



EXPRESSION, PURIFICATION ET ÉTUDE STRUCTURALE PAR RMN DE PEBP DE DROSOPHILE

Rautureau Gilles

► To cite this version:

Rautureau Gilles. EXPRESSION, PURIFICATION ET ÉTUDE STRUCTURALE PAR RMN DE PEBP DE DROSOPHILE. Autre [q-bio.OT]. Université d'Orléans, 2006. Français. NNT : . tel-00139424

HAL Id: tel-00139424

<https://theses.hal.science/tel-00139424>

Submitted on 30 Mar 2007

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

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



THESE PRESENTEE A L'UNIVERSITE D'ORLEANS
POUR OBTENIR LE GRADE DE
DOCTEUR DE L'UNIVERSITE D'ORLEANS

Discipline : Biophysique Moléculaire

par

Gilles RAUTUREAU

**Expression, purification et étude structurale par RMN de
PEBP de drosophile**

Soutenue le 11 décembre 2006

MEMBRES DU JURY :

Françoise Vovelle ; Professeur, Université d'Orléans

Françoise Guerlesquin ; Directrice de recherche, CNRS, Marseille

Didier Marion, Directeur de recherche, INRA, Nantes

Isabelle Callebaut ; Directrice de recherche, CNRS, Paris

Laurence Jouvensal ; Maître de conférence, Université d'Orléans

Françoise Schoentgen ; Directrice de recherche, CNRS, Orléans

Directeur de thèse

Rapporteur

Rapporteur

Examineur

Examineur

Examineur

Remerciements

Ce manuscrit est pour moi l'occasion de faire le bilan de quatre années de travail passées au Centre de Biophysique Moléculaire du CNRS à Orléans dirigé par Monsieur Jean-Claude Beloeil.

Ces années, incluant mon DEA et ma thèse, auront été pour moi une vraie source d'épanouissement tant professionnel que personnel. Aussi je tiens à exprimer ici toute ma reconnaissance aux personnes qui ont contribué à ma formation scientifique et rendu possible cette thèse en me faisant bénéficier de conditions de travail propices à la recherche et de conseils stimulant la réflexion, tout en restant humainement très chaleureuses.

S'il est impossible de nommer ici toutes ces personnes, je tiens néanmoins à souligner tout mon intérêt à côtoyer l'ensemble des membres du CBM qui ont su partager leurs savoir-faire, leurs connaissances et leur enthousiasme. Une équipe que je quitterai avec regret tant pour l'ambiance à la fois motivée et détendue du laboratoire que pour le sens des recherches qui y sont menées.

Je remercie également particulièrement les membres de l'équipe de RMN et de modélisation moléculaire de protéines et de peptides, dirigée avec une grande compétence et une implication de chaque instant par le Professeur Françoise Vovelle et dont les membres sont ou ont été : Baptiste Legrand, Pedro Da Silva, Henri Iabbé, Hervé Meudal, Denise Sy, Laurence Jouvensal, Céline Landon et Christian Damblon. Merci Françoise de m'avoir accueilli au sein de ton équipe, merci pour ta disponibilité et la qualité de ton encadrement qui restent pour moi un modèle de professionnalisme et de constance allié à un esprit d'ouverture marqué par la générosité. Merci Laurence d'avoir accepté de travailler avec moi sur ce sujet ambitieux sans jamais renoncer face aux nombreuses difficultés qui ont émaillées ce parcours. Tu m'as "scientifiquement" démontré tous les bénéfices de la patience, la rigueur, la minutie et l'indispensable refus de l'approximation.

Cette thèse est également issue d'une collaboration passionnante avec différentes équipes du CBM d'Orléans. Aussi, je souhaite ici exprimer ma gratitude à Françoise Schoentgen pour sa confiance, ses encouragements et sa grande disponibilité ; ainsi qu'à Martine Decoville et à Daniel Locker pour leur soutien et la qualité de leurs conseils scientifiques. Françoise, sans votre parfaite connaissance des PEBP et sans votre expérience en biochimie et en

purification des protéines, cette thèse n'aurait sans doute jamais abouti, merci !

Merci aussi à mes amis fidèles, Fabien, Julien, Violaine, Florence... qui n'ont pas hésité à lire et relire l'ensemble de cette thèse pour me prodiguer des conseils judicieux non pas sur les données scientifiques mais sur les conditions permettant de faciliter la lecture, et donc le partage, de ce travail.

Enfin, je voudrais dédier cette thèse à ma famille qui m'a offert toutes les conditions pour « découvrir le monde » et trouver ma voie...

... Ainsi qu'à tous et à toutes. En espérant que ce travail participe, même modestement, à l'idéal scientifique... mieux connaître le monde pour contribuer à son épanouissement !

Sommaire

Introduction	15
Introduction	17
Introduction	17
I. La famille des PEBP	21
1. Les PEBP	21
2. Caractéristiques moléculaires	25
2.1. Taille moléculaire	25
2.2. Modifications post-traductionnelles	25
2.3. Localisation	25
2.4. Fixation de ligands	26
2.5. Stabilité physicochimique	27
2.6. Données structurales	27
2.6.1. La PEBP bovine	29
2.6.2. La PEBP humaine	32
2.6.3. Les PEBP des autres espèces	34
2.6.3.a. La PEBP-2 de souris	34
2.6.3.b. La protéine CEN (1QOU)	36
2.6.3.c. Les protéines TFL1 (1WKO) et FT (1WKP) d'Arabidopsis thaliana	38
2.6.3.d. La protéine TFS1 de la levure Saccharomyces cerevisiae	40
2.6.3.e. Les protéines YBHB et YBCL d'E. coli	42
2.6.4. PEBP et liaisons cis	45
2.6.5. PEBP et dimères	46
3. Activités moléculaires	47
3.1. Inhibition de kinases	47
3.2. Inhibition de protéases	48
3.3. Activité liée à la fixation de ligands anioniques	49
4. Rôles cellulaires et physiologiques	49
4.1. Rôles cellulaires : les voies de transduction du signal	49
4.1.1. RKIP intervient dans plusieurs voies de transduction du signal.	50
4.1.1.a. RKIP inhibe la voie des MAP-kinases Raf/MEK/ERK	50
4.1.1.b. La phosphorylation de RKIP régule son activité	52
4.1.1.c. RKIP inhibe la voie NFκB	52
4.1.1.d. RKIP phosphorylée inhibe la kinase GRK2	53
4.1.1.e. RKIP module l'activité des protéines G hétérotrimériques	55
4.1.2. Au delà des voies de transduction : apoptose survie des cellules	55
4.1.2.a. Rôle de RKIP dans la cytotoxicité des agents chimiothérapeutiques dans les lignées de cellules tumorales.	56
4.1.2.b. Induction de RKIP par le Rituximab et sensibilisation des lymphomes non hodgkiniens aux agents chimiothérapeutiques	57
4.1.3. PEBP et membranes cellulaires	57
4.2. Rôles physiologiques	58
4.2.1. Contrôle de l'inflorescence chez les plantes	58
4.2.1.a. Arabidopsis thaliana	59
4.2.1.b. Antirrhinum centroradialis	59
4.2.1.c. Lycopersicon esculentum (tomate)	60
4.2.2. Spermatogénèse	62
4.2.3. Transport de la morphine	63
4.2.4. Le peptide HCNP : physiologie cardiaque / cerveau	64

5.	PEBP humaine et maladies	65
5.1.	RKIP est un suppresseur de métastases.....	65
5.2.	Maladie d'Alzheimer.....	66
5.3.	Diabète	66
II.	Le projet « PEBP de drosophile »	68
1.	Contexte scientifique.....	68
1.1.	Aspects biologiques.....	68
1.2.	Aspects structuraux	69
1.3.	Les interactions	69
2.	Les PEBP de drosophiles	69
2.1.	Les gènes codant pour des PEBP chez la drosophile	70
2.2.	Les 7 PEBP de drosophiles	71
3.	Le projet d'étude des PEBP de drosophile.....	73
3.1.	Objectifs	73
3.2.	Mon rôle au sein du projet.....	74
III.	Matériels et méthodes.....	77
1.	Dichroïsme circulaire :	77
1.1.	Principe de la méthode	77
1.2.	Dichroïsme circulaire des protéines	79
2.	La RMN pour l'étude de la structure des protéines	80
2.1.	La RMN : progrès technologiques	82
2.1.1.	Caractéristiques du signal RMN	82
2.1.2.	Les progrès instrumentaux et techniques.	84
2.1.3.	Le marquage isotopique et la RMN hétéronucléaire.....	85
2.2.	Stratégie d'attribution pour une protéine de 20 kDa	86
2.2.1.	Attribution des résonances du squelette peptidique d'une protéine.....	87
2.2.2.	Attribution des résonances des chaînes latérales.....	90
2.3.	Paramètres spectroscopiques apportant des informations structurales	91
2.3.1.	Utilisation de l'index de déplacement chimique (Chemical Shift Index, CSI) pour déterminer les structures secondaires	92
2.3.2.	Mesure des distances à partir des NOE.....	93
2.3.3.	Mesure des angles dièdres à partir des constantes de couplage scalaire.....	94
2.3.4.	Vitesse d'échange des protons avec le solvant pour déterminer les liaisons hydrogène.....	97
2.3.5.	Informations sur l'orientation (RDC).....	97
2.4.	Modélisation moléculaire à partir des données RMN.....	98
2.4.1.	Bases de la modélisation moléculaire à partir des données RMN.	98
2.4.1.a.	Les données RMN utilisées dans la modélisation.....	98
2.4.1.a.i.	Calibration des données NOE.	98
2.4.1.a.ii.	Obtention des angles de torsion	99
2.4.1.a.iii.	Autres données	100
2.4.1.b.	La modélisation par recuit simulé.	100
2.4.2.	Le logiciel ARIA.....	102
2.4.2.a.	Contraintes de distance ambiguë.....	103
2.4.2.b.	Calibration des pics NOE	103
2.4.2.c.	Attribution des pics NOE ambigus.....	104
2.4.2.d.	Principe de fonctionnement du logiciel ARIA.....	105
2.4.3.	Le logiciel CYANA	107
2.5.	Etude de la dynamique d'une protéine par la relaxation ¹⁵ N	110
2.5.1.	Les paramètres T ₁ et T ₂	110

2.5.1.a.	La relaxation longitudinale	111
2.5.1.b.	La relaxation transversale	111
2.5.1.c.	Mesure des paramètres T_1 et T_2	112
2.5.2.	Le paramètre NOE	113
2.5.3.	Extraction des paramètres de dynamique	115
IV.	Etude de la protéine CG18594	121
1.	Données de base	122
1.1.	Analyse de la séquence	122
1.2.	Alignements de séquence	123
1.2.1.	Alignement de la protéine CG18594 avec les autres PEBP de drosophile	123
1.2.2.	Alignement de CG18594 avec les PEBP des autres espèces	125
2.	« Cloning, high yield over-expression, purification, and characterization of CG18594, a new PEBP/RKIP family member from <i>Drosophila melanogaster</i> »	127
3.	Résultats et discussion	135
3.1.	Production de la protéine CG18594	135
3.2.	Recherche des conditions RMN	137
3.2.1.	Étude de conditions en gouttes suspendues	138
3.2.2.	Diffusion dynamique de la lumière (DLS)	141
3.2.2.a.	Principe de la méthode	141
3.2.2.b.	Application à la protéine CG18594	143
3.2.3.	Dichroïsme circulaire	145
3.2.3.a.	Effet du pH	145
3.2.3.b.	Comportement de la protéine lors du chauffage, en fonction du pH	147
3.3.	Résonance magnétique nucléaire	149
3.3.1.	Stabilité de la protéine CG18594 dans le temps	149
3.3.2.	RMN à 5°C	151
3.3.3.	Agrégation de la protéine en présence de sel, à 5°C	152
3.3.4.	Effet de la température sur les spectres RMN	153
3.3.5.	Etude de la forme molten globule	155
3.3.6.	Conclusion	156
4.	Conclusion	156
V.	Etude de la protéine CG10298	161
1.	Données de base :	162
1.1.	Comparaison de la protéine CG10298 avec les autres PEBP de drosophile	164
1.2.	Comparaison de la protéine CG10298 avec les PEBP des autres espèces	165
2.	« <i>Drosophila melanogaster</i> CG10298 is a new testis-specific member of the phosphatidylethanolamine binding protein family. »	167
3.	Stabilité de la protéine au cours du temps	181
4.	Conclusion	182
VI.	Etude structurale de la protéine CG7054	187
1.	Données de base	187
1.1.	Comparaison de la protéine CG7054 avec les autres PEBP de drosophile	189
1.2.	Comparaison de la protéine CG7054 avec les PEBP des autres espèces	190
2.	Matériels et méthodes	193
2.1.	Clonage de l'ORF du gène <i>cg7054</i>	193
2.2.	Production et purification de la protéine CG7054	194
2.2.1.	Culture bactérienne et surexpression de la protéine, cas de la protéine non marquée	194
2.2.2.	Culture bactérienne et surexpression de la protéine, cas de la protéine marquée ^{15}N et doublement marquée $^{13}\text{C} / ^{15}\text{N}$	194

2.2.3.	Extraction et purification de la protéine	195
2.2.4.	Chromatographie d'échange d'ions	196
2.2.5.	Dialyses et lyophilisation	196
2.2.6.	Conservation.....	196
2.3.	Recherche des conditions RMN	197
2.3.1.	Etude des conditions en gouttes suspendues	197
2.3.2.	Dichroïsme circulaire	197
2.4.	Résonance magnétique nucléaire	198
2.4.1.	Stabilité de la protéine CG7054 au cours du temps	198
2.4.2.	Acquisition des spectres	198
2.4.3.	Expériences d'échange proton / deutérium	198
2.4.4.	Attribution du squelette peptidique	199
2.4.4.a.	Le logiciel Smartnotebook	199
2.4.4.b.	Le logiciel PACES	199
2.4.4.c.	Le logiciel PISTACHIO	200
2.4.5.	Attribution des chaînes latérales	200
2.4.5.a.	Attribution des chaînes latérales aliphatiques	200
2.4.5.b.	Les groupements NH ₂ des asparagines et glutamines	201
2.4.5.c.	Attribution des groupes aromatiques.....	201
2.4.6.	Détermination des paramètres de relaxation	201
2.4.6.a.	Détermination des temps de relaxation T ₁ et T ₂	201
2.4.6.b.	Détermination de l'effet NOE hétéronucléaire ¹ H- ¹⁵ N	203
2.5.	Calcul de structure.....	204
2.6.	Interaction de la protéine CG7054 avec la phosphoryléthanolamine	204
3.	Résultats et discussion.....	204
3.1.	Clonage de l'ADNC du gène <i>cg7054</i>	204
3.2.	Production et purification de la protéine CG7054	207
3.3.	Recherche des conditions RMN.....	209
3.3.1.	Étude de conditions en gouttes suspendues.....	209
3.3.2.	Dichroïsme circulaire	209
3.3.2.a.	Effet du pH	210
3.3.2.b.	Effet de la température	211
3.3.2.c.	Conclusions pour le choix des conditions de l'échantillon pour l'étude structurale par RMN.....	213
3.4.	RMN.....	214
3.4.1.	Stabilité de la protéine CG7054 au cours du temps	214
3.4.2.	Attribution du squelette peptidique	214
3.4.3.	Attribution des chaînes latérales	216
3.4.4.	Collection des données structurales	217
3.4.4.a.	Identification des structures secondaires et contraintes d'angles.....	217
3.4.4.a.i.	CSI.....	218
3.4.4.a.ii.	Talos	218
3.4.4.a.iii.	Comparaison des différentes méthodes.....	219
3.4.4.b.	Echange H / D	221
3.5.	Détermination de la structure de la protéine CG7054.....	223
3.5.1.	Résultats de l'attribution des corrélations NOE et des contraintes de distances	224
3.5.2.	Analyse des structures.....	226
3.5.2.a.	Qualité des structures	226
3.5.2.b.	Topologie et structure de la protéine CG7054	230

3.5.2.c.	Comparaison de la protéine CG7054 et de la PEBP-1 bovine	231
3.5.2.d.	Comparaison du site d'interaction avec celui de la PEBP bovine.....	233
3.5.2.e.	Comparaison avec les autres PEBP de structure 3D connue.....	236
3.5.2.f.	Conclusion	236
3.6.	Dynamique de la protéine CG7054	237
3.6.1.	Mesures des temps de relaxation longitudinale T_1 et transversale T_2	238
3.6.2.	Mesure de l'effet NOE hétéronucléaire ^1H - ^{15}N	242
3.6.3.	Détermination du temps de corrélation τ_c	243
3.6.4.	Interprétation des paramètres de relaxation par le formalisme de Lipari-Szabo.	243
3.7.	Interaction de la protéine CG7054 avec la phosphoryléthanolamine.....	245
4.	Conclusion	251
VII.	Conclusions, perspectives	255
	Références Bibliographiques	261
	Annexe 1 : Clonage de la région codante du gène <i>cg7054</i>	273
1.	Matériel initial	273
2.	Amplification par PCR du gène CG7054	273
3.	Clonage dans le vecteur pCR 2.1 – TOPO	274
3.1.	Insertion dans le vecteur pCR2.1-TOPO	274
3.2.	Transformation dans les bactéries Top10:.....	275
3.3.	Extraction de plasmide	275
4.	Préparation de l'insert CG7054	276
5.	Transfert dans le vecteur pET31b ouvert, ligation	277
6.	Etape de transformation.....	277
6.1.	Préparation de cellules compétentes.....	277
6.2.	Transformation	278
7.	Analyse finale des clones de bactéries BL21DE3 ayant intégré le plasmide pET31b-CG7054	278
	Annexe 2 : Paramètres d'acquisition des expériences 2D et 3D mises en œuvre pour l'étude structurale de la protéine CG7054	280
	Annexe 3 : Résultats obtenus pour les indices de déplacement chimique de la protéine CG7054.....	282
	Annexe 4 : Angles prédits par le logiciel Talos pour la protéine CG7054.....	284
	Annexe 5 : Table de déplacements chimiques de la protéine CG7054.	285

Abréviations

2D : bidimensionnel(le)
3D : tridimensionnel(le)
ADNc : Acide DésoxyriboNucléique complémentaire
ADR : Ambiguous Distance Restraint
ARIA : Ambiguous distance Restraints for Iterative Assignments
ARNm : Acide RiboNucléique messenger
BLAST : Basic Alignment Search Tool
CD : Circular Dichroism
CNS : Crystallography NMR System
COSY : COrrrelation SpectroscopY
CSA : Chemical Shift Anisotropy
CSI : Chemical Shift Index
DSS : 2,2-Dimethyl-2-silapentane-5-sulfonate de sodium
EDTA : Ethylène diamine tétra-acétate
HCA : Hydrophobic Cluster Analysis
HOHAHA : HOmonuclear HARTman HAHn spectroscopy
HSQC : Heteronuclear Single Quantum Coherence
IEF : Isoélectrofocalisation
INEPT : Insensitive Nuclei Enhanced Polarisation Transfer
NMR : Nuclear Magnetic Resonance
NOE : Nuclear Overhauser Effect
NOESY : Nuclear Overhauser Effect SpectroscopY
Pb : Paire de base
PDB : Protein Data Bank
pI : Point Isolélectrique
RMN : Résonance Magnétique Nucléaire
RMSD : Root Mean Square Deviation
RX : Rayons X
SDS-PAGE : Sodium Dodecyl Sulfate Poly-Acrylamide Gel Electrophoresis
SW : Largeur spectrale
TOCSY : Total Correlation SpectroscopY
TROSY : Transverse Relaxation Optimized SpectroscopY
UV : Ultra-Violet

Introduction

Introduction

Les PEBP ou Phosphatidyléthanolamine-Binding Proteins sont des protéines d'environ 20 kDa conservées depuis les bactéries jusqu'aux mammifères et aux plantes. Elles sont caractérisées par leur capacité de fixer des ligands anioniques comme des phospholipides, des nucléotides ou des dérivés de la morphine. Les PEBP agissent dans un grand nombre de processus cellulaires et physiologiques de l'organisme sain. Elles interviennent par exemple au niveau de plusieurs voies de transduction du signal: elles contrôlent les protéines G, inhibent les voies MAP-kinases et NFkB. Par ailleurs des modifications de l'expression de ces protéines sont corrélées à des situations pathologiques pour lesquelles les stratégies thérapeutiques demeurent fragiles (cancers primaires et métastatiques, diabète, maladie d'Alzheimer, infertilité...). Malheureusement les fonctions, rôles et modes d'action de ces protéines ne sont pas clairement établis ou compris.

Afin de mieux comprendre les relations structure / fonction au sein de cette famille de protéines qui apparaissent majeures à la vue des implications de ces protéines chez l'humain, un projet d'étude des PEBP de drosophile a été mis en place au CBM. La mouche du vinaigre a été choisie car il s'agit d'un organisme modèle pour la détermination des rôles biologiques et des modes d'action des protéines. Il existe 7 gènes codant pour des PEBP chez la drosophile : *cg18594*, *cg7054*, *cg10298*, *cg17919*, *cg6180*, *cg17917* et *cg5430*. Le projet s'articule autour de trois équipes du CBM : l'équipe de Daniel Locker, spécialisée dans l'étude des drosophiles pour déterminer le rôle biologique des PEBP *in vivo*, l'équipe de Françoise Schoentgen pour la partie biologie moléculaire et la purification des protéines et enfin l'équipe de Françoise Vovelle pour les études structurales. Les objectifs sont simples : obtenir de nouvelles informations biologiques et structurales concernant les PEBP en utilisant un modèle biologique reconnu et encore inexploré.

L'étude structurale par RMN, dont l'envergure dépasse le cadre de cette thèse, a été initiée afin de définir dans un premier temps la structure de ces protéines en solution mais aussi leur dynamique interne. Ensuite, en relation avec les données biologiques et biochimiques recueillies, des études d'interaction protéine-protéine ou protéine-ligand seront menées.

Durant la première partie de ma thèse j'ai exprimé la protéine CG18594 (marquée ¹⁵N ou non) dans *Escherichia coli* et mis au point les protocoles de purification pour la produire

en quantité importante. J'ai ensuite testé le comportement de la protéine purifiée dans différentes conditions de tampons et de sels afin de choisir les conditions les plus favorables à sa monodispersion et sa solubilité. Dans les conditions déterminées la protéine produite est soluble et repliée. Cependant elle n'est pas stable plus de quelques jours à 20°C. Elle est stable à 5°C mais les spectres enregistrés sont difficilement utilisables (largeurs de raie importantes, absence de transfert d'aimantation transversale, manque de résolution). Les propriétés physico-chimiques de la protéine CG18594 ne sont donc pas compatibles avec une étude de sa structure tridimensionnelle par RMN. Cette protéine est en cours d'essai de cristallisation au laboratoire, en collaboration avec l'équipe d'Alain Roussel.

J'ai continué cette étude en réalisant le clonage des sections codantes des gènes *cg10298*, *cg7054* et *cg6180* dans *Escherichia coli*. Des protocoles de purification ont été mis au point pour les protéines CG10298 et CG7054. Actuellement, nous pouvons les produire purifiées, marquées ou non, en grande quantité. Une étude de leur stabilité en solution a été réalisée par RMN: la protéine CG10298 n'est stable qu'environ 10 jours à 20°C alors que la protéine CG7054 est stable plusieurs mois. Cette dernière a donc été choisie pour une étude structurale. Après acquisition des spectres hétéronucléaires, nous avons réalisé l'attribution des signaux ^1H , ^{13}C et ^{15}N du squelette peptidique et des chaînes latérales. L'étude de la dynamique du squelette peptidique a été réalisée et la structure 3D a été déterminée.

Ce mémoire s'articule en cinq parties. La première présente la bibliographie des PEBP. La seconde est consacrée à l'exposé des techniques utilisées pour réaliser ce travail : dichroïsme circulaire et RMN. Les trois chapitres suivants sont dédiés à la présentation des résultats obtenus : le premier décrit le travail réalisé sur la protéine CG18594, le second détaille les résultats des études de la protéine CG10298. Enfin, le troisième expose les résultats de l'étude structurale obtenue pour la protéine CG7054.

La famille des PEBP

I. La famille des PEBP

1. Les PEBP

En 1984, Bernier et Jolles isolèrent, à partir du cerveau de bœuf, une protéine soluble, basique et de masse moléculaire apparente 23 kDa (Bernier et Jolles, 1984). Sa séquence ne présentait aucune homologie significative avec celles des autres protéines connues à l'époque (Schoentgen *et al.*, 1987). De plus, il apparut rapidement que la protéine restait associée durant sa purification à des lipides endogènes (triacylglycérides et esters de cholestérol) (Bernier et Jolles, 1984). L'interaction avec des molécules hydrophobes fut donc testée et un début de caractérisation montra qu'elle était capable de lier de manière non covalente les anions organiques comme la bromosulfophtaléine, le sulfate de déhydroépiandrostérone, l'estradiol, le linoléate et l'oléate de cholestérol, mais aussi des phospholipides (Bernier *et al.*, 1986). Ces études mirent en évidence la présence d'un site saturable pour la phosphatidyléthanolamine, tandis que la phosphatidylcholine se liait de façon non spécifique. La protéine fut donc nommée « phosphatidylethanolamine binding protein » (PEBP) (Bernier *et al.*, 1986).

Au fil des années, des protéines homologues à la PEBP bovine ont été isolées et caractérisées. Elles correspondent toutes à des protéines de 20 kDa environ. Une famille s'est ainsi progressivement constituée et une séquence protéique consensus a été retenue. Elle définit l'appartenance à la famille des PEBP (Bernier et Jolles, 1984) :

$$\text{DPD-x-P-x(11)-H-x(28-30)-G-x-HR}$$

Chez certaines bactéries, on retrouve la séquence consensus légèrement modifiée :

$$\text{DPD-x-P-x(6)-H-x(28-30)-G-x-HR}$$

Les protéines de la famille des PEBP présentent une grande conservation des résidus, notamment dans la partie centrale (résidus 65 à 130 de la PEBP bovine). Les extrémités N- et C- terminales apparaissent plus variables d'une protéine à l'autre.

Un début de classification a été réalisé pour les PEBP de mammifères. A partir d'un alignement de séquence, Simister et ses collaborateurs ont distingué, en 2002, 4 sous-

[illegible]

Chez l'homme, les 2 PEBP caractérisées ont été nommées hPEBP-1 (aussi appelée RKIP) et hPEBP-4. La protéine de souris mPEBP-2 présente 84 % d'identité avec hPEBP-1. Les PEBP-2 de rat et de souris (rPEBP-2 et mPEBP-2) présentent 91 % d'identité de séquence entre elles. La protéine mPEBP-3 de souris est légèrement plus éloignée : elle possède 79 % d'identité avec hPEBP-1. La sous-famille PEBP-4 est caractérisée par 2 insertions (entre les résidus 55-56 et 102-103 de hPEBP-1) et une délétion (131 à 134 de hPEBP-1). La protéine mPEBP-4 ne présente que 27 % d'identité avec la protéine hPEBP-1.

22

moins si on recherche une signature complète. Actuellement plus de 450 séquences protéiques de PEBP putatives sont connues.

Au sein d'une même espèce il est habituel de trouver plusieurs gènes putatifs codant pour des homologues de PEBP. Ainsi, on observe chez l'homme 6 gènes putatifs codant pour des protéines portant la signature, 5 chez le bœuf, 6 chez la souris. Il existe chez la drosophile *Drosophila melanogaster* 7 séquences codant pour des PEBP putatives enregistrées dans la banque Pfam, dont l'Odorant Binding Protein 5 (OBP5). Six séquences présentent la signature complète des PEBP, la septième, l'OBP5, présente une signature légèrement altérée (CPDAP et QIHR au lieu de DPDXP et GXHR). La Figure I-2 présente l'alignement des séquences de ces 7 PEBP putatives de drosophile avec la protéine humaine hPEBP-1. Le Tableau I-1 indique les pourcentages d'identité de séquence observés entre ces protéines.

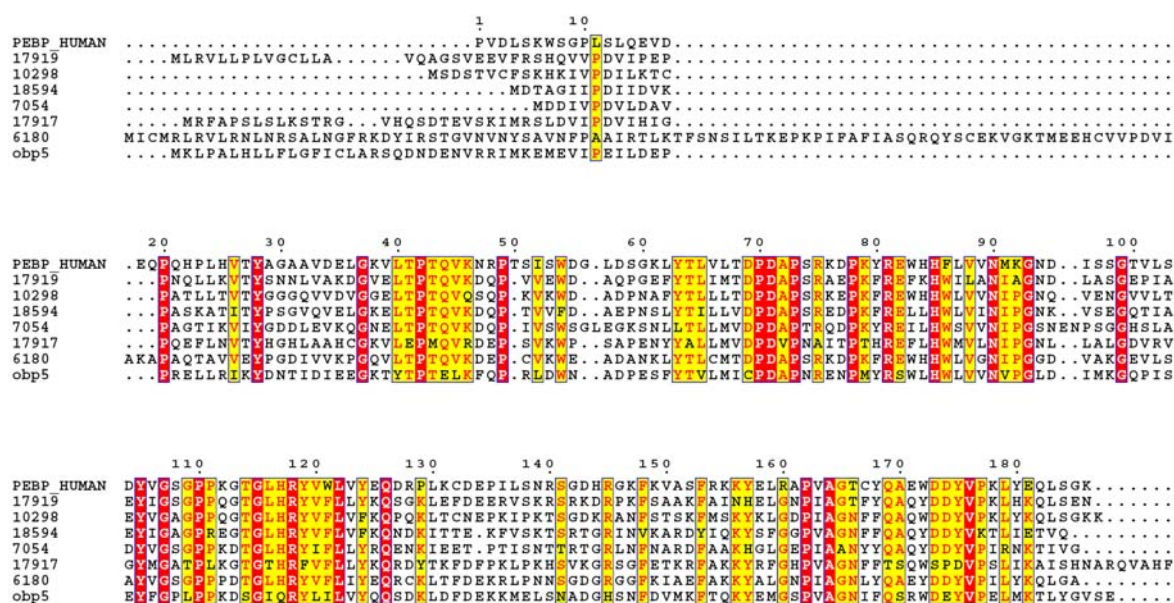


Figure I-2 : Alignement de séquence des PEBP de drosophile avec la protéine humaine hPEBP-1 (ClustalW, (Eddy, 1995)). Les résidus identiques sont indiqués par un cadre rouge, un cadre jaune indique une forte conservation.

	hPEBP-1	CG6180	CG17919	CG10298	CG18594	CG 7054	CG 17917	OBP5
hPEBP-1		53	47	51	42	43	35	38
CG 6180	53		50	55	48	50	37	36
CG 17919	47	50		54	48	49	40	35
CG 10298	51	55	54		55	48	41	40
CG 18594	42	48	48	55		48	40	42
CG 7054	43	50	49	48	48		37	35
CG 17917	35	37	40	41	40	37		32
OBP5	38	36	35	40	42	35	32	

Tableau I-1 : Pourcentages d'identité (nombre de résidus identiques divisé par le nombre de résidus comparés hors insertions) observés entre les PEBP putatives de drosophile et la protéine humaine hPEBP-1.

Malheureusement, on ne dispose à l'heure actuelle d'aucune étude génétique systématique sur l'expression de l'ensemble des gènes présents au sein d'un même organisme. Chez l'homme, 2 protéines ont été étudiées sur les 6 putatives, une seule sur 5 chez le bœuf et 2 sur 6 chez la souris.

Concernant l'expression d'une même protéine au sein d'un organisme, des études d'immunolocalisation ont montré que la protéine hPEBP-1 humaine est retrouvée au niveau de nombreux tissus. Dans le cerveau, hPEBP-1 est présente dans les oligodendrocytes, les cellules de Schwann, les cellules de Purkinje dans le cervelet, ainsi qu'au niveau des six couches de cellules spécifiques du cortex et de l'hippocampe (Frayne *et al.*, 1999). On trouve aussi hPEBP-1 dans des cellules et tissus non neuronaux, comme les spermatozoïdes, les plaquettes sanguines, les cellules interstitielles de Leydig, les entérocytes, les mégacaryocytes de la rate, au niveau de l'épithélium de l'épididyme, de la prostate, du tube contourné proximal du rein, du petit intestin, des bronchioles, des oviductes et de l'utérus, au niveau d'organes entiers comme le cœur, le foie, les ovaires, les glandes mammaires et la thyroïde. hPEBP-1 est observée dans la salive (Hirtz *et al.*, 2005), mais pas dans le lait ni les sécrétions prostatiques (Schoentgen et Jolles, 1995). Par opposition, la protéine de souris mPEBP-2 n'est exprimée qu'au niveau de l'appareil reproducteur male, uniquement dans les cellules germinales.

2. Caractéristiques moléculaires

Nous allons, dans cette partie, détailler les caractéristiques moléculaires primaires des PEBP : leur taille, leur localisation cellulaire, leur capacité de fixation des ligands anioniques, leur stabilité physicochimique et enfin leur structure.

2.1. Taille moléculaire

Les PEBP sont des protéines solubles d'environ 20 à 25 kDa. hPEBP-1 et la PEBP-1 bovine, très proches, sont toutes deux constituées de 186 résidus, pour une masse moléculaire d'environ 21 kDa, alors que la protéine de levure TFS1 présente 219 résidus pour une masse de 24 kDa. Parmi les plus petites, on trouve la protéine d'*Escherichia coli* YBHB qui ne comporte que 158 résidus pour une masse de 17 kDa. Le point isoélectrique des PEBP n'est pas conservé et varie entre 5 et 9.

2.2. Modifications post-traductionnelles

Les PEBP ne présentent habituellement pas de modifications post-traductionnelles constitutives particulières. Cependant, certaines modifications peuvent intervenir dans des situations bien précises, par exemple pour la régulation de l'activité de ces protéines. Ainsi, la phosphorylation de la serine 153 de la PEBP-1 humaine par la protéine kinase C (PKC) permet de moduler ses capacités d'interaction avec d'autres protéines intervenant dans la voie des MAP-kinases et dans l'internalisation des récepteurs liés aux protéines G hétérotrimériques (Corbit *et al.*, 2003). On note aussi que l'acétylation N-terminale de la protéine TFS1 (PEBP de levure) par la protéine NatB est nécessaire pour son activité d'inhibition de la carboxypeptidase Y (Caesar et Blomberg, 2004).

2.3. Localisation

Les PEBP sont le plus souvent observées dans le cytoplasme des cellules mais on peut aussi les trouver sécrétées dans le milieu extracellulaire. Ainsi, la protéine de rat rPEBP-1 est présente dans le liquide séminal (Saunders *et al.*, 1995), la protéine humaine hPEBP-1 dans le liquide céphalorachidien et dans le sérum (Goumon *et al.*, 2004). Il est cependant intéressant de remarquer que les PEBP ne comportent généralement pas de séquence signal de sécrétion caractéristique. Actuellement, sur les 453 séquences de PEBP

putatives disponibles, 55 présentent un signal de sécrétion, dont les protéines OBP5 et CG17919 de la drosophile *Drosophila melanogaster*.

Les PEBP sont solubles dans le cytoplasme, mais on les retrouve majoritairement associées aux membranes cellulaires (surtout au niveau de la membrane plasmique et des membranes du réticulum endoplasmique granuleux) (Roussel *et al.*, 1988). L'étude de la PEBP-1 bovine a montré que cette interaction se fait grâce à des liaisons non covalentes, la protéine interagissant avec les groupements polaires chargés négativement des couches phospholipidiques (Vallee *et al.*, 2001).

2.4. Fixation de ligands

Les PEBP sont caractérisées par leur capacité à lier des anions organiques. Cette propriété a particulièrement été étudiée pour la protéine PEBP-1 de bœuf. Ainsi Bernier et ses collaborateurs ont montré en 1986 que cette protéine liait la phosphatidyléthanolamine d'une manière spécifique, mais pas la phosphatidylcholine. La protéine n'interagit pas avec les acides gras isolés. L'interaction avec d'autres composés et hormones a été testée et il est apparu que la bromosulfophtaléine, le sulfate de déhydroépiandrostérone et l'estradiol-17 β étaient liés avec des constantes d'affinité de 10^6 , 10^4 et 10^3 M⁻¹ (Bernier *et al.*, 1986). Par ailleurs, Bucquoy et ses collaborateurs ont caractérisé l'interaction de la PEBP bovine avec différents ligands et nucléotides. Ainsi les interactions avec la trioléine, le linoléate et l'oléate de cholestérol ont montré des valeurs de Kd de 0,11 pM, 20 nM and 0,29 pM, respectivement. Pour les nucléotides, une forte affinité a été démontrée mais non quantifiée. Ainsi par ordre d'affinité décroissant : FMN > GTP > GDP > GMP > FAD > ATP > NADP > CTP > UTP > ADP. Au contraire, l'AMP, AMPc, GMPc et le NAD ne se lient pas (Bucquoy *et al.*, 1994). Enfin, la PEBP-1 humaine est capable de fixer des dérivés de morphine comme la morphine-6-glucuronide avec une affinité de l'ordre du nanomolaire (Goumon *et al.*, 2006).

Enfin, les PEBP sont caractérisées par leur capacité à interagir avec d'autres protéines. Ainsi un très grand nombre d'interactions a été mis en évidence, par exemple avec de petites protéines G (Kroslak *et al.*, 2001), des kinases (Banfield et Brady, 2000) ou des facteurs de transcription (Yeung *et al.*, 2001) (*cf. infra*).

2.5. Stabilité physicochimique

Seule la PEBP-1 bovine a été étudiée en détail par dichroïsme circulaire, fluorescence statique et dynamique et microcalorimétrie (Vallee *et al.*, 1999). L'étude de sa stabilité thermique a montré un profil de dénaturation irréversible à pH neutre, alors qu'à pH 4 la protéine est correctement repliée et sa dénaturation est parfaitement réversible. Dans les deux cas, la transition se produit à 54°C. L'étude fine du mécanisme de dénaturation par microcalorimétrie à balayage en température a permis de conclure à un modèle de dénaturation classique avec deux états (natif et dénaturé) en équilibre variable en fonction de la température. La variation d'enthalpie à pH 4 est de 119 kcal.mol⁻¹, et la variation d'énergie libre à 25°C est de 5 kcal.mol⁻¹. Ces résultats indiquent que la PEBP-1 bovine présente une stabilité physicochimique relativement faible, ce qui la rend flexible ; sa structure peut donc s'adapter lors des interactions avec ses nombreux partenaires, protéines ou ligands.

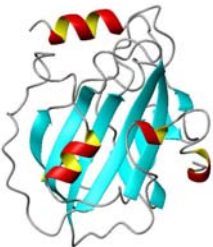
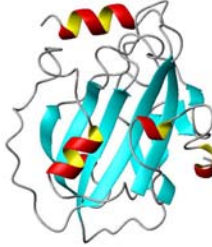
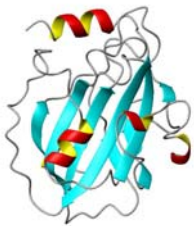
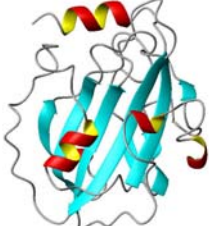
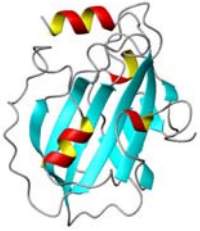
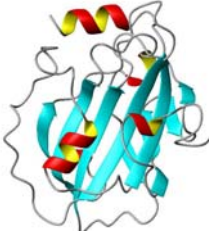
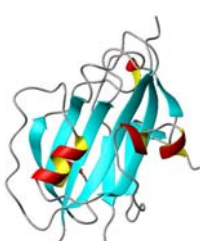
2.6. Données structurales

Les structures de quelques protéines appartenant à la famille des PEBP ont été résolues par cristallographie : les PEBP-1 bovine (Serre *et al.*, 1998) et humaine (Banfield *et al.*, 1998), la PEBP-2 de souris (Simister *et al.*, 2002), la protéine TFS1 de la levure *Saccharomyces cerevisiae* (Mima *et al.*, 2005), les protéines d'*Escherichia coli* YBHB et YBCL (Serre *et al.*, 2001), la protéine CEN d'*Antirrhinum centroradialis* (Banfield et Brady, 2000), les protéines TFL1 et FT d'*Arabidopsis thaliana* (Ahn *et al.*, 2005).

Toutes ces données permettent d'affirmer que les PEBP appartiennent à une famille structurale unique. Malgré un repliement global particulièrement conservé, les protéines de cette famille présentent des différences structurales significatives, notamment au niveau de leurs extrémités N- et C- terminales qui pourraient être un point clé pour leur fonction biologique, notamment pour la régulation de leurs interactions avec différents partenaires (Figure I-3).

Les protéines PEBP-1 humaine et bovine présentent un repliement extrêmement similaire. Dans la mesure où il s'agit des protéines les plus étudiées, nous commencerons par

détailler leurs structures, qui sont souvent considérées comme référence. Nous étudierons ensuite les ressemblances et les différences que nous pouvons observer avec les protéines d'autres espèces.

Protéine	Code PDB	Vue en stéréo	
PEBP-1 bovine	1A44		
PEBP-1 humaine	1BEH		
PEBP-2 souris	1KN3		
FT	1WKP		
TFL1	1WKO		

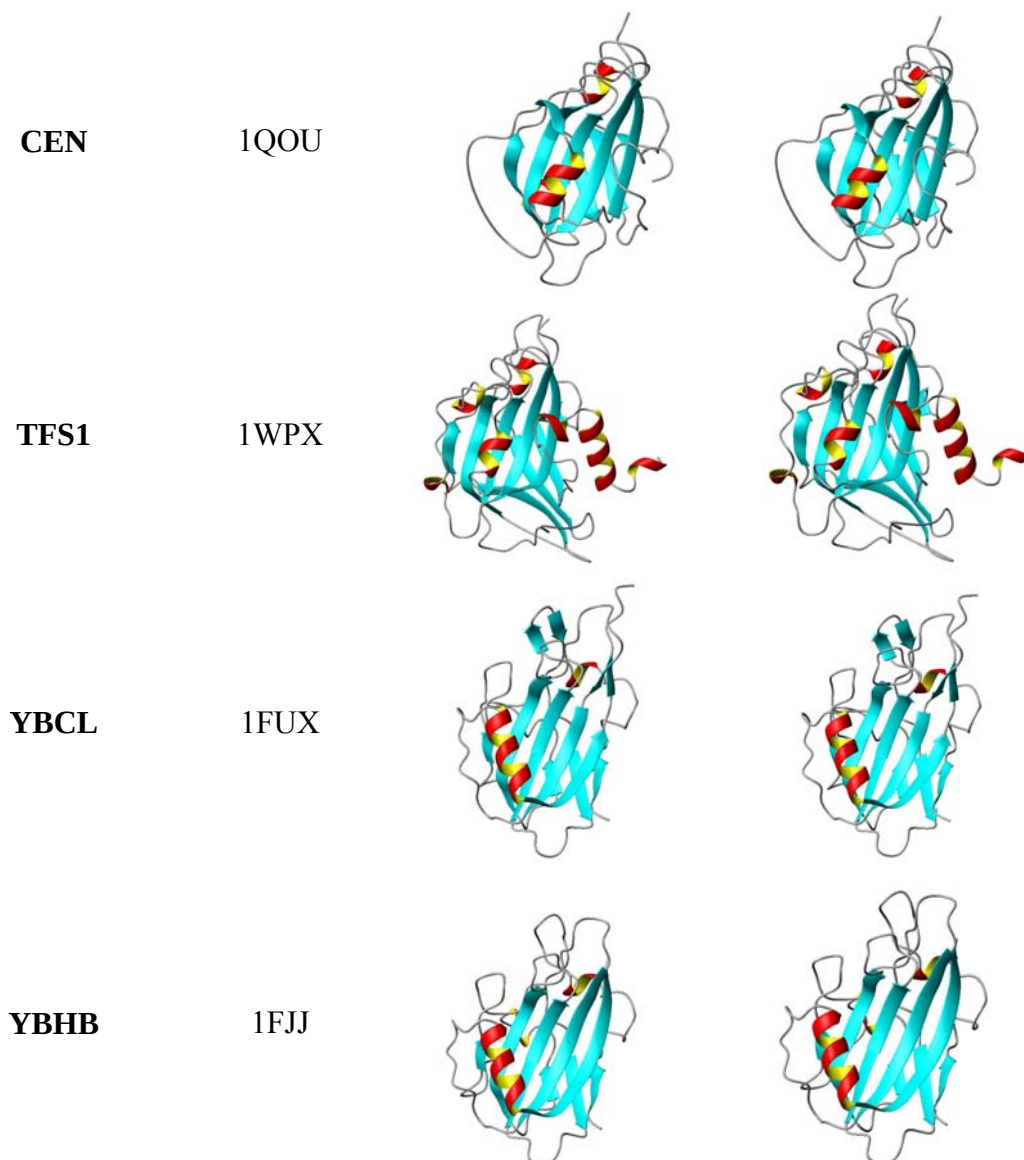


Figure I-3 : Structures des PEBP. Vues en stéréo, les protéines sont orientées par rapport à la PEBP-1 bovine. Figure réalisée avec Molmol (Koradi et al., 1996).

2.6.1. La PEBP bovine

Historiquement, la structure de la PEBP bovine a été la première à être résolue. En 1998 Laurence Serre et ses collaborateurs ont réussi à obtenir des cristaux en présence de trois ligands différents (acétate, phosphate et phosphoryléthanolamine), qui ont permis de déterminer la structure à une résolution de 1,8 Å (Serre *et al.*, 1998).

Elle apparaît composée d'un domaine unique présentant un repliement de type β . Ce domaine est constitué de deux clés grecques formées de feuillettes β antiparallèles (Figure

I-4) et qui interagissent entre elles par une interface hydrophobe. L'extrémité C-terminale correspond quant à elle à un repliement de type $\alpha\beta\alpha$. La structure est aussi caractérisée par de longues boucles qui séparent les différents éléments de structure secondaire. La structure est riche en β -turns et présente trois petites hélices 3_{10} (résidus 5 à 7, 13 à 15 et 96 à 99).

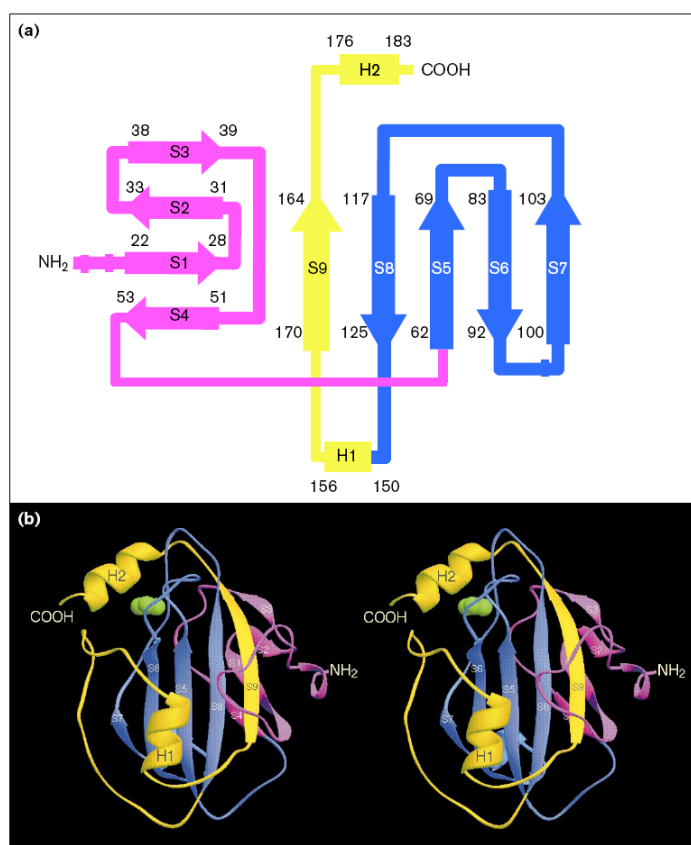


Figure I-4 : Structure de la PEBP-1 bovine d'après Serre (Serre et al., 1998). (a) Diagramme topologique de la PEBP bovine. Les feuillets sont représentés par des flèches, les hélices par des rectangles et les hélices 3_{10} par des barres. Les deux structures en clés grecques sont colorées en magenta et violet. L'extrémité C-terminale (en jaune) adopte un repliement de type $\alpha\beta\alpha$. La numérotation des résidus importants est indiquée. (b) Vue en stéréoscopie sous forme de ruban, le code couleur est conservé. La molécule d'acétate provenant du tampon est représentée en vert.

Fait remarquable, la PEBP-1 bovine possède deux liaisons peptidiques *cis*, au niveau de la proline 73, et entre l'arginine 81 et le glutamate 82. Les liaisons peptidiques *cis* ne faisant pas intervenir une proline sont très rares dans les structures protéiques puisqu'elles interviennent à une fréquence inférieure à 0.03% (statistiques sur les structures cristallographiques de la Protein Data Bank, été 2006). La proline 73 est située au niveau

d'une boucle alors que la deuxième liaison *cis* se situe à l'extrémité d'un brin β . Cette configuration entraîne que les C α de l'arginine 81 et du glutamate 82 se situent à moins de 3 Å l'un de l'autre, et que la liaison peptidique, contrainte, n'est pas tout à fait plane. Les résidus arginine 81 et glutamate 82 sont bien stabilisés dans la structure. En effet, la liaison peptidique peut former deux liaisons hydrogène : avec le carbonyle de la phénylalanine 148 et avec l'amine de la chaîne latérale de la lysine 147. Les chaînes latérales de l'arginine 81 et du glutamate 82 sont aussi stabilisées par un grand nombre de liaisons hydrogène avec les résidus voisins.

La PEBP-1 bovine a été co-cristallisée en présence de trois petits ligands anioniques : un ion acétate, un ion phosphate et un groupement phosphoryléthanolamine (tête polaire de la phosphatidyléthanolamine). Ces ligands ont permis de déterminer la zone de fixation des ligands anioniques, parfois appelée « site actif » par abus de langage en référence aux enzymes. Bien que la structure de la PEBP-1 bovine soit compacte, cette zone forme une légère cavité à la surface de la protéine. Les résidus qui entourent cette cavité intègrent les résidus déterminant la signature de la famille des PEBP. Le site de fixation des ligands anioniques est délimité par deux boucles qui relient des brins β , appelées CR1 (*Connecting Region 1* : région de connection 1) (résidus 68 à 85) et CR2 (résidus 106 à 119), et l'hélice C-terminale. Les deux liaisons peptidiques *cis* appartiennent à la région CR1 et se situent donc à proximité du site de fixation des ligands. Elles semblent jouer un rôle très important pour l'activité de la protéine. Les trois ligands avec lesquels il a été possible de co-cristalliser la PEBP-1 bovine se placent de manière similaire dans le site, en établissant des liaisons hydrogène avec l'aspartate 69, l'histidine 85, la glycine 109 et la tyrosine 119 (Figure I-5).

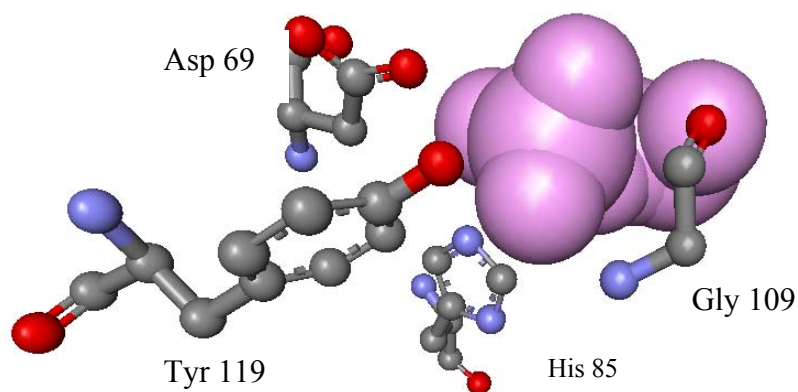


Figure I-5 : Position de la phosphoryléthanolamine dans le « site actif ». L'aspartate 69, l'histidine 85, la glycine 109 et la tyrosine 119 interagissent avec le ligand.

L'analyse de ce site révèle que le volume disponible pour la fixation d'un ligand est faible et, sans déformation de la structure, il ne peut accueillir que la tête polaire anionique des ligands (Figure I-6).

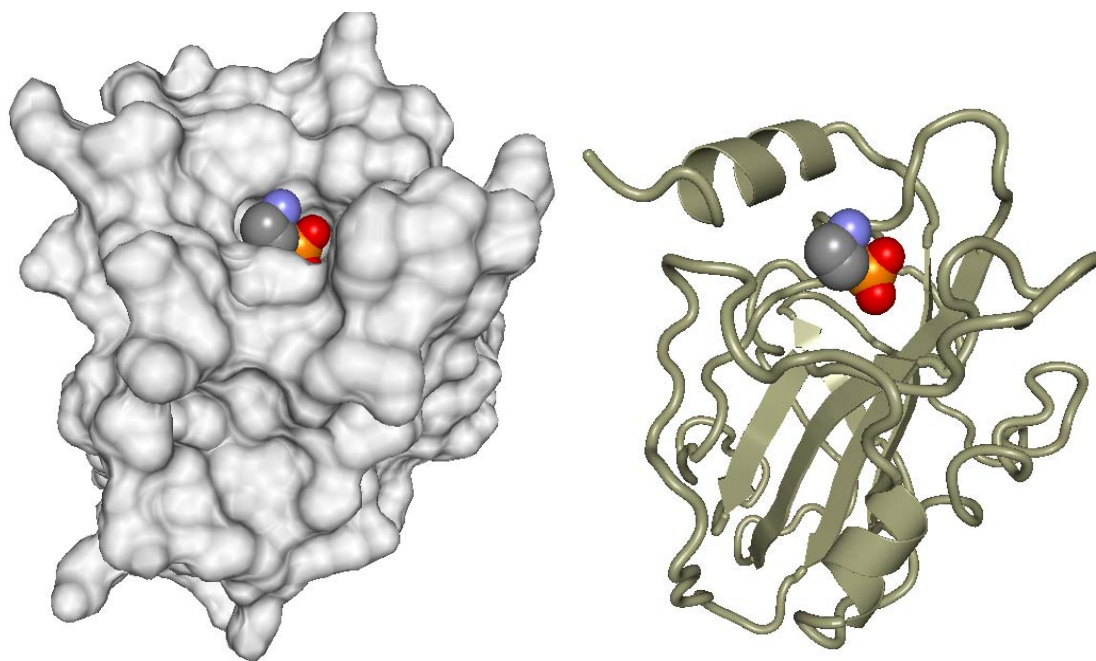


Figure I-6 : Le site de fixation de la phosphoryléthanolamine est étroit. Les 2 représentations sont réalisées avec la même orientation. Figure réalisée avec Ballview.

2.6.2. La PEBP humaine

Les alignements de séquence montrent que la PEBP-1 humaine présente plus de 94 % d'identité avec la PEBP-1 bovine (Figure I-7).

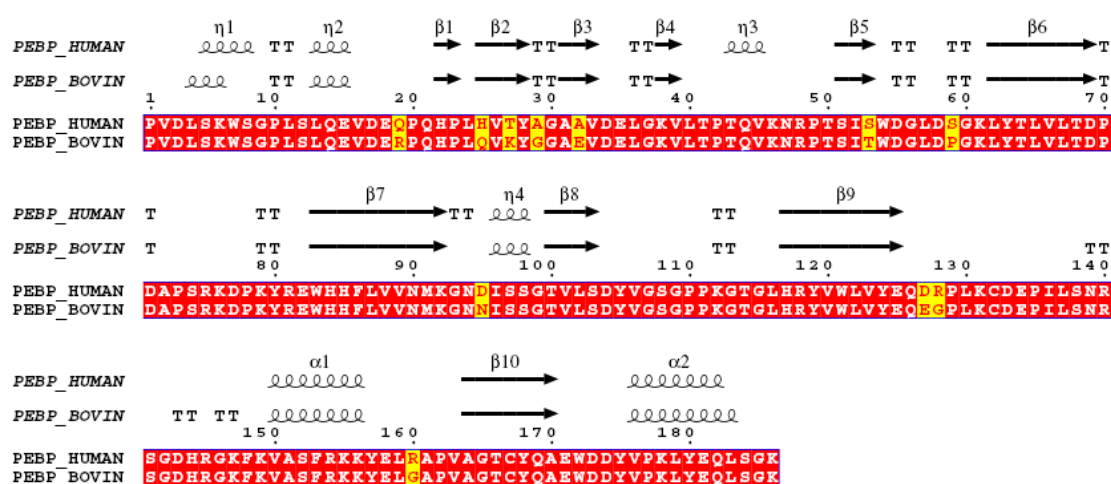


Figure I-7 : Alignement de séquence des PEBP-1 humaine et bovine (CLUSTALW, (Eddy, 1995)). Les cadres rouges indiquent les résidus identiques. Les résidus non conservés, mais similaires, sont indiqués en jaune. Les structures secondaires obtenues sur les modèles cristallographiques sont indiquées.

La structure de la PEBP-1 humaine a été obtenue quasiment au même moment que celle de la PEBP-1 bovine (Banfield *et al.*, 1998) et confirme la grande ressemblance de ces deux protéines : les structures tridimensionnelles sont extrêmement proches, avec un RMSD de 0,38 Å sur le squelette peptidique (Figure I-8).

La co-cristallisation a été réalisée avec un ion cacodylate. Celui-ci se place au niveau du site de fixation des ligands anioniques de la même manière que les ions acétate, phosphate et phosphoryléthanolamine au niveau de la PEBP-1 bovine, en établissant 4 liaisons hydrogène.

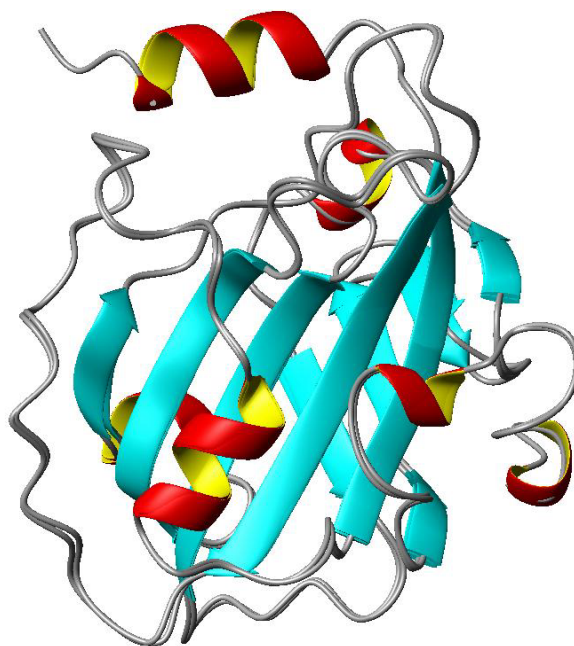


Figure I-8 : Superposition des PEBP-1 humaine et bovine.

2.6.3. Les PEBP des autres espèces

L'observation des différentes structures cristallographiques disponibles permet de mieux caractériser les déterminants structuraux qui définissent la famille des PEBP. Comme nous venons de le voir, les PEBP-1 bovine et humaine sont extrêmement proches. Les protéines des autres espèces présentent des similarités structurales fortes, notamment la présence de brins β assemblés sous la forme de deux clés grecques. Cependant l'orientation de ces deux clés grecques varie légèrement d'une espèce à une autre. De plus on observe des différences significatives au niveau des extrémités ainsi qu'au niveau des positions adoptées par les nombreuses boucles présentes entre les structures secondaires. Une comparaison des différentes structures avec la PEBP-1 bovine est présentée dans ce qui suit.

2.6.3.a. La PEBP-2 de souris

La structure de la PEBP-2 de souris est disponible dans la Protein Data Bank (code 1KN3, (Simister *et al.*, 2002))

L'alignement de la séquence de la protéine PEBP-2 de souris avec celle de la PEBP-1 bovine indique 83 % d'identité (Figure I-9). Les séquences sont donc très proches.

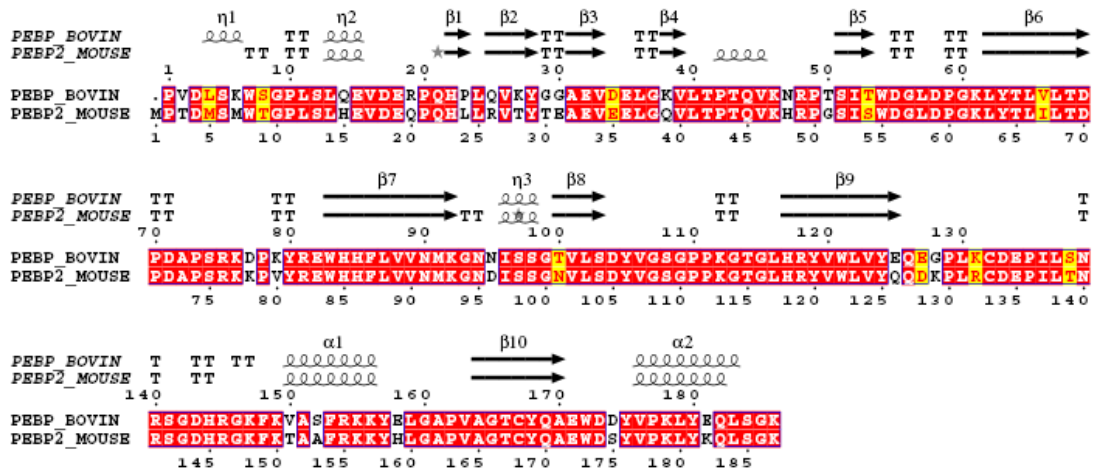


Figure I-9 : Alignement de séquence des PEBP-2 de souris et PEBP-1 bovine (CLUSTALW, (Eddy, 1995)). Les cadres rouges indiquent les résidus identiques ; les résidus non conservés mais similaires sont indiqués en jaune. Les structures secondaires obtenues sur les modèles cristallographiques sont indiquées. (Figure réalisée avec le logiciel Esript).

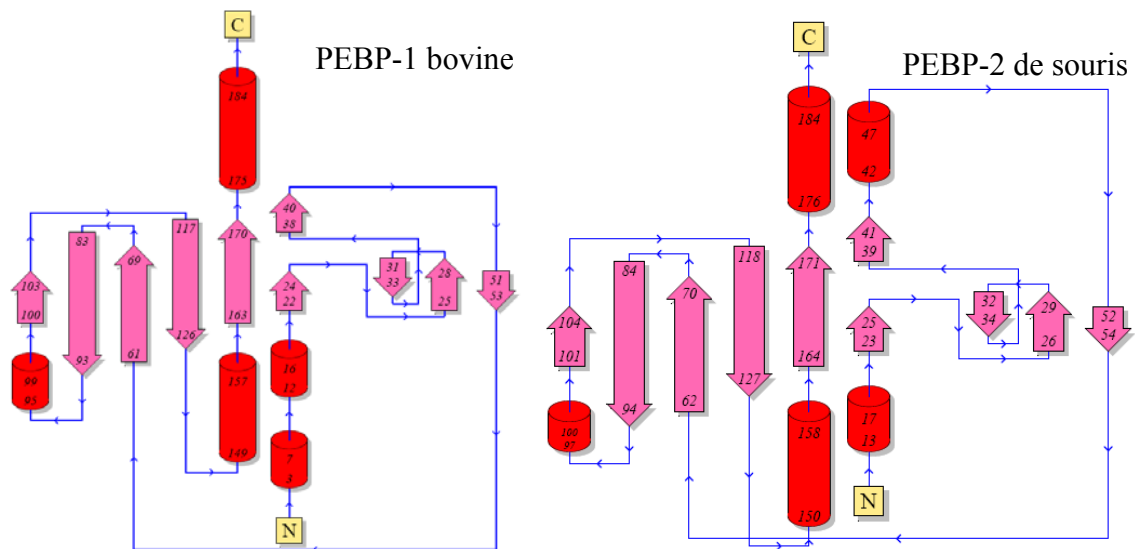


Figure I-10 : Diagramme de topologie des protéines PEBP-1 bovine (à gauche) et PEBP-2 de souris (à droite). (Figure réalisée avec le logiciel PDBsum (Laskowski et al., 2005))

La superposition des structures de ces deux protéines est très bonne avec un RMSD de 0,48 Å pour 180 Cα alignés. Les structures secondaires sont particulièrement conservées à l'exception de quelques turns, par exemple au niveau de l'extrémité N-terminale (Figure I-10). La détection d'une courte hélice allant des résidus 43 à 46 de la PEBP de souris est essentiellement liée au cut-off de détection du logiciel utilisé mais ne constitue pas une différence structurale par rapport à la PEBP-1 bovine. L'orientation des deux clés grecques est parfaitement conservée.

Bien qu'il n'y ait pas de ligand au niveau du site de fixation putatif des ligands anioniques de mPEBP-2, les résidus concernés sont conservés et parfaitement superposables à ceux de la PEBP-1 bovine en complexe avec une molécule d'acétate.

2.6.3.b. La protéine CEN (1QOU)

L'alignement de la séquence de la protéine CEN d'*Antirrhinum centroradialis* (code PDB 1QOU, (Banfield et Brady, 2000)) avec celle de la PEBP-1 bovine indique 20 % d'identité (Figure I-11). Ces deux protéines présentent donc des différences de séquence significatives.

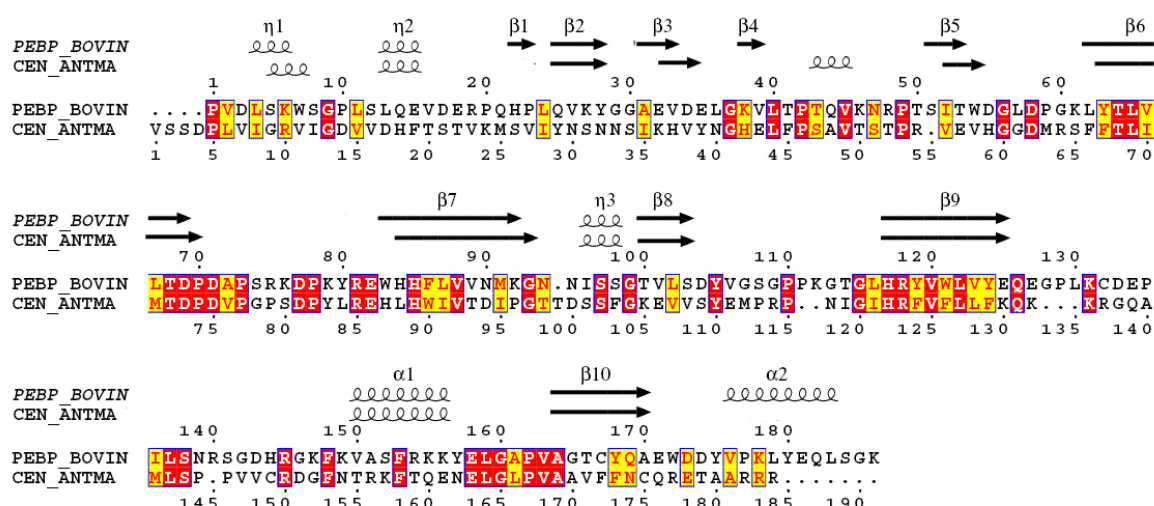


Figure I-11 : Alignement des séquences de la protéine CEN et de la PEBP-1 bovine (CLUSTALW, (Eddy, 1995)). Les cadres rouges indiquent les résidus identiques ; les résidus non conservés mais similaires sont indiqués en jaune. Les structures secondaires obtenues sur les modèles cristallographiques sont indiquées. (Figure réalisée avec le logiciel Esript).

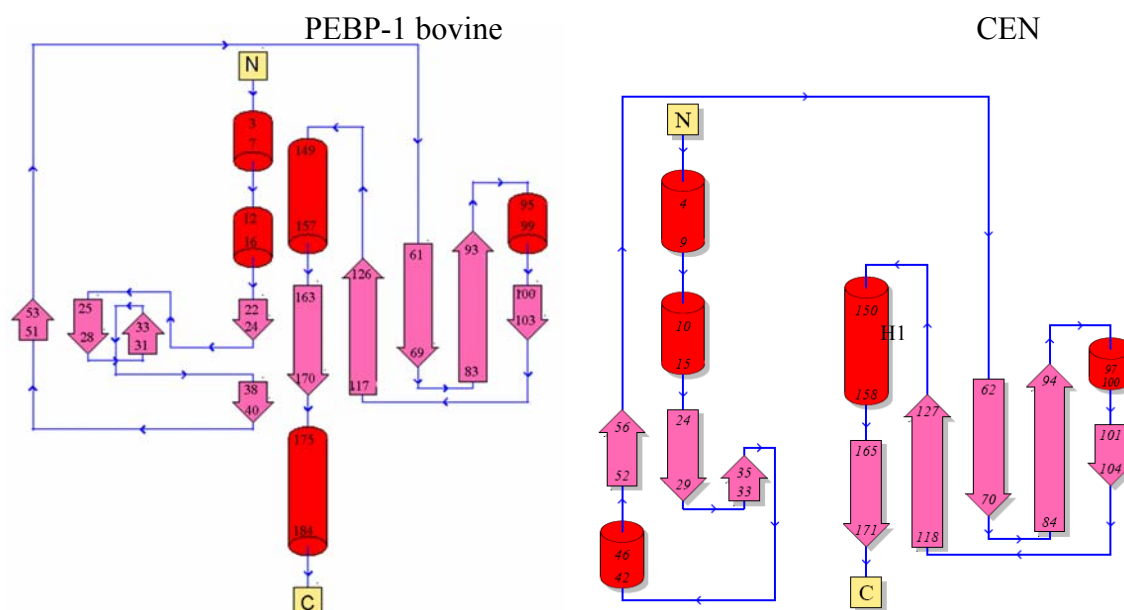


Figure I-12 : Diagramme de topologie des protéines PEBP-1 bovine (à gauche) et CEN (à droite). (Figure réalisée avec le logiciel PDBsum (Laskowski et al., 2005)).

La superposition des structures conduit à un RMSD de 1,41 Å pour 118 C α superposés.

La structure de la protéine CEN a été déterminée par remplacement moléculaire à partir celle de la protéine hPEBP-1. Bien que sa séquence soit plus éloignée, la protéine CEN présente un cœur β conservé et les deux clés grecques habituelles. Toutefois, certains résidus n'ont pas pu être résolus sur la carte de densité électronique : N37, située au niveau d'un turn, et les résidus 130 à 143 qui constituent une grande boucle à la surface de la protéine. Les différences majeures proviennent d'une délétion correspondant aux résidus 111 et 112 de la PEBP-1 bovine qui se trouvent dans une boucle située à proximité du site de fixation des ligands anioniques ; la configuration de cette boucle diffère donc pour CEN. Par ailleurs, la topologie de la clef grecque N-terminale de CEN diffère de celle de la PEBP bovine (Figure I-12). D'autre part, l'hélice H1 est présente mais l'extrémité C-terminale n'a pas été résolue au-delà du résidu thréonine 175 (pour un total de 181 résidus). Nous ne disposons donc pas d'information concernant une éventuelle hélice terminale.

Bien qu'aucun ligand n'ait été retrouvé dans la cavité et que la séquence ne soit pas conservée dans cette zone, les positions des résidus du site putatif de fixation des ligands

anioniques sont extrêmement proches dans les deux protéines, et on peut noter les équivalences (Tableau I-2) :

PEBP-1 bovine	W83	H85	G107	S108	G109	D69	Y119
CEN	H84	H86	M108	P109	R110	D70	F120

Tableau I-2 : Résidus équivalents du site putatif de fixation des ligands anioniques entre la PEBP-1 bovine et la protéine CEN.

2.6.3.c. Les protéines TFL1 (1WKO) et FT (1WKP) d'*Arabidopsis thaliana*

L'alignement des séquences des protéines TFL1 et FT (code PDB 1WKO et 1WKP, (Ahn *et al.*, 2006)) avec celle de la PEBP-1 bovine indique 25 et 28 % d'identité respectivement (Figure I-13). TFL1 et FT présentent 56% d'identité entre elles.

