalement la levure et les différentes lignées cellulaires et tumorales). Ce foisonnement d'informations contraste avec la rareté des données biochimiques, immunologiques et structurales concernant E6. À l'origine de ce décalage, le fait est que E6, malgré sa taille relativement faible (environ 150 acides aminés), a longtemps posé de sérieux problèmes d'expression et de purification. La protéine était en effet systématiquement insoluble ou hautement agrégée, indépendamment du système d'expression utilisé (Lechner & Laimins, 1994). Cette situation a non seulement rendu difficile toute étude structurale de la protéine seule ou avec ses partenaires, mais elle a aussi prévenu toute étude cinétique ou thermodynamique des interactions et activités de E6. La seule publication faisant état d'une purification de E6 ne mentionnait ni les tampons, ni les colonnes utilisés (Daniels, 1997).

En conséquence, il n'existe quasiment aucune stratégie de thérapie basée sur le blocage des activités oncogènes de E6. On peut juste citer quelques recherches actuellement en cours pour l'inhibition de E6 *in vivo* : des peptides aptamères capables d'induire l'apoptose dans des cellules tumorales (Butz, 2000), ou bien l'étude *in vitro* de molécules éjectant les ions zinc de la protéine E6 (Beerheide, 1999). Cette approche est d'ailleurs similaire à celle déjà développée dans le cas du HIV (Tummino, 1996). Enfin, quelques vaccins sont évoqués depuis plusieurs années.

MATÉRIELS ET MÉTHODES

B.I/ TECHNIQUES GENERALES D'EXPRESSION ET DE PURIFICATION DE PROTEINES

B.I.1/ Milieux de culture

♦ Milieu de culture LB (Luria Broth)

Ce milieu est utilisé comme milieu non marqué de culture de bactéries. Un litre de LB (Becton Dickinson Microbiology Systems) est composé de : 10g de digestion pancréatique de caséine, 5g d'extrait de levures, 5g de chlorure de sodium.

• Milieu minimum M9 ^{15}N

1 litre de ce milieu minimum contient : 8 g de Na₂HPO₄ (2 H₂O), 3 g de KH₂PO₄, 0,5 g de NaCl, 0,5 g de ¹⁵NH₄Cl, 8,3 g FeCl₃ (6 H₂O), 0,8 mg ZnCl₂, 0,1 mg CuCl₂, 0,1 mg CoCl₂ (6 H₂O), 0,1 mg H₃BO₃, 11 mg MnCl₂ (2 H₂O), 0,4% de glucose, 120 mg MgSO₄, 36,6 mg CaCl₂, 1 mg de biotine, 1 mg de thiamine.

Dans les deux cas, nous avons additionné les antibiotiques suivants : l'ampicilline à 100 μ g/ml pour les vecteurs pMal, la kanamycine à 15 μ g/ml pour les vecteurs pETM, et le chloramphenicole à 34 μ g/ml pour les vecteurs de co-expression de GroEl (193-335).

B.I.2/ Expressions bactériennes

♦ Souches bactériennes

Nous avons travaillé principalement avec deux types de souches bactériennes : la souche XL1-Blue pour l'amplification des vecteurs dérivant du pET et pMal-c2 (New England Biolabs), et pour l'expression des protéines codées dans pMal-c2 ; la souche BL21 DE3 (Invitrogen life technologies) pour l'expression de protéines codées par les vecteurs pET et pMal-c2.

Les protéines codées dans ces deux vecteurs peuvent être induites par l'ajout d'isopropyl-1- β -Dthiolgalactopyranoside, ou lactose. La différence qui nous intéressait entre ces deux souches est que BL21 DE3 ne posséde pas deux protéases (ompT et 10n) présentes au contraire dans XL1-Blue, conduisant à une diminution de la dégradation et donc à une augmentation de la production de protéine recombinante.

◆ Transformation par électroporation

40 µl de bactéries sont transformées par électroporation avec 70 – 100 ng du vecteur codant pour la protéine désirée, en utilisant un appareil Gene PulserTM (BioRad) réglé à 200 Ω , 2,5 V et 25 µFa. Elles sont ensuite reprises dans 1 ml de LB et stockées 1 h à 37°C. Les bactéries transformées sont culotées par centrifugation à 1800 x g, et resuspendues dans 1 ml du milieu utilisé ensuite comme préculture.

• Les conditions de préculture

Nous nous sommes affranchis de l'étape de culture sur Boite de Petri pour ne pas risquer de sélectionner des clones ayant un défaut d'expression ou de séquence. La solution de transformation est donc directement diluée dans une préculture au 1/200^{ème}. Les précultures en milieu minimum M9¹⁵N sont incubées pendant 3 jours à 25°C et 250 rpm. Celles en LB sont incubées toute une nuit durant à 37°C et 250 rpm.

♦ Les conditions de culture

Les précultures sont ensuite diluées au 1/40^{ème} dans un milieu de culture préalablement équilibré à 37°C. Les deux souches bactériennes sont ensuite traitées de manière distincte :

- XL1-Blue en milieu complet : la densité optique (DO) à 600 nm d'une culture placée sous agitation constante à 37°C est suivie jusqu'à une valeur d'environ 0,2. Après y avoir ajouté une solution de $Zn^{2+}SO_4^{2-}$ à 100 μ M final, la culture est alors transférée dans une armoire chauffante à 27°C jusqu'à une DO de 0,25 0,27, toujours sous agitation constante. L'induction de l'expression protéique se fait grâce à l'IPTG à 0,5 mM final. La valeur maximale que l'on peut espérer atteindre tourne autour de 1,2 à 1,5 de DO après 6 à 15 heures d'induction. Cette souche a été utilisée principalement pour la production de la protéine E6 ou de ses mutants.
- BL21 DE3 en milieu riche : ces bactéries sont induites (toujours avec de l'IPTG) à une DO plus élevée (de l'ordre de 0,6 à 0,7). Le milieu de culture reste à une température constante de 37°C, et contient déjà la quantité de zinc voulue. La durée de l'induction n'excède pas 3 heures de temps. Cette souche a été utilisée pour la production du domaine C-terminal de E6.

Après culture, les bactéries sont ensuite culotées dans des pots de 500 ml dans un rotor J.L.A. 10.500 monté sur une centrifugeuse Beckman (2200 x g). Les culots sont alors soit utilisés directement, soit stockés dans un congélateur à -80° C).

B.I.3/ Purifications de protéines

Dans la suite de ces protocoles, toutes les manipulations de protéines sont effectuées à 5°C. Par ailleurs, tous les tampons, filtrés, sont fraîchement dégasés sur une trompe à eau, puis bullés extensivement avec de l'argon pendant au moins 20 minutes afin de limiter au maximum les risques d'oxydation.

♦ Lyse des bactéries

Quelque soit le système de lyse utilisé (French Press ou sonification), le tampon de cassure des bactéries est composé de : Tris-HCl 50 mM pH 7,4, 150 mM NaCl, 2 mM DTT, 2,5 μ g/ml de Dnase, 2,5 μ g/ml de Rnase, 5% de glycérol, ainsi qu'une tablette pour 50 ml de tampon d'un cocktail sans EDTA d'inhibiteur de protéases (Roche). En moyenne, un litre de culture est resuspendu dans 50 ml de tampon.

La sonification est effectuée sur des volumes de 20 ml successifs avec une sonde de 13 mm d'un sonificateur Vibracell 72412 (Bioblock Scientific). Une puissance de 10 à 20% est appliquée pendant 2 minutes par cycles de 1s d'ultrasons et 1s de pause, puis l'échantillon est replacé sur glace pour éviter un échauffement trop important. Cette étape est répétée généralement 3 à 5 fois jusqu'à ce que la solution passe d'un état opaque à une phase trouble mais légèrement translucide. Lors de la sonification proprement dite, une attention toute particulière est apportée pour limiter absolument toute production de mousse afin de ne pas oxyder les protéines.

Les extraits bruts sont ensuite ultracentrifugés dans des tubes Corex à 30 000 rpm (80 000 x g au milieu du tube) dans un rotor Ti 70 monté dans une centrifugeuse L-70 (Beckman).

♦ *Purification de la protéine fusionnée à MBP : la colonne amylose*

Les surnageants sont filtrés sur des filtres à 0,22 μ M, puis déposés sur une colonne amylose, résine spécifique de la protéine MBP (Maltose Binding Protein). La colonne a été préalablement équilibrée dans un tampon composé de : Tris-HCl pH 6,8 50 mM, NaCl 150 mM, DTT 1mM (tampon T1). La colonne est ensuite lavée avec 8 à 10 volumes de tampon T2 (T1 + cocktail d'inhibiteurs de

protéases). La quantité d'antiprotéase décroît par pas successifs de 100 à 10 % de la concentration recommandée par le fabricant. Le décrochage de la fusion est assuré par l'ajout de maltose à une concentration de 10 mM.

• Séparation des molécules agrégées de celles monomériques : l'ultracentrifugation

Les expériences utilisent des tubes UltraClear[®] de 12,5 ml placés dans un rotor SW41 Ti tournant à 36 000 rpm (200 000 x g) sur une centrifugeuse L-70 (Beckman) réfrigérée à 5°C. Dans notre cas, j'ai constaté expérimentalement que la durée de rotation doit être supérieure à 12 heures, et que le rotor doit s'arrêter sans frein aucun pour éviter au maximum les mouvements de convection. Par sécurité, seuls les 10 ml supérieurs sont prélevés par un pipetage précautionneux (en évitant les tourbillons).

◆ Coupure de la fusion ; séparation de MBP et de la protéine E6, ou d'un de ses domaines. La digestion à la thrombine ou à la TEV se déroule à 5°C dans une centrifugeuse GR412 (Jouan) à 4000 rmp (3000 x g) jusqu'à coupure complète (entre 6 et 36 heures). La solution est ensuite concentrée jusqu'à un volume inférieur à 2 ml puis chargée sur une colonne gel-filtration HiLoad 16/60 Superdex 75 (gamme de séparation entre 3 et 70 kDa) (Amersham Biosciences) préalablement équilibrée avec 240 ml de tampon T1. L'élution de la protéine E6 ou du domaine E6-

C conduit à un pic unique dont le volume correspond à celui attendu pour un monomère d'après la calibration de la colonne.

B.I.4/ Concentrations de protéines

Il existe trois types de concentrateurs (Fig. 29). Tous trois ont été testés, mais seul celui à membrane oblique a permis de limiter la précipitation et l'adsorption de protéines sur les membranes (filtres de 15 ml UltraFree Biomax 5krpm à membrane NMLW de Millipore (MUB 5/15) avec une accélération maximale de 1500 x g).



Fig. 29 : les trois types de concentrateurs. La solution à concentrer (le concentrat) est représentée en gris sombre, l'éluat en gris clair. a/ le concentrat est situé au dessus ; l'éluat est chassé vers le bas par simple centrifugation. b/ le concentrat est placé en dessous ; la rotation presse le réceptacle de l'éluat sur le concentrat, chassant ainsi l'éluat vers le haut. c/ ce type de concentrateurs reprend le principe de a/ mais en utilisant cette fois-ci une membrane oblique : cette construction permet de limiter l'adsorption de protéines sur les membranes de filtration.

B.I.5/ Méthode de séparation de la protéine agrégée de celle monodiperse

Le domaine C-terminal de E6, purifié selon le protocole décrit précédemment, est monodiperse et ne nécessite pas de réaliser l'étape supplémentaire que constitue la colonne échangeuse d'ions comme dans le cas de E6 entière. Cependant, lors de la purification de la protéine E6 entière, il n'était pas rare que l'éluat d'une colonne amylose contienne des fusions agrégées (mais restant toujours en solution) de par leur forte sensibilité à l'agrégation.

Pourtant, le bon déroulement de la purification nécessite absolument l'élimination de ces agrégats. Pour y parvenir, nous avions, dans les publications 1 et 2, utilisé l'une des propriétés permettant de distinguer ces deux types de particules : leurs charges exposées. Ceci nous avait orienté vers l'utilisation de colonnes échangeuses d'ions. Par la suite, nous avons utilisé plutôt leur différence de masses –et donc une ultracentrifugation– pour les séparer. En effet, la vitesse de sédimentation peut se calculer en fonction de la constante de sédimentation s, de la vitesse de rotation ω , et de la distance x par rapport à l'axe de rotation, par l'expression suivante :

$$v_s = s.\omega^2.x$$

Par ailleurs, la constante de diffusion D_t permet de déterminer la distance moyenne parcourue pendant un temps T dans une direction donnée. L'idée est ensuite de trouver la vitesse de rotation angulaire pour laquelle la vitesse de sédimentation sera plus rapide que la vitesse de diffusion dans le cas des agrégats, tandis que ces vitesses seraient du même ordre de grandeur pour les particules monomériques. Il est évident que cette vitesse de rotation est faussée par les approximations utilisées pour la détermination de D_t et de v_s (notamment : molécule de forme sphérique, viscosité identique en toutes positions dans le tube). Cependant, le fait de répéter l'ultracentrifugation pendant des durées variant de 1 à 36 heures augmente les chances de réussir cette séparation.

Nous avons montré que cette méthode, applicable plus particulièrement à la protéine E6 entière, permet de supprimer l'étape de la colonne échangeuse d'ions. Ainsi, la figure 30 montre le rapport d'agrégation (Cf. paragraphe suivant) et la valeur de la DO (ramenée à 1 ml) en fonction de la position dans le tube après centrifugation à 200 000 x g durant 16 heures pour une solution de fusion MBP-E6. On y constate que les courbes du rapport d'agrégation et de DO se superposent bien.



Fig. 30 : mesures d'agrégation (Rag) et de concentration (DO) après ultracentrifugation. Un culot très net apparaît au fond du tube lorsque la solution contient initialement des agrégats. Dans le tube de centrifugation représenté à gauche, le gris sombre correspond à une plus forte concentration de protéines. Ces molécules, d'après le R_{ag} lu sur la figure de droite, se présentent sous une forme agrégée.

Dans le cas de MBP-E6 6C/6S, les dix « premiers » ml renferment environ 20% de la quantité de protéine totale avant centrifugation, tandis qu'une partie non négligeable se trouve enfermée dans le culot. Cette quantité avait déjà été établie en analysant sur gel de protéines le dépôt et l'éluat d'une expérience sur colonne échangeuse d'ions (Publication 2).

B.II/ TECHNIQUES GENERALES DE SPECTROSCOPIES (AUTRES QUE LA RMN)

B.II.1/ Spectrofluorimètrie

Les expériences de spectrofluorimètrie ont été conduites sur un appareil SPEX Fluorolog-2 (SPEX Industries Inc. Edison, NJ) équipé d'une lampe Xenon de 450 W, un monochromateur d'émission, un monochromateur d'excitation, tous trois montés en L, ainsi que de deux photomultiplicateurs (PM) dont un pour la référence (placé après le monochromateur d'entrée). 4 fenêtres sont à régler : une avant (a), et une (b) après le monochromateur d'entrée, une avant (c) et une après (d) le monochromateur de sortie (Fig. 31).



Les largeurs de ces fenêtres étant directement reliées à la bande passante en nanomètre, les ouvertures sont exprimées en nm. Ainsi, une mesure de fluorescence utilise les réglages suivants : a/ 16 nm ; b/ 7,2 nm ; c/ 3,6 nm ; d/ 7,2 nm. La finesse de la fenêtre d'entrée sur le monochromateur d'excitation permet une bonne sélectivité de la longueur d'onde. Les réglages pour une mesure de diffusion diffèrent légèrement : a/ 16 nm ; b/ 3,6 nm ; d/ 7,2 nm. Quant à c/, son ouverture est réglée lors de la première mesure de manière à obtenir un signal de référence reproductible. Ceci permet de comparer directement les valeurs obtenues lors de différentes manipulations. Cette ouverture est généralement de l'ordre de 1 nm, générant un signal de $0,5.10^5$ coups par seconde (cps) : ceci laisse donc une marge importante de mesures de diffusion, le PM saturant à 10^7 cps. Quelle que soit l'ouverture de ces fenêtres, il est impératif de vérifier la non saturation des PM : la haute tension (HV) appliquée aux PM (permettant l'amplification du signal reçu) doit être ajustée de telle sorte

que le signal en sortie du PM se situe aux alentours de 1 μ A. Dans notre cas, les hautes tensions des PM de référence et de sortie sont respectivement de l'ordre de 350 et 800 V.

Avant chaque expérience (fluorescence ou diffusion de lumière), l'appareil est calibré en longueur d'onde. Un premier spectre d'excitation (variation de la longueur d'onde sur le monochromateur de sortie) est enregistré avec une solution tampon : celui-ci permet de connaître le décalage en longueur d'onde entre les monochromateurs d'entrée et de sortie (le signal diffusé doit avoir la même longueur d'onde que celui de l'émission). Un spectre d'émission (variation de la longueur d'onde sur le monochromateur d'entrée) est ensuite enregistré sur une solution contenant des résidus aromatiques (avec une préférence pour le tryptophane). En plaçant le monochromateur de sortie à 340 nm, le maximum d'émission doit se situer à une longueur d'onde de 279 nm. Ces deux spectres permettent ainsi de calibrer de manière absolue les deux monochromateurs.

♦ Mesures de fluorescence

Cette mesure est réalisée à l'aide d'un spectre d'excitation : l'émission étant réglée généralement à 295 nm, l'excitation entre 305 et 400 nm avec un pas de 0,5 nm et un temps d'intégration de 1s. La référence est soustraite à ce spectre.

• Mesures de diffusion de lumière

Pour cette mesure, les longueurs d'ondes d'émission et d'excitation sont conservées à une valeur constante de 350 nm, les molécules biologiques n'absorbant pas ou quasiment pas à cette longueur d'onde. La mesure de diffusion est obtenue en enregistrant la valeur de la diffusion toutes les secondes. Le fait de pipeter dans la solution entraîne souvent des perturbations : il est donc nécessaire d'attendre un minimum de 100s pour être sûr de revenir à l'équilibre. La valeur de la diffusion de la solution sera donc la valeur de ce signal à laquelle nous aurons soustrait la valeur de diffusion de la solution seule.

• Mesures du rapport d'agrégation

La valeur de la diffusion dépend de la concentration en protéine de la solution. Au contraire, le rapport R_{agg} , défini comme ci-dessous, est quasi indépendant de cette concentration :

$$R_{agg} = \frac{\Delta I_{diff}^{350}}{\Delta I_{fluo}}$$
(12)

Le numérateur correspond à la mesure de diffusion de lumière à 350 nm, référence soustraite ; le dénominateur au maximum de fluorescence, référence soustraite. Ce maximum, atteint à une longueur d'onde dépendant de la protéine, varie généralement entre 325 et 350 nm.

Ce rapport n'est indépendant de la concentration en protéines qu'en première approximation. En effet, l'agrégation des protéines risque de perturber le signal de fluorescence, voire même d'entraîner la précipitation des particules, modifiant ainsi la concentration totale de protéines en solution. Cependant, cette indépendance de la concentration de protéine peut être vérifiée en mesurant le rapport à deux concentrations différentes : il ne doit pas varier au delà de l'erreur expérimentale (déterminée en fonction du bruit de fond). Il est important de noter que ce rapport dépend de chaque appareil, de l'utilisateur et de la protéine étudiée. Il est donc nécessaire de calibrer initialement l'appareil avec une solution de protéines monodisperses, et une seconde de particules agrégées à la même concentration de protéines.

Il existe une relation linéaire entre l'intensité de lumière diffusée et la masse des particules en solution. Ainsi, dans notre cas, connaissant les intensités diffusées et les masses des protéines pour une protéine monodisperse ($\Delta I = 0, 1.10^5$, M) et une particule agrégée ($\Delta I = 5, 0.10^5$, M' ≈ 300 x M déterminée par DLS), la connaissance de l'erreur expérimentale ($0, 1.10^5$) nous permet de remonter jusqu'à la taille minimale des agrégats, soit donc de l'ordre de 10 à 15 monomères par particules agrégée. Il serait possible d'améliorer cette sensibilité en utilisant des solutions acqueuses dé-ionisées, et en augmentant l'ouverture des fenêtres du PM de sortie. Cependant, une faible ouverture a l'avantage de laisser une plus grande marge de manœuvre dans la détection d'agrégats.

B.II.2/ Dichroïsme circulaire

Les spectres de dichroïsme circulaire ont été enregistrés sur un modèle de dichrographe CD6 (Jobin Yvon, France) entre 260 nm et la valeur la plus basse permise par l'absorption (généralement comprise entre 195 et 188 nm). Des cellules en quartz de 0,1 mm sont utilisées ; l'intervalle d'enregistrement est de 0,5 nm ; le temps d'intégration est de 5s avant 210 nm, de 2s dans la région au delà de 210 nm. La cuve est conservée à une température constante de $15^{\circ}C$ ($\pm 0,5^{\circ}C$) par une circulation permanente d'eau. Tous les spectres sont la moyenne d'un minimum de 3 spectres, ce nombre pouvant atteindre 10 pour les échantillons ayant un mauvais rapport signal / bruit. La référence est systématiquement soustraite à ces signaux.

B.II.3/ Dénaturation de protéines

♦ Aspects généraux

L'agrégation étant un phénomène consécutif à la dénaturation de la protéine, il est préférable d'évaluer la dénaturation par le suivi de l'exposition du tryptophane au solvant, plutôt que par des mesures de diffusion de lumière.

Les expériences de dénaturation ont été réalisées principalement sur le spectrofluorimètre SPEX Fluorolog-2 (SPEX Industries Inc. Edison, NJ), plus adapté à des mesures de dénaturation en présence d'agents type urée ou chlorure de guanidium, que pour des mesures par dénaturation thermique. Une contrainte liée à l'utilisation de ces agents est que l'ajout d'une quantité de poudre modifie de manière importante le volume de la solution : les régles de proportionnalités dans les dilutions ne peuvent plus s'appliquer directement ici. Le tableau ci-dessous donne la démarche utilisée pour calculer la concentration d'une solution après dilution de y grammes de poudre dans z grammes de solution.

La première étape consiste à déterminer le rapport massique (en %) :

$$W = \frac{1}{100} \cdot \frac{y}{y+z}$$

	Urée	Chlorure de Guanidium
masse molaire	60,06 g/mol	95,5 g/mol
solubilité (à 25°C)	10,49 M	8,54 M
densité relative d/d ₀	$= 1 + 0,2658.W + 0,033 W^2$	$= 1 + 0,271.W + 0,033 W^2$

g de poudre / g H_20 pour préparer une solution de :

8 M	0,755	1,816
10 M	1,103	

On constate donc bien dans ce tableau la non linéarité entre la masse de poudre pour obtenir une solution à 8 M urée de celle pour 10 M, liée à cette non proportionnalité.

Prenons comme exemple la dilution de 6 g d'urée dans 5,2 g d'eau.

W = 52,2 %, et donc $d/d_0 = 1,148$.

Le volume final après dilution totale sera donc : (6 + 5,2) / 1,148 = 9,99 ml. Ce volume final attendu permet de connaître la quantité de poudre du tampon à rajouter à la solution pour obtenir la concentration désirée. La concentration en urée est alors de 10,1 M.

La concentration saline est réajustée après chaque dilution d'urée. La mesure de diffusion de lumière est réalisée à 2 reprises, à 5 minutes d'intervalle afin de s'assurer que l'équilibre a été atteint : si aucune variation n'est observée, le point obtenu est validé. La cuve est thermostatée à 15°C tout au long de l'expérience. La fluorescence est mesurée en enregistrant l'intensité maximale et la longueur d'onde de ce maximum pour une excitation à 295 nm, sélective du tryptophane.

Deux cas sont différenciés selon la quantité de protéines à disposition : les contraintes pour chacun des deux cas sont détaillées ci-dessous.

• Cas de la protéine en quantite suffisante

Une nouvelle solution de référence est utilisée pour chaque point en urée, obtenue après dilution d'une solution stock d'urée à une concentration minimale de 10 M. Après ajout de la quantité désirée de protéines, une attente de 5 minutes avant d'effectuer la mesure permet de s'assurer que le système est à l'équilibre. Le volume total de protéines nécessaire est donc égal au volume utilisé pour chaque point, multiplié par le nombre de points voulus pour la courbe de dénaturation.

◆ Cas de la protéine en quantité limitée

La même solution de protéine est conservée tout au long de l'expérience, en y diluant la quantité de poudre d'urée nécessaire au point suivant. Des masses de 60 mg d'urée –mesurées avec une balance de précision– diluées successivement dans la solution initialement de 2 ml, constituent un bon compromis temps passé / nombre de points obtenus.

Connaissant les masses d'urée ajoutées, la détermination de la concentration d'urée en chaque point est calculée en mesurant le nouveau volume de la solution à l'aide de micropipettes. Cette manipulation, permettant d'ailleurs une homogénéisation de la solution, est effectuée deux fois.

Ces deux protocoles peuvent être ajustés pour l'utilisation d'une micro-cuve de 150 µl.

• Traitement des courbes de dénaturation

Le traitement de données est effectué sous le logiciel MatLab. Les valeurs de longueur d'onde –ou d'intensité de fluorescence– sont converties en fraction de protéines dénaturées. Le signal réel de la protéine tient compte de la dilution induite par l'ajout d'urée, et du signal propre de la solution d'urée (ou du chlorure de guanidium). Théoriquement, une excitation à 295 nm n'induit pas de fluorescence pour l'urée ou la guanidine. Cependant, les impuretés, même à l'état de trace dns les poudres, ne peuvent plus être totalement négligées lorsqu'on employe des solutions d'urée fortement concentrées.