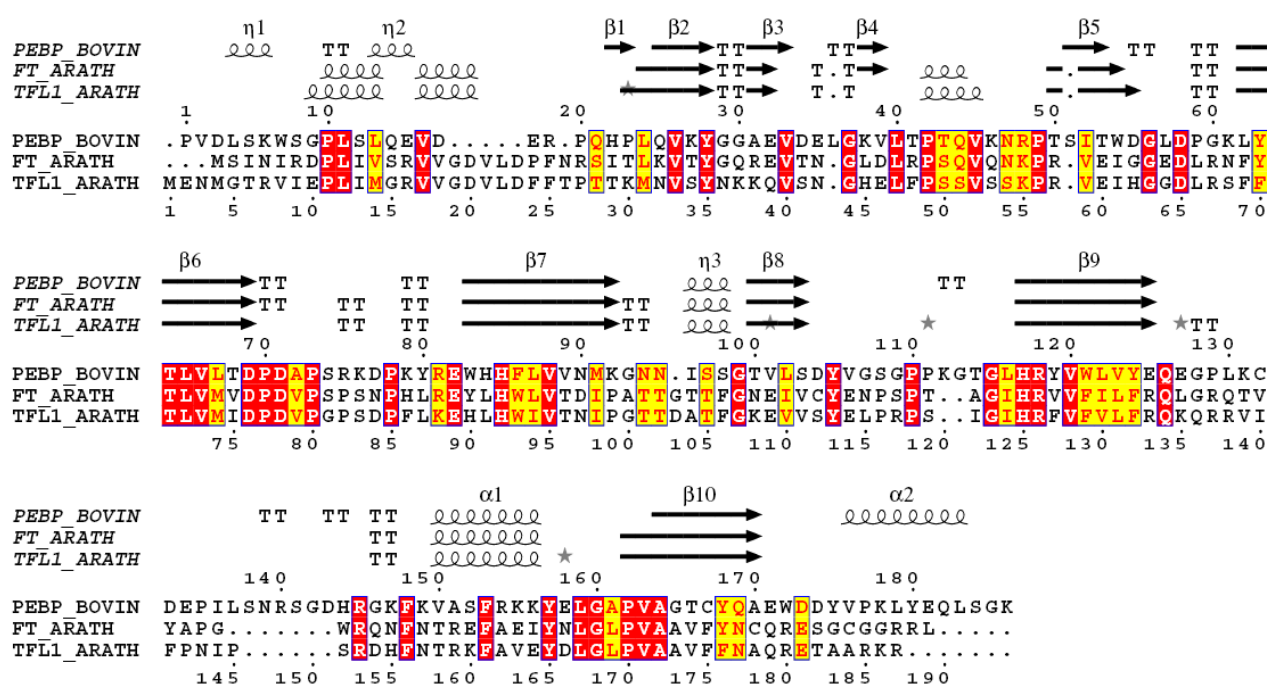


Figure I-13 : Alignement de séquence de TFL1 (1WKO) et FT (1WKP) d'*Arabidopsis thaliana* avec la PEBP-1 bovine (CLUSTALW, (Eddy, 1995)). Les cadres rouges indiquent les résidus identiques ; les résidus non conservés mais similaires sont indiqués en jaune. Les structures secondaires obtenues sur les modèles cristallographiques sont indiquées. (Figure réalisée avec le logiciel Esript).

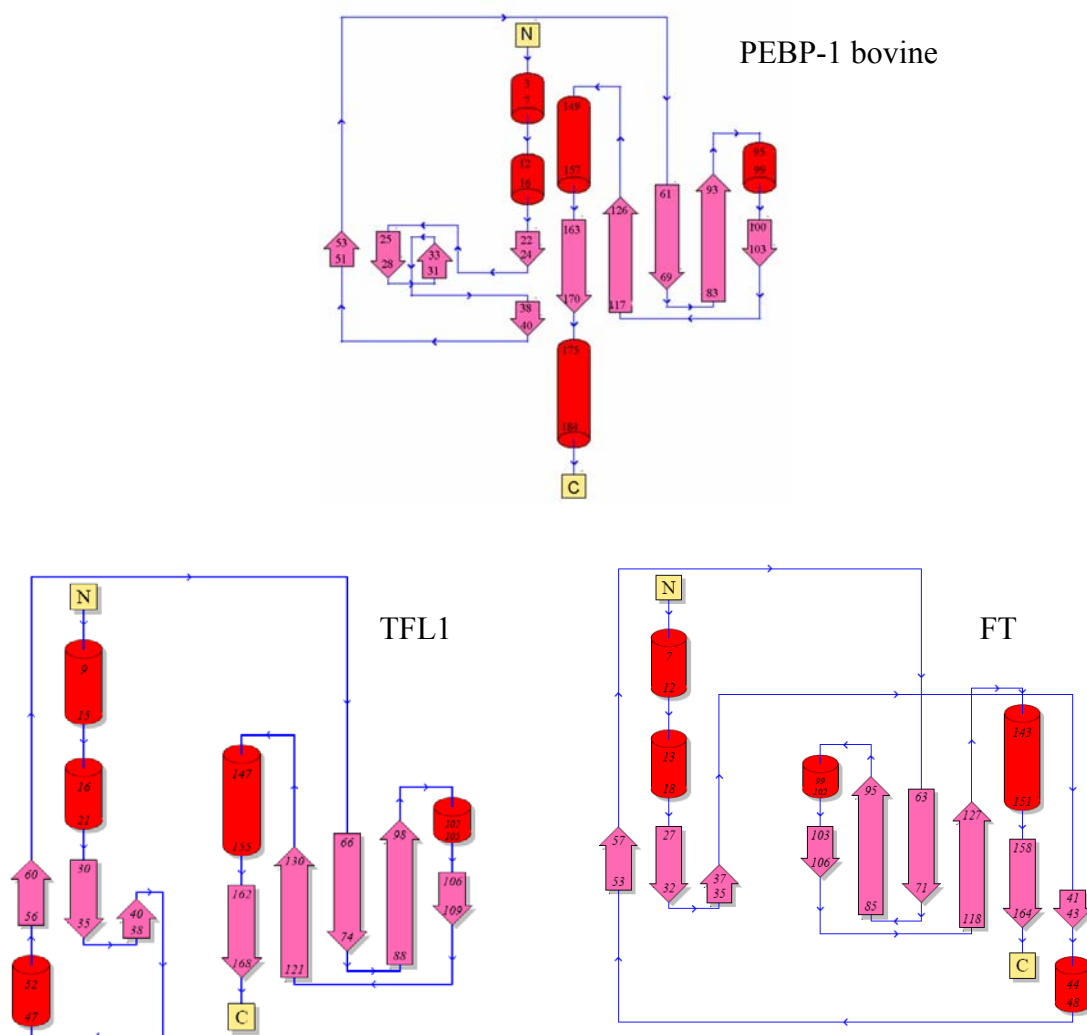


Figure I-14 : Diagramme de topologie des protéines PEBP-1 bovine (page 38), TFL1 (en à gauche) et FT (à droite). (Figure réalisée avec le logiciel PDBsum (Laskowski et al., 2005))

Les structures des protéines TFL1 et FT sont globalement très proches l'une de l'autre (RMSD de 0,78 Å / 158 Cα alignés). Elles sont constituées du cœur classique organisé en deux clés grecques. Mis à part les extrémités N et C-terminales qui sont de longueurs légèrement différentes, les différences sont minimales et résident au niveau du positionnement des boucles correspondant aux résidus 31 à 36, 57 à 59 et 127 à 139 (numérotation pour la protéine FT).

La superposition des structures de ces deux protéines avec celle de la PEBP bovine est bonne, avec des RMSD de 1,14 Å pour TFL1 (127 Cα alignés) et 1,06 Å pour FT (132 Cα alignés). La topologie de la première clé grecque diffère de celle de la PEBP bovine (Figure I-14) et adopte une configuration plus proche de celle de la protéine CEN. L'hélice H1 est conservée par rapport à la protéine PEBP-1 bovine mais la boucle qui précède cette hélice (des résidus 135 à 144, pour la protéine bovine) est absente chez TFL1 et FT, ainsi que l'hélice C-terminale. Il faut remarquer que chez la PEBP-1 bovine ces deux dernières régions qui sont absentes dans les protéines d'*Arabidopsis* interagissent entre elles grâce à des liaisons hydrogène. Leur interaction permet de moduler directement l'accès au site putatif de fixation des ligands anioniques. Ce site semble donc plus accessible pour les protéines TFL1 et FT.

Pour les deux protéines, un ion sulfate a été retrouvé au niveau du site putatif de fixation des ligands anioniques. Les résidus du site adoptent une configuration tout à fait comparable à celle des résidus de la PEBP-1 bovine et on peut noter les équivalences (Tableau I-3) :

PEBP bovine	D69	H85	S108	G109	Y119
TFL1	D74	H90	P114	R115	F123
FT	D73	H87	P111	S112	V120

Tableau I-3 : Résidus équivalents du site putatif de fixation des ligands anioniques entre la PEBP-1 bovine et les protéines TFL1 et FT.

2.6.3.d. La protéine TFS1 de la levure *Saccharomyces cerevisiae*

La structure de la protéine TFS1 de *Saccharomyces cerevisiae* a été résolue en complexe avec la carboxypeptidase Y (code 1WPX, (Mima *et al.*, 2005)).

L'alignement de la séquence de la protéine TFS1 avec la PEBP-1 bovine indique 24 % d'identité (Figure I-15). Ce pourcentage d'identité est calculé par rapport au nombre de résidus alignés, il ne prend donc pas en compte le fait que la protéine TFS1 est plus longue que la protéine bovine (219 résidus contre 186). Ces deux protéines sont donc significativement différentes en terme de séquence.

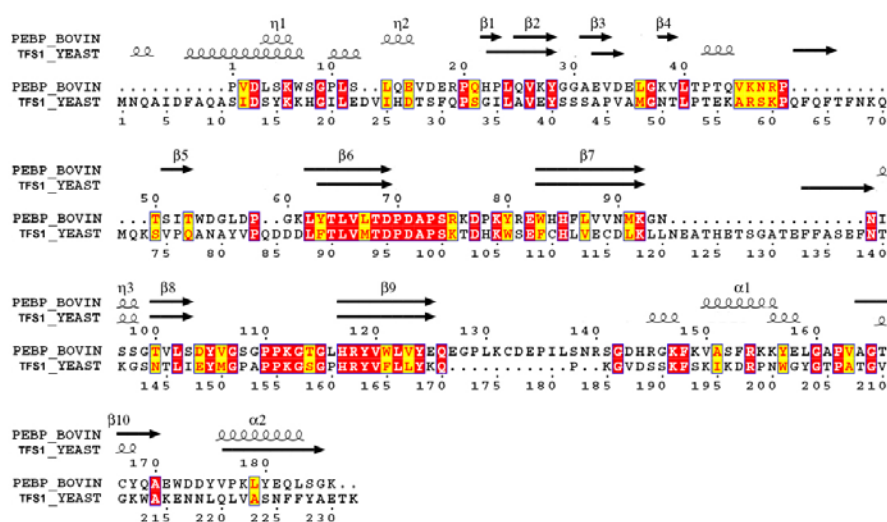


Figure I-15 : Alignement des séquences de la protéine TFS1 et de la PEBP-1 bovine (CLUSTALW, (Eddy, 1995)). Les cadres rouges indiquent les résidus identiques ; les résidus non conservés mais similaires sont indiqués en jaune. Les structures secondaires obtenues sur les modèles cristallographiques sont indiquées. (Figure réalisée avec le logiciel Esript).

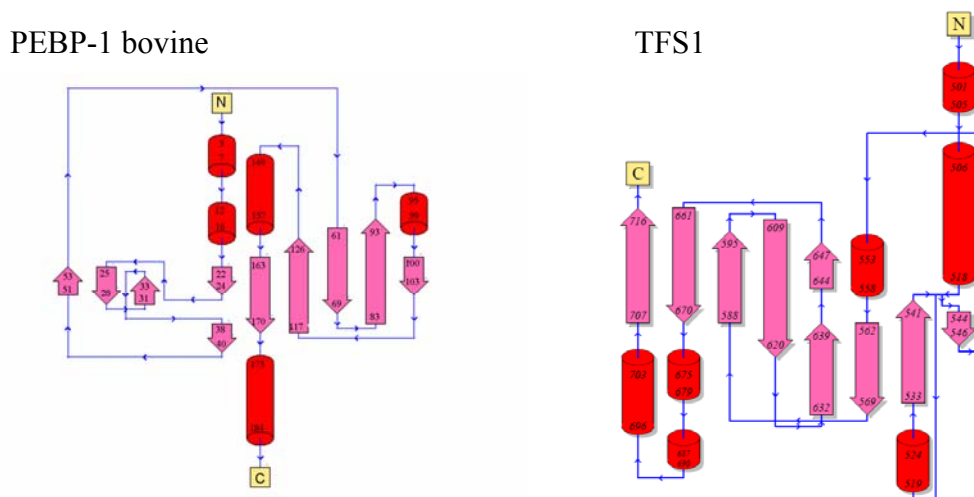


Figure I-16 : Diagramme de topologie des protéines PEBP-1 bovine (à gauche) et TFS1 (à droite ; il faut retirer 500 à la numérotation indiquée pour obtenir la numérotation habituelle). (Figure réalisée avec le logiciel PDBsum (Laskowski et al., 2005)).

La superposition de la PEBP-1 bovine et de TFS1 conduit à un RMSD de 1,12 Å pour 129 Cα alignés. Le repliement de TFS1 présente le cœur central caractéristique de la famille des PEBP mais l'arrangement des deux clés grecques y est plus compliqué : leur orientation varie sensiblement et elles s'entremêlent (Figure I-16). L'hélice H1 est

présente mais plus courte que pour la PEBP-1 bovine. La structure ne présente pas d'hélice au niveau de l'extrémité C-terminale. L'extrémité N-terminale de TFS1 est légèrement plus longue que pour la PEBP-1 bovine et présente un petit fragment d'hélice 3_{10} suivi d'une hélice α . Cette extrémité inhibe la carboxypeptidase Y en s'intégrant dans son site actif (cf infra). La protéine TFS1 est caractérisée par la présence de 2 insertions de grande taille que l'on ne retrouve pas dans les autres PEBP : résidus 62 à 74 et 121 à 138 (numérotation pour TFS1). Ces insertions viennent allonger les brins de la clé grecque la plus superficielle. Les fragments allant du E122 à A130 et de la K73 à A78 n'ont pas pu être résolus dans la carte de densité électronique. Il faut ensuite remarquer une délétion dans la protéine TFS1 correspondant aux résidus E127 jusqu'à R140 de la PEBP-1 bovine. Cette zone interagit directement avec l'hélice C-terminale chez la protéine de bœuf. L'absence de l'hélice C-terminale chez TFS1 pourrait être corrélée à cette délétion, comme pour les PEBP d'*Arabidopsis*.

Un ion sulfate est présent au niveau du site putatif de fixation des ligands anioniques. Les résidus à proximité adoptent une structure proche de celle des résidus du « site actif » de la PEBP-1 bovine. On peut donc définir les résidus équivalents :

PEBP-1 bovine	H85	D69	S108	G109	Y119
TFS1	H111	D95	P152	A153	Y163

Tableau I-4 : Résidus équivalents du site putatif de fixation des ligands anioniques entre la PEBP-1 bovine et la protéine TFS1.

2.6.3.e. Les protéines YBHB et YBCL d'*E. coli*

Les structures des protéines YBHB et YBCL sont disponibles dans la Protein Data Bank.

L'alignement des séquences des protéines YBHB et YBCL d'*Escherichia coli* (codes PDB 1FJJ et 1FUX, (Ahn *et al.*, 2006)) avec la PEBP-1 bovine indique 16 et 14 % d'identité respectivement (Figure I-17). YBHB et YBCL présentent 50 % d'identité entre elles.

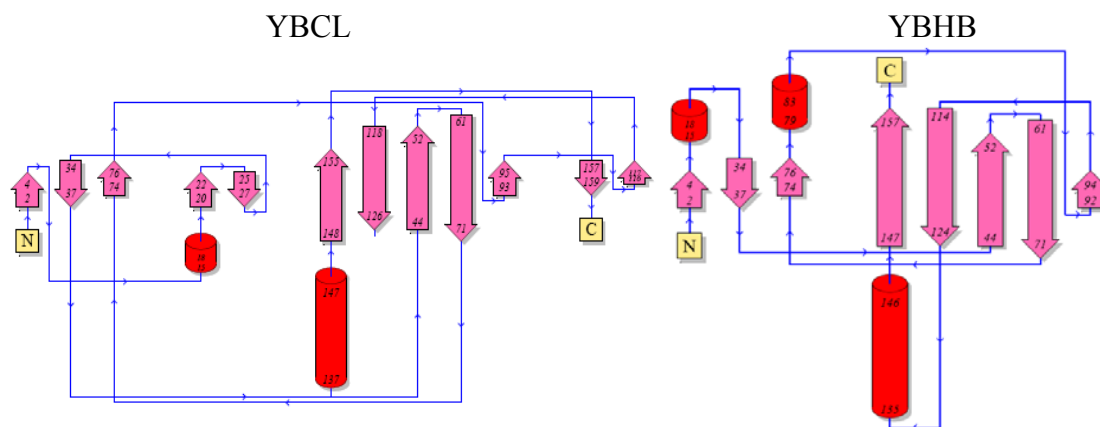


Figure I-18 : Diagramme de topologie des protéines PEBP-1 bovine (page précédente), YBCL (en bas à gauche) et YBHB (en bas à droite). (Figure réalisée avec le logiciel PDBsum (Laskowski et al., 2005))

La protéine YBCL présente un peptide signal en N-terminal et un site de coupure entre l'alanine 21 et la phénylalanine 22. Il a été montré que cette séquence signal était coupée *in vivo*. La structure cristallographique de YBCL a été réalisée sur la version tronquée de la protéine. La numérotation des résidus que nous utiliserons est donc identique pour les 2 protéines d'*Escherichia coli*.

Les structures des deux protéines sont proches et présentent les caractéristiques générales communes à toutes les protéines de la famille des PEBP (Figure I-18). Ainsi on retrouve une organisation composée des deux clés grecques et de la boucle C-terminale de type $\alpha\beta$. Cependant la clé grecque N-terminale est incomplète et on observe la formation d'un petit feuillet supplémentaire entre les résidus 21 et 27 pour YBCL. La superposition des deux structures conduit à un RMSD de 0,83 Å pour 149 C α alignés. Les structures secondaires sont similaires et les différences sont observées au niveau d'insertions ou de boucles. Ainsi, la plus grande déviation est observée pour les résidus 21 à 27 qui sont situés à proximité du site putatif de liaison des ligands anioniques, et pour les résidus 80 à 92.

La superposition de ces deux structures avec celle de la PEBP-1 bovine conduit à un RMSD de 1,38 Å pour YBHB (84 C α alignés) et 1,39 Å pour YBCL (82 C α alignés). La superposition est particulièrement bonne au niveau des structures secondaires du cœur de la protéine. Les différences principales se situent au niveau des boucles et aux extrémités des protéines. Ainsi, aucune des deux PEBP de coli ne présente d'hélice C-terminale.

Dans la PEBP bovine, cette hélice semble pouvoir moduler l'accessibilité des substrats au niveau du site putatif de fixation des ligands anioniques. Chez *coli* cette fonction pourrait être assurée par la boucle définie par les résidus 21 à 27 qui vient se placer juste à proximité du site.

Une molécule d'HEPES a été retrouvée au niveau du site de fixation des ligands anioniques de la protéine YBHB alors qu'aucune molécule n'a été observée dans le site de la protéine YBCL. Cependant, la structure du « site actif » est parfaitement conservée entre les deux protéines, et parfaitement superposable à celle du site de la PEBP-1 bovine. La parfaite correspondance des structures tridimensionnelles dans cette région, ainsi que la position adoptée par le ligand au niveau de la protéine YBHB, permet de définir les résidus impliqués dans l'interaction :

PEBP-1 bovine	H85	D69	S108	G109	Y119
YBCL	H63	D52	A107	C108	Y118
YBHB	H63	D52	A106	A107	Y116

Tableau I-5 : Résidus équivalents du site putatif de fixation des ligands anioniques entre la PEBP-1 bovine et les protéines YBCL et YBHB.

2.6.4. PEBP et liaisons cis

Les PEBP ont la particularité de présenter plusieurs liaisons peptidiques en *cis* qui ont été retrouvées au niveau de toutes les protéines dont la structure est connue (sauf chez les protéines YBHB et YBCL d'*Escherichia coli* qui n'en présentent aucune) :

- Pour une proline qui appartient à la séquence conservée définissant l'appartenance d'une protéine à la famille des PEBP : DPDxP. Ce motif se situe juste à proximité du site de fixation de ligands anioniques. Il a été retrouvé dans toutes les PEBP non bactériennes.
- entre les résidus conservés R81 et E82 de la PEBP-1 bovine. La conservation stricte de cette orientation non favorable souligne un rôle structural important. Ce motif est situé juste à l'entrée du site de fixation des ligands anioniques. La chaîne latérale du glutamate en position 82 adopte une conformation en *cis* en agissant comme un accepteur de proton au niveau d'une liaison hydrogène avec l'amide du squelette peptidique du résidu 150. Ce motif est retrouvé dès lors que la protéine

présente une liaison Arg-Glu 7 résidus après la proline en *cis* du motif conservé. On retrouve donc cette particularité structurale au niveau des protéines PEBP-1 humaine et bovine, murine, CEN, TFL1 et FT.

La conservation et la localisation de ces liaisons peptidiques en configuration *cis* à proximité de la zone putative de fixation des ligands anioniques laisse entrevoir un rôle majeur de ces liaisons dans les relations structure/fonction au sein de la famille des PEBP.

2.6.5. PEBP et dimères

Les structures cristallographiques d'un certain nombre de PEBP ont été résolues sous forme d'homodimères dans l'unité asymétrique: les PEBP-1 bovine (1B7A) et humaine (1BD9 et 1BEH), les protéines d'*Escherichia coli* YBHB (1FJJ) et YBCL (1FUX), la protéine CEN d'*Antirrhinum centroradialis* (1QOU), les protéines TFL1 (1WKO) et FT d'*Arabidopsis thaliana* (1WKP). Par opposition, la PEBP-1 de veau (1A44), la PEBP-2 de souris (1KN3) et la protéine TFS1 de la levure *Saccharomyces cerevisiae* (1WPX), ont été résolues sous forme de monomères.

Parmi les protéines retrouvées dans le cristal sous forme de dimères, seules les protéines de coli YBHB et YBCL semblent capables de former des homodimères en solution (Serre *et al.*, 2001). De plus, la formation du dimère conduit à la formation d'une nouvelle configuration de la zone de fixation des ligands anioniques, à l'interface entre les 2 monomères (Figure I-19) (Serre *et al.*, 2001). Les résidus impliqués dans l'interaction et la structure de l'interface ne sont pas conservés entre procaryotes et eucaryotes.

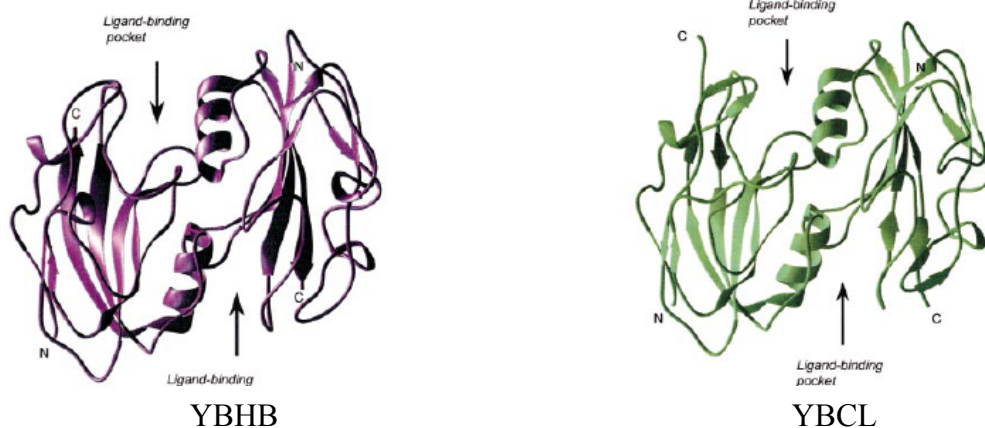


Figure I-19 : Configuration des homodimères de YBHB et YBCL. Le site de fixation des ligands anioniques profite de l'interface créée entre les deux monomères. (image d'après (Serre et al., 2001))

3. Activités moléculaires

En rapport avec les caractéristiques moléculaires primaires des PEBP, un certain nombre d'activités moléculaires ont été rapportées. Ces activités peuvent être regroupées en deux catégories. D'abord une capacité à interagir, moduler et réguler l'activité de différentes enzymes et protéines dans la cellule. La deuxième catégorie d'activité moléculaire est évidemment liée à la capacité de fixer des ligands anioniques comme la phosphatidyléthanolamine ou les dérivés de morphine mais, bien que des travaux récents apportent de nouvelles pistes de réflexion, cette activité n'est pas encore clairement établie.

3.1. Inhibition de kinases

Différentes études ont montré que les PEBP sont capables d'inhiber et donc de réguler un certain nombre de kinases (Banfield et Brady, 2000). Par exemple, la PEBP-1 humaine est un inhibiteur efficace de plusieurs kinases intervenant dans diverses voies de transduction du signal : Raf-1, GRK2 (Keller *et al.*, 2005) ou TAK-1, NIK, IKK α , IKK β (Yeung *et al.*, 2001). Ainsi la protéine hPEBP-1 est aussi appelée Raf Kinase Inhibitor Protein (RKIP) (Yeung *et al.*, 1999). Des activités similaires ont été retrouvées dans différents organismes, par exemple pour les PEBP-2 de souris (Gibbons *et al.*, 2005) et de

rat (Trakul et Rosner, 2005), pour la protéine CEN de la plante *Antirrhinum centroradialis* (Banfield et Brady, 2000). Enfin la protéine TFS1 de levure inhibe entre autres la kinase Ira2p (Chautard *et al.*, 2004). Les détails moléculaires et atomiques de l'interaction ne sont actuellement pas connus, mais il semble que le site de fixation des ligands anioniques intervienne (Hélène Benedetti, communication personnelle).

3.2. Inhibition de protéases

Les PEBP sont capables d'inhiber de manière spécifique un certain nombre de protéases à sérine. Par exemple, la PEBP-3 de souris inhibe, avec une constante d'inhibition de l'ordre du micro-molaire, la thrombine, la neuropsine, la chymotrypsine mais n'inhibe pas la trypsine, l'activateur du plasminogène, ni l'élastase (Hengst *et al.*, 2001).

La protéine de levure TFS1 est quant à elle capable d'inhiber la carboxypeptidase Y, avec une constante d'inhibition de 1,8 nM (Mima *et al.*, 2005). Il a été possible de co-cristalliser la carboxypeptidase Y avec la protéine TFS1 (Mima *et al.*, 2005). La structure du complexe montre que l'interaction a lieu au niveau des 7 résidus N-terminaux de TFS1, la méthionine N-terminale acétylée venant se placer au cœur même du site actif de la carboxypeptidase Y. Si la méthionine N-terminale n'est pas acétylée, la constante d'inhibition passe de l'ordre du nano-molaire au micro-molaire. Le site actif de TFS1 se situe à proximité de l'interface intermoléculaire, et un ion sulfate y a été retrouvé (Figure I-20).

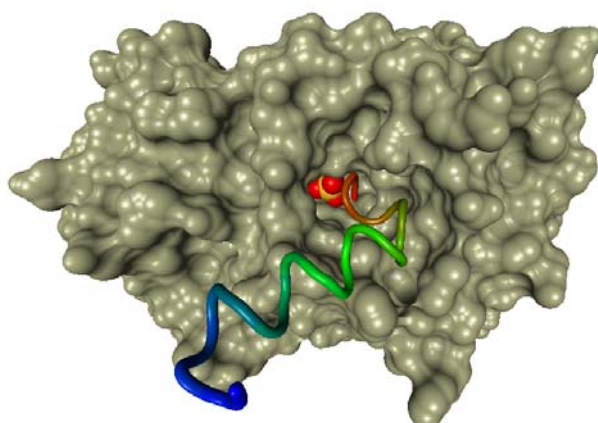


Figure I-20 : Le fragment N-terminal de TFS1 vient se glisser à l'intérieur du site actif de la carboxypeptidase Y. Le fragment de TFS1 est coloré du rouge vers le bleu, le

rouge correspondant au résidu N-terminal. On aperçoit un ion sulfate enfoui au fond de la cavité de la carboxypeptidase Y.

3.3. Activité liée à la fixation de ligands anioniques

Il est apparu rapidement que la famille des PEBP était caractérisée par la capacité de fixer des ligands anioniques, que ce soit des phospholipides, des nucléotides, des dérivés du cholestérol, des dérivés d'hormones, etc. On pense que ces protéines sont impliquées dans le métabolisme lipidique, l'organisation des lipides à l'intérieur des membranes et dans les remodelages membranaires. Malgré le travail de plusieurs équipes dans le monde entier, aucune activité liée à la capacité de fixer des ligands anioniques n'a pour l'instant été mise en évidence. Cependant il a été récemment découvert que la PEBP-1 humaine était capable de fixer *in vivo* des dérivés de la morphine sécrétés dans les granules des cellules chromaffines de la médullosurrénale après stimulation (Goumon *et al.*, 2006). Une fois fixé le dérivé de morphine est protégé contre la dégradation, ce qui lui permet de conserver son pouvoir analgésique. La protection liée à cette capacité de fixation des ligands anioniques constitue peut-être une nouvelle activité de la PEBP-1 humaine.

4. Rôles cellulaires et physiologiques

Les activités moléculaires primaires des PEBP (inhibition de kinases, inhibition de protéases, etc.) s'expriment à différents niveaux cellulaires et physiologiques. Ainsi, ces activités participent à la régulation de plusieurs voies de transduction du signal fondamentales pour les cellules. On retrouve aussi les PEBP impliquées dans un grand nombre de processus physiologiques complexes comme la spermatogénèse (Turner *et al.*, 1996), le contrôle de l'activité cardiaque (Goumon *et al.*, 2004), le développement du cerveau chez les mammifères (Yuasa *et al.*, 2001), ou l'inflorescence chez les plantes (Banfield *et Brady*, 2000).

4.1. Rôles cellulaires : les voies de transduction du signal

L'implication des PEBP dans les voies de transduction du signal a surtout été étudiée pour la PEBP-1 humaine. Cependant les conclusions obtenues semblent être transposables à de nombreuses espèces.

4.1.1. RKIP intervient dans plusieurs voies de transduction du signal.

Comme nous l'avons vu, la protéine hPEBP-1 intervient dans la voie des MAP-kinases, dans la voie NF κ B et dans le contrôle des protéines G hétérotrimériques, sa capacité à d'inhiber la protéine Raf de la voie des MAP-kinases lui ayant valu l'autre nom de Raf Kinase Inhibitor Protein ou RKIP (Yeung *et al.*, 1999).

4.1.1.a. RKIP inhibe la voie des MAP-kinases Raf/MEK/ERK

Les facteurs de croissance comme l'EGF (Epidermal Growth Factor), le PDGF (Platelet Derived Growth Factor), le NGF (Nerve Growth Factor), l'insuline, les hormones de croissance ou les esters de phorbol sont capables de déclencher des réponses intracellulaires en se fixant sur des récepteurs, notamment les récepteurs à activité Tyrosine Kinase : les RTK. La voie des MAP-kinases Raf-1/MEK/ERK (Figure I-21) est une voie de transduction du signal ubiquitaire qui permet la transmission des signaux de prolifération et de différenciation cellulaire depuis les RTK situés au niveau de la membrane plasmique jusqu'au noyau (Elion, 1998). Cette cascade de kinases est organisée autour de la protéine Ras (Moodie *et al.*, 1993). Ras est une petite protéine G, elle peut fixer le GDP et le GTP et possède une activité GTPase. Son activité est reliée à la transduction du signal depuis les RTK, mais aussi depuis les récepteurs couplés aux protéines G (RCPG) et depuis les canaux ioniques Ca²⁺-dépendants. Ras est activée par des RTK ayant fixé un facteur de croissance, par l'intermédiaire de la protéine SOS. Une fois activée, la protéine Ras échange un GDP contre un GTP et se lie à la kinase Raf-1. Cette liaison induit l'activation de Raf-1 par une kinase présente au niveau de la membrane plasmique mais encore non identifiée. Raf-1 activée phosphoryle et active MEK, une kinase qui phosphoryle à son tour la protéine ERK (Marais et Marshall, 1996). ERK activée phosphoryle et active à son tour le facteur de transcription AP-1 qui régule l'expression de nombreux gènes impliqués dans la survie ou la mort cellulaire, comme le gène Bcl-3 (Shaulian et Karin, 2002).

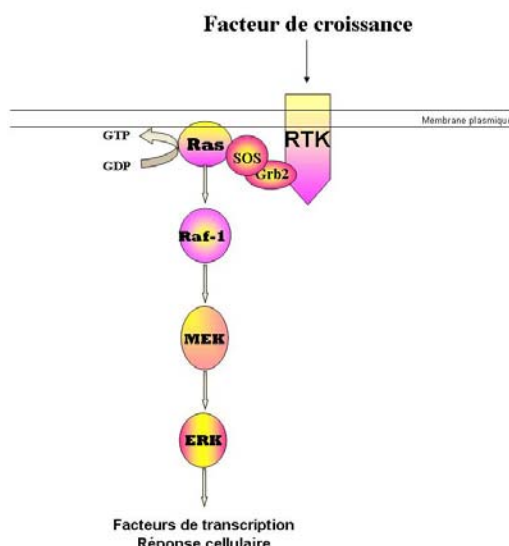


Figure I-21 : La voie des MAP-kinases Raf/MEK/ERK

Une régulation extrêmement fine de l'ensemble des voies de transduction du signal est essentielle pour l'adaptation de la cellule à son environnement, la réponse à un stimulus devant être proportionnée quantitativement, qualitativement et au cours du temps. La régulation de la voie Raf-1/MEK/ERK est complexe et fait intervenir des protéines qui assurent l'efficacité et la spécificité de la transmission du signal.

Dans le but d'identifier des protéines régulant Raf-1, Yeung et ses collaborateurs ont isolé par double hybride la protéine RKIP dans une banque d'ADNc humains (Yeung *et al.*, 1999). Ces auteurs ont montré que RKIP interrompt sélectivement la transmission du signal au-delà de Raf-1 en inhibant la phosphorylation de MEK par Raf-1. RKIP se lie à Raf-1, à MEK et plus faiblement à ERK *in vitro* et *in vivo*, mais ne se lie pas à Ras (Yeung *et al.*, 2000). Les auteurs ont montré par immunofluorescence que RKIP et Raf-1 sont co-localisées, que la cellule soit au repos ou activée par une forme oncogénique de Ras. De plus, RKIP est co-immunoprécipitée avec le complexe Raf-1/MEK/ERK. Lorsque l'expression de la PEBP endogène est réprimée par l'expression d'un ARN anti-sens ou par l'injection d'un anticorps spécifique de RKIP, la transcription des gènes dépendants de MEK, ERK et AP-1 est activée. Dans des conditions de surproduction de RKIP, les protéines MEK et ERK ne sont plus activées, il en va de même pour la transcription des gènes sous le contrôle du facteur de transcription AP-1. De plus, l'activation oncogénique de Raf-1 est abolie par une surproduction de RKIP. La

surproduction de RKIP interfère donc avec la prolifération et la différenciation cellulaire induites par Raf-1 (Yeung *et al.*, 1999).

Les mécanismes moléculaires de l'inhibition de la voie des MAP-kinases par RKIP ont été étudiés par double hybride, la capture par « GST Pulldown » et mutagénèse dirigée (Yeung *et al.*, 2000). RKIP forme des complexes ternaires avec Raf-1, MEK et ERK : alors que MEK et ERK peuvent s'associer simultanément à RKIP, les liaisons de Raf-1 à RKIP ou de Raf-1 à MEK sont mutuellement exclusives. RKIP est capable de dissocier le complexe Raf-1/MEK et agit comme un inhibiteur compétitif de Raf-1 pour la phosphorylation de MEK. La localisation des domaines de liaison montre que MEK et RKIP s'associent à des domaines distincts de Raf-1 et que Raf-1 et RKIP s'associent à des sites différents de MEK. Pour empêcher l'inhibition de la voie MAP-kinase par RKIP, les deux sites de liaison à Raf-1 et à MEK doivent être détruits. Ce résultat montre que la liaison de RKIP soit à Raf-1 soit à MEK est suffisante pour l'inhibition (Yeung *et al.*, 2000). Il semble d'autre part que la cavité conservée des PEBP intervienne dans l'interaction entre RKIP et Raf-1.

4.1.1.b. La phosphorylation de RKIP régule son activité

Plusieurs études ont montré que les sérine/thréonine kinases de la famille des PKC (Protéines Kinases C) sont capables d'activer Raf-1 et/ou MEK, en réponse à des facteurs de croissance (Schonwasser *et al.*, 1998), (Ueda *et al.*, 1996) en supprimant l'activité inhibitrice de RKIP. En effet, Corbit et ses collaborateurs ont montré que la PKC était capable de phosphoryler RKIP sur la sérine 153 (Corbit *et al.*, 2003). Cette phosphorylation de RKIP *in vivo* en réponse à l'EGF (Epidermal Growth Factor) provoque la libération de RKIP lorsqu'elle est liée à Raf-1. La surexpression de la PKC permet donc de supprimer l'inhibition de la voie MAP-kinase par RKIP.

4.1.1.c. RKIP inhibe la voie NFκB

Yeung et ses collaborateurs ont montré que RKIP jouait également un rôle de régulateur négatif de la voie de signalisation NFκB, qui intervient en réponse à la stimulation par des cytokines pro-inflammatoires comme le TNFα ou l'interleukine 1, mais aussi par le

lipopolysaccharide (LPS), les UV ou un choc thermique (Yeung *et al.*, 2001) (Figure I-22). La variation du taux d'expression de RKIP affecte la voie NFκB indépendamment de la voie MAP-kinase. RKIP agit en amont du complexe kinase responsable de la phosphorylation et de l'inactivation de l'inhibiteur de NFκB, IκB (Yeung *et al.*, 2001).

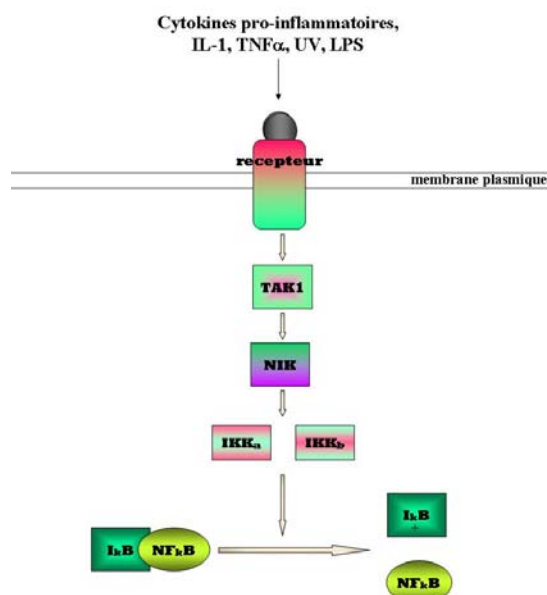


Figure I-22 : la voie NFκB.

RKIP est capable d'interagir avec quatre kinases de la voie NFκB et de les inhiber *in vitro* et *in vivo* : NIK (NFκB Inducing Kinase), TAK-1 (Transforming growth factor beta Activated Kinase-1), IKKα et IKKβ. Cela aboutit à une inhibition de la transduction du signal (Yeung *et al.*, 2001). Ce mode d'action ressemble à celui joué par RKIP avec Raf-1 et MEK dans la voie MAP-kinase.

L'ensemble de ces données suggère que la famille des PEBP représente une nouvelle famille très conservée de régulateurs de kinases (Yeung *et al.*, 2001). Comme les voies MAP-kinase et NFκB ont des rôles physiologiques bien distincts, la fonction de RKIP pourrait être, au moins pour une part, de coordonner la régulation de ces voies.

4.1.1.d. RKIP phosphorylée inhibe la kinase GRK2

Lorenz et ses collaborateurs (2003) sont allés plus loin dans l'étude de la phosphorylation de RKIP par la PKC. Ils ont montré que cette phosphorylation provoquait non seulement la dissociation du complexe Raf-1/RKIP mais également l'association de RKIP avec la kinase GRK2. GRK2 est un élément de contrôle inhibiteur essentiel exercé sur les récepteurs couplés aux protéines G (RCPG). Elle phosphoryle les RCPG activés, les découple des protéines G et initialise leur internalisation (Krupnick et Benovic, 1998). Les auteurs ont donc montré que l'association de RKIP phosphorylée à GRK2 avait pour effet d'inhiber cette dernière. En inhibant l'inhibiteur, RKIP phosphorylée participe donc à l'activation des RCPG (Figure I-23).

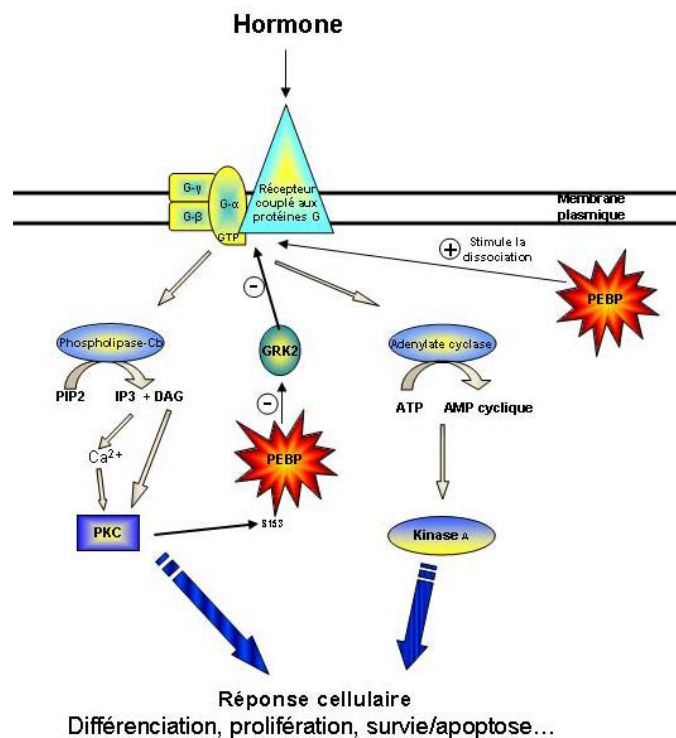


Figure I-23 : Action de RKIP sur GRK2 et les RCPG

En activant la PKC, le signal du RCPG est amplifié par deux processus : RKIP phosphorylée se détache de Raf-1, ce qui lève l'inhibition de la voie MAP-kinase ; RKIP bloque ensuite l'internalisation du récepteur RCPG en inhibant GRK2, ce qui prolonge la durée du signal d'activation à la membrane. En passant de Raf-1 à GRK2 après avoir été phosphorylée par la PKC, RKIP agit comme un régulateur de signal qui augmente la

transduction du signal en inhibant GRK2. Par ailleurs, RKIP non phosphorylée augmente le rapport signal sur bruit de l'activité de Raf-1 en inhibant son activité basale. Ainsi RKIP aide au maintien d'un équilibre entre le signal activateur de croissance dépendant de Raf-1, et la réponse inhibitrice de GRK2.

4.1.1.e. RKIP module l'activité des protéines G hétérotrimériques

Comme RKIP est capable d'interagir avec les petites protéines G (Bucquoy *et al.*, 1994) et peut être purifiée en même temps que les récepteurs μ aux opioïdes sur une colonne d'affinité à la morphine (Grandy *et al.*, 1990), Krosiak et ses collaborateurs ont cherché à identifier un rôle de la PEBP vis-à-vis des récepteurs μ (qui sont des récepteurs couplés aux protéines G, RCPG) (Krosiak *et al.*, 2001). Ces auteurs ont montré que la PEBP-1 humaine augmentait le signal induit par les protéines G hétérotrimériques. Dans des œufs de *Xenopus laevis* où la PEBP, le récepteur μ ou d'autres RCPG ont été co-exprimés, on observe une augmentation de l'activation des canaux spécifiques au potassium couplés au RCPG. Une augmentation de ce signal d'activation est également obtenue par injection de PEBP purifiée dans les ovocytes. La même étude montre qu'on observe une activation des protéines G hétérotrimériques après la surexpression de la PEBP dans les cellules de mammifères. Enfin, Krosiak et ses collaborateurs ont montré que, dans les cellules exprimant la PEBP, la quantité de $\text{GTP}\gamma^{35}\text{S}$ fixée sur la sous-unité $\text{G}\alpha$ était supérieure à celle retrouvée dans les membranes témoins (Krosiak *et al.*, 2001). RKIP est donc capable de réguler *in vivo* les protéines G hétérotrimériques et la voie de transduction du signal des récepteurs couplés aux protéines G. Il est possible que cet effet soit la conséquence de l'inhibition de GRK2 par RKIP : RKIP phosphorylée par la PKC inhibe GRK2 et contribue ainsi à bloquer l'internalisation du RCPG, et donc à activer les protéines G hétérotrimériques (Figure I-23).

4.1.2. Au delà des voies de transduction : apoptose survie des cellules

Bien que la participation de RKIP aux voies de transduction du signal soit maintenant relativement bien caractérisée, on connaît mal les effets biologiques qui sont liés à l'inhibition des voies MAP-kinases et $\text{NF}\kappa\text{B}$. Le rôle inhibiteur de RKIP dans ces deux voies suggère que RKIP pourrait réguler la prolifération et la survie des cellules. Les

recherches actuelles montrent que les cellules stressées ou les cellules traitées avec des drogues génotoxiques sur-expriment la protéine RKIP. Cette augmentation de la quantité de RKIP dans les cellules semble corrélée au passage de la prolifération ou de la survie cellulaire vers l'inhibition de la prolifération et l'induction de l'apoptose.

4.1.2.a. Rôle de RKIP dans la cytotoxicité des agents chimiothérapeutiques dans les lignées de cellules tumorales.

La majorité des drogues anticancéreuses agissent sur les mêmes voies de transduction du signal que RKIP. Il semble donc possible que RKIP puisse réguler l'apoptose en réponse à ces drogues. L'étude de différentes lignées cellulaires cancéreuses traitées avec différentes drogues a montré une surexpression de RKIP (Chatterjee *et al.*, 2004). Cela suggère que l'un des mécanismes possibles de l'action de ces drogues serait l'induction de RKIP. Il a été montré sur la lignée de cellules DU145 provenant d'un carcinome de la prostate qu'en présence de 9-nitrocamptothécine (9NC) (inhibiteur de la topoisomérase I), les cellules subissaient une apoptose généralisée accompagnée d'une augmentation significative de RKIP (Pantazis *et al.*, 2003). Or il existe une lignée de cellules RC1 issue de la lignée DU145 qui est résistante au traitement par le 9NC et pour laquelle aucune augmentation de RKIP n'a été observée lors du traitement (Chatterjee *et al.*, 2004). Les auteurs ont émis l'hypothèse que l'induction de RKIP pourrait donc être nécessaire à l'apoptose induite par le 9NC. Cette hypothèse a été confirmée en montrant que le blocage de RKIP par de petits ARN interférents empêche l'apoptose pour la lignée DU145 traitée par le 9NC, alors qu'à l'inverse l'injection de RKIP restaure la sensibilité des cellules RC1 au 9NC en provoquant leur apoptose. Enfin, le traitement de la lignée DU145 par d'autres stimulants génotoxiques comme l'étoposide ou le cisplatine provoque également l'induction de l'expression de RKIP (Chatterjee *et al.*, 2004). La corrélation directe entre la sensibilité aux drogues pro-apoptotiques et l'induction de RKIP suggère donc que la surexpression de RKIP pourrait être nécessaire pour l'apoptose des lignées de cellules de cancer de la prostate lors de l'utilisation d'agents chimiothérapeutiques utilisés en clinique.

4.1.2.b. Induction de RKIP par le Rituximab et sensibilisation des lymphomes non hodgkiniens aux agents chimiothérapeutiques

Le Rituximab est un médicament autorisé aux Etats-Unis. Il a d'abord été utilisé pour le traitement des lymphomes non hodgkiniens (LHD) primaires, mais il est maintenant largement utilisé pour les cancers qui concernent les lymphocytes B (Huhn *et al.*, 2001). Il est constitué d'un anticorps monoclonal dirigé contre le marqueur CD20 des lymphocytes B (Reff *et al.*, 1994). Lors du traitement par ce médicament, les anticorps reconnaissent et se fixent spécifiquement sur les lymphocytes B qui sont alors détruits par le système immunitaire. Des études ont montré que le Rituximab sensibilise les cellules de LHD aux drogues pro-apoptotiques en induisant une inhibition de la voie NFκB et du facteur de transcription AP-1 (voie des MAP-kinases). Jazirehi et ses collaborateurs ont montré en 2004 que cette inhibition est liée au fait que le Rituximab déclenche une induction de RKIP qui provoque une diminution de la prolifération cellulaire puis l'apoptose en agissant sur les voies de transduction du signal MAP-kinases et NFκB (Jazirehi *et al.*, 2004). Néanmoins le mécanisme qui permet l'induction de RKIP est actuellement inconnu.

4.1.3. PEBP et membranes cellulaires

Les PEBP sont solubles dans le cytoplasme, mais on les retrouve essentiellement associées aux membranes cellulaires (surtout au niveau de la membrane plasmique et des membranes du réticulum endoplasmique granuleux) (Roussel *et al.*, 1988). Cette localisation particulière, ainsi que la capacité des PEBP à fixer des ligands anioniques comme les phospholipides, a très vite amené l'hypothèse de l'implication de cette famille de protéines dans le métabolisme lipidique, la formation et le remodelage des membranes (Schoentgen *et al.*, 1993; Schoentgen et Jolles, 1995).

L'étude de la PEBP-1 bovine a montré que l'interaction avec les membranes se fait par des liaisons non covalentes, la protéine présentant une affinité pour les couches phospholipidiques chargées négativement (Vallee *et al.*, 2001). *In vivo*, la PEBP-1 humaine interagit préférentiellement avec les radeaux lipidiques chargés négativement riches en cholestérol et sphingolipides des cellules chromaffines, ce qui pourrait expliquer

que cette protéine est trouvée excrétée dans le sérum alors qu'elle ne possède pas de séquence signal (Goumon *et al.*, 2004).

Les situations où l'expression des PEBP est corrélée à un processus physiologiques faisant intervenir des remaniements de la structure des membranes cellulaires sont nombreuses. Ainsi, l'examen de cerveaux de nouveaux-nés et de leur chorde spinale a montré que des taux d'expression de RKIP très importants sont observés durant les processus de développement des oligodendrocytes de la matière blanche, des cellules de Schwann entourant les extrémités nerveuses et des neurones en cours d'élongation (Moore *et al.*, 1996). Par ailleurs, les PEBP-1 et -2 de souris sont impliquées dans la formation et la maturation des spermatozoïdes, mais avec des rôles et des localisations spatio-temporelles bien distincts (Hickox *et al.*, 2002). La protéine PEBP-1, quasi ubiquitaire chez la souris, est un facteur de décapacitation des spermatides : la protéine disparaît de la membrane au moment où le spermatozoïde acquiert sa fertilité (Gibbons *et al.*, 2005). La PEBP-2 de souris est quant à elle exprimée uniquement au niveau des cellules germinales précoces testiculaires, et semble participer à la maturation de la membrane des spermatozoïdes (Hickox *et al.*, 2002).

4.2. Rôles physiologiques

Les PEBP sont impliquées dans des processus cellulaires majeurs, mais on les retrouve aussi impliquées dans différents processus physiologiques complexes comme la spermatogénèse ou le contrôle de l'inflorescence chez les fleurs.

4.2.1. Contrôle de l'inflorescence chez les plantes

Les homologues de PEBP jouent un rôle primordial dans l'architecture des plantes. En effet, les plantes sauvages et mutées présentent des différences de phénotypes spectaculaires, notamment au niveau de l'inflorescence. Cet effet a été observé pour 3 espèces de plantes : chez *Arabidopsis*, chez une espèce de gueule de loup et chez la tomate.

4.2.1.a. *Arabidopsis thaliana*

Les protéines FT (Flower locus T) et TFL1 (Terminal Flower 1) d'*Arabidopsis thaliana* appartiennent à la famille des PEBP. Elles comportent 175 et 177 résidus respectivement. Toutes les deux présentent 24% d'identité avec la PEBP-1 humaine.

FT et TFL1 agissent au niveau de promoteurs intervenant dans la formation des fleurs grâce à leur interaction avec un facteur de transcription, FD. Cependant, FT induit la floraison alors que TFL1 la réprime (Bradley *et al.*, 1996). Des études ont montré que la région impliquée dans cette différence fondamentale d'activité des 2 protéines comprend un fragment de 14 résidus très conservé chez les orthologues de FT, mais très peu conservé chez les orthologues de TFL1. Les structures cristallographiques de ces 2 protéines ont été résolues. Il semble que la différence d'activité soit due à un résidu qui bloque partiellement l'accès au site de fixation des ligands chez FT, mais pas chez TFL1 (Tyr85 chez FT / His 88 chez TFL1) (Ahn *et al.*, 2006).

4.2.1.b. *Antirrhinum centroradialis*

La protéine CEN d'*Antirrhinum centroradialis* (gueule de loup ou mufler) appartient à la famille des PEBP, elle comporte 181 résidus et présente 23 % d'identité avec la PEBP-1 humaine.

Les mutants de CEN ont été particulièrement étudiés à cause du phénotype remarquable qu'ils présentent (Bradley *et al.*, 1996). Le mutant *cen* et le type sauvage sont radicalement différents (Figure I-24). Chez le mutant, le méristème apical inflorescent est converti en fleur : l'inflorescence indéterminée est convertie en inflorescence déterminée. L'architecture de la plante mutante pour le gène CEN est complètement modifiée : elle est plus petite et plus touffue car la croissance de sa tige s'arrête prématurément.



Figure I-24 : Inflorescence d'*Anthirinum centroradialis* sauvage (à gauche) et du mutant *cen* (à droite) (Bradley *et al.*, 1996). Contrairement au type sauvage, la mutation *cen* entraîne la formation d'une fleur (inflorescence déterminée) au niveau du méristème apical : la croissance de la plante est stoppée.

L'analyse de ce phénotype tend à montrer que le gène CEN joue un rôle dans la programmation du méristème apical, très certainement en relation avec les gènes de floraison comme le gène FLO (*floricaula*). Dans un contexte sauvage, au sixième jour après l'induction florale, CEN est fortement exprimé dans la région sub-apicale, ce qui inhibe l'expression de FLO dans la zone apicale : le méristème apical conserve alors son caractère d'inflorescence indéterminée. Par contre, chez le mutant *cen*, l'expression de FLO n'est pas inhibée et le produit du gène est présent dans tout le méristème apical (Bradley *et al.*, 1996). Ce dernier est alors converti en fleur, conduisant à un mode de floraison déterminé.

4.2.1.c. *Lycopersicon esculentum* (tomate)

Contrairement à *Antirrhinum* et *Arabidopsis* qui sont des plantes monopodiales, c'est-à-dire qui se développent d'abord selon une phase végétative (méristème végétatif) puis selon une phase reproductive (méristème inflorescent), la tomate est une plante sympodiale, c'est-à-dire pour laquelle les deux phases végétative et reproductive alternent tout au long de la croissance (Figure I-25).

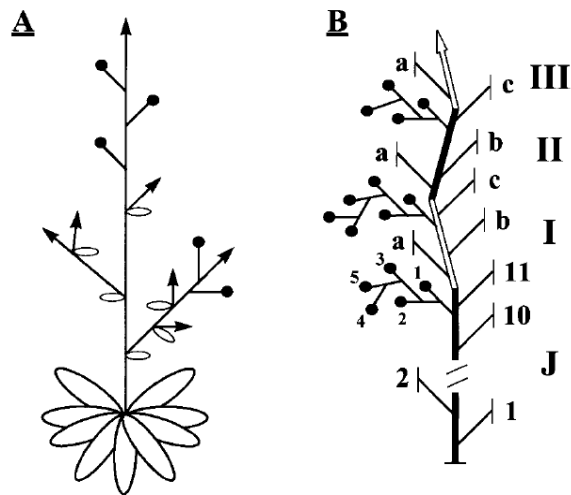


Figure I-25 : Architecture des bourgeons d'Arabidopsis et de la tomate d'après Pnueli (Pnueli et al., 1998). (A) Arabidopsis est une plante monopodiale. Les rameaux latéraux se développent sans qu'il y ait arrêt définitif du fonctionnement du méristème de la tige principale. Les flèches indiquent un développement de feuilles et de fleurs, alors que les cercles noirs indiquent la formation de fleurs solitaires. (B) La tomate est une plante sympodiale. Après une première phase juvénile (J) où la plante développe 8 à 12 feuilles et se termine par une inflorescence, une succession d'unités sympodiales apparaît (I, II, III ...). Chaque unité sympodiale est composée de 3 nœuds végétatifs (feuilles a, b et c) et d'une inflorescence. Le méristème axillaire situé juste en dessous de cette inflorescence se développe alors et la plante poursuit sa croissance.

Cette alternance de phases végétatives et reproductives est régulée par le gène SP (Self Pruning), homologue de CEN et TFL1/FT (Pnueli et al., 1998). Le gène SP code pour une protéine appartenant à la famille des PEBP qui comporte 175 résidus et présente 63 % de similarités avec la PEBP-1 humaine.

Un mutant ponctuel naturel de la tomate correspondant à un allèle récessif de SP a été décrit (Pnueli et al., 1998). Il correspond à la substitution d'un seul acide aminé particulièrement conservé au sein de la famille des PEBP : la proline 76 est remplacée par une leucine (Pnueli et al., 1998). Ce mutant *sp* développe progressivement de moins en moins de feuilles au niveau de l'unité sympodiale (trois, puis deux, puis une), et finalement sa pousse se termine par deux inflorescences consécutives (Figure I-26). Il ne présente pas fondamentalement une structure altérée mais présente plutôt une rupture de la régularité des cycles d'alternance des phases végétative et reproductrice. La croissance de la plante est alors avortée prématurément.



Figure I-26 : Inflorescence de la tomate sauvage et du mutant *sp/sp* (Pnueli et al., 1998). (A) Tomate de type sauvage. Un segment sympodial complet est présenté : il est constitué de 3 branches numérotées et d'une inflorescence terminale qui apparaît sous la branche 3. Les flèches indiquent 2 inflorescences consécutives. (B) Tomate mutante *sp/sp*. Seule une branche sépare les 2 inflorescences. L'inflorescence terminale (TI) est indiquée ainsi qu'un bourgeon axillaire (AS) qui se développe sous l'ancienne inflorescence terminale

Par ailleurs, la surexpression du gène SP conduit à transformer les méristèmes inflorescents en feuilles. D'autre part, l'effet de SP n'est pas immédiat, puisque chez le mutant *sp*, les 11 feuilles de la phase juvénile se développent normalement et que les premières unités sympodiales sont normales. Bien que le mécanisme moléculaire de l'activité de SP ne soit pas encore connu, il semblerait donc que la régulation de SP s'accroît au fur et à mesure du vieillissement de la plante. Or, une dépendance de l'expression de PEBP-1 en fonction du stade de développement a déjà été décrite, notamment pour le rat (Roussel *et al.*, 1988).

4.2.2. Spermatogénèse

Les protéines PEBP-1 et -2 de rat et de souris ont été identifiées dans le testicule où elles jouent un rôle dans la spermatogénèse à un stade précoce (PEBP-2) ou plus tardif, lors de la capacitation (PEBP-1) (Perry *et al.*, 1994; Hickox *et al.*, 2002). Il semble en effet que ces protéines interviennent lors des remaniements morphologiques complexes qui accompagnent la formation et la différenciation des spermatozoïdes. Dans les testicules de rats adultes, mPEBP-2 a été localisée au niveau des spermatides allongées, dans les corps résiduels et dans les cellules interstitielles de Leydig. La protéine mPEBP-1 a quant à elle

été retrouvée dans l'épididyme de rat adulte au niveau des cellules épithéliales des régions *caput*, *corpus* et *cauda* et dans les gouttes cytoplasmiques des spermatozoïdes dans la lumière de l'épididyme. Cependant, dans les testicules pré-pubères, la protéine mPEBP-2 n'est localisée qu'au niveau des cellules de Leydig et dans aucun autre type cellulaire avant la différenciation des spermatides (Hickox *et al.*, 2002). Récemment, Gibbons et ses collaborateurs, ont montré que la PEBP-1 de souris est un facteur de décapacitation, c'est-à-dire une des protéines présente au niveau de la membrane des spermatides mais qui est perdue durant l'étape de capacitation (Gibbons *et al.*, 2005). De par sa localisation au niveau des membranes cellulaires et sa présence particulièrement importante aux stades où les modification morphologiques des spermatozoïdes sont les plus importantes, on peut penser que les PEBP-1 et -2 sont impliquées dans la réorganisation des membranes durant la spermatogénèse.

4.2.3. Transport de la morphine

La morphine-6-glucuronide (M6G), dérivée de la morphine, est actuellement considérée comme un nouveau facteur endocrine (Glattard *et al.*, 2006). En effet, la M6G est présente et sécrétée dans les granules des cellules chromaffines de la médullosurrénale, après stimulation (Goumon *et al.*, 2006). Les mêmes auteurs ont montré qu'au sein des granules des cellules chromaffines, on trouve également une enzyme UDP-glucuronosyltransférase 2B-like, décrite comme étant capable de transformer la morphine en M6G hautement analgésique. En outre, une synthèse *de novo* de morphine-6-glucuronide a été observée, démontrant l'implication de l'UDP-glucuronosyltransférase 2B-like dans le métabolisme de M6G. Une fois sécrétée dans la circulation, la M6G pourrait avoir plusieurs actions systémiques, sur les cellules immunitaires par exemple, en fonction de leur affinité pour les récepteurs μ -opioïdes. Ces activités pourraient être facilitées par RKIP qui est capable de fixer le M6G et agit comme "bouclier moléculaire" en l'empêchant d'être dégradée rapidement (Goumon *et al.*, 2006). Des travaux réalisés chez l'Homme ont montré que certaines cellules étaient capables de synthétiser elles-mêmes de la morphine ((Zhu *et al.*, 2001); (Goumon *et al.*, 2002)). Les implications cliniques sont nombreuses. D'abord parce que le M6G est actuellement en essai de phase III dans les douleurs post-opératoires, et ensuite parce que, depuis 20 ans, des greffes de cellules chromaffines de la médullosurrénale sont expérimentées dans des modèles de

douleurs aiguës et chroniques. Dans ces expériences, le M6G pourrait donc participer activement à l'effet analgésique observé.

4.2.4. Le peptide HCNP : physiologie cardiaque / cerveau

Les PEBP-1 humaine (Tohdoh *et al.*, 1995), bovine (Goumon *et al.*, 2004), de rat (Ojika et Appel, 1984) et de souris (Matsukawa *et al.*, 1999) sont les précurseurs du peptide HCNP (Hippocampal Cholinergic Neurostimulating Peptide) qui correspond aux 11 résidus N-terminaux de ces protéines. La protéine qui assure le clivage de la PEBP de rat a été mise en évidence au niveau de l'hippocampe par Otsuka et Ojika, il s'agit d'une protéine de 68 kDa (Otsuka et Ojika, 1996).

Ojika et ses collaborateurs ont montré qu'HCNP stimule l'assimilation de la choline, la synthèse d'acétylcholine ainsi que l'activité choline acétyl-transférase, et ce de manière dose-dépendante (Ojika et Appel, 1984). Chez l'homme, cette stimulation de l'activité cholinergique est fonction de l'âge, ce qui laisse penser qu'HCNP pourrait être un facteur neurotrophique lors du développement de l'hippocampe qui joue un rôle important pour la mémoire et l'apprentissage (Yuasa *et al.*, 2001). Par ailleurs l'examen de cerveaux néonataux humains et de leur chorde spinale a montré des taux d'expression de RKIP très importants dans le cytoplasme des oligodendrocytes de la matière blanche, les cellules de Schwann entourant les extrémités nerveuses, mais une expression faible au niveau des neurones acétylcholinergiques (Moore *et al.*, 1996).

Le peptide HCNP présente d'autres fonctions physiologiques. Par exemple, sur un cœur isolé et perfusé de grenouille, le peptide HCNP agit avec un effet inotropique négatif, c'est-à-dire qu'il diminue la contractilité (gamme de concentration 10^{-11} à 10^{-7} M). De plus, ce peptide est à même d'inhiber les effets d'un agoniste β -adrénergique, l'isoprotérénol (10^{-8} M) qui possède à l'inverse un effet inotropique positif (l'inhibition totale de cette activité est observée à une concentration de HCNP de 10^{-10} M). Les recherches ont permis de déterminer que HCNP et son précurseur sont aussi sécrétés par les cellules chromaffines de la médullo-surrénale lors d'une stimulation nicotinique. Cela suggère que RKIP et HCNP peuvent être considérés comme des facteurs endocrines impliqués dans la physiologie cardiaque (Goumon *et al.*, 2004).

5. PEBP humaine et maladies

De par son implication prépondérante dans les voies de transduction du signal et dans divers processus physiologiques fondamentaux, la dérégulation de l'expression de RKIP est associée chez l'humain à un grand nombre de pathologies.

5.1. RKIP est un suppresseur de métastases

L'implication de RKIP dans la régulation de la survie et de l'apoptose des cellules suggère que cette protéine pourrait jouer un rôle central dans la progression des tumeurs. Une diminution de son expression a été relevée dans de nombreuses tumeurs : au niveau de la prostate (Fu *et al.*, 2003), du sein (Chatterjee *et al.*, 2004), du mélanome de la peau (Park *et al.*, 2005), du carcinome des cellules de Merkel (Houben *et al.*, 2006), du carcinome des cellules du foie (Schuierer *et al.*, 2006).

Fu et ses collaborateurs ont montré que RKIP était moins exprimée dans les lignées cellulaires métastatiques du carcinome de la prostate que dans les lignées non métastatiques (Fu *et al.*, 2003). Des niveaux d'expression faible, voire l'absence, de RKIP dans les cellules tumorales ont été corrélés au caractère invasif et à la résistance aux stimuli pro-apoptotiques : les cellules cancéreuses qui expriment RKIP en faible quantité sont peu sensibles aux stimuli pro-apoptotiques, de même que les cellules qui n'expriment pas du tout de RKIP (cancers avancés et métastases) apparaissent beaucoup plus résistantes aux stimuli pro-apoptotiques. A l'inverse, la surexpression de RKIP dans les métastases du carcinome de la prostate a diminué largement leur caractère invasif *in vitro* et leur développement dans le poumon *in vivo* chez la souris. La surexpression de RKIP ne semble pas avoir d'effet direct sur la tumeur primaire, mais elle est associée à une diminution très significative (3 à 5 fois) du nombre de vaisseaux qui l'infiltrant, ainsi qu'à une diminution de l'invasion vasculaire par les cellules cancéreuses (Fu *et al.*, 2003). RKIP étant un inhibiteur de la kinase Raf, les auteurs ont vérifié que son absence conduisait à une augmentation de la signalisation induite par Raf. Par ailleurs, ils ont noté que le blocage des cibles de Raf décroît la capacité invasive des cellules du cancer de la prostate, ce qui renforce clairement le rôle de la signalisation de Raf dans le processus

d'invasion métastatique. De la même manière, la diminution de l'expression de RKIP, en utilisant de petits ARN interférents dans des cellules cancéreuses de la prostate non métastatiques renforce le pouvoir invasif de la tumeur. RKIP est donc un suppresseur de métastases (Keller, 2004). Récemment, RKIP a été acceptée comme facteur pronostique pour évaluer le risque métastatique chez les patients atteints d'un cancer de la prostate (Fu *et al.*, 2006).

5.2. Maladie d'Alzheimer

L'implication de RKIP et de son produit de clivage, le peptide HCNP, dans la maladie d'Alzheimer et la démence a été étudié ((Maki *et al.*, 2002); (Mitake *et al.*, 1996); (Tsugu *et al.*, 1998)). Ces recherches s'appuient sur l'observation d'une très forte expression et accumulation de HCNP dans les corps d'Hirano. Il s'agit d'agrégats cytoplasmiques d'actine et de protéines liées à l'actine qui affectent le cytosquelette des cellules nerveuses et qui sont associés à des maladies neurodégénératives telles que la maladie d'Alzheimer. On les rencontre surtout dans la région CA1 de l'hippocampe chez les patients âgés. Les patients qui présentent des corps d'Hirano souffrent de troubles cholinergiques du système septo-hippocampal qui affectent l'apprentissage et la mémoire. De plus, des taux importants de HCNP ont été observés dans le liquide cébrospinal à un stade précoce de la maladie (Tsugu *et al.*, 1998). Au contraire, chez la lignée de souris Tg2576 qui constitue un modèle murin de la maladie d'Alzheimer, on observe une diminution de l'expression de l'homologue mPEBP-1 de RKIP en même temps que l'accumulation de β -amyloïdes. Dans ces conditions, l'implication et le rôle de RKIP et HCNP dans la physiopathologie de la maladie d'Alzheimer sont encore mal cernés.

5.3. Diabète

RKIP pourrait intervenir dans les effets secondaires du diabète par l'intermédiaire de son interaction avec la protéine kinase C (PKC). En effet, il a été montré que la PKC jouait un rôle dans les néphropathies diabétiques ((Kikkawa *et al.*, 2003); (Tuttle et Anderson, 2003)). Lors du diabète la PKC est suractivée, notamment au niveau du rein, ce qui entraîne une modification de la régulation de l'ensemble des voies de transduction du signal. RKIP semble surexprimée dans les cellules présentant ces désordres de régulation (Thongboonkerd, 2004). Des analyses sont actuellement en cours pour savoir si

l'implication de cet excès de PEBP participe au développement de la néphropathie diabétique, ou si au contraire il contribue à l'atténuation des conséquences dues à la sur-activation de la PKC.

II. Le projet « PEBP de drosophile »

Il persiste un grand nombre de questions biologiques et structurales non résolues concernant la famille des PEBP. Afin d'obtenir des informations inédites et ainsi développer de nouveaux axes de recherche, un projet basé sur la collaboration de trois équipes a été mis en place au Centre de Biophysique Moléculaire pour étudier les PEBP de *Drosophila melanogaster*.

1. Contexte scientifique

1.1. Aspects biologiques

Nous avons vu que les PEBP interviennent dans un grand nombre de processus cellulaires et physiologiques de l'organisme sain. Par ailleurs, des modifications de l'expression de ces protéines sont corrélées à plusieurs situations pathologiques qui touchent un pourcentage significatif de la population et pour lesquelles les stratégies thérapeutiques actuelles demeurent fragiles (cancers primaires et métastatiques, diabète, maladie d'Alzheimer, infertilité...). Il est donc apparu rapidement que les PEBP étaient des cibles de choix pour la recherche médicale. Cependant, les rôles et les modes d'action de ces protéines durant ces processus majeurs de la vie ne sont pas tous clairement établis ou compris. Dans l'état actuel des connaissances, la recherche médicale appliquée reste prématurée et difficile, même si les premiers résultats sont disponibles : par exemple la PEBP-1 humaine est actuellement reconnue aux États-Unis comme un facteur pronostique pour évaluer le risque de diffusion des cellules cancéreuses de la prostate dans l'organisme (Fu *et al.*, 2006). Il existe donc une forte demande pour mieux définir les rôles biologiques des PEBP.

Il est généralement admis à l'heure actuelle que les PEBP sont le support de plusieurs activités : ce sont des protéines multifonctionnelles dont les rôles sont encore probablement sous-estimés. Les questions sont donc nombreuses : Pourquoi, au sein d'un même organisme, retrouve-t-on plusieurs gènes codant pour des PEBP ? Quelles sont les différences d'activité entre ces différentes protéines ? Quelle est la fonction liée à la capacité de fixer des ligands anioniques ? Pourquoi les PEBP sont-elles observées à proximité des membranes cellulaires ? Quelle est cette fonction indispensable à la vie qui

explique la persistance des PEBP depuis les bactéries jusqu'aux mammifères en passant par les plantes ?

1.2. Aspects structuraux

Il persiste aussi un grand nombre d'interrogations concernant les relations structure-fonction pour cette famille de protéines. En effet, les structures cristallographiques montrent une très grande conservation des résidus qui entourent le site de fixation putatif des ligands anioniques. Pourquoi cette région est-elle aussi conservée ? Quelle est la fonction associée à cette configuration unique ? En quoi les zones non conservées sont-elles le support d'une activité spécifique ?

1.3. Les interactions

Enfin, il y a aussi un grand nombre de questions ouvertes concernant les interactions des PEBP. Ces protéines sont capables d'interagir avec des ligands anioniques de nature variée (phospholipides, nucléotides, hormones...), mais aussi avec des protéines, comme on l'observe dans les voies de transduction du signal. Les questions posées concernent donc à la fois les partenaires impliqués, qu'il va falloir déterminer et caractériser, mais aussi les modes d'action et les effets de l'interaction.

2. Les PEBP de drosophiles

Pour répondre à ces questions biologiques et structurales qui apparaissent majeures à la vue des implications des PEBP chez l'Homme, un projet d'étude des PEBP de drosophile a été mis en place au CBM. Il faut souligner à ce niveau que l'étude des PEBP de drosophile est un sujet original, encore inexploré. Le modèle de la mouche du vinaigre a été choisi car de nombreux outils génétiques validés sont disponibles pour l'exploration du rôle d'une protéine. Les mouches du vinaigre sont des organismes petits et faciles à élever en laboratoire. Leur cycle de génération est court et les croisements sont généralement très productifs. Enfin leur génome est séquencé, il est constitué de quatre paires de chromosomes : trois paires d'autosomes et une paire de chromosomes sexuels.

2.1. Les gènes codant pour des PEBP chez la drosophile

L'analyse informatique du génome de la drosophile indique que 7 gènes codent pour des PEBP. Ces gènes sont exprimés dans la mesure où les ADN complémentaires provenant des ARN messagers sont disponibles commercialement (<http://flybase.bio.indiana.edu/>) (CG6180, CG17919, CG17917, CG10298, CG7054 et CG5430 (aussi nommée Odorant Binding Protein 5) à partir des ARNm de têtes d'adultes ; CG18594 à partir des ARNm de larves et de têtes d'adultes).

Plusieurs constatations intéressantes peuvent être faites en observant la localisation génomique de ces sept gènes : deux gènes sont seuls et éloignés sur le bras gauche du chromosome 2, les autres gènes sont regroupés au niveau du bras droit du chromosome 3 (Figure II-1).



Figure II-1 : Répartition des sept gènes codant pour des PEBP sur le génome de la drosophile.

On y observe un groupe de deux (*cg18594* et *cg7054*) et un groupe de trois (*cg17917*, *cg17919*, *cg10298*). Il est important de noter la distance séparant les gènes d'un même groupe : par exemple 150 nucléotides entre les gènes *cg18594* et *cg7054* (Figure II-2). Cela pose de nombreux problèmes pour comprendre l'expression de ces protéines. En

effet, il est difficile d'envisager la présence d'un promoteur et de séquences régulatrices au niveau de ces 150 nucléotides. La même situation est rencontrée au niveau du groupe de trois gènes.

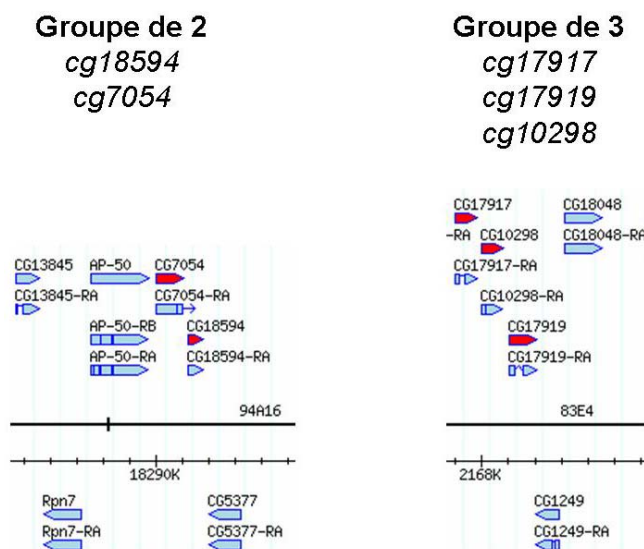


Figure II-2 : Détails de la répartition des gènes en groupes de 2 et de 3 (images issues de Flybase)

2.2. Les 7 PEBP de drosophiles

L'alignement des séquences des 7 protéines correspondant à ces gènes avec les séquences de PEBP de structure 3D connue conduit à l'observation que ces protéines appartiennent à la même famille (Figure II-3). Elles possèdent toutes la signature caractéristique de la famille des PEBP, sauf l'OBP5 qui présente une signature légèrement altérée. Cependant l'analyse de la séquence, la conservation des résidus caractéristiques et la longueur de la protéine nous amène à considérer OBP5 comme une protéine faisant partie de la famille des PEBP. En effet, l'alignement de ces 7 PEBP montre une bonne conservation des résidus, même si on observe des zones où l'alignement est beaucoup moins bon par exemple aux extrémités et au niveau de zones bien localisées correspondant essentiellement à des boucles.

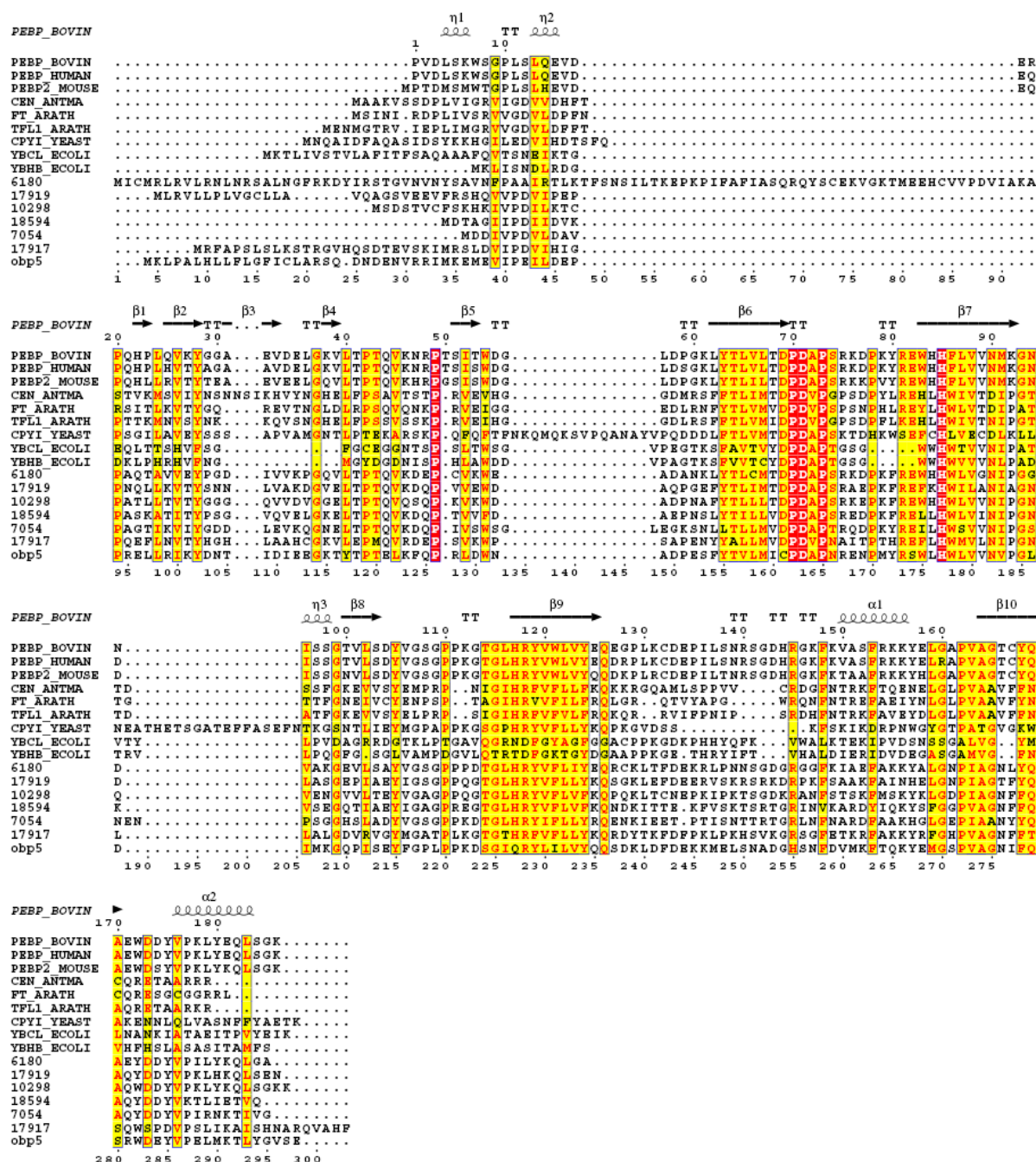


Figure II-3 : Alignement des 7 séquences des PEBP de drosophile (CLUSTALW, (Eddy, 1995)) avec les PEBP de structure tridimensionnelle connue. La séquence et les structures secondaires de la PEBP-1 bovine obtenues sur le modèle cristallographique sont indiquées. Les cadres rouges indiquent les résidus identiques ; les résidus non conservés mais similaires sont indiqués en jaune. Les structures secondaires sont indiquées. (Figure réalisée avec le logiciel Esript).

Comme nous l'avons vu précédemment, le rôle des PEBP n'est pas encore clairement établi. Un rôle nouveau vient cependant d'être suggéré par des études protéomiques ciblant l'immunité innée au niveau de la drosophile (Levy *et al.*, 2004; Vierstraete *et al.*,

2004a; Vierstraete *et al.*, 2004b). En effet, les protéines CG18594 et CG17919 sont surexprimées dans l'hémolymph de drosophile après induction par du lipopolysaccharide ou lors d'une infection à bactéries Gram+. Nous sommes surpris de retrouver deux protéines agissant au même niveau biologique mais ne faisant pas partie du même groupe de gènes. Enfin, les PEBP sont capables de lier la phosphatidylethanolamine, par définition. Hors la 1-lysophosphatidyléthanolamine est un antimicrobien qui a été identifié chez *musca domestica*, la mouche commune (Meylaers *et al.*, 2004).

3. Le projet d'étude des PEBP de drosophile

Le projet d'étude des PEBP de drosophile ne se comprend que dans le cadre d'une collaboration avec une synergie de compétences entre trois équipes du CBM : l'équipe de Daniel Locker, spécialisée dans l'étude *in vivo* des drosophiles pour examiner le rôle biologique des PEBP ; l'équipe de Françoise Schoentgen pour prendre en charge la partie biologie moléculaire et la purification des protéines et enfin l'équipe de Françoise Vovelle pour réaliser les études structurales.

3.1. Objectifs

Le projet s'est fixé des objectifs simples : obtenir de nouvelles informations biologiques et structurales concernant les PEBP en utilisant un modèle biologique reconnu.

La drosophile est le premier organisme où l'étude des différents gènes peut être envisagée d'une manière systématique. On observe plusieurs gènes capables d'exprimer des PEBP au niveau des autres organismes. Par exemple chez l'Homme on retrouve 6 gènes codant pour des PEBP. Cependant aucune information n'est disponible quant à l'expression de ces différents gènes. Chez la drosophile sept gènes différents codent des PEBP et ces gènes sont exprimés. Ces 7 protéines différentes présentent-elles des fonctions distinctes ? Existe-t-il des différences au niveau de la régulation spatiotemporelle de leur expression ? Quelles sont les relations structure fonction fines qui expliquent le maintien au cours de l'évolution de 7 PEBP au sein du même organisme ? Les nouvelles fonctions mises en évidence sont-elles différentes des fonctions mises en évidence dans les autres espèces ? Y a-t-il de nouvelles fonctions ? Est-ce que cela crée des pistes de recherches intéressantes ?

Enfin, une étude structurale par RMN, dont l'envergure dépasse le cadre de cette thèse, a été initiée afin de définir dans un premier temps la structure de ces protéines en solution mais aussi leur dynamique interne. Ensuite, en relation avec les données biologiques et biochimiques recueillies, des études d'interaction seront menées que ce soit dans le contexte interaction protéines-protéines ou protéines-ligands, par exemple l'interaction des PEBP avec la 1-lyso-phosphatidyléthanolamine. L'objectif est aussi de suggérer des mutants dont les caractéristiques pourront être évaluées *in vivo* ou *in vitro*.

3.2. Mon rôle au sein du projet

Ma thèse s'intègre dans l'étude structurale des PEBP de drosophile et a impliqué le clonage des différents gènes, la mise au point des protocoles de purification et de marquage des protéines, la détermination des conditions d'études pour la RMN, un début de caractérisation physico-chimique en utilisant les techniques disponibles au laboratoire, l'acquisition des spectres RMN hétéronucléaires, le traitement des spectres, l'attribution des signaux, la modélisation de la structure en utilisant les contraintes expérimentales déduites de la RMN et enfin l'étude de la dynamique. Il faut signaler que les PEBP sont les premières grosses protéines étudiées par RMN au laboratoire, ce qui a impliqué de nombreuses mises au point expérimentales et méthodologiques. Par ailleurs les protéines purifiées ont été utilisées pour obtenir des anticorps spécifiques, particulièrement utiles pour les recherches de l'équipe de Daniel Locker.

Matériels et méthodes

III. Matériels et méthodes

1. Dichroïsme circulaire :

1.1. Principe de la méthode

Le dichroïsme circulaire est basé sur la capacité des molécules chirales d'interagir différemment avec la lumière polarisée circulairement gauche et droite. Une onde plane polarisée rectilignement peut être considérée comme étant la résultante d'une onde lumineuse polarisée circulairement gauche et d'une onde polarisée circulairement droite, de même intensité et de même fréquence. Les deux composantes traversent une solution contenant une espèce chimique douée d'une activité optique rotatoire avec des célérités différentes, ce qui a pour conséquence de faire tourner le plan de polarisation d'un angle α , et sont absorbées différemment (amplitude E_g différente de E_d). Après passage dans la solution, la lumière est donc polarisée elliptiquement (Figure III-1).

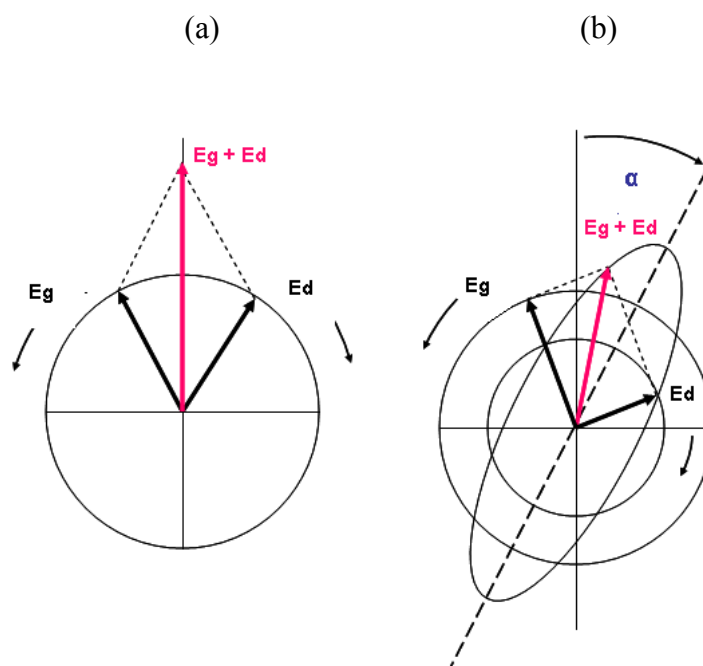


Figure III-1 : (a) une onde polarisée linéairement peut être considérée comme la somme de deux ondes polarisées circulairement droite et gauche, de même amplitude et de même phase. (b): la vitesse et l'absorption différentes des composantes droite et gauche durant la traversée de l'échantillon conduisent à la formation d'une onde polarisée elliptiquement.

En pratique, dans le spectropolarimètre, un faisceau linéairement polarisé est divisé en deux composantes polarisées circulairement, l'une droite, l'autre gauche, de même amplitude E_0 , lors de son passage à travers un modulateur (généralement un cristal piézo-électrique tel que le quartz soumis à un champ électrique alternatif). Le modulateur transmet successivement chacune des deux composantes vers l'échantillon puis le photomultiplicateur.

Le dichroïsme circulaire reflète la différence d'amplitude entre les deux composantes droite et gauche après traversée de l'échantillon, et permet de mesurer le pouvoir dichroïque d'une substance à une longueur d'onde donnée. Ce dernier peut être exprimé :

- En différence d'absorbance ΔA des deux composantes gauche (A_g) et droite (A_d) :

$$\Delta A = A_g - A_d \quad [III-1]$$

Avec

$$A_g = \log \left(\frac{I_0}{I_g} \right) = \log \left(\frac{E_0^2}{E_g^2} \right) \quad [III-2]$$

- En différence de coefficient d'extinction molaire $\Delta \epsilon$:

$$\Delta \epsilon = \epsilon_g - \epsilon_d \quad [III-3]$$

où $\epsilon = A/(l.c)$ est exprimé en $(\text{mol.L}^{-1})^{-1} \cdot \text{cm}^{-1}$

- En ellipticité θ (Exprimée en degrés):

$$\theta = \tan^{-1} \cdot \frac{b}{a} \quad [III-4]$$

où b et a sont respectivement le petit axe et le grand axe de l'ellipse décrite par l'onde résultante :

$$b = \left| \overrightarrow{E_d} - \overrightarrow{E_g} \right| \quad a = \left| \overrightarrow{E_d} + \overrightarrow{E_g} \right|$$

On montre que θ et ΔA sont liés par la relation :

$$\theta = 32.98 \Delta A \sim 33.00 \Delta A \quad [III-5]$$

L'unité couramment utilisée est l'ellipticité molaire, $[\theta]$, qui permet de s'affranchir de la contribution du trajet optique (l) et de la concentration (c). L'ellipticité et l'ellipticité molaire sont liées par la relation :

$$[\theta] = \frac{100 \times \theta}{l \times c} = 3298 \times \frac{\Delta A}{l \times c} = 3298 \Delta \epsilon \quad [III-6]$$

$[\theta]$ est exprimé en $\text{deg} \cdot \text{dmol}^{-1} \cdot \text{cm}^2$ (remarque : $1 \text{ deg} \cdot \text{dmol}^{-1} \cdot \text{cm}^2 = 1 \text{ deg} \cdot (\text{mol} \cdot \text{L}^{-1})^{-1} \cdot \text{m}^{-1}$), θ en deg, l en cm et c en $\text{mol} \cdot \text{L}^{-1}$.

Le signal de dichroïsme circulaire est observé au niveau des bandes d'absorption.

1.2. Dichroïsme circulaire des protéines

Le dichroïsme circulaire est une méthode très sensible à la conformation des protéines, donc appropriée pour obtenir rapidement la structure secondaire globale d'une protéine ou d'un peptide et pour étudier les changements structuraux provoqués par exemple par la dénaturation ou par la mutation d'un acide aminé. La rapidité et la simplicité de la mesure en font une technique de choix pour une première approche de la structure secondaire des protéines. En effet :

- Un rayonnement dans l'UV lointain (190 à 250 nm) permet de caractériser qualitativement et quantitativement les structures secondaires. Dans cette région, la liaison peptidique (CO-NH) est le chromophore. Le groupe amide possède deux transitions électroniques parfaitement définies : la transition $n \rightarrow \pi^*$, de faible intensité mais large, centrée vers 210-220 nm, et la transition $\pi \rightarrow \pi^*$, intense et proche de 190 nm. Chaque type de structure secondaire donne un spectre caractéristique (Figure III-2).
- La zone de l'UV proche (250 à 290 nm) rend compte de la structure tertiaire des protéines. Les chaînes latérales des résidus aromatiques Phe, Tyr et Trp possèdent des bandes d'absorption $\pi \rightarrow \pi^*$ caractéristiques entre 250 et 300 nm lorsqu'elles sont placées dans des environnements chiraux. Les ponts disulfure et certains groupes prosthétiques émettent aussi dans cette région. Il est ainsi possible, à partir du spectre de l'UV proche, de déduire l'angle dièdre d'une liaison disulfure. De même des renseignements sur la conformation et l'environnement des groupes prosthétiques, qui

jouent un rôle clef dans la fonction de la protéine, sont aussi accessibles dans cette gamme de longueurs d'ondes.

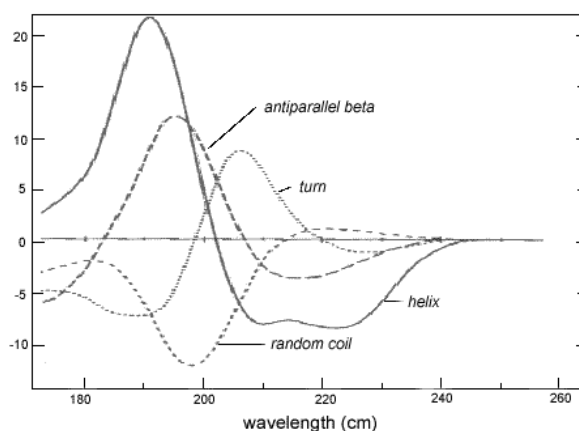


Figure III-2 : Signatures spectrales caractéristiques des différentes structures secondaires en dichroïsme circulaire dans l'UV lointain.

Dans ce travail, les spectres CD ont été réalisés dans le domaine de l'UV lointain. En faisant l'hypothèse que le spectre dichroïque expérimental est une somme pondérée de trois spectres élémentaires correspondant respectivement aux liaisons peptidiques d'hélices α , de feuillets β et de régions non structurées, il est possible d'obtenir à partir de l'analyse du spectre de dichroïsme des informations structurales concernant la protéine. Dans notre cas, nous avons choisi d'utiliser le programme de déconvolution « Dicroprot » avec l'algorithme « self consistent method 3 » (Sreerama *et al.*, 2000). Il s'agit de la méthode qui donne la meilleure correspondance entre les résultats de déconvolution et la structure tridimensionnelle de la protéine PEBP de bœuf déterminée par cristallographie (Serre *et al.*, 1998).

2. La RMN pour l'étude de la structure des protéines

De nos jours, deux techniques sont couramment utilisées pour déterminer la structure des macromolécules biologiques à une résolution de l'ordre de l'angström : la radiocristallographie des rayons X (RX) et la résonance magnétique nucléaire (RMN).

La cristallographie des rayons X nécessite des cristaux de protéine bien ordonnés. Les positions des atomes sont déduites des cartes de densité électronique obtenues après

traitement des diagrammes de diffraction des rayons X. La première structure de protéine résolue par cette technique fut la myoglobine en 1958 (Kendrew *et al.*, 1958). Depuis, grâce à des progrès techniques (sources synchrotron, robots de cristallisation) et méthodologiques, le nombre de structures résolues par cristallographie n'a cessé d'augmenter. A ce jour, plus de 30000 structures de protéines, de virus et de peptides résolus par RX sont déposées dans la banque de données PDB (Protein Data Bank, <http://www.rcsb.org/pdb>).

L'utilisation de la RMN est beaucoup plus récente puisque la première structure de protéine (un inhibiteur de protéase) a été résolue par cette technique en 1985 (Williamson *et al.*, 1985). Cette méthode exploite les propriétés magnétiques du spin nucléaire. Les spins précessent autour du champ magnétique B_0 à une vitesse angulaire ω_0 correspondant à la fréquence de Larmor qui dépend du rapport gyromagnétique γ du noyau.

$$\omega_0 = \gamma B_0 \quad [III-7]$$

La fréquence de résonance d'un spin est très sensible à l'environnement chimique de l'atome considéré, et chaque spin possède donc une fréquence de résonance particulière, on parle de déplacement chimique δ exprimé en ppm.

Lorsque l'échantillon est placé dans un champ magnétique B_0 , les spins s'orientent parallèlement ou antiparallèlement au champ magnétique : les niveaux d'énergie des noyaux possédant un spin nucléaire ne sont plus dégénérés et se séparent. Une faible majorité des spins (la polarisation est de l'ordre de 1,4 pour 10 000 spins dans un champ magnétique de 18,8 Tesla) est orientée parallèlement au champ et génère une aimantation macroscopique. L'expérience RMN consiste à exciter les noyaux par des impulsions radiofréquence pour rompre l'équilibre thermodynamique et modifier les populations des niveaux d'énergie. Cette excitation permet de modifier l'orientation de l'aimantation macroscopique qui passe d'une orientation parallèle au champ magnétique (à l'équilibre) vers une orientation perpendiculaire à ce même champ. Lors du retour à l'équilibre, la précession de l'aimantation va générer un courant détectable au niveau d'une bobine. Ce signal correspond à une sinusoïde amortie dans le temps appelée FID (Free Induction Decay, signal de précession libre). Pour passer du domaine temporel au domaine des fréquences, une transformée de Fourier est appliquée. En modulant le type (forme, durée),

l'intensité et la séquence des impulsions radiofréquences, on pourra mettre en évidence différents types d'interactions entre des paires de noyaux : interactions à travers des liaisons, c'est-à-dire des couplages 1J , 2J , 3J (corrélations scalaires) ; interactions à travers l'espace entre protons distants de moins de 5 Å (corrélations dipolaires). Différentes expériences permettront donc d'une part de procéder à l'attribution des déplacements chimiques des noyaux de la molécule, d'autre part d'obtenir des informations structurales qui seront utilisées pour reconstruire les structures par des méthodes de modélisation moléculaire.

Actuellement plus de 5000 structures de protéines et peptides ont été résolues par RMN. C'est aussi un outil efficace pour étudier les interactions intermoléculaires, l'environnement des noyaux étant modifié lors d'une interaction. La limitation majeure de cette technique correspond à la taille des molécules étudiées : pour les objets de grande taille, les propriétés de relaxation entraînent une baisse de qualité des signaux enregistrés et la résolution des spectres n'est plus suffisante pour distinguer les signaux. Ainsi, bien que l'attribution ainsi que le repliement d'une protéine de 753 résidus (malate synthase G) aient été déterminés par RMN (Tugarinov *et al.*, 2005), la limite actuelle pour une étude classique se situe aux alentours de 30 kDa.

2.1. La RMN : progrès technologiques

Pour que la structure d'une protéine puisse être étudiée par RMN, de nombreuses conditions doivent être respectées. La protéine doit être pure, soluble (idéalement de l'ordre du millimolaire), monodisperse et sa masse moléculaire ne doit pas être trop importante sous peine d'observer des signaux de mauvaise qualité. La RMN n'est pas une technique très sensible du fait de la faible polarisation de l'échantillon placé dans le champ magnétique. Elle nécessite donc l'utilisation d'échantillons concentrés et stables car l'enregistrement des différentes expériences peut durer plusieurs semaines.

2.1.1. Caractéristiques du signal RMN

Le rapport signal sur bruit (S/N) d'une expérience est proportionnel aux paramètres suivants :

$$\frac{S}{N} \propto nA \frac{1}{T} B_0^{3/2} \gamma_{exc} \gamma_{obs}^{3/2} T_2^* \sqrt{N_{scans}} \quad [III-8]$$

où n est le nombre de molécules dans le volume d'observation, A l'abondance des spins actifs, T la température, B_0 le champ magnétique statique, γ_{exc} et γ_{obs} les rapports gyromagnétiques des spins excités et observés, T_2^* le temps de relaxation transversale (* indique que l'on tient compte de l'inhomogénéité du champ magnétique : $T_2^* < T_2$) et N_{scans} le nombre de scans.

Cette relation montre que la sensibilité augmente avec l'intensité du champ magnétique et le nombre de spins observables (concentration de l'échantillon).

En revanche, la taille des molécules observées est un facteur de diminution de la sensibilité, par le biais de la largeur de raie. En effet, la largeur à mi-hauteur d'un pic est proportionnelle à l'inverse du temps de relaxation transversale T_2^* . Or T_2 décroît quand le temps de corrélation τ_c augmente. τ_c est lui-même proportionnel à la taille de la molécule :

$$\tau_c = \frac{\eta V}{kT} \quad (\text{pour une molécule globulaire}) \quad [III-9]$$

où η est le coefficient de viscosité du milieu, k la constante de Boltzmann, T la température et V le volume de la molécule qui est directement relié à sa masse moléculaire.

Il en résulte que la largeur des signaux augmente avec la taille de la molécule. En conséquence, la sensibilité et la résolution d'une expérience diminuent avec l'augmentation de la taille de la molécule étudiée (Figure III-3). De plus, la gamme des déplacements chimiques des différents signaux détectables est constante et limitée : ils s'étendent habituellement sur 12 ppm en proton, environ 30 ppm en azote. L'augmentation du nombre de signaux observables dans des molécules de grande taille comme les protéines entraîne une superposition des signaux qui, ajoutée à la diminution de la résolution, rend les spectres difficilement interprétables.

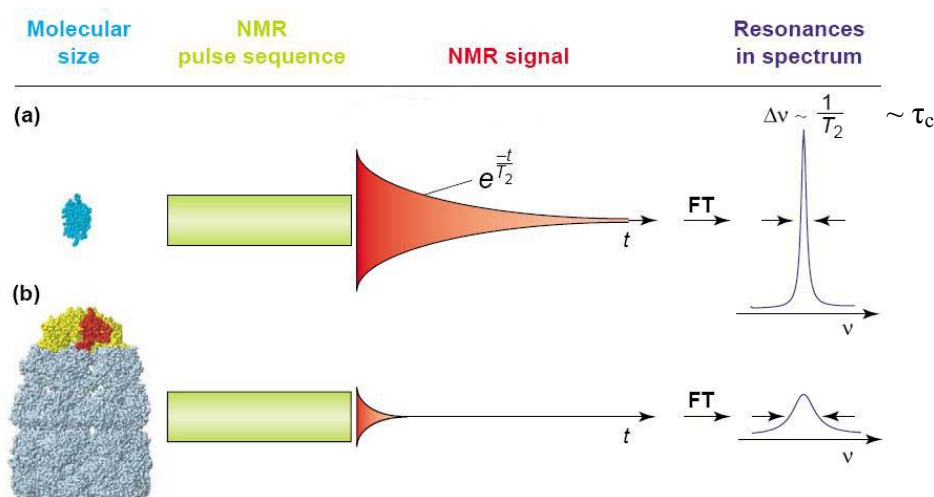


Figure III-3 : Spectroscopie RMN selon le poids moléculaire. (a) Le temps de relaxation transverse (T_2) pour une petite molécule est long (le signal persiste longtemps), ce qui mène à des pics fins et intenses. (b) Le temps de relaxation transverse (T_2) pour une grosse molécule est court ce qui mène à des pics larges et, évidemment, moins intenses. Tiré de (Fernandez et Wider, 2003).

Néanmoins, de nombreuses avancées ont été réalisées ces dernières années pour améliorer les performances de la technique et étendre la gamme des tailles moléculaires étudiables. Elles concernent les méthodes de marquage et de préparation des échantillons mais aussi les spectromètres et les méthodes d'analyse.

2.1.2. Les progrès instrumentaux et techniques.

Nous avons vu que la sensibilité d'une expérience augmente avec le champ magnétique en $B_0^{3/2}$. L'intensité du champ magnétique produit par les aimants n'a cessé d'augmenter pour passer de 2,1 Tesla (90 MHz en ^1H) en 1962 à 22,1 Tesla (950 MHz) en 2005. Un spectre enregistré à 18,8 Tesla (800 MHz) est 6,4 fois plus sensible et deux fois mieux résolu qu'un spectre enregistré à 9,4 Tesla (400 MHz). L'augmentation de l'intensité du champ permet d'augmenter l'intensité des raies de résonance et de diminuer le nombre de recouvrements.

S'il n'est pas possible de baisser considérablement la température des échantillons observés en solution, l'arrivée des cryosondes constitue cependant une autre avancée technique majeure. Dans ces dernières, les bobines de détection sont plongées dans un bain d'hélium liquide, ce qui permet de diminuer le bruit d'origine thermique et ainsi d'augmenter la sensibilité des expériences par un facteur deux ou trois par rapport à une

sonde classique, à condition que la force ionique de l'échantillon soit faible (faible concentration saline). L'utilisation de cryosondes permet donc de diminuer la concentration de l'échantillon, et ainsi de travailler sur de plus petites quantités de protéines.

2.1.3. Le marquage isotopique et la RMN hétéronucléaire

Des avancées importantes proviennent des progrès concomitants réalisés dans le marquage isotopique et le développement des techniques de RMN hétéronucléaire multidimensionnelle.

Les protéines sont des molécules organiques constituées d'hydrogène, d'azote, de carbone, d'oxygène et de soufre. Parmi ces éléments, seul le proton présente un spin, une abondance naturelle et un rapport gyromagnétique adéquats pour être détecté directement d'une manière satisfaisante. Les techniques de marquage isotopique actuelles permettent d'introduire, de manière partielle ou totale, d'autres noyaux présentant un spin $\frac{1}{2}$ et qui sont par conséquent détectables (^{13}C , ^{15}N). Depuis les années 1980, des expériences hétéronucléaires multidimensionnelles ont été développées pour exploiter les couplages scalaires de type ^1J (atomes voisins séparés par une seule liaison chimique) entre hétéronoyaux ou entre proton et hétéronoyau, plutôt que les couplages $^3\text{J}_{\text{HH}}$. Selon les noyaux considérés, et le nombre de liaisons qui les séparent, la valeur du couplage spin-spin varie. On peut donc sélectionner le cheminement des cohérences et les noyaux édités. Ainsi les expériences HSQC à deux dimensions relient le déplacement chimique d'un proton à celui de l'atome ^{15}N ou ^{13}C auquel il est lié. D'autres expériences hétéronucléaires à trois dimensions permettent de corrélérer spécifiquement plusieurs atomes entre eux, selon les critères suivants : i) le type d'atome, ii) sa position dans le résidu, iii) la nature de la corrélation qui peut s'effectuer soit via les liaisons chimiques (couplage scalaire), soit à travers l'espace (couplage dipolaire). L'attribution consistera à combiner les informations provenant des différents spectres pour extraire l'ensemble des résonances désirées. L'approche hétéronucléaire apporte donc deux avantages majeurs:

1. la corrélation de noyaux reliés par liaisons covalentes qui facilite l'attribution des résonances.

2. L'augmentation de la résolution des spectres. En effet, les résonances des protons vont être séparées selon une 3ème voire une quatrième dimension en fonction de la nature chimique de leurs voisins.

D'autres techniques telles que la méthode TROSY (Transverse Relaxation Optimized Spectroscopy) ont été développées afin de contourner certaines limitations liées à la diminution du temps de relaxation transversale T_2 des molécules de tailles importantes (Pervushin *et al.*, 1997). Elles ne seront pas décrites dans ce manuscrit car nous n'avons pas eu besoin d'y recourir.

2.2. Stratégie d'attribution pour une protéine de 20 kDa

Une protéine de 20 kDa possède environ 1400 atomes d'hydrogène, 900 de carbone et 240 d'azote. Compte tenu des recouvrements sur les spectres proton, pour déterminer la fréquence de résonance de l'ensemble des protons, il est nécessaire d'avoir recours aux techniques de RMN hétéronucléaire multidimensionnelle. Elles permettent de réaliser l'attribution séquentielle des résidus le long du squelette peptidique en suivant l'enchaînement des liaisons covalentes (exploitation des couplages scalaires). Avec une protéine doublement marquée $^{15}\text{N} / ^{13}\text{C}$, les couplages scalaires homonucléaires et hétéronucléaires utiles sont $^1J_{\text{C-C}}$, $^1J_{\text{H-C}}$, $^1J_{\text{H-N}}$, $^1J_{\text{C-N}}$, $^2J_{\text{C-N}}$ et $^2J_{\text{C-C}}$. Leurs valeurs sont indiquées sur la Figure III-4 : elles sont uniformes et dépendent peu de la conformation de la protéine. Cette propriété permet d'éditer spécifiquement certains noyaux. Ces constantes de couplages sont supérieures aux constantes $^3J_{\text{HH}}$ et permettent donc que le transfert d'aimantation soit plus efficace lorsque la largeur de raie est importante pour de grosses molécules.

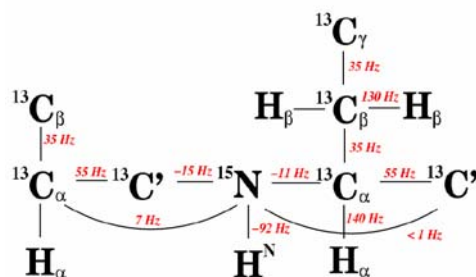


Figure III-4 : Constantes de couplage entre les noyaux d'un peptide.

Dans un premier temps on réalise l'attribution des atomes du squelette peptidique (N, H_N, C_α, H_α, C'), puis l'attribution est étendue aux chaînes latérales.

2.2.1. Attribution des résonances du squelette peptidique d'une protéine

Pour attribuer les résonances des atomes du squelette peptidique, ainsi que des atomes C_β et H_β des chaînes latérales, il est nécessaire de combiner les informations provenant de plusieurs expériences tridimensionnelles. On utilise des couples d'expériences dont le principe est de corréler l'azote et le proton amide d'un résidu (i) avec un troisième atome (ou plus) (¹H ou ¹³C) appartenant soit au même résidu (i), soit au résidu précédent (i-1). En combinant ces informations intra et inter-résidu, l'enchaînement des atomes du squelette peptidique se reconstitue.

Le nom des expériences provient du type de noyau édité, correspondant donc à une des trois dimensions, et des chemins de cohérence suivis. Les parenthèses autour d'un noyau signifient que le chemin de cohérence passe par ce noyau mais qu'il n'est pas édité. Ainsi, l'expérience HNCA corréle les noyaux H_{N(i)}-N_(i)-C_{α(i)} et les noyaux H_{N(i)}-N_(i)-C_{α(i-1)}. Au contraire l'expérience HN(CO)CA ne corréle que les noyaux H_{N(i)}-N_i-C_{α(i-1)} car le couplage ²J_{N(i)CO(i)} est trop faible (Bax et Ikura, 1991). Par comparaison, il est alors facile d'exploiter les informations séquentielles ou intra-résiduelles (Figure III-5).

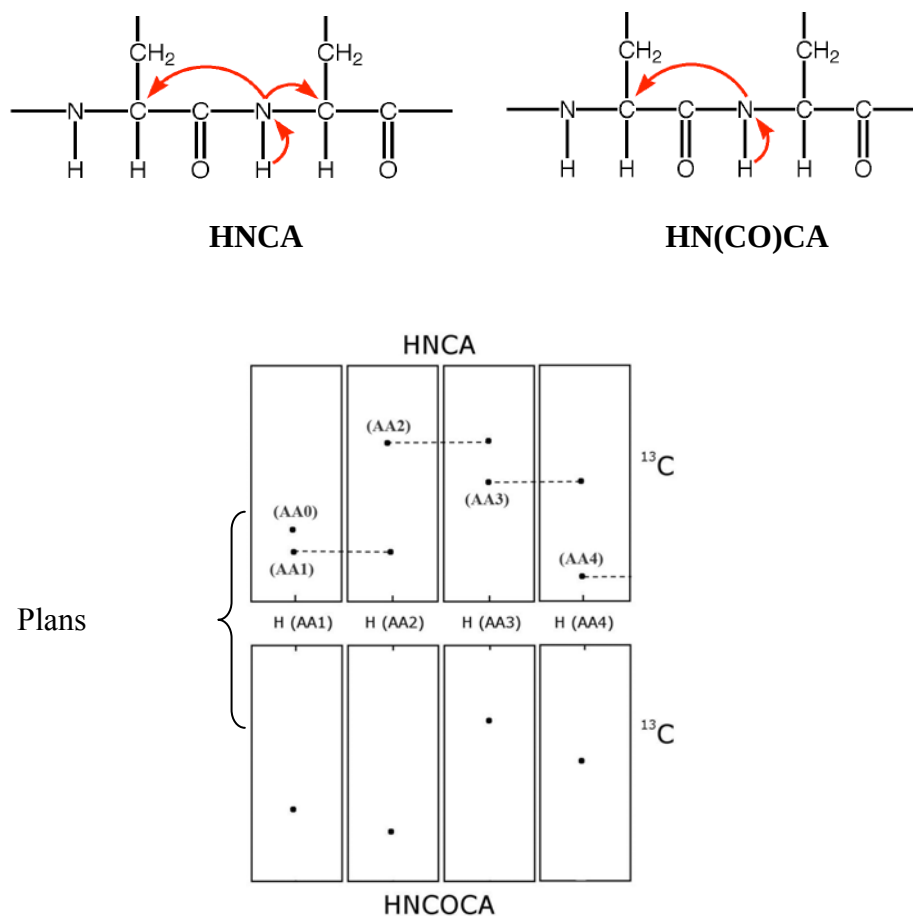


Figure III-5 : les expériences HNCA et HN(CO)CA (Bax et Ikura, 1991). L'expérience HNCA corrèle les noyaux $H_{N(i)}-N_i-C_{\alpha(i)}$ et $H_{N(i)}-N_i-C_{\alpha(i-1)}$: dans le plan correspondant au déplacement chimique de l'azote d'un résidu i et au niveau du déplacement chimique de son H_N , on observe 2 corrélations dans la dimension carbone qui correspondent aux C_{α} du résidu i et du résidu précédent. Dans la même zone du spectre HN(CO)CA, on n'observe qu'une seule corrélation dans la dimension carbone, qui correspond au C_{α} du résidu précédent : cette expérience ne corrèle que les noyaux $H_{N(i)}-N_i-C_{\alpha(i-1)}$.

Pour une protéine de 20 kDa doublement marquée $^{13}\text{C}/^{15}\text{N}$, les superpositions des signaux des atomes du squelette peptidique sont nombreuses. L'utilisation de plusieurs couples d'expériences permet de combiner les informations séquentielles et intra-résiduelles pour lever un maximum d'ambiguïtés. Ainsi, 9 expériences sont nécessaires pour l'attribution du squelette peptidique d'une protéine d'environ 200 résidus. Le tableau iii-1, ci-dessous, détaille ces expériences, ainsi que corrélations observées :

Expérience	Corrélations observées	Références bibliographiques
HSQC ^1H-^{15}N	$\text{H}_{\text{N}(i)}, \text{N}_i$	Kay, Keifer and Saarinen, JACS, 114, 10663 (1992)
HNCO	$\text{H}_{\text{N}(i)}, \text{N}_i, \text{C}_{\text{CO}(i-1)}$	Ikura, Kay and Bax, Biochem, 29, 4659 (1990)
HN(CA)CO	$\text{H}_{\text{N}(i)}, \text{N}_i, \text{C}_{\text{CO}(i-1)}$ $\text{H}_{\text{N}(i)}, \text{N}_i, \text{C}_{\text{CO}(i)}$	Yamazaki, Lee, Arrowsmith, Muhandiram, Kay, JACS, 116, 11655 (1994)
HNCA	$\text{H}_{\text{N}(i)}, \text{N}_i, \text{C}_{\alpha(i-1)}$ $\text{H}_{\text{N}(i)}, \text{N}_i, \text{C}_{\alpha(i)}$	Ikura, Kay and Bax, Biochem, 29, 4659 (1990)
HN(CO)CA	$\text{H}_{\text{N}(i)}, \text{N}_i, \text{C}_{\alpha(i-1)}$	Yamazaki, Lee, Arrowsmith, Muhandiram, Kay, JACS, 116, 11655 (1994)
HNCACB	$\text{H}_{\text{N}(i)}, \text{N}_i, \text{C}_{\alpha(i)}$ $\text{H}_{\text{N}(i)}, \text{N}_i, \text{C}_{\alpha(i-1)}$ $\text{H}_{\text{N}(i)}, \text{N}_i, \text{C}_{\beta(i)}$ $\text{H}_{\text{N}(i)}, \text{N}_i, \text{C}_{\beta(i-1)}$	Wittekind and Mueller, JMR, B101, 201 (1993)
CBCA(CO)NH	$\text{H}_{\text{N}(i)}, \text{N}_i, \text{C}_{\alpha(i-1)}$ $\text{H}_{\text{N}(i)}, \text{N}_i, \text{C}_{\beta(i-1)}$	Grzesiek and Bax, JACS 114, 6291 (1992)
HNHA	$\text{H}_{\text{N}(i)}, \text{N}_i, \text{H}_{\alpha(i)}$	Vuister, G. W. & Bax, A., JACS, 115, 7772-7777 (1993)
HBHA(CO)NH	$\text{H}_{\text{N}(i)}, \text{N}_i, \text{H}_{\alpha(i-1)},$ $\text{H}_{\text{N}(i)}, \text{N}_i, \text{H}_{\beta(i-1)}$	Grzesiek and Bax, J. Biomolecular NMR, 3185-204 (1993)

Tableau III-1 : expériences nécessaires à l'attribution du squelette peptidique d'une protéine de 20 kDa.

L'expérience HNCO est la plus sensible du fait des grandes valeurs des couplages impliqués (Figure III-4, page 86). Cependant, son utilisation reste limitée dans la mesure où la dispersion des déplacements chimiques des carbonyles est relativement faible (environ 12 ppm) et où l'expérience HN(CA)CO, qui contient l'information séquentielle, est bien moins sensible. La gamme de déplacement chimique des carbones C_{α} est plus étendue (environ 30 ppm). Elle est exploitée au niveau des couples d'expériences HNCA / HN(CO)CA et HNCACB / CBCA(CO)NH. Ce dernier couple intègre également les déplacements chimiques des C_{β} mais la sensibilité de l'expérience est diminuée par rapport à celle du couple HNCA / HN(CO)CA. La combinaison des informations séquentielles apportées par les carbonyles, C_{α} et C_{β} permet de lever la grande majorité des ambiguïtés. Enfin, les déplacements chimiques des C_{α} et C_{β} présentent une dépendance avec le type d'acide aminé auquel ils appartiennent, ce qui constitue une aide durant le processus d'attribution (Figure III-6). Il faut cependant remarquer que les structures secondaires peuvent influencer de manière importante sur les déplacements chimiques des noyaux impliqués. Cette propriété est mise à profit pour l'identification des structures secondaires (*cf. infra* 2.3.1 : Utilisation de l'index de déplacement chimique (Chemical Shift Index, CSI) pour déterminer les structures secondaires, page 92).

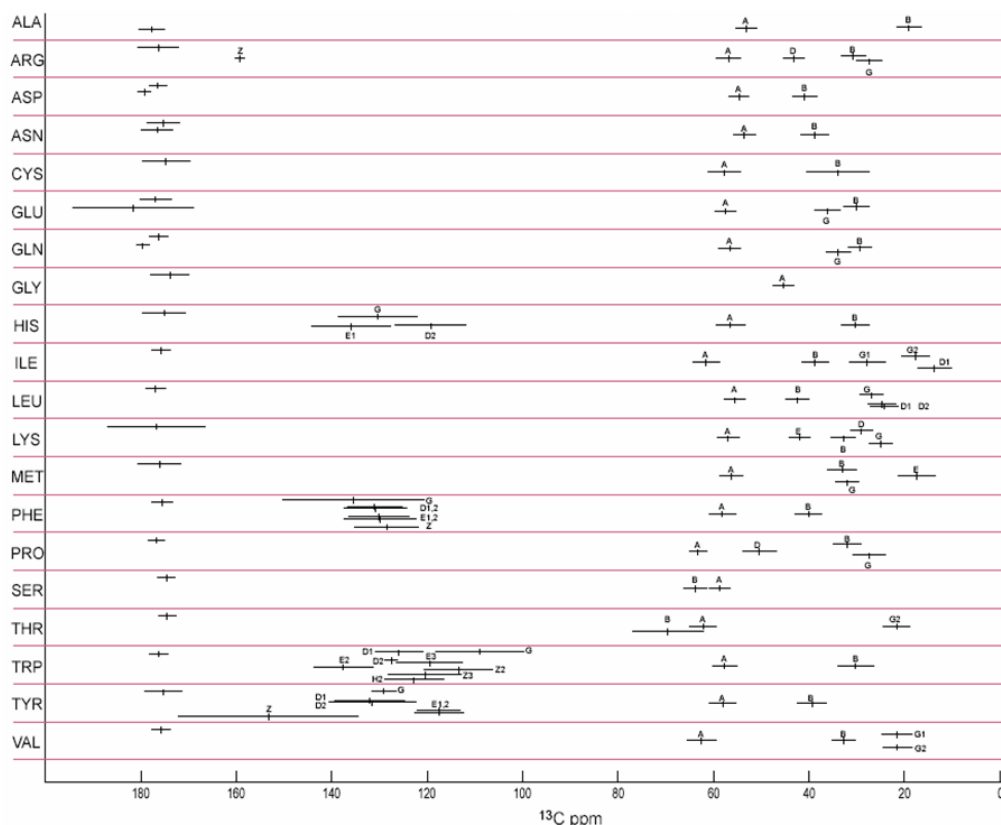


Figure III-6 : Intervalles de déplacements chimiques observés pour les carbones de chaque type d'acide aminé. Le type de carbone est indiqué. Les intervalles non annotés correspondent aux C_{CO} du squelette peptidique et de la chaîne latérale pour la glutamine, l'acide glutamique, l'asparagine et l'acide aspartique. (Source : BMRB).

2.2.2. Attribution des résonances des chaînes latérales

L'attribution des résonances des chaînes latérales est réalisée grâce à 3 expériences hétéronucléaires : HCCH-TOCSY, H(CC)(CO)NH-TOCSY et C(CO)NH TOCSY. Le Tableau III-2 ci-dessous détaille ces expériences ainsi que les informations obtenues :

Expérience	Corrélations observées	Références bibliographiques
HCCH-TOCSY	$H_{\text{aliph}}(i)$, $C_{\text{aliph}}(i)$	Kay et. al. J. Magn. Reson. B 101, 333-337 (1993)
H(CCO)NH	$H(i)$, $N(i)$, $H_{\text{aliph}}(i-1)$	G.T. Montelione, B.A. Lyons, S.D. Emerson & M. Tashiro, ; J. Am. Chem. Soc. 114, 10974-75 (1992)
C(CO)NH	$H(i)$, $N(i)$, $C_{\text{aliph}}(i-1)$	G.T. Montelione, B.A. Lyons, S.D. Emerson & M. Tashiro, ; J. Am. Chem. Soc. 114, 10974-75 (1992)

Tableau III-2 : Expériences nécessaires à l'attribution des chaînes latérales d'une protéine de 20 kDa.

L'expérience $H(CCO)NH$ permet d'obtenir l'ensemble des résonances 1H de la chaîne latérale du résidu précédent. Par ailleurs, l'expérience $C(CO)NH$ fournit l'ensemble des carbones de la chaîne latérale du résidu précédent. En couplant ces deux expériences, il devient possible d'analyser l'expérience $HCCH$ -TOCSY qui permet de relier chaque proton à chaque carbone d'une chaîne latérale. Le résultat de cette dernière expérience contient la même information qu'une expérience 3D TOCSY-HSQC éditée ^{13}C . Toutefois cette expérience est plus sensible car le transfert $HCCH$ -TOCSY est bien plus efficace qu'un transfert par les hydrogènes. La Figure III-7 montre les déplacements moyens observés les protons aliphatiques.

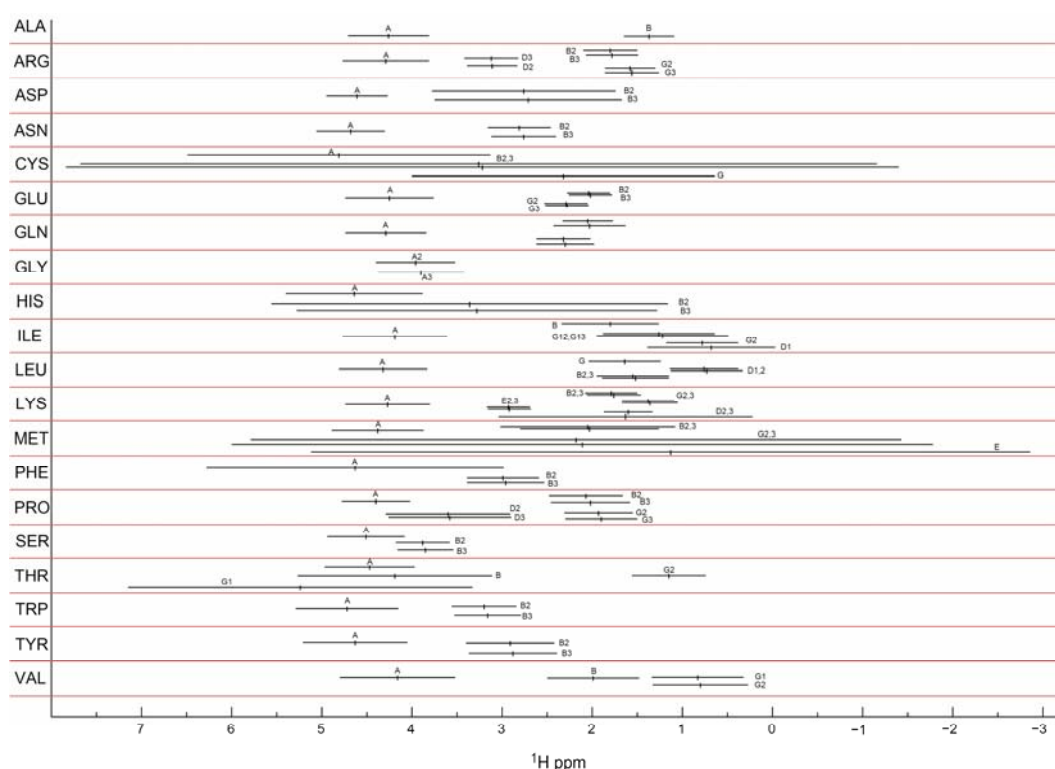


Figure III-7 : déplacements moyens observés pour les protons aliphatiques. (Source : BMRB).

2.3. Paramètres spectroscopiques apportant des informations structurales

Le développement considérable des techniques de la RMN au cours des dernières années en a fait une méthode efficace pour la détermination des structures tridimensionnelles des molécules biologiques. La RMN ne permet pas de déterminer directement la structure 3D de la molécule. Après attribution de l'ensemble des noyaux observables, les paramètres

expérimentaux recueillis sont les intensités des pics NOE obtenus à partir d'une carte NOESY (Nuclear Overhauser Spectroscopy) et les constantes de couplages $^3J_{HH}$. Ces données expérimentales doivent ensuite être converties en paramètres structuraux : distances inter protons d_i (protons distants de moins de 5 Å) et angles dièdres.

2.3.1. Utilisation de l'index de déplacement chimique (Chemical Shift Index, CSI) pour déterminer les structures secondaires

Wishart et ses collaborateurs ont observé que le déplacement chimique moyen des H_α des résidus impliqués dans une structure secondaire s'éloigne des valeurs moyennes observées sur des acides aminés en pelote statistique (Wishart *et al.*, 1991). Cette déviation dépend du type de structure secondaire dans laquelle les résidus sont impliqués : dans une hélice α , les fréquences de résonance des H_α se trouvent en moyenne décalées vers les basses fréquences (variation de déplacement chimique de -0,39 ppm par rapport au random coil), dans un brin β , le décalage se fait en moyenne vers les hautes fréquences ($\Delta\delta = +0,37$ ppm par rapport au random coil). Pour l'identification des structures secondaires d'une protéine à partir des déplacements chimiques des H_α , Wishart a proposé l'utilisation des index de déplacement chimique (CSI). Le principe est d'associer à chaque résidu un nombre (-1, 0, +1) en fonction de l'écart du déplacement chimique de son H_α par rapport au déplacement chimique du même résidu en pelote statistique :

- 0 si le déplacement chimique observé est compris dans un intervalle de ± 0.1 ppm autour du random coil : $|\delta_{\text{obs}} - \delta_{\text{RC}}| < 0.1$ ppm
- +1 si le déplacement chimique observé est plus grand que la borne supérieure de l'intervalle : $\delta_{\text{obs}} - \delta_{\text{RC}} > 0.1$ ppm
- -1 si le déplacement chimique observé est plus petit que la borne inférieure de l'intervalle : $\delta_{\text{obs}} - \delta_{\text{RC}} < -0.1$ ppm

Chaque index indique donc la propension d'un acide aminé à être structuré en hélice α (-1) ou en brin β (+1) ou autre (0). Une succession de 4 index "-1" ou plus, non interrompue par un "+1", indique une hélice α . Une succession de 3 "+1" ou plus non interrompue par un "-1", détermine la présence d'un brin β . Les autres régions ne sont pas

impliquées dans une structure secondaire canonique. Les extrémités des structures secondaires sont identifiées par un index opposé ou par la succession d'index nuls.

Cette technique a ensuite été étendue aux $C\alpha$, $C\beta$ et CO (Wishart et Sykes, 1994), l'index étant défini de la même façon que pour les $H\alpha$. La procédure est identique pour les $C\alpha$ et CO en inversant le signe des index. Le cas des $C\beta$ est différent dans la mesure où l'index associé à ce noyau ne peut être corrélé à l'identification des hélices α , la procédure précédente reste cependant valable pour l'identification des brins β , en utilisant le même signe d'index que pour le $H\alpha$.

2.3.2. Mesure des distances à partir des NOE

L'effet Overhauser nucléaire (NOE) représente la principale source d'information pour la détermination structurale en RMN. Il repose sur le transfert d'aimantation entre 2 noyaux proches dans l'espace (corrélation dipolaire). En première approximation, à faible temps de mélange, le volume η_i d'un pic de corrélation NOESY entre 2 protons est relié à la distance d_i séparant ces 2 protons par la relation (Abragam, 1961) :

$$\eta_i = \frac{K}{d_i^6} \quad [III-10]$$

où K est une constante qui dépend du temps de corrélation τ_c de la molécule.

En raison de la décroissance en $1/d^6$ du signal de corrélation, seules les informations à courte distance (inférieure à $5\text{\AA}$) sont mesurables. Les effets NOE fournissent donc des informations très locales. L'utilisation d'un grand nombre de distances permet de reconstruire la structure de la molécule, et il est donc important de recueillir un maximum de données pour affiner la structure de la molécule.

L'utilisation des couplages résiduels dipolaire permet d'obtenir des informations à plus longue distance. Nous n'y avons pas eu recours, cette technique ne sera donc pas développée dans ce manuscrit.

2.3.3. Mesure des angles dièdres à partir des constantes de couplage scalaire

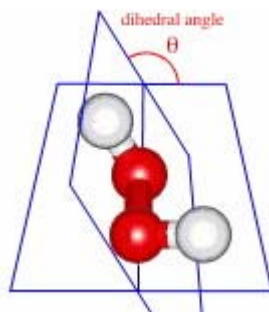


Figure III-8 : Définition de l'angle de torsion θ .

Les constantes de couplage scalaires 3J entre protons séparés par 3 liaisons covalentes permettent d'obtenir une estimation des angles dièdres. Ces constantes de couplage sont en effet reliées à l'angle de torsion θ (Figure III-8) autour de la liaison centrale par l'intermédiaire de la formule Karplus (1959) :

$$^3J(\theta) = A\cos^2 \theta + B\cos \theta + C \quad [III-11]$$

où, A, B et C sont des constantes empiriques dont la valeur dépend des noyaux impliqués dans le couplage.

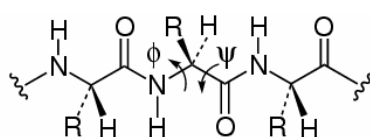


Figure III-9 : Définition des angles de torsion phi (Φ) et psi (Ψ) du squelette peptidique.

La constante de couplage $^3J_{\text{HN-H}\alpha}$ fournit une mesure de l'angle dièdre Φ (Figure III-9) du squelette peptidique des protéines. Plusieurs jeux de valeurs des coefficients A/B/C de la relation de Karplus où ($\theta = |\Phi - 60^\circ|$) ont été proposés par différents auteurs : A/B/C = 6.51/-1.76/1.60 (Vuister *et al.*, 1993); A/B/C = 6.41/-1.46/1.9 (Pardi *et al.*, 1984), et A/B/C = 6.98/-1.38/1.72 (Wang et Bax, 1996). Des valeurs de l'angle Φ aux

alentours de -60 et -120 degrés sont caractéristiques des structures secondaires en hélice α et feuillet β respectivement (Figure III-10).

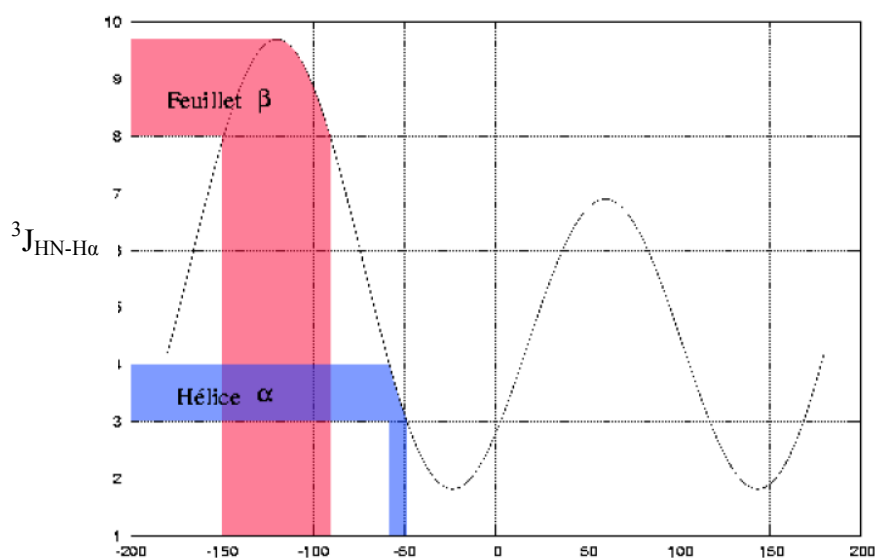


Figure III-10: Courbe de Karplus reliant la constante de couplage $^3J_{\text{HN-H}\alpha}$ à la valeur de l'angle dièdre Φ ($A, B, C = 6.41, -1.46, 1.9$ (Pardi et al., 1984)). Les intervalles pour la constante de couplage des hélices α droites et des feuillets β canoniques sont indiqués en fonction des valeurs angulaires permises.

Différentes méthodes permettent de mesurer ces constantes de couplages (mesure directe sur les cartes COSY, mesures déduites des cartes NOESY à l'aide du programme INFIT (Ludvigsen et al., 1991)). Dans notre travail, disposant d'une protéine marquée ^{15}N , la constante de couplage $^3J_{\text{HN-H}\alpha}$ a été déterminée expérimentalement à partir de l'expérience HNHA (Vuister et Bax, 1993). Cette expérience est largement utilisée lors de l'étude RMN de la structure 3D de protéines simplement marquées ^{15}N . Elle met en corrélation les résonances $^1\text{H}_\text{N}$, ^{15}N et $^1\text{H}_\alpha$ d'un même résidu. Dans le plan ^{15}N et au niveau du déplacement chimique d'un $^1\text{H}_\text{N}$, nous observons un pic positif sur la diagonale et une corrélation croisée négative au déplacement chimique d'un $^1\text{H}_\alpha$ (Figure III-11).

$\delta N = 121.08 \text{ ppm}$

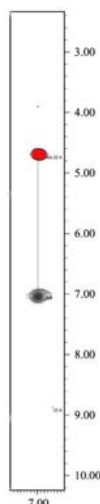


Figure III-11 : Extrait du spectre HNHA de l'ubiquitine présentant le résidu isoleucine 30. L'abscisse correspond aux déplacements chimiques 1H_N , l'ordonnée aux déplacements chimiques 1H . Le pic négatif ($^1H_\alpha$) est indiqué en rouge, le pic positif (diagonale $^1H_N/^1H_N$) en noir. (Source : BMRB)

Le rapport d'intensité de ces deux pics est relié à la constante de couplage $^3J_{HN-H\alpha}$ par la relation :

$$\frac{I_c}{I_d} = -\tan^2(2\pi(^3J_{HN-H\alpha})\zeta) \quad [III-12]$$

I_c et I_d correspondent respectivement aux intensités des pics de corrélation croisée et de la diagonale. ζ est un délai calculé à partir de la séquence d'impulsion de l'expérience 3D HNHA (Vuister et Bax, 1993). La courbe $^3J_{HN-H\alpha}$ en fonction de $\left| \frac{I_c}{I_d} \right|$ permet d'obtenir les constantes de couplage $^3J_{HN-H\alpha}$ (Figure III-12) :

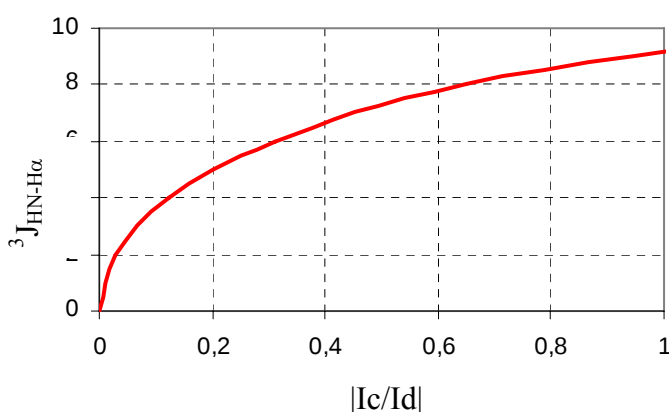


Figure III-12 : Courbe $^3J_{HN-H\alpha}$ en fonction de $\left| \frac{I_c}{I_d} \right|$. I_c et I_d , qui correspondent respectivement aux intensités des pics de corrélation croisée et de la diagonale, sont mesurées sur le spectre HNHA.

Il faut signaler que la présence de deux protons H_α dans les glycines ne permet pas de conclure sur la définition de l'angle dièdre Φ pour ces résidus. Par ailleurs il faut souligner que l'analyse n'est valable que si les signaux des pics de corrélation croisée et de la diagonale sont parfaitement résolus et sans superposition avec des signaux voisins pour éviter toute erreur dans la mesure des intensités. Cette méthode d'analyse conduit généralement à une sous estimation des constantes de couplage $^3J_{HN-H\alpha}$. Vuister et Bax ont proposé d'appliquer un facteur correctif global de 1,11 à l'ensemble des couplages mesurés. Cette correction appliquée à la nucléase du staphylocoque (SNase) est justifiée pour les résidus présentant peu de mobilité interne (Vuister *et al.*, 1993).

D'autres méthodes combinées à celle-ci permettent en outre d'estimer les angles de torsion Ψ (voir III.2.4.1.a.ii : Obtention des angles de torsion, page 99).

2.3.4. Vitesse d'échange des protons avec le solvant pour déterminer les liaisons hydrogène

La détermination des protons amide impliqués dans une liaison hydrogène a été réalisée à partir d'expériences d'échange de protons après dissolution de la protéine dans un solvant deutéré. Les protons amide exposés au solvant s'échangent rapidement avec les atomes de deutérium et leurs signaux disparaissent donc rapidement du spectre. Les protons engagés dans une liaison hydrogène ou inaccessibles au solvant s'échangent lentement : leurs signaux persistent plus longtemps. Expérimentalement, les signaux des protons amide de la protéine sont suivis en réalisant des spectres 1H - ^{15}N HSQC successifs rapidement après la mise en solution de l'échantillon dans le solvant deutéré.

2.3.5. Informations sur l'orientation (RDC)

Les protéines et les acides nucléiques, tout comme leurs complexes, sont constitués de différents domaines dont la structure peut être déterminée séparément. Toutefois, la connaissance de la position relative de ces éléments reste souvent peu précise. Une solution à ce problème consiste à exploiter les couplages dipolaires résiduels (RDCs) mesurés par des techniques de RMN en solution, en utilisant des échantillons faiblement orientés. En s'appuyant sur une approche basée sur des fragments rigides, l'orientation

relative des différents éléments structuraux peut être optimisée afin d'être en accord avec les données de RDCs.

2.4. Modélisation moléculaire à partir des données RMN.

La modélisation de la PEBP CG7054 de drosophile a été initiée en utilisant le logiciel ARIA, puis poursuivie, pour des raisons d'efficacité, en utilisant le logiciel CYANA. Ce dernier logiciel est en effet beaucoup plus performant sur le plan des temps de calcul et pour l'attribution *ab initio* des cartes NOESY. Ces deux logiciels reposent sur l'utilisation des méthodes de recuit simulé à partir des données RMN et combinent calculs de structures et attribution des NOE dans un processus itératif. Les bases de la technique de reconstruction des structures seront rappelées, puis les spécificités du logiciel ARIA seront présentées dont de nombreuses fonctionnalités ont été reprises dans CYANA et sont donc fondamentales. Les particularités de CYANA seront ensuite exposées.

2.4.1. Bases de la modélisation moléculaire à partir des données RMN.

2.4.1.a. Les données RMN utilisées dans la modélisation

2.4.1.a.i. Calibration des données NOE.

Comme nous l'avons déjà remarqué, le paramètre RMN le plus important pour la détermination de la structure des macromolécules est l'intensité des pics NOE obtenue à partir d'une expérience NOESY.

L'effet NOE est lié aux interactions dipôle-dipôle entre les moments magnétiques de spin de deux protons voisins dans l'espace et dépend, en première approximation, de l'inverse de la puissance 6^{ème} de la distance entre les deux atomes.

$$\eta_i = K/d_i^6 \quad [III-13]$$

Si on connaît une distance de référence r_{ref} , par exemple la distance entre les protons H_{B1} et H_{B2} des chaînes latérales des acides aminés dans une protéine, on pourra obtenir une estimation de la distance d_i en utilisant la relation :

$$d_i = d_{ref} (\eta_{ref}/\eta_i)^{1/6} \quad [III-14]$$

où η_{ref} est l'intensité NOE du pic de référence.

Cette étape dite de « calibration » des NOE est essentielle pour réaliser la modélisation. Cependant, compte tenu de l'incertitude sur la mesure de l'intensité et de la dynamique interne de certaines régions de la molécule qui peut entraîner une diminution de l'effet NOE, on utilisera des intervalles de valeurs permises pour les distances inter protons : $d_{min} < d_i < d_{max}$.

2.4.1.a.ii. Obtention des angles de torsion

Une estimation des angles de torsion du squelette peptidique est obtenue :

- Des valeurs de couplage $^3J_{HNH\alpha}$ déduites de l'expérience HNHA comme vu au paragraphe 2.3.3.
- D'autres méthodes qui combinent différents types d'informations : HABAS, TALOS

Là encore, les valeurs des angles de torsion ne sont pas déterminées de façon précises et on définira un intervalle de valeurs possibles pour cet angle dièdre : $\phi_{min} < \phi < \phi_{max}$

Certains logiciels comme HABAS (Günter P. et al 1989), permettent de combiner données NOEs et constantes de couplage pour estimer l'ensemble des angles de torsion (ϕ, ψ) du squelette. De même, le logiciel TALOS, que nous avons utilisé, permet d'obtenir un intervalle de valeurs possibles pour les angles (ϕ, ψ) à partir des valeurs des déplacements chimiques de cinq atomes d'un résidu: $H\alpha$, $C\alpha$, $C\beta$, CO, N (Cornilescu *et al.*, 1999). TALOS considère chaque résidu dans l'environnement de ses 2 voisins et compare, pour chaque tripeptide, les déplacements chimiques mesurés à ceux de tripeptides dans une banque de données. Cette banque, actualisée récemment, regroupe les données de 78 protéines de structures connues, représentant environ 9000 triplets d'acides aminés. Les dix tripeptides comportant à la fois une bonne similarité de séquence et les déplacements chimiques les plus proches de ceux de la protéine cible sont conservés. Les angles ϕ et ψ correspondants sont représentés sur une carte de Ramachandran. L'utilisateur valide le résultat : si dans au moins 9 cas sur 10, ceux-ci se situent au niveau de la même région favorable du diagramme de Ramachandran, la prévision est considérée comme satisfaisante.