Enfin, de fortes concentrations d'urée diffusant largement la lumière, il est nécessaire de soustraire l'intensité diffusée pour l'urée seule au signal mesuré pour la protéine.

B.III/ TECHNIQUES GENERALES DE RMN

B.III.1/ Acquisition des spectres RMN

Les spectres de RMN ont été enregistrés sur un tube Shigemi contenant un échantillon de E6-C 4C/4S à 1 mM, marqué ou non à l'azote 15 (Tris deutéré 20 mM, pH 6,8, NaCl 50 mM, DTT 1 mM, D₂0 10%).

Les données ont été enregistrées sur un spectromètre Bruker DR600 pour les expériences 2D homonucléaires, et sur un spectromètre Bruker DR500 pour les expériences 2D et 3D hétéronucléaires. Les séquences d'impulsions utilisées pour les mesures de la vitesse de relaxation transversale $R_N(N_{x,y})$ et des NOEs hétéronucléaires {1H}-¹⁵N sont basées sur des expériences hétéronucléaires détectées sur le proton (Kay, 1989). Le signal de l'eau est supprimé à l'aide de la séquence WATERGATE avant l'acquisition sur le proton (Piotto, 1992). Pendant l'acquisition des expériences hétéronucléaires, l'utilisation de la séquence GARP assure le découplage de l'azote N¹⁵ (Shaka, 1985).

Les expériences 2D homonucléaires TOCSY (Bax & Davis, 1985) (temps de mélange de 60 et 80 ms avec un cycle d'impulsions composites DIPSI), et NOESY (Kumar, 1980) (temps de mélange de 200 ms) ont été acquises à 5 et 15°C avec une largeur spectrale de 8 kHz (soit 13,3 ppm) dans les deux dimensions. 48 transitoires de 2048 points ont été accumulées pour chacun des 512 points dans la dimension f1. Le délai de relaxation était fixé à 2,25 s.

Les expériences hétéronucléaires 3D N¹⁵-TOCSY-HSQC (temps de mélange de 60 ms) et N¹⁵-NOESY-HSQC (temps de mélange de 200 ms) ont été enregistrées à 15°C avec des largeurs spectrales de 7,5 kHz (soit 15,4 ppm) dans les deux dimensions proton, et 1,7 kHz (soit 33,7 ppm) dans la dimension azote. 8 transitoires de 1024 points ont été accumulées pour chacun des 240 (proton) X 64 (azote) FID. Le noyau N¹⁵ est découplé du proton durant l'acquisition. Le délai de relaxation était fixé à 2,5 s.

Les mesures de relaxation sur E6-C 4C/4S₈₇₋₁₅₈ ont été enregistrées avec un total de 200 incréments t1 et 16 transitoires de 2048 points par incrément. Le délai de relaxation était réglé à 2,5 s ; les largeurs spectrales de 6 kHz pour le proton et 1,7 kHz pour l'azote. Pour la mesure de $R_N(N_{x,y})$, 13 expériences ont été enregistrées avec les délais de relaxation de 16, 32, 48, 63, 79, 101, 143, 175, 207, 238, 286, 318, 365 ms. Le point à 79 ms a été enregistré une seconde fois afin d'estimer le niveau de bruit.

Les NOEs hétéronucléaires {1H}-¹⁵N sont mesurés à l'aide de deux expériences, avec et sans la saturation du proton. Chaque spectre 2D a été enregistré avec 32 transitoires de 2048 points pour chacun des 200 FID. Le délai de relaxation était fixé à 3,0 s.

Les spectres sont référencés sur le DSS pour les dimensions proton. La référence azote a été calculée à partir de la référence proton et des rapports gyromagnétiques du proton et de l'azote. Les spectres 2D et 3D ont été traités respectivement par les programmes UXNMR (Bruker Inc.) et NMRPipe (Delaglio, 1995).

B.III.2/ Calculs de structure

• Collecte des contraintes de distances

L'attribution manuelle, ainsi que l'intégration des pics, ont été réalisées à l'aide du programme XEASY (Bartels, 1995). L'attribution des résonances des protons de E6-C 4C/4S a été réalisée en utilisant tout d'abord la méthode classique de Wüthrich (Wüthrich, 1986) sur les spectres 2D homonucléaires. Celle-ci consiste dans un premier temps à identifier les différents systèmes de spins sur les cartes TOCSY, puis à corréler ces systèmes de spins entre eux par l'observation de proximités spatiales $H_{\alpha}(i) - H_N(i+1)$ et $H_N(i) - H_N(i+1)$ sur les cartes NOESY. Ces corrélations étaient validées par leur observation simultanée sur les spectres à 15°C et à 5°C. Un report de cette première attribution sur les cartes 3D hétéronucléaires nous a permis de lever les ambiguités qui subsistaient du fait d'un grand recouvrement de pics. Les contraintes utilisées dans les calculs de

structure proviennent systématiquement des cartes 3D pour les taches de corrélation bien résolues. Toutes les autres ont été extraites des spectres 2D à 15°C. L'attribution séquentielle des 3 prolines a été obtenue par les effets nOes observés entre les protons H_{δ} des prolines et les protons H_{α} du résidu précédent, ainsi qu'entre les protons H_{α} des prolines et le H_N du résidu suivant. Pour chacun des deux spectres NOESY 2D et 3D, le classement des intensités en trois catégories fort, moyen, faible a permis d'obtenir un premier jeu de distances limites supérieures classées en 2.5, 3.7 et 5.0 Å. Les déplacements chimiques du domaine E6-C 4C/4S sont reportés en annexe de ce manuscrit.

L'utilisation de ces contraintes de distances dans les calculs de structure a révélé une incohérence de ce premier jeu, liée à des erreurs d'attributions. Celles-ci étaient principalement dues à des élargissements localisés de raies pour certaines parties de la protéine, conduisant à une absence de convergence du protocole de recuit simulé (SA) (Nilges, 1988, 1999) calculé dans le logiciel XPLOR 3.8 (Brünger, 1992). Cette observation, ainsi que la difficulté à identifier ces erreurs nous a conduit à modifier légèrement la stratégie de calcul de structure. Celle-ci repose sur l'identification préalable de régions de la protéine pour lesquelles l'ensemble des observations RMN indique de façon fiable la présence de structures II. Cet ensemble cohérent de mesures comprenant :

- les déplacements chimiques secondaires : la méthode de Wishart (Wishart, 1992), appliquée avec la base de données établie par Mertuka (Merutka, 1995), a permis d'identifier les résidus susceptibles de se situer dans une structure de type α ou β ,
- les angles dièdres ϕ ont été déduits de la mesure des constantes de couplage ${}^{3}J_{HN-H_{\alpha}}$ déterminées à partir d'un spectre HNHA (Vuister & Bax, 1993),
- une recherche systématique des corrélations nOes caractéristiques de structures type α ou β (Wüthrich, 1986) est effectuée. Cette recherche a permis de fixer les trois brins du feuillet β les uns par rapport aux autres. Les contacts H_{α}/H_{α} ont été obtenus grâce à un spectre NOESY D₂O à 15°C,

a été utilisé pour définir un jeu de contraintes « dures » permettant de rigidifier ces régions. Les premières structures calculées ont permis également d'identifier de manière non ambiguë les 4 Cys conservées dans les alignements comme étant les coordinants du zinc. Ceci a justifié l'ajout de 10 contraintes de distances entre ces 4 cystéines et le zinc.

• Procédure itérative sur le recuit simulé (SA)

Une procédure de calcul itérative a ensuite été utilisée pour assurer progressivement la cohérence de l'ensemble du jeu de données, puis identifier les erreurs d'attribution responsables de la non convergence (Fig. 32).



Fig. 32 : représentation schématique de la procédure itérative d'analyse des contraintes. Les traits en pointillés correspondent à l'entrée et à la sortie de la procédure.

Cette procédure se déroule au maximum en 10 boucles. A chaque boucle, l'ordinateur calcule un jeu de structures, analyse les violations[#], modifie en conséquence les contraintes expérimentales, puis relance le calcul d'un nouveau jeu de structures. Si les jeux de contraintes de deux tours successifs sont identiques, le programme s'arrête.

Cette méthode a été utilisée à deux reprises : dans un premier temps pour détecter des erreurs d'attribution, et dans un second temps pour affiner la calibration.

[#] Une contrainte de distance est violée lorsque la distance cible correspondant à cette contrainte n'est pas respectée dans les structures calculées.

Dans cette première étape, la contrainte est supprimée des fichiers utilisés dans le tour suivant si la contrainte est violée dans plus de 6 structures parmi les 30 calculées pour chaque jeu, et si la moyenne des écarts entre la distance cible et la distance calculée est supérieure à 0.5 Å. L'analyse des contraintes exclues du calcul par cette procédure itérative nous a permis d'identifier les erreurs d'attribution en utilisant un concept analogue à celui du « Network anchoring » utilisé dans NOAH. Ainsi, lorsque plusieurs contraintes revenaient systématiquement sur un même résidu, ceci était généralement corrélé avec une erreur d'attribution de ce résidu. Au contraire, lorsque les contraintes supprimées se répartissaient dans une région confinée de l'espace de la structure, l'analyse de l'ensemble des contraintes de distance non supprimées de cette région, révélait généralement plusieurs d'attribution, ces erreurs ayant entrainées des violations sur des contraintes avoisinantes, et non sur elles-mêmes. Ceci a concerné environ 10% des pics sur les 900 initiaux, mais n'a pu entièrement lever l'incohérence du jeu de nOes.

Dans une deuxième étape, il s'agissait dans ce cas, non plus de supprimer la contrainte violée, mais d'augmenter la distance cible de 0,5 Å. Bien que cette étape n'a concerné que 24 contraintes au total, elle a permis d'obtenir quelques structures convergentes sans aucune violation, ni de distances, ni d'angles dièdres.

Les différentes contraintes (soit un total de 979, toutes classes confondues) utilisées pour les calculs sont résumées dans les tables ci-après : la table 6 indique, pour l'ensemble des contraintes de distance, la distribution suivant la séquence, la distance et le type de protons ; la table 7 liste les différents fichiers de contraintes avec leur nombre.

SÉQUENCE	DISTANCE	CONTACTS
i/i 0	< 3,5 Å 61	HN / HN 87
i / i+1 370	> 3,5Å et $< 4,5$ Å 156	HA / ali 124 (10)
i / i<5 229	>4,5 Å 762	HN / ali 348
i / i>=5 380		HN/HA 113
		ali / ali 297

Table 6 : statistiques des contraintes. <u>Séquence</u> : contacts entre les résidus i et j avec j=i, j=i+1,i+1 < j < i+5 ou j > i+5. <u>Distance</u> : la distance exprimée en angstræms séparant les résidus i et j.<u>Contacts</u> : type d'atomes avec HN : proton amide, HA : proton attaché au C α ; ali : tous les autresprotons. Le nombre entre parenthèse correspond aux nombres de contraintes H α /H α .

CONTRAINTES	CARACTÉRISTIQUES	NOMBRE	CLASSE
	contraintes Sy(Cys) / Zn et Sy(Cys) / Sy(Cys)	10	forte
Contraintes	liaisons hydrogènes présentes dans les hélices α droite	12	forte
de distances	liaisons hydrogènes présentes dans les feuillets β	14	forte
de distances	contraintes observées caractéristiques d'un feuillet β	10	forte
	autres contraintes observées (spectres 2D et 3D)	925	faible
Contraintes	angles ϕ et ψ caractéristiques des hélices α droite	34	unique
d'angles	angles ϕ et ψ caractéristiques des feuillets β	24	unique
dièdres	autres angles φ expérimentaux	8	unique

 Table 7 : fichiers de contraintes utilisés pour les calculs de structure. Il est indiqué pour chaque fichier ses caractéristiques, le nombre de contraintes et la classe (uniquement pour les contraintes de distances).

Seuls 3 atomes d'azote de la chaîne principale n'ont pas pu être attribués : les deux premiers et celui de la Serine 149. 77% de tous les protons et azotes de la séquence ont été attribués. Ce chiffre monte à 84 si l'on ne tient pas compte des atomes de bout de chaînes latérales non attribués. A noter que les protons n'ont pas été stéréospécifiés

• Un affinement des structures basé sur le protocole SA.

Dans une dernière étape, les structures obtenues ont été affinées en utilisant un algorithme de recuit simulé modifié. Dans ce calcul, la température de l'étape exploratoire est réduite à 1000 K, les structures initiales correspondent à celles issues des premiers calculs et enfin, la géométrie du site de liaison au zinc est définie dans la topologie de la protéine (topologie tétraédrique avec une longueur de liaison entre les atomes de soufre et le zinc de 2.3 Å).

B.IV/ TECHNIQUES LIEES A L'INSTRUMENTATION BIACORE.

B.IV.1/ Préparation des réactifs

• Le ligand (molécule accrochée à la surface)

L'hybridation des ADN cruciformes et double brins est réalisée par un lent refroidissement d'oligonucléotides de 30 bases chacun dont les séquences correspondent soit à l'ADN cruciforme dit de Lilley (Duckett, 1988), soit à celui dit de Bianchi (Bianchi, 1989). L'un des oligonuclétides est biotinylé en son extrémité 5'. La vérification de la masse et de la pureté des ADN est réalisée par électrophorèse d'ADN à l'aide d'un dépôt, sur le même gel, de fragments simple brin, double brin et cruciforme, chacun sous une forme biotinylée ou non.

• L'analyte (molécule en solution)

La qualité de la protéine E6, préparée selon le protocole décrit précemment, est vérifiée par une mesure d'agrégation avant chaque nouvelle série d'expériences BIAcore. Seuls les échantillons présentant le rapport d'agrégation le plus faible sont utilisés ensuite.

B.IV.2/ Immobilisation des ligands à la surface

Toutes les expériences BIAcore ont été réalisées sur un instrument BIAcore 2000 (Biacore AB, Uppsala, Suède). Le système est entièrement nettoyé de toutes les protéines adsorbées selon le protocole recommandé par le constructeur. Les surfaces *sensor* employées sont du type CM5 (la plus classique). Le tampon de réaction (ou tampon de run) dans lequel baigne en permanence la surface est constitué de Tris 50 mM pH 6.8, NaCl 150 mM Glycérol 5% filtré sur un filtre 0,22 μ m puis dégazé pendant 20 minutes minimum.

La concentration de ligands à la surface est ajustée, soit pour maximiser le transport de masse afin de déterminer la concentration active en analytes, soit pour le minimiser afin de mesurer des constantes cinétiques.

Dans le cas d'une cinétique, l'immobilisation se déroule en 5 étapes, les 4 premières en mode multicanal, la dernière en monocanal.

- i/ Activation des charges du dextran par un passage de 70 μl à un flux de 10 $\mu l/min$ (F10) d'une solution de EDC / NHS.
- ii/ Accrochage de la streptavidine diluée dans du tampon HBS (60 µl à F10).
- iii/ Désactivation des charges non utilisées dans une liaison avec la streptavidine par 100 μl d'une solution d'éthanolamine (F10).
- iv/ Lavage de la surface par 1 ml (F10) de NaCl 1M, suivi de 16 lavages de 5 μl d'une solution de NaOH 50 mM diluée dans de l'eau (stabilisation de la ligne de base), et enfin lavage extensif

pendant un minimum de 8 heures à F50 par le tampon de réaction (élimination de toute trace de soude).

v/ Immobilisation proprement dite de l'ADN biotinylé : cette étape se déroule en mode monocanal par injection manuel de l'ADN biotinylé (concentration comprise entre 25 et 100 nM) à un flux lent de 1 µl/min afin d'accrocher le plus précisemment la quantité voulue (aux environs de 200 RU de signal). L'immobilisation des 4 cellules de la surface suit généralement le schéma cidessous :

Cellule	Ligand
1	streptavidine seule
2	streptavidine + ADN double brin biotinylé
3	streptavidine + ADN cruciforme Lilley biotinylé
4	streptavidine + ADN cruciforme Bianchi biotinylé

Pour éviter toute contamination d'un ligand sur la cellule suivante, le système microfluidique ainsi que la boucle d'injection sont lavés extensivement après chaque immobilisation avec du tampon de réaction. Par la suite, une injection de SDS à 0,05% simultanément sur les 4 cellules permet de stabiliser la ligne de base. Celle-ci ne doit plus varier de plus de 10 RU par heure tout au long de l'expérience.

B.IV.3/ Conditions de régénérations

Le tampon de régénération permet de retrouver à un signal identique à celui qui existait sur les cellules avant l'injection de l'analyte. Dans notre cas, il est constitué de : Tris 50 mM, NaCl 850 mM, Magnésium 10 mM et Glycérol 5%. Ce tampon ne doit pas provoquer la précipitation de la protéine qui risquerait de boucher un canal. L'ajout ou non de SDS à 0,05% ne change rien quant à la régénération.

B.IV.4/ Acquisition des données expérimentales

L'injection de l'analyte est réalisée selon le mode « COINJECT » permettant une injection de l'analyte en deux phases. La première contient l'analyte proprement dit (la protéine E6 6C/6S) diluée dans le tampon pour lequel on veut tester l'effet des composants (principalement NaCl 150 ou 500 mM, Mg à 1 mM ou EDTA à 1 mM, Dextran à 1 ou 0 mg/ml). La seconde reprend strictement ce même tampon mais sans l'analyte, et ce afin de vérifier l'effet de ces constituants sur la cinétique de dissociation. La protéine E6 est utilisée à 5 concentrations différentes : 1000, 500, 250, 125 et 62,5 nM.

Les cinétiques sont normalement obtenues à des flux compris entre 10 et 50 μ l/min. Dans notre cas, des phénomènes de déplétion de l'analyte (non expliqués de manière satisfaisante) nous ont obligé à travailler à un flux de 10 μ l/min. Pour éviter toute dispersion de l'échantillon dans le circuit microfluidique, les volumes injectés sont réduits à 20 μ l dans la phase d'association et 100 μ l dans la phase de dissociation. Chaque cycle est prolongé jusqu'à 25 minutes en présence du tampon de réaction pour permettre un lavage le plus complet de l'ADN (Sun JC., Secondes Rencontres Françaises BIA, Paris Institut Curie, 30 mai 2000).

B.IV.5/ Traitement des données expérimentales

Le traitement des courbes est effectué grâce au logiciel BIAevaluation 3.0. La première étape consiste à ajuster précisemment le début de l'injection de l'analyte à zéro en temps et en signal. Les courbes suivent ensuite un double référençage [(Karlsson & Falt, 1997); (Myszka, 1999)]. La première référence, produite par la cellule contenant la streptavidine seule, est soustaite au signal de chacune des autres cellules. La seconde référence correspond au signal du tampon seul sur chaque cellule. L'ensemble des courbes obtenues aux différentes concentrations est systématiquement traité par ajustement global.

B.V/ AJUSTEMENT NON LINEAIRE DE COURBES EXPERIMENTALES SOUS MATLAB

Le premier paragraphe détaille les techniques de résolution numérique utilisées pour l'ajustement non linéaire de courbes, sigmoïdales pour des cinétiques ou des dénaturations, ou exponentielles pour la diffusion translationnelle la vitesse de relaxation transversale en RMN. Le second développe rapidement le principe utilisé pour l'étude numérique de l'interaction E6 / ADN cruciforme.

♦ Ajustement non linéaire

Lorsque les courbes sigmoïdales vont de 0 vers 1 pour des concentrations croissantes de produit [A] (cas z), je « renverse » ces courbes, qui vont ainsi de 1 vers 0 (cas z' = 1 - z).

$$z = \frac{\left[A\right] \cdot K_d}{1 + \left[A\right] \cdot K_d} \Leftrightarrow z' = 1 - z = \frac{1}{1 + \left[A\right] \cdot K_d}$$

Le paramètre ajusté (en l'occurence ici K_d) est toujours le même. Cependant, il apparaît deux fois dans l'expression de z, rendant l'ajustement moins stable. Au contraire, ce terme n'apparaît plus qu'une seule fois dans le cas z'.

Un ajustement non linéaire par moindres carrés est réalisé dans le logiciel MatLab (The Matworks Inc.). L'écart quadratique moyen δe qui ressort d'un tel ajustement permet de quantifier sa qualité. Cependant, il ne reflète en rien la contribution de l'erreur expérimentale de chaque point. Une procédure de type Monte-Carlo est employée afin d'évaluer l'impact de cette erreur expérimentale sur l'erreur globale de l'ajustement : les points utilisés lors du premier ajustement sont considérés comme étant la meilleure approximation de la valeur moyenne statistique des mesures expérimentales. Une distribution normale de valeurs est ensuite générée autour de ces points expérimentaux, et la procédure d'ajustement répétée 100 fois. Ceci permet de déterminer une valeur moyenne du K_D, ainsi que l'écart-type δK_D .

• Etude numérique de l'interaction E6 / ADN cruciforme

La méthode permettant d'affirmer que le domaine C-terminal de E6 se lie sur deux sites de l'ADN, identiques et indépendants, est basée sur l'approche des poids statistiques établis pour des interactions entre proches voisins. Cette approche est décrite bien plus en détails dans le livre de Tinoco (Tinoco, 2002).

Le modèle général pour une interaction à deux sites indépendants est :

$$X + 2 \ P \iff XP + P \iff XP_2$$

où X et P correspondent respectivement à l'ADN et à la protéine.

L'idée est en fait d'affecter un poids statistique pour passer de l'état X + P à XP, et de l'état XP + P à XP₂. Dans le premier cas, ce poids est égal à S, dans le second à S. τ où S = K.[P], K étant la constante de dissociation à l'équilibre. τ est un facteur contrôlant le passage à l'état XP₂ à partir de XP. Il permet de « rendre » une interaction coopérative ($\tau > 1$, le passage de XP à XP₂ étant alors favorisé), indépendante ($\tau = 1$) ou anti-coopérative ($\tau < 1$).

Les quantités X, XP et XP_2 ont été déterminées numériquement à l'aide du logiciel MatLab, en fonction de la concentration en protéines P. Ces quantités, comparées aux fractions X, XP et XP_2 déduites d'un gel retard, laissent apparaître deux sites indépendants.

RÉSULTATS et DISCUSSIONS

Les résultats présentés se divisent en trois parties. La première concerne l'obtention de l'oncoprotéine E6 repliée, stable, active et monomérique, ainsi que de ses deux domaines N et C-terminaux. Les résultats de l'étude structurale du domaine E6-C 4C/4S sont développés dans la seconde. Enfin, la troisième expose les résultats obtenus sur l'interaction entre cette protéine E6 et un type particulier d'ADN, l'ADN cruciforme.

_ RESULTATS, DISCUSSIONS

<u>PARTIE I</u>

Obtention de l'oncoprotéine E6 de HPV 16 sous une forme active et repliée, ainsi que de ses deux domaines N et C-terminaux ; caractérisation biophysique de ces particules.

La présentation des résultats de cette partie se traduit par 3 publications :

- Publication 1 : Formation of soluble inclusion bodies by HPV 16 E6 oncoprotein fused to Maltose Binding Protein. <u>Nominé Y.</u>, Ristriani T., Laurent C., Lefèvre J-F., Weiss E., Travé G. (2001) *Protein Expression and Purification* 23, 22-32.
- Publication 2 : A strategy for optimising the monodispersity of fusion proteins: application to purification of recombinant HPV 16 E6 oncoprotein. <u>Nominé Y.</u>, Ristriani T., Laurent C., Lefèvre J-F., Weiss E., Travé G. (2001) *Protein Engineering* 14(4) 297-305
- Publication 3 : Domain substructure of HPV E6 oncoprotein: Biophysical characterization of E6 C-terminal DNA-binding domain. <u>Nominé Y.</u>, Charbonnier S., Ristriani T., Stier G., Masson M., Cavusoglu N., Van Dorsselaer A., Weiss E., Kieffer B., Travé G. (Publication soumise à *Biochemistry*)

I.1/ Publication 1

Nominé Y., Ristriani T., Laurent C., Lefèvre J-F., Weiss E., Travé G. (2001) : Formation of soluble inclusion bodies by HPV 16 E6 oncoprotein fused to Maltose Binding Protein. *Protein Expression and Purification* 23, 22-32

Ce travail concerne la qualité des protéines de fusion. Celles-ci sont constituées de la protéine d'intérêt (dans notre cas, E6) fusionnée à l'extrémité C-terminale d'une protéine porteuse MBP (Maltose Binding Protein). Ces fusions ont la propriété remarquable de « solubiliser » la protéine d'intérêt, même si celle-ci a tendance à précipiter lorsqu'elle est produite en dehors de la fusion [Kapust, 1999]. Nous démontrons ici que ces fusions restent effectivement en solution, mais contiennent une majorité de protéines E6 non repliées et biologiquement inactives. Ces molécules de E6 s'agglomèrent entre elles par des interactions hydrophobes, tandis que la présence des protéines MBP, repliées, stabilisent ces agrégats et évitent ainsi leur précipitation. Ces molécules agrégées constituent donc des particules de hauts poids moléculaires, mais parfaitement solubles. Nous avons appelé ces objets des « corps d'inclusion solubles » par analogie avec les corps d'inclusion insolubles retrouvés dans les bactéries. L'une des propriétés de ces objets est de se co-purifier sur une résine amylose en même temps qu'une fusion ne contenant pas de protéine agrégée, contaminant ainsi la protéine « pure ». Ainsi, le fait que la fusion soit soluble ne garantit en rien que la protéine d'intérêt soit active et correctement repliée.