La méthode de TALOS, bien que particulièrement efficace, demeure empirique et il a été montré que les prédictions étaient fausses pour 3% des angles (Cornilescu *et al.*, 1999).

2.4.1.a.iii. Autres données

Certains logiciels comme CNS et ARIA permettent l'utilisation directe des constantes de couplage, sans passer par la conversion en valeurs angulaires. Il est également possible d'introduire en données les déplacements chimiques qui peuvent être recalculés de façon empirique à partir de la structure (Williamson MP *et al* 1995, Asakura T *et al* 1995), en particulier les déplacements chimiques du ^{13}C dont le calcul est considéré comme le plus fiable (Oldfield, 1995). Enfin, les liaisons hydrogène dont l'existence a été établie à partir des expériences d'échange des protons amide avec les protons du solvant sont souvent utilisées sous forme de contraintes de distance lors de la modélisation.

Dans ce travail, seules les données NOE et les données relatives aux angles de torsion ont été prises en compte pour la reconstruction des structures.

2.4.1.b. La modélisation par recuit simulé.

Le principe de la modélisation est de générer des structures compatibles avec les données expérimentales et qui présentent une géométrie covalente réaliste, en essayant d'explorer au mieux l'espace des conformations. La meilleure technique à l'heure actuelle pour explorer l'espace conformationnel est le recuit simulé. Il s'agit d'une méthode de dynamique moléculaire à haute température et température constante permettant le franchissement de barrières énergétiques importantes. Cette dynamique peut être effectuée dans l'espace des coordonnées cartésiennes ou dans l'espace des angles dièdres (coordonnées internes). Dans le second cas, les variables sont les angles de torsion, les longueurs de liaisons, les angles de valence demeurant fixes. Le nombre de variables est donc bien inférieur au nombre de coordonnées cartésiennes (soit $3N$ pour une molécule de N atomes) et les temps de calcul sont alors notablement réduits.

Pour effectuer un calcul de dynamique moléculaire, il est nécessaire d'utiliser une fonction analytique pour représenter l'énergie potentielle de la molécule. Ici, compte tenu de la nécessité d'améliorer les temps de calcul, les fonctions d'énergie potentielle utilisées sont simplifiées et ne comportent que les termes essentiels :

- un terme d'énergie de contrainte E_c , destiné à prendre en compte les données expérimentales. Généralement cette énergie est dérivée d'un potentiel harmonique, avec des constantes de force définies par l'utilisateur et pouvant varier au cours du recuit simulé.
- un terme de répulsion E_{rep} , destiné à éviter les mauvais contacts stériques :
 - o si $d_i < D$, somme des rayons de van der Waals des atomes mis en jeux, le paramètre s , $0 < s < 1$, permet de faire varier le rayon de van der Waals au cours du recuit simulé.

$$E_{rep} = \sum K_{rep} (d_i - sD_i)^2 \quad [III-15]$$

- o $E_{rep} = 0$ dans le cas contraire

Enfin, dans le cas d'une simulation dans l'espace des coordonnées cartésiennes, un terme est ajouté pour maintenir la géométrie covalente de la molécule au voisinage des valeurs standards. Les constantes de force intervenant dans les fonctions harmoniques sont augmentées par rapport aux valeurs des champs de force réalistes afin d'assurer le maintien de la géométrie covalente aux températures élevées (donc irréalistes) auxquelles sont effectuées les simulations.

Fonction cible

L'ensemble des logiciels développés par l'équipe de Wüthrich puis P. Günter (DIANA, DYANA et CYANA), utilise la notion de fonction cible V plutôt que celle d'énergie potentielle :

$$V = \sum_{c=u,l,v} w_c \sum_{(\alpha,\beta) \in l_c} (d_{\alpha\beta} - b_{\alpha\beta})^2 + w_a \sum_{i \in l_a} \left[1 - \frac{1}{2} \left(\frac{\Delta_i}{\Gamma_i} \right)^2 \right] \Delta_i^2 \quad [III-16]$$

Si la contrainte de distance est vérifiée entre les atomes α et β séparés d'une distance $d_{\alpha\beta}$, alors la contribution est nulle, sinon l'expression est calculée. $b_{\alpha\beta}$ correspond alors soit aux contraintes supérieures (u), inférieures (l), ou de van der Waals (v) s'il n'y a pas de contrainte expérimentale.

Δ_i représente la violation de contrainte angulaire et $\Gamma_i = \pi - (\phi_{\max} - \phi_{\min})/2$, demi-intervalle de valeurs interdites pour un angle de torsion ϕ_i . Les facteurs w sont des facteurs de pondération et V s'exprimera en $\text{\AA}^2$.

Afin d'éviter les mauvais contacts stériques, pour les couples d'atomes pour lesquels il n'y a pas de données NOE, on choisit $d_{\min} = D_i$, somme des rayons de van der Waals des 2 atomes et $d_{\max}$ est fixé arbitrairement à 999 $\text{\AA}$.

Les étapes de recuit simulé sont ensuite réalisées en utilisant des méthodes classiques de dynamique moléculaire dans l'espace des coordonnées cartésiennes (logiciel CHARMM (Brooks *et al.*, 1983) pour la suite de logiciels XPLOR (Brunger et Nilges, 1992), CNS (Brunger *et al.*, 1998), ARIA (Linge *et al.*, 2001)) ou des méthodes plus complexes mises au point dans le système des angles de torsion pour CYANA((Guntert *et al.*, 1997), (Stein *et al.*, 1997)). Le protocole de recuit simulé comporte plusieurs étapes durant lesquelles les diverses interactions sont pondérées de façon différentes afin de permettre à la structure de s'adapter aux contraintes expérimentales. Par exemple, les forces de van der Waals sont de type répulsives avec des rayons réduits en début de calcul (paramètre s petit) et qui augmentent au cours des étapes. Le poids des contraintes expérimentales augmente aussi au fur et à mesure du calcul. Enfin, les potentiels électrostatiques ne sont pas pris en compte sauf durant l'étape finale réalisée dans une boîte d'eau (ARIA).

2.4.2. Le logiciel ARIA

ARIA est un logiciel d'attribution semi-automatique des pics NOE et de calcul de structures qui fonctionne en interface avec le logiciel CNS, lui-même dérivé du logiciel XPLOR développé par Brunger et Nilges (Brunger et Nilges, 1992). ARIA intègre le concept de contrainte de distance ambiguë (ambiguous distance restraint ou ADR) qui permet de prendre en compte toutes les attributions possibles pour un pic de corrélation. Le logiciel ARIA a été développé dans le but de réduire de façon itérative le nombre de paires contribuant à chaque NOE observé, en éliminant les paires ne contribuant pas physiquement à l'intensité. L'utilisation des contraintes de distance ambiguë permet de s'affranchir des corrections de pseudo-atomes car les protons équivalents (méthyles, protons aromatiques) sont traités comme des NOE ambigus. Enfin l'attribution

stéréospécifique est laissée flottante durant la procédure par échange des déplacements chimiques des hydrogènes géminaux ou des méthyles de valine ou isoleucine.

2.4.2.a. Contraintes de distance ambiguë

Avec la notion d'ADR, un pic NOE est traité comme la superposition des signaux de corrélation de plusieurs paires de protons dont les fréquences de résonance sont compatibles avec la position du pic. Ainsi, un pic NOE avec une attribution multiple ($n > 1$, n : nombre d'attributions) est considéré comme la somme de n pics NOE et converti en contrainte de distance ambiguë $\bar{d}$ d'après l'équation suivante :

$$\bar{d} = \frac{1}{\sqrt[n]{\sum_{i=1}^n \frac{1}{d_i^6}}} \quad [III-17]$$

où chaque distance d_i est la distance entre une paire i de protons et n représente le nombre total d'attributions possibles. Cette valeur $\bar{d}$ est toujours inférieure à d_i , quel que soit i , une mauvaise attribution est donc sans influence ; c'est la notion de contribution C_k , avec $1 \leq k \leq n$ (cf. *infra*, 2.4.2.c : Attribution des pics NOE ambigus, page 104).

2.4.2.b. Calibration des pics NOE

Aria prend comme constante de calibration $C=1/K$, où K est la constante qui intervient dans la relation $\eta_i = K/d_i^6$ reliant l'intensité NOE η_i à la distance d_i entre les deux atomes. C est déterminée de façon simplifiée par l'expression :

$$C = \frac{1}{K} = \frac{1}{n} \sum_{NOEs} \frac{d_{obs}^{-6}}{\eta_i} \quad [III-18]$$

avec n le nombre total d'attributions.

Pour chaque paire de protons correspondant à un pic NOE de volume η_i on calcule une distance moyenne $\langle d_i \rangle$ sur l'ensemble des structures générées. On en déduit les contraintes de distance observées (d_i^{obs}) :

$$d_i^{obs} = \frac{1}{(C \times \eta_i)^{1/6}} \quad [III-19]$$

L'erreur sur l'évaluation des contraintes de distance (Δ) est estimée à 12,5 % du carré de la distance observée d_i^{obs} . On définit donc un intervalle de distance faisant intervenir les limites d_{min} et d_{max} :

$$d_{min} = d_i^{obs} - \varepsilon (d_i^{obs})^2 \quad [III-20]$$

$$d_{max} = d_i^{obs} + \varepsilon (d_i^{obs})^2 \quad [III-21]$$

Avec $\varepsilon = 0,125 \text{ \AA}^{-1}$

2.4.2.c. Attribution des pics NOE ambigus

Pour chaque pic NOE ambigu, c'est-à-dire dont une coordonnée de déplacement chimique correspond à la fréquence de résonance de plusieurs protons, une contrainte de distance ambiguë est créée. Le pic NOE est donc la somme de n contributions C_k correspondant aux différentes attributions possibles. Chaque contribution C_k est calculée de la manière suivante :

$$C_k = \frac{d_k^{-6}}{\sum_{i=1}^n d_i^{-6}} \quad [III-22]$$

avec $0 < C_k < 1$ et $\sum C_k = 1$

où d_k est la distance moyenne de la paire de protons k calculée sur les 7 structures de plus basse énergie de l'itération précédente. On classe ensuite les contributions C_k par ordre décroissant et on exclut les contributions C_k de faible poids dès lors que :

$$\sum_{i=1}^{n_p} C_k > p$$

où p constitue la valeur du seuil d'attribution et n_p est le nombre de contributions C_k participant majoritairement à la définition du pic NOE.

La valeur du seuil d'attribution est proche de 1 à la première itération et décroît progressivement au fur et à mesure des itérations pour atteindre la valeur de 0,8 à l'itération 8.

De plus, au fur et à mesure des itérations, le traitement ambigu des pics NOE permet d'exclure de la modélisation des pics dont les contraintes correspondantes sont systématiquement violées sur plus de la moitié des structures analysées. Ces pics exclus sont vérifiés manuellement après chaque cycle itératif et peuvent résulter de situations diverses : problèmes de correction de ligne de base, bruit t1, mauvaise suppression du solvant, diffusion de spin.

L'utilisateur d'ARIA peut régler à volonté tous les paramètres du calcul. Cependant, seuls deux paramètres semblent être réellement dépendants de la protéine : le seuil p (voir ci-dessus) et la tolérance $\Delta\delta_{\text{ppm}}$.

2.4.2.d. Principe de fonctionnement du logiciel ARIA

L'utilisation simultanée des logiciels ARIA et CNS permet de réaliser l'attribution des pics NOE et le calcul des structures 3D de manière itérative (Figure III-13) :

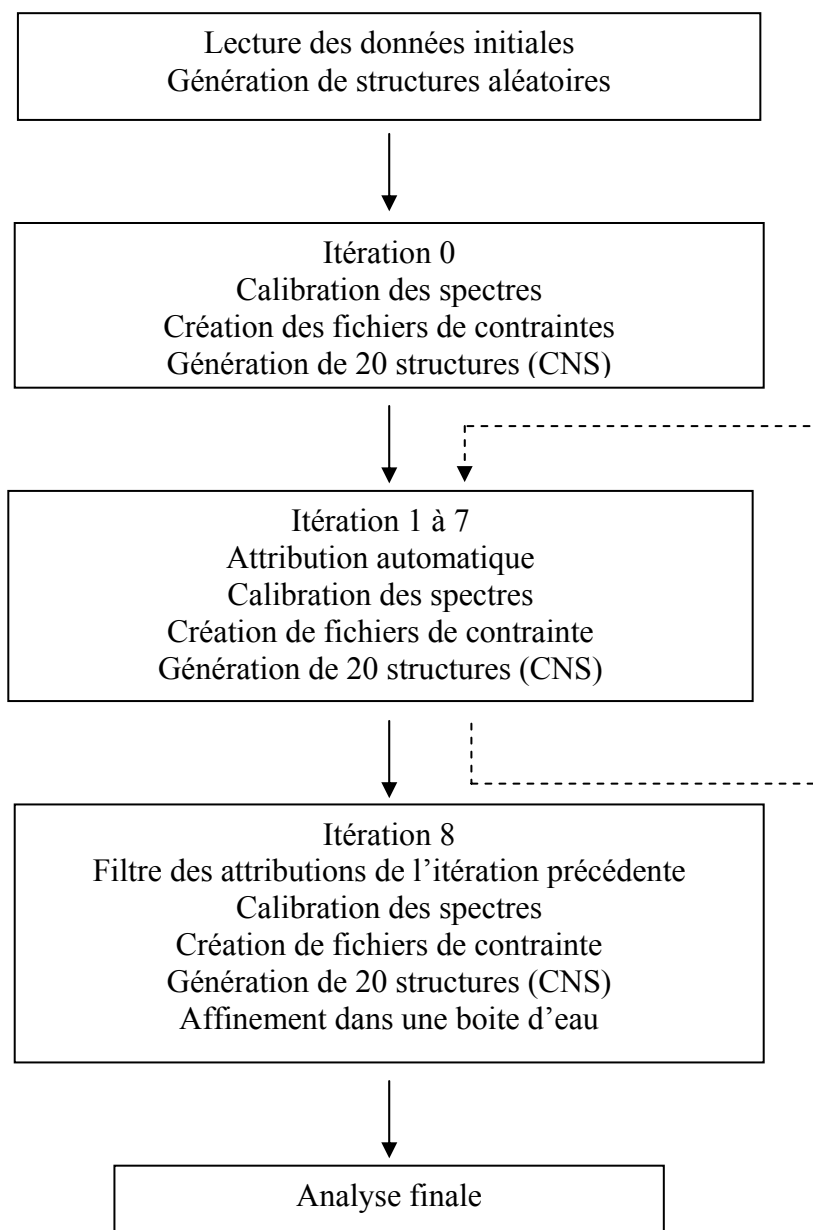


Figure III-13: Description simplifiée du processus d'attribution et de calcul de structure itératif ARIA/CNS.

Les données initiales comportent les tables de déplacements chimiques et les listes de pics NOE, éventuellement partiellement attribués ainsi que les angles dièdres connus. Pour chaque pic NOE, un intervalle de tolérance $\Delta\delta_{\text{ppm}}$ sur les déplacements chimiques permet la détermination des n attributions possibles.

A chaque itération :

1. Les x structures de plus basse énergie de l'itération $i-1$ sont sélectionnées (le nombre x de structures est défini par l'utilisateur, généralement 7).
2. Les distances moyennes d_i de chaque paire de protons sont calculées à partir de ces 7 structures.
3. Les volumes NOE sont convertis en contraintes de distance en utilisant une calibration automatique des spectres.
4. Les contributions les plus faibles aux contraintes ambiguës sont éliminées en fonction d'un paramètre de sélection.
5. Un ensemble de 20 structures est calculé à partir des nouvelles contraintes.

Le schéma précédent est appliqué jusqu'à l'itération 8 qui inclut une dynamique dans une boîte d'eau utilisant un champ de force réaliste.

2.4.3. Le logiciel CYANA

Le logiciel CYANA, comme ARIA, permet de réaliser simultanément l'attribution des pics NOE et le calcul de structure 3D. Ce logiciel est basé sur la suite de programmes DIANA (Guntert *et al.*, 1991; Guntert et Wuthrich, 1991) puis DYANA (Guntert *et al.*, 1997) qui introduit la notion de dynamique dans l'espace des angles de torsion. Les calculs font intervenir les notions de fonction cible et de contraintes ambiguës. Globalement, CYANA utilise un processus itératif voisin de celui d'ARIA. L'étape finale de raffinement dans une boîte d'eau, qui utilise un champ de forces réaliste, n'existe pas dans CYANA. Enfin, de nouvelles fonctionnalités très efficaces sont utilisées dans CYANA, par exemple pour procéder aux attributions et pondérer ces attributions au travers de l'algorithme CANDID (Herrmann *et al.*, 2002).

Principe de l'algorithme CANDID de CYANA

Les données initiales sont identiques à celle d'ARIA : tables de déplacements chimiques et listes des NOEs. A chaque itération, l'attribution est effectuée sur la base des déplacements chimiques en utilisant un intervalle de tolérance et utilise le jeu de structures générées à l'itération précédente (sauf à la 1^{ère} itération). Ces attributions sont également pondérées par l'utilisation de la notion de « Network Anchoring » qui vérifie la cohérence des attributions et leur compatibilité avec la géométrie covalente (Figure III-14). Un score N_{AB} est calculé pour chaque attribution d'un NOE à une paire d'atomes

(A, B), en considérant toutes les connectivités mettant en jeu les atomes A et B et éventuellement un 3^{ème} atome C. Tous ces NOEs doivent être compatibles sur le plan géométrique et doivent bien sûr être compatibles avec la géométrie covalente. N_{AB} est calculé en utilisant des expressions analytiques assez complexes (Herrmann *et al.*, 2002).

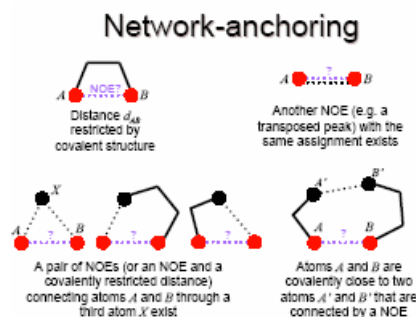


Figure III-14 : Chaque attribution doit être confirmée par l'existence d'autres données expérimentales compatibles avec elle et pouvant correspondre à l'une des 4 situations illustrées ici (d'après Guntert, cours EMBO 2003) (<http://www.embl-heidelberg.de/nmr/sattler/embo/>). En violet la connectivité évaluée, en noir les autres connectivité faisant intervenir les atomes A et B.

Trois critères sont pris en compte pour qu'une attribution soit validée (Figure III-15). Les attributions ayant une N_{AB} trop faible sont éliminées.

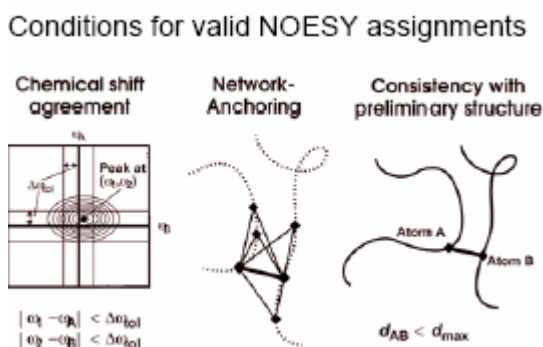


Figure III-15: Conditions pour qu'une attribution soit prise en compte (d'après Guntert).

Les contraintes ainsi validées sont calibrées et utilisées dans les calculs de structure. Toutefois, des contraintes erronées peuvent provenir de divers artefacts spectraux (bruit, pics en troncature...) confondus avec des pics de corrélation. Ceci peut se produire en

particulier pour les NOE longue distance lorsqu'aucune structure 3D n'est disponible pour filtrer les attributions. Pour minimiser leur impact, lors des deux premières itérations, un processus de combinaison de contraintes est utilisé (Figure III-16). Deux modes sont possibles : dans le premier, deux contraintes longues distances sont remplacées par une seule contrainte ambiguë ; dans le second, 4 contraintes longues distances sont remplacées par 4 contraintes ambiguës combinant ces contraintes deux à deux. Dans ce cas, la contrainte erronée n'aura pas d'influence sur le calcul de structure.

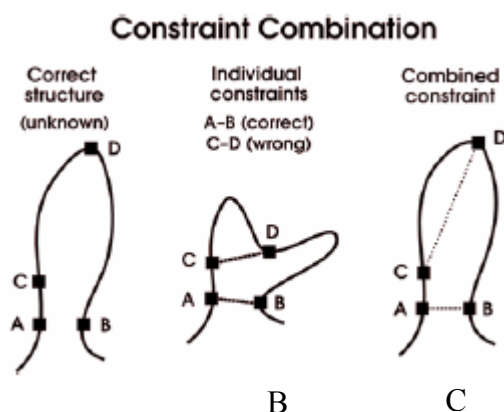


Figure III-16 : combinaison de 2 contraintes (Guntert). Effet de la contrainte erronée entre C et D ; si elle est prise individuellement (B), la structure est déformée car les contraintes sont satisfaites simultanément ; si la contrainte erronée entre dans une combinaison de contraintes sous forme de contrainte ambiguë (C), la structure est correcte.

Lors de la dernière itération, contrairement à ARIA, un filtre supplémentaire ne conserve que les NOE ayant une attribution unique à une paire de protons.

Pour la calibration, plusieurs options sont possibles. Elle peut être effectuée manuellement par l'utilisateur ou automatiquement en utilisant ou non un jeu de structures (Herrmann *et al.*, 2002).

Le logiciel ARIA est efficace pour compléter l'attribution des cartes NOESY lorsqu'un repliement initial de la protéine est déjà connu ou lorsque qu'une partie de la carte a été attribué manuellement. L'algorithme CANDID, intégré à CYANA, s'avère extrêmement efficace dans l'attribution *ab initio* des cartes NOESY grâce à l'utilisation du « Network anchoring » et des combinaisons de contraintes. Cette méthode a été évaluée dans des

calculs test pour l'attribution *de novo* et la détermination de structures de protéines et a fourni d'excellents résultats (Herrmann *et al.*, 2002).

2.5. Etude de la dynamique d'une protéine par la relaxation ^{15}N

Une des possibilités les plus intéressantes offertes par la RMN est sa capacité à sonder les mouvements moléculaires au niveau atomique, sur une gamme de temps très large. En effet, les spins nucléaires ont des propriétés remarquables pour analyser la dynamique interne des molécules. Ils sont sensibles à leur environnement proche et leur relaxation s'étale sur une gamme de fréquences très large qui traduit celle des mouvements des macromolécules. Depuis le développement du marquage isotopique des protéines avec les noyaux ^{15}N et ^{13}C , les études de dynamique par relaxation nucléaire se sont multipliées afin de comprendre la relation entre la structure, la dynamique et la fonction d'une protéine. La dynamique interne des protéines est étudiée essentiellement à partir de la relaxation des spins nucléaires ^{15}N ou ^{13}C qui portent un seul proton. Les informations sur la dynamique sont obtenues par la mesure de 3 paramètres expérimentaux : les temps de relaxation longitudinale T_1 et transversale T_2 , et la relaxation croisée NOE. Les noyaux ^{15}N constituent des sondes idéales : ils sont présents de manière homogène tout le long de la chaîne principale et leur source de relaxation principale provient du proton qui leur est attaché.

2.5.1. Les paramètres T_1 et T_2

Après excitation des spins par une impulsion radiofréquence, le système retourne à l'équilibre par le phénomène de relaxation. Ce retour à l'équilibre peut être décrit par une loi mono-exponentielle avec des temps caractéristiques T_1 et T_2 .

$$M_z(t) = M_{z_0} \cdot e^{\frac{-t}{T_1}}$$

$$\text{Et } M_{xy}(t) = M_{xy_0} \cdot e^{\frac{-t}{T_2}}$$

où $M_z(t)$ et $M_{xy}(t)$ sont les composantes longitudinale et transversale de l'aimantation au temps t , M_{z0} et M_{xy0} sont ces composantes au temps $t=0$.

2.5.1.a. La relaxation longitudinale

La relaxation longitudinale est liée aux échanges d'énergie entre les spins et leur environnement, elle décrit le retour à l'équilibre de l'aimantation M_z le long de l'axe z , associé au passage des spins de l'état β à l'état α (et réciproquement). Cette relaxation est favorisée par des fluctuations de champ magnétique à haute fréquence qui fluctuent à des fréquences ω de l'ordre de ω_0 . Le temps de relaxation longitudinale T_1 est donc minimum pour $\tau_c = 1/\omega_0$ (Figure III-17).

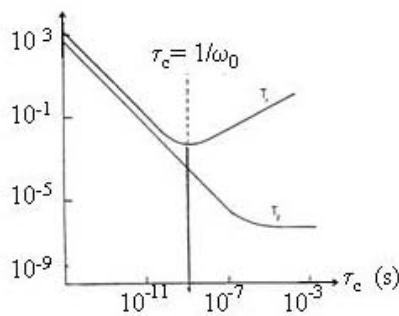


Figure III-17 : Variations des temps de relaxation T_1 et T_2 en fonction du temps de corrélation τ_c .

2.5.1.b. La relaxation transversale

La relaxation transversale provient d'un échange d'énergie entre les spins sans interaction avec le réseau. Elle décrit le retour à zéro de l'aimantation transversale $M_{xy}(t)$ située dans le plan xy perpendiculairement au champ magnétique B_0 . La vitesse de relaxation transversale correspond au déphasage des aimantations nucléaires sous l'effet des fluctuations du champ local à toutes les fréquences : les mouvements « lents » (devant les fréquences de transition RMN) y contribuent aussi bien que les mouvements haute fréquence. T_2 est inférieur à T_1 puisqu'un plus grand nombre de phénomènes contribuent au T_2 . La courbe décrivant les variations de T_2 en fonction de τ_c est décroissante,

indiquant que la relaxation est d'autant plus efficace que les mouvements sont lents (Figure III-17).

2.5.1.c. *Mesure des paramètres T_1 et T_2*

La mesure de ces paramètres se fait par la mesure des volumes ou des intensités des pics de corrélation dans des expériences enregistrées avec des temps de relaxation tr différents.

Les séquences d'impulsions pour mesurer les temps de relaxation longitudinale T_1 et transversale T_2 sont basées sur la séquence d'impulsions ^1H - ^{15}N HSQC. Cette séquence commence par la série INEPT pour transférer l'aimantation du proton vers le noyau ^{15}N . Ensuite une série d'impulsions amène le système hors de son état d'équilibre, dans l'un des états dont on veut mesurer la relaxation. Suit un délai tr de relaxation, avec une série d'impulsions qui permettent de conserver le simple processus de relaxation en supprimant les effets de relaxation croisée dipôle-dipôle et d'anisotropie de déplacement chimique. Enfin l'aimantation du noyau ^{15}N est traitée afin de transférer sa polarisation vers le proton à travers une série INEPT-inverse. Une série d'expériences est enregistrée en faisant varier le délai tr afin d'échantillonner au mieux la courbe de relaxation.

On obtient donc une série de spectres de type ^1H - ^{15}N HSQC pour les différentes valeurs de tr . L'intensité des pics croisés décroît de manière exponentielle et tend vers zéro quand les temps de relaxation tendent vers l'infini. Les valeurs de R_1 et R_2 sont déterminées en réalisant un ajustement de l'intensité des pics croisés à l'aide d'une fonction monoexponentielle décroissante :

$$I(tr) = I_0 e^{\left(\frac{-tr}{T}\right)} \quad \text{[III-23]}$$

où $I(tr)$ est l'intensité du pic croisé au délai de relaxation tr et I_0 l'intensité au temps $tr = 0$ et T correspond au temps de relaxation T_1 ou T_2 .

L'optimisation des valeurs de I_0 et T est obtenue en minimisant la fonction cible χ_2 dont l'expression est la suivante :

$$\chi^2 = \sum_{i=1}^N \frac{(I(tr)_{calc} - I(tr)_{exp})^2}{\sigma_I^2} \quad [III-24]$$

où $I(tr)_{calc}$ et $I(tr)_{exp}$ sont respectivement les intensités des pics calculées et mesurées expérimentalement, et σ_I correspond à l'incertitude de mesure sur l'intensité du pic. Plus la valeur du χ^2 est faible, plus la correspondance entre le modèle et les valeurs expérimentales est correcte.

Les incertitudes sur les valeurs des paramètres de relaxation R_1 et R_2 sont évaluées dans le module *Hetnoe* du logiciel Nmrview par la méthode de Monte Carlo (Press *et al.*, 1986). Les intensités calculées grâce à l'équation III-23 sont perturbées aléatoirement selon une distribution normale avec un écart type correspondant à celui de l'erreur sur la mesure. Un échantillonnage statistique de 200 vitesses de relaxation par acide aminé a été obtenu et a permis d'aboutir à une vitesse de relaxation moyenne avec un écart type.

2.5.2. Le paramètre NOE

Comme nous l'avons déjà vu, l'effet NOE correspond à un échange d'aimantation entre deux spins (I (par exemple le proton) et S (par exemple l'azote ^{15}N)) proches dans l'espace lorsque les populations d'un des deux spins (par exemple I) sont perturbées par rapport à l'équilibre. Il en résulte une modification de l'intensité de la résonance du spin S. Pour rappel, l'effet NOE (ou facteur d'amplification) peut être déterminé grâce à l'équation ci-dessous :

$$NOE_{IS} = \frac{S_{sat} - S_{eq}}{S_{eq}} = \frac{S_{sat}}{S_{eq}} - 1 \quad [III-25]$$

Avec S_{eq} et S_{sat} les intensités du pic de résonance du spin S à l'équilibre et pendant la saturation du spin I.

D'un point de vue théorique, le NOE peut être défini comme le rapport de la vitesse de relaxation croisée σ_{IS} sur la vitesse de relaxation dipolaire longitudinale totale ρ_{IS} :

$$NOE_{IS} = \frac{\gamma_I}{\gamma_S} \frac{\sigma_{IS}}{\rho_{IS}} = \frac{\gamma_I}{\gamma_S} \left[\frac{W_2 - W_0}{W_0 + W_2 + 2W_1^S} \right]$$

Où W_2 et W_0 sont les vitesses des transitions à deux et zéro quanta respectivement et W_1^S est la vitesse de transition à un quantum du spin I.

Comme on peut le constater, le signe du NOE dépend du signe des rapports gyromagnétiques des noyaux impliqués et de la différence ($W_2 - W_0$).

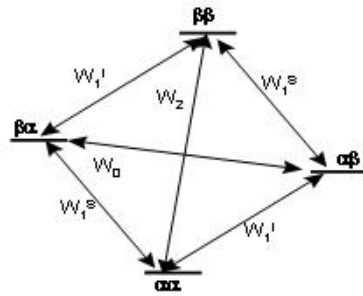


Figure III-18 : Transitions possibles dans un système à deux spins

Les vitesses de transitions W_2 , W_0 et W_1^S (Figure III-18) sont exprimées par :

$$W_0 = \gamma_I^2 \gamma_S^2 \left[\frac{2\tau_c}{r_{IS}^6 (1 + (\omega_S - \omega_I)^2 \tau_c^2)} \right]$$

$$W_2 = \gamma_I^2 \gamma_S^2 \left[\frac{12\tau_c}{r_{IS}^6 (1 + (\omega_S + \omega_I)^2 \tau_c^2)} \right]$$

$$W_1^S = \gamma_I^2 \gamma_S^2 \left[\frac{3\tau_c}{r_{IS}^6 (1 + \omega_S^2 \tau_c^2)} \right]$$

et sont donc dépendantes de la distance r_{IS} et de τ_c , le temps de corrélation de la protéine.

Si la distance r_{IS} est connue (exemple liaison $^1\text{H}-^{15}\text{N} \approx 1.01 \text{ \AA}$), alors le NOE va directement renseigner sur le τ_c du vecteur observé.

Lorsque le NOE hétéronucléaire est observé entre le ^1H (noyau I saturé) et un noyau possédant un rapport gyromagnétique négatif (ex S: ^{15}N), le NOE observé sera négatif et aboutira à une diminution de l'intensité du signal. Le facteur de diminution dépend du τ_c du vecteur, comme on peut le voir Figure III-19 dans le cas du vecteur H-N.

D'un point de vue pratique, pour mesurer cet effet, on enregistre dans un premier temps une expérience où l'on irradie le ^1H pendant quelques secondes afin de rompre l'équilibre et de créer un état stationnaire (expérience de saturation), et où on regarde l'effet de cette perturbation sur l'aimantation longitudinale du spin ^{15}N . Une seconde expérience est enregistrée sans saturation (expérience de référence). En pratique l'irradiation est en fait réalisée hors résonance afin de produire le même échauffement de l'échantillon que dans l'expérience de saturation. Le paramètre NOE est déterminé en faisant le rapport des intensités mesurées dans l'expérience de saturation et dans l'expérience de référence.

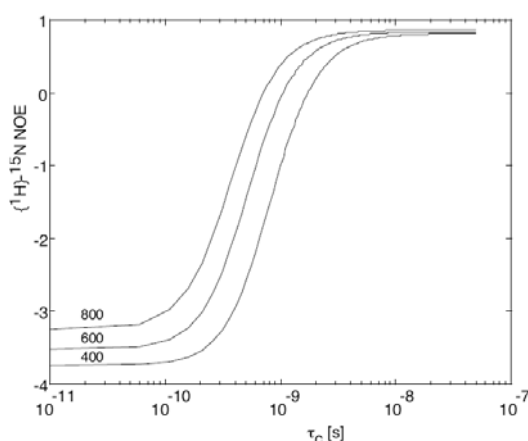


Figure III-19: Variation du paramètre NOE ^1H - ^{15}N en fonction du τ_c du vecteur N-H et du champ statique B_0 . Le NOE hétéronucléaire ^1H - ^{15}N permet d'identifier des mouvements dans la gamme $[10^{-10}; 10^{-8}]$ s.

2.5.3. Extraction des paramètres de dynamique

Les paramètres de dynamique sont déterminés à l'aide de l'approche sans modèle géométrique développée par Lipari et Szabo (Lipari et Szabo, 1982a; Lipari et Szabo, 1982b). Cette approche sépare la diffusion rotationnelle globale de la molécule des mouvements internes plus rapides. Ces deux phénomènes sont considérés comme

indépendants car les échelles de temps de ces mouvements sont différentes d'au moins un ordre de grandeur. Le temps de corrélation de la molécule τ_c se situe dans la gamme des nanosecondes alors que les mouvements internes τ_i vont de la picoseconde à la nanoseconde.

Les paramètres de dynamique sont déterminés à partir de la fonction de densité spectrale $J(\omega)$ (Figure III-20) qui, selon l'approche de Lipari et Szabo, est de la forme :

$$J(\omega) = \frac{2}{5} \left\{ \frac{S^2 \tau_c}{1 + (\omega \tau_c)^2} + \frac{(1 - S^2) \tau}{1 + (\omega \tau)^2} \right\} \quad [2.31]$$

$$\text{avec } \tau = \frac{\tau_c \tau_e}{\tau_c + \tau_e} \text{ soit } \frac{1}{\tau} = \frac{1}{\tau_c} + \frac{1}{\tau_e}$$

où τ_c est le temps de corrélation global de la molécule, τ_e le temps de corrélation effectif et S^2 le paramètre d'ordre qui donne une mesure de la restriction d'amplitude des mouvements internes ($0 < S^2 < 1$; plus S^2 tend vers 1, plus le mouvement de la liaison est restreint). Cette fonction permet de connaître l'amplitude des mouvements des vecteurs étudiés dans la molécule.

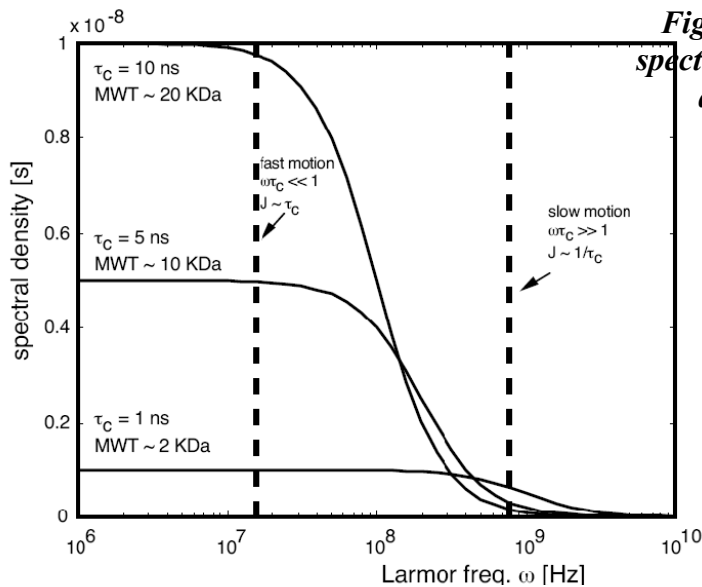


Figure III-20 : Fonction de densité spectrale obtenue en fonction du temps de corrélation de la molécule.

La détermination de cette fonction est possible car les 3 paramètres mesurés, R_1 , R_2 et NOE, l'échantillonnent à 5 fréquences différentes (Peng et Wagner, 1992). En effet :

$$R_1 = d^2 [J(\omega_H - \omega_N) + 3J(\omega_N) + 6J(\omega_H + \omega_N)] + c^2 J(\omega_N)$$

$$R_2 = \frac{d^2}{2} [4J(0) + J(\omega_H - \omega_N) + 3J(\omega_N) + 6J(\omega_H + \omega_N) + 6J(\omega_H)] + c^2 \left[\frac{2}{3} J(0) + \frac{1}{2} J(\omega_N) \right]$$

$$NOE = 1 + \frac{\gamma_H d^2}{\gamma_N R_1} [6J(\omega_N + \omega_H) - J(\omega_H - \omega_N)]$$

$$\text{Avec } d^2 = \frac{\gamma_H^2 \gamma_N^2 \hbar^2}{4r_{NH}^6} \text{ et } c^2 = \frac{(\sigma_{\parallel} - \sigma_{\perp})^2 \omega_N^2}{3}$$

$\hbar$ est la constante de Planck, γ_H et γ_N sont les rapports gyromagnétiques du 1H et de ^{15}N , ω_H et ω_N sont les fréquences de Larmor 1H et ^{15}N et r_{NH} est la distance internucléaire. $\sigma_{\parallel}$ - $\sigma_{\perp}$ sont les composantes parallèle et perpendiculaire du tenseur de déplacement chimique de ^{15}N .

En faisant l'approximation $J(\omega_H \pm \omega_N) \approx J(\omega_H)$ du fait que $\gamma_N \approx -0.1 \gamma_H$, on obtient un système de 3 équations à 3 inconnues que l'on peut résoudre.

Pour une molécule ayant une diffusion isotrope, en l'absence de mouvements internes et d'échange conformationnel, le terme en $(1-S^2)$ tend vers 0 et la fonction de densité spectrale devient :

$$J(\omega) = \frac{2}{5} \left[\frac{S^2 \tau_c}{1 + (\omega \tau_c)^2} \right]$$

Dans ce cas, le rapport R_2/R_1 est indépendant du paramètre d'ordre et permet de déterminer le temps de corrélation global τ_c de la molécule (Kay *et al.*, 1989).

Dans le cas d'une diffusion anisotrope, le traitement est un peu plus compliqué. Il devient nécessaire d'avoir une structure pour déterminer le tenseur de diffusion et extraire les paramètres de dynamique.

La détermination des paramètres de dynamique peut être faite par le programme Model-Free (Lipari et Szabo, 1982a; Lipari et Szabo, 1982b).

Etude de la protéine

CG18594

IV. Etude de la protéine CG18594

Nous avons initié le projet d'étude des PEBP de drosophiles en étudiant la protéine CG18594. Cette cible a été choisie car il existe une lignée de drosophiles qui possède un élément P transposable juste en amont du gène, ce qui permet de créer des lignées de mouches mutantes par délétion de tout ou partie du gène. Cette situation est particulièrement favorable pour établir le rôle d'une protéine car les phénotypes mutants, s'ils existent, fournissent souvent des informations sur le rôle de la protéine sauvage. De plus le gène *cg18594* ne présente pas d'intron, ce qui facilite les étapes de clonage de sa portion codante. Enfin le gène *cg18594* se situe juste à proximité (moins de 150 paires de bases) du gène *cg7054*, qui code lui aussi pour une protéine appartenant à la famille des PEBP. La très courte distance séparant les deux gènes pose de nombreuses questions concernant la régulation de l'expression du gène *cg18594* car il est difficile de concevoir la présence de séquences régulatrices entre les deux. Il semble donc possible que les deux protéines soient liées fonctionnellement. Cette situation n'a pour l'instant été observée que chez la drosophile.

Lors de mon arrivée au laboratoire le clonage dans *Escherichia coli* de la région codante du gène *cg18594* dans le vecteur d'expression pET31b avait déjà été réalisé par Estelle Rapine, étudiante en maîtrise de biologie. J'ai donc commencé par mettre au point le protocole de purification et de marquage de la protéine CG18594. Actuellement, nous pouvons la produire purifiée, marquée ou non, en grande quantité. En vue de l'étude structurale par RMN, j'ai ensuite testé le comportement de la protéine dans différentes conditions de tampons et de sels, et réalisé un début de caractérisation physico-chimique. Ces résultats sont présentés dans l'article publié dans le « *Journal of Protein Expression and Purification* ». Ces études nous ont permis de choisir les conditions les plus favorables à la monodispersion de la protéine et à sa solubilité. Dans ces conditions la protéine produite est soluble et structurée. Cependant elle n'est pas stable plus de quelques jours à une concentration de 1 mM et à 20°C. Elle est stable à 5°C mais les spectres enregistrés ne sont pas exploitables. En utilisant la β -lactoglobuline, qui est une protéine de taille proche de celle de la protéine CG18594 (162 résidus, 18.3 kDa), nous avons montré l'effet majeur de la température sur la qualité des spectres. Nous nous sommes donc heurtés à la non faisabilité de l'étude de la structure tridimensionnelle de la protéine CG18594 par RMN.

1. Données de base

1.1. Analyse de la séquence

La séquence du gène *cg18594* est disponible sur la banque de données Flybase (<http://flybase.bio.indiana.edu/>). Ce gène est situé sur le bras droit du chromosome 3, en position 94A16, à proximité du gène *cg7054* (Figure IV-1).

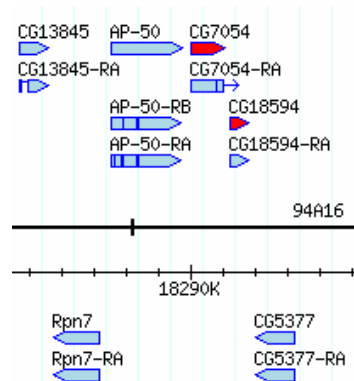


Figure IV-1 : Localisation des gènes *cg7054* et *cg18594* sur le bras droit du chromosome III (image issue de Flybase).

Le gène *cg18594* a une longueur de 598 nucléotides sur l'ADN génomique de la souche de *Drosophila melanogaster* Oregon R (<http://flybase.bio.indiana.edu/>). Il ne présente pas d'intron. La traduction de la portion codante fournit une séquence protéique qui présente la signature caractéristique de la famille des PEBP (Figure IV-2).

1	11	21	31	41	51	
1	MDTAGIIPDI	IDVKPASKAT	ITYPSGVQVE	LGKELTPTQV	KDQPTVVFDA	EPNSLYTILL
61	VDPDAPSRED	PKFRELLHWL	VINIPGNKVS	EGQTIAEYIG	AGPREGSGLH	RYVFLVFKQN
121	DKITTEKFVS	KTSRTGRINV	KARDYIQKYS	FGGPVAGNFF	QAQYDDYVKT	LIETVQ

Figure IV-2 : Séquence de la protéine *CG18594*. La séquence consensus de la famille des PEBP est indiquée en rouge.

L'analyse de cette séquence est présentée dans le tableau ci dessous (outil Protparam du site Expasy, (Gasteiger *et al.*, 2003)) (Figure IV-3).

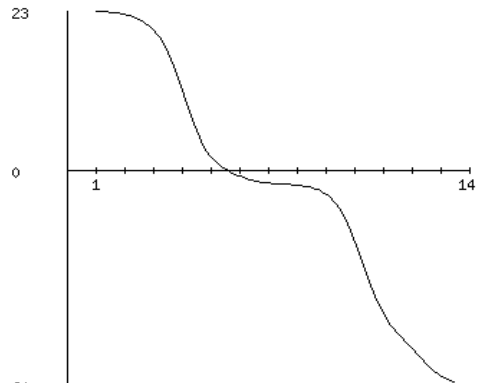
Nombre d'acides aminés	176		
Masse moléculaire théorique	19594,2		
Point isoélectrique théorique	5,72		
Composition en acides aminés	Ala (A)	10	5,7%
	Arg (R)	7	4,0%
	Asn (N)	6	3,4%
	Asp (D)	12	6,8%
	Cys (C)	0	0,0%
	Gln (Q)	9	5,1%
	Glu (E)	10	5,7%
	Gly (G)	13	7,4%
	His (H)	2	1,1%
	Ile (I)	14	8,0%
	Leu (L)	11	6,2%
	Lys (K)	13	7,4%
	Met (M)	1	0,6%
	Phe (F)	8	4,5%
	Pro (P)	12	6,8%
	Ser (S)	9	5,1%
	Thr (T)	14	8,0%
	Trp (W)	1	0,6%
	Tyr (Y)	8	4,5%
	Val (V)	16	9,1%
Courbe de titration théorique (charge théorique de la protéine en fonction du pH)			

Figure IV-3 : Analyse de la structure primaire de la protéine CG18594 avec l'outil Protparam (<http://us.expasy.org/tools/protparam.html>).

1.2. Alignements de séquence

1.2.1. Alignement de la protéine CG18594 avec les autres PEBP de drosophile

La séquence de la protéine CG18594 a été alignée avec celle des 6 autres PEBP de drosophile (Figure IV-4).

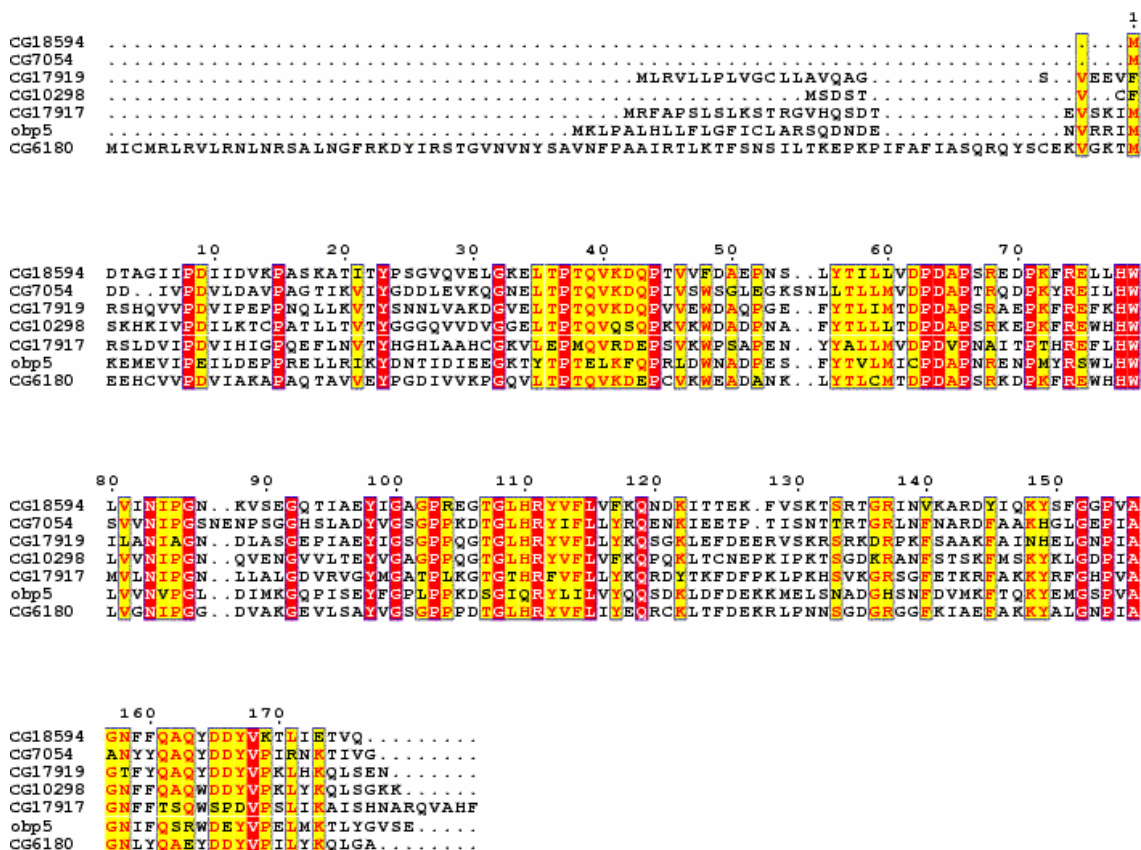


Figure IV-4 : Alignement de la protéine CG18594 avec les 6 autres PEBP de drosophiles. L'alignement a été réalisé avec le logiciel CLUSTALW, (Eddy, 1995). Les cadres rouges indiquent les résidus identiques, les résidus non conservés, mais similaires, sont indiqués en jaune. (Figure réalisée avec le logiciel Espript).

À partir de cet alignement nous pouvons calculer le pourcentage d'identité entre la protéine CG18594 et les six autres PEBP de drosophile. Les valeurs obtenues sont indiquées dans le Tableau IV-1:

Protéine	% d'identité avec la protéine CG18594
CG6180	48
CG17919	48
CG10298	55
CG7054	48
CG17917	40
OBP5	42

Tableau IV-1 : Pourcentage d'identité de la protéine CG18594 avec les autres PEBP de drosophile. L'alignement a été réalisé avec le logiciel CLUSTALW.

La protéine CG18594 montre donc un pourcentage d'identité globalement proche avec les autres PEBP de drosophile (de 40 % d'identité avec la protéine CG17917 à 55 % d'identité avec la protéine CG10298). L'observation des alignements de séquences montre clairement que certaines zones sont conservées pour l'ensemble des PEBP de drosophile alors que d'autres zones sont beaucoup plus variables. Ainsi, les extrémités N-terminal ou C-terminal sont moins bien alignées. La protéine CG6180 est très différente par sa longue extension N-terminale. Les PEBP de drosophile, bien que présentant un pourcentage d'identité important, possèdent donc des différences de séquence significatives.

1.2.2. Alignement de CG18594 avec les PEBP des autres espèces

La séquence de la protéine CG18594 a été alignée avec celle des autres PEBP étudiées sur le plan structural (Figure IV-5).

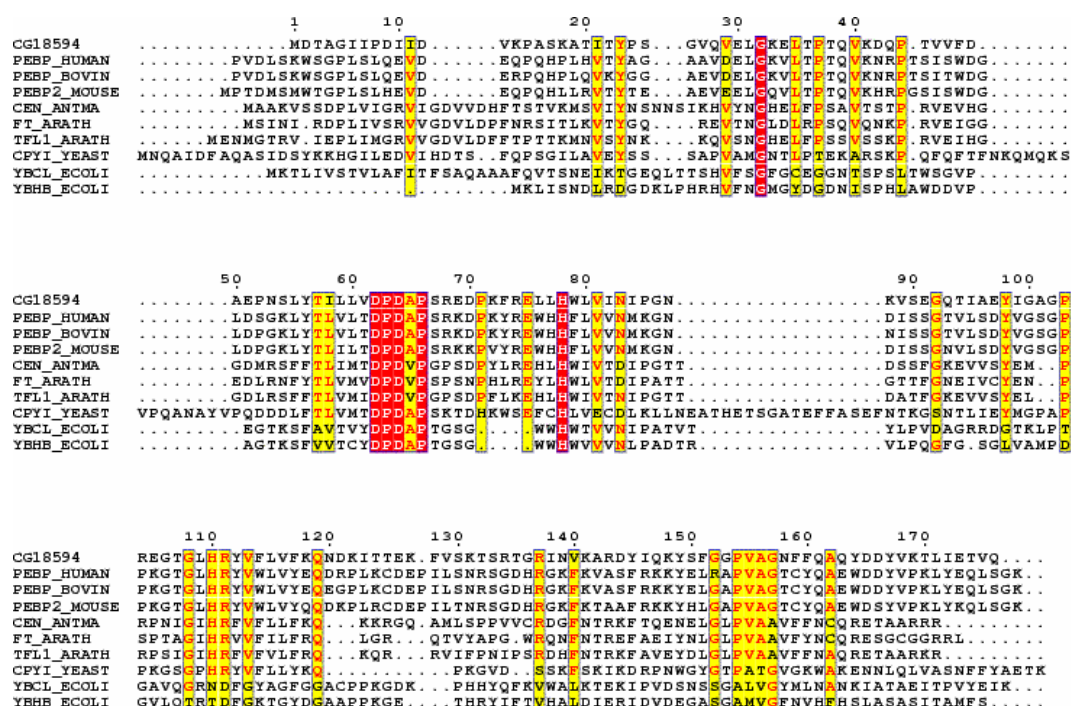


Figure IV-5 : Alignement de la protéine CG18594 avec différentes PEBP dont la structure 3D est connue. L'alignement a été réalisé avec le logiciel CLUSTALW, (Eddy, 1995). Les cadres rouges indiquent les résidus identiques, les résidus non conservés, mais similaires, sont indiqués en jaune. (Figure réalisée avec le logiciel Esprit).

Cet alignement montre bien que les zones conservées entre différentes espèces sont moins nombreuses qu'entre les 7 PEBP de drosophile.

Le tableau ci-dessous indique les pourcentages d'identité observée entre la protéine CG18594 et ces différentes PEBP (Tableau IV-2) :

	% d'identité avec la protéine CG18594
PEBP-1 humaine	42
PEBP-1 bovine	42
PEBP-2 souris	39
Antirrhinum Cen	28
Arabidopsis FT1	29
Arabidopsis TFL1	28
Saccharomyces TFS1	23
E. coli YBCL	11
E. coli YBHB	13

Tableau IV-2 : Pourcentages d'identité de la protéine CG18594 avec les PEBP dont la structure 3D est connue. L'alignement a été réalisé avec le logiciel CLUSTALW.

A partir de ces alignements, nous pouvons conclure que sur la base de sa séquence, la protéine CG18594 appartient bien à la famille des PEBP. En effet, elle possède la séquence consensus caractéristique des PEBP mais en plus on observe des conservations de résidus significatives avec les autres membres de la famille. Les PEBP les plus proches sont les PEBP humaines et bovines (42 % d'identité). L'alignement avec les protéines de plantes, levure (TFS1) ou de bactéries (YBHB et YBCL) montre un pourcentage d'identité beaucoup plus faible.

Mon travail sur la protéine CG18594 a d'abord consisté à mettre au point les techniques d'expression et de purification permettant de produire la protéine, marquée ou non, dans le but d'une étude structurale. J'ai ensuite réalisé une caractérisation physico-chimique de cette protéine afin de déterminer quelles étaient les meilleures conditions pour l'étude structurale par RMN. Les principaux résultats sont présentés dans l'article « *Cloning, high yield over-expression, purification, and characterization of CG18594, a new PEBP/RKIP family member from Drosophila melanogaster* », publié dans le journal « *Protein expression and purification* ».

2. « Cloning, high yield over-expression, purification, and characterization of CG18594, a new PEBP/RKIP family member from *Drosophila melanogaster* »

ARTICLE IN PRESS



Available online at www.sciencedirect.com

SCIENCE @ DIRECT®

Protein Expression and Purification xxx (2006) xxx–xxx

Protein
Expression
& Purification

www.elsevier.com/locate/yprep

Cloning, high yield over-expression, purification, and characterization of CG18594, a new PEBP/RKIP family member from *Drosophila melanogaster*

Gilles Rautureau, Laurence Jouvensal, Martine Decoville, Daniel Locker, Françoise Vovelle, Françoise Schoentgen *

Centre de Biophysique Moléculaire, UPR 4301 CNRS affiliated to Orléans University and to INSERM, Rue Charles Sadron, 45071 Orléans cedex 2, France

Received 24 November 2005, and in revised form 18 January 2006

Abstract

The phosphatidylethanolamine-binding protein (PEBP) family is widely distributed in various species, from bacteria to mammals. These proteins seem to modulate important cell mechanisms: they control heterotrimeric G-proteins, inhibit the MAP-kinase and NFκB signaling pathways, and also serine proteases (thrombin, neuropsin, and chymotrypsin). In order to establish structure–function relationships for this family of proteins, our study focuses on PEBPs expressed within a single organism: *Drosophila melanogaster*, which constitutes a model system that lends itself well to establishing links between genes' expression and the corresponding proteins' functions, and to studying physiological mechanisms such as development. Here, we describe an optimized protocol for high level over-expression and high yield/high purity production of CG18594, one of *Drosophila* six putative PEBPs, for biophysical studies. The yield of the purified ¹⁵N labeled protein is estimated to be 60 mg/L of M9 minimal medium. Analysis of the secondary structure using circular dichroism indicates that the protein comprises mainly β-sheets at pH 7. The good dispersion of the crosspeaks on the ¹H–¹⁵N HSQC spectrum provides evidence of a proper folding of the purified protein, though its time evolution suggests a tendency to denature. Taken together, these data are consistent with the assumption that the CG18594 protein belongs to the PEBP family.

© 2006 Elsevier Inc. All rights reserved.

Keywords: RKIP; Phosphatidylethanolamine-binding protein; *Drosophila*; CG18594; CG7054; CG10298; CG17917; CG17919; CG6180; Purification; Isotope labeling; NMR

The phosphatidylethanolamine-binding protein (PEBP)¹ family is a highly conserved group of proteins found in a variety of tissues from a wide range of organisms, from bacteria to mammals and plants. PEBPs are proteins of about 20 kDa that were originally purified as cytosolic proteins from bovine brain [1]. Binding studies have shown that PEBPs have an affinity for anionic ligands, such as phosphatidylethanolamine, opioids, nucleotides such as GTP or

GDP, and small GTP-binding proteins [2]. A number of biological functions have been proposed for PEBPs from various organisms, including inhibition of kinases of the MAP kinase [3] and NFκB signaling pathways [4], modulation of heterotrimeric G proteins [5], and also inhibition of serine proteases (thrombin, neuropsin, chymotrypsin) [6]. In humans, PEBP has been renamed Raf-1 kinase inhibitor protein (RKIP) [3]. Human PEBP/RKIP is implicated in cancer, and acts as a suppressor of metastases [7]. During cancer treatment, it sensitizes the cells to chemotherapy and immunotherapy [8]. Human PEBP/RKIP is also implicated in other pathologies: Alzheimer's disease [9], infertility [10,11], diabetes [12], etc. However, the molecular mechanisms by which PEBPs act remain obscure.

* Corresponding author. Fax: +33 (0) 238 63 15 17.

E-mail address: schoent@cnrs-orleans.fr (F. Schoentgen).

¹ Abbreviations used: PEBP, phosphatidylethanolamine-binding protein; RKIP, Raf-1 kinase inhibitor protein; ESTs, expressed sequence tags; LPS, lipopolysaccharides; SDS, sodium dodecyl sulfate; PCR, polymerase-chain reaction.

Crystal structures have been determined for a few members of the PEBP family (human (PDB code: 1BD9) [13], bovine (PDB code: 1B7A) [14], murine (PDB code: 1KN3) [15], yeast (PDB code: 1WPX) [16], *Antirrhinum* (PDB code: 1QOU) [17], *Arabidopsis* (PDB code: 1WKP) [18], and *Escherichia coli* (PDB code: 1FJJ) [19]), all of which show extensive fold conservation. The structure consists of a large central β -sheet flanked by a smaller one on one side, and by an α -helix on the other side. Sequence alignments show a conserved central region, D-P-D-x-P-x(11)-H-x(28,30)-H-R, which forms a consensus signature for the PEBP family [14]. This region forms part of a ligand-binding site, which can accommodate various anionic groups. The N- and C-terminal regions of the protein are the least conserved, and may be involved in interactions with different protein partners. In humans and rats, the N-terminal residues 2–12 form the natural cleavage peptide HCNP involved in neuronal development [20,21]. The C-terminal region is deleted in plant and bacterial PEBP homologues while it seems to help control accessibility toward the active site.

To establish structure–function relationships for this family of proteins, our study focuses on PEBPs expressed within a single organism: *Drosophila melanogaster*, which is one of the most intensively studied organisms in biology. Owing to its easy-to-manipulate genetic system, relatively low cost, and biological complexity comparable to that of a mammal, it is used as a model system for the investigation of many developmental and cellular processes common to higher eukaryotes, including humans. Indeed, many organ systems in mammals have well-conserved homologs in *Drosophila*, and research on *Drosophila* has already provided new insights into cancer, neurodegenerative diseases, behavior, immunity, aging, multigenic inheritance, and development. The study of the fruit fly genome (<http://www.fruitfly.org/>) has led to the identification of six PEBP paralogs. These six putative genes are transcribed since: full length cDNAs are commercially available for five of them (<http://dgrc.cgb.indiana.edu/>) (CG6180, CG17919, CG10298, CG7054 from adult head mRNAs, and CG18594 from larvae/pupae mRNA), and the

last one, CG17917, has been detected as expressed sequence tags (ESTs) (<http://flybase.bio.indiana.edu/>). Moreover, CG18594 has been recently identified in the hemolymph of *D. melanogaster* [22], where it was proven to be over-expressed in larvae induced by *E. coli* lipopolysaccharides (LPS) [23] or by a suspension of Gram-positive bacteria or yeasts [24].

A sequence alignment of the six PEBPs from *D. melanogaster* is shown in Fig. 1. The six proteins show sequence differences (12% of identity for the six proteins) and their pI range from 5.0 to 9.7 for CG7054 and CG17917, respectively, indicating major differences in their physicochemical properties and probably in their modes of action. Understanding the reason why six different PEBPs have been maintained through evolution in *D. melanogaster* should be critical in deciphering the role of the PEBP family.

We have initiated a biophysical study on one of *Drosophila* putative PEBPs. In this paper, we present an improved method for high yield/high purity purification of CG18594, as well as an initial characterization of the purified protein.

Materials and methods

Cloning of *cg18594* ORF into *pET31b* vector to produce untagged CG18594 and N- or C-terminally 6-His-tagged CG18594

Since the *cg18594* gene is intron free, purified genomic DNA from *D. melanogaster* Oregon R was used as a direct template for PCR instead of mRNA. Oligonucleotide primers corresponding to the *cg18594* ORF were designed and synthesized by MWG Biotech AG, based on the published genome sequence of *D. melanogaster* (<http://flybase.bio.indiana.edu/>). For the untagged protein, the forward primer (5'-CATATGGACACCGCCGGCATTAT-3') included a *NdeI* restriction site at the 5' end of the ORF (underlined) while the reverse primer (5'-CTCGAGTTACTGGACCGTCTCG-3') added a *XhoI* restriction site at the 3' end of the ORF (underlined). The N-terminally 6-His-tagged protein was obtained

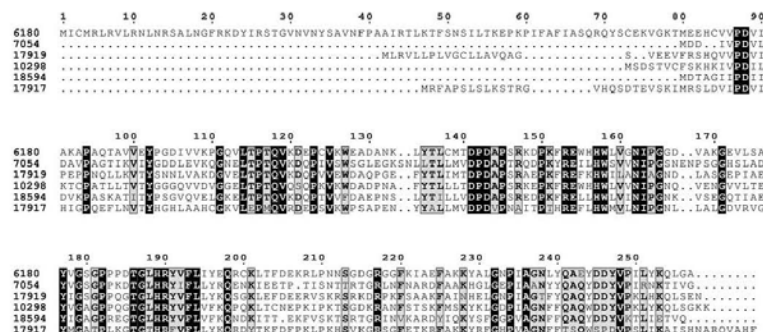


Fig. 1. Sequence alignment of the six PEBPs from *D. melanogaster*. Multiple sequence alignments were carried out using the sequence analysis program CLUSTALW. Black boxes indicate identities and clear boxes homologies.

by using the forward primer 5'-CAGCATAT GACCATC ATCATCATCATGACACCGCCGGCATTAT-3', which included a *NdeI* restriction site at the 5' end of the ORF (underlined) and a 6-His-encoding sequence (italics), and the reverse primer 5'-CTCGAGTTACTGGACCGTCTCG-3', which added a *XhoI* restriction site at the 3' end of the ORF (underlined). The C-terminally 6-His-tagged protein was obtained by using the forward primer 5'-CATATGGACACCGCCGGCATTAT-3', which included a *NdeI* restriction site at the 5' end of the ORF (underlined), and the reverse primer 5'-CGCTCGAGCTGGACCGTCTCGATGAGG-3', which added a *XhoI* restriction site at the 3' end of the ORF (underlined) with no stop codon, so that the resulting recombinant protein has a C-terminal 6-His-tag encoded by the pET31b vector.

The PCR protocol for 30 cycles was as follows: 94°C for 1 min, 52°C for 1 min, 70°C for 1 min. *Taq* DNA polymerase *Dynazyme* (Ozyme) was used. The PCR product was cloned into plasmid *pCRII-TOPO* (Invitrogen kit), and transformed into *Top10* cells using the commercial standard protocol. The *pCRII-TOPO-18594* plasmid was purified using a standard protocol [25], digested with *NdeI* and *XhoI* restriction endonucleases (New England Biolabs), and ligated with the *NdeI*-*XhoI* digested expression vector (pET-31b(+)) (Novagen) using T4 DNA ligase (Invitrogen). The product was transformed into *E. coli* C600 competent cells, which were then plated on LB plates supplemented with ampicillin 100 µg mL⁻¹. Plasmid DNA from a few transformants was extracted using the Qiaprep Spin Miniprep kit (Qiagen). The sequence of the cloned ORF (in the pET-31b expression vector) was confirmed by DNA sequencing (MWG Biotech). One clone was selected and transformed into *E. coli* strain BL21(DE3) (Novagen) for protein expression.

Preliminary tests of non-tagged and N- or C-terminally 6-His-tagged CG18594 expression

A few transformants of *Escherichia coli* BL21(DE3) transformed by plasmid pET31b-18594 were picked out from an LB agar plate containing ampicillin 100 µg mL⁻¹, and inoculated in 2 mL of LB medium supplemented with the same antibiotic for cultivation at 30°C overnight. The culture was then diluted 100-fold with LB medium, and cultured at 37°C to an OD₆₀₀ of ~0.6 (optical cell density at 600 nm). Expression was then induced with IPTG (isopropyl-β-D-thiogalactoside; 1 mM final concentration). Samples were collected at 30 min, 1, 2, 3 and 4 h, mixed with 2× SDS–PAGE loading buffer, boiled, and sonicated (5 min each), and finally harvested (10 min, 15,000g) before being loaded onto a 15% SDS–PAGE gel to screen for the best transformants showing high level expression of tagged or untagged CG18594.

Preparative over-expression of unlabeled non-tagged and N- or C-terminally 6-His-tagged CG18594

For over-expression of unlabeled CG18594, 1.5 L of LB medium with ampicillin (100 µg mL⁻¹) were inoculated with

a 15 mL overnight culture of *E. coli* BL21 (DE3, pET31b-18594), and incubated at 30°C with shaking at 250 rpm. Incubation continued until OD₆₀₀ reached 0.7; then IPTG was added to a final concentration of 1 mM to induce expression of CG18594. After further incubation at 30°C and 250 rpm for 2 h 30, the cells were pelleted by centrifugation at 4000g for 5 min at 4°C, and then frozen at –20°C.

Preparative over-expression of ¹⁵N-labeled non-tagged CG18594

The general protocol for isotope labeling is as follows (adapted from the published method [26]): 15 mL of an *E. coli* BL21 (DE3, pET31b-18594) overnight preculture were inoculated in 2 L of LB shaken at 250 rpm at 30°C. Upon reaching OD₆₀₀ ~0.7, the cells were pelleted by a 4 min centrifugation at 5000g. They were then washed and pelleted again using 500 mL of M9 salt solution, excluding all nitrogen and carbon sources. The cell pellet was resuspended in an isotopically labeled M9 medium with ¹⁵NH₄Cl (99%) (Cambridge Isotope Laboratories), ampicillin 50 µg mL⁻¹, and then incubated to allow the recovery of growth and the clearance of unlabeled metabolites. After 1 h, protein expression was induced by the addition of IPTG to a final concentration of 1 mM. After a 3-h incubation period, the cells were harvested at 4000g for 5 min and frozen at –20°C.

Purification of non-tagged and N- or C-terminally 6-His-tagged CG18594

The frozen cell pellet (4 g) was thawed in 60 mL of lysis buffer (Tris–HCl 50 mM, pH 8.5, EDTA 1 mM, Na₂S₂O₃ 0.05%, DTT 0.5 mM, PMSF 0.5 mM, lysozyme 1 mg mL⁻¹), and incubated 30 min at 37°C before 5 cycles of freeze/thaw steps. Then, 50 mM of MgCl₂ and 0.1 mg mL⁻¹ DNase I were added before a 30-min incubation at room temperature. The disrupted cell lysate was centrifuged at 15,000g for 30 min at 4°C.

For untagged CG18594, the clear supernatant was collected and precipitated with 32.6 g/100 mL of ammonium sulfate during 30 min at 4°C. After centrifugation at 15,000g, 4°C for 30 min, the supernatant was dialysed overnight against QAE buffer (Tris–HCl 50 mM, pH 8, NaCl 50 mM, EDTA 1 mM, Na₂S₂O₃ 0.05%), and concentrated to 15 mL on an Amicon stirred cell (Millipore) with a YM10 10 kDa membrane. The sample was loaded on a 30 cm/2.4 cm QAE-Sephadex A-50 column. The flow was 35 mL h⁻¹ and 3 mL fractions were collected. CG18594 was eluted and detected by 15% SDS–PAGE. The peak fractions were concentrated to 10 mL (Amicon cell). Five milliliters of samples were submitted to preparative isoelectric focusing (IEF) in sucrose gradient (LKB column 8101) with Ampholines™ pH 4–6 and 6–8 (Amersham Pharmacia Biotech). The IEF was performed at 1600 V during 18 h. The 1.5 mL fractions were tested for pH and optical density at 280 nm. The CG18594 peak was pooled, dialysed against

20 mM Na_2HPO_4 , pH 7, NaCl 0.2 M, and concentrated to 10 mL. Two milliliters of samples were then loaded onto a gel filtration Superdex 75 16/60 column (Pharmacia). Fractions of 1 mL were collected and checked by 15% SDS-PAGE. Pooled peak fractions were finally dialysed overnight against 20 mM Na_2HPO_4 , pH 7, concentrated to 1 mM (Amicon cell) and frozen at -20°C . The quantity of CG18594 was evaluated using an extinction coefficient of $15,930\text{ M}^{-1}\text{ cm}^{-1}$ at 280 nm (theoretical estimation from <http://us.expasy.org/>). The final protein purity was assessed according to 15% SDS-PAGE gel and mass spectrometry. The purified protein was submitted to automated Edman degradation performed in a Procise 492 protein sequencer (Applied Biosystems). The phenylthiohydantoines of amino acids were identified by online HPLC.

Purification of 6 His-tagged CG18594 was performed using TalonTM resin (Co^{2+} , Clontech) following the standard commercial protocol. The final protein was dialysed against 20 mM Na_2HPO_4 , pH 7, NaCl 0.2 M, and concentrated (Amicon cell).