Mis en évidence dans le cas de la protéine E6, ces corps d'inclusion solubles ne semblent pourtant pas se limiter à cette unique protéine, et seraient fréquemment retrouvés lorsqu'une protéine, sensible aux problèmes de repliement, est fusionnée à une protéine porteuse. Cette dernière empêche les particules agrégées d'atteindre la masse critique de sédimentation.


[Signalement bibliographique ajouté par : ULP – SCD – Service des thèses électroniques]

Formation of soluble inclusion bodies by HPV 16 E6 oncoprotein fused to Maltose Binding Protein.

Nominé Y., Ristriani T., Laurent C., Lefèvre J-F., Weiss E., Travé G.

Protein Expression and Purification, 2001, Vol. 23, Pages 22-32

Pages 22 à 32 :

La publication présentée ici dans la thèse est soumise à des droits détenus par un éditeur commercial.

Pour les utilisateurs ULP, il est possible de consulter cette publication sur le site de l'éditeur : <u>http://dx.doi.org/10.1006/prep.2001.1451</u>

Il est également possible de consulter la thèse sous sa forme papier ou d'en faire une demande via le service de prêt entre bibliothèques (PEB), auprès du Service Commun de Documentation de l'ULP: peb.sciences@scd-ulp.u-strasbg.fr

I.2/ Publication 2

Nominé Y., Ristriani T., Laurent C., Lefèvre J-F., Weiss E., Travé G. (2001) :

A strategy for optimising the monodispersity of fusion proteins: application to purification of recombinant HPV 16 E6 oncoprotein. *Protein Engineering* **14**(4) 297-305

Dans la première publication, nous avions démontré que les préparations de MBP-E6 étaient contaminées par une majorité de corps d'inclusion solubles. Nous avons donc élaboré une stratégie basée sur différentes conditions d'expression et de purification à petites échelles pour améliorer la qualité –principalement la monodispersité– de ces protéines de fusion. Pour cela, nous nous sommes notamment aidés de mesures, sur un simple spectrofluorimètre, de diffusion statique de lumière à un angle de 90°. L'une des étapes déterminantes a été de séparer les protéines de fusion monomériques, des corps d'inclusion solubles. Ces protéines de fusion, de meilleure qualité, coupées ensuite par une protéase spécifique, permettent d'obtenir une protéine E6 repliée, monomérique et biologiquement active.

En parallèle, ces résultats nous ont conduits à concevoir un mutant de E6, appelé E6 6C/6S, pour lequel les cystéines non impliquées dans la liaison aux ions zinc ont été mutées en sérines. Ce mutant offre l'avantage, par rapport à la protéine sauvage, d'être plus stable dans le temps. La protéine E6 a ainsi pu livrer ses premières mesures biophysiques.



[Signalement bibliographique ajouté par : ULP – SCD – Service des thèses électroniques]

A strategy for optimising the monodispersity of fusion proteins: application to purification of recombinant HPV 16 E6 oncoprotein.

Nominé Y., Ristriani T., Laurent C., Lefèvre J-F., Weiss E., Travé G.

Protein Engineering, 2001, Vol. 14, Pages 297-305

Pages 297 à 305 :

La publication présentée ici dans la thèse est soumise à des droits détenus par un éditeur commercial.

Pour les utilisateurs ULP, il est possible de consulter cette publication sur le site de l'éditeur : <u>http://peds.oxfordjournals.org/cgi/content/full/14/4/297</u>

Il est également possible de consulter la thèse sous sa forme papier ou d'en faire une demande via le service de prêt entre bibliothèques (PEB), auprès du Service Commun de Documentation de l'ULP: peb.sciences@scd-ulp.u-strasbg.fr I.3/ Publication 3 (en cours de soumission)

 Nominé Y., Charbonnier S., Ristriani T., Stier G., Masson M., Cavusoglu N., Van Dorsselaer A., Weiss E., Kieffer B., Travé G. Domain substructure of HPV E6 oncoprotein: Biophysical characterization of E6 C-terminal DNA-binding domain. (Publication soumise à *Biochemistry*)

Les résultats décrits dans les deux publications précédentes nous ont permis d'accumuler de nombreux échantillons stables, repliés et biologiquement actifs de l'oncoprotéine E6 de HPV16. Parmi ceux stockés sur une période de plusieurs mois, une protéolyse non spécifique de la protéine entière nous a permis, par spectrométrie de masse et séquençage protéique en N-terminal, d'isoler deux fragments : un fragment N-terminal (résidus 7 à 83) et un fragment C-terminal (résidus 87 à 158). Ici, nous analysons en particulier le domaine C-terminal de E6 HPV16 qui, après clonage, sur-expression et purification, a pu être amené à une concentration de l'ordre du milli-molaire. Tout comme la protéine entière, ce domaine reconnaît également l'ADN cruciforme. Par ailleurs, en se basant sur des résultats d'absorption et de fluorescence UV, ainsi que de dichroïsme circulaire, nous montrons que ce peptide est replié et qu'il contient une proportion identique de structures secondaires en hélice α et en feuillet β . Enfin, des mesures de diffusions translationnelles par RMN nous permettent de conclure que ce fragment est monomérique, et ce même à des concentrations de l'ordre du milli-molaire.

Cette publication est suivie d'un paragraphe montrant l'évolution de la qualité du domaine E6-C 4C/4S au cours des différentes purifications.

Domain substructure of HPV E6 oncoprotein: Biophysical characterization of E6 C-terminal DNA-binding domain.

Yves Nominé[‡], Sebastian Charbonnier[‡], Tutik Ristriani[‡], Gunter Stier[§],

Murielle Masson[‡], Nukhet Cavusoglu^{||}, Alain Van Dorsselaer^{||},

Étienne Weiss[‡], Bruno Kieffer[⊥] and Gilles Travé[‡]*

(t)This work was supported by a grant from the Association pour la Recherche sur le Cancer and by a "PCV" grant from the CNRS. Tutik Ristriani was a recipient of a grant from the Ligue Nationale Contre le Cancer.

(‡) Laboratoire d'Immunotechnologie, UMR CNRS 7100, École Supérieure de Biotechnologie de Strasbourg, 67400 Illkirch, France.

(§) EMBL, Meyerhofstrasse 1, D-69117 Heidelberg, Germany.

(⊥) Laboratoire de RMN, UMR CNRS 7104, École Supérieure de Biotechnologie de Strasbourg, 67400 Illkirch, France.

(||) Laboratoire de Spectrométrie de Masse Bio-Organique, UMR CNRS 7509, Faculté de Chimie, 67087 Strasbourg, France

* Corresponding author

tel: 0033 3 90 24 47 65, fax: 0033 3 90 24 47 70, e-mail: trave@esbs.u-strasbg.fr

Running title : The C-terminal domain of HPV-16 E6

¹Abbreviations : HPV, human papillomavirus ; CD, circular dichroism ; UV, ultraviolet ; NMR, nuclear magnetic resonance ; DTT, dithiothreitol ; LB, Luria Broth ; SDS-PAGE, sodium dodecyl sulfate-polyacrylamide gel electrophoresis ; IPTG, isopropyl 1-thio β -D-galactopyranoside.

ABSTRACT.

E6 is a viral oncoprotein implicated in cervical cancers, produced by high-risk human papillomaviruses (HPVs). Structural data concerning this protein are scarce due to the difficulty of producing recombinant E6. Recently, we described the expression and purification of a stable, folded and biologically active HPV16 E6 mutant called E6 6C/6S. Here, we analyzed the domain substructure of this mutated E6. Non-specific proteolysis of full-length E6 6C/6S (158 residues) yielded N-terminal and Cterminal fragments encompassing residues 7-83 and 87-158, respectively. The C-terminal fragment 87-158 was cloned, overexpressed and purified at concentrations as high as 1 mM. The purified domain retains the selective four-way DNA junction recognition activity of full-length E6 protein. Using UV absorption, UV fluorescence, circular dichroism and nuclear magnetic resonance, we show that the peptide is monomeric and folded with equal proportions of α -helix and β -sheet secondary structure.

E6 is one of the two oncoproteins produced by "highrisk" human papillomaviruses (HPVs) responsible for cervical cancers (1). E6 is thought to promote tumorigenesis by stimulating cellular degradation of the tumour suppressor p53 via formation a trimeric complex comprising E6, p53 and the cellular ubiquitination enzyme E6AP (2, 3). However, recent findings suggest that E6 displays other activities unrelated to p53. These include recognition of a variety of other cellular proteins : transcription co-activators p300/CBP (4, 5) and ADA3 (6), transcription factors c-Myc (7) and IRF3 (8), replication protein hMCM7 (9), DNA repair proteins MGMT (10) and XRCC1 (11), protein kinases PKN (12) and Tyk2 (13), Rap-GTPase activating protein E6TP1 (14), tumor necrosis factor receptor TNF-R1 (15), apoptotic protein Bak (16), clathrin-adaptor complex AP-1 (17), focal adhesion component paxillin (18, 19), calcium-binding proteins E6BP (20) and fibulin-1 (21), and several members of the PDZ protein family such as hDLG (22), hScrib (23), MAGI-1 (24) and MUPP1 (25). In addition, E6 activates or represses several cellular or viral transcription promotors (8, 26, 27, 28). In particular, E6 induces transcriptional activation of the gene encoding the retrotranscriptase of human telomerase (29-31). Finally, we have recently demonstrated that E6 is a DNA-binding protein which recognizes four-way DNA junctions (32, 33). In contrast to the large amount of functional data, there are very few biophysical and structural studies concerning E6 due to the difficulty of expressing and purifying this protein -or its domainsin a stable folded form. Most efforts have focussed on E6 from HPV16 since this is the highest-risk HPV strain for cervical cancers. Two HPV16 E6 polypeptides are thought to be produced in vivo depending on which methionine is used as a start codon : a 158-residue form and a 151-residue form

(34) (here, we will use the 158-residue numbering). Sequence alignements of E6s from numerous HPV subtypes suggested the presence of two zinc-binding motifs (35). Both motifs are 37 residues-long and contain four cysteines distributed as follows : CxxC-(29x)-CxxC. Zinc binding to these motifs was experimentally demonstrated by zinc-blotting assays (35-38), sitedirected mutagenesis (39) and atomic and electronic absorbtion methods (40). On this basis, several authors (41-43) divide E6 sequence into five "regions": Nterminal loop (residues 1-36), N-terminal zinc-binding motif (residues 37-73), linker region (residues 74-110), C-terminal zinc-binding motif (residues 110-146) and Cterminal loop (residues 147-158). However, the structural boundaries of the zinc-binding domains may include additional residues provided by the N-terminal and/or C-terminal regions flanking the 37-residue motifs. Recently, we expressed and purified two fragments containing respectively the N-terminal and the Cterminal zinc-binding motifs of wild-type HPV16 E6 (33). Based on multiple sequence alignments, we deliberately included 9 residues upstream of the first cysteine and 5 residues downstream of the fourth cysteine of each motif. The N-terminal and the Cterminal fragments each contained 51 residues and spanned residues 28-78 and 101-151, respectively. Both constructs were produced as soluble fusions to the Cterminus of Maltose-Binding Protein (MBP). The MBP-ZD2 fusion containing the C-terminal zinc-binding motif was monomeric and was shown to retain the DNA recognition activity of full-length HPV16 E6 (33). However, in that work, we did not attempt to separate the ZD2 peptide from the MBP carrier. In another work, Lipari et al. (40) produced an N-terminal fragment of wild-type HPV16 E6 spanning residues 9-84. This fragment was therefore 75 residues-long, and it included 28 residues upstream of the first cysteine and 9 residues downstream of the fourth cysteine. The fragment was expressed and purified in a soluble form which could be raised at high concentrations. It was folded with a high content of α -helical and β -sheet secondary structures and contained a single zinc ion (40). The same authors also expressed and purified a longer construct which spanned almost the entire sequence of wild-type HPV16 E6 (residues 9-149). This construct was also folded with α helical and β -sheet structure, but it displayed low solubility (40). The authors concluded that the Cterminal region (85-149), containing the second zincbinding motif, was probably unstable or insoluble and contributed to the low solubility of full-length HPV 16 E6

Recently, we described a stabilized mutant of HPV 16 E6 in which six non-conserved cysteines were changed into serines (44, 45). This mutant, called E6 6C/6S, can be purified as a stable, folded and biologically active form (45). Here, we have investigated the domain substructure of the E6 6C/6S mutant. Non-specific proteolysis of the folded protein yielded two soluble fragments of about 75 residues each which were

separated by chromatography and analyzed by Nterminal sequencing and mass spectrometry. One of these fragments corresponds to the N-terminal domain already described for wild-type HPV16 E6 (40). More interestingly, the second fragment corresponds to the complementary C-terminal half. This C-terminal domain, derived from our E6 6C/6S mutant, is soluble and monomeric at concentrations as high as 1 mM. Circular dichroism and NMR data show that the protein is properly folded and contains a high content of α -helical and β -sheet secondary structures.

EXPERIMENTAL PROCEDURES.

Expression, purification of E6 6C/6S, E6-C 4C/4S.

Full-length E6 6C/6S (158 residues) was expressed as a fusion to the C-terminus of MBP and purified as described previously (45). The MBP-E6-C 4C/4S construct corresponds to residues 87-158 of E6 6C/6S fused to the Cterminus of His6-MBP via a TEV protease-sensitive linker. The sequence coding for E6-C 4C/4S was PCR-amplified with appropriate oligomers and inserted into the NcoI/KpnI sites of PETM-41, a modified pET24d expression vector (Novagen) containing a N-terminal His6-MBP tag followed by a TEV protease cleavage site. The sequence was checked using Thermo Sequenase Cy5.5 dye terminator cycle sequencing kit (Amersham Pharmacia Biotech). Sequences were run on a Seq4x4 Basecaller and processed using ALFwin sequence analyser. BL21 DE3 E. Coli cells freshly electroporated with the MBP-E6-C 4C/4S expression construct were directly transfered in LB, 15ug/ml kanamycin, and grown overnight for 12-14 hrs at 37°C. The preculture was diluted 40-fold in fresh LB/kanamycin and grown at 37°C until OD(600 nm) = 0.6-0.7. Cells were harvested by centrifugation at 2,300 g and 25°C during 10 minutes, transferred to fresh medium preequilibrated at 37°C and grown at 37°C until OD(600 nm) = 0.6-0.7. The culture was then adjusted to 0.5 mM IPTGand grown for three more hours at 22 °C. Cells from 2 l of either expression culture were harvested by centrifugation at 2,300 g and 4°C during 20 minutes. To minimize oxidation problems, all buffers used for purification were degassed using a water vacuum pump then bubbled extensively with argon. The pellet was resuspended in 150 ml of buffer A (Tris 50 mM, NaCl 150 mM, DTT 2 mM, pH 6.8) containing 5% glycerol, 1 µg/ml DNase I, 1 µg/ml RNase I, and 3 tablets of anti-protease cocktail (EDTAfree) (Boehringer-Mannheim). Cells were broken by sonication on ice, then centrifuged at 18,000 g at 6°C for 30 minutes. The supernatant was filtered (Millipore 0.22 µm) and loaded on a 80 ml column of amylose resin (New England Biolabs) pre-equilibrated with buffer A. The column was washed stepwise with 1 volume, 3 volumes and 12 volumes of buffer A containing respectively 100%, 50%, and 12% of initial anti-protease concentration. Remarkably, MBP-E6-C 4C/4S eluted as a relatively pure form by leaking from the column in the late washing steps. The protein was incubated 12-24 hrs at 6°C with an appropriate concentration of recombinant TEV protease (Stier G, unpublished results) until full separation of E6-C 4C/4S from the MBP tag was achieved. The TEV cleavage site results in two additional residues (Gly-Ala) on the N-

terminus of the construct, prior to the methionine. The digestion product was concentrated to a 1 ml volume using a 10kDa-centriprep concentration device (Amicon) and applied on a Hiload 16/60 superdex 75 gel-filtration column (Amersham Biosciences) pre-equilibrated with buffer A. Pure monomeric E6-C 4C/4S peptide eluted as a single peak at the volume expected for a monomer according to the column's calibration. Sample concentration was raised to 1-1.4 mM using a 15ml Ultrafree Biomax 5K NMWL Membrane (Millipore).

Mass and sequence analysis.

Mass measurements were carried out on a Bruker BIFLEX IIITM Matrix-Assisted Laser Desorption Time-Of-Flight Mass Spectrometer (Bremen, Germany) equipped with the SCOUTTM. High Resolution Optics with X-Y multisample probe and griddles reflector. This instrument was used at a maximum accelerating potential of 19 kV and was operated in linear mode. Ionization was accomplished with a 337 nm beam from a nitrogen laser with a repetition rate of 3 Hz. The output signal from the detector was digitized at a sampling rate of 2 GHz. The mass accuracy was approximately of 0.1%. A saturated solution of α -cyano-4-hydroxycinnamic acid in acetone was used as a matrix. A first layer of fine matrix crystals was obtained by spreading and fast evaporation of 0.5 µl of matrix solution. On this fine layer of crystals, a droplet of 0.5 µl of aqueous HCOOH (5%) solution was deposited. Afterwards, 0.5 µl of sample solution was added and a second droplet 0.2 ul of matrix saturated solution (in 50% H_2O / 50% ACN) was added. The preparation was dried under vacuum. The instrument was calibrated in the mass range 5000-20000 Da using a mixture of chicken cytochrome c and lysozyme (Sigma). To perform a linear calibration in this mass range, mono and bi-charged species at 6181 and 12361 m/z for cytochrome c and 7153 and 14306 m/z for lysozyme were used successively.

For N-terminal sequence analysis the peptides were separated by SDS-gel and transferred onto a nitrocellulose membranes. The bands were isolated and subjected to Edman degradation method.

Spectroscopic measurements.

Fluorescence measurements were performed with a SPEX Fluorolog-2 spectrofluorimeter (SPEX Industries, Inc., Edison, NJ) equipped with a 450 Watt Xe lamp, a double grating excitation monochromator and a single grating emission monochromator. Data were acquired with a photon counting photomultiplier (linear up to 10^7 counts/sec) with high voltages fixed at 800 Volts. Slit widths were adjusted to avoid saturation of detectors. 2 ml of solution was placed in a cuvette maintained at 15°C in a thermostatted cuvette handler. Emission spectra recordings were typically sampled at every half-nanometer. In all experiments, slit widths were set to 1.8 nm both for excitation and emission. Ultraviolet absorption measurements were performed with a Perkin Elmer 1-2 spectrophotometer. Circular dichroism was performed using a Jobin-Yvon spectropolarimeter. Secondary structure content was estimated by the VARiable SELECtion method (VARSELEC) (46).

Nuclear Magnetic Resonance experiments.

¹H NMR measurements were obtained from a 1 mM sample of E6-C 4C/4S in 20 mM deuterated Tris, pH 6.8, 50 mM NaCl, 1 mM DTT, 10% D_2O . The 2D [¹H,¹H] NOESY spectrum was recorded at 15 °C on a Bruker DRX-600 spectrometer with a mixing time of 200 ms and a spectral width of 8 kHz in both

dimensions. The water signal was suppressed using the WATERGATE sequence (47). 48 transients of 2K data points were acquired per t_1 value, with 512 t_1 increments and a relaxation delay of 2.25 s. Data were processed using NMR-pipe (48) and analysed with XEASY (49).

The translational diffusion constants were measured on a Bruker AMX-500 spectrometer equipped with a Z-gradient 5 mm probe using a modified version of the Longitudinal Encoded Diffusion experiment (LED) (50, 51). The protein signal decay was followed by recording 32 LED spectra with a gradient length δ varying from 1 ms to 16 ms and other constant delays set to the following values: $\tau=20$ ms, Δ =50 ms. Square gradient pulses were applied with an intensity of 0.28 T/m. This intensity was calibrated by imaging a hole of known length in a Teflon phantom and by measuring the diffusion of a water sample. An additional check was performed by measuring the selfdiffusion coefficient of lysozyme and comparing it to previously reported values (52). The measured value was similar to the reported value with a deviation lower than 16%. Moreover, the value which could be extrapolated from our curve for haemoglobin was also consistent with the experimental value (53). FELIX 2.1 (Accelrys Inc., Burlington, MA) software was used to process the spectra and to measure the peak height of ten well-resolved protein resonances. The diffusion coefficients were obtained by fitting the signal intensities with a three-parameter monoexponential decay using a simplex algorithm implemented within the Matlab software (The Mathworks Inc., Natick, MA). The expression of the exponential was:

 $R(x) = R_{\infty} + R_0 \exp(-D_t x)$ where $x = \gamma H^2 G_Z^2 \delta^2 (\Delta - \delta/3)$

Errors on the diffusion coefficients were estimated by calculating the standard devation of the ten diffusion values obtained at each temperature.

Construction of the DNA probe for gelretardation assay.

A cruciform based upon the sequence of junction 1 originally described by Duckett *et al.* (54), was constructed by hybridizing four oligonucleotides each of 30 nucleotides referred to as b, h, r and x, respectively, one of which was 5'-[32P]-labelled. These were:

b-strand:5'-TCCGTCCTAGCAAGCCGCTGCTACCGGAAG-3' h-strand:5'-CTTCCGGTAGCAGCGGGAGAGCGGTGGTTGAA-3' r-strand:5'-TTCAACCACCGCTCTTCTCAACTGCAGTCT-3' x-strand:5'-AGACTGCAGTTGAGAGCCTTGCTAGGACGGA-3'

Gel retardation analysis.

Purified proteins were mixed with labelled DNA junction in 25 mM Tris-HCl pH 7.5, 100 mM NaCl, 10 mM KCl, 1.5 mM MgCl₂, 100 μ g/ml genomic DNA and 5% glycerol in a total volume of 10 μ l. Samples were then incubated for 15 min on ice and electrophoresed in 6.5% polyacrylamide gels (29:1 acryl:bisacryl) with 45 mM Tris-Borate in the presence of 1 mM EDTA, at 10 V/cm and 8°C. Gels were dried on Whatman 3M paper and analysed on a phosphorimager (Amersham Pharmacia Biotech).

RESULTS.

Non-specific proteolysis of E6 6C/6S yields soluble fragments corresponding to the Nterminal and C-terminal halves of E6 6C/6S.

We routinely purified samples of the E6 6C/6S mutant at a final concentration of 5-10 μ M (45). When stored at 4 °C in high-salt phosphate buffer (20 mM sodium phosphate, pH 7.4, 500 mM NaCl) samples did not undergo detectable proteolysis for several months. However, when the purified protein was stored in highsalt Tris buffer (20 mM Tris, pH 7.4, 500 mM NaCl) instead of phosphate, the samples underwent nonspecific proteolysis, generating several soluble fragments of about 8-10 kDa (Figure 1). Over five months of storage, the full-size protein gradually disappeared to be completely replaced by the fragments (Figure 1, lanes 1 to 4). At low salt conditions, the slowest-migrating fragment (peptide 3) and a faster minor band (peptide 2) bound to anion-exchange resin and could be eluted at higher ionic strength (lane 6). The fastest-migrating band (peptide 1) did not bind to resin and was thus recovered in the flow-through of the low-salt incubation step (lane 7). The purified peptides were subjected to N-terminal sequencing and Maldi-TOF mass spectrometry. The results of this analysis are summarized in Table I. Peptide 1, which did not bind to anion-exchange at low salt, starts at alanine 7 and displayed an experimental mass of 9251 \pm 9 Da. Peptides 2 and 3, which bound resin at low salt and were eluted at high-salt, start both at serine 87 and display experimental masses of 8011 ± 8 Da and 8637 ± 8 Da, respectively. Therefore, as compared to the 158 residues of full-length E6 6C/6S, peptide 1 corresponds to residues 7-83, peptide 2 to residues 87-153, and peptide 3 to residues 87-158, respectively. For each peptide, the theoretical mass computed from sequence is in good agreement with the experimental mass (Table 1). The calculated isoelectric points of the peptides also fit with their behavior on anion-exchange resin. Peptide 1 is neutral at pH 7.4 and therefore it does not bind S-sepharose resin. By contrast, peptides 2 and 3 are positively charged at pH 7.4 which explains their strong binding to the resin.

Cloning, expression and purification of E6-C 4C/4S (residues 87-158 of E6 6C/6S).

The results presented above pointed to C-terminal peptide 3 (residues 87-158 of E6 6C/6S) as an interesting candidate for biophysical studies. Residues 87-158 were fused to the C-terminus of Maltose Binding Protein (MBP) via a linker sensitive to TEV protease. The resulting MBP-fusion (called MBP-E6-C 4C/4S) was overexpressed (Figure 2, lane 1), purified by affinity on MBP-specific amylose resin (lanes 2-3), then subjected to thrombin digestion yielding a 40 kDa protein corresponding to MBP and an approximatively 10 kDa protein corresponding to the E6-C 4C/4S domain (lane 4). E6-C 4C/4S was then separated from MBP and TEV by gel-filtration (lane 5). E6-C 4C/4S was eluted as a

single peak which corresponded to a monomer according to the column's calibration. This purified E6-C 4C/4S sample was completely soluble and could be concentrated up to 1 mM without undergoing any detectable aggregation.

In parallel, we tried a similar approach with a shorter construct, called MBP-ZD2, which consists of residues 101-151 of wild-type HPV16 E6 fused to the C-terminus of MBP (33). We have previously shown that MBP-ZD2 retains the DNA-binding properties of full-length E6 (33). The ZD2 construct includes the C-terminal zinc-binding motif, but it lacks 14 Nterminal residues (S₈₇YSLYGTTLEQQYN₁₀₀) when compared to E6-C 4C/4S. In addition, ZD2 corresponds strictly to a fragment of the sequence of wild-type HPV16 E6 whereas E6-C 4C/4S contains 4 mutations of non-conserved cysteines into serines. However, when ZD2 was separated from the MBP carrier by thrombin digestion, it became prone to non-specific proteolysis and could not be purified further (data not shown).

E6-C 4C/4S retains monomeric binding to four-way DNA junction.

We have previously demonstrated that E6 is a DNAbinding protein which recognizes four-way junctions (32), and that this activity is localized in the Cterminus of the protein (33). The purified E6-C 4C/4S sample retains this activity as shown by gelretardation assay (Figure 3, lanes 5 to 8). In our previous work, we have shown that only one monomer of E6 6C/6S can bind to the junction (32). Interestingly, E6-C 4C/4S generated only one retardation band even at the highest concentration tested (lane 8). In addition, this band migrated faster than the band generated by E6 6C/6S (lanes 2 to 4). This strongly suggests that the E6-C 4C/4S domain binds to the junction only as a monomer, as observed for full-length E6 6C/6S.

Biophysical characterization of E6-C 4C/4S.

To check the purity and quality of the E6-C 4C/4S preparation, the sample was first subjected to UV absorption measurements (Figure 4A). This spectrum has a particular shape due to the aromatic residues of E6-C 4C/4S: 3 tyrosines (absorption maximum 274 nm) and 1 tryptophan (absorption maximum 280 nm). The CD (Circular dichroism) spectrum spanning the far-UV region is characteristic of a folded protein with defined secondary structure elements (Figure 4B). Based on that spectrum, back-calculations of the secondary structure content of E6-C 4C/4S give a proportion of $22\pm1\%$ of α -helix, 22±3% of antiparallel β -sheet, $3\pm1\%$ of parallel β -sheet and $19\pm 2\%$ of turn. The fluorescence spectrum (Figure 4C) was recorded at an excitation wavelength of 280 nm. The maximum of intrinsic fluorescence was found at 331 nm, which is shifted to lower wavelength as compared to the normal maximum

emission of tryptophan in water (about 350 nm). This indicates that the single tryptophan of E6-C 4C/4S is buried inside the protein in an hydrophobic environment, suggesting that the protein is folded. Finally, the protein was subjected to NOESY NMR measurements (Figure 5). Spectral dispersion in both amide and aliphatic region confirms that the protein is folded. The occurrence of numerous HN-HN NOESY cross-peaks indicates the presence of α -helical regions, whereas several H α resonances with chemical shifts greater than 4.8 ppm are indicative of β -sheet structure. However, a detailed analysis of the spectra shows a large heterogeneity in peak line width along the sequence, spanning from 16 to 47 Hz. This suggests the occurrence of intermediate usms time-scale motions within the domain. Alternatively, heterogeneous line broadening could be due to proteinprotein interactions leading to formation of dimers or oligomers of higher order.

The E6-C 4C/4S domain diffuses as a monomer at 1 mM concentration.

To check whether heterogeneous line broadening originated from oligomerization, the diffusion properties of the domain were investigated at 1mM concentration by NMR. Pulsed field gradient LED experiments were used to measure the translational diffusion coefficient (D_t) of E6-C 4C/4S for temperatures ranging from 15 $^{\circ}$ C to 50 °C. Figure 6A shows the Dt values plotted as a function of the temperature to viscosity ratio (T/η) . The observed linear relationship between D_t and T/η reflects the accuracy of the diffusion measurement and indicates that the shape and the oligomeric state of the protein was conserved over this range of temperatures. For temperatures above 50 °C, the accuracy of the diffusion measurements was diminished by convection. In order to determine the oligomeric state of E6-C 4C/4S, we compared normalized D_t values (20°C in water) measured for proteins with sizes ranging from 1100 Da (a peptidic siderophore) to 14500 Da (lysozyme) (Figure 6B). For a given molecular weight, the normalized D_t values depend on the mass and the shape of the proteins. The comparison of normalized D_t values indicates that E6-C 4C/4S behaves as a monomeric protein and that its hydrodynamic shape does not differ significantly from the spherical model. Therefore, heterogeneous line broadening cannot due to oligomerization but rather to the occurrence of particular motions in the monomeric domain.

DISCUSSION.

In the present work, we have analyzed the domain substructure of a E6 6C/6S, a stabilized mutant of HPV 16 E6. Non-specific proteolysis of the protein generated three soluble peptidic fragments which span almost the entire sequence of E6 (Table 1). One fragment spans the N-terminal half (residues 7-83) and includes the first cluster of conserved cysteines known to bind zinc (36, 37, 39, 40) (positions 37, 40, 70, 73). The two other fragments span the C-terminal half (residues 87-153 and 87-158) and include the second zincbinding motif (positions 110, 113, 143, 146). Proteolysis was extremely selective and attacked only a few positions on the protein sequence, i.e. peptide bonds 6-7, 83-84, 86-87 and 153-154. Under limiting conditions, proteases are known to operate preferentially on interdomain linkers as well as Nterminal and C-terminal unstructured regions (55). Our results therefore suggest that E6 6C/6S consists mainly of two folded zinc-binding domains corresponding to residues 7-83 and 87-153, connected by a three-residue linker (residues 83-86) and flanked by an N-terminal loop (residues1-6) and a C-terminal loop (residues 154-158). This result confirms the recent work by Lipari et al. (40) which also suggested a two-domain substructure for E6 proteins. Therefore, the familiar division of E6 sequence into five "regions" (41-43) is probably inaccurate in terms of 3-dimensional structure. Remarkably, the circular dichroism spectrum of E6-C 4C/4S domain is very similar to that previously obtained for the E6-N domain (40) and backcalculations from both spectra give comparable estimates of secondary structure content for both domains. This supports previous suggestions, based on sequence alignements, that the two domains have a similar fold and originate from duplication of a common structural ancestor (35).

E6 proteins and E6 domains have long resisted expression and purification attempts. One of the reasons for this might be their anomalously high cysteine content (45, 56). We have observed that whenever E6 and E6 fragments are handled without extreme control of redox conditions, aggregation and precipitation systematically occur. It is this particularity of E6 proteins which led us to design the E6 6C/6S mutant, in which six non-conserved cysteine residues were changed into serines (45). The present E6-C 4C/4S construct is derived from E6 6C/6S mutant and therefore contains only four cysteines instead of the eight cysteines present in the corresponding region of wild-type E6. The four remaining cysteines of E6-C 4C/4S are all implicated in zinc binding, which should reduce their ability to form disulfide bridges. Despite this fact the purification of E6-C 4C/4S again required particular care against oxidation, as indicated in the Experimental Procedures. In this regard, it is also interesting to compare our results with the work of Lipari et al. (40). By analyzing the proteolysis products of a "minimal core construct" of wild-type HPV16 E6, the authors obtained only one stable peptide corresponding to residues 9-84. This peptide (called E6-N) is nearly identical to our N-terminal peptide 1 (residues 7-83 of E6 6C/6S). In their report, Lipari et al. mentioned their failure to detect a

complementary C-terminal peptide, which should have been situated between residues 85-150. The authors deduced that the C-terminal half of E6 was "unstable or insoluble" and contributed to the instability and insolubility of full-length E6. By contrast, our C-terminal E6-C 4C/4S construct (residues 87-158) is folded and highly soluble. It therefore appears that the four cysteine/serine mutations present in our E6-C 4C/4S construct have dramatically improved the biochemical stability, probably by decreasing the cys-bonding reactivity of the domain. In fact, the better results obtained by Lipari et al. for the E6-N domain compared to the E6-C domain may also be explained by differences in cys-bonding reactivity. Six cysteine residues are present in wild-type HPV 16 E6-N compared to eight in E6-C. If we take into account the four zinc-binding cysteines of each domain, we are left with two free cysteines in E6-N against four free cysteines in E6-C. Therefore, two molecules of wild-type E6-N can in principle establish four distinct intermolecular cys-cys bonds, whereas two molecules of wild-type E6-C can form sixteen distinct intermolecular cys-cys bonds, thereby increasing the possibility of oligomerization leading to precipitation. It is also possible that one or several cysteines in E6-C display anomalously high cysbonding reactivity promoted by a particular chemical environment in the folded domain.

Our NMR diffusion data show that E6-C 4C/4S remains monomeric up to NMR concentrations (1 mM). Lipari et al. (40) also have found that the "minimal core" of wildtype HPV16 (residues 9-149) as well as the E6-N domain (residues 9-84) were monomeric at physiological concentrations (i.e., 1-10 uM). In contrast, older works often suggested that E6 proteins form dimers or oligomers of higher order (57, 58). Considering the sensitivity of E6 proteins to oxidation, we think that reported E6 dimers or oligomers were probably associated via cysteine bonds, and that this was a purification artefact. We have recently studied by immunogold electron microscopy the intracellular localisation of wild-type HPV-16 E6 protein in HPVpositive cervical carcinoma cells (Masson et al., submitted to publication). E6 is predominantly found in the nucleus but a fraction of the protein is also found in cytoplasm. These two compartments are reducing environments which cannot promote E6 oligomerization via cysteine bond formation.

E6-C 4C/4S domain appears as the minimal autonomous folding unit retaining the DNA-junction recognition activity of full-length E6. In a previous paper, we demonstrated that this activity was also retained by the MBP-ZD2 construct (*33*). The ZD2 fragment also includes the C-terminal zinc-binding motif but it lacks 14 N-terminal residues ($S_{87}YSLYGTTLEQQYN_{100}$) as compared to E6-C 4C/4S. In contrast to E6-C 4C/4S, the ZD2 construct binds DNA only in the presence of the MBP carrier, and ZD2 is not stable after proteolytic separation from MBP. Therefore, the S_{87} -N₁₀₀ region is not necessary for DNA-binding but it must contain secondary structure elements essential for the folding and/or stability of the domain. E6-C 4C/4S and ZD2 also differ in their DNA-binding stoichiometry. The natural stoichiometry of full-length E6/DNA complexes is 1:1 (32). We have shown here that E6-C 4C/4S also binds junctions as a single monomer consistent with the behavior of full-length E6. In contrast, two monomers of MBP-ZD2 bind to one junction on independent and equivalent sites (33). Thus, when one monomer of E6 4C/4S is bound to DNA, a second interaction site should be available on the junction for addition of a second monomer. However, access to this site is probably masked because of steric hindrance by the S₈₇-N₁₀₀ region. This situation of indirect steric hindrance has already been raised to interpret monomeric binding of HMG box proteins to four-way junctions (59).

Our NMR data have also indicated slow motions (nano- to millisecond time scale) within localized regions of the monomeric domain. The occurrence of such motions has been reported for several DNAbinding proteins (60-62). New NMR approaches have recently been developped for detailed investigation of protein motions (63). It will be interesting to apply such methods to analyze the dynamic behavior of E6 and E6 domains, and to study the possible biological relevance of these motions. More generally, the present data constitute a definitive step towards obtaining structural information on papillomavirus E6 protein. We have delimited the C-terminal domain of E6 and have produced samples allowing biophysical analysis by a variety of methods including NMR. Completion of the solution structure of the E6-C 4C/4S domain will provide precious information on the molecular basis of E6 activities.

Acknowledgements.

We are grateful to Dr Andrew Atkinson for critical reading of the manuscript.

REFERENCES

- 1. zur Hausen, H. (1991) Virology 184, 9-13.
- 2. Scheffner, M., Werness, B. A., Huibregtse, J. M., Levine, A. J., and Howley, P. M. (1990) *Cell 63*, 1129-1136.
- 3. Scheffner, M., Huibregtse, J. M., Vierstra, R. D., and Howley, P. M. (1993) *Cell* 75, 495-505.
- 4. Patel, D., Huang, S. M., Baglia, L. A., and McCance, D. J. (1999) *EMBO J 18*, 5061-5072.
- 5. Zimmermann, H., Degenkolbe, R., Bernard, H. U., and O'Connor, M. J. (1999) *J Virol* 73, 6209-6219.
- Kumar, A., Zhao, Y., Meng, G., Zeng, M., Srinivasan, S., Delmolino, L., Gao, Q., Dimri, G., Weber, G., Wazer, D., Band, H., and Band, V. (2002) *Mol Cell Biol 22*, 5801-5812.
- 7. Gross-Mesilaty, S., Reinstein, E., Bercovich, B., Tobias, K., Schwartz, A., Kahana, C., and

Ciechanover, A. (1998) Proc Natl Acad Sci U S A 95, 8058-8063.

- Ronco, L. V., Karpova, A. Y., Vidal, M., and Howley, P. M. (1998) *Genes Dev 12*, 2061-2072.
- 9. Kukimoto, I., Aihara, S., Yoshiike, K., and Kanda, T. (1998) *Biochem. Biophys. Res. com. 249*, 258-262.
- 10. Srivenugopal, K., and Ali-Osman, F. (2002) *Oncogene 21*, 5940-5945.
- Iftner, T., Elbel, M., Schopp, B., Hiller, T., Loizou, J., Caldecott, K., and Stubenrauch, F. (2002) *EMBO J.* 21, 4741-4748.
- 12. Gao, Q., Kumar, A., Srinivasan, S., Singh, L., Mukai, H., Ono, Y., Wazer, D. E., and Band, V. (2000) *J Biol Chem 275*, 14824-30.
- 13. Li, S., Labrecque, S., Gauzzi, M. C., Cuddihy, A. R., Wong, A. H., Pellegrini, S., Matlashewski, G. J., and Koromilas, A. E. (1999) *Oncogene 18*, 5727-5737.
- 14. Gao, Q., Srinivasan, S., Boyer, S. N., Wazer, D. E., and Band, V. (1999) *Mol Cell Biol 19*, 733-744.
- 15. Filippova, M., Song, H., Connolly, J., Dermody, T., and Duerksen-Hughes, P. (2002) *J Biol Chem.():*. 277, 21730-21739.
- 16. Thomas, M., and Banks, L. (1998) Oncogene 17, 2943-2954.
- 17. Tong, X., Boll, W., Kirchhausen, T., and Howley, P. M. (1998) *J Virol* 72, 476-482.
- 18. Tong, X., and Howley, P. M. (1997) *Proc Natl Acad Sci U S A. 94*, 4412-4417.
- 19. Vande Pol, S. B., Brown, M. C., and Turner, C. E. (1998) *Oncogene 16*, 43-52.
- 20. Chen, J., Reid, C. E., Band, V., and Androphy, E. J. (1995) *Science 269*, 529-532.
- 21. Du, M., Fan, X., Hong, E., and Chen, J. (2002) *Biochem Biophys Res Commun 296*, 962-969.
- 22. Kiyono, T., Hiraiwa, A., Fujita, M., Hayashi, Y., Akiyama, T., and Ishibashi, M. (1997) *Proc Natl Acad Sci U S A 94*, 11612-11616.
- 23. Nakagawa, S., and Huibregtse, J. M. (2000) *Mol Cell Biol 20*, 8244-53.
- 24. Glaunsinger, B., Lee, S., Thomas, M., Banks, L., and Javier, R. (2000) *Oncogene 19*, 5270-5280.
- 25. Lee, S., Glaunsinger, B., Mantovani, F., Banks, L., and Javier, R. (2000) *J Virol.* 74, 9680-9693.
- 26. Sedman, S. A., Barbosa, M. S., Vaa, W. C., Hubbert, N. L., Haas, J. A., Lowy, D. R., and Schiller, J. T. (1991) *Journal of Virology* 65, 4860-4866.
- 27. Morosow, A., Phelps, W. C., and Raychaudhuri, P. (1994) *Journal Biological Chemistry 2 6 9*, 18434-18440.
- 28. Dey, A., Atcha, I. A., and Bagchi, S. (1997) *Virology* 228(2), 190-199.
- 29. Gewin, L., and Galloway, D. A. (2001) J Virol 75, 7198-201.
- 30. Veldman, T., Horikawa, I., Barrett, J. C., and Schlegel, R. (2001) *J Virol 75*, 4467-72.
- 31. Oh, S. T., Kyo, S., and Laimins, L. A. (2001) *J Virol* 75, 5559-66.
- 32. Ristriani, T., Masson, M., Nominé, Y., Laurent, C., Lefèvre, J. F., Weiss, E., and Travé, G. (2000) *J. Mol. Biol.* 296, 1189-1203.

- 33. Ristriani, T., Nomine, Y., Masson, M., Weiss, E., and Trave, G. (2001) *J Mol Biol 305*, 729-39.
- 34. Androphy, E. J., Hubbert, N. L., Schiller, J. T., and Lowy, D. R. (1987) *EMBO Journal* 6, 989-992.
- 35. Cole, S. T., and Danos, O. (1987) *Journal of Molecular Biology 193*, 599-608.
- 36. Grossman, S. R., and Laimins, L. A. (1989) *Oncogene 4*, 1089-93.
- 37. Barbosa, M. S., Lowy, D. R., and Schiller, J. T. (1989) *J Virol 63*, 1404-1407.
- 38. Barbosa, M. S., Lowy, D. R., and Schiller, J. T. (1989) *Journal of Virology* 63, 1404-1407.
- 39. Kanda, T., Watanabe, S., Zanma, S., Sato, H., Furuno, A., and Yoshiike, K. (1991) *Virology 185*, 536-43.
- 40. Lipari, F., McGibbon, G. A., Wardrop, E., and Cordingley, M. G. (2001) *Biochemistry* 40, 1196-204.
- 41. Foster, S. A., Demers, G. W., Etscheid, B. G., and Galloway, D. A. (1994) *Journal of Virology* 68, 5698-5705.
- 42. Pim, D., Storey, A., Thomas, M., Massimi, P., and Banks, L. (1994) *Oncogene 9*, 1869-76.
- 43. Thomas, M., Pim, D., and Banks, L. (1999) Oncogene 18, 7690-7700.
- 44. Nominé, Y., Ristriani, T., Laurent, C., Lefèvre, J. F., Weiss, E., and Travé, G. (2001) *Protein Expr Purif.* 23, 22-32.
- 45. Nominé, Y., Ristriani, T., Laurent, C., Lefèvre, J. F., Weiss, E., and Travé, G. (2001) *Protein Eng. 14*, 297-305.
- 46. Johnson, W. C. (1999) Proteins 35, 307-312.
- 47. Piotto, M., and Saudek, V. (1992) *J Biomol NMR* 2, 661-665.
- 48. Delaglio, F., Grzesiek, S., Vuister, G. W., Zhu, G., Pfeifer, J., and Bax, A. (1995) *J. Biol. NMR* 6, 277-293.
- 49. Bartels, C., Xia, T. H., Billeter, M., Güntert, P., and Wüthrich, K. (1995) *J. Biomolecular NMR* 5, 1-10.
- 50. Gibbs, S. J., Morris, K. F., and Johnson Jr, C. S. (1991) *Journal of Magnetic resonance 94*, 165-169.
- 51. Dingley, A. J., Mackay, J. P., Chapman, B. E., Morris, M. B., Kuchel, P. W., Hambly, B. D., and King, G. F. (1995) *J Biomol NMR* 6, 321-328.
- 52. Haner, R., and Schleich, T. (1989) *Methods Enzymol.* 176, 418-46.
- 53. Everhart, C. H., and Johnson, C. S. J. (1982) J. Magn. Reson. 50, 466-474.
- 54. Duckett, D. R., Murchie, A. I. H., Diekmann, S., Von Kitzing, E., Kemper, B., and Lilley, D. M. J. (1988) *Cell* 55, 79-89.
- 55. Pisani, F. M., Manco, G., Carratore, V., and Rossi, M. (1996) *Biochemistry* 35, 9158-9166.
- 56. Dell, G., and Gaston, K. (2001) Cell Mol Life Sci. 58, 1923-1942.
- 57. Pim, D., Massimi, P., and Banks, L. (1997) *Oncogene 15*, 157-264.
- Daniels, P. R., Sanders, C. M., Coulson, P., and Maitland, N. J. (1997) *FEBS Letters 416*, 6-10.

- 59. Pohler, J. R. G., Norman, D. G., Brahmam, J., Bianchi, M. E., and Lilley, D. M. J. (1998) *EMBO J.* 17, 817-826.
- 60. Zhu, L., Hu, J., Lin, D., Whitson, R., Itakura, K., and Chen, Y. (2001) *Biochemistry 40*, 9142-9150.
- 61. Ramboarina, S., Srividya, N., Atkinson, R. A., Morellet, N., Roques, B. P., Lefevre, J. F., Mely, Y., and Kieffer, B. (2002) *J Mol Biol. 316*, 611-627.
- 62. McIntosh, P., Taylor, I., Frenkiel, T., Smerdon, S., and AN, L. (2000) *J Biomol NMR 16*, 183-96.
- 63. Cavanagh, J., and Venters, R. A. (2001) Nature Structural Biology 8, 912.

Table 1. Characteristics of the three proteolytic peptides : binding to S-sepharose resin at pH 7.4 and low salt, N-terminal peptidic sequence, molecular weight obtained by mass spectrometry, sequence deduced from experimental data, theoretical mass re-calculated from sequence, and Isoelectric Point re-calculated from sequence.

	Peptide 1	Peptide 2	Peptide 3
S-seph. Binding	not bound	bound	bound
N-term seq.	AMFQDPQE	SYSLYGTT	SYSLYGTT
Exp. Mass	9251 ± 9 Da	8011 ± 8 Da	8637 ± 8 Da
Sequences	⁷ AMFQDPSKISEY ⁸³	87SYSLYGRSSRTR153	⁸⁷ SYSLYGRRETQL ¹⁵⁸
Calc. Mass	9242.7 Da	8002.2 Da	8629.9 Da
Calc. p.I.	7.04	10.8	10.8

TABLE.

FIGURES LEGENDS.

Figure 1. Non-specific proteolysis of E6 6C/6S yields soluble fragments separable by ion-exchange chromatography. A/ Time-dependent proteolysis of E6 6C/6S in Tris buffer. E6 6C/6S was purified as previously described (45) but was kept in Tris buffer, pH 7.4, instead of phosphate buffer. SDS gel analysis of samples after 1 day, 9 days, 1 month and five months of storage at 4°C (lanes 1 to 4, respectively). B/ Fragment separation. Soluble fraction from E6 6C/6S stored 5 months in Tris buffer was diluted into low-salt Tris buffer (20 mM NaCl) then allowed to incubate with Ssepharose anion-exchange resin (Amersham Biosciences) overnight. The resin was washed three times with low salt buffer then incubated with high-salt Tris buffer (750 mM NaCl) to allow elution of the bound protein. Lane 5: total protein before incubation with resin, lane 6 : high-salt eluate containing peptides 2 and 3 which did bind to resin at low salt conditions; lane 7: flow-through containing peptide 1 which did not bind to resin at low salt conditions.

Figure 2 : Expression and purification of E6-C 4C/4S domain. Pellets of XL1 Blue bacteria expressing the MBP-E6-C 4C/4S (lane 1) construct were sonicated then ultracentrifugated. The cleared supernatant (lane 2) was loaded on amylose column. Pure MBP-E6-C 4C/4S

fusion leaked from column in the late washing steps (lane 3). Thrombin incubation led to the proteolytic separation of the MBP moiety and the E6-C 4C/4S moiety (lane 4). Digestion products were concentrated and separated by gel-filtration. Fractions containing E6-C 4C/4S were pooled and concentrated to 50 μ M (lane 5).

Figure 3. Optical characteristics of E6-C 4C/4S domain. (A) UV-absorption spectrum (10 μ M). (B) Circular dichroism spectrum (10 μ M). (C) UV-fluorescence emission spectrum (excitation wavelength at 280 nm) (0.5 μ M).

Figure 4. Binding of pure E6-C 4C/4S domain to four-way DNA junctions. 1nM of radioactively $[^{32}P]$ -labelled junction 1 (Duckett *et al.*, 1988) with four arms of 12 base pairs each, was incubated with increasing concentrations of either pure E6 6C/6S or pure E6-C 4C/4S in presence of 100 µg/ml genomic DNA and 1 mM magnesium ions. The mixtures were electrophoresed in the presence of 1 mM EDTA, then analysed by phosphorimaging. Lanes 1 to 8: 1 nM cruciform DNA was incubated with 0, 80, 125 and 250 nM of E6 6C/6S (lanes 1 to 4, respectively) and with 125, 250, 500, 1000 and 0 nM of E6-C 4C/4S (lanes 5 to 9, respectively). The labelled junction is schematized as a square cross, the E6-C 4C/4S domain as a single sphere, and full-length E6 6C/6S as a pair of spheres (representing the two domains).

Figure 5. 2D-NOESY spectrum of E6-C 4C/4S. Amide-aliphatic region of a 600 MHz NOESY spectrum recorded on a 1 mM sample of E6-C 4C/4S at pH 6.8 and 15° C. The line width of four NOE cross-peaks are indicated to illustrate the heterogeneous relaxation behavior along the sequence.