Characterization of CG18594 by DLS and CD

Dynamic light scattering (DLS) analyses were carried out on a temperature controlled Dynapro-MSTC DLS (Protein Solutions). The effects of protein concentration (0.09, 0.2, 0.25, 0.67, 1, and 1.5 mM), of salts (NaCl (0, 0.1, 0.2, 0.3, 0.4, and 0.5 M), LiCl (0, 50, 100, 150, and 200 mM), KSCN (0, 2, 5, and 15 mM), and arginine (0, 200, and 400 mM)), of various buffer conditions (citrate phosphate from pH 3.5 to 7, potassium phosphate from pH 2 to 9, sodium phosphate from pH 2 to 9, and potassium acetate from pH 4.5 to 6), or lyophilization, were studied on aggregation of CG18594. All measurements were performed at 5°C .

CD measurements were recorded on a Jasco J-810 CD spectropolarimeter. Spectra were measured in the far-UV range (190–260 nm) in 1-mm pathlength quartz cells at a protein concentration of 0.2 mg mL^{-1} . The buffer used was 10 mM KH_2PO_4 adjusted from pH 2 to 9 with 0.1 M KOH. All measurements were performed at 5°C . Each spectrum was the average of five consecutive scans, and all spectra were corrected by subtracting the buffer baseline. The results were expressed in terms of mean residue ellipticity $[\theta]$ in units of $\text{deg cm}^2\text{ dmol}^{-1}$.

NMR spectroscopy

The NMR sample was prepared by addition of 50 μL D_2O , 5 μL 1% NaN_3 , and 5 μL of 100 mM 2,2-dimethyl-2-silapentane-5-sulfonate sodium salt (DSS) to 580 μL of protein solution (final concentrations: 1.0 mM protein in 18 mM Na_2HPO_4 , pH 7). NMR spectra were acquired on a 600 MHz Varian INOVA spectrometer. The sample temperature was maintained at 5 or 20°C . ^1H – ^{15}N HSQC spectra were recorded with 1472 points in t_2 and 128 complex data points in t_1 . ^1H chemical shifts were referenced relative to internal DSS (0.00 ppm), and ^{15}N was set indirectly using

the ^1H – ^{15}N frequency ratio [27]. The NMR spectra were processed with NMRPipe [28] and analyzed using NMRView [29] software.

Results and discussion

Cloning of *cg18594* ORF into pET31b vector to produce untagged CG18594, and N- or C-terminally 6-His-tagged CG18594

The ORF of *cg18594* (FlyBase ID: FBgn0038973) is located on chromosome-arm 3R, and has a length of 598 bp on the genomic sequence. It does not show any intron. Its cytologic location is 94A16.

To over-express CG18594, we cloned the *cg18594* ORF into the *NdeI* and *XhoI* restriction sites of pET-31b. This expression vector uses the T7 promoter, potentially allowing for high level expression. A stop codon was incorporated in the reverse primer so that no histidine tag would be expressed at the C-terminus of CG18594.

For 6-His-tag strategies, either we incorporated 6-His at the N-terminus of CG18594, or we used the C-terminal 6-His-tag encoded by vector pET31b, so that the recombinant protein displayed a 6-His-tag at the C-terminus. DNA sequences were verified in expression vectors before protein expression.

Protein expression

Escherichia coli BL21 (DE3, pET31b-18594) were grown either in LB medium, for unlabeled protein, or first in 2 L of LB medium and then in 500 mL of M9 minimal medium, for labeled protein. This latter method permits efficient isotopic labeling of recombinant proteins (more than 95% ^{15}N incorporation) and a considerable improvement in production compared to standard protocols in M9 medium [26]. The level of over-expression induced by IPTG was very high for both methods, labeled or unlabeled. Measured by densitometry scanning, the amount of synthesized CG18594 was up to 20% of the total cellular proteins for the untagged protein, and up to 10% for the tagged protein. The apparent molecular weight was about 22 kDa, a little larger than the value expected from the sequence (19.6 kDa). This phenomenon is commonly observed among members of the PEBP family [30].

Protein purification

Recombinant untagged CG18594 was purified from the supernatant of the bacterial lysate using sulfate precipitation, QAE-Sephadex ion chromatography, preparative isoelectric focusing in sucrose gradient, and finally gel filtration on Superdex 75 column. All steps were performed at 4°C , and the entire purification procedure was accomplished in about 10 days. Figs. 2A–D show the elution profiles obtained at different steps of the purification procedure. SDS-PAGE analyses carried out at each purifi-

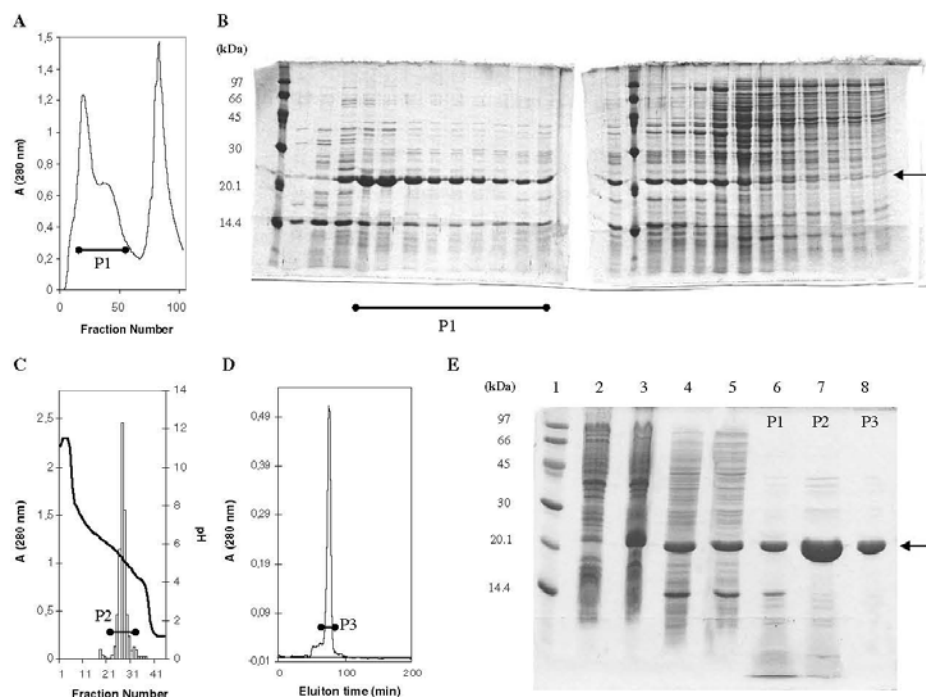


Fig. 2. Purification of untagged CG18594. (A) QAE Sephadex A-50 column elution profile at 280 nm. (B) Selected fractions were subjected to SDS-PAGE. The order of elution is conserved. CG18594 appeared in the first fractions (arrow). The horizontal bar indicates the fractions pooled for further purification (P1) (see (E), lane 6). (C) Electrofocusing profile of CG18594. The protein forms a well-resolved major peak around its pI (5.7), which was pooled (P2) (see (E), lane 7). (D) Gel filtration chromatography on Superdex 75 16/60 column. CG18594 was in the major peak, which was pooled (P3) (see (E), lane 8). (E) 15% SDS-PAGE analysis at different steps of CG18594 purification. Lane 1, molecular weight markers. Lane 2, *E. coli* (BL21, DE3) culture before induction. Lane 3, 3 h after IPTG induction. Lane 4, soluble fraction of the cell lysate after centrifugation. Lane 5, supernatant after ammonium sulfate precipitation. Lane 6, pool P1 after QAE sephadex ion exchange chromatography. Lane 7, pool P2 after preparative isoelectric focusing in sucrose gradient. Lane 8, pool P3 after gel filtration.

cation step during the preparation of non-isotopically labeled CG18594 are shown in Fig. 2E. The purity of the final CG18594 is higher than 95% as determined by SDS-PAGE. The final yield of purified recombinant CG18594 was about 15 mg/L of LB medium, in the case of the unlabeled protein, or 30 mg per 0.5 L of M9 medium (prepared from 2 L of LB), in the case of the labeled protein.

Alternatively, we tried to use the 6-His-tag strategy at the N- or C-terminus. With the 6-His-tag, the over-expression rate was half that of the untagged protein, and affinity purification resulted in high purity but very low yields (~1 mg/L of LB medium). Furthermore, CD and NMR spectra showed that both N- or C-terminally tagged proteins were unfolded (data not shown). Consequently, since purification by classical methods could be achieved with high yields, we chose to over-express CG18594 without affinity tag.

The primary structures of the purified proteins were verified by mass spectrometry and N-terminal Edman sequencing, which confirmed the 10 first amino acids

(MDTAGIIPDI for the untagged protein and for the C-terminal 6-His-tagged protein, and MHHHHHHHTA for the N-terminal 6-His-tagged protein).

Dynamic light scattering

Preliminary NMR experiments suggested a tendency of the protein to aggregate. To investigate and optimize CG18594 monodispersion in solution, we performed a screen of various conditions by DLS experiments at 5°C. Whatever the conditions tested (buffer, salt; for details, see Materials and methods), the lyophilized protein always appeared as large heterogeneous aggregates, masking the possible presence of CG18594 monomers. For the non-lyophilized protein, monomers could be detected when no salt was added to the buffered protein solution, and, assuming the protein has a spherical shape, hydrodynamic radius values were derived, which lie between two and three nanometers depending on the solution composition. For subsequent experiments, we chose phosphate buffer pH 7, which

appeared to be the most favorable for CG18594 monodispersion in solution.

CD spectroscopy and secondary structure

CD spectroscopy was used to examine the conformational state of CG18594 at 5 °C and various pHs. The spectra displayed in Fig. 3 indicate that the secondary structure of CG18594 is highly pH-dependent. At pH 7, the CD spectrum shows a minimum at 221 nm. Since no negative band is observed at 209 nm, the structure appears to be largely β -sheet in nature, though the presence of small helices might account for the shift in minimum from 218 (purely β -sheet) to 221 nm. These data were used to evaluate the percentage of secondary structure elements in CG18594 using Dichroprot software and the “self consistent method 3” algorithm [31]. The CD spectrum at pH 7 corresponds to the higher content of secondary structure elements, with α -helix and β -sheet contents of 10.5 and 27.8%, respectively, in agreement with the structural features known for human, bovine, plant and bacterial PEBPs [32]. Secondary structure prediction from the amino acid sequence was performed with Psipred [33]. The locations of the predicted α -helix and β -sheet regions are shown Fig. 4. In agreement with the analysis of the CD spectrum at pH 7, CG18594 was predicted to be formed of approximately 12% of α -helices and 28% of β -sheets. When the pH is lowered below 7, the secondary structure content decreases. Interestingly, at pH 2, deconvolution of the CD spectrum indicates the persistence of approximately 9% of α -helices and 19% of β -sheets, whereas the turn content rises from 4% at pH 7 to 13%. However, the 1D ^1H NMR spectrum at pH 2.3 unambiguously shows the absence of tertiary structure in the protein (Fig. 3). Taken together, the persistence of secondary structures and the absence of tertiary structure demonstrate that CG18594 forms a molten globule at low pH.

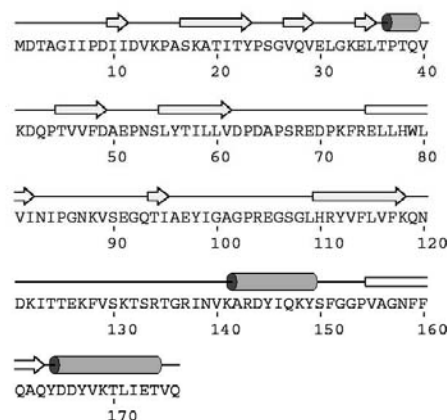


Fig. 4. Amino acid sequence and secondary structure prediction for CG18594 using Psipred (<http://bioinf.cs.ucl.ac.uk/psipred/>). Ribbons indicate α -helices and arrows β -strands.

NMR spectroscopy and tertiary structure

To examine whether the protein was properly folded at pH 7, a ^1H - ^{15}N HSQC spectrum was recorded. According to the primary structure, 164 amide proton–nitrogen correlations are expected on the HSQC spectrum. At 20 °C, about 167 crosspeaks could be found on the spectrum, but the peaks progressively shifted and the chemical shift dispersion was considerably reduced. After 4 days, the protein was completely unfolded. At 5 °C (Fig. 5) the linewidths were consistent with a 20 kDa protein, and the spectra remained unchanged for at least one week. The large chemical shift dispersion is consistent with the presence of a tertiary structure and is characteristic of a folded globular protein.

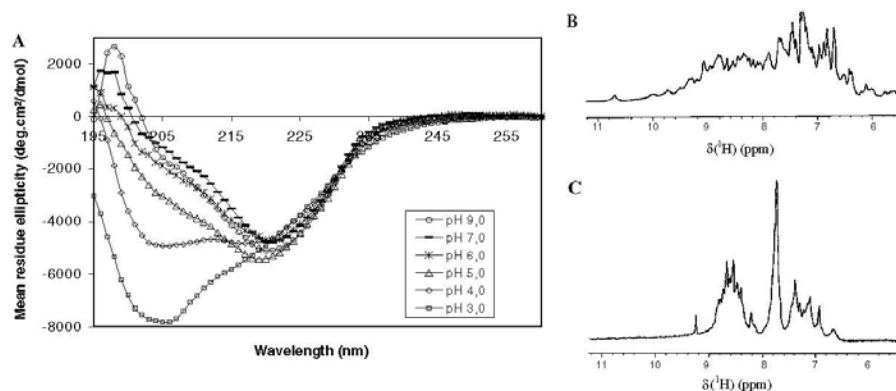


Fig. 3. (A) Circular dichroism spectra of CG18594 at 5 °C and different pH values. (B) Amide region of CG18594 ^1H NMR spectrum at pH 7, 5 °C: the large chemical-shift dispersion is indicative of a well-ordered structure. (C) Amide region of CG18594 ^1H NMR spectrum at pH 2.3, 5 °C: the poor chemical-shift dispersion indicates that the protein is unfolded.

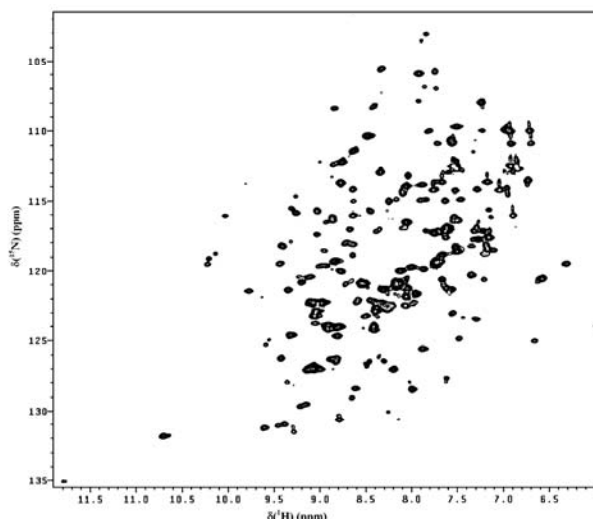


Fig. 5. ^1H – ^{15}N HSQC spectrum at 5 °C of 1 mM ^{15}N uniformly labeled CG18594 in 20 mM Na_2HPO_4 , pH 7, D_2O 10%, azide 0.1%.

Conclusion

We have described an optimized protocol for the over-expression and high yield/high purity production of CG18594 with a view to biophysical and functional studies. Interestingly, the stability and solubility of non-tagged CG18594 were better than those of 6-His-tagged CG18594. Analysis of the secondary structure using circular dichroism indicates that β -sheets are predominant at 5 °C and pH 7. Furthermore, the dispersion seen in the crosspeaks of the ^1H – ^{15}N HSQC spectrum indicates the presence of a well-ordered structure which is stable at low temperature during several days. Taken together, these data support the fact that CG18594 putative protein belongs to the PEPB family.

Acknowledgments

The authors are grateful to Dr. Jean Claude Maurizot for providing access to the circular dichroism facility and for helpful discussions, as well as to Estelle Rapine and Fabienne Godin for their excellent technical assistance in molecular biology experiments.

References

- [1] F. Schoentgen, N. Seddiqi, S. Bucquoy, P. Jolles, L. Lemesle-Varloot, K. Provost, J.P. Mornon, Main structural and functional features of the basic cytosolic bovine 21 kDa protein delineated through hydrophobic cluster analysis and molecular modelling, *Protein Eng.* 5 (4) (1992) 295–303.
- [2] S. Bucquoy, P. Jolles, F. Schoentgen, Relationships between molecular interactions (nucleotides, lipids and proteins) and structural features of the bovine brain 21-kDa protein, *Eur. J. Biochem.* 225 (3) (1994) 1203–1210.
- [3] K. Yeung, T. Seitz, S. Li, P. Janosch, B. McFerran, C. Kaiser, F. Fee, K.D. Katsanakis, D.W. Rose, H. Mischak, J.M. Sedivy, W. Kolch, Suppression of Raf-1 kinase activity and MAP kinase signalling by RKIP, *Nature* 401 (6749) (1999) 173–177.
- [4] K.C. Yeung, D.W. Rose, A.S. Dhillon, D. Yaros, M. Gustafsson, D. Chatterjee, B. McFerran, J. Wyche, W. Kolch, J.M. Sedivy, Raf kinase inhibitor protein interacts with NF- κ B-inducing kinase and TAK1 and inhibits NF- κ B activation, *Mol. Cell. Biol.* 21 (21) (2001) 7207–7217.
- [5] T. Krosiak, T. Koch, E. Kahl, V. Holtt, Human phosphatidylethanolamine-binding protein facilitates heterotrimeric G protein-dependent signaling, *J. Biol. Chem.* 276 (43) (2001) 39772–39778.
- [6] U. Hengst, H. Albrecht, D. Hess, D. Monard, The phosphatidylethanolamine-binding protein is the prototype of a novel family of serine protease inhibitors, *J. Biol. Chem.* 276 (1) (2001) 535–540.
- [7] Z. Fu, P.C. Smith, L. Zhang, M.A. Rubin, R.L. Dunn, Z. Yao, E.T. Keller, Effects of raf kinase inhibitor protein expression on suppression of prostate cancer metastasis, *J. Natl. Cancer Inst.* 95 (12) (2003) 878–889.
- [8] E.T. Keller, Metastasis suppressor genes: a role for raf kinase inhibitor protein (RKIP), *Anticancer Drugs* 15 (7) (2004) 663–669.
- [9] A.J. George, R.M. Holsinger, C.A. McLean, S.S. Tan, H.S. Scott, T. Cardamone, R. Cappai, C.L. Masters, Q.X. Li, Decreased phosphatidylethanolamine binding protein expression correlates with Abeta accumulation in the Tg2576 mouse model of Alzheimer's disease, *Neurobiol. Aging*, 2005, in press.
- [10] T.G. Cooper, A. Wagenfeld, G.A. Cornwall, N. Hsia, S.T. Chu, M.C. Orgebin-Crist, J. Drevet, P. Vernet, C. Avram, E. Nieschlag, C.H. Yeung, Gene and protein expression in the epididymis of infertile cros receptor tyrosine kinase-deficient mice, *Biol. Reprod.* 69 (5) (2003) 1750–1762.
- [11] D.M. Hickox, G. Gibbs, J.R. Morrison, K. Sebire, K. Edgar, H.H. Keah, K. Alter, K.L. Loveland, M.T. Hearn, D.M. de Kretser, M.K. O'Bryan, Identification of a novel testis-specific member of the phosphatidylethanolamine binding protein family, *pebp-2*, *Biol. Reprod.* 67 (3) (2002) 917–927.

- [12] E.T. Keller, Z. Fu, M. Brennan, The role of Raf kinase inhibitor protein (RKIP) in health and disease, *Biochem. Pharmacol.* 68 (6) (2004) 1049–1053.
- [13] M.J. Banfield, J.J. Barker, A.C. Perry, R.L. Brady, Function from structure? The crystal structure of human phosphatidylethanolamine-binding protein suggests a role in membrane signal transduction, *Structure* 6 (10) (1998) 1245–1254.
- [14] L. Serre, B. Vallee, N. Bureaud, F. Schoentgen, C. Zelwer, Crystal structure of the phosphatidylethanolamine-binding protein from bovine brain: a novel structural class of phospholipid-binding proteins, *Structure* 6 (10) (1998) 1255–1265.
- [15] P.C. Simister, M.J. Banfield, R.L. Brady, The crystal structure of PEBP-2, a homologue of the PEBP/RKIP family, *Acta Crystallogr. D. Biol. Crystallogr.* 58 (Pt 6 Pt 2) (2002) 1077–1080.
- [16] J. Mima, M. Hayashida, T. Fujii, Y. Narita, R. Hayashi, M. Ueda, Y. Hata, Structure of the carboxypeptidase Y inhibitor IC in complex with the cognate proteinase reveals a novel mode of the proteinase-protein inhibitor interaction, *J. Mol. Biol.* 346 (5) (2005) 1323–1334.
- [17] M.J. Banfield, R.L. Brady, The structure of Antirrhinum centroradialis protein (CEN) suggests a role as a kinase regulator, *J. Mol. Biol.* 297 (5) (2000) 1159–1170.
- [18] J.H. Ahn, D. Miller, V.J. Winter, M.J. Banfield, J.H. Lee, S.Y. Yoo, S.R. Henz, R.L. Brady, D. Weigel, A divergent external loop confers antagonistic activity on floral regulators FT and TFL1, *EMBO J.* 25 (3) (2006) 605–614.
- [19] L. Serre, K. Pereira de Jesus, C. Zelwer, N. Bureaud, F. Schoentgen, H. Benedetti, Crystal structures of YBHB and YBCL from *Escherichia coli*, two bacterial homologues to a Raf kinase inhibitor protein, *J. Mol. Biol.* 310 (3) (2001) 617–634.
- [20] S. Mitake, K. Ojika, E. Katada, Y. Otsuka, N. Matsukawa, O. Fujimori, Distribution of hippocampal cholinergic neurostimulating peptide (HCNP) immunoreactivity in the central nervous system of the rat, *Brain Res.* 706 (1) (1996) 57–70.
- [21] B.S. Vallee, G. Coadou, H. Labbe, D. Sy, F. Vovelle, F. Schoentgen, Peptides corresponding to the N- and C-terminal parts of PEBP are well-structured in solution: new insights into their possible interaction with partners in vivo, *J. Pept. Res.* 61 (2) (2003) 47–57.
- [22] M. Guedes Sde, R. Vitorino, K. Tomer, M.R. Domingues, A.J. Correia, F. Amado, P. Domingues, *Drosophila melanogaster* larval hemolymph protein mapping, *Biochem. Biophys. Res. Commun.* 312 (3) (2003) 545–554.
- [23] E. Vierstraete, P. Verleyen, G. Baggerman, W. D'Hertog, G. Van den Bergh, L. Arckens, A. De Loof, L. Schoofs, A proteomic approach for the analysis of instantly released wound and immune proteins in *Drosophila melanogaster* hemolymph, *Proc. Natl. Acad. Sci. USA* 101 (2) (2004) 470–475.
- [24] E. Vierstraete, P. Verleyen, F. Sas, G. Van den Bergh, A. De Loof, L. Arckens, L. Schoofs, The instantly released *Drosophila* immune proteome is infection-specific, *Biochem. Biophys. Res. Commun.* 317 (4) (2004) 1052–1060.
- [25] T. Maniatis, E.F. Fritsch, J. Sambrook, *Molecular Cloning: A Laboratory Manual*, second ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989.
- [26] J. Marley, M. Lu, C. Bracken, A method for efficient isotopic labeling of recombinant proteins, *J. Biomol. NMR* 20 (1) (2001) 71–75.
- [27] D.S. Wishart, C.G. Bigam, A. Holm, R.S. Hodges, B.D. Sykes, I.H. 13C and 15N random coil NMR chemical shifts of the common amino acids. I. Investigations of nearest-neighbor effects, *J. Biomol. NMR* 5 (1) (1995) 67–81.
- [28] F. Delaglio, S. Grzesiek, G.W. Vuister, G. Zhu, J. Pfeifer, A. Bax, NMRPipe: a multidimensional spectral processing system based on UNIX pipes, *J. Biomol. NMR* 6 (3) (1995) 277–293.
- [29] B.A. Johnson, Using NMRView to visualize and analyze the NMR spectra of macromolecules, *Methods Mol. Biol.* 278 (2004) 313–352.
- [30] F. Schoentgen, F. Saccoccio, J. Jolles, I. Bernier, P. Jolles, Complete amino acid sequence of a basic 21-kDa protein from bovine brain cytosol, *Eur. J. Biochem.* 166 (2) (1987) 333–338.
- [31] N. Sreerama, S.Y. Venyaminov, R.W. Woody, Estimation of protein secondary structure from circular dichroism spectra: inclusion of denatured proteins with native proteins in the analysis, *Anal. Biochem.* 287 (2) (2000) 243–251.
- [32] B. Vallee, C. Teyssier, R. Maget-Dana, J. Ramstein, N. Bureaud, F. Schoentgen, Stability and physicochemical properties of the bovine brain phosphatidylethanolamine-binding protein, *Eur. J. Biochem.* 266 (1) (1999) 40–52.
- [33] L.J. McGuffin, K. Bryson, D.T. Jones, The PSIPRED protein structure prediction server, *Bioinformatics* 16 (4) (2000) 404–405.

3. Résultats et discussion

3.1. Production de la protéine CG18594

Dans une première partie de mon travail de thèse, j'ai mis au point une méthode permettant d'obtenir rapidement une quantité importante de protéine (de l'ordre de la dizaine de milligrammes), repliée et la plus pure possible. Cet objectif a été atteint : les protocoles décrits permettent maintenant d'obtenir en deux semaines environ 15 mg de protéine par litre de milieu de culture non marqué (LB) ou 60 mg par litre de milieu marqué (M9).

Les mises au point réalisées ont principalement concerné les étapes suivantes :

- Optimisation de la sur-expression : conditions d'induction, la quantité d'IPTG, durée de surexpression, température d'incubation.
- Détermination des seuils utilisés pour les précipitations au sulfate d'ammonium.
- Adaptation de la technique d'isoélectrofocalisation.
- Choix des types de colonnes, optimisation des conditions de chromatographie, de l'enchaînement des différentes chromatographies d'échange d'ions et de filtration sur gel.

Les bactéries *Escherichia coli* BL21DE3 contenant le plasmide pET31b pourvu de l'insert d'intérêt surexpriment la protéine CG18594 par ajout d'IPTG. La quantité d'IPTG nécessaire pour obtenir une bonne surexpression est supérieure dans le cas du milieu M9 car il contient du glucose, qui est un répresseur métabolique connu de l'opéron lactose dont nous utilisons le promoteur. Différents essais nous ont montré que la surexpression était optimale à 30°C pendant 2h30 dans le milieu classique LB ou 3h dans le milieu marqué. Les cellules sont ensuite lysées par plusieurs cycles de congélation / décongélation en présence de lysozyme. Nous avons montré que la sonication entraînait une dénaturation de la protéine CG18594 qui était alors non retenue en cellule Amicon sur membrane YM10.

Dans le but de réaliser une précipitation sélective des protéines du surnageant obtenu, nous avons testé la précipitation de la protéine CG18594 en fonction de la concentration

en sulfate d'ammonium, entre 20% et 90%. Nous observons bien l'absence de la protéine dans le surnageant à partir de 60% et sa présence dans le culot. Par prudence nous avons donc choisi de réaliser l'étape de précipitation à 55% de sulfate d'ammonium. Nous éliminons de cette manière une grande quantité de protéines, d'acides nucléiques et autres contaminants. Afin d'éliminer le sulfate d'ammonium présent, la solution est rapidement dialysée contre du tampon QAE.

La protéine est ensuite concentrée par ultrafiltration et déposée sur une résine de QAE-Sephadex sur laquelle elle n'est pas retenue. En effet, d'après la courbe de titration théorique de la protéine CG18594 (Figure IV-3, page 122), elle n'est que très faiblement chargée négativement à pH 8. La faible concentration en NaCl (50 mM) du tampon QAE suffit pour que la protéine ne soit pas retenue sur la colonne. Les fractions d'intérêt (détermination par gel SDS PAGE et lecture du profil d'élution à 280 nm) sont rassemblées et concentrées pour être intégrées dans le gradient de saccharose de l'isoélectrofocalisation.

En sortie d'isoélectrofocalisation, le pH et la densité optique à 280 nm de chaque fraction sont déterminés. La protéine est retrouvée dans la zone de pH entourant son pI (5,7), généralement entre pH 4,8 et 5,8 (Figure IV-6) :

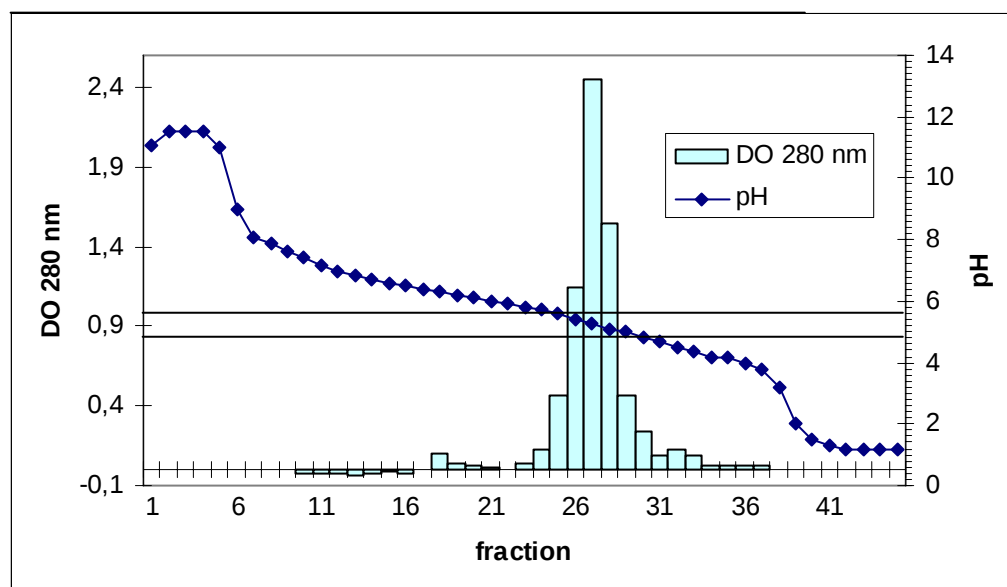


Figure IV-6 : Profil d'élution de la protéine CG18594 en sortie d'isoélectrofocalisation. La densité optique à 280 nm ainsi que le pH sont reportés pour chaque fraction.

En sortie d'IEF, la protéine est concentrée / dialysée en cellule Amicon sur membrane YM10 contre du tampon phosphate de sodium pH 6,9, NaCl 200 mM puis injectée sur une colonne de gel filtration Superdex 75 16/60. Le temps de rétention est de 75 minutes (Figure IV-7). La protéine forme un pic d'élution très majoritaire situé dans la zone théorique attendue pour une protéine d'environ 20 kDa.

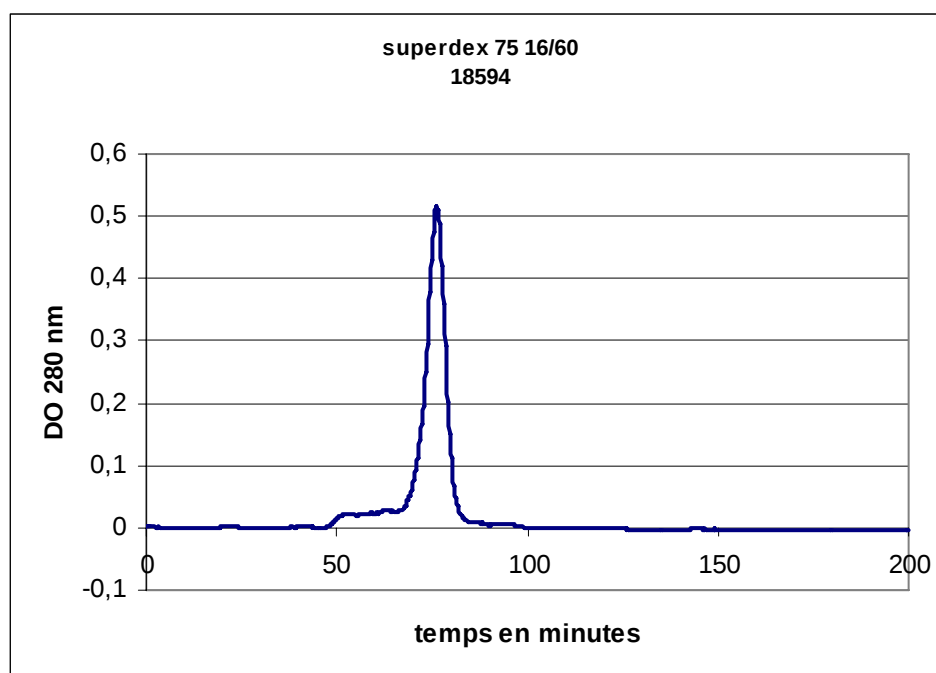


Figure IV-7 : Diagramme d'élution de la protéine CG18594 sur colonne Superdex 75 16/60 en tampon phosphate 50 mM, NaCl 0,2M. Débit : 1 mL/min. Injection 2mL.

Aucun pic ne laisse suspecter la présence d'oligomère de la protéine.

Le pic collecté est ensuite concentré par ultrafiltration jusqu'à une concentration de 1 mM déterminée par densité optique à 280 nm en considérant un coefficient d'extinction de $15930 \text{ M}^{-1} \cdot \text{cm}^{-1}$ (Expasy, Protparam). La protéine est ensuite aliquotée et congelée à -20°C , la lyophilisation entraînant la formation d'agrégats irréversibles lors de la mise en solution de la poudre (cf. RMN et DLS *infra*).

3.2. Recherche des conditions RMN

La détermination précise des meilleures conditions de tampon et de sels est un point critique et absolument incontournable avant toute étude structurale par RMN. En effet la protéine étudiée doit être repliée, rester parfaitement stable le plus longtemps possible et

sa solubilité doit être importante. Pour cela, la composition de la solution doit être optimisée : nature et concentration du tampon, pH, sels, additifs divers, co-facteurs etc. Généralement de nombreuses informations proviennent des études biochimiques qui ont été menées sur la protéine, par exemple des tests d'activité dans différentes conditions. La caractérisation biochimique de la protéine est donc une aide considérable pour la détermination des meilleures conditions pour l'étude structurale.

L'étude de la protéine CG18594 est récente, aucune caractérisation biochimique n'est pour le moment disponible. Ainsi, pour rechercher les meilleures conditions nous avons mis en oeuvre et adapté plusieurs techniques. Nous avons d'abord utilisé la technique des gouttes suspendues, empruntée aux cristallographes, pour cribler un grand nombre de conditions différentes de tampon et de sels. Nous avons ainsi pu déterminer les conditions dans lesquelles la protéine ne précipitait pas. Ensuite, nous avons testé les meilleures conditions par diffusion dynamique de la lumière afin d'optimiser la mono dispersion de la protéine en solution. Enfin le dichroïsme circulaire et des essais préliminaires de RMN ont permis de vérifier le repliement correct de la protéine.

3.2.1. Étude de conditions en gouttes suspendues

La méthode des gouttes suspendues, très utilisée en cristallographie a été adaptée afin de tester dans un volume minimal le comportement de la protéine CG18594 dans des conditions proches de celles de la RMN : concentration en protéine de l'ordre du millimolaire, différentes conditions de solvant, sels, etc. Une observation à la loupe binoculaire permet de visualiser facilement la solubilité totale ou une précipitation partielle ou totale de la protéine dans la goutte. Les gouttes de 4 μ L sont réalisées en mélangeant 3 μ L de protéine CG18594 à 1,3 mM dans l'eau avec 1 μ L de tampon quatre fois concentré. La goutte déposée sur une lamelle est retournée au dessus d'un réservoir contenant le tampon seul, à la concentration identique à celle de la goutte (Figure IV-8).

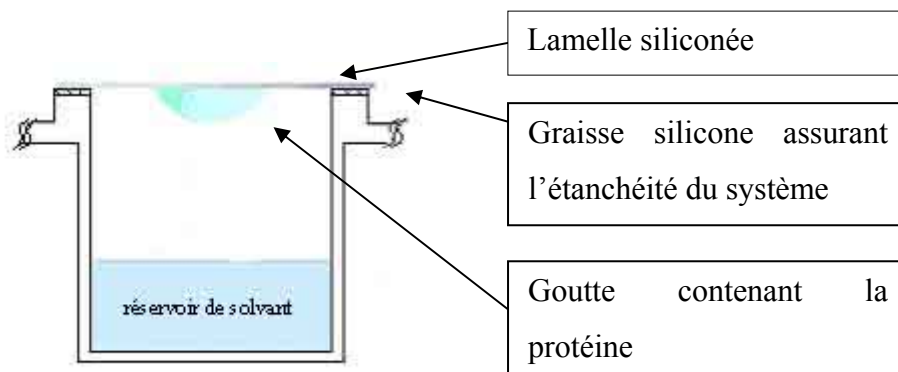


Figure IV-8 : Schéma du montage d'une goutte suspendue.

Ainsi, les conditions suivantes ont été testées :

- 20°C, Acétate de potassium 100 mM pH 4,5, 5 ou 5,5 en présence de 0, 50 ou 100 mM NaCl
- 20°C, phosphate de potassium 100 mM pH 6, 6,5 ou 7 en présence de 0, 50 ou 100 mM NaCl
- 20°C, tampon dit « physiologique », dont la composition est détaillée dans le tableau ci-dessous :

Tampon KPi pH 7,4	50 mM
NaCl	140 mM
KCl	5 mM
MgCl ₂	0,8 mM
CaCl ₂	1,8 mM

Tableau IV-3 : composition du tampon dit « physiologique »

- 20°C, eau milliQ
- 20°C, CHAPS 5, 7,5, 12 ou 20 mM dans KPi 100 mM pH 7
- 20°C, LiCl 50, 100 ou 200 mM dans KPi 100 mM pH 7
- 20°C, arginine 200 mM ou 400 mM dans KPi 100 mM pH 7
- 5°C, tampon citrate-phosphate 100 mM pH 3,5, 5,5, 6, 6,5 ou 7.

Les gouttes suspendues ont été observées régulièrement à la loupe binoculaire équipée d'un polarisateur de lumière, durant 4 mois. Nous avons obtenus les résultats suivants :

Les conditions dans lesquelles on ne distingue pas de précipité :

- 20°C, tampon phosphate de potassium 100 mM pH 6,5 et 7 en absence de NaCl.
- 20°C, eau milliQ.
- 20°C, LiCl 100 mM, tampon KPi 100 mM pH 7.
- 20°C, arginine 400 mM, tampon KPi 100 mM pH 7.

Les conditions dans lesquelles on distingue une précipitation légère, qui apparaît plus ou moins rapidement (quelques jours) :

- 20°C, tampon dit « physiologique ».
- 5°C, tampon citrate phosphate 100 mM pH 3,5, 5,5, 6, 6,5 et 7.
- 20°C, LiCl 50 et 200 mM, tampon KPi 100 mM pH 7.
- 20°C, arginine 200 mM, tampon KPi 100 mM pH 7.

Les conditions dans lesquelles on distingue une précipitation totale, qui apparaît rapidement (en moins d'une heure) :

- 20°C, tampon acétate de potassium 100 mM pH 4,5, 5, 5,5 en présence de 0, 50 ou 100 mM NaCl.
- 20°C, tampon phosphate de potassium 100 mM pH 6, 6,5, 7 en présence de 50 ou 100 mM NaCl.
- 20°C, CHAPS 5, 7.5, 12, 20 mM dans KPi 100 mM pH 7 (concentration en protéine CG18594 : 0,5mM)

Ces études en gouttes suspendues permettent donc une première approche efficace pour tester les conditions de la RMN, caractérisées par une forte concentration en protéine de l'ordre du millimolaire. Cette méthode est relativement économe en protéine et facilement mise en oeuvre mais ne permet qu'un résultat macroscopique du comportement de la protéine en solution. Cependant nous pouvons d'ores et déjà écarter les conditions où nous observons la formation de précipité et nous concentrer sur les conditions les plus

favorables. Nous avons donc retenu les conditions dans lesquelles la protéine CG18594 n'a pas précipité.

Ces résultats n'étant que « macroscopiques », nous devons caractériser l'état des protéines en solution dans ces conditions.

3.2.2. Diffusion dynamique de la lumière (DLS)

Les études de DLS sont utilisées notamment lorsque l'on suspecte la présence d'agrégats dans des solutions de protéines. La diffusion dynamique de la lumière a été utilisée afin d'observer la monodispersion des espèces de la protéine CG18594 en solution. Cette méthode permet d'obtenir aussi une estimation du pourcentage de chaque espèce en solution ainsi que leur rayon hydrodynamique.

3.2.2.a. Principe de la méthode

La DLS est une méthode importante pour la caractérisation des états oligomériques des protéines. Cette technique est fondée sur la mesure du mouvement brownien des particules (mouvement aléatoire des particules provoqué par les impacts des molécules du solvant sur la surface de la particule). La fréquence et l'amplitude de ce mouvement brownien dépendent de la taille de la particule et de la viscosité du solvant : à une viscosité donnée, plus la particule est petite et plus son mouvement brownien présente une fréquence et une amplitude relative importante. De même, plus la valeur de la viscosité du solvant est faible (température plus élevée par exemple), plus ce mouvement brownien est rapide et ample pour une particule de taille donnée. Il est donc important de maîtriser la viscosité et donc la température du milieu. Le déplacement des particules sous l'effet de fluctuations de concentrations est caractérisé par le coefficient de diffusion de translation (D_T) qui, d'après la loi de Stokes-Einstein, est lui-même proportionnel au diamètre hydrodynamique de la particule :

$$d_h = \frac{k_b T}{3\pi\eta_0 D_T}$$

Où k_b est la constante de Boltzmann, T la température, η_0 la viscosité du solvant, D_T le coefficient de diffusion et d_h le diamètre hydrodynamique de la particule considérée.

Le montage optique est constitué d'une source laser monochromatique (laser HeNe, longueur d'onde $\lambda=632.8$ nm), d'une cellule de mesure et d'un photomultiplicateur (choisi pour la mesure d'intensités diffusées très faibles générées par des particules de très petites tailles). Le laser est équipé d'un filtre spatial et d'une optique de focalisation du faisceau au milieu de la cellule. En général, le photomultiplicateur (PM) est placé sur un goniomètre afin d'obtenir un angle d'incidence variable. Cependant, dans nos expériences, nous travaillons à angle fixe, le détecteur étant immobilisé dans une direction perpendiculaire à la direction de l'onde incidente.

Le signal mesuré par le photomultiplicateur correspond aux fluctuations de la lumière diffusée dues au mouvement brownien des particules. Ce signal est ensuite traité par un corrélateur numérique afin d'obtenir la fonction d'autocorrélation du signal. Cette fonction représente la probabilité de retrouver une particule à un instant $t+\tau$ au même point de l'espace où elle se trouvait à l'instant t : plus une particule est petite et plus cette probabilité est faible, plus une particule possède un grand diamètre et plus cette probabilité est forte. Cette fonction d'autocorrélation (G) est une exponentielle décroissante du temps dans le cas d'une distribution monomodale de particules. Sa forme est beaucoup plus complexe dans le cas d'une distribution de particules de tailles différentes. On ajuste alors différents modèles mathématiques à la fonction d'autocorrélation issue de la mesure. Le résultat de ce traitement mathématique permet de donner une bonne approximation de la dispersion de l'échantillon si celui-ci n'est pas exagérément polydisperse.

Cependant l'analyse devient très difficile, voire impossible si l'échantillon contient une distribution de très grandes (plusieurs micromètres) et de très petites particules (quelques nanomètres) colloïdales. En effet, les grandes particules, même présentes en faible concentration, contribuent de façon majoritaire à l'intensité diffusée et peuvent de ce fait occulter la contribution des petites particules, pourtant majoritaires, à l'intensité totale diffusée. En effet, d'après la théorie de Rayleigh et dans la limite des tailles de particules petites devant la longueur d'onde de la lumière, l'intensité diffusée est proportionnelle au diamètre de la particule élevé à la puissance 6 (une particule de 100 nm diffuse un million de fois plus qu'une particule de 10 nm).

Pour la plupart de nos expériences, nous avons adopté le mode d'analyse monomodal qui consiste à décrire la décroissance de la fonction d'autocorrélation comme s'il n'existait

qu'une sorte de particules sphériques et à inclure le faible écart à la décroissance exponentielle parfaite dans un coefficient de polydispersité (analyse des cumulants) :

$$\log(G) = a + b\tau + c\tau^2 + d\tau^3 + \dots (2)$$

où b représente la taille moyenne des particules, et $2c/b^2$ la variance de la distribution ou polydispersité.

La méthode des cumulants offre des résultats intéressants pour une distribution monomodale étroite, définie par le diamètre moyen des particules et leur polydispersité.

3.2.2.b. Application à la protéine CG18594

Après avoir déterminé grâce aux gouttes suspendues un certain nombre de conditions dans lesquelles la protéine CG18594 ne précipitait pas, nous avons vérifié et optimisé la mono dispersion de la protéine en solution. Par ailleurs, les premiers spectres RMN tracés laissaient pressentir un phénomène d'agrégation (cf infra), nous avons voulu tester cette hypothèse et étudier l'effet de la lyophilisation sur la protéine. Les études de diffusion dynamique de la lumière ont été réalisées au laboratoire de Virologie Moléculaire & Structurale UMR CNRS 2472 - INRA 1157 à Gif sur Yvette. Le spectromètre utilisé est le modèle Dynapro – MSTC DLS de chez Protein Solutions.

Le tableau suivant reprend les résultats obtenus, en détaillant pour chaque espèce détectée, son rayon hydrodynamique (Å), le pourcentage de l'intensité lumineuse détectée et la masse correspondante (Tableau IV-4) :

Echantillon	Rayon hydrodynamique (Å)	% d'intensité	% de masse
Protéine lyophilisée 0,67 mM	195,43	75,8	99,9
	1128,35	24,2	0,1
Protéine lyophilisée 0,25 mM	194,20	85	100
	1370,29	15	0
Protéine lyophilisée 0,22 mM, tampon KPi 60 mM pH 7	218,38	95,6	100
	1482,13	4,4	0
Protéine lyophilisée 0,22 mM, tampon KPi 60 mM pH 7 (2eme mesure)	163,11	78,3	99,9
	1173,75	21,7	0,1
Protéine lyophilisée 0,22 mM, tampon KPi 60 mM pH 7, Arginine 200 mM	284,50	66,3	99,7
	1080,41	33,7	0,3
Protéine lyophilisée 0,22 mM, tampon KPi 60 mM pH 7, Arginine 400 mM	157,72	84,8	100
	1356,30	15,5	0
Protéine lyophilisée 0,67 mM, tampon Tris 66 mM pH 9	Non mesurable	Non mesurable	Non mesurable
Protéine lyophilisée 0,67 mM, KSCN 15 mM	188,22	78,8	99,9
	1254	21,2	0,1
Protéine lyophilisée 0,5 mM KSCN 25 mM	227,49	87,5	100
	1308,21	12,5	0
Protéine lyophilisée 0,67 mM, 150 mM LiCl	144,26	73,6	99,9
	1251,92	26,4	0,1
Protéine non lyophilisée 0,18 mM, NaCl 430 mM	1,99	16,5	100
	86,32	76,2	
	1438,76	7,3	
Protéine non lyophilisée 0,09mM	0,04	?	?
	2,44	83,5	100
	44,80	15,4	0
Protéine non lyophilisée 0,29 mM	2,35	53,5	100
	78,8	30,3	0
	1406,6	16,1	0
Protéine non lyophilisée 0,29 mM, Arg 200 mM	2,91	40,4	100
	11,71	40,5	0
	1224,44	19,1	0

Tableau IV-4 : Résultats des expériences de DLS réalisées sur la protéine CG18594 lyophilisée (en bleu) ou non (en rouge) dans différentes conditions de tampon

Ces résultats indiquent très clairement que la protéine CG18594 ne supporte pas la lyophilisation : quelque soit la concentration ou la composition de la solution, la protéine ayant été lyophilisée montre toujours une majorité d'agrégats, masquant le signal d'éventuels monomères (dont le rayon hydrodynamique est de l'ordre de 2 à 3 Å en fonction du milieu) alors que pour la protéine non lyophilisée les monomères sont détectables.

L'agrégation des échantillons lyophilisés semble irréversible. La protéine qui a été lyophilisée forme des agrégats que l'on n'a pas réussi à dissocier, ni en diluant la protéine, ni avec KSCN, ni avec du LiCl, ni avec de l'arginine 200 mM/400 mM, ni avec du tampon Tris pH 9. L'arginine est connue pour ses capacité de solubilisation et de stabilisation des protéines à long terme (Golovanov *et al.*, 2004).

Pour la protéine n'ayant pas été lyophilisée, la concentration entraîne une augmentation de la quantité d'agrégats, même s'ils restent en quantité minime. L'arginine ne semble pas avoir un effet important. Le sel à forte concentration semble favoriser la formation d'agrégats.

Nous retiendrons essentiellement de cette étude de diffusion dynamique de la lumière que la protéine CG18594 ne doit pas être lyophilisée.

3.2.3. Dichroïsme circulaire

Le dichroïsme est une méthode de choix pour étudier la structure secondaire des protéines. Nous avons mis à profit cette méthode afin d'observer le comportement de la protéine dans différentes conditions physico-chimiques. Nous avons ainsi observé le comportement de la protéine lors du chauffage de 5°C à 85°C en fonction du pH.

3.2.3.a. Effet du pH

Les spectres ont été enregistrés entre 195 et 260 nm à l'aide d'un dichrographe Jasco 810 en utilisant une cuve de trajet optique 1 mm, avec une vitesse de défilement de 50 nm.min⁻¹. La concentration de la protéine était de 0,2 mg/mL dans du tampon KPi 10 mM, pour des valeurs de pH variant d'une unité entre 2 et 13. La Figure IV-9 ci-

dessous présente les spectres obtenus à 5°C en fonction du pH. A pH 13 la qualité du signal n'a pas permis l'enregistrement complet des spectres.

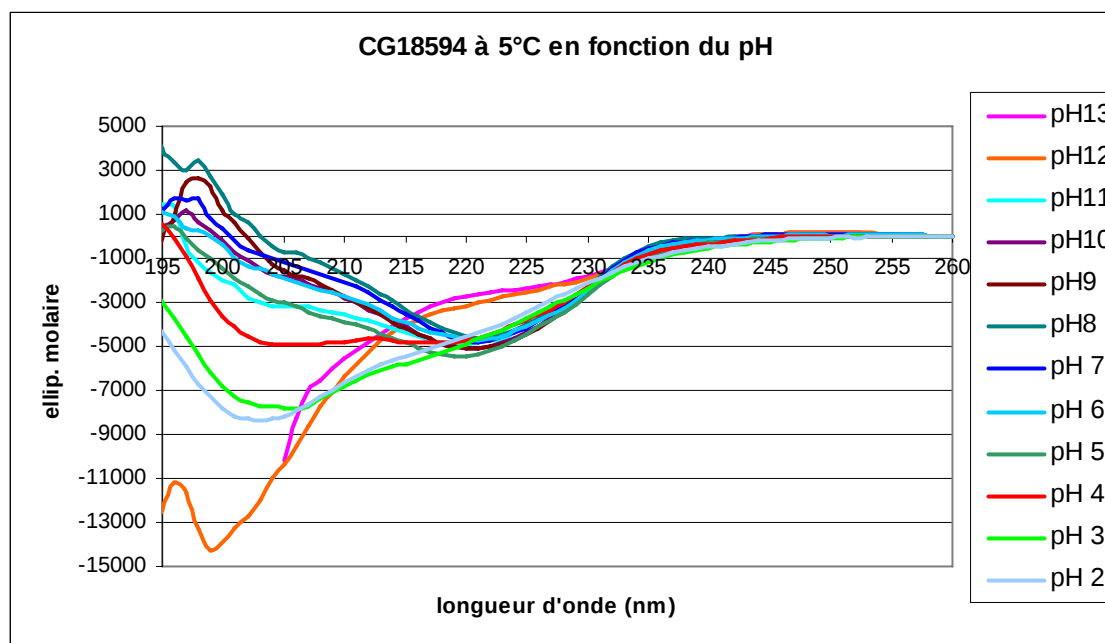


Figure IV-9 : Spectres de dichroïsme circulaire dans l'UV lointain en fonction du pH à 5°C

A pH 7 le spectre mesuré est caractérisé par un minimum à 221 nm. La déconvolution de ce spectre avec l'algorithme « self consistent method 3 » indique 10,5% d'hélices α et 27,8% de feuillets β .

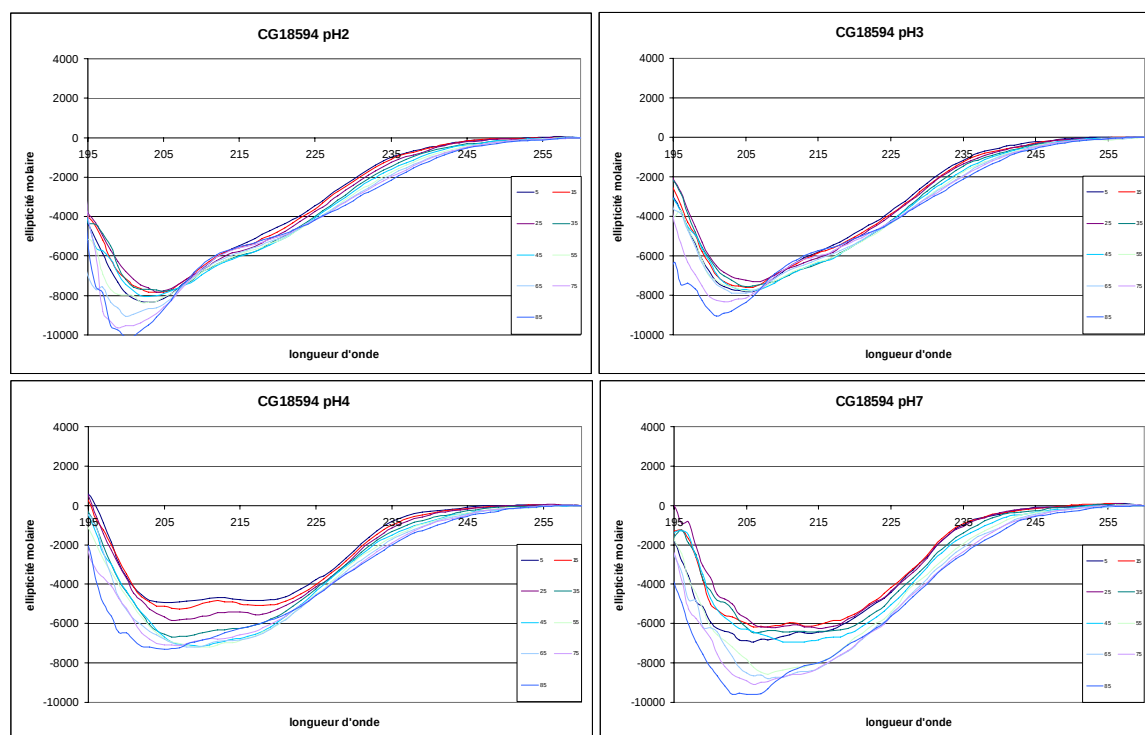
Au-delà de pH 12 nous pouvons observer un spectre caractéristique d'une protéine dénaturée alors qu'à pH 2 le spectre mesuré est caractérisé par un minimum à 203 nm. Ce spectre ne correspond pas à un spectre de protéine dénaturée classique. La déconvolution avec l'algorithme « self consistent method 3 » indique 9% d'hélice α et 19% de feuillet β . Ainsi, la protéine semble avoir gardé certaines structures secondaires, mais nous n'avons pas d'informations quant à sa structure tertiaire. En combinant ces données avec les résultats de la RMN (*cf infra*) il semble probable que nous ayons affaire à pH 2 à une structure de type « molten globule ».

Les spectres tracés aux pH intermédiaires entre 7 et 2 semblent indiquer une transition progressive depuis la forme native à pH neutre jusqu'à la forme molten globule à pH très acide. Il n'a pas été possible de déterminer si le passage « molten globule » → « native »

était réversible ou non dans la mesure où la protéine précipite lorsque l'on tente de remonter le pH après passage à pH acide

3.2.3.b. Comportement de la protéine lors du chauffage, en fonction du pH

Nous avons chauffé la protéine CG18594 de 5°C à 85°C à différents pH (Figure IV-10). Il est à remarquer que la protéine CG18594 a précipité lors du chauffage à pH 5 et 6. Ces spectres ne peuvent donc pas être interprétés dans la mesure où la concentration finale de protéine en solution n'est pas connue avec exactitude. Ceci confirme la plus faible solubilité de la protéine lorsque le pH de la solution s'approche du point iso-électrique (5,72). A pH 13 la qualité du signal n'a pas permis l'enregistrement complet des spectres.



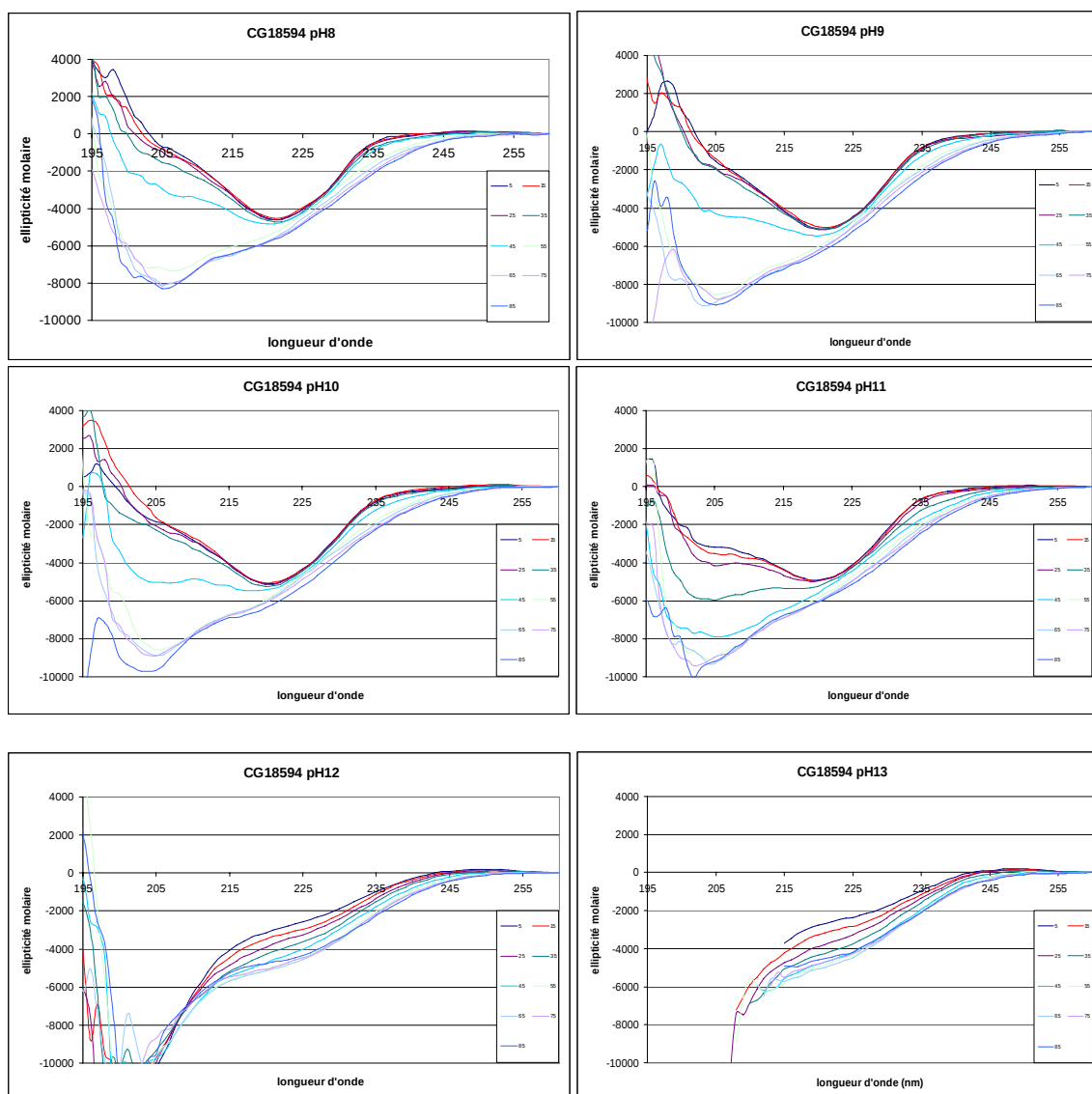


Figure IV-10 : spectres de dichroïsme circulaire obtenus lors de la dénaturation thermique de la protéine CG18594 à différents pH. Pour chaque spectre la température est indiquée en degrés Celsius, l'ellipticité molaire en $\text{degré} \cdot \text{dmol}^{-1} \cdot \text{cm}^2$ et la longueur d'onde en nanomètres. La même échelle est utilisée pour tous les spectres.

Globalement nous pouvons observer que quelque soit le pH, on tend à haute température vers une même structure caractérisée par un minimum entre 202 et 205 nm (les spectres réalisés à pH 5 et 6 ne sont pas présentés pour cause de précipitation). Il est intéressant de remarquer que les spectres à haute température se rapprochent des spectres obtenus à des pH très acides : il ne s'agit pas de spectres communément obtenus pour des formes de protéines dénaturées. La déconvolution des spectres à haute température avec l'algorithme « self consistent method 3 » indique environ 9% d'hélices α et 20% de feuillets β . Nous sommes donc en présence d'un cas de dénaturation non conventionnelle.

Le dichroïsme circulaire dans l'UV lointain est une méthode très utile pour l'observation des structures secondaires mais qui ne permet pas d'obtenir d'information quant à la structure tertiaire des protéines. Par la suite, les études RMN réalisées à pH très acide (pH 2,3) ont montré l'absence de structure tertiaire (cf. infra). Nous pouvons donc conclure que nous avons affaire à une forme de globule en fusion ou « molten globule », c'est-à-dire une protéine pour laquelle certaines structures secondaires persistent malgré la perte de la structure tertiaire.

Cette étude nous indique qu'en dessous de pH 7 nous risquons de rencontrer en solution un mélange de formes « native » et « molten globule ». Ceci n'est pas favorable à une étude structurale par RMN. Il faut aussi remarquer que les résultats du dichroïsme circulaire ne peuvent pas être adaptés directement pour la RMN dans la mesure où la différence de concentration nécessaire pour réaliser ces deux études est significative (0,2 mg/mL pour le dichroïsme circulaire contre 20 mg/mL pour la RMN).

3.3. Résonance magnétique nucléaire

Les études de conditions par gouttes suspendues et par diffusion dynamique de la lumière nous ont rapidement orientés vers le tampon phosphate de potassium à pH 7, sans sel. Le dichroïsme circulaire nous a démontré qu'à pH inférieur à 7 nous risquons de trouver au moins deux espèces en solution. Le pH pour les études RMN nous est donc imposé.

3.3.1. Stabilité de la protéine CG18594 dans le temps

Dans un premier temps nous avons testé la stabilité de la protéine dans le temps. En effet une étude RMN nécessite une protéine parfaitement stable pendant au moins une dizaine de jours. Nous avons donc observé la protéine à 20°C en fonction du temps. Voici le spectre ^1H - ^{15}N HSQC initial de la protéine CG18594 à 20°C (Figure IV-11) :

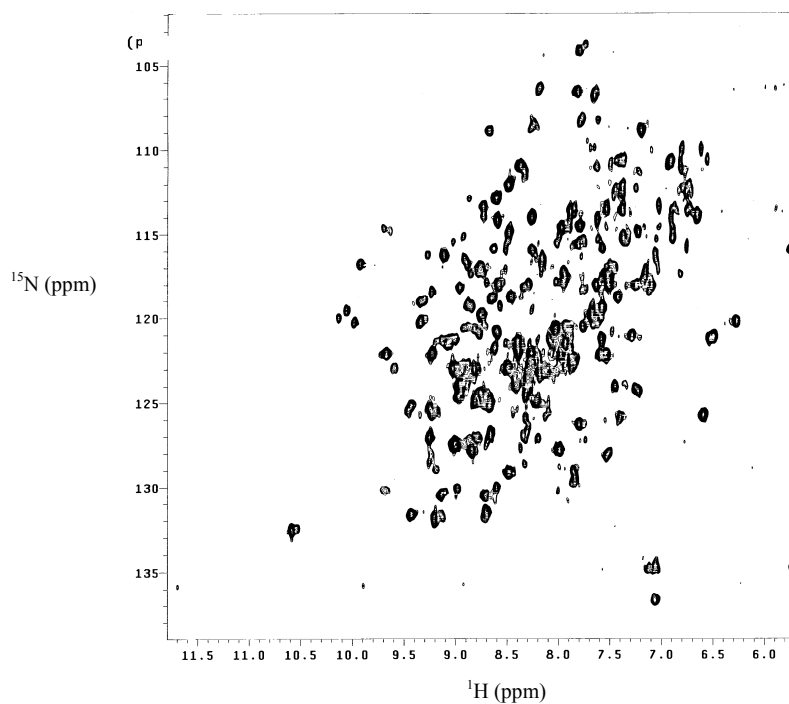


Figure IV-11 : Spectre ^1H - ^{15}N HSQC initial de la protéine CG18594 à 20°C, tampon KPi 100 mM pH 7.

Voici le spectre ^1H - ^{15}N HSQC obtenu après 4 jours à 20°C (Figure IV-12) :

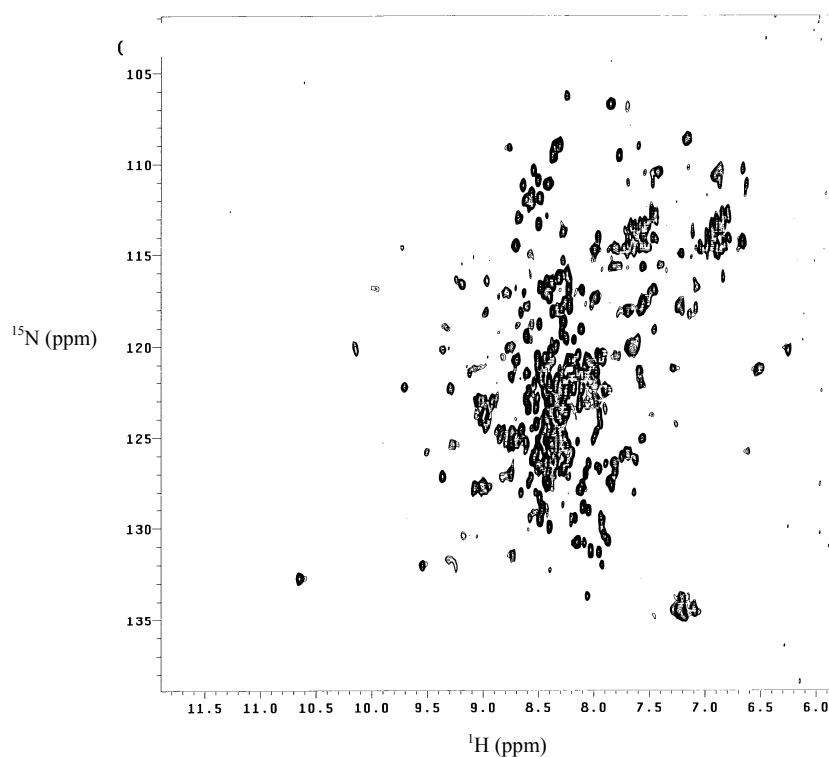


Figure IV-12 : Spectre ^1H - ^{15}N HSQC de la protéine CG18594 après 4 jours à 20°C, tampon KPi 100 mM pH 7.

La comparaison de ces deux spectres ^1H - ^{15}N HSQC nous permet de tirer les conclusions suivantes :

- La protéine CG18594 est bien structurée à pH 7 : les corrélations ^1H - ^{15}N sont bien dispersées sur le spectre ^1H - ^{15}N HSQC.
- Les raies sont larges et non homogènes, même pour une protéine de 20 KDa, ce qui laisse supposer que la protéine est sous forme agrégée.
- La protéine est dénaturée en 4 jours à 20°C.

3.3.2. RMN à 5°C

Face à ce constat, nous avons été contraints de travailler à 5°C : la protéine est alors stable durant au moins une semaine. Cependant, cette température n'est pas optimale pour les études RMN car la diminution de la température conduit à une augmentation du temps de corrélation τ_c de la protéine. Ceci modifie les propriétés de relaxation longitudinale et transversale du signal RMN et conduit à un élargissement des raies, il y a perte d'aimantation durant les expériences de transferts d'aimantation transversale (TOCSY).

A 5°C, nous avons observé :

- L'absence de signal TOCSY : le système relaxe extrêmement rapidement, le signal est perdu durant la durée du spin lock.
- Nous n'observons aucun transfert d'aimantation hétéronucléaire au niveau des spectres HNHA.
- La largeur des raies des spectres ^1H - ^{15}N HSQC semble démontrer, malgré l'effet de la température, que la protéine CG18594 est sous forme agrégée.

Ces résultats indiquent que la détermination de la structure 3D de la protéine CG18594 n'est pas réalisable dans les conditions actuelles. Les spectres protons ne sont pas interprétables par manque de signal (TOCSY) et par manque de résolution (NOESY). Les techniques de RMN hétéronucléaire ne semblent pas utilisables dans la mesure où l'expérience de base de transfert HNHA ne donne aucun résultat satisfaisant (absence de signal).

3.3.3. Agrégation de la protéine en présence de sel, à 5°C

La largeur des raies des différents spectres ainsi que la vitesse de relaxation du signal lors des expériences TOCSY nous ont amenés à émettre l'hypothèse d'une agrégation de la protéine CG18594. Cette hypothèse, étudiée par DLS, a été confirmée pour la protéine lyophilisée mais pas pour la protéine non lyophilisée. Cependant, la protéine non lyophilisée présente des spectres RMN très proches de la protéine lyophilisée en terme de largeur de raie. La protéine non lyophilisée semble donc également agrégée.

Nous avons observé l'effet de l'ajout de sel NaCl sur l'agrégation de la protéine en réalisant une gamme de concentration (0, 100, 200, 300, 400, 500 mM) par RMN, et en traçant pour chaque concentration un spectre ^1H - ^{15}N HSQC et la première expérience du spectre HNHA. Lorsque la concentration en sel augmente, le signal ^1H - ^{15}N HSQC diminue progressivement (Figure IV-13). Couplé aux expériences de diffusion dynamique de la lumière, nous pouvons dire que l'ajout de sel NaCl entraîne la formation d'agrégats par la protéine CG18594. Il faut remarquer que la formation d'agrégats n'est pas synonyme de précipitation pour la protéine CG18594, celle-ci reste soluble dans une solution parfaitement limpide.

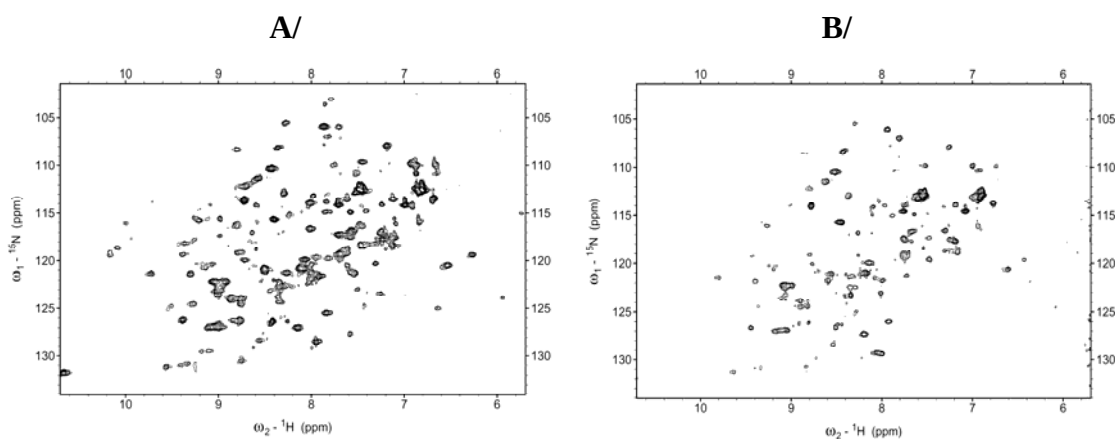


Figure IV-13 : Spectre ^1H - ^{15}N HSQC de la protéine CG18594 (1mM dans KPi 100 mM pH 7, 5°C) en absence de NaCl (A) et avec 500 mM de NaCl (B).