Figure 6. Diffusion of E6-C 4C/4S as measured using NMR Pulsed Field Gradients. A/.Temperature dependance of the diffusion coefficients of E6-C 4C/4S from 5°C to 40°C. The linear fit of the data points is shown by a solid line and the correlation coefficient of the fit is 0.996. B/.Comparison of normalized Dt values (20°C in water) measured for proteins of various sizes. The theoretical mass dependent diffusion value of a sphere is shown by a solid line using the measured lysosyme diffusion coefficient as a reference for a monomeric protein. The dotted line corresponds to diffusion coefficients that would be obtained for dimers. Circles and triangles indicate respectively values deduced from our experiments (unpublished results) and values published by other groups. The peptides shown on the figure are: 1/ G4R4 (1100 Da), 2/ conotoxin MVIIA (2646 Da), 3/ PMPD2 (3750 Da), 4/ E6-C 4C/4S (8630 Da), 5/ Lysosyme (14500 Da), 6/ Lysosyme (measured by Haner et al. (*52*)), 7/ Human hemoglobin (65500 Da) (measured by Everhart et al. (*53*)).

Figure 1









131

Figure 4





Figure 6





Evolution de la qualité du domaine E6-C 4C/4S en fonction des purifications

L'amélioration de la qualité du domaine C-terminal de E6 nous a permis de produire le domaine Cterminal de E6n sous une forme soluble et monomérique jusqu'à une concentration de 1 mM. La figure 33 montre l'évolution de la qualité de la protéine suivie par RMN (A/) et par DC (B/). L'éclatement des raies observé sur le spectre RMN (courbe rouge dans A/), en particulier entre 6 et 10,5 ppm et audelà de 0 ppm, dénote bien le caractère replié du domaine. Quant aux spectres de dichroïsme circulaire, la comparaison aux courbes de références de Greenfield (Fig. 14) indique clairement une augmentation significative du pourcentage d'hélice α entre les échantillons 2 et 3, ou 2 et 1, alors même que ces trois échantillons étaient monodisperses d'après le rapport d'agrégation mesuré au fluorimètre. Hormis le type de colonnes utilisées (gel filtration ou résine échangeuse d'ions), la seule différence entre ces 3 échantillons réside dans l'attention encore plus grande portée à diminuer les conditions d'oxydation et les conséquences des protéases. On constate également sur ces spectres que les échantillons mesurés à 45°C contiennent un pourcentage de conformation aléatoire plus élevé que ceux à 15°C.



-5,0E-04

-1,0E-03



Fig. 33 : comparaisons par RMN et DC de spectres du domaine E6-C 4C/4S.

A/ Progression de la qualité du repliement de la protéine, suivie par RMN, entre un premier échantillon produit en 1998 contenant le domaine déplié (courbe bleue) et l'échantillon obtenu en 2001 à partir duquel nous avons calculé la structure (courbe rouge). B/ Evolution du pourcentage de structures secondaires suivie par DC pour différents lots de protéines (la température d'acquisition est indiquée entre parenthèses). Échantillon 1 : gelfiltration (décembre 2000) ; échantillon 2 : gel-filtration (octobre 2000) ; échantillon 3 : colonne SP-Sépharose FF (octobre 2000). L'échantillon 1 est celui qui a été utilisé pour enregistré les spectres RMN homonucléaires.

<u>PARTIE II</u>

Etude structurale par RMN du domaine C-terminal de E6 HPV16

Les résultats de cette partie sont présentés dans :

- Publication 4 : Assignment and secondary structure of the C-terminal domain of HPV16 E6 oncoprotein. <u>Nominé Y.</u>, Charbonnier S., Kieffer B., Travé G. (Publication soumise à *Journal of Biomolecular NMR*).
- un chapitre sur le calcul par RMN de la structure du domaine C-terminal de E6.

II.2/ Publication 4 (en cours de soumission)

Nominé Y., Charbonnier S., Kieffer B., Travé G. : ¹H, ¹⁵N resonances assignment and secondary structure of the C-terminal domain of human papillomavirus 16 oncoprotein E6. Publication soumise à *Journal of Biomolecular NMR*.

L'obtention précédemment décrite du domaine C-terminal de E6, sous une forme repliée, monomérique, stable et à une concentration de l'ordre du milli-molaire, nous a permis d'entamer son étude structurale par RMN 2D. Cependant, en raison du recouvrement d'un trop grand nombre de pics, l'attribution complète des protons de la molécule n'a été rendu possible qu'avec l'utilisation conjuguée d'un échantillon marqué à l'azote 15 et de la RMN 3D. Cette publication présente quelques-unes des étapes déterminantes pour cette attribution, en particulier la reconnaissance sur les spectres des deux histidines de la protéine grâce une HMQC à faible constante de couplage.

Une version papier de la table d'attribution se situe en annexe de ce manuscrit.

¹H, ¹⁵N resonances assignment and secondary structure of the C-terminal domain of human papillomavirus 16 oncoprotein E6.

Yves Nominé[‡], Sebastian Charbonnier[‡], Gilles Travé[‡]* and Bruno Kieffer[⊥]

(‡) Laboratoire d'Immunotechnologie, UMR CNRS 7100, Ecole Supérieure de Biotechnologie de Strasbourg, 67400 Illkirch, France.

 (\perp) Laboratoire de RMN, UMR CNRS 7104, Ecole Supérieure de Biotechnologie de Strasbourg, 67400 Illkirch, France.

* Corresponding author

tel: 0033 3 90 24 47 65, fax: 0033 3 90 24 47 70, e-mail: trave@esbs.u-strasbg.fr

Running title : Assignment of the C-terminal domain of HPV-16 E6

Keywords : heteronuclear NMR assignments, HPV16 oncoprotein E6, C-terminal domain

Biological context

E6 is one of two oncoproteins produced by "highrisk" human papillomaviruses (HPVs) responsible for cervical cancers (zur Hausen, 1991). This protein promotes tumorigenesis by stimulating cellular degradation of the tumour suppressor p53 via formation a trimeric complex comprising E6, p53 and the cellular ubiquitination enzyme E6AP (Scheffner et al., 1993). Two forms of HPV16 E6 are produces in vivo, 158 and 151 residues-length, depending on which methionine is used as a start codon (Androphy et al., 1987). Although the first E6-encoding DNA sequences was described in 1985 (Schwarz et al., 1985), no structural data are available for the wild-type E6 protein. Sequence alignments of E6s from numerous HPV subtypes suggested the presence of two zinc-binding motifs (Cole & Danos, 1987). Recently, we designed and produced a stable, folded and monomeric form of the C-terminal domain of HPV 16 E6, called E6-C 4C/4S, in which 4 cysteines not implicated in zinc binding are mutated into serines (Nominé et al. paper submitted). This domain encompasses residues 87 - 158 of the longest transcript of fulllength protein and allowed us to define residues involved in the interaction with p53 (Ristriani et al. in press). Furthermore, we have shown that this domain of E6 interacts specifically with a cruciform DNA (Ristriani et al., 2001).

Here we present the nearly complete ¹H and ¹⁵N resonance assignments for the C-terminal domain of E6-C 4C/4S from HPV 16 and its secondary structure.

Methods and experiments

We produced unlabelled and uniformly ¹⁵N-labelled samples of E6-C 4C/4S by growing *E. coli* in LB or M9 mimimal media, respectively. Details of the expression and purification will be published elsewhere. Sample containing 1 mM E6-C 4C/4S was dissolved in 20 mM deuterated Tris at pH 7.0 in H_2O/D_2O (9:1).

All heteronuclear experiments were recorded at 288 K on a Bruker DRX500 spectrometer equipped with a z-gradient triple resonance probe. Homonuclear proton spectra were acquired at 278, 288 and 298 K on a Bruker DRX600 spectrometer. 2D and 3D data were processed using UXNMR (Bruker Inc.) and NMRPipe (Delaglio *et al.*, 1995) respectively. Spectra were assigned manually and peaks integration was performed using XEASY software (Bartels *et al.*, 1995). Dihedral ϕ angles were obtained from ³J_{NH-Ha} coupling constants measured in a HNHA spectrum (Vuister & Bax, 1993). The both samples allowed us to obtained complete full assignment of the proton and nitrogen frequencies (Figure 1).

The nitrogen resonance assignments of the His125 ring and subsequently its protonation state was determined using a long range HMQC spectrum (Pelton *et al.*, 1993) in which the nitrogen spectral width was set to 200 ppm.

Extent of assignment and data deposition

Backbone resonances $({}^{1}H^{N}, {}^{15}N \text{ and } {}^{1}H^{\alpha})$ have been assigned for all but the first two residues of the protein and the nitrogen frequencies of Thr93 and Arg153. Complete but non-stereospecific assignments were obtained for side-chains protons. The resonance assignments for Asn134 (H α : 4.28; Hβ:1.83 & -0.30; Hδ : 5.75 & 3.86 and Nδ: 107.0) were particulary striking being far from random coil values. The reason for these effects re not yet clear. Analysis of short-range contacts, torsion angles and chemical shifts allows to identified two helices encompassing residues 95-101, 120-129 and a β -stranded encompassing residues 88-90, 132-135 and 138-140 (Figure 2). The percentages of α -helical and β -sheet content are consistent with estimations obtained by circular dichroism, $22\pm1\%$ and $22\pm3\%$, respectively (Nomine et al. paper submitted). Characterization of the backbone mobility of the E6-C 4C/4S domain showed a strongly heterogeneous dynamic behaviour which may explain the difficulties reached to obtain the tridimensional structure. In particular, in the vicinity of the first stretch of the zinc-liganding cysteines residues are subjected either to high-frequency motions (D105, L106, C110 and C113) or to proton and nitrogen line broadening (L107, I108 and I111). The ¹H and ¹⁵N chemicals shifts have been deposited in the BioMagResBank (http://www.bmrb.wisc.edu) under accession number AAAA.

Acknowledgements

This work was supported by a grant from the Association pour la Recherche sur le Cancer and by a "PCV" grant from the CNRS.

Bibliography

- Androphy, E., Hubbert, N., Schiller, J. & Lowy, D. (1987). Identification of the HPV-16 E6 protein from transformed mouse cells in human cervical carcinoma cell lines. *EMBO J.* **6**, 989-992.
- Bartels, C., Xia, T., Billeter, M., Güntert, P. & Wütrich, K. (1995). The programm XEASY for computersupported NMR spectral analysis of biological macromolecules. *J. Biomol. NMR* **5**, 1-10.
- Cole, S. & Danos, O. (1987). Nucleotide sequence and comparative analysis of the human papillomavirus type 18 genome. Phylogeny of papillomaviruses and repeated structure of the E6 and E7 gene products. *J. Mol. Biol.* **193**(4), 599-608.
- Delaglio, F., Grzesiek, S., Vuister, G., Zhu, G., Pfeifer, J. & Bax, A. (1995). NMRPipe: a multidimensional

spectral processing system based on UNIX pipes. J. Biomol. NMR 6(3), 277-93.

- Merutka, G., Dyson, H. & Wright, P. (1995). 'Random coil' 1H chemical shifts obtained as a function of temperature and trifluoroethanol concentration for the peptide series GGXGG. *J. Biomol. NMR* **5**(1), 14-24.
- Pelton, J., Torchia, D., Meadow, N. & Roseman, S. (1993). Tautomeric states of the active-site histidines of phosphorylated and unphosphorylated IIIGlc, a signaltransducing protein from Escherichia coli, using twodimensional heteronuclear NMR techniques. *Protein Sci.* 2(4), 543-58.
- Ristriani, T., Nomine, Y., Masson, M., Weiss, E. & Trave, G. (2001). Specific recognition of four-way DNA junctions by the C-terminal zinc-binding domain of HPV oncoprotein E6. *J. Mol. Biol.* **305**(4), 729-39.
- Scheffner, M., Huibregtse, J., Vierstra, R. & Howley, P. (1993). The HPV-16 E6 and E6-AP complex functions as a ubiquitin-protein ligase in the ubiquitination of p53. *Cell* **75**, 495-505.
- Schwarz, E., Freese, U., Gissmann, L., Mayer, W., Roggenbuck, B., Stremlau, A. & zur Hausen, H. (1985). Structure and transcription of human papillomavirus sequences in cervical carcinoma cells. *Nature* **314**, 111-114.
- Vuister, G. & Bax, A. (1993). Quantitative J correlation: a new approach for measuring homonuclear three-bond J(HN-Hα) coupling constants in 15N-enriched proteins. *J. Am. Chem. Soc.* **115**, 7772-7777.
- zur Hausen, H. (1991). Human papillomaviruses in the pathogenesis of anogenital cancer. *Virology* **184**, 9-13.

Figure 1. Representative 2D [¹H,¹⁵N]-HSQC spectrum recorded at pH 6.8, 288 K and 500 MHz on ¹⁵N-labelled E6-C 4C/4S.. Cross peaks are marked by residue number, according to the numbering of full-length HPV16 E6 (158-residue form). Side chain NH₂ resonances of Asn and Gln residues are connected by horizontal bars, excepted for Asn134, for which only one proton appears in this spectrum.

Figure 2. Sequence of E6-C 4C/4S with a summary of short-range (| i-j | \leq 4) NOEs, ³J_{HNHa} coupling constants, secondary chemical shifts, and R2 values. Residue numbering is that of full-length HPV16 E6 sequence (158-residue form). Residues 84-86 (GAM) result from cloning and TEV proteolysis and do not appear in the wild-type sequence. The presence of a NOE cross peak is indicated by a line connecting the two residues. The thickness of the line is proportional to the intensity of the NOE. ³J_{HNHa} coupling constants < 6.5 Hz and > 7.5 Hz are indicated by closed and open squares, respectively. Shift(α) represents the difference between H_a chemical shifts values and values extracted from the database of Merutka (Merutka *et al.*, 1995) for each amino acid: differences > 0.2 ppm and < -0.2 ppm are indicated by a closed square below and above the line, respectively. In the dynamics line, R2 values that deviate significantly from the trimmed mean value of 12 s⁻¹ are indicated by black (R2 > 13 s⁻¹) and white (R2 < 11 s⁻¹) circles. Crosses indicate residues for which no data could not be obtained (overlapping or unobserved peaks).

Figure 1



Figure 2



II.2/ Calculs de la structure du domaine C-terminal de E6 HPV 16

Nominé Y., Kieffer B., Travé G.

<u>Avertissement</u> : je précise que ce chapitre présente uniquement la méthode d'obtention de la structure tri-dimensionnelle du domaine C-terminal de E6, mais en aucun cas l'analyse de cette structure. J'ai en effet finalisé ce travail quelques jours avant d'entamer la rédaction de ce manuscrit.

Qualité des modèles

Après la procédure d'affinement, les structures ont été sélectionnées sur la base de leur énergie globale. Les structures ne répondant pas aux critères suivants ont été écartées : i/ pas de violations de la géométrie définie par le champ de force, ii/ pas de violations de contraintes de distance supérieures à 0,5Å, iii/ pas de violations d'angles dièdres excédant 10°. Ce tri a permis de retenir la totalité des 20 structures calculées.

L'accord entre les modèles calculés et les données expérimentales de distance est décrit dans la figure 34. En particulier, cette représentation met clairement en évidence l'arrangement antiparallèle des brins β compris entre les résidus 88-90 et 132-135. D'autre part, les nombreux contacts observés entre les régions 106-113 et 140-144 correspondent aux proximités spatiales imposées par la coordination du zinc.

La superposition des 10 structures affinées nous renseigne sur la précision du modèle (Fig. 35A). Après superposition des atomes de la chaine carbonée principale des 10 meilleures structures, une structure moyenne est générée sous Xplor, suivie d'une minimisation. L'écart-type entre les 10 modèles et cette structure moyenne dans la région 88-146 est de 1,2 \pm 0,2 Å (Tableau 8). Les écart-types par résidu calculés localement et globalement, ainsi que le nombre de contraintes par résidu sont présentés dans la figure 36A. La courbe d'ajustement local (superposition sur une fenêtre de 5 résidus) indiquent que les éléments de structures secondaires sont bien définis, en particulier les hélices et les deux derniers brins beta. Cependant, cette courbe met en évidence plusieurs régions de la structure moins bien définies correspondant à des résidus présentant un faible nombre de contacts expérimentaux. C'est le cas en particulier des deux régions de séquence à proximité du site de coordination du zinc : la région 114 à 116 qui contient un résidu proline, ainsi que l'extrêmité C-terminale. Cette observation est corrélée avec la position excentrée de ce site par rapport au cœur de la structure. La variabilité de position de ce site se répercute sur la qualité de la superposition globale.

L'analyse de la relaxation transversale des noyaux azote 15 le long de la séquence (Fig. 36B) montre que plusieurs résidus sont affectés par des phénomènes d'échanges conformationnels lents. Ces résidus sont localisés au sein du premier brin β , ainsi que sur la partie C-terminale de l'hélice 120-127. Ces résidus sont proches dans la structure tri-dimensionnelle (Fig. 35B) et sont sans doute sensibles à un équilibre conformationnel localisé dont la nature reste à déterminer.

Rmsd moyen entre les 10 structures et la r autres que les protons dans la chaine carbonée	Tableau 8 : Résumé de l'analyse statistique des	
résidus 96-101 (hélice α1)	$0,37 \pm 0,09$	structures calculées par recuit simulé. Econ
résidus 120-128 (hélice α2)	$0,33 \pm 0,06$	$E_{NOE}, E_{CDIH}, E_{VdW}, E_{liaison}$
résidus 88-90, 132-140 (feuillet β)	$0,\!48\pm0,\!09$	et représentent les énergies des angles de
résidus 110-113, 143-146 (site à zinc)	1,0 ±0,4	torsion, des contraintes
idus 88-146 $1,2 \pm 0,2$		NOE, des contraintes des constantes de couplage, des termes de répulsion de Van der
Rmsd par rapport à la géométrie idéale		
NOEs (Å)	$0,038 \pm 0,007$	Waals et des energies de liaison dans la fonction
Contraintes dièdres (deg)	2 ± 1	du recuit simulé.
Liaisons covalentes (Å)	$0,\!0048 \pm 0,\!0009$	
Angles (deg)	$0,8 \pm 0,1$	
Impropres (deg)	$0,7 \pm 0,2$	
Energies (kcal.mol ⁻¹)		
Etor	180 ± 20	
E _{NOE}	59 ± 13	
E _{CDIH}	10 ± 7	
E _{VdW}	52 ± 13	
E _{liaison}	19 ± 3	

Description de la structure

Le repliement de E6-C 4C/4S est organisé en trois courtes hélices α (α_1 sur les résidus 96 à 99, α_2 sur les résidus 103 à 106, et α_3 sur 120 à 127) qui se plaquent du même côté d'un feuillet β constitué de trois brins anti-parallèles (β_1 sur les résidus 88 à 90, β_2 sur 132 à 135 et β_3 sur 138 à 140 avec un coude de type II centré sur les résidus R₁₃₆ et G₁₃₇). Les 4 cystéines conservées se rassemblent pour former le site de liaison au zinc, qui remarquablement occupe une position plutôt périphérique par rapport au reste de la structure. Les deux cystéines C-terminales (C₁₄₃ et C₁₄₆) ont tendance à former un début d'hélice α . Les résidus au delà de l'Arg₁₅₁ ne présentant aucun contact autre que des interactions entre proches voisins, ne sont pas pris en compte dans ces calculs de structures. Le repliement laisse apparaître un cœur hydrophobe constitué principalement des acides aminés Leu₉₀, Leu₁₀₃, Ile₁₀₈, Leu₁₁₇, Leu₁₂₆, Phe₁₃₂ et Trp₁₃₉. Notons également que trois résidus non-hydrophobes Glu₉₆, Lys₁₂₂ et Asn₁₃₄ sont enfouis à l'intérieur de la structure. À l'inverse, trois résidus hydrophobes de la partie Nterminale du domaine sont exposés au solvant : Tyr₈₈, Tyr₉₁, et Leu₉₅, faisant apparaître un patch hydrophobe caractéristique, situé à l'opposé du site à zinc (Fig. 37 A et B). Ces positions –et particulièrement la Leu₉₅– étant conservées dans tous les HPVs, laissent penser que cette surface hydrobe exposée constitue probablement l'interface avec le domaine N-terminal de E6. Une autre propriété remarquable du domaine est la répartition polarisée du potentiel électrostatique de surface (Fig. 37C). Les charges négatives, minoritaires dans le domaine, sont regroupées sur une même face (orientée vers le haut sur cette représentation) alors que les charges positives, majoritaires, sont regroupées sur la face opposée (orientée vers le bas). Cette répartition favorise probablement l'interaction avec l'ADN. Parmi les résidus basiques portant ces charges positives, plusieurs sont apparus indispensables pour l'interaction avec l'ADN lors de notre étude par mutagénèse dirigée (Cf. Partie III des résultats). La répartition surfacique de ces résidus est représentée sur la figure 37D.



Fig. 34 : Comparaison des matrices de contacts déduites, soit des contraintes expérimentales (A), soit des structures calculées. L'intensité (par rapport à l'échelle de niveau de gris) de chaque "pic" (i,j) est proportionelle au nombre de contacts entre les résidus i et j. (l'échelle des gris n'est pas le même en A/ et en B/).



Fig. 35 : La structure RMN du domaine C terminal de E6 6C/6S. A/ Superposition de la chaîne principale des 20 structures calculées. B/ Vue stéréographique d'une représentation en ruban de la structure tridimensionnelle (même orientation qu'en A). Les éléments de structure secondaires ont été définis sur la structure moyenne par le logiciel MOLMOL (Koradi, 1996). Les chaînes latérales des résidus enfouis à l'intérieur du domaine sont représentées (hydrophobes en rose, basiques en bleu, acides en rouge, cystéines en marron) ainsi que l'atome de zinc.



Fig. 36 : Données RMN pour la structure du domaine. A/ La courbe représente la moyenne du RMSD entre les 20 structures calculées et la structure moyenne (calculé pour les N,C, O de la chaine principale pour chaque résidu). L'histogramme représente le nombre de distances NOE par résidu, colorées en rouge, bleu clair et jaune pour les distances impliquant des résidus consécutifs (i=1), moyennement proches (i<5), ou éloignés (i>=5) dans la séquence. B/ Valeurs de relaxation transversale pour chaque résidu (les barres indiquent l'erreur de la mesure).


Fig. 37 : Différentes propriétés de surface du domaine. La structure est chaque fois représentée dans deux orientations opposées, reliées par une rotation de 180° autour de l'axe vertical. Les représentations ribbon sur les côtés permettent de visualiser l'orientation du domaine dans les représentations de surfaces. Entre les représentations de B sont conservées dans les vues C et D. A/ Répartition surfacique des résidus hydrophobes (histidines en bleu fonçé, tyrosines en vert, autres hydrophobes en bleu clair). B/ Visualisation du patch hydrophobe sur deux vues ayant subi une rotation additionelle de -90°. Noter la localisation du patch hydrophobe sur la face située à l'opposé du site à zinc. C/ Répartition surfacique du potentiel électrostatique (charges négatives en rouge et charges positives en bleu). D/ Répartition surfacique des résidus basiques (en bleu) dont la mutagénèse a affecté fortement l'interaction avec l'ADN.

PARTIE III

Identification et caractérisation de l'interaction entre l'oncoprotéine E6 et l'ADN cruciforme.

Les résultats de cette partie sont présentés dans :

- 3 publications :
 - Publication 5 : HPV Oncoprotein E6 is a structure-dependant DNA-binding protein that recongnizes four-way DNA junction. Ristriani T., Masson M., <u>Nominé Y.</u>, Laurent C., Lefèvre J-F., Weiss E., Travé G. (2000) *Journal of Molecular Biology* 296 1189-1203
 - Publication 6 : Specific recognition of four-way DNA junctions by the C-terminal zincbinding domain of HPV oncoprotein E6. Ristriani T., <u>Nominé Y.</u>, Masson M., Weiss E., Travé G. (2001) *Journal of Molecular Biology* **305**, 729-739
 - Publication 7 : Protein mutagenesis with monodispersity-based quality probing: selective inactivation of p53 degradation and DNA binding properties of HPV E6 oncoprotein. Ristriani T., Nominé Y., Laurent C., Weiss E., Travé G. (2002) *Protein Expression and Purification* (sous presse).
- un chapitre sur l'étude de l'interaction E6 / ADN cruciforme par la technique de Résonance Plasmonique de Surface (technologie BIAcore).

III.1/ Publication 5

Ristriani T., Masson M., **Nominé Y**., Laurent C., Lefèvre J-F., Weiss E., Travé G. (2000) : HPV Oncoprotein E6 is a structure-dependant DNA-binding protein that recongnizes four-way DNA junction. *Journal of Molecular Biology* **296** 1189-1203

L'activité de liaison et de dégradation de p53 par E6 est connue depuis de nombreuses années. Pourtant, E6 pourrait avoir un rôle dans d'autres activités non liées à p53. En parallèle, une similarité de séquence a été décrite entre E6 et l'endonucléase VII, une protéine du phage T4 qui reconnaît spécifiquement et coupe les jonctions de « Holidays ». Cette similarité laissait envisager une interaction entre E6 et l'ADN cruciforme Cette hypothèse n'avait cependant jamais été confirmée. Nous pensons *à posteriori* que ce problème était lié au problème d'obtention d'une protéine de fusion de qualité suffisante.

Dans cet article, nous démontrons qu'une interaction existe entre cet ADN cruciforme et l'oncoprotéine E6 des souches malignes. Cette interaction est spécifique de la structure de l'ADN, et non de sa séquence. D'autres paramètres tels que l'affinité, la stœchiométrie, l'influence des ions magnésiums, ou encore la conformation de l'ADN, ont également été étudiés.

Ce travail est la première démonstration d'une liaison spécifique entre E6 et un ADN, suggérant de nouveaux modes d'actions de E6, particulièrement importants pour son rôle oncogène.



[Signalement bibliographique ajouté par : ULP – SCD – Service des thèses électroniques]

HPV Oncoprotein E6 is a structure-dependant DNA-binding protein that recongnizes four-way DNA junction.

Ristriani T., Masson M., Nominé Y., Laurent C., Lefèvre J-F., Weiss E., Travé G.

Journal of Molecular Biology, 2000, Vol. 296, Pages 1189-1203

Pages 1189 à 1203 :

La publication présentée ici dans la thèse est soumise à des droits détenus par un éditeur commercial.

Pour les utilisateurs ULP, il est possible de consulter cette publication sur le site de l'éditeur : http://dx.doi.org/10.1006/jmbi.2000.3527

Il est également possible de consulter la thèse sous sa forme papier ou d'en faire une demande via le service de prêt entre bibliothèques (PEB), auprès du Service Commun de Documentation de l'ULP: peb.sciences@scd-ulp.u-strasbg.fr

III.2/ Publication 6

Ristriani T., **Nominé Y**., Masson M., Weiss E., Travé G. (2001) : Specific recognition of four-way DNA junctions by the C-terminal zincbinding domain of HPV oncoprotein E6. *Journal of Molecular Biology* **305**, 729-739

Nous mettons ici en évidence que le domaine de liaison au zinc C-terminal de E6 est seul responsable de l'activité de liaison avec l'ADN cruciforme. Je me suis en particulier intéressé à la stœchiométrie de cette interaction. J'ai démontré, en analysant les variations d'intensités observées sur gels retard par un modèle développé sur le logiciel MatLab, que ce domaine se lie à deux sites identiques et indépendants, et ce contrairement à la protéine E6 entière.

Par ailleurs, d'après les alignements de séquences, ce domaine semblerait représenter une nouvelle classe de module à zinc se liant à l'ADN. En précisant la région de E6 responsable de cette interaction, ces résultats ouvrent donc la voie à la conception de mutants ponctuels ayant perdu cette activité.



[Signalement bibliographique ajouté par : ULP – SCD – Service des thèses électroniques]

Specific recognition of four-way DNA junctions by the C-terminal zincbinding domain of HPV oncoprotein E6.

Ristriani T., Nominé Y., Masson M., Weiss E., Travé G.

Journal of Molecular Biology, 2001, Vol. 305, Pages 729-739

Pages 729 à 139 :

La publication présentée ici dans la thèse est soumise à des droits détenus par un éditeur commercial.

Pour les utilisateurs ULP, il est possible de consulter cette publication sur le site de l'éditeur : <u>http://dx.doi.org/10.1006/jmbi.2000.4330</u>

Il est également possible de consulter la thèse sous sa forme papier ou d'en faire une demande via le service de prêt entre bibliothèques (PEB), auprès du Service Commun de Documentation de l'ULP: peb.sciences@scd-ulp.u-strasbg.fr

III.3/ Publication 7

Ristriani T., **Nominé Y**., Laurent C., Weiss E., Travé G. (2002) : Protein mutagenesis with monodispersity-based quality probing: selective inactivation of p53 degradation and DNA binding properties of HPV E6 oncoprotein. *Protein Expression and Purification* (sous presse).

L'idée de ce travail est de séparer, pour la protéine E6, l'activité de liaison à l'ADN cruciforme, de celle de dégradation de la protéine gardienne du génome p53. Ce travail est une étape obligatoire pour l'étude *in vivo* de l'activité de liaison à l'ADN par E6 dans le processus oncogène de E6.

Nous utilisons la protéine de fusion MBP-E6 6C/6S comme protéine de référence, puisque nous savons que ce mutant dégrade p53 tout en interagissant avec l'ADN cruciforme ; il présente de plus l'avantage d'être monodisperse après purification. Pour l'ensemble des 19 mutants analysés, 3 paramètres sont mesurés. Le premier concerne la détermination de l'agrégation de la protéine. La recherche d'un état monodisperse nous semble en effet constituer un facteur à respecter préalablement à tout test biologique : lorsqu'un mutant est plus agrégé que la protéine de référence, nous considérons que cette molécule a perdu tout ou partie de son repliement, rendant les conclusions de tests biologiques totalement caduques. Les mutants ayant ainsi conservé leur intégrité structurale sont ensuite soumis aux tests de dégradation de p53, ainsi que celui de liaison à l'ADN cruciforme.

Les résultats de cette étude confirment dans un premier temps que le domaine C-terminal de E6 est responsable de l'activité de liaison au cruciforme. Plus intéressant, plusieurs mutations ont pu ainsi être mises en évidence : d'un côté F54R, F54D et Q114A sont déterminants pour l'activité de liaison à l'ADN, tandis que K115A affecte la dégradation de p53. Ces mutants constituent donc des candidats idéaux pour l'étude *in vivo* de l'activité oncogène de E6.



[Signalement bibliographique ajouté par : ULP – SCD – Service des thèses électroniques]

Protein mutagenesis with monodispersity-based quality probing: selective inactivation of p53 degradation and DNA binding properties of HPV E6 oncoprotein.

Ristriani T., Nominé Y., Laurent C., Weiss E., Travé G.

Protein Expression and Purification, 2002, Vol. 26, Pages 357-367

Pages 357 à 367 :

La publication présentée ici dans la thèse est soumise à des droits détenus par un éditeur commercial.

Pour les utilisateurs ULP, il est possible de consulter cette publication sur le site de l'éditeur : <u>http://dx.doi.org/10.1016/S1046-5928(02)00570-3</u>

Il est également possible de consulter la thèse sous sa forme papier ou d'en faire une demande via le service de prêt entre bibliothèques (PEB), auprès du Service Commun de Documentation de l'ULP: peb.sciences@scd-ulp.u-strasbg.fr

III.4/ Étude de l'interaction E6 / ADN cruciforme par BIAcore

Nominé Y., Ristriani T., Choulier L., Altschuh D., Travé G. : Interaction of HPV E6 Oncoprotein with Cruciform DNA: a BIAcore Study

Nous avons montré, par la technique de gel retard, que E6 interagit avec des ADN de structure cruciforme. Cependant, cette technique ne permet pas de déterminer l'influence du magnésium sur l'interaction. Ici, grâce au BIAcore, j'ai pu montrer que cette interaction existe également en présence d'ions magnésium. Ce point est d'autant plus crucial pour l'interaction que cet ion est un ion physiologique présent dans les noyaux des cellules. Par ailleurs, en développant un modèle non standard de traitement des données, j'ai mis en évidence deux constantes d'association distinctes, impliquant deux sites de fixation de la protéine sur l'ADN cruciforme : la constante la plus élevée correspond à l'interaction structure - dépendante observée par gel retard.

Ce projet a été initié dans le cadre d'une collaboration avec le groupe de Danièle Altschuh à l'IBMC fin 1998.

Ces résultats ont ensuite fait l'objet d'un poster lors du IIIème congrès Européen de Biophysique en 2000 à Munich, et de deux présentations orales (2^{ème} journée nationale du BIAcore en mai 2000 à Paris ; rencontre d'utilisateurs BIAcore en mars 2001 à Tours).

Au laboratoire, Ristriani *et al.* ont démontré, par la méthode des gels électrophorétiques de retard d'ADN, que E6 lie sélectivement un type particulier d'acide nucléique (l'ADN cruciforme), et ce même en présence d'un large excès d'ADN double-brins. Cette interaction est dépendante de la structure de l'ADN et non de sa séquence (Cf. publication 5).

Puisque le magnésium est présent dans les noyaux des cellules, il est important de savoir si E6 interagit toujours avec cet ADN en présence de cet ion. Or justement, les expériences par gel retard d'ADN ne répondent pas entièrement et de manière satisfaisante à cette question. Dans ces expériences, la protéine E6 et l'ADN cruciforme sont tous deux incubés en présence d'ions magnésium. Si ce mélange est chargé sur un gel électrophorétique contenant de l'EDTA –un complexant des ions bivalents–, une interaction est détectée. Au contraire, lorsque ce même mélange est chargé sur un gel contenant du magnésium, les complexes ne sont plus détectés (Cf. publication 5). Ces résultats suggèrent que la présence du magnésium n'affecte pas la formation du complexe, mais uniquement la détection du complexe par cette technique. Des expériences de chromatographie allaient d'ailleurs dans le même sens. Il restait donc une ambiguité sur le rôle du magnésium, qui oblige à confirmer l'existence de cette interaction en présence de magnésium par une autre technique.

Nous nous sommes alors portés vers la technologie BIAcore qui offrait pour ce problème plusieurs avantages : i/ la possibilité de travailler avec des petits volumes (quelques dizaines de micro-litres) à des concentrations inférieures à 0,1 mM. ii/ un contrôle strict des conditions expérimentales permettant d'évaluer l'effet sur l'interaction d'une large gamme de tampon, de la température, du flux. iii/ l'acquisition en temps réel, permettant une évaluation des constantes de cinétiques. Cette technologie est très sensible à des paramètres tels que la pureté de l'échantillon ou son agrégation. La qualité de la protéine nouvellement purifiée nous laissait donc envisager de telles études.

Cependant, alors que la démonstration de l'existence d'une interaction par BIAcore est relativement facile, la détermination des paramètres cinétiques peut s'avérer délicate. La qualité des données expérimentales est alors cruciale, surtout lorsque l'on travaille sur un système différent d'interaction différent du modèle de Langmuir (A + B \iff AB), comme ce sera le cas.

Afin d'obtenir des données de la meilleure qualité possible, je me suis basé principalement sur deux articles : l'un énumère, pour chaque étape, les conditions expérimentales optimales (Myszka, 1999),

le second s'attarde en particulier sur les paramètres déterminants lorsque le système ne suit plus le modèle de Langmuir (A + B <==> AB) (Morton, 1998).

C.III.4.a/ Conception de l'expérience

Il est absolument nécessaire de travailler avec des réactifs de bonne qualité. Les molécules sont donc systématiquement analysées avant chaque nouvelle série d'expériences : soit par un gel pour la pureté et la masse des ADN, soit par un gel et par une mesure d'agrégation pour la protéine.

Il nous a semblé plus opportun que ce soit l'ADN qui constitue le ligand et E6 l'analyte. La première raison est que l'acide nucléique est moins fragile que la protéine, facilitant ainsi la recherche des conditions de régénération. De plus, en utilisant E6 comme ligand, l'ADN (alors analyte) aurait pu éventuellement se fixer simultanément à deux molécules de E6 (ligand) : cet état complique beaucoup l'interprétation des résultats.

Le système détectant toute modification à la surface, il est nécessaire de multiplier les témoins pour s'assurer que les réponses observées soient effectivement reliées directement à l'interaction E6 / ADN cruciforme. Parmi les 4 cellules disponibles sur une même surface, les références sont choisies afin de tester une unique condition : streptavidine seule et ADN cruciforme pour tester l'affinité de E6 sur la streptavidine ; ADN simple-brin, double-brin et cruciforme pour vérifier l'affinité de E6 sur ces différents ADN ; ADN cruciformes biotinylés sur deux bras différents (X ou B) pour analyser l'effet de l'orientation de ces ADN. Enfin, pour éviter les problèmes de transport de masse, il est préférable de limiter la quantité de ligands accrochés à la surface. Expérimentalement, nous nous limitons à une quantité d'ADN de l'ordre de 200 RU, donnant un signal d'analyte de l'ordre de quelques dizaines de RU. Cette valeur est bien supérieure à la sensibilité du BIAcore 2000 (1 à 2 RU).

C.III.4.b/ Détermination de la concentration active de protéine E6

En se placant dans des conditions de transport de masse, nous avons été en mesure de déterminer la concentration de protéine active sur un lot de E6. Typiquement, pour E6 interagissant avec l'ADN cruciforme, la protéine active représente environ 3 à 5% de la concentration totale de protéine. Cette valeur peut paraître de prime abord très faible. Il faut cependant remarquer que les pourcentages établis par BIAcore n'excèdent que très rarement 50% (Zeder-Lutz, 1999).

C.III.4.c/ Analyse de la phase de dissociation

Parmi les différents modèles proposés (voir Tableau 4 dans le chapitre « *Techniques biophysiques* »), l'analyse de la phase de dissociation permet d'estimer celui à utiliser pour traiter l'ensemble des courbes de cinétiques.



Fig. 38 : ajustement global pour deux modèles (Langmuir ou parallèle) des courbes de dissociation obtenues pour l'ADN cruciforme de Bianchi en présence de Dextran (1 mg/ml) et de magnésium (1 mM).

Ainsi, la comparaison des ajustements selon les modèles parallèle ou de Langmuir des courbes de dissociation permet d'affirmer qu'il existe deux vitesses de dissociation distinctes correspondant probablement à deux sites de liaison de la protéine E6 sur l'ADN cruciforme de type Bianchi (Fig. 38).

C.III.4.d/ Analyse globale des phases d'association et de dissociation

Pourtant, les résultats obtenus par gel retard indiquent, en présence d'un large excès d'ADN génomique, qu'une seule protéine de E6 est liée à une molécule d'ADN cruciforme. Le modèle décrit ci-après (Fig. 39), adapté du modèle parallèle classique, permet de tenir compte des deux sites de liaisons décrits précédemment : l'un sur les bras du cruciforme (à priori de moins forte affinité) nommé site « double-brin » ou site 1, et l'autre à la jonction des deux brins du cruciforme, nommé site « spécifique du cruciforme » ou site 2.

Le modèle parallèle comporte 6 paramètres à déterminer par ajustement, et utilise deux analytes distincts en solution (Fig. 39). Au contraire, le modèle adapté fait l'hypothèse d'un unique analyte interagissant sur deux sites distincts du ligand. L'introduction d'un unique paramètre R_{max} , pondéré par les nombres n ou 1, permet de redéterminer les deux paramètres $R_{max}1$ et $R_{max}2$. Lorsque le paramètre n est laissé libre, l'ajustement donne des valeurs de n comprises entre 1000 et 2000. Pour y remédier, ce paramètre est fixé lors de chaque ajustement, ce dernier étant répété pour des valeurs de n comprises entre 1 et 20. Cette modification permet par la même occasion de réduire le nombre de paramètres à ajuster de 6 à 5, et donc augmente la rigueur de l'ajustement.



Fig. 39 : illustration du modèle parallèle adapté. *Ce modèle tient compte de deux sites de liaison distincts par ADN cruciforme : n sites « doubles-brin » et 1 site « cruciforme ».*

Les différents modèles proposés sont ajustés globalement sur l'ensemble des courbes cinétiques. La figure 40 illustre l'efficacité accrue de ce modèle adapté (appelé désormais modèle protéine / ADN) par rapport au modèle parallèle quant à l'ajustement des courbes d'association et de dissociation de E6 sur l'ADN cruciforme. La différence très nette visuellement se traduit également par un χ^2 sensiblement de meilleure qualité.

La figure 41 donne un exemple des variations de k_{a1} , k_{a2} et χ^2 en fonction des valeurs de n allant de 1 à 20. Il ressort une gamme de valeurs de n pour lesquelles les paramètres k_{a1} et k_{a2} sont relativement constants, et la valeur de χ^2 minimale. C'est donc dans cette gamme que sont choisies les valeurs de n utilisées dans les ajustements suivants.



Fig. 40 : illustration de l'ajustement par le modèle parallèle adapté. Conditions d'enregistrements : ADN cruciforme de Bianchi, EDTA 1 mM, Dextran 1 mg/ml. Les indices de réfraction, ainsi que les fenêtres d'ajustement sont identiques dans les deux cas. Valeur de n pour cet ajustement : 5.

C.III.4.e/ Effets des conditions expérimentales

La comparaison des paramètres cinétiques calculés en moyennant les constantes obtenues pour des valeurs de n comprises entre 4 et 10 (Fig. 42) donne plusieurs renseignements sur cette interaction. Par exemple, E6 se lie de la même manière à l'ADN de Bianchi qu'à l'ADN de Lilley, signe que cette interaction ne dépend pas de la séquence. De plus, non seulement cette interaction existe également en présence de magnésium –effet particulièrement visible sur le paramètre cinétique d'association–, mais en plus il existe un facteur 10 entre les constantes d'association à l'équilibre en magnésium ou en EDTA. Nous pouvons par ailleurs constater que les cinétiques du site 1 (site « double-brin ») ne changent guère avec les conditions.

C.III.4.f/ Comparaison des résultats de gels retard et de BIAcore

Cette comparaison est intéressante, les deux techniques étant ici complémentaires. Les expériences de gels retard indiquent une interaction E6 / ADN cruciforme, spécifique de la structure, et non de la séquence. Ce résultat est confirmé en BIAcore par les observations suivantes :

la séquence d'ADN utilisé (Lilley et Bianchi) ne change pas les valeurs des paramètres cinétiques au-delà de l'incertitude expérimentale, laissant supposer l'implication des groupements phosphate de l'ADN (chargés positivement) dans la liaison avec E6 (chargée négativement).



Fig. 41 : variations des paramètres ka1, ka2, c2 en fonction de n. (*cas de l'ADN cruciforme Bianchi, EDTA à 1 mM, Dextran à 1 mM, NaCl à 150 mM*).



Fig. 42 : valeurs des constantes cinétiques d'association, de dissociation, ainsi que la constante d'association à l'équilibre. La comparaison de l'ADN cruciforme de Lilley (barres bleues) et de l'ADN cruciforme de Bianchi (barres vertes) permet de vérifier le rôle de la séquence de l'ADN dans cette interaction. La comparaison des barres vertes (présence d'EDTA) et rouges (Mg) permet de juger du rôle de l'EDTA et du magnésium dans cette interaction.

- une concentration de NaCl à 500 mM inhibe tout signal (résultat non décrit), dénotant le caractère ionique de cette interaction. Cette inhibition à haut sel, compatible avec le point précédent, a d'ailleurs été utilisée pour mettre au point le tampon de régénération de la surface.
- l'interaction n'est pas sensible à l'orientation du cruciforme. Ce résultat (non décrit) est également en accord avec l'idée d'une interaction spécifique de la structure.

Par ailleurs, le BIAcore nous permet de compléter les résultats du gel retard : l'interaction est plus spécifique en présence de magnésium qu'en présence d'EDTA. Ce point laissé en suspend par la technique de gel retard est pourtant crucial pour transférer cette interaction dans un contexte *in vivo*. En effet, l'ion magnésium est présent dans le noyau, compartiment cellulaire contenant les jonctions de « Holliday ». De plus, ces expériences prouveraient :

- l'existence de deux sites de liaison pour E6 sur l'ADN cruciforme,
- la spécificité de l'interaction en un site unique du cruciforme, très certainement au croisement des deux brins du cruciforme,

A première vue, un point semble pourtant contradictoire entre ces deux techniques : un retard apparaît en présence d'un large excès d'ADN génomique, alors que des quantités comparables d'ADN double-brin injectées simultanément avec la protéine E6 inhibent quasiment tout signal détecté par BIAcore. Cette contradiction pourrait en fait provenir des temps de contact entre l'ADN et la protéine : une incubation d'1 heure serait suffisante pour atteindre l'équilibre, tandis qu'une durée de 120 secondes de la phase d'injection se révèlerait être trop courte pour atteindre un seuil détectable par BIAcore.

Enfin, les constantes de dissociation à l'équilibre déterminées par BIAcore (5 nM) et par gel retard (100 nM ; Cf. publication 5) semblent différentes. Cette apparente contradiction ne serait en fait que le reflet du pourcentage de protéines actives déterminées par la technologie SPR. Ce pourcentage, de l'ordre de 3 à 5% pour E6 interagissant avec l'ADN cruciforme est à rapprocher du facteur 1/20 constaté entre les deux constantes à l'équilibre.

CONCLUSIONS

ET

PERSPECTIVES

La production bactérienne de protéines d'intérêt fusionnées à des protéines porteuses hautement solubles est souvent utilisé comme moyen de sur-expression de protéines (Georgiou & Valax, 1996) (Baneyx, 1999). Parmi les protéines porteuses, la MBP a retenu un intérêt particulier, en tant que moyen efficace pour solubiliser des protéines récalcitrantes (Kapust & Waugh, 1999). Certains auteurs avaient pourtant mentionné le comportement anormal d'une préparation de protéine de fusion soluble : ces fusions présenteraient une activité biologique très réduite voire inexistante comparée à la protéine non fusionnée (Sachdev & Chirgwin, 1998 & 1999).

Dès le début de nos travaux, nous avions également constaté un comportement anormal de nos préparations solubles de MBP-E6 : la précipitation de la protéine E6 après séparation protéolytique de la protéine MBP. Dans le but de comprendre ce phénomène, nous avons utilisé la technique de diffusion de lumière, et mis en évidence la formation de corps d'inclusion solubles (publication 1). Bien que les phénomènes d'agrégation solubles de protéines avaient été étudiées dès les années 1950 (Doty & Edsall, 1951), ce problème n'avait apparemment pas été abordé dans le cas des protéines de fusion. Nous pensons en fait que ce manquement provient d'une confusion sur le critère de solubilité utilisé généralement pour vérifier le repliement d'une protéine non fusionnée : les protéines ne précipitant pas après centrifugation sur paillasse sont considérées comme correctement repliées. Ce principe de solubilité a été étendu aux protéines fusionnées (MBP, GST ou Thiorédoxine) : lorsque la fusion apparaît soluble avec ce test, la protéine d'intérêt est alors considérée comme correctement repliée. Cette confusion entre solubilité et repliement correct, se rencontre fréquemment dans des études utilisant les fusions MBP, certains auteurs concluant d'ailleurs au rôle possible de MBP comme chaperonine [(Kapust & Waugh, 1999), (Fox, 2001)]. Nos résultats tendent à contredire cette interprétation puisqu'ils montrent que les fusions MBP n'aident pas E6 à se replier, mais maintiennent simplement en solution des états mal repliés.

L'agrégation soluble comme l'un des états intermédiaires entre monodisperse et précipité

Avant de précipiter, une protéine monodisperse doit obligatoirement passer par un état transitoire intermédiaire : un agrégat soluble. Cet agrégat est une particule de masse variable, dont la distribution de taille peut être hétérogène. Sa taille continue d'augmenter tant qu'il existe un apport de matériel : sa durée de vie est d'autant plus courte que le phénomène d'agrégation qui génère cette particule est coopératif. Cependant, cette durée peut être prolongée si l'architecture intrinsèque

de l'agrégat limite sa progression : sa taille reste inférieure à la valeur seuil au delà de laquelle l'agrégat aurait précipité. C'est cet état que l'on retrouve dans le cas des corps d'inclusion solubles de MBP-E6 : ces fusions s'agrègent de par la précipitation de la protéine E6 mal repliée, rejetant ainsi vers l'extérieur de cet agglomérat les protéines MBP, elles, correctement repliées. Le nombre de monomères de MBP-E6 est donc limité par l'encombrement stérique des protéines MBP dans l'enveloppe externe de l'agrégat. Dans cet état, la transition entre les états monodisperse et précipité dépend d'un équilibre entre sédimentation et diffusion de particules (Fig. 43). Cet équilibre peut être déplacé vers la sédimentation en utilisant la force d'accélération engendrée par une ultracentrifugation.



Fig. 43 : l'agrégation est un état métastable entre les états précipité et monodisperse. Cet équilibre est régulé par la sédimentation et la diffusion.

• Les causes de l'agrégation ou de la précipitation dans le cas de E6

Ces analyses biophysiques nous ont permis de distinguer au moins deux processus distincts : (1) agrégation de la protéine mal repliée dès l'expression dans la bactérie, et (2) agrégation de la protéine repliée au cours de la purification.

(1) Le mauvais repliement de MBP-E6 lors de l'expression conduit à former des agrégats solubles. A ce propos, nous avons remarqué que l'expression des domaines N- et C-terminaux de E6 ne conduit pas à la formation de corps d'inclusion. Ces résultats laissent penser qu'aucun de ces deux domaines n'est isolément responsable de l'agrégation, tandis que leur combinaison dans la protéine entière provoque l'agrégation. Selon Garel JR (Garel, 1992), la cinétique de repliement de protéines multidomaines se décomposerait en une première phase de structuration rapide des domaines isolés, et une seconde phase, plus lente, d'association des domaines repliés menant à la structure native finale. Dans le cas de E6, il est possible que cette deuxième phase soit perturbée par des phénomènes d'association *intermoléculaire* de domaines prenant le pas sur leur association *intramoléculaire*.

(2) La protéine, même correctement repliée et séparée de MBP, est sensible à l'oxydation. Ce phénomène peut entraîner, pendant ou après purification, la précipitation de E6.

Ceci illustre l'importance d'effectuer des tests de détection des agrégats tant en sortie d'expression, qu'au cours de la purification ou après une période de stockage.

• La nature physique des corps d'inclusion solubles.