Le NaCl semble donc renforcer l'agrégation de la protéine CG18594. Ces données sont en accord avec les observations précédentes faites par la méthode des gouttes suspendues.

3.3.4. Effet de la température sur les spectres RMN

Face à ces observations, nous avons voulu savoir si la mauvaise qualité des spectres obtenus à 5°C, où la protéine est stable, était due à la température ou si elle était due à l'agrégation probable de la protéine. Pour répondre à cette question nous avons recherché un témoin de protéine de la même taille que la protéine CG18594, et dont la structure et le comportement en solution était bien connus. Nous avons utilisé pour cela la β -lactoglobuline (18,3 kDa, 162 résidus), monomérique et correctement repliée à pH 2, à 5°C et à 40°C.

Les spectres 1D démontrent parfaitement l'affinement des raies avec l'augmentation de la température (Figure IV-14) :

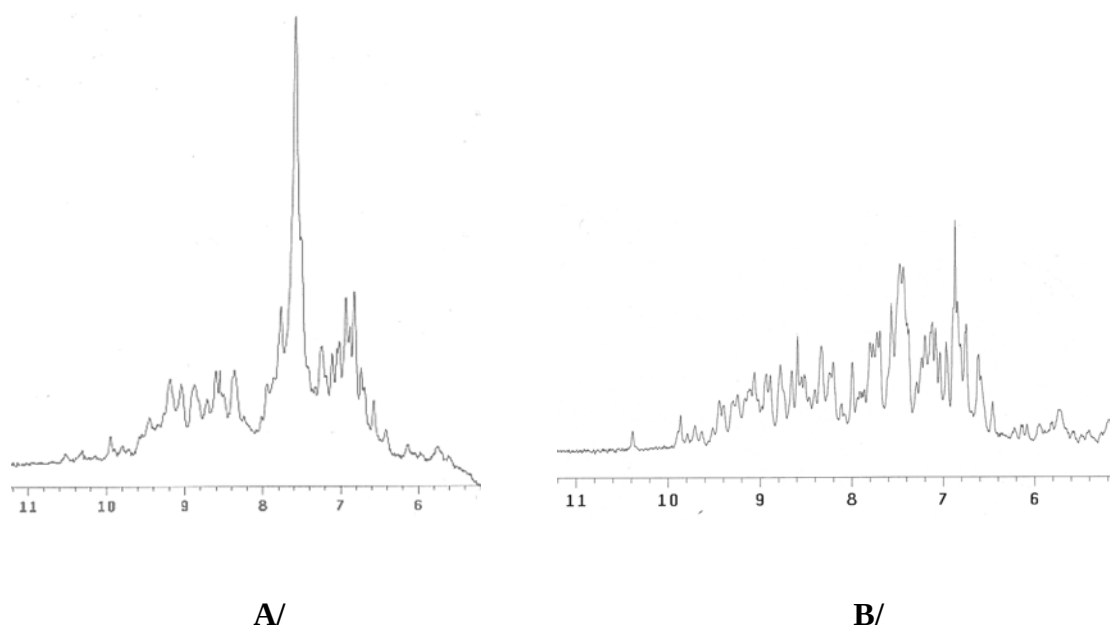


Figure IV-14 : Effet de la température sur la largeur des raies. Spectres proton 1D de la β -lactoglobuline (2mM, pH1,9, HCl 20mM) à 5°C (A/) et 40°C (B/).

L'effet de la température sur la largeur des raies est donc spectaculaire. Toutefois, à 5°C les raies de la β -lactoglobuline sont plus fines que celle de la protéine CG18594 à 5°C (Figure IV-15). Ceci tend à confirmer l'agrégation de la protéine CG18594.

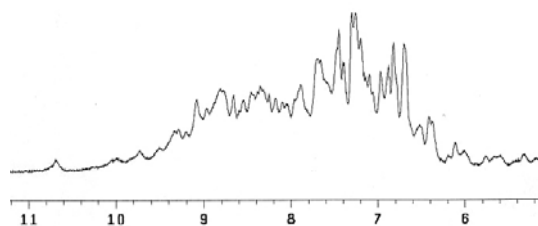


Figure IV-15 : Spectres proton 1D de la CG18594 (1mM, NaPi 60 mM pH 6,9) à 5°C.

De la même manière l'effet de la température est aussi particulièrement visible au niveau des propriétés de relaxation de la β -lactoglobuline : cette protéine présente un transfert TOCSY acceptable à 40°C alors qu'à 5°C la relaxation se fait très rapidement, entraînant la perte rapide du signal lorsqu'on allonge la durée du spin lock (Figure IV-16).

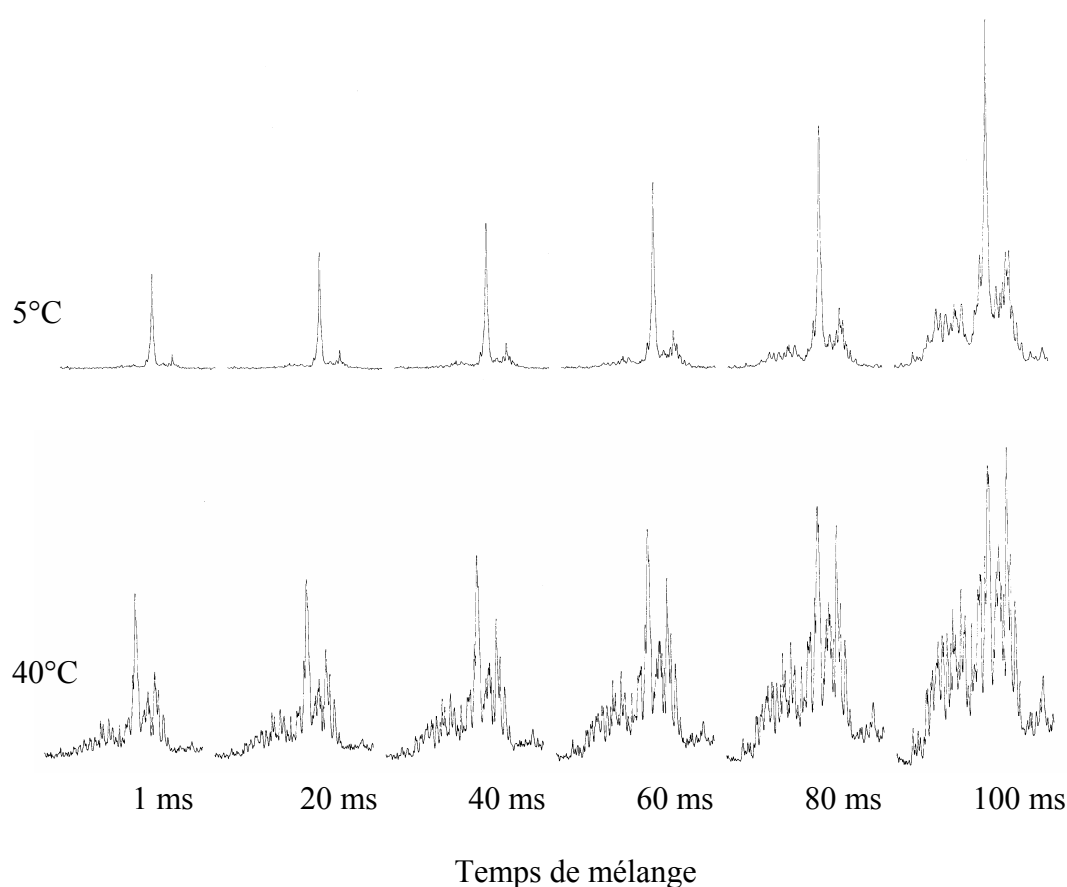


Figure IV-16 : Effet de la température sur les propriétés de relaxation. Spectres proton TOCSY 1D de la β -lactoglobuline (2 mM, pH 1,9, HCl 20mM) en fonction du temps de mélange à 5°C et 40°C.

Nous montrons donc que la mauvaise qualité des spectres obtenus pour la protéine CG18594 est principalement liée à la température : les spectres de la β -lactoglobuline, protéine de référence, ne sont pas exploitables s'ils ont été acquis à 5°C, alors qu'ils sont

parfaitement interprétables à 40°C. A cet effet de température s'ajoute l'effet de l'agrégation de la protéine.

A pH 7 et 5°C, l'étude de la structure tridimensionnelle de la protéine CG18594 par RMN semble donc compromise.

3.3.5. Etude de la forme molten globule

Les études de dichroïsme circulaire décrites précédemment montrent que la protéine CG18594 conserve des structures secondaires lorsqu'elle est chauffée à haute température ou placée à pH très acide (*cf. supra*). Nous avons voulu déterminer si une structure tertiaire était conservée, ou si nous avons affaire à une forme « molten globule », caractérisée par la présence de structures secondaires mais l'absence d'une structure tertiaire bien définie. Pour cela, nous avons étudié un échantillon de protéine CG18594 à pH 2,3.

La Figure IV-17 montre le spectre ^1H - ^{15}N HSQC que nous avons obtenu à pH 2,3 :

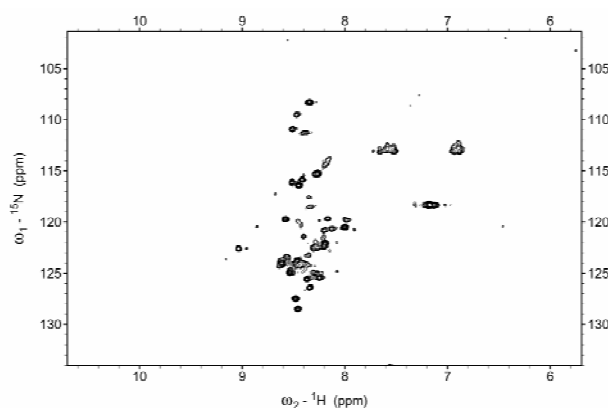


Figure IV-17 : Spectre ^1H - ^{15}N HSQC de la protéine CG18594 à pH 2,3 : la protéine est totalement dénaturée (0.5 mM dans KPi 100 mM pH 2,3).

L'absence de dispersion des corrélations $^{15}\text{N}/^1\text{H}$ montre sans ambiguïté l'absence de structure tertiaire de la protéine CG18594 à pH 2,3. Nous pouvons donc conclure à la présence d'une forme « molten globule » à pH très acide dans la mesure où les spectres de dichroïsme circulaire mettent en évidence la présence de structures secondaires (environ

9% d'hélices alpha et 20% de feuillets β à pH 2) et où la RMN démontre l'absence de structure tertiaire. Les formes de globule en fusion ne sont pas fréquemment observées chez les protéines. On considère actuellement qu'elles sont des intermédiaires de repliement dont l'étude fondamentale fournit des informations importantes pour la compréhension des mécanismes qui dirigent le repliement des protéines.

3.3.6. Conclusion

L'étude RMN de la β -lactoglobuline nous permet donc de conclure que la température de 5°C où la protéine CG18594 est stable, n'est pas compatible avec une étude structurale par RMN. En effet, à basse température le signal RMN relaxe très rapidement : le signal disparaît rapidement durant les périodes d'évolution des séquences RMN.

Par ailleurs nous avons montré que la protéine CG18594 formait un globule en fusion ou « molten globule » à haute température ou à pH acide.

4. Conclusion

Nous avons réussi durant ce travail à exprimer la protéine CG18594 (marquée ^{15}N ou non) et à la purifier en quantité importante. La protéine produite est soluble à pH 7 (solubilité supérieure à 1,33 mM) et repliée. Cependant la protéine n'est pas stable plus de quelques jours à une concentration de 1 mM et à 20°C. La protéine est stable à 5°C mais les spectres enregistrés sont difficilement utilisables (largeur de raie importante, absence de transfert d'aimantation transversale, manque de résolution). En utilisant la β -lactoglobuline, qui est une protéine dont la taille (162 résidus, 18,3 kDa) est proche de celle de la protéine CG18594, nous avons réussi à montrer l'effet majeur de la température sur la qualité des spectres. Nous constatons donc la non faisabilité de l'étude de la structure tridimensionnelle de la protéine CG18594 par RMN dans les conditions actuelles.

Les études par dichroïsme circulaire ont par ailleurs montré un processus de dénaturation non conventionnel de la protéine CG18594 avec persistance de structures secondaires à

haute température ou à pH très acide. L'étude par RMN à pH 2,3 de la protéine montre l'absence de structure tertiaire : nous avons affaire à une forme « molten globule ».

La RMN ne semble donc pas être une méthode adaptée à l'étude structurale de la protéine CG18594 puisque sa stabilité ne permet pas l'enregistrement serein de spectres de qualité. La protéine est instable à température ambiante, a tendance à s'agréger et à adopter une forme de globule en fusion. Cette protéine est en cours d'essai de cristallisation dans le cadre d'une collaboration au laboratoire avec l'équipe d'Alain Roussel.

Etude de la protéine CG10298

V. Etude de la protéine CG10298

Face à l'impossibilité de réaliser une étude structurale de la protéine CG18594 par RMN, nous avons décidé de réaliser le clonage des sections codantes des gènes *cg10298* et *cg7054*. Nous avons choisi le gène *cg10298* car la protéine correspondante est *a priori* la plus proche des PEBP-1 bovine et humaine en termes d'homologie de séquence (environ 50 % d'identité) et de conservation des structures secondaires. De plus le point isoélectrique de cette protéine est basique (9,15), tout comme celui des protéines bovine et humaine, ce qui est empiriquement favorable à une étude structurale par RMN.

Les PEBP-1 humaine et bovine sont les protéines les mieux caractérisées de la famille des PEBP, tant au niveau biologique qu'au niveau physico-chimique et structural. La comparaison de leurs relations structure / fonctions avec la protéine CG10298 semble donc prioritaire.

Après avoir cloné dans *Escherichia coli* la région codante du gène *cg10298* dans le vecteur d'expression pET31b, j'ai mis au point le protocole de purification et de marquage de la protéine CG10298. Actuellement, nous pouvons la produire purifiée, marquée ou non, en grande quantité. Une étude initiale de sa stabilité en solution a été réalisée par RMN: la protéine CG10298 n'est stable qu'environ 10 jours à 20°C. Dans la mesure où la protéine CG7054, qui a été produite à la même période, s'est avérée immédiatement très stable, nous avons décidé de réaliser l'étude structurale sur cette dernière. Cependant l'étude de la protéine CG10298 n'est pas abandonnée, la stabilité de la protéine CG10298 pourra probablement être améliorée par la suite en jouant sur les conditions de tampon et de sel dans l'échantillon.

La protéine purifiée a été mise à profit pour produire des anticorps de lapin spécifiques. Ces anticorps ont permis la localisation de la protéine chez la drosophile : elle est uniquement exprimée dans les testicules.

L'ensemble du travail réalisé est présenté dans l'article soumis au « Journal of Biochemistry (Tokyo) » : «*Drosophila melanogaster CG10298 is a new testis-specific member of the phosphatidylethanolamine binding protein family.* »

1. Données de base :

La séquence du gène *cg10298* est disponible sur la banque de données Flybase (<http://flybase.bio.indiana.edu/>). Ce gène est situé sur le bras droit du chromosome III, en position 83E4, entre les gènes *cg17917* et *cg 17919* (Figure V-1).

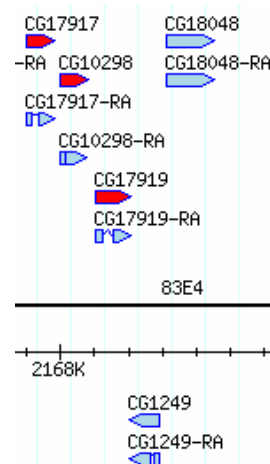


Figure V-1: localisation du gène *cg10298* sur le bras droit du chromosome III (image issue de Flybase).

Le gène *cg10298* a une longueur de 799 nucléotides sur l'ADN génomique, il est composé de deux exons séparés par un intron. La traduction de la portion codante fournit une séquence protéique qui présente la signature caractéristique de la famille des PEBP (Figure V-2).

1	11	21	31	41	51	
1	11	21	31	41	51	
1	MSDSTVCFSK	HKIVPDILKT	CPATLLTVTY	GGGQVVDVGG	ELTPTQVQSQ	PKVKWDADPN
61	AFYTLLLTDP	DAPSRKEPKF	REWHHWLVVN	IPGNQVENG	VLTEYVGAGP	PQGTGLHRYV
121	FLVFKQPQKL	TCNEPKIPKT	SGDKRANFST	SKFMSKYKLG	DPIAGNFFQA	QWDDYVPKLY
181	KQLSGKK					

Figure V-2: séquence de la protéine CG10298. La séquence consensus de la famille des PEBP est indiquée en rouge.

L'analyse de cette séquence a été réalisée avec l'outil Protparam du site Expasy (Gasteiger *et al.*, 2003), (Figure V-3).

Nombre d'acides aminés: 187

Masse moléculaire: 20871,9 Da

pI théorique: 9,15

Composition en acides aminés:

Acides aminés	Code	Nombre	Fréquence	Acides aminés	Code	Nombre	Fréquence
Ala	(A)	8	4,3%	Leu	(L)	15	8,0%
Arg	(R)	4	2,1%	Lys	(K)	19	10,2%
Asn	(N)	7	3,7%	Met	(M)	2	1,1%
Asp	(D)	11	5,9%	Phe	(F)	9	4,8%
Cys	(C)	3	1,6%	Pro	(P)	16	8,6%
Gln	(Q)	11	5,9%	Ser	(S)	10	5,3%
Glu	(E)	6	3,2%	Thr	(T)	14	7,5%
Gly	(G)	15	8,0%	Trp	(W)	4	2,1%
His	(H)	4	2,1%	Tyr	(Y)	7	3,7%
Ile	(I)	5	2,7%	Val	(V)	17	9,1%

Figure V-3: Analyse de la structure primaire de la protéine CG10298 avec l'outil Protparam (<http://us.expasy.org/tools/protparam.html>)

La séquence présente 3 cystéines dont il faut tenir compte. Les cystéines 7 et 21 sont engagées dans un pont disulfure (*cf. infra*), la troisième est donc réactive et il faut se placer dans un milieu réducteur pour éviter la formation de dimères.

A partir de cette analyse des acides aminés de la protéine CG10298, nous pouvons tracer la charge nette théorique moyenne de la protéine en fonction du pH : (Expasy, titration curve) (

Figure V-4). Cette courbe nous sera particulièrement utile lors du choix des méthodes de purification : à pH 8 la protéine est chargée positivement, contrairement à une bonne partie des protéines cellulaires.

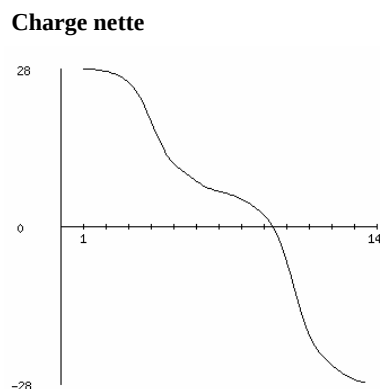


Figure V-4: Charge nette théorique moyenne de la protéine CG10298 en fonction du pH

1.1. Comparaison de la protéine CG10298 avec les autres PEBP de drosophile

La séquence de la protéine CG10298 a été alignée avec celles des 6 autres PEBP de drosophile (Figure V-5).

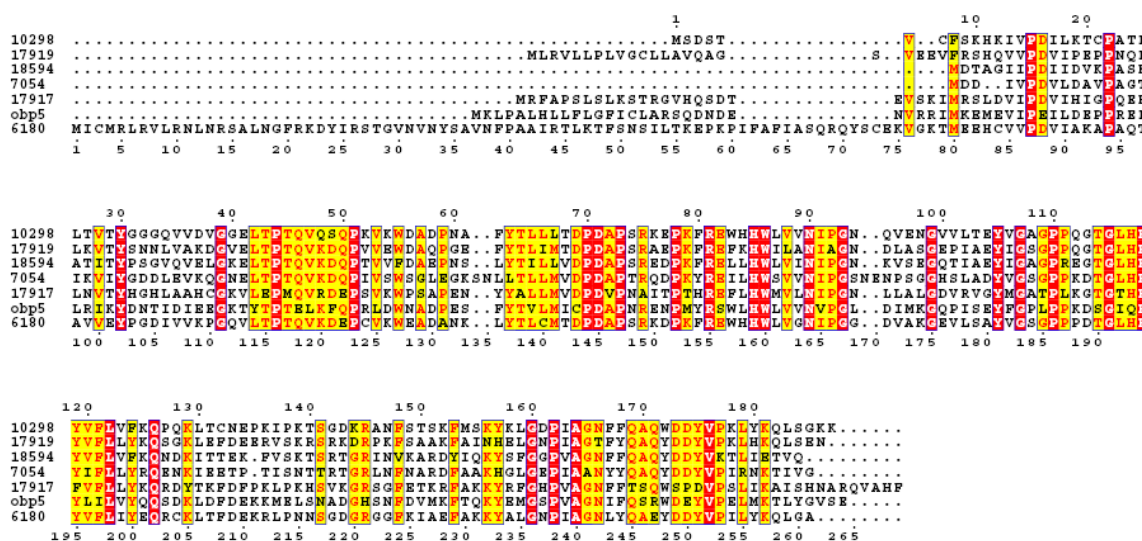


Figure V-5: Alignement de la protéine CG10298 avec les 6 autres PEBP de drosophiles. L'alignement a été réalisé avec le logiciel CLUSTALW, (Eddy, 1995). Les cadres rouges indiquent les résidus identiques, les résidus non conservés, mais similaires, sont indiqués en jaune. Les structures secondaires obtenues sur les modèles cristallographiques sont indiquées. (Figure réalisée avec le logiciel Esript).

À partir de cet alignement nous pouvons calculer le pourcentage d'identité entre la protéine CG10298 et les six autres PEBP de drosophile. Les valeurs obtenues sont indiquées dans le Tableau V-1 :

	% d'identité avec la protéine CG10298
CG6180	55
CG17919	54
CG18594	55
CG7054	48
CG17917	41
OBP5	40

Tableau V-1 : Pourcentage d'identité de la protéine CG10298 avec les autres PEBP de drosophile. L'alignement a été réalisé avec le logiciel CLUSTALW.

La protéine CG10298 présente un pourcentage d'identité proche de 55% avec les protéines CG6180, CG17919, et CG18594. L'alignement est par ailleurs un peu moins bon avec les protéines CG17917 et l'OBP5.

1.2. Comparaison de la protéine CG10298 avec les PEBP des autres espèces

La séquence de la protéine CG10298 a été alignée avec celle des autres PEBP de structures tridimensionnelles connues (Figure V-6).

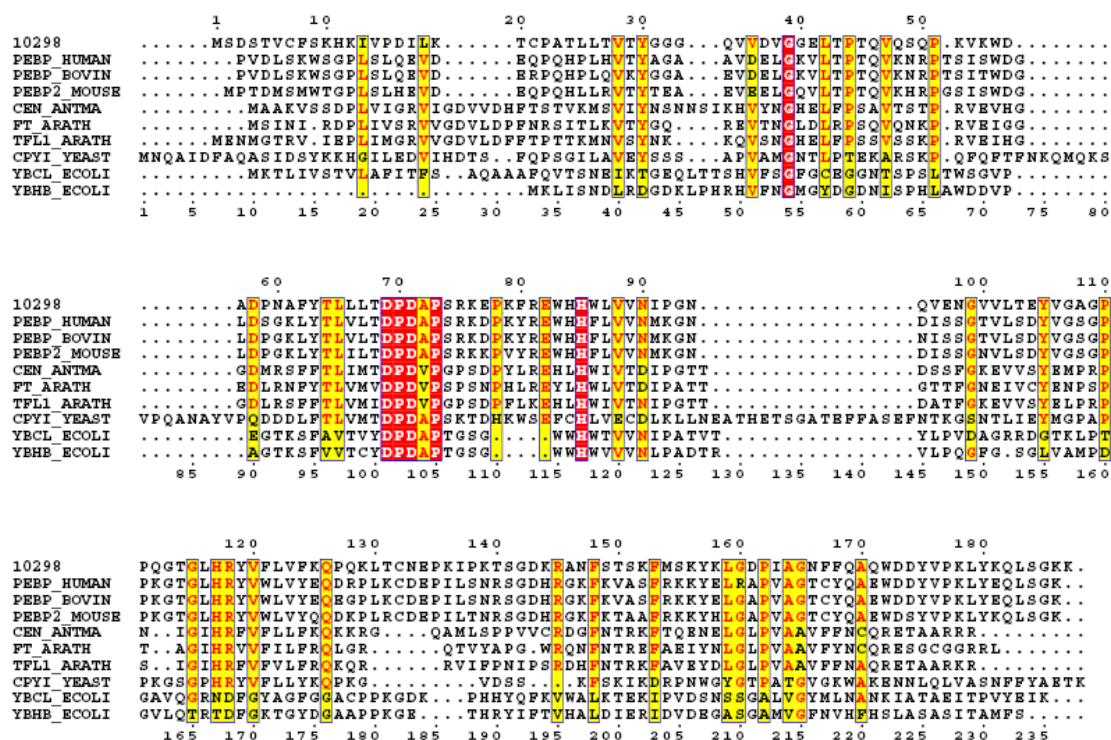


Figure V-6: Alignement de la protéine CG10298 avec différentes PEBP dont la structure 3D est connue. L'alignement a été réalisé avec le logiciel CLUSTALW, (Eddy, 1995). Les cadres rouges indiquent les résidus identiques, les résidus non conservés, mais similaires, sont indiqués en jaune. Les structures secondaires obtenues sur les modèles cristallographiques sont indiquées. (Figure réalisée avec le logiciel Esript).

Cet alignement montre bien que les zones conservées entre différentes espèces sont moins nombreuses qu'entre les 7 PEBP de drosophile. Les protéines YBHB et YBCL de la bactérie *Escherichia coli* sont les plus éloignées.

Le tableau V-2 indique les pourcentages d'identité observée entre la protéine CG10298 et ces différentes PEBP :

	% d'identité avec la protéine CG10298
PEBP-1 humaine	51
PEBP-1 bovine	52
PEBP-2 souris	52
Antirrhinum Cen	28
Arabidopsis FT1	33
Arabidopsis TFL1	30
Saccharomyces TFS1	26
E. coli YBCL	12
E. coli YBHB	13

Tableau V-2 : Pourcentage d'identité de la protéine CG10298 avec les PEBP de structure 3D connue. L'alignement a été réalisé avec le logiciel CLUSTALW

A partir de ces alignements, nous pouvons conclure que sur la base de sa séquence la protéine CG10298 appartient bien à la famille des PEBP. Elle possède la séquence consensus caractéristique des PEBP mais aussi une conservation de résidus significative avec les autres membres de la famille. Les PEBP les plus proches sont les PEBP humaines, bovines et murines (43 % d'identité). L'alignement avec les protéines de plantes, levure (TFS1) ou de bactéries (YBHB et YBCL) présente un pourcentage d'identité plus faible. Les PEBP de drosophile sont donc plus proches entre elles qu'avec celles des autres espèces.

2. «*Drosophila melanogaster* CG10298 is a new testis-specific member of the phosphatidylethanolamine binding protein family. »

Drosophila melanogaster CG10298 is a new testis-specific member of the phosphatidylethanolamine binding protein family.

Gilles Rautureau, Laurence Jouvensal, Françoise Vovelle, Françoise Schoentgen, Daniel Locker and Martine Decoville¹.

¹ To whom correspondence should be addressed (email decovil@cnrs-orleans.fr).

Centre de Biophysique Moléculaire, UPR 4301 CNRS affiliated to Orléans University and to INSERM, Rue Charles Sadron, 45071 Orléans cedex 2, France

SUMMARY

The phosphatidylethanolamine binding proteins (PEBPs) family is evolutionary conserved and involved in different physiological phenomena. PEBPs were found in many species from bacteria to mammals. Despite numerous studies, PEBPs biological function and mode of action remain elusive. Based on sequence homology, seven genes were detected in the *Drosophila* genome. Only one of them, the odorant binding proteins (OBP) has been characterized. To date nothing is known concerning the expression pattern and biological roles of the six other PEBP genes. By comparing structural parameters of the six PEBP proteins with those of human PEBP-1, also named raf kinase inhibitor protein (RKIP), we have established that PEBP CG10298 is most closely related to human PEBP/RKIP. Additionally, we have shown by RT-PCR and Western blotting in wild type and male-sterile mutants that CG10298 is expressed only in the testes during spermatid differentiation.

Keyword : PEBP, RKIP, Testis, *Drosophila*, Expression

The phosphatidylethanolamine binding protein (PEBP) family is an evolutionary conserved group of proteins found in a wide range of species, including bacteria (1), yeast (2), plants (3, 4), nematodes (5, 6), insects (7) and mammals (8-12). Initially PEBP was described in bovine brain as a 23 kDa cytosolic protein (13). It was reported to bind hydrophobic ligands and especially phosphatidylethanolamine (14). Sequence alignment of all known PEBPs revealed a specific signature: DPDXP-(X)₁₁-H present in the whole family, except for bacteria which display the signature DPDXP-(X)₆-H (15). Based on sequence homology with human PEBP and on the presence of the signature, several genes encoding PEBPs have been identified in most species. Indeed, two PEBPs are present in *E. coli* and in most bacteria (1), two PEBPs were also identified in the yeast *Saccharomyces cerevisiae* (2). Six genes encoding putative PEBPs are encountered in human and mouse genomes. The presence of PEBP homologues in such a diverse range of species

suggests that PEBPs may perform essential biological roles and in the same organism the different paralogues may have specific action. The PEBP family appears to be multifunctional. However, very few information is known concerning the accurate role of the different family members. In plants, PEBPs have been shown to be implicated in morphogenesis of inflorescence (3, 4, 16): in tomato, when Pro73 is replaced by Leu (spontaneous mutation) in SELF-PRUNING protein (a member of PEBP family), the normal indeterminate inflorescence of the plant is changed in determinate inflorescence (16). In mammals, only two members of the PEBP family have been encountered to be expressed in adult tissues: one of them PEBP-1 is present in numerous tissues including liver, brain, spleen, stomach, heart, blood vessels, intestine and muscles (17). PEBP-1 seems to be implicated in cellular morphogenesis and differentiation. Particularly, it has been shown to be implicated in the growth and activation of acetylcholinergic neurones during brain

development (18) and is implicated in spermatogenesis; mainly it controls the elongation of late spermatids (19) and is a sperm decapacitation factor receptor (20). Another member of the PEBP family, PEBP-2 is expressed exclusively within the testis in rat and mouse and is involved in male fertility (21). PEBP-2 mRNA and protein are expressed at high levels within germ cells at stages of spermatogenesis temporally distinct from those of PEBP-1, suggesting a functional specificity and an integral role in either spermiogenesis or posttesticular sperm maturation. The mode of action of the different PEBPs in cells remains to elucidate. However, in human, it has been demonstrated that PEBP-1 controls MAP-kinase (22) and NFkB (23) signal pathways by inhibiting several kinases implicated in these pathways. Thus, PEBP-1 interacts directly with Raf1, MEK and ERK to control the MAP-kinase pathway and with TAK1, NIK, IKK α and IKK β to control the NFkB pathway. In all cases, PEBP-1 binds to the kinases preventing interactions with their partners. As Raf1 was the main kinase to be inhibited by PEBP-1, some authors have renamed human PEBP-1 as Raf-Kinase-Inhibitor-protein (RKIP) (22). To date, PEBP/RKIP is considered to be a metastases suppressor in cancer as the presence of PEBP/RKIP in tumors decreases metastases development (24). Moreover, PEBP/RKIP sensitizes tumor cells to chemotherapy (25) and immunotherapy (26). Thus, PEBP/RKIP is considered as a new therapeutic target (24) and its amount, in tumor cells, is a factor of prognostic (27).

Seven genes belonging to the PEBP family are present in the whole genome of *Drosophila melanogaster*. However one of them, named Odorant Binding Protein gene (OBP) is expressed in the insect antenna (7). Consistent with the particular role of OBP, its sequence does not display the complete signature of the PEBP family; thus, we will not consider OBP in this paper. Concerning the six other genes encoding PEBP paralogues, two clusters of PEBPs genes are localised on the right arm of chromosome III: one containing three genes (CG17917, CG10298 and CG17919), the other one containing two genes (CG7054 and CG18594). The sixth gene being localised on the left arm of chromosome II (CG6180). Recently, a proteomic analysis of the immune response of *Drosophila* has revealed that CG18594 (28) and CG17919 (29) are up regulated after infection with Gram-positive bacteria. However, nothing is known concerning the expression and the biological function of these six genes in *Drosophila*.

The present work compares the biophysical properties of the six *Drosophila* PEBPs and human PEBP-1, also called RKIP. Sequence alignments, hydrophobic cluster analysis (HCA), CD and

molecular modelling comparisons suggest that CG10298 is the most similar to human PEBP/RKIP. We report herein the expression pattern of CG10298. Our findings based on Western blotting and RT-PCR suggest that CG10298 is expressed only in testes, during spermatid differentiation.

MATERIALS AND METHODS

Comparative modelling- Structural models of CG10298 were built using MODELLER 6.2 (30). MODELLER is a program based on the satisfaction of spatial restraints generated on the target structure (structure to model) from its sequence alignment with templates of known 3D structures. The PROCHECK program (31) was used to assess the stereochemical quality of the structures. Further evaluations were completed using the programs ANOLEA (32) and VERIFY3D (33), (http://www.doe-mbi.ucla.edu/Services/Verify_3D/) and also by visual examination of the structures on a graphic display. Depending on the assessment scores, a representative model was selected and energy minimised with the SYBYL software (TRIPOS Inc., St Louis, MO). The secondary structures of the model were analysed with the MOLMOL program (34).

Cloning of CG10298 ORF into pET31b vector- cDNA from *Drosophila melanogaster* Oregon-R purchased from Berkeley *Drosophila* Genome Project (<http://www.fruitfly.org/DGC/>) (clone ID: GH28351) was used as a direct template for PCR. In order to produce a highly soluble protein, it was decided to produce CG10298 without any tag (35). The forward primer (5'-CATATGTCGATTCCACCGTGTGC-3') included a *Nde* I restriction site at the 5' end of the ORF (underlined) while the reverse primer (5'-CTCGAGCTACTTCTTGCCAGATAGTTGC-3') added a *Xho* I restriction site at the 3' end of the ORF (underlined). The PCR protocol for 30 cycles was as follows: 94°C for 1 min, 52°C for 1 min, 70°C for 1 min. The Taq DNA polymerase Dynazyme (Ozyme) was used. The PCR product was cloned into plasmid pCRII-TOPO (Invitrogen kit) and transformed into Top 10 cells using the commercial standard protocol. The pCRII-TOPO-10298 plasmid was purified using a standard protocol (36), digested with *Nde* I and *Xho* I restriction endonucleases (New England Biolabs), and ligated with the *Nde* I-*Xho* I digested expression vector (pET-31b(+)) (Novagen) using T4 DNA ligase (Invitrogen). The product was transformed into *E. coli* C600 competent cells and the sequence of the cloned ORF (in the pET-31b expression vector) was confirmed by DNA sequencing (MWG Biotech). One clone was

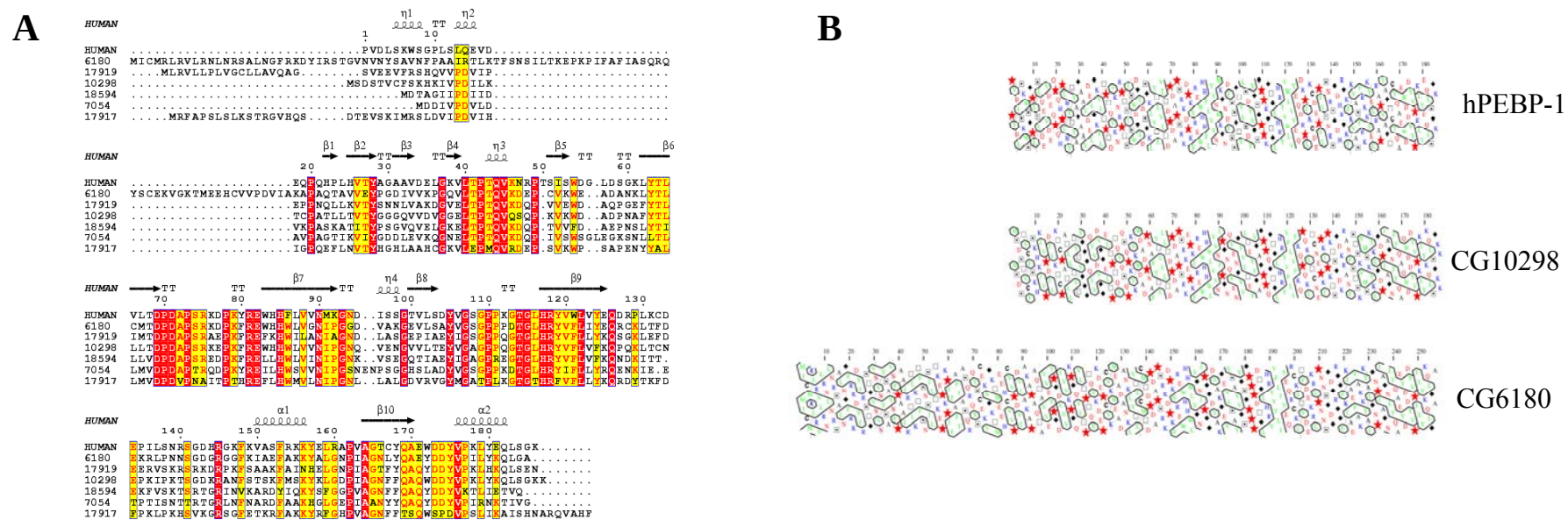


Figure 1. CG10298 is closely related to hPEBP-1. (A) Sequence alignment of the 6 PEBPs from *Drosophila melanogaster* with hPEBP-1 (CLUSTALW). Red boxes indicate sequence identities and yellow boxes conservative changes. Secondary structures of hPEBP-1 are indicated in cartoon form above the alignment: arrow: β -strand, helix: α -helix. (B) HCA plot of hPEBP-1, CG10298 and CG6180. Similarities between the distribution of clusters and their features (amino acids indicated in green letters and surrounded by a blue line) are readily observed between the different proteins confirming the excellent conservation of secondary structures. The symbols used in the plot are as follows: open square, threonine; square with dot inside, serine; diamond, glycine; and star, proline.

selected and transformed into *E. coli* strain BL21(DE3) (Novagen) for protein expression.

Purification of CG10298- The frozen cell pellet was thawed in 30ml ice-cold QAE buffer (Tris HCl 50 mM pH 8, NaCl 50 mM, EDTA 1 mM, NaN₃ 0.05%). The suspension was sonicated and centrifuged. The supernatant, divided in two aliquots, was loaded on two 30 cm x 2.4 cm QAE Sephadex A-50 columns (the theoretical isoelectric point is 9.15). The flow was 35ml h⁻¹ and 3ml fractions were collected. CG10298 was eluted and detected by 15% SDS-PAGE. CG10298 was pooled, dialysed against 20mM Na₂HPO₄, pH 7, concentrated to 1mM (Amicon cell) and frozen at -20°C. The final protein purity was assessed according to 15 % SDS-PAGE gel and mass spectrometry. The quantity of CG10298 was evaluated using an extinction coefficient of 31,720 M⁻¹ cm⁻¹ at 280nm (theoretical estimation from <http://us.expasy.org/>). The protocol for over-expression in bacteria of ¹⁵N-labeled CG10298 was adapted from (35).

NMR spectroscopy -The NMR sample contained 2 mM of ¹⁵N-labelled CG10298 in 50 mM KH₂PO₄ buffer, 5mM mercaptoethanol, 1mM DSS (2,2-dimethyl-2-silapentane-5-sulfonate sodium salt) in H₂O/D₂O (90:10). The pH was adjusted to 5.5. NMR spectra were acquired on a 600 MHz Varian INOVA spectrometer. ¹H-¹⁵N HSQC spectra were recorded at 20°C and referenced relative to the internal DSS signal (0.0 ppm). The NMR spectra were processed with NMRPipe (37) and analyzed using NMRView (38) software.

Characterization of CG10298 by Circular Dichroism- CD measurements were recorded on a Jasco J-810 CD spectropolarimeter. Spectra were measured at 5°C, in the far-UV range (195-260 nm) in 1-mm pathlength quartz cells at a protein

concentration of 0.2 mg.ml⁻¹ (10 mM KH₂PO₄ buffer, pH 7). Each spectrum was the average of 5 consecutive scans, and all spectra were corrected by subtracting the buffer baseline.

Drosophila Strains- Flies were raised on standard medium at 22°C. *sa¹red¹/TM3*, *Sb¹*, *twe¹cn¹bw¹/CyO* and *P{ry(+t7.2)=PZ}bol¹ry⁵⁰⁶/TM3*, *ry(RK)Sb¹Ser¹* were obtained from the Bloomington Fly Stock Center. Oregon-R was used as wild-type reference strain.

Western blot analysis- Tissues were crushed in 10 µl of water, and SDS protein loading buffer (50 mM tris-HCl pH 7.5, 300 mM β-mercaptoethanol, 4% SDS, 12% glycerol, bromophenol blue) was immediately added. Samples were boiled for 5 minutes prior to loading on a 12% SDS polyacrylamide gel. Western blotting was performed using the standard method. The anti-CG10298 was used at 1/4000 dilution in 5% milk/PBST. The secondary antibody was an alkaline phosphatase-conjugated anti-rabbit IgG (Promega). For the detection, the filter was stained with NBT and BCIP reagents as described by Promega. Antibodies directed against CG10298 were obtained by injecting rabbits with the recombinant CG10298.

RNA preparation and RT-PCR amplification- Total RNA was purified using the RNeasy Plus Mini kit (Qiagen) according to the manufacturer. cDNA was synthesized with the Omniscript RT kit (Qiagen) according to the manufacturer using the primer: 5'-CTTTGTAGGGATTGGGAGCCTAATTA-3'. PCR was then performed using half of the RT product with the two primers: forward 5'-CAGGTGACTTATGGTGGTGACA-3', reverse 5'-CTACTTCTTGCCAGATAGTTGCTTGT-3'.

The PCR protocol for 45 cycles was as follows: 95°C for 1 min, 55°C for 1min, 70°C for 1 min. As a control RT-PCR was performed with primers

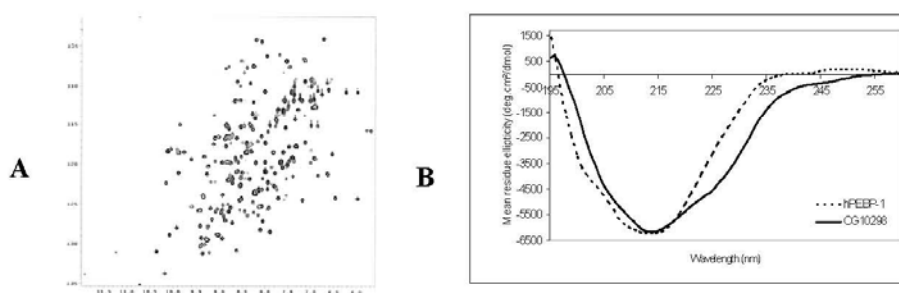


Figure 2. NMR and circular dichroism spectra of CG10298 (A) 2D NMR map (B) Circular dichroism spectra at 5°C and pH7 of hPEBP-1 (dotted line) and CG10298 (plain line).

specific of the ribosomal protein RP49: forward 5'-TTGGTGAGCGGACCGACAGCTGC-3', reverse 5'-GTATCGACAACAGAGTCGGTCGC-3'. The reverse primer was used for RT. The PCR protocol for 40 cycles was as follows: 70°C for 1 min, 40 cycles 94°C for 1 min, 40°C for 1min, 70°C for 1 min, and 70°C for 10 min.

RESULTS

Comparative structure of PEBPs- The usual way to compare proteins is to look at their sequences and at basic chemico-physical properties such as the isoelectric point. A sequence alignment of the 6 fruit fly PEBPs with hPEBP-1 performed using the ClustalW program (29) (figure 1A) reveals that the two *Drosophila* proteins that present the higher identity scores with hPEBP-1 are CG6180 (53%)

and CG10298 (51%). These two proteins are basic and their theoretical pIs estimated by ExPASy ProtParam are similar to that of hPEBP-1 (9.0 and 9.1, respectively, versus 8.3). A more elaborated approach to compare protein sequences is to use the Hydrophobic Cluster Analysis (HCA) method (39). This graphical method allows to easily visualise α -helices and β -strands, and to align the conserved hydrophobic clusters. Comparison of the HCA profiles of *Drosophila* PEBPs and of hPEBP-1 (Figure 1B) reveals that there is an almost perfect agreement between the various clusters of CG10298 and hPEBP-1, in particular in the strand regions. The only region showing a discrepancy is the N-terminus. Such agreement is not observed for the other *Drosophila* PEBPs, and especially for CG6180: though it presents the highest percentage of identity with hPEBP-1, it shows a distribution of

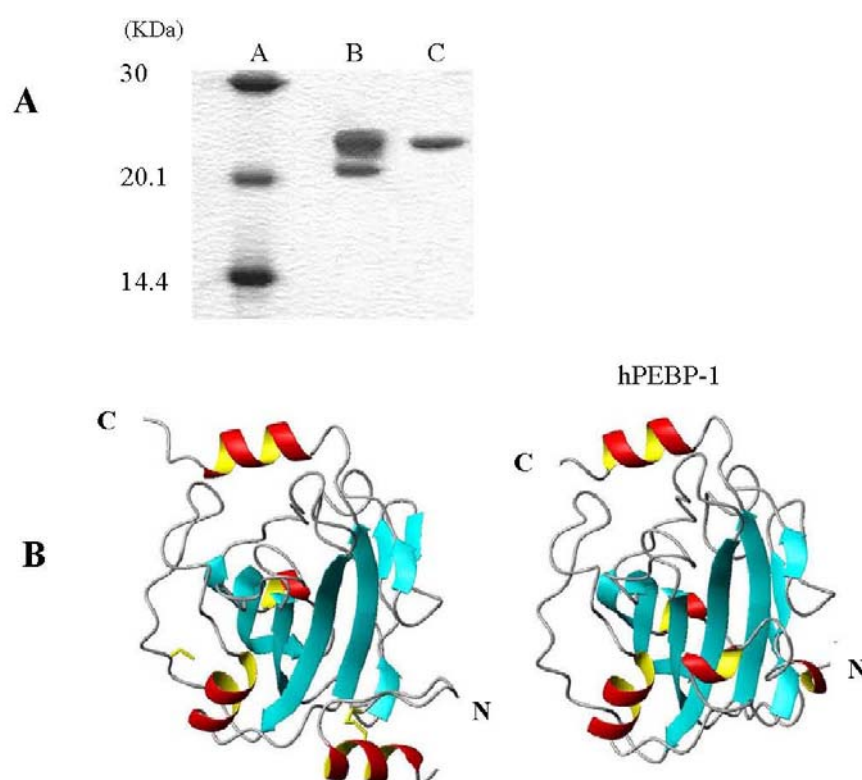


Figure 3. Modeling of CG10298 (A) SDS PAGE migration of the recombinant CG10298. Lane1: molecular markers, lane 2 purified CG10298 loaded without mercaptoethanol, lane 3 purified CG10298 loaded with 1M mercaptoethanol. Even in absence of mercaptoethanol, the sample contains the reduced form of CG10298 (upper band) revealing the presence of a disulfure bridge. (B) Comparison of hPEBP-1 with the homology model of CG10298. The model leading to the best ANOLEA score is presented on the figure.

hydrophobic clusters rather different. In addition, CG10298 exhibits a long N-terminal extension of approximately 80 residues which may be of functional importance and has never been observed in the sequence of other known PEBPs.

Molecular modelling of CG10298. The ORF was cloned, and the corresponding ^{15}N -labelled protein was over-expressed in *Escherichia coli* and purified. In order to examine whether the purified protein was correctly folded at pH 5.5, a ^1H - ^{15}N HSQC spectrum was recorded (fig. 2A). The good dispersion of the cross-peaks on this spectrum provided evidence of a proper folding of the protein. However, at room temperature, the HSQC time evolution suggested a tendency to denaturation, which prevented the determination of its 3D structure by NMR.

CD spectroscopy at 5°C and pH 7 was used to compare the conformational states of CG10298 and hPEBP-1. The CD spectra of CG10298 and hPEBP-1 (fig. 2B) show a minimum around 215 nm indicative of predominantly β -sheet structures. These data were used to evaluate the percentages of secondary structure elements of both proteins using the Dichroprot software and the "self consistent method 3" algorithm (40). This algorithm leads to 8.4% of α -helices and 21.3% of β -sheets for hPEBP-1, in total agreement with the X-ray structure (41). These values can be easily compared to 10.2% of α -helices and 23.9% of β -sheets for CG10298, indicating a slightly higher content of secondary structures in CG10298. Taken together, these data confirm the close relation of CG10298 with hPEBP-1.

Using the program MODELLER and taking hPEBP-1 as template (PDB code 1BEH), we built a structural model of CG10298. The sequence alignment proposed by the sequence recognition program mGenTHREADER (42) was used, except for the N-terminal residues: 1 to 20 that showed no sequence similarities with hPEBP-1. In order to model the N-terminal segment twenty first residues, we used data obtained from secondary structure predictions. Several methods were tested, and three of them (Pspred (43), Jpred (44) and Sopma (45)) correctly predicted the β -sheet content and the expected positions of the β -sheets and helices of hPEBP-1. These three methods predict the first ten residues of CG10298 as mainly α -helical, though the exact length of the helix varies from one method to the other. Thus, several models were built with helical constraints at the N-terminus corresponding to helices of different lengths and positions. Examination of preliminary structures evidenced that the sulfur atoms of Cys7 and Cys21 were in van der Waals contact which might indicate the presence of a disulfide bridge between these two residues. The formation of this disulfide bridge was investigated by SDS-PAGE (without reducing agent) experiments. As shown on fig 3A, the SDS-

PAGE migration profiles of the recombinant CG10298 are different under oxidative and reductive conditions. The migration profile of CG10298 encountered in testes was comparable to that of the reduced recombinant PEBP (result not shown) confirming the presence of a disulfide bridge between Cys7 and Cys21. The presence of this disulfide bridge was therefore taken into account in further structure calculations. According to evaluations performed with ANOLEA and Verify3D, the best model was found when the first helix spanned residues 3 to 10, corresponding to a helical content in agreement with the CD data (12.3% versus 10.2%). Among the models generated by MODELLER, different orientations of the N-terminal helix with respect to the rest of the protein were found. The global fold of the model leading to the best ANOLEA score is presented on figure 3B and compared to that of hPEBP-1.

Expression pattern of CG10298. In order to compare the expression pattern of CG10298 and hPEBP-1, we looked first for the presence of mRNA in various adult tissues, larvae and embryos by examining the products of RT-PCR using specific primers. As a control, we amplified by RT-PCR the RP49 mRNA, which encodes a ribosomal protein. The results presented on figure 4A revealed the presence of RT-PCR product only in testes. The RT-PCR was controlled by digestion with two restriction enzymes specific of CG10298 (*Taq* I and *Ava* II).

Expression of CG10298 protein was also investigated using antibodies raised against a recombinant CG10298 protein. Prior to use, we confirmed the specificity of the anti-10298 polyclonal antibody by Western blotting, using several recombinant *Drosophila* PEBPs. The anti-10298 antibody detected the 23 kDa band corresponding to CG10298, but none other tested PEBP (data not shown). This antibody was used to detect CG10298 protein in various tissues of larvae or adults, or in embryos. As shown in figure 4B, CG10298 is only detected in adult testes. The detected band migrated at the same position as the recombinant protein, suggesting no gross post-translational modification of the protein.

To specify the temporal distribution of CG10298 protein during spermatogenesis, we made use of various male-sterile mutants: *spermatocyte arrest* (*sa*), *boule* (*bol*) and *twine* (*twe*) mutants. Function of the *sa* gene is required both for meiotic cell-cycle progression and for the onset of spermatid differentiation. In *sa* mutants, the early stages of spermatogenesis are present, germ cell differentiation proceeds through the spermatocyte growth phase, but arrests at the G2/M transition of meiosis I, and the arrested cells fail to undergo spermatid differentiation (46). Wild-type function of *bol*, a homologue of the human Y-linked *Deleted in Azoospermia* (*DAZ*) gene (47), is also required

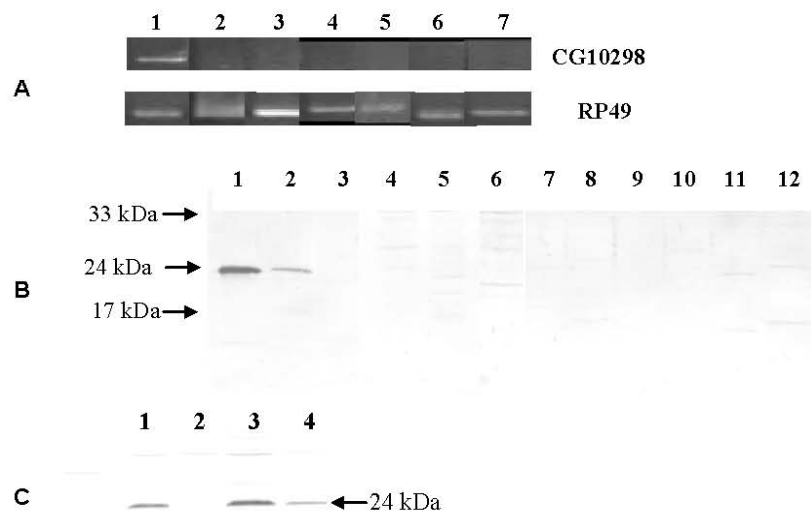


Figure 4. CG10298 is only expressed in adult testis. (A) RT-PCR using primers specific of CG10298 or of the ribosomal protein RP49 as control with RNA extracted from (1) adult testes, (2) ovary, (3) head, (4) intestine, (5) thorax, (6) larvae, (7) embryo. (B) Western blot of total proteins extracted from various tissues carried out using CG10298 antibodies. (1) recombinant CG10298 protein, (2) adult testes, (3) accessory glands, (4) ovary, (5) head, (6) intestine, (7) larval brain, (8) imaginal discs, (9) salivary glands, (10) fat body, (11) larval intestine, (12) Malpighi tubules. (C) Western blot of total testes proteins extracted from various homozygous male sterile mutants. (1) Wild type, (2) *sa¹/sa¹*, (3) *twe¹/twe¹*, (4) *bol¹/bol¹*.

for the G2/M transition. In *bol* mutants, spermatocytes are also arrested at the G2/M transition, but some of them undergo spermatid differentiation (48). Wild-type function of *twe* is also involved in the G2/M transition but not in spermatid differentiation. In *twe* mutants, spermatocytes are able to skip crucial events of meiosis, yet the spermatid differentiation program proceeds (48, 49). Protein extracts from homozygous mutant adult testes were probed by Western blot with anti-10298 antibodies (fig. 4C). Clearly, CG10298 is absent in *sa* mutant testes, but present in *bol* and *twe* mutant testes. This result suggests that CG10298 is involved in spermatid differentiation.

DISCUSSION

In this paper, we describe the characterisation of CG10298, a new member of the PEBP family in *Drosophila melanogaster*. Six ORFs encoding putative PEBPs and displaying the complete PEBP family signature have been identified in the *Drosophila* genome; sequence analysis reveals that

the higher identity scores with hPEBP-1 are found for CG6180 and CG10298. Sequence based alignments are however rather rough methods to compare proteins considering that the 3D structures are frequently much more conserved than the sequences. In the absence of 3D structure an efficient approach is to compare the HCA profiles of proteins. This method allows to delineate hydrophobic clusters and to easily identify regular secondary structures (α -helices and β -strands). These clusters corresponding to the internal faces of secondary structure elements are much more conserved than the primary structures and allow to detect significant similarities between proteins even at low sequence identity. Among the six *Drosophila melanogaster* PEBPs, only CG10298 displays a HCA profile comparable to that of hPEBP-1. The regions corresponding to the strands of the two β -sheets and to the C-terminal α -helix can be easily identified on the CG10298 HCA profile and compared to the corresponding hydrophobic regions in hPEBP-1 (Figure 1A and B), indicating that the two proteins share the same secondary structure elements. These structural similarities are

further confirmed by CD experiments (Figure 1C) revealing approximately the same content of β -sheets and α -helices in both proteins. Comparative modelling based on the X-ray structure of hPEBP-1 and on secondary structure predictions for the N-terminal part of CG10298 that does not align properly with hPEBP-1 leads to a good quality model. The global fold of the model is of course very close to that of hPEBP-1 structure: the root mean square deviation between the coordinates of the C α atoms of residues 22 to 185 is only 0.45Å. In addition, other features appear to be common to hPEBP-1 and CG10298. For example, the binding site for negatively charged ligands is well preserved, with the conservation of the three residues involved in the binding interaction: Asp69, His85 and Tyr119 (Asp70, His86 and Tyr120 in hPEBP-1 numbering). Among *Drosophila* PEBPs, Asp69 and Tyr119 are well conserved, but His85 is found only in CG6180 and CG10298 sequences. In addition, the conservation of the sequence Ala72-Pro73 and Arg81-Glu82 suggests the formation of *cis*-peptide bonds between these residues, insuring the proper geometry of the binding site. This indicates that CG10298 may interact with negatively charged ligands in the same manner as other PEBPs (50). Such features are also observed for mouse PEBP-2, which is specifically expressed in testes (51). Taken together, these results evidence structure similarities between CG10298 and hPEBP-1 that may indicate functional relationships. However, the large variability of the N-terminal amino acids among PEBPs should be considered and may be determinant for the regulation and the specificity of the various PEBPs. In this work, the absence of sequence similarity of the 20 first residues of CG10298 with PEBPs of known structure has not allowed a proper determination of the structure of this segment, which was modelled on the basis of secondary structure predictions. This region was predicted in helical conformation, a conformation already found for the N-terminal segment of TFS1, a yeast PEBP (52). Even if the modelling of this region is not fully reliable, it does not interfere with the structure of the ligand binding site, which is well preserved. In order to understand the biological function of CG10298, we have determined its pattern of expression at the transcriptional and translational levels. We found that CG10298 is only expressed in adult testes and seems to be specific of late spermatogenesis. As a matter of fact, CG10298 protein is only detected when spermatids differentiate. Mutants blocked at the G2 phase without spermatid differentiation do not express CG10298 protein. These results strongly suggest that CG10298 might have a role in late spermatogenesis.

Previously, in mammalian testes, the presence of PEBPs was very early detected, especially in rat

where the protein is very strongly overexpressed (18). However, it appeared that two PEBPs named PEBP-1 and PEBP-2 are expressed in testes. To date, it seems that PEBP-1 is expressed in the elongated late spermatids of both rat and mouse testes (18, 53), it is also present in secretions of the testis and epididymis and on the surface of spermatozoa (54, 55). Recently, PEBP-1 was described to be a sperm decapacitation factor receptor associated to the membrane (20). It is located on both the acrosomal and post-acrosomal regions of the head and along the whole of the flagellum in both human and mouse spermatozoa. The other testicular PEBP, named PEBP-2, is specific of testes (21). PEBP-2 is encountered only in germ cells. It is expressed especially in late spermatocytes, and lower concentrations are seen within round and late spermatids. It is therefore expressed earlier than PEBP-1 which is expressed in late spermatids (21).

Then two major questions remain opened: the first one concerns the possible existence of another PEBP expressed in *Drosophila* testes. To answer this question further investigations are needed and we have undertaken the search for others *Drosophila* PEBPs in testes. The second question is: what is the function of CG10298 in spermatogenesis? Based on the structural homology with the human RKIP/PEBP, the presence of CG10298 in spermatids may indicate an implication in the regulation of MAP kinase signalling during elongation or individualisation of spermatids. This hypothesis is supported by the data of Bergeret et al. (56) who have shown that a Ras-MAP kinase pathway is involved during *Drosophila* sperm differentiation. A second hypothesis, is a function of inhibitor of serine proteases involved in spermatogenesis. Such a function has already been described in mammal brain (57).

In conclusion, we have demonstrated that among the six *Drosophila* PEBPs, CG10298 is the most related to hPEBP-1. We have shown that CG10298 is only expressed during spermatogenesis, suggesting a function in late differentiation of spermatids. The molecular mode of action of CG10298 remains to be determined. So, further works focusing on relationships between CG10298, Ras-MAP kinase pathway and/or serine proteases will probably contribute to the elucidation of CG10298 function during spermatogenesis opening a wide range of perspectives. Another important question is to know if several PEBPs are expressed in *Drosophila* testes as in mammals and in this case, if they are involved in similar or different processes.

ACKNOWLEDGEMENTS

We are grateful to A. Soulas for her excellent technical assistance. Special thanks to J. L. Dacheux for providing CG10298 antibodies.

REFERENCES

1. Serre, L., Pereira de Jesus, K., Zelwer, C., Bureaud, N., Schoentgen, F. and Bénédicti, H. (2001) Crystal structures of YBHB and YBCL from *Escherichia coli*, two bacterial homologues to a Raf kinase inhibitor protein. *J. Mol. Biol.* **310**, 619-636
2. Robinson, L.C. and Tatchell, K. (1991) TFS1: a suppressor of *cdc25* mutations in *Saccharomyces cerevisiae*. *Mol. Gen. Genet.* **230**, 241-250
3. Bradley, D., Carpentier, R., Copsey, L., Vincent, C., Rothstein, S. and Coen E. (1996) Control of inflorescence architecture in *Antirrhinum*. *Nature* **379**, 791-797
4. Bradley, D., Ratcliffe, O., Vincent, C., Carpenter, R. and Coen, E. (1997) Inflorescence commitment and architecture in *Arabidopsis*. *Science* **275**, 80-83
5. Lobos, E., Altmann, M., Mengod, G., Weiss, N., Rudin, W. and Karam, M. (1990) Identification of an *Onchocerca volvulus* cDNA encoding a low-molecular-weight antigen uniquely recognized by onchocerciasis patient sera. *Mol. Biochem. Parasitol.* **39**, 135-146
6. Gems, D., Ferguson, C.J., Robertson, B.D., Nieves, R., Page, A.P., Blaxter, M.L. and Maizels, R.M. (1995) An abundant trans-spliced mRNA from *Toxocara canis* infective larvae encodes a 26-kDa protein with homology to phosphatidylethanolamine-binding proteins. *J. Biol. Chem.* **270**, 18517-18522
7. Pikielny, C.W., Hasan, G., Rouyer, F. and Rosbash, M. (1994) Members of a family of *Drosophila* putative odorant-binding proteins are expressed in different subsets of olfactory hairs. *Neuron* **12**, 35-49
8. Seddiqi, N., Bollengier, F., Alliel, P.M., Périn, J.P., Bonnet, F., Bucquoy, S., Jollès, P. and Schoentgen, F. (1994) Amino acid sequence of the *Homo sapiens* brain 21-23kDa protein (neuropeptide h3), comparison with its counterparts from *Rattus norvegicus* and *Bos taurus* species and expression of its mRNA in different tissues. *J. Mol. Evol.* **39**, 655-660
9. Moore, C., Perry, A.C.F., Love, S. and Hall, L. (1996) Sequence analysis and immunolocalisation of phosphatidylethanolamine binding protein (PBP) in human brain tissue. *Mol. Brain Res.* **37**, 74-78
10. Perry, A.C.F., Hall, L., Bell, A.E. and Jones, R. (1994) Sequence analysis of a mammalian phospholipid-binding protein from testis and epididymis and its distribution between spermatozoa and extracellular secretions. *Biochem. J.* **301**, 235-242
11. Schoentgen, F., Saccoccio, F., Jollès, J., Bernier, I. and Jollès, P. (1987) Complete amino acid sequence of a basic 21-kDa protein from bovine brain cytosol. *Eur. J. Biochem.* **166**, 333-338
12. Grandy, D.K., Hanneman, E., Bunzow, J., Shih, M., Machida, C.A., Bidlack, J.M. and Civelli, O. (1990) Purification, cloning and tissue distribution of a 23kDa rat protein isolated by morphine affinity chromatography. *Mol. Endo.* **4**, 1370-1376
13. Bernier, I. and Jollès, P. (1984) Purification and characterization of a basic 23kDa cytosolic protein from bovine brain. *Biochim. Biophys. Acta* **790**, 174-181
14. Bernier, I., Tresca, J.P. and Jollès, P. (1986) Ligand-binding studies with a 23 kDa protein purified from bovine brain cytosol. *Biochim. Biophys. Acta* **871**, 19-23
15. Schoentgen, F. and Jollès, P. (1995) From structure to function: possible biological roles of a new widespread protein family binding hydrophobic ligands and displaying a nucleotide binding site. *FEBS Letters* **369**, 22-26
16. Pnueli, L., Carmel-Goren, L., Hareven, D., Gutfinger, T., Alvarez, J., Ganai, M., Zamir, D. and Lifschitz, E. (1998) The SELF-PRUNING gene of tomato regulates vegetative to reproductive switching of sympodial meristems and is ortholog of CEN and TFL1. *Development* **125**, 1979-1989
17. Bollengier, F. and Mahler, A. (1988) Localization of the novel neuropeptide h3 in subsets of tissues from different species. *J. Neurochem.* **50**, 1210-1214
18. Taiji, M., Tohdoh, N. and Ojika, K. (1996) Neuronal expression of hypocampal cholinergic neurostimulating peptide (HCNP)-precursor mRNA in rat brain. *J. Neurosci. Res.* **45**, 202-215
19. Seddiqi, N., Segretain, D., Bucquoy, S., Pineau, C., Jégou, B., Jollès, P. and Schoentgen, F. (1996) Characterization and localization of the testicular phosphatidylethanolamine binding protein in rat, mouse and human testis. *Experientia* **52**, 101-110
20. Gibbons, R., Adeoya-Osiguwa, S.A. and Fraser, L.R. (2005) A mouse sperm decapacitation factor receptor is phosphatidylethanolamine-binding protein 1. *Reproduction* **130**, 497-508.
21. Hickox, D.M., Gibbs, G., Morrison, J.R., Sebire, K., Edgar, K., Keah, H.H., Alter, K., Loveland, K.L., Hearn, M.T., de Kretser, D.M. and O'Bryan, M.K. (2002) Identification of a novel testis-specific member of the phosphatidylethanolamine binding protein family, pebp-2. *Biol. Reprod.* **67**, 917-927
22. Yeung, K., Seitz, T., Li, S., Janosch, P., McFerran, B., Kaiser, C., Fee, F., Katsanakis, K.D., Rose, D.W., Mischak, H., Sedivy, J.M. and Kolch, W. (1999) Suppression of Raf-1 kinase activity and MAP kinase signalling by RKIP. *Nature* **401**, 173-177.
23. Yeung, K.C., Rose, D.W., Dhillon, A.S., Yaros, D., Gustafsson, M., Chatterjee, D., McFerran, B., Wyche, J., Kolch, W. and Sedivy, J.M. (2001) Raf kinase inhibitor protein interacts with NF- κ B inducing kinase and TAK1 and inhibits NF- κ B activation. *Mol. Cell. Biol.* **21**, 7207-7217.
24. Keller, E.T. (2004) Metastasis suppressor genes: a role for Raf Kinase Inhibitor Protein (RKIP). *Anti-Cancer Drugs* **15**, 663-669

25. Chatterjee, D., Bai, Y., Wang, Z., Beach, S., Mtt, S., Roy, R., Braastad, C., Sun, Y., Mukhopadhyay, A., Aggarwal, B.B., Darnowski, J., Pantazis, P., wyche, J., fu, Z., Kitagawa, Y., Keller, E.T., Sedivy, J.M. and Yeung, K.C. (2004) RKIP sensitizes prostate and breast cancer cells to drug-induced apoptosis. *J. Biol. Chem.* **279**, 17515-17523
26. Jazirehi, A.R. and Bonavida B. (2005) Cellular and molecular signal transduction pathways modulated by rituximab (rituxan, anti-CD20 mAb) in non-Hodgkin's lymphoma: implications in chemosensitization and therapeutic intervention. *Oncogene* **24**, 2121-2143
27. Fu, Z., Kitagawa, Y., Shen, R., Shah, R., Mehra, R., Rhodes, D., Keller, P.J., Mizokami, A., Dunn, R., Chinnaiyan, A.M., Yao, Z. and Keller, E.T. (2006) Metastasis suppressor gene Raf Kinase Inhibitor Protein (RKIP) is a novel prognostic marker in prostate cancer. *Prostate* **66**, 248-256.
28. Vierstraete, E., Verleyen, P., Baggerman, G., D'Hertog, W., Van den Bergh, G., Arckens, L., De Loof, A and Schoofs, L.A. (2004) proteomic approach for the analysis of instantly released wound and immune proteins in *Drosophila melanogaster* hemolymph. *Proc. Natl. Acad. Sci. USA* **101**, 470-475
29. Levy, F., Bullet, P., and Ehret-Sabatier, L. (2004) Proteomic analysis of the systemic immune response of *drosophila*. *Mol. Cell. Proteomics* **3**, 156-166.
30. Sali, A. and Overington, J.P. (1994). Derivation of rules for comparative protein modelin from a database of protein structure alignment. *Protein Sci.* **3**,1582-1596.
31. Laskowski, R.A., Rullmann, J.A., MacArthur, M.W., Kaptein R. and Thornton, J.M. (1996) AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR. *J. Biomol. NMR* **8**, 477-486.
32. Melo, F., Devos, D., Depiereux, E. and Feytmans, E. (1997) ANOLEA: a www server to assess protein structures. *Proc. Int. Conf. Intell. Syst. Mol. Biol.* **5**, 187-190.
33. Luthy, R., Bowie, J.U. and Eisenberg, D. (1992) Assessment of protein models with three-dimensional profiles. *Nature* **356**, 83-85.
34. Koradi, R., Billeter, M. and Wuthrich, K. (1996) MOLMOL: a program for display and analysis of macromolecular structures. *J. Mol. Graph.* **14** 51-55, 29-32.
35. Rautureau, G., Jouvensal, L., Decoville, M., Locker, D., Vovelle, F. and Schoentgen, F. (2006) Cloning, high yield over-expression, purification and characterization of CG18594, a new PEBP/RKIP family member from *Drosophila melanogaster*. *Protein Expr. Purif.* **48**, 90-97
36. Maniatis, T., Fritsch, E.F. and Sambrook, J., *Molecular cloning: a laboratory manual*. 2nd edition ed. 1989, Cold Spring Harbor N. Y.: Cold Spring Harbor Laboratory Press.
37. Delaglio, F., Grzesiek, S., Vuister G.W., Zhu G., Pfeifer J. and Bax A. (1995) NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J. Biomol. NMR* **6**, 277-293.
38. Johnson, B.A. (2004) Using NMRView to visualize and analyze the NMR spectra of macromolecules. *Methods Mol. Biol.* **278**, 313-52.
39. Gaboriaud, C., Bissery, V., Benchetrit, T. and Mornon, J.P. (1987) Hydrophobic cluster analysis: an efficient new way to compare and analyse amino acid sequences. *FEBS Letters* **224**, 149-155.
40. Sreerama, N., Venyaminov, S.Y. and Woody, R.W. (2000) Estimation of protein secondary structure from circular dichroism spectra: inclusion of denatured proteins with native proteins in the analysis. *Anal. Biochem.* **287**, 243-251.
41. Banfield, M.J., Barker, J.J., Perry, A.C. and Brady, R.L. (1998) Function from structure? The crystal structure of human phosphatidylethanolamine-binding protein suggests a role in membrane signal transduction. *Structure* **6**, 1245-1254.
42. Jones, D.T., Tress, M., Bryson, K. and Hadley C. (1999) Successful recognition of protein folds using threading methods biased by sequence similarity and predicted secondary structure. *Proteins*, Suppl 3, 104-111.
43. McGuffin, L.J., Bryson, K. and Jones, D.T. (2000) The PSIPRED protein structure prediction server. *Bioinformatics* **16**, 404-405.
44. Cuff, J.A., Clamp, M.E., Siddiqui, A.S., Finlay, M. and Barton, G.J. (1998) JPred: a consensus secondary structure prediction server. *Bioinformatics* **14**, 892-893.
45. Geourjon, C. and Deleage, G. (1995) SOPMA: significant improvements in protein secondary structure prediction by consensus prediction from multiple alignments. *Comput. Appl. Biosci.* **11**, 681-684
46. Lin, T.Y., Viswanathan, S., Wood, C., Wilson, P.G., Wolf, N. and Fuller, M.T. (1996) Coordinate developmental control of the meiotic cell cycle and spermatid differentiation in *Drosophila* males. *Development* **122**, 1331-1341.
47. Reijo, R., Lee, T.-Y., Salo, P., Alagappan, R., Brown, L.G., Rosenberg, M., Rozen, S., Jaffe, T., Straus, D., Hovatta, O., de la Chapelle, A., Silber, S. and Page, D.C. (1995) Diverse spermatogenic defects in humans caused by Y-chromosome deletions encompassing a novel RNA binding protein gene. *Nat. Genet.* **10**, 383-393.
48. Eberhart, C.G., Maines, J.Z. and Wasserman, S.A. (1996) Meiotic cell cycle requirement for a fly homologue of human Deleted in Azoospermia. *Nature* **381**, 783-785.
49. Courtot, C., Frankhauser, C., Simanis, V. and Lehner, C. (1992) The *drosophila* cdc25 homolog

- twine is required for meiosis. *Development* **116**, 405-416.
50. Serre, L., Vallee, B., Bureau, N., Schoentgen, F. and Zelwer, C. (1998) Crystal structure of the phosphatidylethanolamine-binding protein from bovine brain: a novel structural class of phospholipid-binding proteins. *Structure* **6**, 1255-1265.
51. Simister, P.C., Banfield, M.J. and Brady, R.L. (2002) The crystal structure of PEBP-2, a homologue of the PEBP/RKIP family. *Acta Crystallogr. D Biol. Crystallogr.* **58**, 1077-1080.
52. Mima, J., Hayashida, M., Fujii, T., Narita, Y., Hayashi, R., Ueda, M. and Hata Y. (2005) Structure of the carboxypeptidase Y inhibitor IC in complex with the cognate proteinase reveals a novel mode of the proteinase-protein inhibitor interaction. *J. Mol. Biol.* **346**, 1323-1334.
53. Saunders, P.T.K., McKinnell, C., Millar, M.R., Gaughan, J., Turner, K.J., Jegou, B., Syed, V. and Sharpe, R.M. (1995) Phosphatidylethanolamine binding protein is an abundant secretory product of haploid testicular germ cells in the rat. *Mol. Cell. Endocrinol.* **107**, 221-230.
54. Brooks D.E. (1985) Characterization of a 22 kDa protein with widespread tissue distribution but which is uniquely present in secretions of the testis and epididymis and on the surface of spermatozoa. *Biochim. Biophys. Acta* **841**, 59-70.
55. Dacheux, J.L., Belghazi, M., Lanson, Y. and Dacheux, F. (2006) Human epididymal secretome and proteome. *Mol. Cell. Endocrinol.* **250**, 36-42.
56. Bergeret, E., Pignot-Paintrand, I., Guichard, A., Raymond, K., Fauvarque, M.O., Cazemajor, M. and Griffin-Shea, R. (2001) RotundRacGAP functions with Ras during spermatogenesis and retinal differentiation in *Drosophila melanogaster*. *Mol. Cell. Biol.* **21**, 6280-6291.
57. Hengst, U., Albrecht, H., Hess, D. and Monard, D. (2001) The phosphatidylethanolamine-binding protein is the prototype of a novel family of serine protease inhibitors. *J. Biol. Chem.* **276**, 535-540.