Nos résultats actuels favorisent l'idée d'une structure micellaire pour les corps d'inclusion solubles. Il sera intéressant de compléter l'étude de la structure de ces agrégats. Ces agrégats présentent l'intérêt de maintenir en solution une forme mal repliée de E6, la rendant accessible à des techniques impossibles à mettre en œuvre sur des corps d'inclusion solides, telles que spectroscopies DC, de fluorescence, de BIAcore, de RMN, de diffusion dynamique de lumière, d'IR, voire même de microscopie électronique.

De tels travaux nous permettraient également d'étudier une question qui s'est posée plusieurs fois au cours de notre travail : les agrégats de la protéine E6 mal repliée sont-ils capables de recruter la protéine correctement repliée ? Un tel comportement rappellerait celui des protéines prion (Schlumpberger, 2000).

Perspectives pour l'étude structure-fonction de E6.

L'obtention de la structure tri-dimensionelle du domaine C-terminal ouvre la voie à de passionnantes études structurales. De très nombreuses études de complexes peuvent être envisagées, impliquant d'un côté, le domaine C-terminal, et de l'autre, ses différents ligands cellulaires : ADN cruciforme, protéines cellulaires ciblées par E6, domaines ou peptides issus de ces protéines. Hormis ces ligands « naturels », le laboratoire a produit différents inhibiteurs recombinants potentiels de E6, sous la forme soit d'anticorps recombinants, soit de peptides. De plus, nous pensons que nos progrès dans la compréhension du repliement et de la solubilité du domaine C-terminal permettront à terme de produire des échantillons RMN du domaine N-terminal ainsi que de la protéine entière.

L'intérêt majeur des structures de complexes est de révéler le détail atomique des interfaces d'interaction. Les complexes de E6 ou de ses domaines avec leurs cibles « biologiques » (ADN,

protéines cellulaires entières ou fragmentées) ouvriront la voie à la modélisation d'inhibiteurs chimiques prévus pour s'immiscer dans ces interfaces (inhibiteurs compétititifs). Par ailleurs, l'analyse des interfaces de E6 avec les ligands peptidiques « non biologiques » (anticorps recombinants et peptides issus de criblage) permettra aussi de modéliser des inhibiteurs chimiques de E6 conçus par analogie aux ligands protéiques. La compréhension structurale du mode de liaison de scFvs et de peptides issus de criblage permettra aussi de raffiner l'affinité de ces ligands et d'améliorer leur pouvoir inhibiteur éventuel. En parallèle, des études BIAcore permettront d'établir les constantes cinétiques des complexes. Cette double analyse, cinétique et structurale, fournira les données nécessaires pour améliorer l'affinité des ligands peptides ou anticorps recombinant par mutagénèse dirigée. Une fois ces mutations effectuées, leurs caractéristiques cinétiques seront mesurées, et les mutants les plus intéressants seront à nouveau soumis à l'analyse RMN, complexés à E6 ou l'un de ses domaines.

CONCLUSION GÉNÉRALE.

Depuis plus de quinze ans, l'analyse structurale de E6 était empêchée par des problèmes techniques ayant trait à sa production par voie recombinante. Nous sommes parvenus à lever cette barrière en appliquant différentes méthodologies biophysiques pour analyser et améliorer l'agrégation des protéines de fusion. Ces méthodes appliquées à E6 nous ont permis de purifier la protéine recombinante et ses domaines sous forme repliée et stable, puis de résoudre la structure en solution du domaine à zinc C-terminal. Ces progrès ouvrent l'accès à tout un champ d'études structurales jusqu'ici interdit. Ces études sont essentielles car elles aboutiront à la compréhension atomique du mode d'action de E6. Elles permettront de plus la conception et l'amélioration de molécules capables d'inhiber l'action oncogénique de cette oncoprotéine majeure. Ces molécules ouvriront la voie à de nouvelles stratégies de thérapie du cancer du col de l'utérus.

ANNEXE

Table des déplacements chimiques de E6-C 4C/4S

seq.	aa	N	HN	Ηα	Ηβ2	Ηβ3	others		
-3	G								
-2	А								
- 1	М	121.05	8.54	4.49	2.04		Qδ 2.55		
87	S	118.42	8.29	4.67	3.83				
88	Y	122.11	8.34	4.89	3.25	3.02	QδQε7.04 6.77		
89	S	117.89	8.32	5.59	3.50	3.35			
90	L	122.11	8.48	4.77	1.60	1.52	Ηγ 1.22	Qõ 0.73 1.08	
91	Y	119.73	8.81	4.70	3.31	2.86	Qδ Qε 7.20 6.81		
92	G	111.30	7.61	4.20					
93	Т		7.52	4.36					
94	Т	119.21	7.12	4.19	4.07		Qγ 1.51		
95	L	123.42	7.58	3.73	1.31		Qγ Qδ 1.21 0.6	8	
96	Е	117.63	7.80	4.27	2.26		Qγ 2.51		
97	Q	118.15	7.43	4.05	2.21	2.12	Ηγ 2.54 2.37	Νε2 112.88	Ηε 6.84 7.53
98	Q	119.73	8.32	3.88	1.92	1.38	Ηγ 2.16 1.75	Νε2 111.57	Ηε 6.68 7.28
99	Y	114.73	7.80	4.37	3.01	2.59	Qδ, Qε 7.16 6.69	9	
100	Ν	119.73	7.98	4.38	3.13	2.65	Ηδ 7.53 6.79	Nô2 113.04	
101	K	117.63	7.65	4.65	1.70	1.38	Qγ 1.32	Hδ 1.59 1.51	Hε 2.93 2.75
102	Р			4.40	2.22	1.91	Ηγ 1.86 1.68	Hδ 3.62 3.57	
103	L	113.15	8.63	3.66	1.62	1.31	Ηγ 1.66	Hδ 0.67 0.60	
104	S	123.42	9.91	4.40					
105	D	121.84	7.81	4.58	2.72	2.46			
106	L	122.90	7.26	4.15	2.06	0.90	Ηγ 1.31	Hô1 0.28 0.24	
107	L	126.32	8.33	4.23	1.76	1.41	Ηγ 0.84	Qô1 -0.45 -0.54	
108		130.80	8.38	4.18	1.22		Qγ2 0.62	Ηγ1 1.10 0.53	Qô1 -0.17
109	R	125.79	7.78	4.85	1.45	1.27	Ηγ 1.18 0.83	Hð 3.21 2.96	Ηε 7.40
110	C	123.69	8.75	4.24	2.70	3.41			
111		132.38	9.52	4.10	2.07		Qγ2 1.01	Ηγ1 1.45 1.31	Q81 0.90
112	N	123.16	8.90	4.94	3.43	2.86	Ho2 8.33 7.09	Nð2 115.52	
113	C	118.15	8.02	4.88	3.24	2.54	11 0 15 0 00	N 0 1 1 0 07	
114	Q	116.05	7.80	4.10	2.26	2.16	Ηγ 2.45 2.36	Nε2 113.67	Ηε2 7.25 6.55
115	ĸ	123.92	8.37	4.56	1.99	1.00	Ηγ Ι.6Ι Ι.54		QE 3.05
115	P	107.00	0.40	4.42	2.13	1.96	Ηγ 2.19 1.79	H0 3.98 3.60	
117	L	127.38	9.42	4.54	1.96	1.42	Πγ 1.02	Q0 0.00 0.03	
811	2	122.37	9.76	4.//	3.99	1 0 0			
100		110.04	0 00	4.15	2.39	1.93	Ηγ 2.24 2.04	H0 3.97 3.85	
120		118.94	8.89	4.09	2.10	1.90	Ηγ 2.39 2.28		
121		122.03	0.04	4.00	2.54	1 60	$\alpha\gamma 2.41$	08176	
122	r. O	110 72	0.37	3.99	1.90	1.60	$\alpha\gamma$ 1.01 $\alpha\gamma$ 2.10	Q01.70 N₀2117.62	118 2.74 2.49 μορ g 10 g 07
123	P	101 05	0.40	3.00	2.43		$\alpha \gamma 2.10$	He 3 16	1182 0.10 0.97
124	н П	117.00	0.03	4.20	2.00	0 70		TRE J. TU ,	
125	Н	117.89	1.01	4.08	3.53	2.70	/ כנ.ס אח סרו		

```
ANNEXES
```

126	L	118.42	7.52	4.33	2.09	1.79	Ηγ 1.64	Qδ0.95	
127	D	122.37	9.04	4.30	2.71				
128	Κ	116.57	8.25	4.31	2.01		Qγ 1.66	Qδ1.51	
129	Κ	118.94	7.75	4.20	2.09	1.82	Ηγ 1.31 1.22	Hð 1.60 1.53	
130	Q	125.79	7.92	4.22			Ηγ 2.13 1.88	Nε2 110.51	Ηε2 6.99 6.63
131	R	125.53	8.80	4.13	2.41	2.20	Qγ 1.51	Ho 2.19 2.07	Ηε 3.79
132	F	118.68	8.56	4.92	3.28	2.64	Qδ Qε Ηζ 6.95	6.57 6.54	
133	Н	119.21	8.82	5.29	2.79				
134	Ν	130.27	9.18	4.32	1.81	-0.25	H ₀ 2 5.76 3.85	Nδ2 107.35	
135	Ι	128.43	8.54	4.22	1.70		Ηγ1 1.32 1.09	Qγ2 0.81	Qδ1 0.74
136	R	127.38	9.45	3.78	1.74	1.55	Qγ 1.97	Ho 3.22 3.12	
137	G	104.45	7.91	Hα 4.09	3.58				
138	R	120.53	7.37	4.55	1.78	1.66	Qγ 1.51	Qδ 3.11	
139	W	123.42	8.82	4.59	3.01	2.79	Hε 7.18 10.19	Hδ1 7.27	Nε1 123.27
							Ηζ 7.55 6.90	Hη2 6.78	
140	Т	115.78	9.06	5.88	4.32		Qγ 1.17		
141	G	109.72	7.96	$H\alpha 4.95$	3.63				
142	R	129.22	10.55	5.73	1.69		Qγ 1.73	Qδ3.10	Qη 7.56
143	С	129.48	9.54	4.42	3.90	2.59			
144	М	121.05	9.78	4.09	2.26	2.17	Ηγ 2.79 2.71	Qε 2.41	
145	S	118.15	8.64	4.22	4.02				
146	С	124.48	8.80	4.13	3.07	2.83			
147	S	116.57	8.00	4.29	3.97	3.85			
148	R	121.84	7.64	4.26	1.84	1.90	Ηγ 1.74 1.64		
149	S	116.83	7.95	4.32	3.96	3.85			
150	S	118.53	7.94	4.56	4.03				
151	R	124.48	8.33	4.38	1.86	1.74	Ηγ 1.77 1.60	Qδ3.15	
152	Т	115.78	8.14	4.28	4.15		Qγ 1.22		
153	R		8.34	4.29	2.11	1.98	Ηγ 2.42 2.24		
154	R	123.42	8.45	4.26	1.81	1.73	Qγ 1.59	Qδ3.16	
155	Е	122.90	8.60	4.27	2.04	1.92	Qγ 2.29		
156	Т	115.52	8.14	4.30	4.16		Qγ 1.16		
157	Q	124.48	8.39	4.30	2.08	1.96	Qγ 2.32	Hε2 6.88 7.66	Νε2 113.94
158	L	130.80	8.02	4.13	1.57		Ηγ 0.84 0.79		

BIBLIOGRAPHIE

BIBLIOGRAPHIE

BIBLIOGRAPHIE

- Ababou, A. & Bombarda, E. (2001) On the Involvement of Electron Transfer Reactions in the Fluorescence Decay Kinetics Heterogeneity of Proteins. *Protein Sci.* **10**;(10), 2102-2113.
- Adler, A., Greenfield, N. & Fasman, G. (1973) Circular Dichroism and Optical Rotatory Dispersion of Proteins and Polypeptides. *Methods Enzymol.* **27**;675-735.
- Agashe, V. & Hartl, F. (2000) Roles of Molecular Chaperones in Cytoplasmic Protein Folding. Semin. Cell. Dev. Biol. 11(1); 15-25.
- Altamirano, M., Blackburn, J., Aguayo, C. & Fersht, A. (2000) Directed Evolution of New Catalytic Activity Using the Alpha / Beta-Barrel Scaffold. *Nature* **403**;(6770), 617-622.
- Arrondo, J. & Goni, F. (1999) Structure and Dynamics of Membrane Proteins as Studied by Infrared Spectroscopy. *Prog. Biophys. Mol. Biol.* **72**;(4), 367-405.
- Baldwin, R. (1986) Temperature Dependence of the Hydrophobic Interaction in Protein Folding. *Proc. Natl. Acad. Sci. USA* 83;(21), 8069-8072.
- Baneyx, F. (1999) Recombinant Protein Expression in Escherichia Coli. *Curr. Opin. Biotechnol.* **10**;(5), 411-421.
- Bartels, C., Xia, T., Billeter, M., Güntert, P. & Wütrich, K. (1995) The Programm XEASY for Computer-Supported NMR Spectral Analysis of Biological Macromolecules. *J. Biomol. NMR* **5**;1-10.
- Bauer, R., Carrotta, R., Rischel, C. & Ogendal, L. (2000) Characterization and Isolation of Intermediates in Beta-Lactoglobulin Heat Aggregation at High Ph. *Biophys. J.* **79**;(2), 1030-1038.
- Bax, A. & Davis, D. (1985) Mlev-17-Based Two Dimensional Homonuclear Magnetization Transfer Spectroscopy. J. Magn. Reson. 61;306-320.
- Beerheide, W., Bernard, H. U., Tan, Y. J., Ganesan, A., Rice, W. G. & Ting, A. E. (1999) Potential Drugs against Cervical Cancer: Zinc-Ejecting Inhibitors of the Human Papillomavirus Type 16 E6 Oncoprotein. *J. Natl. Cancer Inst.* **91**;(14), 1211-1220.
- Bell, S., Hansen, S. & Buchner, J. (2002) Refolding and Structural Characterization of the Human P53 Tumor Suppressor Protein. *Biophys. Chem.* **96**;(2-3), 243-257.
- Betton, J., Sassoon, N., Hofnung, M. & Laurent, M. (1998) Degradation Versus Aggregation of Misfolded Maltose-Binding Protein in the Periplasm of Escherichia Coli. J. Biol. Chem. 273;(15), 8897-8902.
- Bianchi, M., Beltrame, M. & Paonessa, G. (1989) Specific Recognition of Cruciform DNA by Nuclear Protein HMGL. *Science* 243;(4894 Pt 1), 1056-1059.

Brünger, A. XPLOR: A System for Crystallographiy and NMR. (1992).

- Butz, K., Denk, C., Ullmann, A., Scheffner, M. & Hoppe-Seyler, F. (2000) Induction of Apoptosis in Human Papillomaviruspositive Cancer Cells by Peptide Aptamers Targeting the Viral E6 Oncoprotein. *Proc. Natl. Acad. Sci. U S A* **97**;(12), 6693-6697.
- Cantor, C. & Schimmel, P. Biophysical Chemistry. Part II: Techniques for the Study of Biological Structure and Function, (W.H. Freeman & Company, San Francisco, 1980).
- Carrio, M. M. & Villaverde, A. (2001) Protein Aggregation as Bacterial Inclusion Bodies Is Reversible. *FEBS Letters* **489**;(1), 29-33.
- Carrio, M. M. & Villaverde, A. (2002) Construction and Deconstruction of Bacterial Inclusion Bodies. J. Biotechnol. 96;(1), 3-12.
- Chen, J., Reid, C. E., Band, V. & Androphy, E. J. (1995) Interaction of Papillomavirus E6 Proteins with a Putative Calcium Binding Protein. *Science* **269**;529-532.
- Chen, J. J., Hong, Y., Rustamzadeh, E., Baleja, J. D. & Androphy, E. J. (1998) Identification of an Alpha Helical Motif Sufficient for Association with Papillomavirus E6. *J. Biol. Chem.* **273**;(22), 13537-13544.
- Chothia, C. & Lesk, A. (1986) The Relation between the Divergence of Sequence and Structure in Proteins. *EMBO J.* **5**;(4), 823-826.
- Christensen, L. (1997) Theoretical Analysis of Protein Concentration Determination Using Biosensor Technology under Conditions of Partial Mass Transport Limitation. *Anal. Biochem.* 249;(2), 153-164.
- Cole, S. & Danos, O. (1987) Nucleotide Sequence and Comparative Analysis of the Human Papillomavirus Type 18 Genome. Phylogeny of Papillomaviruses and Repeated Structure of the E6 and E7 Gene Products. *J. Mol. Biol.* **193**;(4), 599-608.
- Corchero, J., Cubarsi, R., Enfors, S. & Villaverde, A. (1997) Limited in Vivo Proteolysis of Aggregated Proteins. *Biochem. Biophys. Res. Com.* 237;(2), 325-330.
- Costantino, H., Langer, R. & Klibanov, A. (1994) Solid-Phase Aggregation of Proteins under Pharmaceutically Relevant Conditions. J. Pharm. Sci. 83;(12), 1662-1669.
- Cowburn, D. (1997) Peptide Recognition by PTB and PDZ Domains. *Curr. Opin. Struct. Biol.* 7;(6), 835-838.
- Creighton, T. (1993) Stability of the Folded Conformation. In T. Creighton (ed.) *Proteins: Structure and Molecular Properties*. W.H. Freeman and Compagny, New York, pp. 287-309.
- Daniels, P., Sanders, C., Coulson, P. & Maitland, N. (1997) Molecular Analysis of the Interaction between HPV Type 16 E6 and Human E6-Associated Protein. *FEBS Lett.* **416**;(1), 6-10.
- Dayie, K. T., Wagner, G. & Lefevre, J. F. (1996) Theory and Practice of Nuclear Spin Relaxation in Proteins. *Annu. Rev. Phys. Chem.* **47**;243-282.

Debye, P. (1915) Zertreuung Von Röntgenstrahlen. Ann. Physik 46;809-823.

- Delaglio, F., Grzesiek, S., Vuister, G., Zhu, G., Pfeifer, J. & Bax, A. (1995) NMRPIPE: A Multidimensional Spectral Processing System Based on Unix Pipes. J. Biomol. NMR 6;(3), 277-293.
- Dell, G. & Gaston, K. (2001) Human Papillomaviruses and Their Role in Cervical Cancer. *Cell. Mol. Life Sci.* **58**;(12-13), 1923-1942.
- Dey, A., Atcha, I. A. & Bagchi, S. (1997) HPV16 E6 Oncoprotein Stimulates the Transforming Growth Factor-Beta 1 Promoter in Fibroblasts through a Specific GC-Rich Sequence. *Virology* **228**;(2), 190-199.
- Dobson, C. (1999) Protein Misfolding, Evolution and Disease. *Trends Biochem. Sci.* 24;(9), 329-332.
- Doolittle, R. (1995) The Multiplicity of Domains in Proteins. Annu. Rev. Biochem. 64;287-314.
- Doty, P. & Edsall, J. (1951) Light Scattering in Protein Solutions. Adv. Protein Chem. 6;35-121.
- Du, M., Fan, X., Hong, E. & Chen, J. (2002) Interaction of Oncogenic Papillomavirus E6 Proteins with Fibulin-1. *Biochem. Biophys. Res. Com.* **296**;(4), 962-969.
- Duckett, D., Murchie, A., Diekmann, S., von Kitzing, E., Kemper, B. & Lilley, D. (1988) The Structure of the Holliday Junction, and Its Resolution. *Cell* **55**;(1), 79-89.
- Effink, M. (1998) The Use of Fluorescence Methods to Monitor Unfolding Transitions in Proteins. *Biochemistry (Mosc.)* **63**;(3), 276-284.
- Enfors, S. (1992) Control of *in vivo* Proteolysis in the Production of Recombinant Proteins. *Trends Biotechnol* **10**;(9), 310-315.
- Filippova, M., Song, H., Connolly, J., Dermody, T. & Duerksen-Hughes, P. (2002) The Human Papillomavirus 16 E6 Protein Binds to Tumor Necrosis Factor (TNF) R1 and Protects Cells from TNF-Induced Apoptosis. *J. Biol. Chem.* **277**;(24), 21730-21739.
- Fox, J., Kapust, R. & Waugh, D. (2001) Single Amino Acid Substitutions on the Surface of Escherichia Coli Maltose-Binding Protein Can Have a Profound Impact on the Solubility of Fusion Proteins. *Protein Sci.* **10**;(3), 622-630.
- Gao, Q., Kumar, A., Srinivasan, S., Singh, L., Mukai, H., Ono, Y., Wazer, D. E. & Band, V. (2000) PKN Binds and Phosphorylates Human Papillomavirus E6 Oncoprotein. *J. Biol. Chem.* **275**;14824-14830.
- Gao, Q., Srinivasan, S., Boyer, S. N., Wazer, D. E. & Band, V. (1999) The E6 Oncoproteins of High-Risk Papillomaviruses Bind to a Novel Putative Gap Protein, E6TP1, and Target It for Degradation. *Mol. Cell. Biol.* **19**;733-744.

- Gardiol, D., Kuhne, C., Glaunsinger, B., Lee, S. S., Javier, R. & Banks, L. (1999) Oncogenic Human Papillomavirus E6 Proteins Target the Discs Large Tumour Suppressor for Proteasome-Mediated Degradation. *Oncogene* 18;(40), 5487-5496.
- Garel, J. (1992) Folding of Large Proteins: Multidomain and Multisubunit Proteins. In T. Creighton (ed.) *Protein Folding*. W. H. Freeman & Compagny, New York, pp. 405-454.
- Georgiou, G. & Valax, P. (1996) Expression of Correctly Folded Proteins in Escherichia Coli. *Curr. Opin. Biotechnol.* 7;(2), 190-197.
- Gerard, D. (1975) Fluorescence Intrinsèque Des Protéines. Mécanismes Et Applications à L'étude Conformationnelle De Quelques Protéines., Université Louis Pasteur.
- Gewin, L. & Galloway, D. A. (2001) E Box-Dependent Activation of Telomerase by Human Papillomavirus Type 16 E6 Does Not Require Induction of c-Myc. J. Virol. **75**;(15), 7198-7201.
- Gill, S. & von Hippel, P. (1989) Calculation of Protein Extinction Coefficients from Amino Acid Sequence Data. *Anal Biochem.* **182**;(2), 319-326.
- Glaunsinger, B., Lee, S., Thomas, M., Banks, L. & Javier, R. (2000) Interactions of the PDZ-Protein MAGI-1 with Adenovirus E4-Orf1 and High-Risk Papillomavirus E6 Oncoproteins. *Oncogene* **19**;(46), 5270-5280.
- Greenfield, N. (1996) Methods to Estimate the Conformation of Proteins and Polypeptides from Circular Dichroism Data. *Anal. Biochem.* **235**;(1), 1-10.
- Gross-Mesilaty, S., Reinstein, E., Bercovich, B., Tobias, K., Schwartz, A., Kahana, C. & Ciechanover, A. (1998) Basal and Human Papillomavirus E6 Oncoprotein-Induced Degradation of Myc Proteins by the Ubiquitin Pathway. *Proc. Natl. Acad. Sci. U S A* **95**;(14), 8058-8063.
- Grossman, S. & Laimins, L. (1989) E6 Protein of Human Papillomavirus Type 18 Binds Zinc. Oncogene 4;(9), 1089-1093.
- Hawley-Nelson, P., Vousden, K. H., Hubbert, N. L., Lowy, D. R. & Schiller, J. T. (1989) HPV16 E6 and E7 Proteins Cooperate to Immortalize Human Foreskin Keratinocytes. *EMBO J.* **8**;(3905-3910.
- Iftner, T., Elbel, M., Schopp, B., Hiller, T., Loizou, J., Caldecott, K. & Stubenrauch, F. (2002) Interference of Papillomavirus E6 Protein with Single-Strand Break Repair by Interaction with XRCC1. *EMBO J.* **21**;(17), 4741-4748.
- Janin, J. & Chothia, C. (1985) Domains in Proteins: Definitions, Location, and Structural Principles. *Methods Enzymol.* **115**;420-430.
- Jiang, M. & Milner, J. (2002) Selective Silencing of Viral Gene Expression in HPV-Positive Human Cervical Carcinoma Cells Treated with Sirna, a Primer of RNA Interference. *Oncogene* **21**;(39), 6041-6048.