3. Stabilité de la protéine au cours du temps

La méthode des gouttes suspendues a été utilisée pour tester dans un volume minimal le comportement de la protéine CG10298 dans des conditions proches de celles de la RMN. Une observation à la loupe binoculaire permet de visualiser facilement la solubilité totale ou une précipitation partielle ou totale de la protéine dans la goutte. Les gouttes de 4 μL sont réalisées en mélangeant 3 μL de protéine CG10298 à 1,3 mM dans l'eau avec 1 μL de tampon quatre fois concentré. La goutte déposée sur une lamelle est retournée au dessus d'un réservoir contenant le tampon seul, à la concentration identique à celle de la goutte.

Notre expérience avec la protéine CG18594 nous a amené à tester en priorité les conditions suivantes:

- 20°C, acétate de potassium 100 mM pH 4,5 ; 5 ; 5,5
- 20°C, phosphate de potassium 100 mM pH 6 ; 6,5 ; 7

La protéine s'est avérée particulièrement soluble dans toutes les conditions testées. L'étude de la stabilité au cours du temps a donc été réalisée sur la protéine CG10298 marquée ^{15}N dans du tampon phosphate de potassium 50 mM pH 5,5, β -mercaptoéthanol 4 mM, DSS 1mM, azide 0,01%, D_2O 10%. La concentration en protéine était de 2 mM. En dehors des périodes d'acquisition des spectres, la protéine a été maintenue à 20°C. Nous avons tracé des spectres ^1H - ^{15}N HSQC de la protéine CG10298 à intervalles réguliers.

Après une dizaine de jours, certains pics ont commencé à changer de position. Le phénomène s'est amplifié par la suite (Figure V-7). Cette situation n'est pas idéale pour une étude structurale complète de la protéine CG10298 car l'enregistrement de l'ensemble des spectres nécessaires à l'attribution et à la détermination de la structure peut durer plusieurs semaines. Cependant la bonne dispersion des résonances dans les deux dimensions indique que la protéine reste repliée. Enfin, les largeurs de raie des différentes résonances sont relativement homogènes et fines, contrairement à celles que nous avons observées pour la protéine CG18594. La largeur de raie tend à montrer que la protéine

CG10298 est sous forme monomérique. Il est donc probable qu'en ajustant les conditions de l'échantillon on parvienne à stabiliser la protéine.

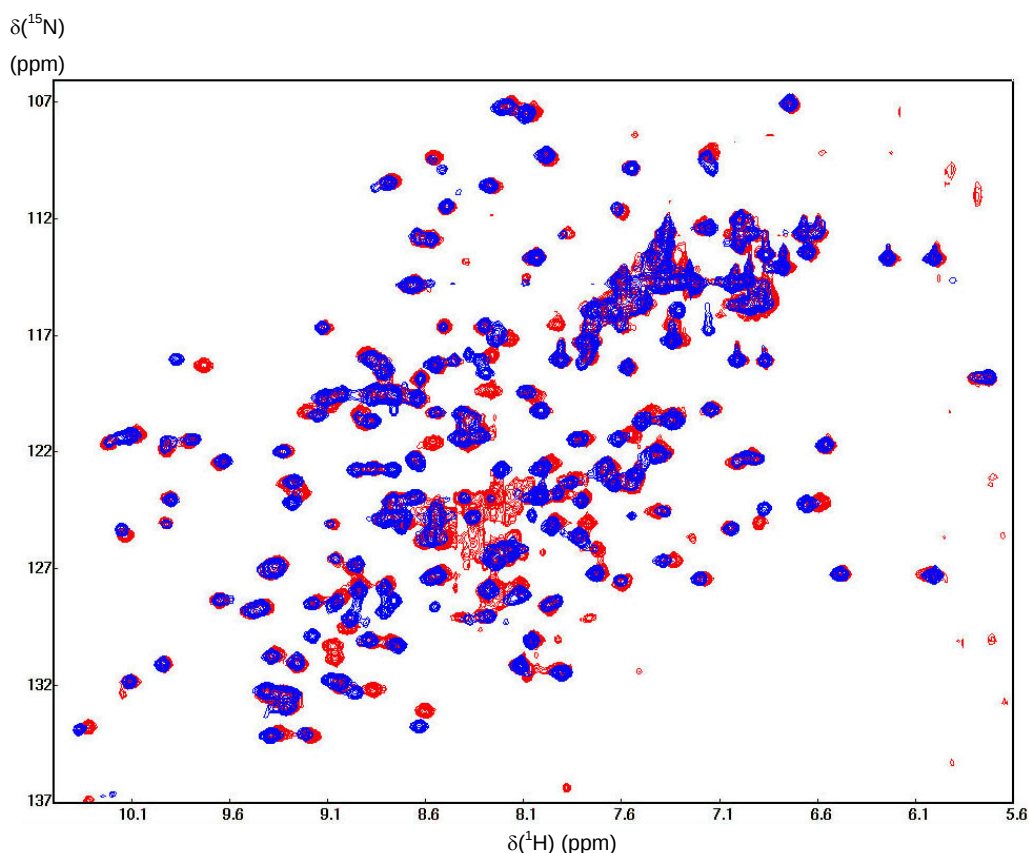


Figure V-7 : Superposition des spectres ^1H - ^{15}N HSQC de la protéine CG10298 juste après la mise en solution (en bleu) et après 7 semaines (en rouge).

4. Conclusion

Nous avons réussi dans ce travail à exprimer la protéine CG10298 (marquée ^{15}N ou non) et à la purifier en quantité importante. La protéine produite est soluble et repliée. Cependant les spectres RMN évoluent lentement au cours du temps à 20°C. Dans la mesure où nous avons obtenu la protéine CG7054 à la même période et où cette dernière s'est avérée parfaitement stable, nous l'avons choisie pour réaliser l'étude structurale. Néanmoins les spectres obtenus pour la protéine CG10298 sont prometteurs et la durée de stabilité de cette protéine (environ 10 jours sans aucune évolution) n'est pas incompatible avec une étude structurale. En effet, il est possible de réaliser rapidement l'acquisition de l'ensemble des spectres nécessaires à l'attribution et à la résolution de la structure sur un

échantillon concentré (la protéine est soluble jusqu'à 2 mM) avec un spectromètre 800MHz équipé d'une cryosonde. Le rendement de production de la protéine étant très bon, on peut aussi envisager facilement l'utilisation de plusieurs échantillons. Enfin, il est probable que l'optimisation des conditions de l'échantillon (pH, sel, ...) puisse résoudre ce problème de stabilité et permette une étude sereine par RMN.

Ce travail a également permis d'avancer dans l'élucidation du rôle de cette PEBP. A partir de la protéine purifiée, il a été possible de produire des anticorps qui ont été utilisés par l'équip de Daniel Locker pour établir de mode d'expression de cette protéine. Nous avons en outre montré que cette PEBP de drosophile est la plus proche de la protéine bovine. Ceci nous a permis de construire un modèle par homologie. Même si ce modèle n'est pas aussi précis qu'une structure expérimentale, il permettra une étude des propriétés de surface (potentiels électrostatiques et de lipophilie) qui pourront donner des pistes quant aux modes d'interaction de la protéine avec des partenaires, une fois que ceux-ci seront connus, en attendant de disposer d'une structure 3D RMN.

Etude de la protéine

CG7054

VI. Etude structurale de la protéine CG7054

Le clonage de la région codante du gène *cg7054* a été réalisé en même temps que celui du gène *cg10298*. Nous avons choisi le gène *cg7054* car il est situé juste à proximité du gène *cg18594*. Ce groupe de 2 gènes est étudié par l'équipe de Daniel Locker qui dispose de différentes lignées mutantes pour ces gènes.

Après avoir cloné dans *Escherichia coli* la région codante du gène *cg7054* dans le vecteur d'expression pET31b, j'ai mis au point le protocole de purification et de marquage de la protéine CG7054. Actuellement, nous pouvons la produire purifiée, marquée ou non, en grande quantité. Une étude initiale de leur de stabilité en solution a été réalisée par RMN: la protéine CG7054 est parfaitement stable à 20°C. Cette dernière a donc été choisie pour l'étude structurale présentée dans cette partie. Après acquisition des spectres hétéronucléaires, nous avons réalisé l'attribution des signaux ^1H , ^{13}C et ^{15}N du squelette peptidique et des chaînes latérales. L'étude de la dynamique du squelette peptidique a été initiée et la structure a été déterminée. Par ailleurs, nous avons réalisé des études d'interaction de la protéine CG7054 avec un ligand anionique : la phosphoryléthanolamine.

1. Données de base

La séquence du gène *cg7054* est disponible sur la banque de données Flybase (<http://flybase.bio.indiana.edu/>). Ce gène est situé sur le bras droit du chromosome 3, en position 94A16, à proximité du gène *cg18594* (Figure VI-1).

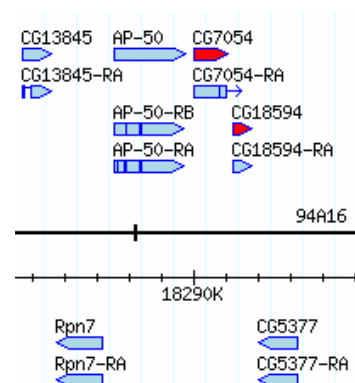


Figure VI-1 : Localisation des gènes *cg7054* et *cg18594* sur le bras droit du chromosome III (image issue de Flybase).

Le gène *cg7054* a une longueur de 1026 nucléotides sur l'ADN génomique, il est composé de deux exons séparés par un intron. La traduction de la portion codante fournit une séquence protéique qui présente la signature caractéristique de la famille des PEBP (Figure VI-2).

```

      1      11      21      31      41      51
      |      |      |      |      |      |
1  MDDIVPDVLD AVPAGTIKVI YGDDLEVQKG NELTPTQVKD QPIVSWSGLE GKSNNLTLLM
61 VDPDAPTRQD PKYREILHWS VVNIPGSNEN PSGGHSADY VSGGPPKDTG LHRIFYLLYR
121 QENKIEETPT ISNTTRTGRL NFNARDFAAK HGLGEPIAAN YYQAQYDDYV PIRNKTIIV

```

Figure VI-2 : Séquence de la protéine CG7054. La séquence consensus de la famille des PEBP est indiquée en rouge.

L'analyse de cette séquence a été réalisée avec l'outil Protparam du site ExPASy (Gasteiger *et al.*, 2003), Figure VI-3.

Nombre d'acides aminés: 179

Masse moléculaire: 19822,1 Da

pI théorique: 5,02

Composition en acides aminés:

Acides aminés	Code	Nombre	Fréquence	Acides aminés	Code	Nombre	Fréquence
Ala	(A)	10	5,6%	Leu	(L)	15	8,4%
Arg	(R)	8	4,5%	Lys	(K)	9	5,0%
Asn	(N)	11	6,1%	Met	(M)	2	1,1%
Asp	(D)	15	8,4%	Phe	(F)	3	1,7%
Cys	(C)	0	0,0%	Pro	(P)	14	7,8%
Gln	(Q)	7	3,9%	Ser	(S)	9	5,0%
Glu	(E)	9	5,0%	Thr	(T)	12	6,7%
Gly	(G)	15	8,4%	Trp	(W)	2	1,1%
His	(H)	4	2,2%	Tyr	(Y)	9	5,0%
Ile	(I)	12	6,7%	Val	(V)	13	7,3%

Figure VI-3 : Analyse de la structure primaire de la protéine CG7054 avec l'outil Protparam (<http://us.expasy.org/tools/protparam.html>)

A partir de cette analyse des acides aminés de la protéine CG7054, nous pouvons tracer la charge nette théorique moyenne de la protéine en fonction du pH (Expasy, titration curve) (

Figure VI-4). Cette courbe est particulièrement utile lors du choix des méthodes de purification : la protéine est chargée positivement en dessous de son point isoélectrique. Elle demeure faiblement chargée négativement entre pH 5 et 9.

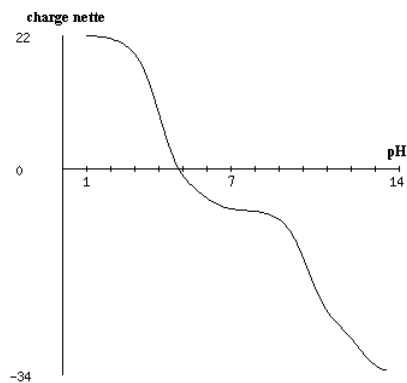


Figure VI-4: Charge nette théorique moyenne de la protéine CG7054 en fonction du pH.

1.1. Comparaison de la protéine CG7054 avec les autres PEBP de drosophile

La séquence de la protéine CG7054 a été alignée avec celles des 6 autres PEBP de drosophile (Figure VI-5).

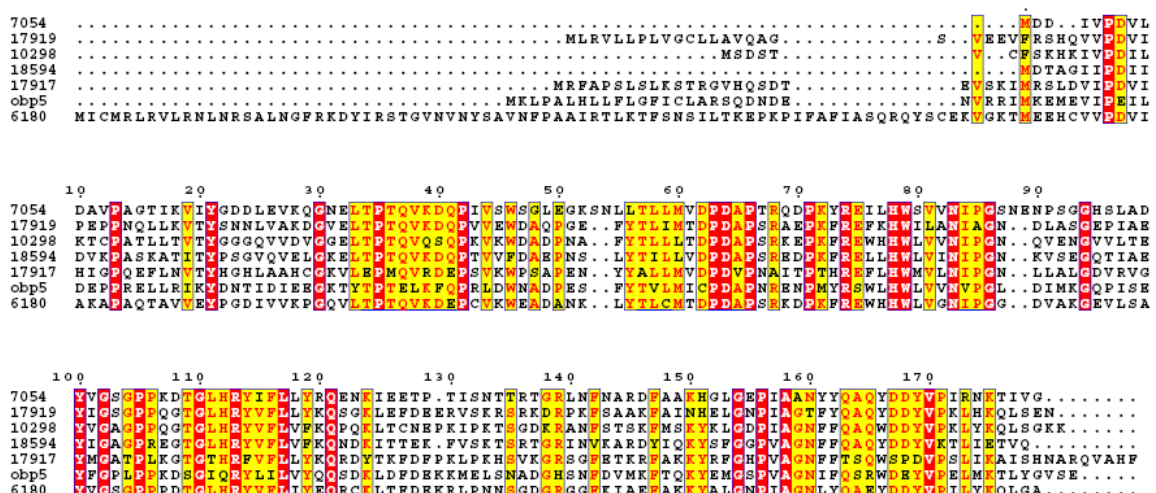


Figure VI-5: Alignement de la protéine CG7054 avec les 6 autres PEBP de drosophiles. L'alignement a été réalisé avec le logiciel CLUSTALW, (Eddy, 1995). Les cadres rouges indiquent les résidus identiques, les résidus non conservés, mais similaires, sont indiqués en jaune. (Figure réalisée avec le logiciel Esript).

À partir de cet alignement nous pouvons calculer le pourcentage d'identité entre la protéine CG7054 et les six autres PEBP de drosophile. Les valeurs obtenues sont indiquées dans le tableau suivant:

	% d'identité avec la protéine CG7054
CG6180	50
CG17919	49
CG10298	48
CG18594	48
CG17917	37
OBP5	35

Tableau VI-1 : Pourcentage d'identité de la protéine CG7054 avec les autres PEBP de drosophile. L'alignement a été réalisé avec le logiciel CLUSTALW.

La protéine CG7054 présente un pourcentage d'identité proche de 50% avec les protéines CG6180, CG17919, CG10298 et CG18594. L'alignement est par ailleurs un peu moins bon avec les protéines CG17917 et l'OBP5.

1.2. Comparaison de la protéine CG7054 avec les PEBP des autres espèces

La séquence de la protéine CG7054 a été alignée avec celle des autres PEBP de structure tridimensionnelle connue (Figure VI-6).

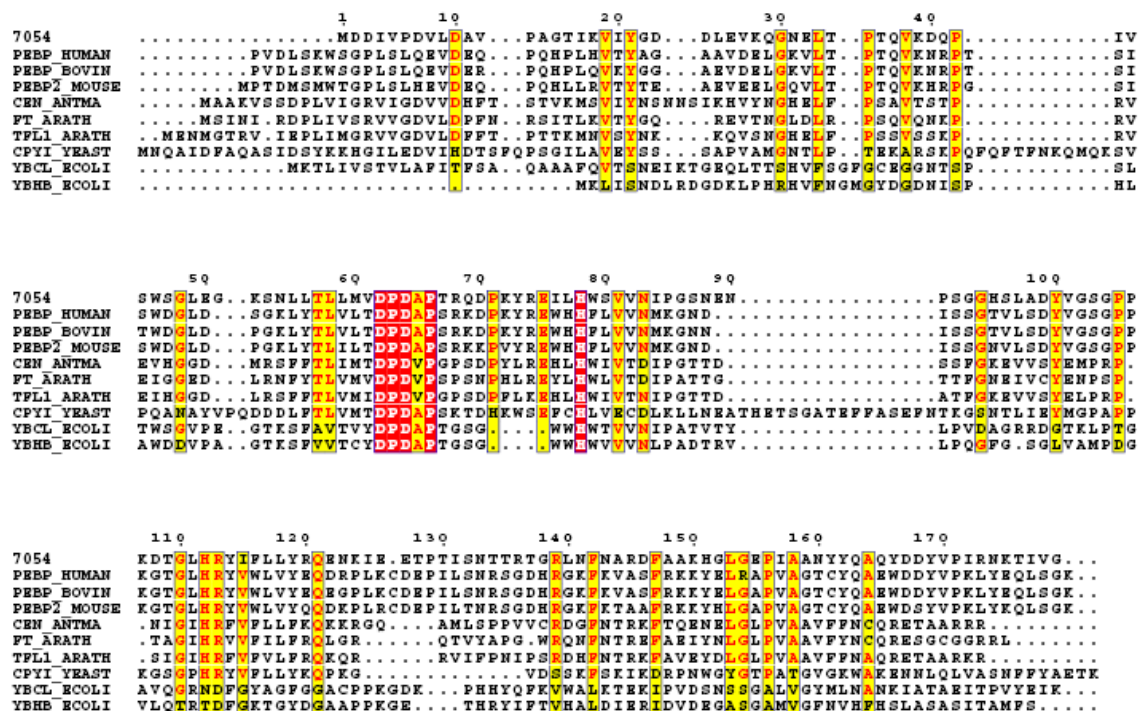


Figure VI-6 : Alignement de la protéine CG7054 avec différentes PEBP dont la structure 3D est connue. L'alignement a été réalisé avec le logiciel CLUSTALW, (Eddy, 1995). Les cadres rouges indiquent les résidus identiques, les résidus non conservés, mais similaires, sont indiqués en jaune. (Figure réalisée avec le logiciel Esprpt).

Cet alignement montre bien que les zones conservées entre les différentes espèces sont moins nombreuses qu'entre les 7 PEBP de drosophile. Les protéines YBHB et YBCL de la bactérie *Escherichia coli* sont les plus éloignées.

Le tableau ci-dessous indique les pourcentages d'identité observés entre la protéine CG7054 et ces différentes PEBP (Tableau VI-2) :

	% d'identité avec la protéine CG7054
PEBP-1 humaine	43
PEBP-1 bovine	43
PEBP-2 souris	41
Antirrhinum Cen	29
Arabidopsis FT1	32
Arabidopsis TFL1	33
Saccharomyces TFS1	22
E. coli YBCL	18
E. coli YBHB	15

Tableau VI-2 : Pourcentage d'identité de la protéine CG7054 avec les PEBP dont la structure 3D est connue. L'alignement a été réalisé avec le logiciel CLUSTALW.

A partir de ces alignements, nous pouvons conclure sur la base de sa séquence que la protéine CG7054 appartient bien à la famille des PEBP. Elle possède la séquence consensus caractéristique des PEBP mais aussi une conservation significative des résidus avec les autres membres de la famille. Les PEBP les plus proches sont les PEBP humaines et bovines (43 % d'identité). L'alignement avec les protéines de plantes, de levures (TFS1) ou de bactéries (YBHB et YBCL) présente un pourcentage d'identité beaucoup plus faible. Les PEBP de drosophile sont donc plus proches entre elles qu'avec celles des autres espèces.

Mon travail sur la protéine CG7054 a d'abord consisté à mettre au point les techniques d'expression et de purification permettant de produire la protéine, marquée ou non, dans le but d'une étude structurale. J'ai ensuite réalisé une caractérisation physico-chimique de cette protéine afin de déterminer quelles étaient les meilleures conditions pour l'étude RMN. L'étude structurale a pu être menée jusqu'à la détermination de la structure.

2. Matériels et méthodes

2.1. Clonage de l'ORF du gène *cg7054*

La stratégie de clonage de la région codante du gène *cg7054* dans le plasmide d'expression pET31b(+) est représentée sur la Figure VI-7. Les détails expérimentaux sont détaillés dans l'annexe 1.

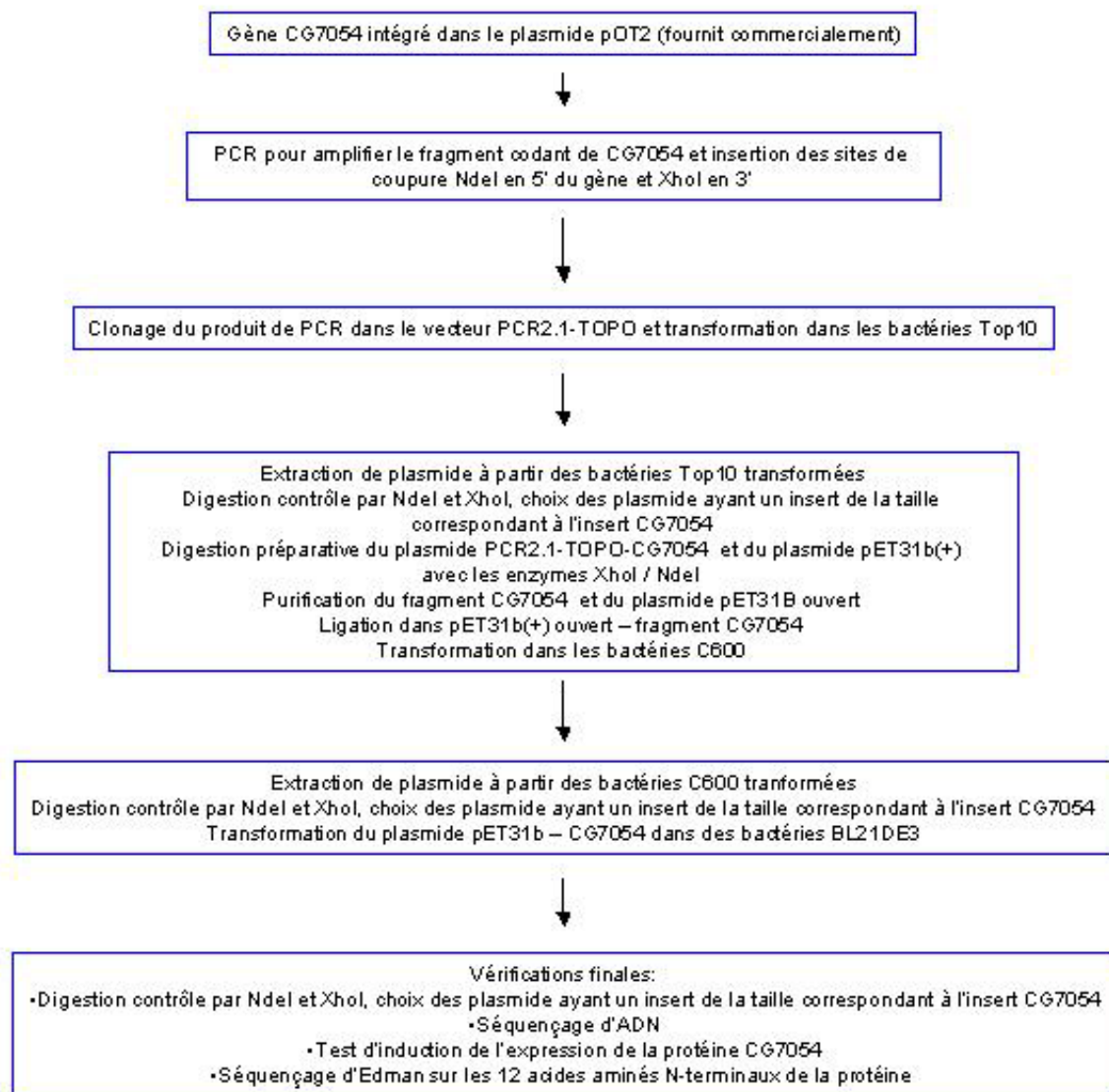


Figure VI-7 : Stratégie de clonage suivie pour intégrer l'ADNc du gène *cg7054* dans le plasmide d'expression pET31b(+).

2.2. Production et purification de la protéine CG7054

Nous disposons donc maintenant de clones d'*Escherichia coli* BL21DE3 exprimant la protéine CG7054 grâce à un plasmide de type pET31b(+) dans lequel l'ADNc du gène *cg7054* a été intégré au niveau des sites NdeI en 5' et XhoI en 3'. La sélection des bactéries possédant le plasmide se fait par l'ampicilline, un gène de résistance étant présent sur le plasmide pET31b. L'induction de l'expression est réalisée par ajout d'IPTG.

2.2.1. Culture bactérienne et surexpression de la protéine, cas de la protéine non marquée

A partir d'une préculture du clone d'*E. coli* BL21DE3 exprimant la protéine CG7054, réalisée pendant une nuit à 30°C sous agitation mécanique en milieu LB, ampicilline 100 µg.mL⁻¹, on ensemence 1,2 litre de milieu LB et ampicilline 100 µg.mL⁻¹ par dilution au centième. On incube à 30°C sous agitation mécanique jusqu'à obtenir une absorbance de 0,7 à 600 nm. La surexpression de la protéine est induite par l'addition d'IPTG (0,1 mM final) à 30°C, sous agitation mécanique pendant 5 heures. Les cultures sont centrifugées 5 minutes à 5000 g et les bactéries sont reprises dans 10 mL de tampon QAE par gramme de cellules obtenu (Tris HCl 50 mM pH 8, NaCl 0,05 M, EDTA 1 mM, NaN₃ 0,05%,) et congelées immédiatement.

2.2.2. Culture bactérienne et surexpression de la protéine, cas de la protéine marquée ¹⁵N et doublement marquée ¹³C /¹⁵N

Le milieu utilisé est un milieu M9 modifié (quantités pour 1 litre de milieu de culture):

Milieu M9 salin de base, pH ajusté à 7,2 et autoclavé (Tableau I-1):

Na₂HPO₄ anhydre (ou 12,8g Na₂HPO₄.7H₂O)	6,8 g
KH₂PO₄	3 g
NaCl	0,5 g
¹⁵NH₄Cl	1 g
H₂O	QSP 1000 mL

Tableau VI-3 : Composition de la base saline du milieu M9.

Préparation du milieu M9 final (Tableau VI-4):

M9 salin de base autoclavé	1000 mL
20% (w/v) Glucose stérile, marqué ¹³C ou non	20 ml
1 M MgSO₄ stérile	1 ml
1 M CaCl₂ stérile	0,3 ml
Thiamine (1 mg.mL⁻¹)	1 ml

Tableau VI-4 : Composition du milieu de culture M9 complet.

A partir d'une préculture de 20 mL du clone d'*E. coli* BL21DE3 exprimant la protéine CG7054, réalisée pendant une nuit à 30°C sous agitation mécanique en milieu LB, ampicilline 100 µg.mL⁻¹, on ensemence 2 litres de milieu LB et ampicilline 100 µg.mL⁻¹ par dilution au centième. On incube à 30°C sous agitation mécanique jusqu'à obtenir une absorbance de 0,7 à 600 nm. Les cellules sont ensuite centrifugées stérilement 4 minutes à 5000 g, reprises délicatement dans 500 mL de milieu M9 ne contenant aucune source d'azote ni de carbone, ampicilline 50 µg.mL⁻¹, centrifugées 4 minutes à 4000 rpm et reprises dans 500 mL de milieu M9 marqué, ampicilline 50 µg.mL⁻¹. Après une heure d'incubation à 30°C servant à la chasse des pools d'azote et de carbone non marqués, l'induction est faite avec 1 mM final d'IPTG pendant 5 à 6 heures. Les cultures sont centrifugées 10 minutes à 4000 rpm, reprises dans 10 mL de tampon QAE par gramme de cellules obtenu (Tris HCl 50 mM pH8, NaCl 0,05M, EDTA 1mM, NaN₃ 0,05%) et congelées immédiatement.

2.2.3. Extraction et purification de la protéine

Les cellules décongelées sont soniquées 3 fois pendant une minute puis centrifugées à 15000 g pendant 45 minutes à 4°C.

2.2.4. Chromatographie d'échange d'ions

La protéine CG7054 est chargée négativement à pH 8. En présence de 50 mM de NaCl, elle est retenue sur une résine greffée d'amines quaternaires, forte échangeuse d'anions (Himmelhoch, 1971). La résine utilisée (QAE Sephadex A-50) provient de chez Pharmacia Fine Chemicals (Uppsala, Suède). Nous la faisons d'abord gonfler dans un grand volume de tampon QAE (Tris HCl 50 mM pH 8, NaCl 0,05 M, EDTA 1 mM, NaN₃ 0,05%). L'équilibrage du gel se fait directement dans la colonne une fois montée. Nous utilisons une colonne de 30 cm * 2,4 cm reliée à une pompe Gilson « Minipuls II » et un collecteur de fractions. La charge correspond à un volume maximum de 20 mL. Le débit est de 35 mL/heure. La protéine CG7054 est retenue sur la colonne et lavée avec environ 200 mL de tampon. Ensuite, la protéine est éluée avec du tampon QAE auquel on ajoute 0,2M de NaCl (soit une concentration finale dans le tampon d'élution de 0,25M NaCl). La protéine est éluée rapidement.

2.2.5. Dialyses et lyophilisation

Les fractions (4 mL) d'intérêt sont rassemblées puis concentrées en cellule Amicon sur une membrane d'ultrafiltration YM10 (limite à 10 kDa), sous une pression de 2 bars. La protéine est ensuite lavée 6 fois avec de l'eau ultrapure : le pool concentré à 10 mL est dilué à 50 mL avec de l'eau ultra-pure et reconcentré à 10 mL. Nous éliminons de cette manière les différents constituants du tampon QAE (Tris, NaCl ...). Il faut remarquer durant cette étape la précipitation progressive de matériel protéique contaminant lorsque la force ionique diminue. La protéine CG7054 reste quant à elle bien soluble.

Après vérification sur un spectre RMN 1D ¹H que la concentration des protons du tampon Tris issu du tampon QAE est négligeable par rapport à la concentration du pool de protéine CG7054, la solution est aliquotée, congelée et lyophilisée. Un échantillon est par ailleurs systématiquement analysé par spectrométrie de masse pour estimer le taux de marquage.

2.2.6. Conservation

La protéine est conservée lyophilisée au congélateur à -20°C dans des tubes hermétiques.

2.3. Recherche des conditions RMN

2.3.1. Etude des conditions en gouttes suspendues

La méthode des gouttes suspendues a été mise à profit afin de tester dans un volume minimal le comportement de la protéine CG7054 dans des conditions proches de celles de la RMN. Une observation à la loupe binoculaire permet de visualiser facilement la solubilité totale ou une précipitation partielle ou totale de la protéine dans la goutte. Les gouttes de 4 μL sont réalisées en mélangeant 3 μL de protéine CG7054 à 1,3 mM dans l'eau avec 1 μL de tampon quatre fois concentré. La goutte déposée sur une lamelle est retournée au dessus d'un réservoir contenant le tampon seul, à la concentration identique à celle de la goutte.

Notre expérience avec la protéine CG18594 nous a amené à tester en priorité les conditions suivantes:

- 20°C, acétate de potassium 100 mM pH 4,5 ; 5 ; 5,5
- 20°C, phosphate de potassium 100 mM pH 6 ; 6,5 ; 7

2.3.2. Dichroïsme circulaire

La technique de dichroïsme circulaire a été utilisée pour observer la stabilité thermique de la protéine CG7054 à différents pH. Les spectres ont été enregistrés entre 185 et 260 nm à l'aide d'un dichrographe Jasco 810 en utilisant une cuve de trajet optique 1 mm, avec une vitesse de défilement de 50 nm.min⁻¹ et dans une gamme de température allant de 5°C à 75°C. Un spectre a été tracé tous les 5°C. Une mesure de l'ellipticité molaire $[\theta]$ à 215 nm a été réalisée tous les 0,2°C. La ligne de base obtenue pour le tampon a été soustraite de tous les spectres enregistrés.

Les mesures ont été effectuées dans les conditions suivantes : tampon phosphate de potassium 10 mM pH 2, 4, 5, 6, 7, 8, 10 et 12. Le pH des solutions étudiées a été vérifié avant et après mesure. La concentration en protéine était de 0,2 mg.mL⁻¹ et une goutte d'huile minérale a été déposée à la surface des solutions dégazées afin de limiter les phénomènes d'évaporation et de formation de bulles à haute température.

2.4. Résonance magnétique nucléaire

L'étude a été réalisée sur la protéine CG7054 (non marquée, simplement marquée ^{15}N ou doublement marquée $^{15}\text{N}/^{13}\text{C}$) dans du tampon phosphate de potassium 50 mM pH 6,5, DSS 1 mM, azide 0,01%, D_2O 10%. La concentration en protéine était de 1,5 mM dans les 3 échantillons. En dehors des périodes d'acquisition des spectres, la protéine est conservée à 4°C.

2.4.1. Stabilité de la protéine CG7054 au cours du temps

Nous avons commencé par observer la stabilité de la protéine au cours du temps. En effet une étude RMN nécessite une protéine parfaitement stable pendant au moins une semaine. Nous avons donc suivi la protéine à 20° en fonction du temps. Pour cela, nous avons tracé des spectres ^1H - ^{15}N HSQC de la protéine CG7054 à intervalles réguliers.

2.4.2. Acquisition des spectres

Les expériences ont été enregistrées à 20°C. La majorité des spectres nécessaires à l'attribution des résonances puis à la détermination de la structure de la protéine CG7054 ont été acquis sur le spectromètre 600 MHz VARIAN UNITY INOVA du laboratoire. Afin de gagner en résolution et en sensibilité, plusieurs spectres importants ont été réalisés sur le spectromètre 800 MHz équipé d'une cryosonde de la plateforme européenne de l'Institut de Biologie Structurale à Grenoble. Les expériences réalisées ainsi que les paramètres d'acquisition sont détaillés dans l'annexe 2. Les spectres ont été traités avec NMRPipe (Delaglio *et al.*, 1995). Le logiciel NMRView (Johnson, 2004) m'a permis d'analyser les spectres.

2.4.3. Expériences d'échange proton / deutérium

Afin d'estimer la vitesse d'échange des protons amide, nous avons enregistré une série de spectres ^1H - ^{15}N HSQC sur la protéine CG7054 fraîchement mise en solution dans le tampon habituel préparé avec du D_2O .

2.4.4. Attribution du squelette peptidique

L'attribution du squelette peptidique a été réalisée en utilisant conjointement plusieurs logiciels.

2.4.4.a. Le logiciel *Smartnotebook*

Smartnotebook (Slupsky *et al.*, 2003) est un outil d'attribution semi-automatique qui fonctionne comme un module intégré à NMRView (Johnson, 2004). Smartnotebook réalise une recherche exhaustive des connectivités possibles et cohérentes entre les différents spectres. L'utilisateur observe les connectivités et décide de leur attribution éventuelle. Smartnotebook compare aussi les déplacements chimiques observés à une banque statistique pour aider à déterminer quel est le type de résidu impliqué. Les alanines, qui présentent des résonances dans la dimension carbone très caractéristiques, sont par exemple aisément identifiées.

Pour la recherche assistée des connectivités, nous avons utilisé les couples de spectres HNCB/HNCACO, HNCACB/CBCACONH, HNCA/HNCOCA.

Smartnotebook est aussi particulièrement utile pour naviguer dans l'espace des spectres à 3 dimensions possédant une dimension ^{15}N -HSQC. Il s'agit d'un outil précieux pour l'observation des spectres.

2.4.4.b. Le logiciel *PACES*

Le logiciel PACES (Protein sequential Assignment by Computer-assisted Exhaustive Search) (Coggins et Zhou, 2003) a également été utilisé. Cet outil se présente sous la forme de macros intégrées au programme Excel (Microsoft Office). PACES réalise une recherche exhaustive des connectivités mais aussi une identification des systèmes de spin. Le processus est itératif : l'utilisateur confirme les attributions non ambiguës des atomes du squelette peptidique et décide de la manière de gérer chaque ambiguïté. En relançant le calcul, le logiciel prend en compte des résultats du cycle précédent pour affiner la nouvelle recherche. Au fur et à mesure des cycles, les ambiguïtés sont écartées.

2.4.4.c. Le logiciel PISTACHIO

Le logiciel PISTACHIO (Probabilistic Identification of Spin systems and their Assignments including Coil-Helix Inference as Output) (Eghbalnia *et al.*, 2005a), récemment renommé PINE, est disponible sur Internet : (<http://miranda.nmrfam.wisc.edu/PINE/>). On fournit un ensemble de listes de pics correspondant aux spectres dont on dispose. Le logiciel retourne par email une liste des déplacements chimiques attribués, une évaluation de la probabilité que ces attributions soient correctes et une identification des structures secondaires est réalisée par le logiciel PECAN (Eghbalnia *et al.*, 2005b). L'algorithme utilisé par Pistachio est complexe, il repose sur l'optimisation itérative de matrices d'attributions, ce qui explique que les temps de calcul soient généralement longs (> 24h). Le logiciel accepte par défaut 15 expériences différentes mais il peut être facilement adapté à un grand nombre d'expériences différentes, y compris en présence de marquages sélectifs. Enfin, on peut fournir une attribution manuelle partielle pour aider le logiciel.

Si on intègre les données adéquates, PISTACHIO peut aussi réaliser l'attribution des chaînes latérales.

2.4.5. Attribution des chaînes latérales

2.4.5.a. Attribution des chaînes latérales aliphatiques

Les chaînes latérales aliphatiques ont été attribuées manuellement à partir des spectres HCCH-TOCSY, HCCH-COSY, H(CCO)NH, C(CO)NH, HNHA, HBHACONH. Le logiciel Smartnotebook a été utilisé pour l'observation des spectres à 3 dimensions possédant une dimension ^{15}N -HSQC : H(CCO)NH, C(CO)NH, HNHA, HBHACONH. Les carbones ont été attribués par comparaison de leur déplacement chimique avec les données statistiques issues de la BMRB (Biological Magnetic Resonance Data Bank, <http://www.bmrb.wisc.edu/>).

2.4.5.b. Les groupements NH₂ des asparagines et glutamines

Les protons amino des résidus asparagine et glutamine sont en général visibles lorsqu'ils ne sont pas en échange trop rapide avec l'eau. L'attribution de ces hydrogènes est délicate. Elle ne peut se faire qu'à travers l'interprétation des spectres NOESY (il existe des corrélations spatiales vers les hydrogènes intrarésiduels les plus proches : les H_β (Asn) et les H_γ (Gln)) ou grâce à des expériences hétéronucléaires particulières (HNCO optimisé pour les amines, COCAH...)).

Dans le cadre de notre étude, l'attribution des protons amino a été réalisée grâce au spectre ¹⁵N NOESY-HSQC.

2.4.5.c. Attribution des groupes aromatiques

L'attribution des résidus tryptophane, histidine, tyrosine et phénylalanine est particulièrement intéressante car ces résidus sont souvent impliqués dans la formation d'un cœur hydrophobe. Ils représentent donc une source de contraintes particulièrement riche.

L'attribution des Trp, Tyr et Phe a été réalisée grâce aux spectres 2D hbCBcgcdHD (corrèle le C_β (connu) avec le H_δ du cycle aromatique), hbCBcgcdceHE (corrèle le C_β (connu) avec le H_ε du cycle aromatique) et ¹³C HSQC (permet de déterminer le déplacement chimique du carbone lié directement au proton aromatique).

2.4.6. Détermination des paramètres de relaxation

2.4.6.a. Détermination des temps de relaxation T₁ et T₂

Les séquences d'impulsion pour mesurer les temps de relaxation longitudinale T₁ et transversale T₂ sont basées sur la séquence HSQC.

Un échantillonnage correct de la courbe de relaxation longitudinale impose de prendre des points répartis de façon homogène jusqu'à l'extinction du signal. Nous avons choisi de prendre les valeurs suivantes de: 0, 250, 500, 750, 1000, 1250, 1500, 1750, 2000, 2250, 2500, 2750 et 3000 ms. Le gain lors de l'acquisition doit être identique quelque soit

le délai pour une comparaison possible des intensités relatives. L'échantillonnage de la courbe mono-exponentielle de relaxation transversale a consisté en 8 points : 10, 30, 50, 70, 110, 150, 190, et 230 ms. Comme la vitesse de relaxation transversale est supérieure à la vitesse de relaxation longitudinale, la décroissance du signal est plus rapide. C'est pourquoi nous avons utilisé une échelle de valeurs plus réduite pour la mesure des T_2 (0 à 230ms) que pour la mesure des T_1 (0 à 3000ms).

Les spectres ont été enregistrés sur le spectromètre Varian 600 MHz du Centre de Biophysique Moléculaire. Chaque expérience a été réalisée avec 128 points complexes en dimension indirecte et 1472 points en dimension d'acquisition. Pour les mesures de T_1 et T_2 , 8 et 16 passages ont été enregistrés respectivement.

Le module *Rate Analysis* du logiciel NmrView a été utilisé pour extraire les paramètres T_1 et T_2 à partir d'un jeu de données expérimentales. L'équation de la relaxation utilisée est une monoexponentielle à trois paramètres ajustables : $I(0)$, T et C qui corrige une ligne de base imparfaite.

$$I(\delta) = I(0)e^{-\frac{\delta}{T}} + c$$

L'avantage principal du module *Rate Analysis* est de permettre une analyse très rapide d'un grand nombre de spectres à différents délais de relaxation δ . Il suffit du peak-picking et de l'attribution d'un seul spectre (celui correspondant au délai le plus petit, le plus intense) pour extraire l'intensité des pics des autres spectres. L'intensité des pics est calculée sur la base du point d'intensité la plus élevée situé dans une limite de $\pm 25\%$ de la taille de la boîte autour du centre de la boîte (Jitter). Cette procédure est très efficace car elle autorise de faibles déplacements des fréquences d'un spectre à l'autre. L'incertitude sur l'intensité est déterminée par NmrView par calcul de l'écart type du bruit sur une boîte ne contenant pas de signal. D'après Palmer et al., les temps de relaxation sont estimés avec une plus grande précision en utilisant l'intensité des pics au lieu du volume (Palmer III *et al.*, 1991). Pour chaque résidu, la valeur de l'intensité est représentée en fonction du délai δ . L'incertitude sur les paramètres extraits a été déterminée en utilisant une simulation de type Monte-Carlo. Le nombre de simulations a été fixé à 200. Les 200 jeux de paramètres ainsi calculés suivent théoriquement une distribution gaussienne. Les extrémités de l'intervalle de confiance sont fixées à 65%,

c'est-à-dire les valeurs extrémales des 2*65 jeux de paramètres les plus proches des paramètres calculés à partir des intensités mesurées. La procédure de minimisation est automatisée sur l'ensemble des résidus. Les résultats sont ensuite vérifiés manuellement, en particulier, l'extraction des intensités et la forme de l'exponentielle.

2.4.6.b. Détermination de l'effet NOE hétéronucléaire ^1H - ^{15}N

Pour la mesure de l'effet nOe hétéronucléaire, la séquence gNnoe (BioPack, Varian Inc.) a été utilisée (Farrow *et al.*, 1994). Le délai d_1 d'établissement de l'équilibre sous irradiation des protons amides a été fixé à 5 s et un délai d_0 de 10 s a été laissé entre deux scans. Deux expériences ont été réalisées : avec et sans irradiation. Chaque expérience a été réalisée avec 128 points complexes dans la dimension ^{15}N et 1280 points en dimension d'acquisition ^1H .

Les valeurs des NOE hétéronucléaires η ^1H - ^{15}N sont déterminées par le rapport suivant :

$$\eta = \frac{I_{irr}}{I_{eq}}$$

Où I_{irr} et I_{eq} correspondent respectivement aux intensités des pics croisés dans l'expérience avec et sans irradiation du proton.

Le module *HetNOE* de NMRView a été utilisé afin d'automatiser la mesure de l'effet nOe η . L'expérience non saturée, qui contient les intensités I_{eq} les plus importantes, est "peak-pickée" et attribuée. Le module mesure ensuite l'intensité I_{sat} des pics dans l'expérience saturée en mode « Jitter » (paragraphe 2.4.6.a). L'incertitude sur l'effet nOe est calculée sur la base de la déviation standard $sdev$ du bruit déterminée dans une région vide de signal (commande *getval*) et de l'intensité du pic non saturé. Cette approche fait l'approximation que le bruit est identique dans les deux expériences et que les incertitudes sur I_{sat} et I_{eq} sont identiques.

2.5. Calcul de structure

Le logiciel CYANA version 2.1 a été utilisé avec les paramètres par défaut pour les macros *noeassign* et *autoaco*. Les données d'entrée ont été : la liste des déplacements chimiques, les contraintes d'angles obtenues par le logiciel Talos, les listes de pics non attribuées des spectres 3D NOESY-HSQC ^{15}N et ^{13}C . Les tolérances de déplacement chimique ont été fixées à 0,05ppm dans la dimension proton et 0,75ppm dans les dimensions ^{15}N et ^{13}C . Le calcul a été réalisé avec 20000 pas de dynamique et le calcul de 200 structures.

2.6. Interaction de la protéine CG7054 avec la phosphoryléthanolamine

L'interaction de la protéine CG7054 avec la phosphoryléthanolamine ($\text{NH}_2\text{CH}_2\text{CH}_2\text{OPO}_3\text{H}_2$) a été testée par RMN sur un échantillon doublement marqué $^{15}\text{N}/^{13}\text{C}$, concentré à 0,5 mM dans le tampon habituel (voir le paragraphe 2.4, page 198). La phosphoryléthanolamine (Sigma Aldrich) a été ajoutée par paliers sous forme de poudre lyophilisée pour atteindre : 0 (spectre de référence), 0,5, 1, 2, 4, 8, 16 et 32 mM. Le pH a été vérifié et le cas échéant corrigé par ajout de soude diluée. Des spectres ^1H - ^{15}N HSQC ont été enregistrés.

3. Résultats et discussion

3.1. Clonage de l'ADNC du gène *cg7054*

L'ADN complémentaire correspondant au gène *cg7054* de *Drosophila melanogaster* a été obtenu commercialement à partir du consortium international « Berkeley Drosophila Genome Project » (clone GH14779). Le gène est fourni inclus dans le plasmide pOT2, entre le site 5' EcoRI et 3' XhoI.

Nous avons ensuite réalisé une PCR pour amplifier le fragment d'intérêt et insérer les sites de coupures aux endonucléases. Le choix des amorces s'est porté sur les séquences présentes en 3' et en 5' de la région codante auxquelles ont été ajoutés les sites de restriction XhoI et NdeI, qui sont absents dans la séquence du gène d'intérêt. Ces sites de restriction ont permis une insertion en phase avec le promoteur situé sur le plasmide

pET31b. Ainsi, d'après la séquence donnée par le séquençage complet de l'ADNc du gène *cg7054*, les amorces choisies présentent les séquences suivantes :

Forward CG7054 : 5' – **CA/T ATG** GAT GAC ATA GTA CCC GAT G 3'

Site NdeI

Reverse CG7054 : 5' – **C/TC GAG** **CTA** CCC TAC GAT CGT TTT ATT CC –3'

Site XhoI **codon stop**

La PCR s'est déroulée de manière classique et les produits de PCR ont ensuite été purifiés sur gel d'agarose 1%. La bande correspondant à l'ADNc du gène *cg7054* (547 pb) a été extraite du gel.

Le fragment purifié a ensuite été traité grâce au kit pCR2.1-TOPO Cloning : les produits de PCR sont inclus dans le vecteur pCR2.1-TOPO puis transformés dans les cellules Top10. Ce kit a permis l'obtention d'un excellent rendement de transformation des produits de PCR et n'a pas nécessité de digérer le vecteur avant l'insertion. Après transformation, la sélection des bactéries Top10 ayant intégré le plasmide pCR2.1-TOPO a été réalisée en présence d'ampicilline, la résistance étant induite par un gène porté par le plasmide. Un second système de sélection faisant intervenir la β -galactosidase a été inclus dans le vecteur pCR2.1-TOPO : les colonies possédant un plasmide pCR2.1-TOPO ayant intégré un fragment d'ADN sont blanches en présence de X-Gal et d'inducteur IPTG. Les bactéries possédant le vecteur pCR2.1-TOPO sans fragment sont alors bleues.

Plusieurs colonies blanches ont ensuite été repiquées en milieu liquide LB et ampicilline 100 $\mu\text{g.mL}^{-1}$. L'ADN plasmidique a été extrait et une fraction a subi une digestion enzymatique par les enzymes XhoI et NdeI afin de vérifier la présence d'un fragment de la taille de l'insert *cg7054* (547pb). Les clones pour lesquels une bande de la taille désirée a été visualisée sur gel d'agarose 1% en présence de bromure d'éthidium ont été retenus et le reste du plasmide disponible digéré de manière préparative par les mêmes enzymes de restriction. Le plasmide pET31b(+) a été ouvert par digestion par ces mêmes enzymes.

Le fragment *cg7054* et le plasmide ont été purifiés sur gel d'agarose 1%, extraits du gel par kit et ont subi une étape de ligation puis de transformation dans les cellules E. Coli

C600. La transformation directe des bactéries E. Coli BL21 DE3 est très délicate et le rendement est habituellement très faible. Il est donc plus facile d'utiliser comme intermédiaire les bactéries E. Coli C600 pour amplifier le plasmide. Leur transformation et culture sont plus aisées. Après la transformation, la sélection des clones de C600 ayant reçus le plasmide {pET31b-cg7054} a été réalisée en présence d'ampicilline $100 \mu\text{g.mL}^{-1}$. L'identification des plasmides ayant intégré le fragment d'intérêt est obtenue par restriction enzymatique par les enzymes NdeI et XhoI qui ont libéré le fragment recherché (547pb), le cas échéant.

Une partie de l'ADN plasmidique des clones positifs a ensuite été envoyée à séquencer (MWG Biotech) et l'autre partie utilisée pour transformer les bactéries E. Coli BL21DE3.

La séquence obtenue a permis un alignement complet avec la séquence de l'ADNc fournie par le consortium *Berkeley Drosophila Genome Project* : aucune mutation ne s'est intégrée durant les différentes étapes du clonage du fragment codant du gène *cg7054*.

Après la transformation, la sélection des clones de E. Coli BL21DE3 ayant reçus le plasmide {pET31b – cg7054} est réalisée en présence d'ampicilline $100 \mu\text{g.mL}^{-1}$, et l'identification des plasmides ayant intégré le fragment d'intérêt s'est faite par restriction enzymatique par les enzymes NdeI et XhoI.

La présence du fragment a ensuite été vérifiée en réalisant un test d'induction avec 1mM d'IPTG sur une culture en phase de croissance exponentielle. La Figure VI-8 montre ce que nous avons obtenu après dépôt sur gel SDS-PAGE : la surexpression de la protéine CG7054 est spectaculaire et bien visible pour tous les clones testés. Cette protéine de 19,8 kDa migre à environ 25 kDa. Les PEBP en général ont la propriété de présenter une masse moléculaire apparente supérieure à la masse moléculaire (Schoentgen *et al.*, 1993). Par exemple la PEBP de bœuf migre à une taille de 23 kDa alors qu'elle ne fait que 20,9 kDa. Il semble que ce phénomène serait du à la présence résiduelle de structures secondaires.

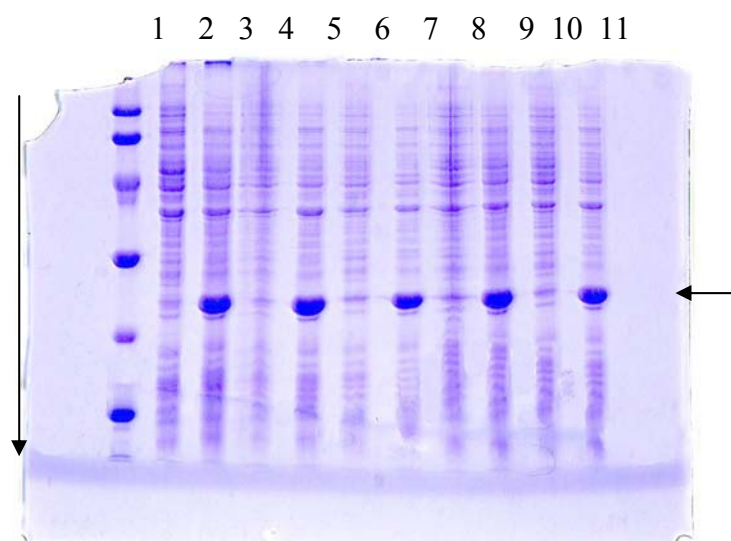


Figure VI-8 : Test de surexpression de la protéine CG7054 par 5 clones de bactéries *E. coli* BL21DE3 ayant reçus le plasmide {pET31b – cg7054}. Les dépôts ont été réalisés dans l'ordre non induit / induit pour chaque clone. (1) marqueur de masse moléculaire (97, 66, 45, 30, 20,1, 14,4 kDa, de haut en bas). (2, 4, 6, 8,10) avant induction. (3, 5, 7, 9, 11) après induction. La flèche horizontale indique la protéine CG7054 et la flèche verticale le sens de migration. Gel SDS PAGE 15%, coloration au bleu de Coomassie.

3.2. Production et purification de la protéine CG7054

Nous avons optimisé les étapes de purification pour obtenir un protocole simple, économique et facile à mettre en œuvre. Nous avons par ailleurs choisi les méthodes les plus douces pour préserver la protéine CG7054. Les protocoles décrits permettent d'obtenir en une semaine environ des rendements records : 100 mg de protéine par litre de milieu de culture non marqué (LB) ou 160 mg par litre de milieu marqué (M9).

Les bactéries *Escherichia coli* BL21DE3 contenant le plasmide pET31b pourvu de l'insert d'intérêt sur-expriment la protéine CG7054 par ajout d'IPTG. Différents essais nous ont montré que la surexpression était optimale à 30°C pendant 5h dans le milieu classique LB ou 6h dans le milieu marqué. La méthode développée par Marley (Marley *et al.*, 2001) a donné d'excellent résultats tant pour le rendement d'intégration des isotopes (la protéine est marquée à 95% en ^{15}N et 94% pour le ^{13}C (estimation réalisée par spectrométrie de masse), que pour la quantité de protéine produite par rapport à la quantité de sulfate d'ammonium marqué ^{15}N et de glucose marqué ^{13}C investie

Les cellules sont ensuite lysées par trois cycles courts de sonication (1 minute) et centrifugées. La surexpression de la protéine CG7054 étant particulièrement bonne dans le milieu riche et le milieu minimum, le surnageant est très riche en protéine CG7054.

Après la lyse, le surnageant est rapidement déposé sur une colonne de QAE-sephadex. La protéine CG7054 possède un point isoélectrique acide voisin de 5, à pH 8 (pH de la colonne d'échange d'ions) elle est donc chargée négativement et reste fixée sur la colonne QAE chargée positivement. La colonne est ensuite largement rincée, puis la protéine est éluée par ajout de 0,2 M de NaCl dans le tampon de chromatographie. La protéine CG7054 est éluée rapidement en formant un pic particulièrement concentré (Figure VI-9).

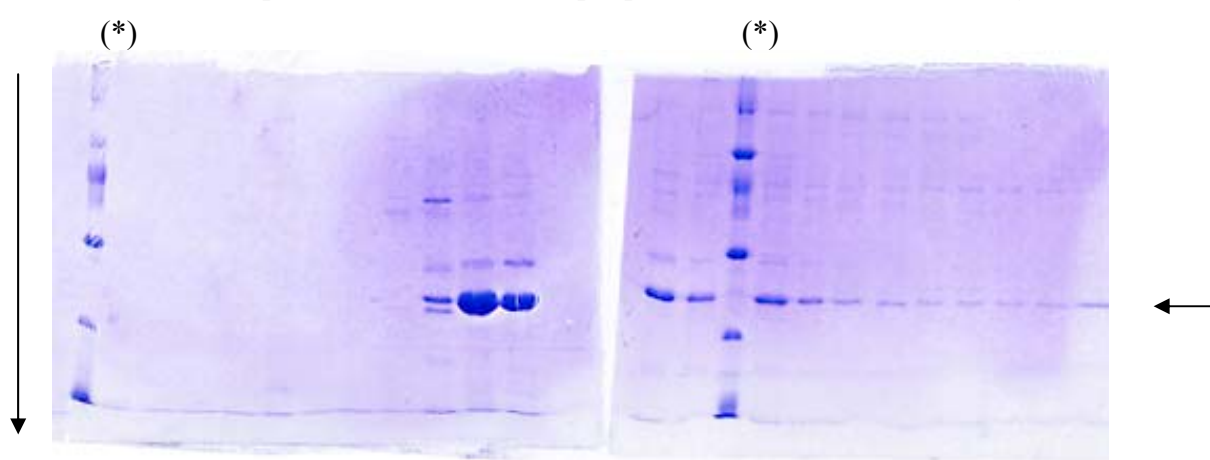


Figure VI-9 : Colonne QAE Sephadex : Elution de la protéine CG7054 après ajout de 0,2M NaCl. Une fraction sur 3 a été déposée chronologiquement de gauche à droite après l'ajout de sel dans le tampon d'éluion. (*) Marqueur de masse moléculaire (97, 66, 45, 30, 20,1, 14,4 kDa, de haut en bas). La flèche horizontale indique la protéine CG7054 et la flèche verticale le sens de migration. Gel SDS PAGE 15%, coloration au bleu de Coomassie.

Les principales fractions d'intérêt sont rassemblées et le pool est dialysé contre de l'eau pure avant lyophilisation. Il faut remarquer que durant la dialyse de nombreuses espèces protéiques contaminantes précipitent, alors que la protéine CG7054 demeure soluble. Cette étape de dialyse finale participe donc à la purification. Au final, nous obtenons une protéine très enrichie dont la pureté de l'ordre de 95% est satisfaisante pour les études RMN (Figure VI-10).

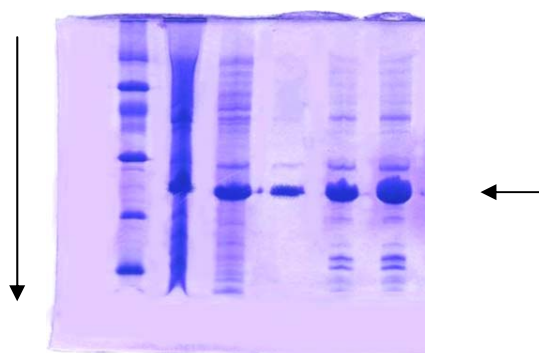


Figure VI-10 : Purification de la protéine (1) marqueur de masse moléculaire (97, 66, 45, 30, 20, 14,4 kDa, de haut en bas). (2) Après induction. (3) Après lyse. (4) Après la colonne d'échange d'ion (QAE). (5,6) Protéine finale, 10 µg et 20 µg. La flèche horizontale indique la protéine CG7054 et la flèche verticale le sens de migration. Gel SDS PAGE 15%, coloration au bleu de Coomassie.

Il faut remarquer que les protéines constitutives des bactéries sont produites essentiellement durant la croissance des bactéries dans le milieu LB, avant le passage dans le milieu marqué. Ces protéines présentent donc un taux de marquage plus faible que la protéine d'intérêt, ce qui a tendance à minimiser leur impact sur les spectres hétéronucléaires.

3.3. Recherche des conditions RMN

3.3.1. Étude de conditions en gouttes suspendues

En écartant les conditions (pH, sel, etc.) dans laquelle la protéine précipite à forte concentration, la technique des gouttes suspendues, courante dans le domaine de la cristallographie, a constitué une première étape simple à mettre en œuvre et économique en quantité de protéine pour la recherche des conditions RMN. Nous avons commencé par tester les conditions qui avaient été favorables pour la protéine CG18594 : le tampon phosphate de potassium à différentes valeurs de pH sans ajout de sel. Il est apparu que la protéine restait parfaitement soluble à forte concentration au dessus de pH 5,5.

3.3.2. Dichroïsme circulaire

Le dichroïsme est une méthode de choix pour étudier la structure secondaire des protéines. Nous avons mis à profit cette méthode afin d'observer le comportement de la

protéine dans différentes conditions physico-chimiques. Nous avons de cette manière observé le comportement de la protéine lors du chauffage de 5°C à 85°C en fonction du pH.

3.3.2.a. Effet du pH

Les spectres ont été enregistrés entre 195 et 260 nm en utilisant une cuve de trajet optique 1 mm. La concentration de la protéine était de 0,2 mg.mL⁻¹ dans du tampon phosphate de potassium 10 mM, pour des valeurs de pH variant d'une unité entre 2 et 12. La protéine précipite à pH inférieur à 3. L'absorption du tampon ne permet pas de présenter les spectres de dichroïsme avec une qualité raisonnable en dessous de 205 nm. La Figure VI-11 présente les spectres obtenus à 5°C en fonction du pH.

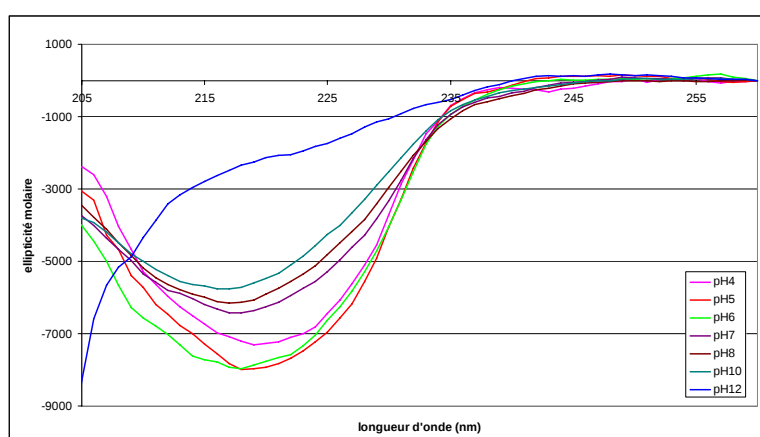


Figure VI-11 : Spectres de dichroïsme circulaire dans l'UV lointain de la protéine CG7054 en fonction du pH, à 5°C.

La déconvolution des spectres avec l'algorithme Selcon3 permet de mieux comprendre les spectres observés (Tableau VI-5) :

pH	% d'hélices α	% de feuillets β	Turns
4	16	28	12
5	18	25	11
6	18	21	13
7	13	21	12
8	9	23	13
10	9	23	12
12	0	3	5

Tableau VI-5 : Résultats de la déconvolution des spectres de dichroïsme circulaire de la protéine CG7054 en fonction du pH avec l'algorithme Selcon 3.

Nous observons nettement une variation du taux d'hélices qui semble être maximal entre pH 5 et 6. Par ailleurs la quantité de feuillets semble augmenter quand le pH diminue.

A pH 12 nous pouvons observer le spectre caractéristique d'une protéine dénaturée.

Ces résultats sont tout à fait en accord avec les résultats obtenus pour les autres PEBP (Schoentgen *et al.*, 1987). La protéine semble donc correctement repliée, avec un taux de structures secondaires important. Enfin il semble qu'un pH raisonnablement acide (entre 4 et 7) favorise la formation des structures secondaires.

3.3.2.b. Effet de la température

Les études de dichroïsme circulaire ont montré que la structure de la protéine CG7054 variait sensiblement en fonction du pH. Nous avons alors voulu savoir si cela se traduisait par des modifications de sa stabilité thermique. Nous avons donc suivi la dénaturation thermique de la protéine par dichroïsme circulaire à pH 5, 6, 7 et 8 (Figure VI-12 et Figure VI-13).

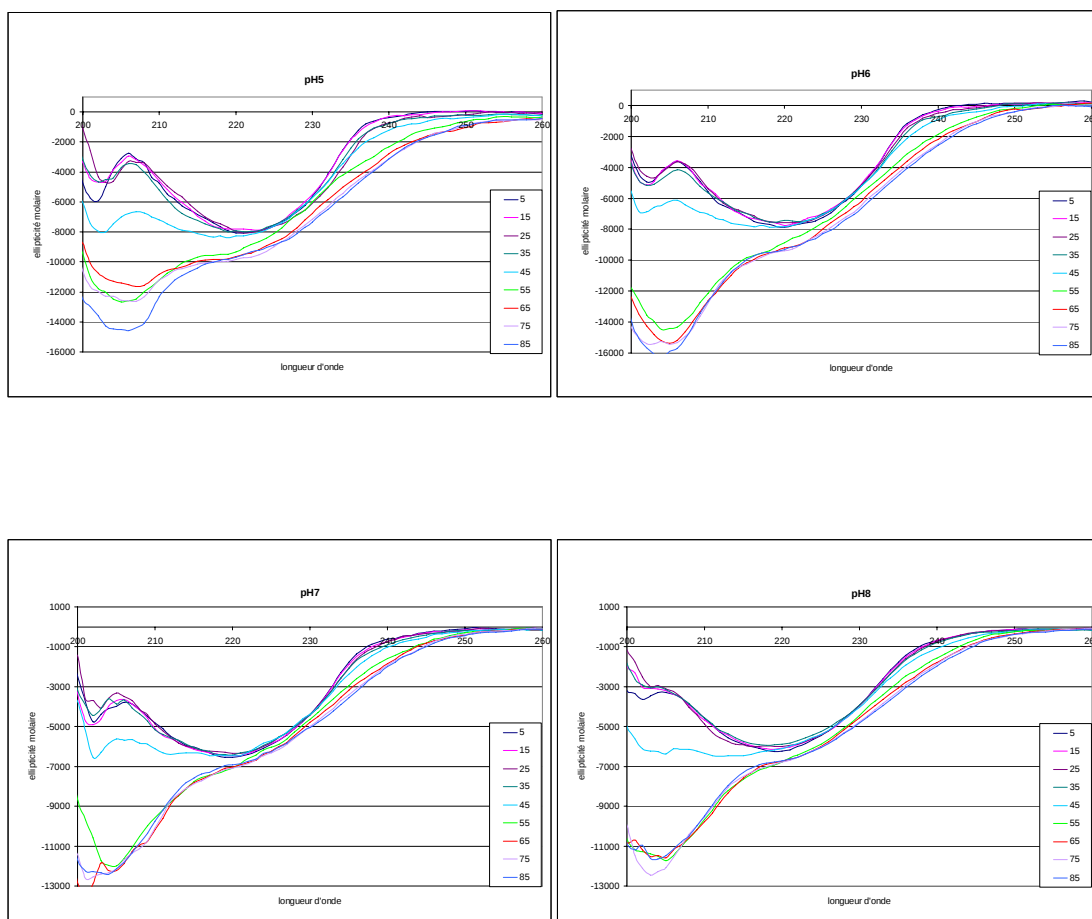


Figure VI-12 : Suivi par dichroïsme circulaire de la dénaturation thermique de la protéine CG7054 à différents pH.