- Kampranis, S., Gormley, N., Tranter, R., Orphanides, G. & Maxwell, A. (1999) Probing the Binding of Coumarins and Cyclothialidines to DNA Gyrase. *Biochemistry* **38**;(7), 1967-1976.
- Kapust, R. & Waugh, D. (1999) Escherichia Coli Maltose-Binding Protein Is Uncommonly Effective at Promoting the Solubility of Polypeptides to Which It Is Fused. *Protein Sci.* **8**;(8), 1668-1674.
- Karlsson, R. & Falt, A. (1997) Experimental Design for Kinetic Analysis of Protein-Protein Interactions with Surface Plasmon Resonance Biosensors. J. Immunol. Methods **200**;(1-2), 121-133.
- Kay, L., Torchia, D. & Bax, A. (1989) Backbone Dynamics of Proteins as Studied by 15N Inverse Detected Heteronuclear NMR Spectroscopy: Application to Staphylococcal Nuclease. *Biochemistry* **28**;(23), 8972-8979.
- Kendrew, J. C. (1960) Structure of Myoglobin. Nature 185;422-427.
- Kiyono, T., Hiraiwa, A., Fujita, M., Hayashi, Y., Akiyama, T. & Ishibashi, M. (1997) Binding of High-Risk Human Papillomavirus E6 Oncoproteins to the Human Homologue of the Drosophila Discs Large Tumor Suppressor Protein. *Proc. Natl. Acad. Sci. U S A* **94**;(21), 11612-11616.
- Kopito, R. R. (2000) Aggresomes, Inclusion Bodies and Protein Aggregation. *Trends Cell. Biol.* **10**;(12), 524-530.
- Koradi, R., Billeter, M. & Wuthrich, K. (1996) MOLMOL: A Program for Display and Analysis of Macromolecular Structures. J. Mol. Graph. 14;(1), 51-55, 29-32.
- Krimm, S. & Bandekar, J. (1986) Vibrational Spectroscopy and Conformation of Peptides, Polypeptides and Proteins. *Adv. Protein Chem.* **38**;181-364.
- Kukimoto, I., Aihara, S., Yoshiike, K. & Kanda, T. (1998) Human Papillomavirus E6 Protein Binds to the C-Terminal Region of Human Minichromosome Maintenance 7 Protein. *Biochem. Biophys. Res. Com.* **249**;(258-262.
- Kumar, A., Ernst, R. & Wuthrich, K. (1980) A 2D nOe Experiment for the Elucidation of Complete Proton-Proton Cross-Relaxation Networks in Biological Macromolecules. *Biochem. Biophys. Res. Com.* **95**;(1-6.
- Kumar, A., Zhao, Y., Meng, G., Zeng, M., Srinivasan, S., Delmolino, L., Gao, Q., Dimri, G., Weber, G., Wazer, D., Band, H. & Band, V. (2002) Human Papillomavirus Oncoprotein E6 Inactivates the Transcriptional Coactivator Human ADA3. *Mol. Cell. Biol.* **22**;(16), 5801-5812.
- Kumar, S., Talis, A. L. & Howley, P. M. (1999) Identification of HHR23A as a Substrate for E6-Associated Protein- Mediated Ubiquitination. *J. Biol. Chem.* **274**;(26), 18785-18792.
- Kunitani, M., Wolfe, S., Rana, S., Apicella, C., Levi, V. & Dollinger, G. (1997) Classical Light Scattering Quantitation of Protein Aggregates: Off-Line Spectroscopy Versus HPLC Detection. J. Pharm. Biomed. Anal. 16;(4), 573-586.

- Lechner, M. & Laimins, L. (1994) Inhibition of p53 DNA Binding by Human Papillomavirus E6 Proteins. J. Virol. 68;(7), 4262-4273.
- Lee, S. S., Glaunsinger, B., Mantovani, F., Banks, L. & Javier, R. T. (2000) Multi-PDZ Domain Protein MUPP1 Is a Cellular Target for Both Adenovirus E4-ORF1 and High-Risk Papillomavirus Type 18 E6 Oncoproteins. J. Virol. 74;(20), 9680-9693.
- Li, J., Uversky, V. N. & Fink, A. L. (2001) Effect of Familial Parkinson's Disease Point Mutations A30P and A53T on the Structural Properties, Aggregation, and Fibrillation of Human Alpha-Synuclein. *Biochemistry* **40**;(38), 11604-11613.
- Li, S., Labrecque, S., Gauzzi, M. C., Cuddihy, A. R., Wong, A. H., Pellegrini, S., Matlashewski, G. J. & Koromilas, A. E. (1999) The Human Papilloma Virus (HPV)-18 E6 Oncoprotein Physically Associates with Tyk2 and Impairs Jak-Stat Activation by Interferon-Alpha. *Oncogene* **18**;(42), 5727-5737.
- Lipari, F., McGibbon, G. A., Wardrop, E. & Cordingley, M. G. (2001) Purification and Biophysical Characterization of a Minimal Functional Domain and of an N-Terminal Zn2+-Binding Fragment from the Human Papillomavirus Type 16 E6 Protein. *Biochemistry* **40**;(5), 1196-1204.
- London, J., Skrzynia, C. & Goldberg, M. E. (1974) Renaturation of *Escherichia Coli* Tryptophanase after Exposure to 8 M Urea. Evidence for the Existence of Nucleation Centers. *Eur. J. Biochem.* 47;(2), 409-415.
- Luo, S., Huang, C., McClelland, J. & Graves, D. (1994) A Study of Protein Secondary Structure by Fourier Transform Infrared/Photoacoustic Spectroscopy and Its Application for Recombinant Proteins. *Anal. Biochem.* **216**;(1), 67-76.
- Manavalan, P. & Johnson, W., Jr. (1987) Variable Selection Method Improves the Prediction of Protein Secondary Structure from Circular Dichroism Spectra. *Anal Biochem* **167**;(1), 76-85.
- Marston, F. (1986) The Purification of Eukaryotic Polypeptides Synthesized in Escherichia Coli. *Biochem J.* **240**;(1), 1-12.
- Maruyama, T., Katoh, S., Nakajima, M. & Nabetani, H. (2001) Mechanism of Bovine Serum Albumin Aggregation During Ultrafiltration. *Biotechnol. Bioeng.* **75**;(2), 233-238.
- Mathevet, P. (2001) [Viruses and Cervical Cancers]. Rev. Prat. 51;(13), 1413-1416.
- Merutka, G., Dyson, H. & Wright, P. (1995) 'Random Coil' 1H Chemical Shifts Obtained as a Function of Temperature and Trifluoroethanol Concentration for the Peptide Series Ggxgg. J. *Biomol. NMR* **5**;(1), 14-24.
- Morosow, A., Phelps, W. C. & Raychaudhuri, P. (1994) Activation of c-Fos Gene by the HPV16 Oncoproteine Depends Upon the Camp-Response Element at -60. *J. Biol. Chem.* **269**;18434-18440.
- Morton, T. & Myszka, D. (1998) Kinetic Analysis of Macromolecular Interactions Using Surface Plasmon Resonance Biosensors. *Methods Enzymol.* **295**;268-294.

- Münger, K., Phelps, W. C., Bubb, V., Howley, P. M. & Schlegel, R. (1989) The E6 and E7 Genes of the Human Papillomavirus Type 16 Together Are Necessary and Sufficient for Transformation of Primary Human Keratinocytes. *Journal of Virology* **63**;4417-4421.
- Myszka, D. (1999a) Improving Biosensor Analysis. J. Mol. Recognit. 12;(5), 279-284.
- Myszka, D. (1999b) Survey of the 1998 Optical Biosensor Literature. J. Mol. Recognit. 12;(6), 390-408.
- Nakagawa, S. & Huibregtse, J. M. (2000) Human Scribble (Vartul) Is Targeted for Ubiquitin-Mediated Degradation by the High-Risk Papillomavirus E6 Proteins and the E6AP Ubiquitin-Protein Ligase. *Mol. Cell. Biol.* **20**;(21), 8244-8253.
- Nandi, P. K. (1996) Protein Conformation and Disease. Vet. Res. 27;(4-5), 373-382.
- Nilges, M., Gronenborn, A., Brunger, A. & Clore, G. (1988) Determination of Three-Dimensional Structures of Proteins by Simulated Annealing with Interproton Distance Restraints. Application to Crambin, Potato Carboxypeptidase Inhibitor and Barley Serine Proteinase Inhibitor 2. *Protein Eng.* **2**;(1), 27-38.
- Nilges, M., Kuszewski, J. & Brünger, A. Computational Aspects of the Study of Biological Macromolecules by NMR, (Plenum Press, New York, 1991).
- Nojima, H., Ikai, A., Oshima, T. & Noda, H. (1977) Reversible Thermal Unfolding of Thermostable Phosphoglycerate Kinase. Thermostability Associated with Mean Zero Enthalpy Change. J. Mol. Biol. 116;(3), 429-442.
- Oh, S. T., Kyo, S. & Laimins, L. A. (2001) Telomerase Activation by Human Papillomavirus Type 16 E6 Protein: Induction of Human Telomerase Reverse Transcriptase Expression through Myc and Gc-Rich Sp1 Binding Sites. *J. Virol.* **75**;(12), 5559-5566.
- Orth, G. (1997) Virus Et Cancer Du Col De L'utérus. Fondamental, Magazine trimestriel d'Information de l'Association pour la Recherche sur le Cancer 74;4-6.
- Pace, C., Shirley, B. & JA., T. (1994) Measuring the Conformational Stability of a Protein. In T. Creighton (ed.) *Protein Structure: A Practical Approach*. IRL Press, Oxford, Vol. 174, pp. 311-329.
- Patel, D., Huang, S. M., Baglia, L. A. & McCance, D. J. (1999) The E6 Protein of Human Papillomavirus Type 16 Binds to and Inhibits Co-Activation by CBP and p300. *EMBO J.* **18**;5061-5072.
- Peabody, D. & Al-Bitar, L. (2001) Isolation of Viral Coat Protein Mutants with Altered Assembly and Aggregation Properties. *Nucleic Acids Res.* **29**;(22), E113.
- Piotto, M., Saudek, V. & Sklenar, V. (1992) Gradient-Tailored Excitation for Single-Quantum NMR Spectroscopy of Aqueous Solutions. *J. Biomol. NMR* **2**;(6), 661-665.
- Privalov, P. (1979) Stability of Proteins: Small Globular Proteins. Adv. Protein Chem. 33;167-241.

- Privalov, P. (1992) Physical Basis of the Stability of the Folded Conformations of Proteins. In T. Creighton (ed.) *Protein Folding*. W. H. Freemann and Compagny, New York, pp. 83-126.
- Privalov, P. (1996) Intermediate States in Protein Folding. J. Mol. Biol. 258;(5), 707-725.
- Provencher, S. & Glockner, J. (1981) Estimation of Globular Protein Secondary Structure from Circular Dichroism. *Biochemistry* **20**;(1), 33-37.
- Przybycien, T., Dunn, J., Valax, P. & Georgiou, G. (1994) Secondary Structure Characterization of Beta-Lactamase Inclusion Bodies. *Protein Eng.* 7;(1), 131-136.
- Ptitsyn, O. (1992) The Molten Globule. In T. Creighton (ed.) *Protein Folding*. W. H. Freeman & Company, New York, pp. 243-300.
- Quinn, J. & O'Kennedy, R. (2001) Biosensor-Based Estimation of Kinetic and Equilibrium Constants. *Anal. Biochem.* **290**;(1), 36-46.
- Rajan, R. S., Illing, M. E., Bence, N. F. & Kopito, R. R. (2001) Specificity in Intracellular Protein Aggregation and Inclusion Body Formation. *Proc. Natl. Acad. Sci. U S A* **98**;(23), 13060-13065.
- Ristriani, T. (2001) Production of Recombinant E6 Protein of HPV Type 16 Responsible for Cervical Cancer. Identification and Characterization of Specific Interaction between E6 and Four-Way DNA Junctions., Université Louis Pasteur.
- Ronco, L. V., Karpova, A. Y., Vidal, M. & Howley, P. M. (1998) Human Papillomavirus 16 E6 Oncoprotein Binds to Interferon Regulatory Factor-3 and Inhibits Its Transcriptional Activity. *Genes Dev.* **12**;2061-2072.
- Sachdev, D. & Chirgwin, J. (1998) Solubility of Proteins Isolated from Inclusion Bodies Is Enhanced by Fusion to Maltose-Binding Protein or Thioredoxin. *Protein Expr. Purif.* **12**;(1), 122-132.
- Sachdev, D. & Chirgwin, J. (1999) Properties of Soluble Fusions between Mammalian Aspartic Proteinases and Bacterial Maltose-Binding Protein. J. Protein Chem. 18;(1), 127-136.
- Scheffner, M., Huibregtse, J., Vierstra, R. & Howley, P. (1993) The HPV-16 E6 and E6-AP Complex Functions as a Ubiquitin-Protein Ligase in the Ubiquitination of p53. *Cell* **75**;95-505.
- Scheffner, M., Werness, B., Huibregtse, J., Levine, A. & Howley, P. (1990) The E6 Oncoprotein Encoded by Human Papillomavirus Types 16 and 18 Promotes the Degradation of p53. *Cell* 63;1129-1136.
- Schlumpberger, M., Wille, H., Baldwin, M., Butler, D., Herskowitz, I. & Prusiner, S. (2000) The Prion Domain of Yeast Ure2p Induces Autocatalytic Formation of Amyloid Fibers by a Recombinant Fusion Protein. *Protein Sci.* **9**;(3), 440-451.
- Schmidt, F. (1994) Measuring the Conformational Stability of a Protein. In T. Creighton (ed.) *Protein Structure: A Practical Approach*. IRL Press, Oxford, Vol. 174, p. 355.

- Schwarz, E., Freese, U., Gissmann, L., Mayer, W., Roggenbuck, B., Stremlau, A. & zur Hausen, H. (1985) Structure and Transcription of Human Papillomavirus Sequences in Cervical Carcinoma Cells. *Nature* **314**;111-114.
- Sedman, S. A., Barbosa, M. S., Vaa, W. C., Hubbert, N. L., Haas, J. A., Lowy, D. R. & Schiller, J. T. (1991) The Full-Length E6 Protein of Human Papillomavirus Type 16 Has Transforming and Transactivating Activities and Cooperates with E7 to Immortalize Keratinocytes in Culture. *J. Virol.* **65**;4860-4866.
- Shaka, A., Barker, P. & Freeman, R. (1985) Computer Optimized Decoupling Scheme for Wide-Band Applications and Low Level Operation. J. Magn. Reson. 64;547-552.
- Soto, C., Castano, E. M., Frangione, B. & Inestrosa, N. C. (1995) The Alpha-Helical to Beta-Strand Transition in the Amino-Terminal Fragment of the Amyloid Beta-Peptide Modulates Amyloid Formation. *J. Biol. Chem.* **270**;(7), 3063-3067.
- Speed, M. A., Wang, D. I. & King, J. (1996) Specific Aggregation of Partially Folded Polypeptide Chains: The Molecular Basis of Inclusion Body Composition. *Nat. Biotechnol.* **14**;(10), 1283-1287.
- Srivenugopal, K. & Ali-Osman, F. (2002) The DNA Repair Protein, O(6)-Methylguanine-DNA Methyltransferase Is a Proteolytic Target for the E6 Human Papillomavirus Oncoprotein. *Oncogene* **21**;(38), 5940-5945.
- Stenberg, E., Persson, B., Roos, H. & Urbaniczky, C. (1991) Quantitative Determination of Surface Concentration of Protein with Surface Plasmon Resonance Using Radiolabeled Proteins. *Journal of Colloid and Interface Science* **143**;(2), 513-526.
- Suttnar, J., Dyr, J. E., Hamsikova, E., Novak, J. & Vonka, V. (1994) Procedure for Refolding and Purification of Recombinant Proteins from Escherichia Coli Inclusion Bodies Using a Strong Anion Exchanger. J. Chromatogr. B. Biomed. Appl. 656;(1), 123-126.
- Tanford, W. (1967) Light Scattering. Physical Chemistry of Macromolecules., pp. 275-316.
- Thomas, J. G. & Baneyx, F. (1996) Protein Misfolding and Inclusion Body Formation in Recombinant Escherichia Coli Cells Overexpressing Heat-Shock Proteins. J. Biol. Chem. **271**;(19), 11141-11147.
- Thomas, M. & Banks, L. (1998) Inhibition of Bak-Induced Apoptosis by HPV-18 E6. *Oncogene* 17;2943-2954.
- Thomas, M., Laura, R., Hepner, K., Guccione, E., Sawyers, C., Lasky, L. & Banks, L. (2002) Oncogenic Human Papillomavirus E6 Proteins Target the MAGI-2 and MAGI-3 Proteins for Degradation. *Oncogene* **21**;(33), 5088-5096.
- Tinoco, I., Sauer, K., Wang, J. & Puglisi, J. (2002) Molecular Distribution and Statistical Thermodynamics. In J. Challice (ed.) *Physical Chemistry: Principles and Applications in Biological Sciences*. Prentice Hall, New Jersey, pp. 614-665.
- Tong, X., Boll, W., Kirchhausen, T. & Howley, P. M. (1998) Interaction of the Bovine Papillomavirus E6 Protein with the Clathrin Adaptor Complex AP-1. J. Virol. 72;(1), 476-482.
- Tong, X. & Howley, P. M. (1997) The Bovine Papillomavirus E6 Oncoprotein Interacts with Paxillin and Disrupts the Actin Cytoskeleton. *Proc. Natl. Acad. Sci. USA* **94**;(9), 4412-4417.
- Tran, P. B. & Miller, R. J. (1999) Aggregates in Neurodegenerative Disease: Crowds and Power? *Trends Neurosci.* 22;(5), 194-197.
- Tummino, P., Scholten, J., Harvey, P., Holler, T., Maloney, L., Gogliotti, R., Domagala, J. & Hupe, D. (1996) The *in vitro* Ejection of zinc from Human Immunodeficiency Virus (HIV) Type 1 Nucleocapsid Protein by Disulfide Benzamides with Cellular Anti- Hiv Activity. *Proc. Natl. Acad. Sci. USA* 93;(3), 969-973.
- Vande Pol, S. B., Brown, M. C. & Turner, C. E. (1998) Association of Bovine Papillomavirus Type 1 E6 Oncoprotein with the Focal Adhesion Protein Paxillin through a Conserved Protein Interaction Motif. *Oncogene* **16**;(1), 43-52.
- Veldman, T., Horikawa, I., Barrett, J. C. & Schlegel, R. (2001) Transcriptional Activation of the Telomerase Htert Gene by Human Papillomavirus Type 16 E6 Oncoprotein. J. Virol. 75;(9), 4467-4472.
- Voet, D., Gratzer, W., Cox, R. & Doty, P. (1963) Absorption Spectra of Nucleotides, Polynucleotides and Nucleic Acids in the Far Ultraviolet. *Biopolymers* 1;(193-208.
- Vuister, G. & Bax, A. (1993) Quantitative J Correlation: A New Approach for Measuring Homonuclear Three-Bond J(HN-Hα) Coupling Constants in 15N-Enriched Proteins. J. Am. Chem. Soc. 115;7772-7777.

Wagner, G. (1997) An Account of NMR in Structural Biology. Nat. Struct. Biol. 4 Suppl;841-844.

- Walboomers, J. M., Jacobs, M. V., Manos, M. M., Bosch, F. X., Kummer, J. A., Shah, K. V., Snijders, P. J., J., P., Meijer, C. J. & Munoz, N. (1999) Human Papillomavirus Is a Necessary Cause of Invasive Cervical Cancer Worldwide. *J. Pathol.* **189**;(1), 12-19.
- Wen, J., Arakawa, T. & Philo, J. (1996) Size-Exclusion Chromatography with on-Line Light-Scattering, Absorbance, and Refractive Index Detectors for Studying Proteins and Their Interactions. *Anal. Biochem.* **240**;(2), 155-166.
- Wickner, S., Maurizi, M. & Gottesman, S. (1999) Posttranslational Quality Control: Folding, Refolding, and Degrading Proteins. *Science* **286**;(5446), 1888-1893.
- Wider, G. & Wuthrich, K. (1999) NMR Spectroscopy of Large Molecules and Multimolecular Assemblies in Solution. *Curr. Opin. Struct. Biol.* **9**;(5), 594-601.
- Williams, R. (1983) Estimation of Protein Secondary Structure from the Laser Raman Amide I Spectrum. J. Mol. Biol. 166;(4), 581-603.

- Wishart, D., Sykes, B. & Richards, F. (1992) The Chemical Shift Index: A Fast and Simple Method for the Assignment of Protein Secondary Structure through NMR Spectroscopy. *Biochemistry* **31**;(6), 1647-1651.
- Wishart, D., Sykes, B. D. & Richards, F. (1991) Relationship between Nuclear Magnetic Resonance Chemical Shift and Protein Secondary Structure. J. Mol. Biol. 222;(2), 311-333.
- Worn, A. & Pluckthun, A. (1998) An Intrinsically Stable Antibody Scfv Fragment Can Tolerate the Loss of Both Disulfide Bonds and Fold Correctly. *FEBS Letters* **427**;(3), 357-361.
- Wüthrich, K. NMR of Proteins & Nucleic Acids, (Wiley-Intersciences Pulbication, New York, 1986).
- Wyatt, P. (1993) Light Scattering and the Absolute Characterization of Macromolecules. *Rev. Anal. Chim. Acta* 272;1-40.
- Yon, J. (1996) The Specificity of Protein Aggregation [New; Comment]. Nat. Biotechnol. 14;(10), 1231.
- Zeder-Lutz, G., Benito, A. & Van Regenmortel, M. (1999) Active Concentration Measurements of Recombinant Biomolecules Using Biosensor Technology. J. Mol. Recognit. 12;(5), 300-309.
- Zimmermann, H., Degenkolbe, R., Bernard, H. U. & O'Connor, M. J. (1999) The Human Papillomavirus Type 16 E6 Oncoprotein Can Down-Regulate p53 Activity by Targeting the Transcriptional Coactivator Cbp/P300. *J. Virol.* **73**;6209-6219.
- zur Hausen, H. (1991) Human Papillomaviruses in the Pathogenesis of Anogenital Cancer. *Virology* **184**;9-13.
- zur Hausen, H. (1996) Papillomavirus Infections--a Major Cause of Human Cancers. *Biochim. Biophys. Acta.* **1288**;(2), F55-78.
- zur Hausen, H. (1999) Papillomaviruses in Human Cancers. Proc. Assoc. Am. Physicians 111;(6), 581-587.
- zur Hausen, H. (2000) Papillomaviruses Causing Cancer: Evasion from Host-Cell Control in Early Events in Carcinogenesis. J. Natl. Cancer Inst. 92;(9), 690-698.

RÉSUMÉ :

L'oncoprotéine E6 des papillomavirus humains (HPV) à « haut-risque » est à l'origine du cancer du col de l'utérus. Elle participe à l'oncogénèse notamment en dégradant p53, la protéine cellulaire suppresseur de tumeurs. Le but de la thèse était d'obtenir la structure tridimensionnelle par Résonance Magnétique Nucléaire (RMN) de la protéine E6.

Le manuscrit revient d'abord sur les différents états (micro- et macroscopique) d'une protéine globulaire en solution, et sur des notions telles que repliement et stabilité. Il rappelle ensuite les principes et limites de différentes techniques biophysiques appliquées à l'étude de protéines. Enfin, il décrit brièvement le contexte biologique de E6.

Bien que sa séquence primaire soit connue depuis 1985, la protéine E6 n'avait pas encore été purifiée. Nous décrivons ici un protocole d'expression et de purification de cette molécule, en montrant tout d'abord que l'expression bactérienne de fusions MBP-E6 (Maltose Binding Protein) génère des « corps d'inclusion solubles ». Ces particules, engendrés par l'agrégation de protéines E6 mal repliées, restent toutefois en solution grâce à la haute solubilité de MBP. Grâce à ces observations, nous avons pu produire des échantillons de la protéine E6 entière soluble et repliée, puis de ses deux domaines à zinc. Nous avons finalement obtenu la détermination de la structure tridimensionnelle par RMN du domaine C-terminal de E6. L'amélioration de la qualité de E6 recombinante a permis également de démontrer l'interaction de E6 avec un motif structural d'ADN présent dans l'ADN cruciforme, puis de déterminer les paramètres cinétiques de cette interaction par BIAcore.

Ce travail ouvre donc la voie à une meilleure compréhension moléculaire des modes d'actions de E6, et donc à de nouvelles stratégies de thérapie du cancer du col de l'utérus. De plus, nos méthodologies de contrôle et d'optimisation de la qualité des protéines de fusion sont généralisables à l'étude de toute protéine récalcitrante.

ABSTRACT :

E6 is an oncoprotein produced by "high risk" Human Papillomaviruses (HPVs) involved in cervical cancers. E6 participates oncogenesis through different pathways, in particular by degrading the cellular tumor suppressor protein p53. The aim of this thesis was to solve the solution structure of E6 by Nuclear Magnetic Resonance (NMR).

The introduction chapter first focuses on the different states (micro- and macroscopic) adopted by globular proteins in solution, including notions such as folding and stability. Then we record the principles, applications and limits of various biophysical techniques for the study of proteins. Finally, we briefly introduce the biological context of E6 protein.

At the start of this work, E6 had never been purified although its sequence was known since 1985. In the results section, we first demonstrate that bacterial expression of E6 fused to the C-terminus of MBP (Maltose Binding Protein) generates « soluble inclusion bodies ». These particles, which originate from agregation of misfolded E6 moieties, remain soluble thanks to the high solubility of MBP moities. These observations have allowed us to produce soluble and folded samples of full-length E6 as well as its two zinc-binding domains. Finally, we have managed to solve the NMR structure of the C-terminal zinc-binding domain. On another hand, the optimized quality of our E6 samples have allowed us to demonstrate E6 binding to a particular DNA structural motif found in four-way DNA junctions. The kinetic parameters of this interaction have been determined by BIAcore.

This work will provide a better understanding of the molecular pathways of E6 action and opens the way for new therapeutic strategies against cervical cancers. Furthermore, our methods for control and optimization of protein fusion quality can be generalized for future studies of other recalcitrant proteins.