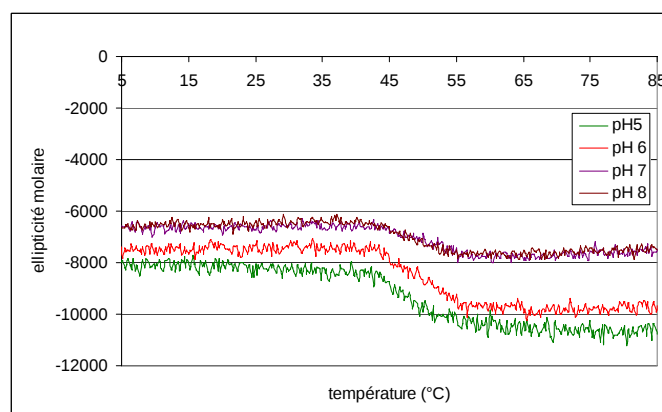


Figure VI-13 : Evolution du signal de dichroïsme circulaire à 215 nm en fonction de la température à pH 5, 6, 7 et 8.

Classiquement, ces courbes montrent que la protéine CG7054 se dénature lorsque la température augmente. Les variations d'intensité du signal observées sont liées à la modification du taux de structures secondaires en fonction du pH. Il est intéressant de remarquer que les courbes d'évolution du signal à 215 nm sont parfaitement superposables à pH 7 et 8. La température de dénaturation aux alentours de 52°C à pH 5 et 6 et 49°C à pH 7 et 8, ainsi que l'allure générale des courbes (pente de la tangente à la courbe sigmoïde de dénaturation au point de transition plus importante à pH 5 et 6) tend à montrer que la protéine CG7054 est légèrement plus stable en milieu acide qu'au dessus de pH 7.

3.3.2.c. Conclusions pour le choix des conditions de l'échantillon pour l'étude structurale par RMN.

L'étude réalisée grâce à la technique des gouttes suspendues a indiqué que la protéine CG7054 était parfaitement soluble à haute concentration dans du tampon phosphate de potassium au dessus de pH 6. L'emploi de ce tampon est par ailleurs tout à fait indiqué pour les PEBP dans la mesure où ces protéines sont caractérisées par leur capacité à lier les anions organiques, notamment ceux porteurs d'un groupement phosphate. En utilisant ce type de tampon, il est donc probable qu'un phosphate va se fixer au niveau du site actif et ainsi stabiliser la protéine, le tampon servant donc à la fois de ligand et de régulateur de pH. Cependant, bien que favorable pour réaliser l'étude de la structure 3D de la protéine, cette situation peut causer des problèmes de compétition lors des études d'interaction.

Les études par dichroïsme circulaire nous ont par ailleurs montré que la protéine était repliée et que les structures secondaires étaient favorisées par un pH acide. Nous avons donc choisi de réaliser le premier tube RMN à pH 6,5 avec du tampon phosphate de potassium 50 mM.

3.4. RMN

3.4.1. Stabilité de la protéine CG7054 au cours du temps

Le premier spectre ^1H - ^{15}N HSQC a été enregistré dans les conditions déterminées précédemment. Cette expérience permet de corrélérer les protons et les azotes directement liés. On y observe donc les signaux des protons amide du squelette peptidique, sauf pour le premier résidu et les prolines, mais aussi des chaînes latérales des asparagines, glutamines, arginines et tryptophanes. Théoriquement nous devons donc compter sur le spectre 210 pics, nous en avons dénombré 207.

La bonne dispersion des résonances dans les deux dimensions indique par ailleurs que la protéine est repliée. Enfin, les largeurs de raie des différentes résonances sont homogènes et fines, contrairement à celles que nous avons observées pour la protéine CG18594. La largeur de raie semble indiquer que la protéine CG7054 est sous forme monomérique.

Le spectre ^1H - ^{15}N HSQC se révélant particulièrement prometteur, nous avons suivi son évolution au cours du temps pour vérifier la stabilité de la protéine. Après une dizaine de jours à 20°C, le spectre n'avait absolument pas évolué et le tube RMN restait limpide. Nous avons donc décidé de débiter l'étude structurale. Le premier échantillon concentré a été réalisé fin avril 2005 et conservé à 4°C en dehors des périodes d'acquisitions des spectres ; son spectre HSQC n'a actuellement pas du tout évolué. La protéine CG7054 est donc extraordinairement stable dans les conditions déterminées.

3.4.2. Attribution du squelette peptidique

La stratégie décrite dans le chapitre II a permis l'attribution de la totalité des résonances attribuables du squelette peptidique. Les logiciels d'attribution automatiques PACES, Pistachio et Smartnotebook ont fourni des pistes intéressantes dans la recherche des premières connectivités. Ces trois outils ont permis l'attribution rapide d'environ 60% de la chaîne principale. Le reste correspond soit à des résidus dont les corrélations sont faibles voire absentes ou alors à des résidus dont les corrélations présentent des superpositions. Il a alors fallu combiner manuellement les informations de l'ensemble des spectres disponibles pour gérer la faiblesse des signaux, les ambiguïtés et résoudre les recouvrements (Figure VI-14).

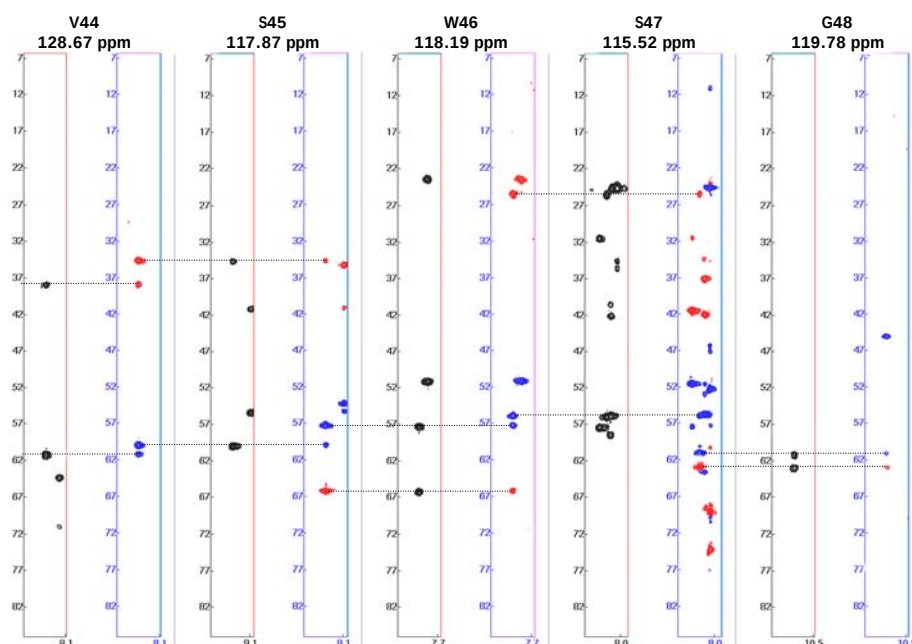


Figure VI-14 : Extraits des spectres CBCACONH (en noir, 600MHz) et HNCACB (en bleu et rouge, 800 MHz) présentant l'attribution des $C\alpha$ et $C\beta$ des résidus 44 à 48

Certains résidus semblent situés dans des zones flexibles, les corrélations correspondantes sont alors faibles et traduisent une dynamique locale importante. Il s'agit essentiellement : des 5 résidus N-terminaux, de la thréonine 36, des résidus 75 à 80, des résidus 100 à 104, de la tyrosine 162 et des résidus 167 à 175.

Les difficultés avec les résidus situés en bout de chaîne sont courantes et traduisent généralement un échange conformationnel. Il est tout particulièrement intéressant de remarquer que les résidus 36, 75 à 80, 100 à 104 et 167 à 175 appartiennent à des boucles qui entourent ou constituent le site de fixation putatif des ligands anioniques (cf. infra, Détermination de la structure de la protéine CG7054, page 223).

Le spectre ^1H - ^{15}N HSQC avec l'attribution est représenté sur la Figure VI-15.

être avancée et fait intervenir la désamidation de l'asparagine 133. Ce phénomène constitue la modification post-traductionnelle non enzymatique des protéines la plus fréquente. Il a longtemps été considéré comme un artefact de purification lié au vieillissement de l'échantillon, mais plusieurs études ont montré son existence *in vivo* (Flatmark, 1967). L'observation d'asparagines labiles conservées au cours de l'évolution amène plusieurs auteurs à proposer ce phénomène comme marqueur du vieillissement des protéines dans la cellule (Weintraub et Manson, 2004). Cette modification spontanée se produit sur les asparagines très exposées au solvant, comme c'est le cas pour l'asparagine 133 qui est située au niveau d'une boucle à la surface de la protéine, de plus l'enchaînement des résidus dans cette zone (SNTT) forme un site favorable pour la désamidation (Robinson et Robinson, 2004). Cette séquence correspond aussi à un site de N-glycosylation. Cependant la protéine a été produite chez *Escherichia coli* qui ne réalise pas cette modification post traductionnelle, et aucune glycolysation n'a pour l'instant été observée dans la famille des PEBP.

L'attribution des protons aromatiques a été beaucoup plus problématique et demeure actuellement partielle. En effet, la sensibilité des spectres hbCBcgcdHD et hbCBcgcdceHE est faible et les incertitudes sur les déplacements chimiques des carbones β (incertitudes d'attribution liées à la résolution des spectres 3D dans les dimensions hétéronucléaires) conduisent à des ambiguïtés que l'observation des cartes NOESY-HSQC ^{13}C et ^{15}N n'a pas réussi à lever.

3.4.4. Collection des données structurales

3.4.4.a. Identification des structures secondaires et contraintes d'angles

Les déplacements chimiques de la protéine CG7054 ont été exploités pour déterminer les structures secondaires et générer des contraintes d'angles dont l'importance lors du calcul de structure est primordiale. Plusieurs méthodes ont été utilisées et les résultats combinés pour extraire le maximum d'informations.

3.4.4.a.i. CSI

La méthode du *Chemical Shift Index* a été appliquée à la protéine CG7054. Les résultats sont présentés en annexe 3. Nous observons la présence de plusieurs feuillets β : I17 à I20, E32 à -T34, L55 à D62, W79 à V82, I115 à Y119 et I157 à Q165 (Figure VI-16). Deux hélices apparaissent de N141 à G152 et de I172 à K175. Ces résultats sont en accord avec les motifs structuraux généralement rencontrés dans la famille des PEBP.

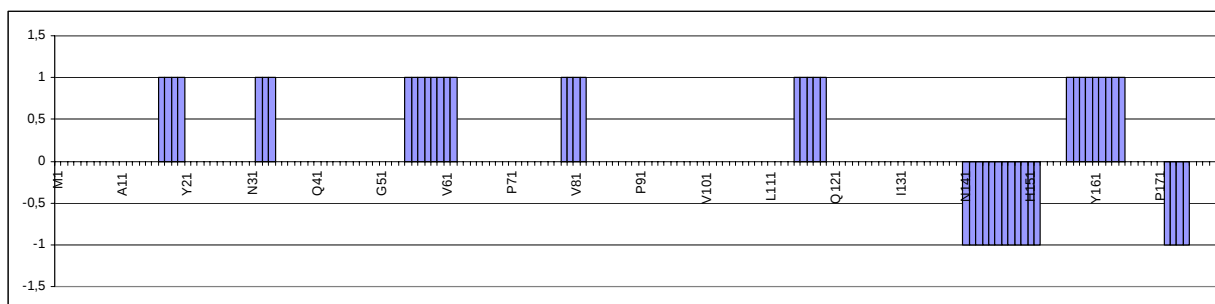


Figure VI-16 : Consensus sur les valeurs du CSI obtenu pour la protéine CG7054.

3.4.4.a.ii. Talos

Le logiciel Talos est largement utilisé pour générer des contraintes sur les angles dièdres ϕ et ψ du squelette peptidique (Figure VI-17). Comme nous l'avons déjà décrit ce logiciel utilise une banque de données contenant les déplacements chimiques d'un ensemble de protéines. Cette banque a été mise à jour récemment et comporte actuellement 78 protéines au lieu des 20 initiales. Les prédictions ambiguës sont ainsi moins fréquentes et l'intervalle des angles prédits plus fin.

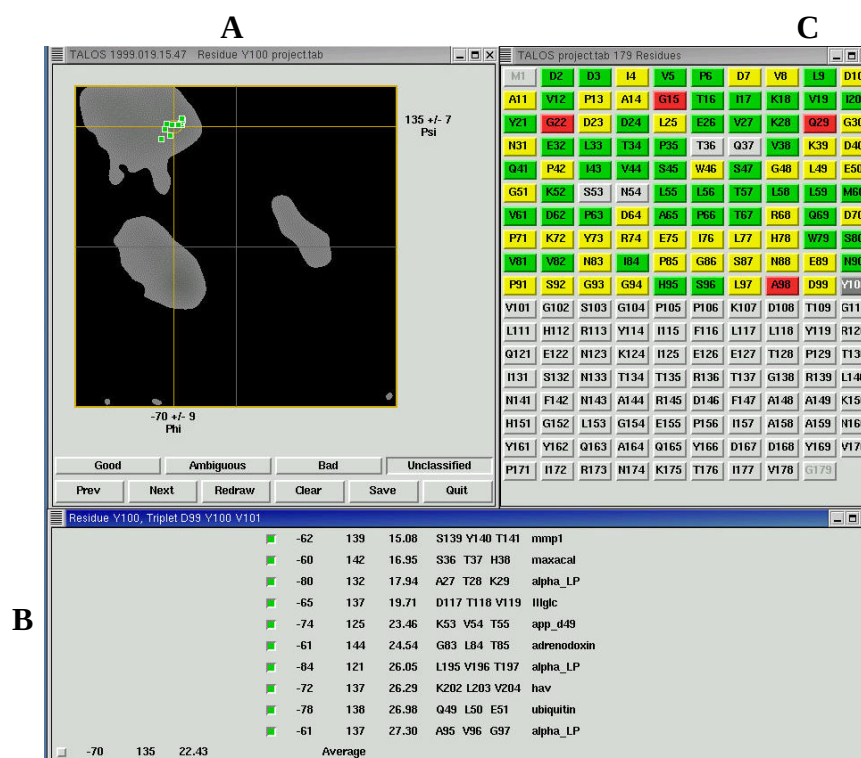


Figure VI-17 : Graphique montrant le résultat du logiciel TALOS pour la tyrosine 100 de la protéine CG7054. Les déplacements chimiques des $^1\text{H}_\alpha$, $^{13}\text{C}_\alpha$, $^{13}\text{C}_\beta$, CO et N ont été utilisés. A/ Graphique de Ramachandran illustrant les angles dièdres du résidu central des dix tripeptides de la banque de données ayant des déplacements chimiques secondaires similaires à ceux du tripeptide D99-Y100-V101 (B). Cette analyse permet d'estimer les angles dièdres ϕ / ψ de Y100 à -70 ± 9 / 135 ± 7 . C/ Les régions vertes représentent les résidus où TALOS suggère une structure secondaire avec un niveau de confiance élevé. Les régions rouges représentent les régions où la structure secondaire est contradictoire (Bad). Les régions jaunes signifient des régions ambiguës et les régions grises correspondent aux régions non classées.

Talos a permis la prédiction de 128 couples d'angles (ϕ , ψ) avec des intervalles allant de $\pm 3^\circ$ à $\pm 34^\circ$. Ces angles sont présentés dans l'annexe 4. L'utilisation de contraintes d'angles à $\pm 3^\circ$ n'est pas réaliste dans le cadre des étapes de modélisation moléculaire, nous avons donc utilisé des intervalles deux fois supérieurs à ceux fournis par Talos, avec un minimum de $\pm 10^\circ$.

3.4.4.a.iii. Comparaison des différentes méthodes

Les résultats en provenance du CSI et de Talos sont globalement compatibles entre eux. Talos détecte deux petits feuillets en plus par rapport au CSI (résidus 24 à 27 et 41 à 45). Quelques différences apparaissent cependant au niveau des limites des structures

secondaires (Figure VI-18). Par exemple, l'hélice C-terminale apparaît légèrement plus longue avec Talos (résidus 169 à 175) qu'avec le CSI (résidus 172 à 175). La méthode Talos est plus sensible que le CSI, il convient donc surtout de vérifier que toutes les structures secondaires détectées par le CSI apparaissent au niveau des prédictions de Talos. Il faut remarquer que Talos apporte par ailleurs des informations concernant les zones situées hors des structures secondaires.

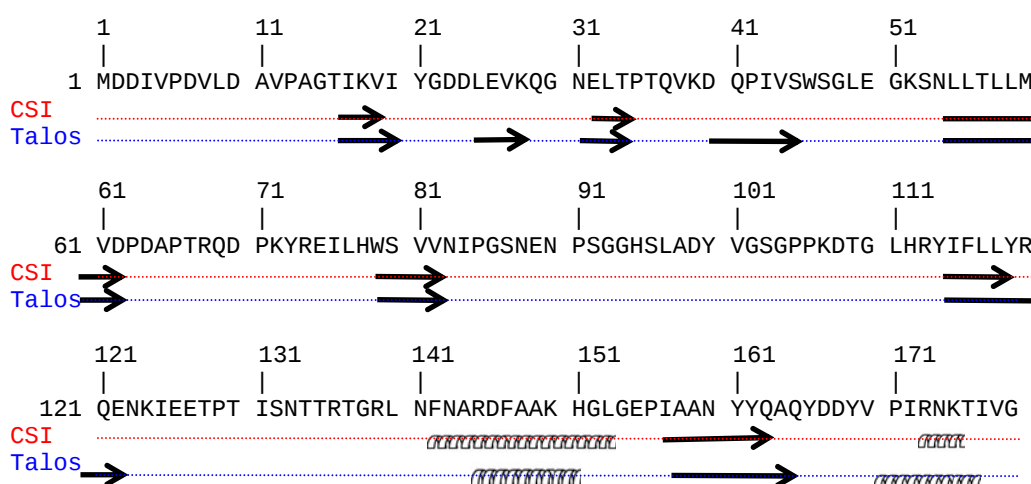


Figure VI-18 : Détermination des structures secondaires de la protéine CG7054 à partir des données RMN grâce aux méthodes CSI et Talos. Les flèches correspondent aux zones en feuillets β , les hélices aux zones en hélices α .

Si on réalise un bilan des informations obtenues (Figure VI-19), on observe que les structures secondaires de la protéine CG7054 sont globalement conservées, en nombre et en position, par rapport aux PEBP humaine et bovine. Seul le feuillet $\beta 8$ de la PEBP humaine ne semble pas présent, mais si on observe les prédictions de Talos, on se rend compte que les angles se situent à la limite de la zone des feuillets β de la carte de Ramachandran. Certains feuillets apparaissent plus courts, par exemple le brin $\beta 7$. Cette différence observée peut être due à la technique employée plutôt qu'à des différences réelles. En effet, les structures de PEBP actuellement connues ont été résolues par cristallographie. Dans un cristal les protéines sont empilées et stabilisées par des contacts intermoléculaires dont l'importance n'est pas négligeable. Cela peut favoriser la formation des structures secondaires en limitant la dynamique interne de la protéine. La

RMN des protéines étudie les molécules en solution et il est courant d'observer des structures secondaires plus courtes que dans les modèles cristallographiques.

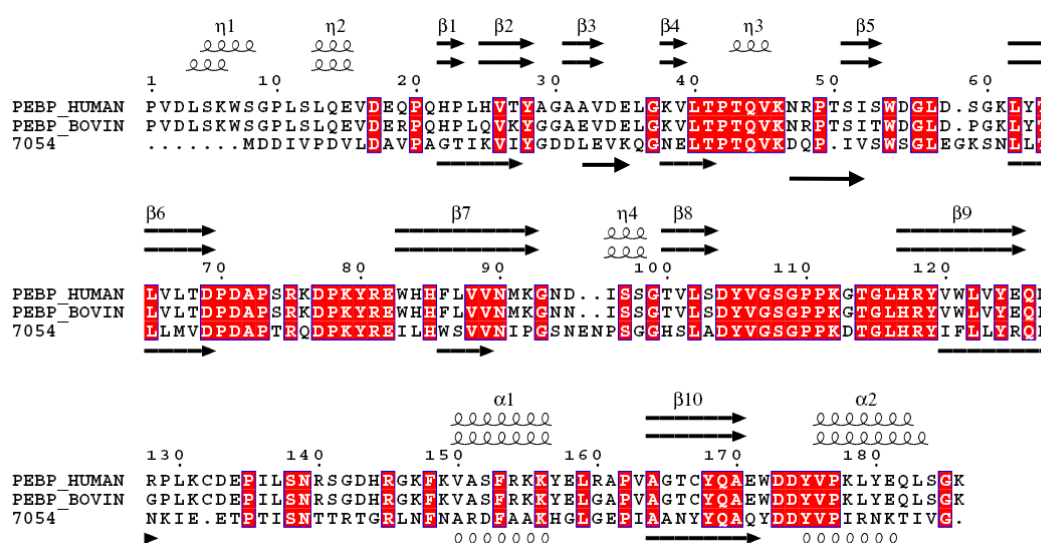


Figure VI-19 : Comparaison des structures secondaires déterminées par Talos de la protéine CG7054 avec celles des protéines PEBP-1 humaine et bovine. Les séquences ont été alignées avec CLUSTALW, (Eddy, 1995). Les cadres rouges indiquent les résidus identiques. Les structures secondaires obtenues sur les modèles cristallographiques sont indiquées. (Figure réalisée avec le logiciel Esprimo)

3.4.4.b. Echange H/D

L'échange H/D met en évidence l'accessibilité et/ou l'implication d'un proton amide dans une liaison hydrogène. En dissolvant la protéine CG7054 doublement marquée ^{13}C - ^{15}N dans du D_2O , j'ai suivi par une série de spectres ^1H - ^{15}N HSQC la disparition des signaux des protons amides (Figure VI-20). Sur les 165 signaux correspondant aux protons amides du squelette peptidique, 5 sont de trop faibles intensités pour être suivis correctement, 74 disparaissent immédiatement (en moins de 15 minutes après la mise en solution de l'échantillon), 24 disparaissent dans un délai de 24 heures, 5 dans le délai 24/72 heures, 6 dans le délai de 72 heures/1 mois et 51 persistent au-delà d'un mois après la mise en solution. Cela montre que la protéine est particulièrement stable : les protons amides engagés dans les structures secondaires situées au cœur de la protéine sont protégés de l'échange avec un atome de deutérium.

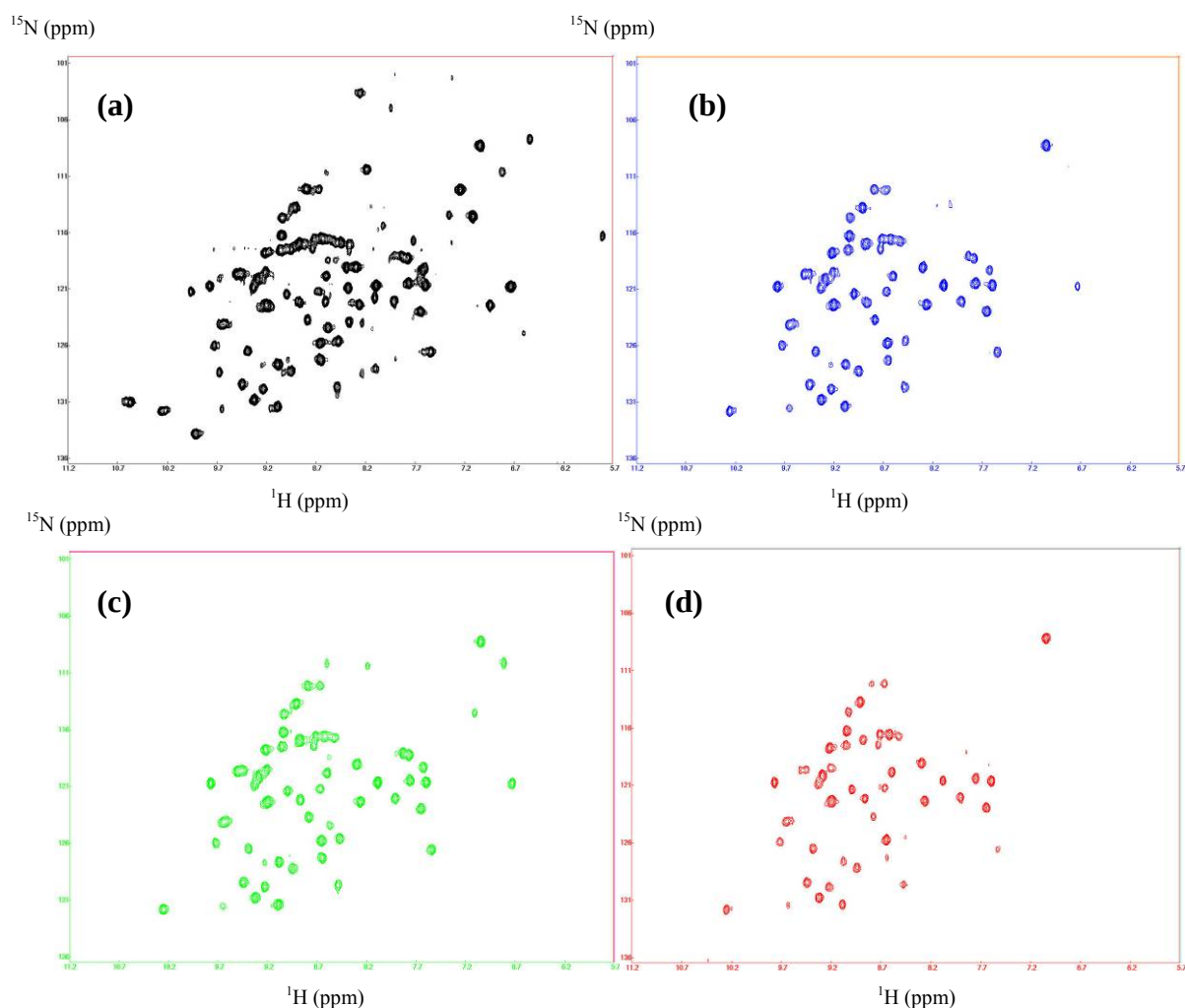


Figure VI-20 : Spectres ^1H - ^{15}N HSQC obtenus lors des expériences d'échange H/D. (a) 15 minutes après la mise en solution (t_0), (b) $t_0 + 24$ heures, (c) $t_0 + 72$ heures, (d) $t_0 + 30$ jours.

L'attribution des protons amides restant n'a pas posé de problème malgré le fait que quelques pics aient bougé, sans doute à cause d'une légère différence de pH de l'échantillon.

L'analyse du temps de persistance de chaque proton amide permet de faire apparaître différentes régions au sein de la protéine CG7054 (Figure VI-21). Dans la mesure où les résidus dont les pics sont superposés sont peu nombreux (L59 et I84, Y21 et D64, A164 et R120) et tous situés dans des structures secondaires bien identifiées, le même délai de persistance du pic a été attribué pour chaque résidu.

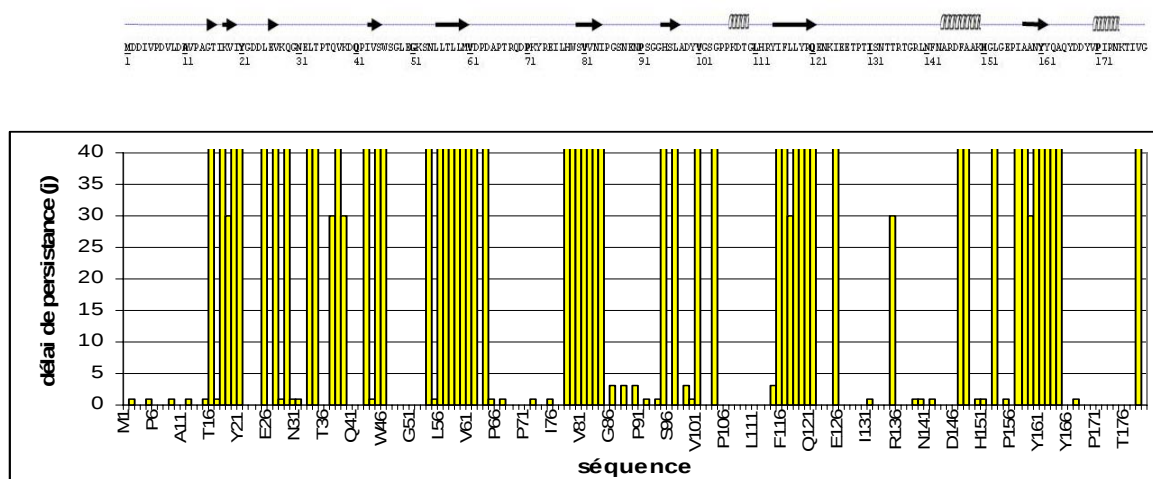


Figure VI-21 : Délai de persistance de chaque proton amide en fonction de la séquence. Les structures secondaires déterminées par Talos sont présentées au dessus de graphique.

La correspondance entre les structures secondaires et les zones où les protons amide persistent longtemps est globalement évidente. On remarque cependant une certaine hétérogénéité au niveau des 3 petits feuillets N-terminaux qui peut être expliquée par le fait que les résidus concernés se situent en surface de la protéine, en contact avec le solvant. Il est d'ailleurs étonnant d'observer des résidus en échange si lent dans cette région. On observe également des zones où les résidus sont protégés mais qui ne sont pourtant pas situées dans des structures secondaires mais dans des boucles (région 36-46, 96-106). La région 96-106 est impliquée dans la région de fixation des ligands anioniques présumée, établie par analogie avec les PEBP humaine et bovine. Enfin l'hélice C-terminale semble en échange rapide sauf pour le résidu Val178 dont le pic intense et non ambigu sur les spectres persiste plus d'un mois.

3.5. Détermination de la structure de la protéine CG7054

Le calcul de la structure de la protéine CG7054 a été réalisé avec le logiciel CYANA, version 2.0 (Guntert, 2004). Comme données source, nous avons utilisé la liste des déplacements chimiques (annexe 5), 128 contraintes d'angles obtenues par le logiciel Talos (annexe 3), et deux listes de pics NOE obtenus par « peak-picking » automatique dans NMRView sur les expériences 3D de type NOESY éditées ^{15}N et ^{13}C .

La philosophie de CYANA est de n'imposer aucune attribution des pics NOE en début de calcul. Si certaines attributions sont disponibles, le logiciel les prend en compte lors de la calibration des distances, mais ne les conserve pas pour le calcul de structure. Après différents essais, nous avons choisi d'utiliser une tolérance sur les déplacement chimiques de 0,05 ppm pour les résonances ^1H et 0,5 ppm pour les hétéronoyaux ^{13}C ou ^{15}N .

Les résultats présentés sont issus de CYANA qui n'utilise pas une fonction d'énergie potentielle réaliste. Le protocole de raffinement dans une boîte d'eau avec une fonction énergie potentielle réaliste d'ARIA sera utilisé ultérieurement pour améliorer les structures, notamment sur le plan des interactions de van der Waals et électrostatiques.

3.5.1. Résultats de l'attribution des corrélations NOE et des contraintes de distances

La majorité des pics obtenus par « peak-picking » sous NMRView a pu être attribués par CYANA, soit de manière ambiguë, soit de manière non ambiguë. Une moyenne de 74% des pics ont été attribués dans les fichiers .xpk (format NMRView). Les 26 % restants non attribués correspondent à des pics de très faible intensité (qui peuvent alors être interprétés comme du bruit expérimental) ou à des pics correspondant à des noyaux non attribués, ou enfin à la superposition de plusieurs taches de corrélation que le logiciel n'a pas réussi à gérer. Plus de 90% des pics ont été attribués de manière non ambiguë (un seul couple d'hydrogènes en interaction). Au final, et après élimination des contraintes redondantes, nous obtenons 3179 contraintes de distance.

Ces contraintes de distance se répartissent de la façon suivante (Tableau VI-6) :

Contraintes de distance	3179
Longue distance $ i-j \geq 5$	1598
Moyenne distance $1 < i-j < 5$	529
Courte distance $ i-j \leq 1$	1052
Contraintes d'angles dièdres (φ et ϕ)	128

Tableau VI-6 : Répartition des contraintes de distances obtenues expérimentalement.

La Figure VI-22 montre la distribution du nombre de contraintes impliquant chaque résidu : contraintes intrarésiduelles, à courte distance (H_i-H_j , avec $|i-j| = 1$), à moyenne distance ($1 \leq |i-j| < 5$) et à longue distance ($|i-j| \geq 5$).

Le nombre moyen de contraintes par résidu est de 18. La majorité des résidus sont impliqués dans des contraintes à moyenne et longue distance ce qui traduit la bonne structuration globale de la protéine. Par contre, sur les segments T34-Q37 et K52-L55, le nombre total de contraintes est beaucoup plus faible et aucune contrainte à longue distance n'implique le segment V101-P106. De même le nombre de contraintes longue distance sur la partie N-terminale est faible. A l'opposé, sur certains résidus comme K18, Y21, L58, I84 et P85 le nombre de contraintes est considérable (plus de 75). Ces résidus sont donc importants pour le repliement de la structure.

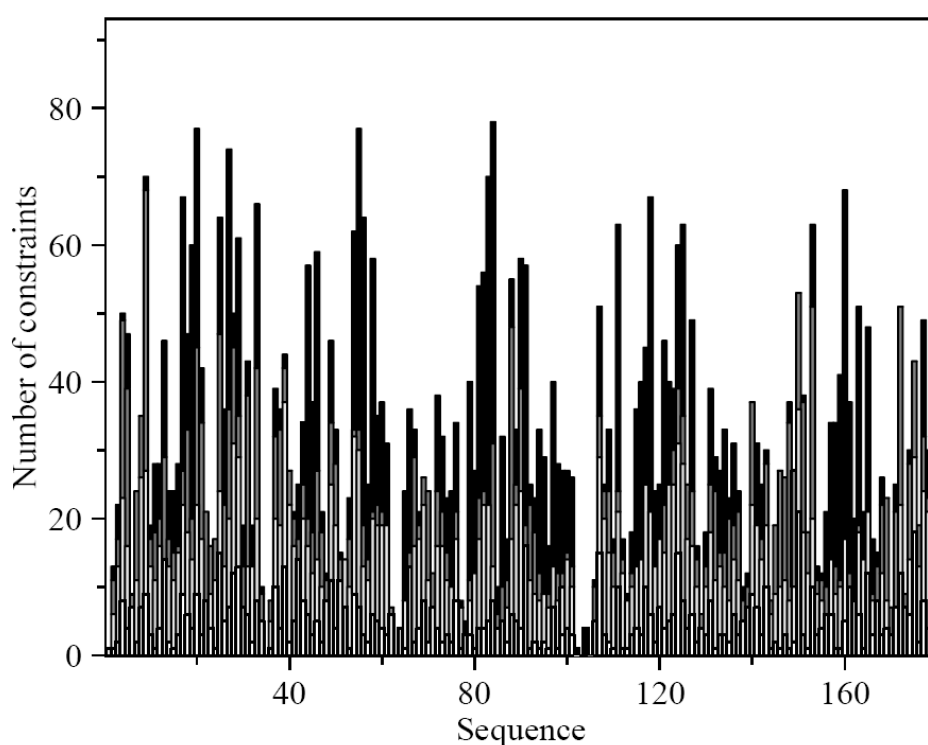


Figure VI-22 : Nombre de contraintes des distances par résidu. Contraintes intrarésiduelles (blanc), à courte distance (H^i-H^j , avec $|i-j| = 1$) (gris clair), à moyenne distance ($1 \leq |i-j| \leq 5$) (gris foncé) et à longue distance ($|i-j| \geq 5$) (noir).

Il faut souligner la présence de liaisons peptidiques cis au niveau des résidus P35, P66 et E75. Ces liaisons cis ont été détectées à partir de la valeur du déplacement chimique ^{13}C

des H α , C α et C β (Schubert et al., 2002) La proline 66 appartient au motif conservé DPDxP et elle a été retrouvée en configuration *cis* chez toutes les PEBP non bactériennes. De même, la configuration de E75 est retrouvée en *cis* chez toutes les PEBP présentant une liaison Arg-Glu 7 résidus après le proline en *cis* du motif conservé (P66), c'est à dire chez les PEBP humaine, bovine, murine, chez CEN, TFL1 et FT.

3.5.2. Analyse des structures

Deux cents structures ont été calculées à l'aide du logiciel CYANA et les 15 structures présentant les meilleures fonctions cibles ont été sélectionnées pour l'analyse.

3.5.2.a. *Qualité des structures*

Le tableau suivant résume les statistiques sur ces 15 structures.

Fonction cible finale ($\text{\AA}^2$)	12,04 $\pm$ 0,69
Ecarts quadratiques moyens (RMSD) sur les coordonnées atomiques ($\text{\AA}$)	
Sur les atomes du squelette peptidique (Par rapport à la structure moyenne)	0,65 $\pm$ 0,15
Sur l'ensemble des atomes lourds (Par rapport à la structure moyenne)	1,15 $\pm$ 0,13
Par paires de structures sur les C α ($\text{\AA}$)	0,83 $\pm$ 0,13
Par paires sur les C α des structures secondaires ($\text{\AA}$)	0,53 $\pm$ 0,1
Statistiques sur la carte de Ramachandran (en % de l'ensemble des résidus)	
Régions les plus favorables	67,7
Régions supplémentaires	29,4
Régions généreusement autorisées	3,0
Régions interdites	0,0

Tableau VI-7 : Statistiques sur les structures de la protéine CG7054

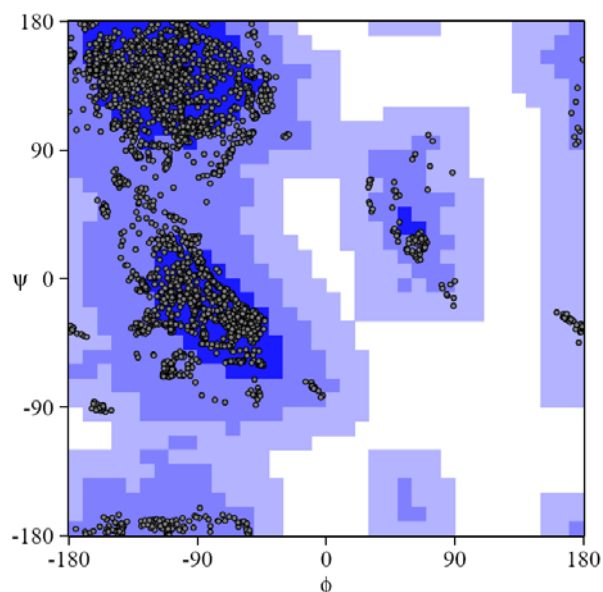


Figure VI-23 : Carte de Ramachandran pour l'ensemble des 15 meilleurs structures de CG7054 générées par CYANA. Aucun résidu ne présente de couple d'angle ϕ/ψ dans les régions interdites.

La fonction cible finale demeure assez élevée en raison de la persistance de violations des contraintes. En effet, il reste 5 violations supérieures à 0,3 Å, dont une supérieure à 0,5 Å qui est violée systématiquement sur toutes les structures. Cette contrainte se situe entre les H_δ de la leucine 56 et les H_γ de l'isoleucine 84. Deux violations d'angles apparaissent sur 5 structures parmi les 15 retenues.

Ces violations peuvent provenir de l'utilisation d'une carte NOESY-HSQC ^{13}C dont la qualité n'est pas suffisante. En effet ce spectre a été acquis au laboratoire à 600 MHz. L'intensité des pics étant faible, on y observe majoritairement des corrélations intrarésiduelles ou à courte distance. L'acquisition d'un nouveau spectre à 800 MHz est nécessaire et va être réalisée prochainement. Nous en attendons un nombre de corrélations plus important, des pics mieux résolus et l'observation de NOE longue distance en plus grand nombre.

On remarque cependant une bonne convergence des structures et les écarts quadratiques moyens sur les coordonnées atomiques (RMSD) sont satisfaisants. Une représentation en boudin de la molécule (Figure VI-29) met en évidence la bonne convergence des structures au niveau des feuillets β , mais aussi la moins bonne définition des segments N- et C-terminaux, des boucles (61-77) et (101-112) et de l'hélice centrale.

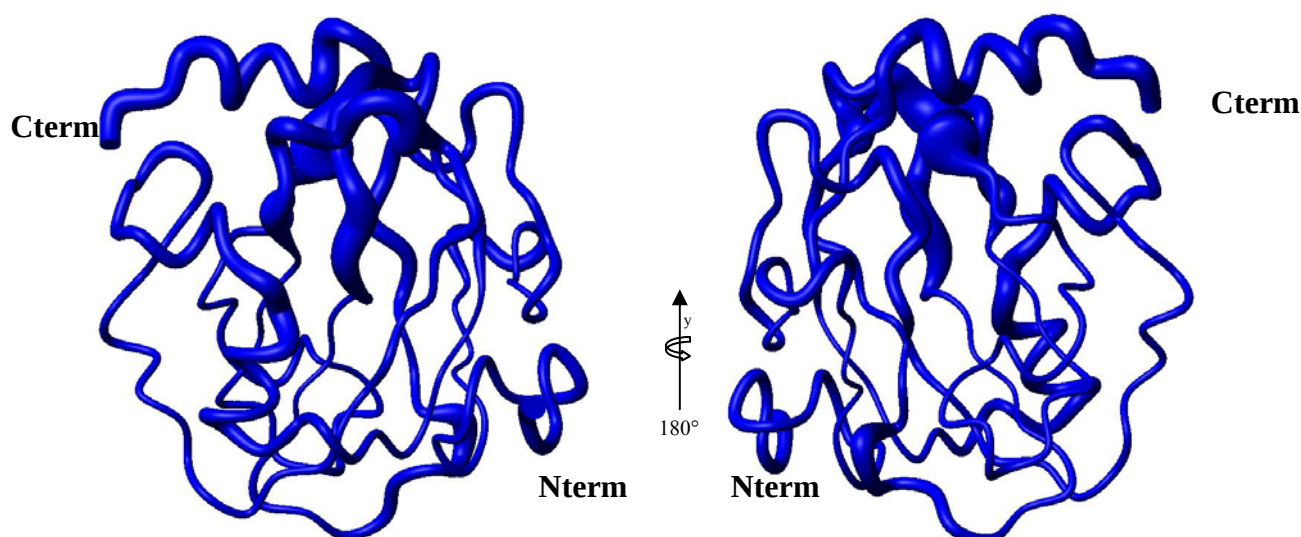


Figure VI-24: Représentation « en forme de boudin » de 15 structures 3D de la protéine CG7054. Le rayon du cylindre est proportionnel au RMSD par rapport à la structure moyenne.

Cette variabilité est liée au faible nombre de contraintes dans ces régions comme le souligne la Figure VI-22 qui donne une représentation des contraintes moyenne et longue distance le long de la séquence. Une analyse réalisée avec Procheck_NMR, montre par ailleurs une bonne convergence des angles ϕ et ψ du squelette peptidique, sauf au niveau des boucles qui présentent une plus grande variabilité (Figure VI-23). Les résidus qui présentent les variations les plus importantes des angles de torsion ϕ et ψ du squelette sont le résidus : D62, I76, L77, H78, G102, S103 et G104. Il est intéressant de remarquer que tous ces résidus entourent le site putatif de fixation des ligands anioniques.

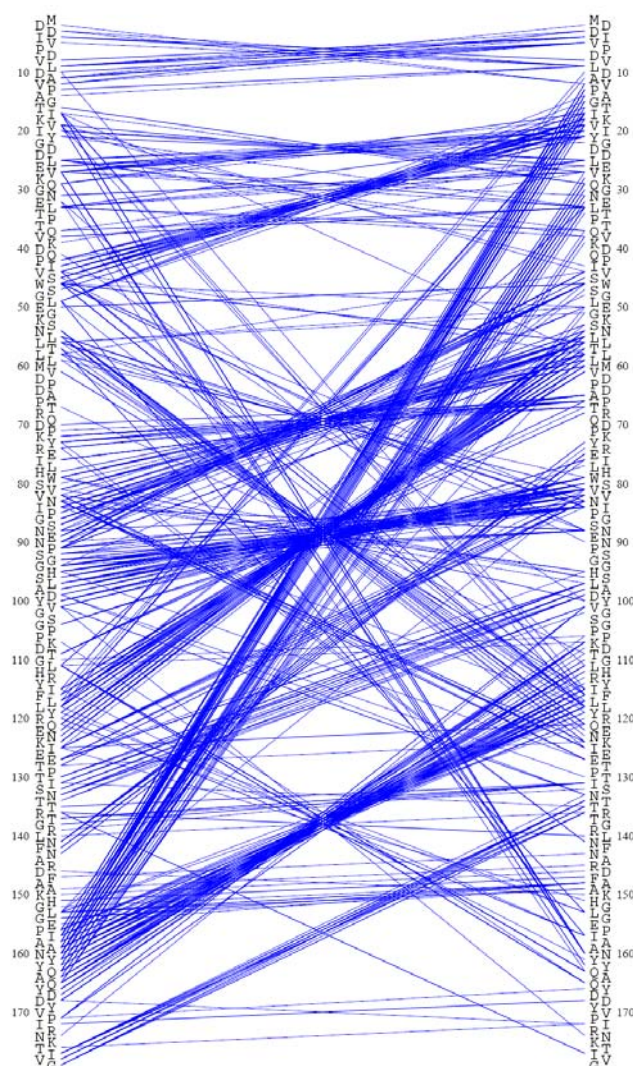


Figure VI-25: Représentation des contraintes longue distance déterminées grâce aux cartes NOESY-HSQC éditées ^{13}C et ^{15}N . Les traits dirigés du haut à gauche vers le bas à droite correspondent aux contraintes entre chaînes latérales, alors que les traits dirigés du haut à droite vers le bas à gauche correspondent aux contraintes qui impliquent les atomes du squelette peptidique.

Pour ce qui concerne les chaînes latérales, l'analyse des variances circulaires des angles χ_1 et χ_2 obtenues avec Procheck_NMR (Laskowski et al., 1996) montre une variabilité importante pour les résidus polaires et chargés. Les résidus hydrophobes sont d'une façon générale mieux définis. On remarque en particulier que les résidus D62 et V101, situés à proximité du « site actif », présentent une variance circulaire importante. H78 et Y114 sont aussi situés dans la même zone, ils présentent une variance circulaire très importante pour les angles χ_2 , mais pas pour les angles χ_1 : ce qui signifie que la position du cycle est variable, mais conserve toujours la même orientation.

3.5.2.b. Topologie et structure de la protéine CG7054

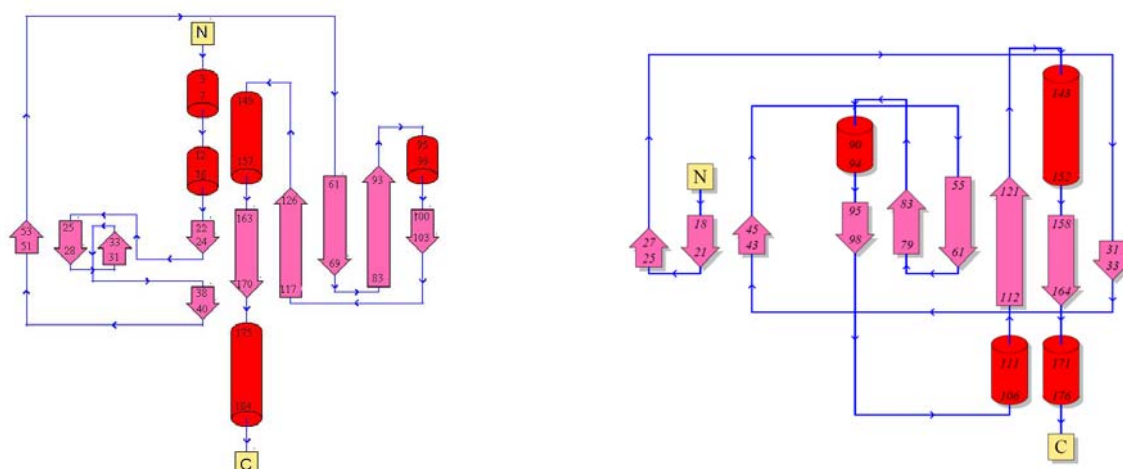


Figure VI-27 : Diagramme de topologie des protéines PEBP-1 bovine (à gauche) et CG7054 (à droite). (Figure réalisée avec l'outil PDBsum (Laskowski et al., 2005)).

La topologie de la protéine CG7054 permet de retrouver les éléments caractéristiques de la famille des PEBP : la présence de deux clés grecques avec une boucle C-terminale de type $\alpha\beta(\alpha)$. Les structures secondaires sont globalement moins étendues que sur la PEBP-1 bovine. Cela peut provenir de l'absence de prise en compte des interactions électrostatiques dans CYANA : les liaisons hydrogène ne sont pas optimisées, ce qui influe sur la formation des structures secondaires. Un raffinement préliminaire dans une boîte d'eau avec le logiciel ARIA permet d'améliorer considérablement la géométrie des liaisons hydrogène et donc la longueur des feuillet β . Comme pour d'autres PEBP, l'absence de certains feuilletts ou d'hélices 310 courts, qui sont à la limite de détection par les programmes de détermination de structures secondaires, donne l'impression d'une différence de topologie. En réalité la topologie du cœur de la protéine est identique à celle des PEBP de mammifères et de plantes, les différences n'intervenant qu'aux extrémités N- et C-terminales.

Il faut remarquer que l'extrémité N-terminale (résidus 1 à 15) adopte une conformation mal définie et une position très variable. De plus, elle n'est pas bien empilée sur la structure, probablement en raison du manque de contraintes longue distance sur cette partie (Figure VI-25), ce qui explique cette mauvaise définition. L'acquisition d'une nouvelle carte ^{13}C NOESY devrait permettre d'améliorer la base de données RMN et

remédier à ce problème. Cependant, chez les autres PEBP cette zone est généralement flexible et est présumée intervenir lors d'interaction avec d'autres protéines, comme dans le cas de la protéine TFS1 ou cette extrémité est insérée dans le site actif de la carboxypeptidase Y (Mima et al 2005). Cette dynamique particulière pourrait être à l'origine du manque de données NOE sur ce segment.

Comme pour les PEBP de mammifères, CG7054 comporte une hélice C-terminale qui n'est pas présente dans les protéines de plantes CEN (en l'état actuel des connaissances), FT et TFL1, à la protéine de levure TFS1, et aux protéines d'*E. coli* YBHB et YBCL. Cette hélice, lorsqu'elle est présente, semble être importante pour réguler l'accès au « site actif ». Elle se comporte comme un « couvercle » qui vient protéger le site actif. Là encore, sa mauvaise définition sur les structures RMN peut traduire une dynamique importante. Dans le cas de la protéine de tomate CEN (Bandfield et Brady 2000), la structure des résidus C-terminaux n'a pu être résolue, probablement en raison d'une dynamique importante de cette extrémité. Chez *E. coli*, où l'hélice C-terminale n'est pas retrouvée, une boucle N-terminale (21-27) vient se placer à proximité du site pour en diminuer l'accessibilité.

3.5.2.c. Comparaison de la protéine CG7054 et de la PEBP-1 bovine

Depuis le chapitre I, nous avons choisi de comparer l'ensemble des PEBP de structure connue à la PEBP bovine (1A44) qui nous sert ainsi de référence.

Un alignement structural de la protéine CG7054 avec la PEBP-1 bovine, effectué avec le logiciel DeepView (Guex et Peitsch, 1997), conduit à un RMSD de 1,70 Å (sur les atomes du squelette), pour 100 résidus superposés. Les différences majeures entre les deux protéines se situent au niveau des boucles : le RMSD réalisé sur les structures secondaires sans l'hélice C-terminale est de 0,53 Å et de 0,88 Å si on inclut cette hélice (Figure VI-26).

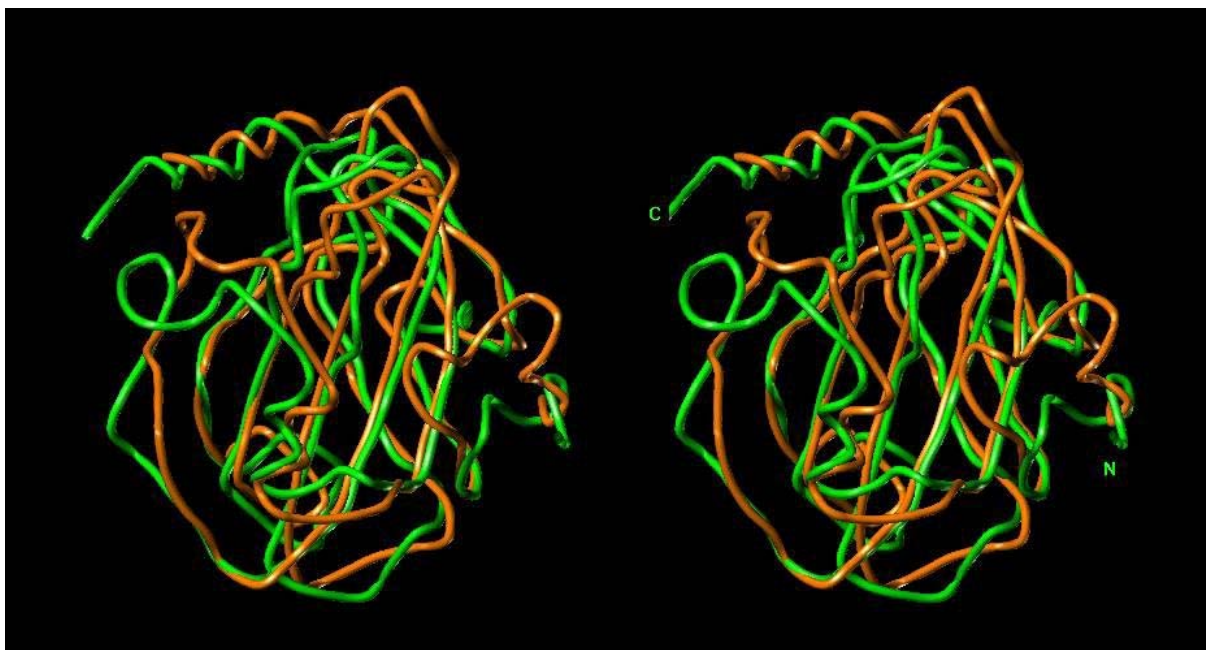
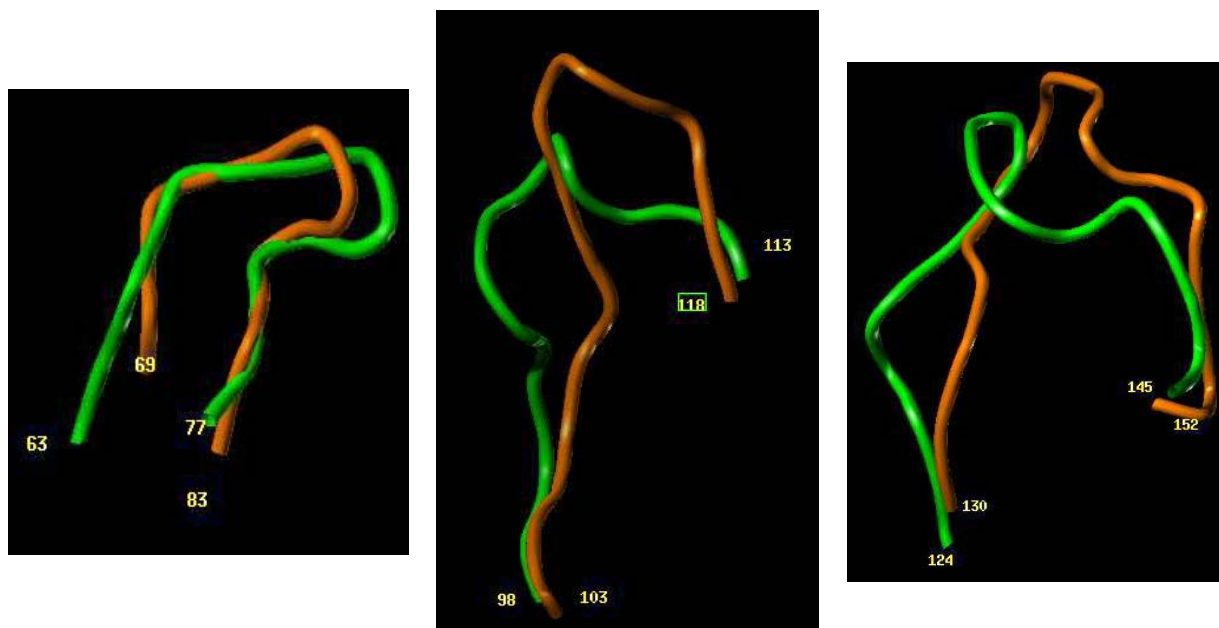


Figure VI-26 : Vue en stéréo de la superposition des protéines CG7054 (en vert) et PEBP-1 bovine (en orange). Figure réalisée avec Sybyl).

Les différences les plus importantes sont observées au niveau (Figure VI-27):

- des boucles 63-77 (CG7054) et 69-83 (PEBP-1 bovine) qui adoptent des positions différentes mais une géométrie comparable.
- des boucles 98-113 (CG7054) et 103-118 (PEBP-1 bovine) avec un turn en hélice 3_{10} pour la protéine de drosophile.
- de la conformation de la boucle 124-145 (CG7054) qui interagit avec l'hélice C-terminale et diffère fortement de celle la PEBP-1 bovine (boucle 130-152). Cette modification est corrélée à une différence de position de l'hélice C-terminale pour permettre son empilement sur le turn de la boucle (résidus 133-134) en partie en raison de la longueur de l'hélice C-terminale.
- l'extrémité N-terminale, comme nous l'avons déjà noté qui forme deux courtes hélices 3_{10} chez la protéine bovine et qui est non structurée chez CG7054.
- l'hélice C-terminale qui en plus de son décalage, présente un caractère d'hélice 3_{10} , alors que chez les PEBP de mammifères l'hélice est une hélice α .



A

B

C

Figure VI-27: Superposition des boucles. A : 63-77 (CG7054, en vert) et 69-83 (PEBP-1 bovine, en orange), B : 98-113 (CG7054, en vert) et 103-118 (PEBP-1 bovine, en orange) C : 124-145 (CG7054, en vert) et 130-152 (PEBP-1 bovine, en orange). Figure réalisée avec Sybyl.

La superposition des structures des deux protéines permet de définir les résidus du site putatif de la protéine CG7054 (Tableau VI-8) :

CG7054	D62	I76	H78	V101	G104	P106	Y114
PEBP-1 bovine	D69	W83	H85	V106	G109	P111	Y119

Tableau VI-8 : Résidus comparables du site putatif de fixation des ligands anioniques entre la protéine CG7054 et la PEBP-1 bovine.

3.5.2.d. Comparaison du site d'interaction avec celui de la PEBP bovine

On constate donc sur le plan de la séquence une bonne conservation des résidus du site. Cependant, sur le plan structural, ces résidus adoptent des positions souvent légèrement décalées. Certains sont situés au niveau des boucles citées précédemment dont la conformation et/ou la position diffèrent de celle de la PEBP bovine. La position du résidu D62/D69 (CG7054/bovine) est la plus éloignée de celle observée dans la PEBP bovine.

On peut toutefois se demander si ce résidu intervient directement dans l'interaction avec un ligand anionique dans la mesure où d'une part il est chargé négativement et d'autre part il se trouve à une distance supérieure à 5 Å du ligand. Le squelette peptidique de l'histidine 78 est correctement positionné, mais sa chaîne latérale est orientée dans une direction complètement opposée à celle du site. Les résidus hydrophobes I76 et V101, adoptent une position beaucoup plus comparable avec celle observée chez la PEBP bovine. Enfin, il serait intéressant d'observer la configuration du site actif en présence d'un ligand. En effet, les positions des chaînes latérales des résidus impliqués peuvent subir des modifications importantes lors de l'interaction.

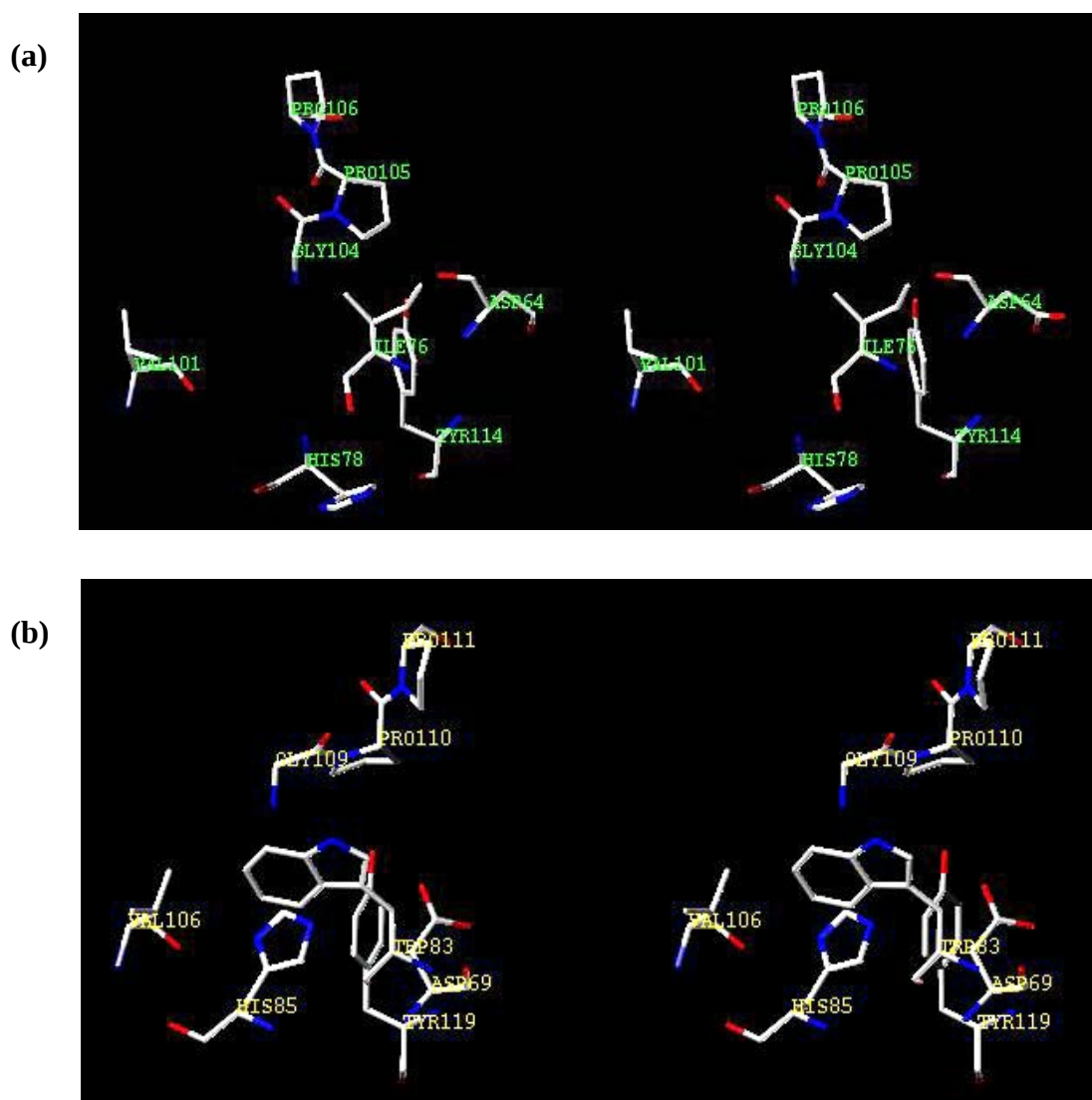


Figure VI-28 : Vue en stéréo des positions des résidus du « site actif » conservé des PEBP. (a) CG7054. (b) PEBP-1 bovine. Figure réalisée avec Sybyl.

L'hypothèse que nous avons émise lors du choix du tampon phosphate, de la possibilité de piéger un ion phosphate dans le site, semble donc mise en défaut. Ni l'analyse de la structure du site putatif, ni la variabilité des structures dans cette région ne permettent de soutenir cette hypothèse.

Bien que la géométrie du site putatif ne soit pas conservée, nous observons la formation d'une cavité en surface de la protéine, comme dans le cas des autres PEBP.

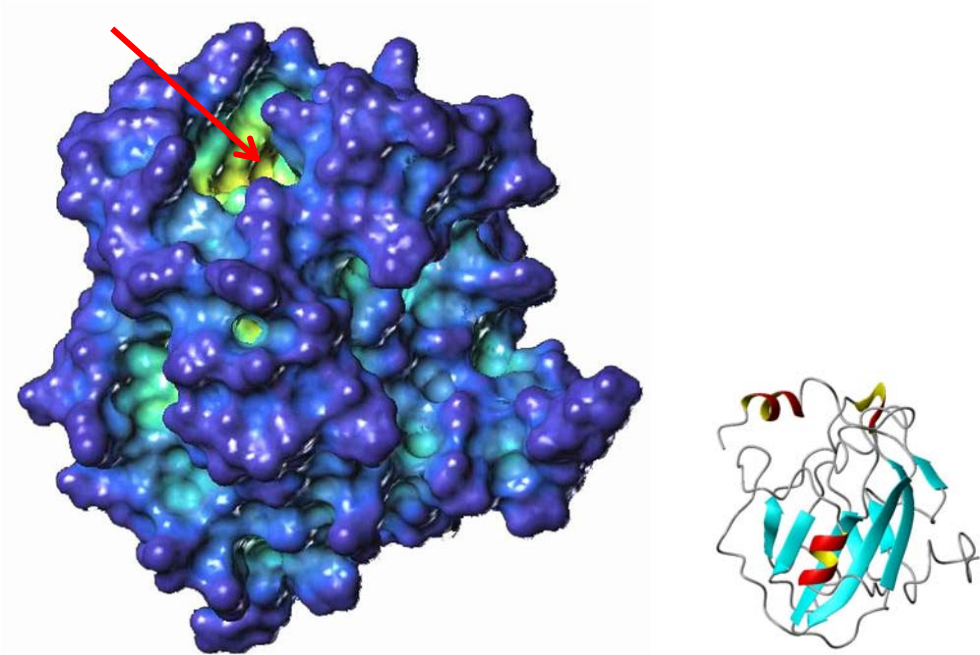


Figure VI-29 : Surface de la protéine CG7054. La flèche rouge indique la cavité conservée des PEBP. L'orientation de la surface est indiquée avec la figure en ruban. La surface a été réalisée avec Sybyl et la figure en ruban avec Molmol.

Cette cavité est plus ouverte que celle de la PEBP bovine. Les résidus situés dans la cavité présentent une accessibilité plus grande que chez la PEBP-1 bovine (Tableau VI-9) :

Résidu de la protéine CG7054 / PEBP-1 bovine	CG7054	PEBP-1 bovine
D62/D69	57	35
I76/W83	53	31
H78/H85	57	45
V101/V106	61	46
G104/G109	57	53
P106/P111	48	56
Y114/Y119	56	37

Tableau VI-9 : Accessibilité des résidus de la cavité de la protéine CG7054 et PEBP-1 bovine calculée par le logiciel MolMol (Koradi et al., 1996)

3.5.2.e. Comparaison avec les autres PEBP de structure 3D connue

La protéine CG7054 a été également comparée aux autres PEBP dont la structure est connue. La superposition des structures a été réalisée et le Tableau VI-10 résume les valeurs des RMSD obtenus le nombre de résidus superposés (résidus alignés), ainsi que les pourcentages d'identité :

Protéine	Code PDB	Nombre de résidus alignés	RMSD (Å)	Pourcentage d'identité
PEBP-1 bovine	1A44	100	1,70	43
PEBP-1 humaine	1BEH	102	1,73	43
PEBP-2 souris	1KN3	97	1,69	41
FT	1WKP	98	1,66	32
TFL1	1WKO	98	1,67	33
CEN	1QOU	90	1,64	29
TFS1	1WPX	43	2,09	22
YBCL	1FUX	63	1,76	18
YBHB	1FJJ	36	2,24	15

Tableau VI-10 : Comparaison de la structure de la protéine CG7054 avec les autres PEBP dont la structure 3D connue

Ces RMSD sont globalement corrélés aux pourcentages d'identité entre les séquences, les protéines de mammifères étant les plus proches et celles de bactéries les plus éloignées. Les valeurs obtenues sont sensiblement supérieures à celles généralement observées dans la famille des PEBP. Cette observation n'est pas due à des différences importantes de topologie : le repliement global des PEBP avec une organisation en deux clés grecques suivi d'une boucle C-terminale de type $\alpha\beta(\alpha)$ est conservé. Les boucles dont les positions varient par rapport à la PEBP-1 bovine ne sont pas conservées non plus chez les autres PEBP. L'écart observé provient essentiellement du fait que les structures secondaires de la protéine CG7054 sont pour le moment moins bien déterminées.

3.5.2.f. Conclusion

En conclusion, l'analyse de la structure de la protéine CG7054 montre qu'elle présente les caractéristiques des PEBP de structure connue et en particulier des PEBP de mammifères

et de plantes. Toutefois, en l'état actuel, la qualité de la structure reste insuffisante. Même si les RMSD sur les atomes du squelette ont des valeurs satisfaisantes, la fonction cible reste trop élevée. Un raffinement dans une boîte d'eau avec ARIA, déjà mentionné, permet d'améliorer la qualité des structures secondaires et conduit à de meilleurs diagrammes de Ramachandran. Les structures obtenues ne présentent plus qu'une seule violation, mais c'est au détriment des RMSD entre les structures qui augmentent considérablement (1,2 Å sur les atomes du squelette). Il est bien connu que la dynamique dans l'espace cartésien (ARIA) permet le franchissement de barrières énergétique qui ne peuvent être franchis dans l'espace des dièdres (CYANA), ce qui explique la diminution du nombre de violations. Toutefois, le fait que les RMSD augmentent montre que la structure n'est pas suffisamment bien définie par l'ensemble des contraintes. Il paraît donc important d'améliorer le nombre de données en particulier les NOEs longue distance par l'enregistrement d'une meilleure carte NOESY éditée ^{13}C .

Notre étude souligne également la possibilité d'une dynamique importante de certaines régions de la protéine, reflétée par une variabilité importante des structures, elle-même liée à un faible nombre de NOEs dans ces régions. Les régions concernées sont, classiquement, les extrémités N- et C-terminales mais également le site actif. Une étude de dynamique s'impose donc pour vérifier si cette dynamique est bien réelle ou simplement liée à des problèmes expérimentaux.

3.6. Dynamique de la protéine CG7054

L'attribution des déplacements chimiques de la protéine CG7054 a permis d'observer que certaines zones du squelette peptidique, notamment celles à proximité du site putatif de fixation des ligands anioniques, présentaient des corrélations dont les intensités étaient plus faibles que pour le reste de la protéine. Ce phénomène est généralement le signe d'une flexibilité importante des régions correspondantes. Afin de tester cette hypothèse, nous avons entrepris l'étude des propriétés dynamiques du squelette peptidique de la protéine CG7054.

3.6.1. Mesures des temps de relaxation longitudinale T_1 et transversale T_2

Pour déterminer les valeurs des temps de relaxation longitudinale T_1 et transversale T_2 , une série d'expériences de type HSQC a été enregistrée en faisant varier le temps de relaxation tr (cf. Matériels et Méthodes). Différentes valeurs de ce délai ont été choisies afin d'échantillonner les courbes de relaxation correspondant à T_1 et à T_2 . Pour T_1 , il s'agit de $tr = 0, 250, 500, 750, 1000, 1250, 1500, 1750, 2000, 2250, 2500, 2750$ et 3000 ms. Pour T_2 , $tr = 10, 30, 50, 70, 110, 150, 190$, et 230 ms.

La Figure VI-30 présente deux spectres utilisés pour déterminer les valeurs des temps de relaxation T_1 .

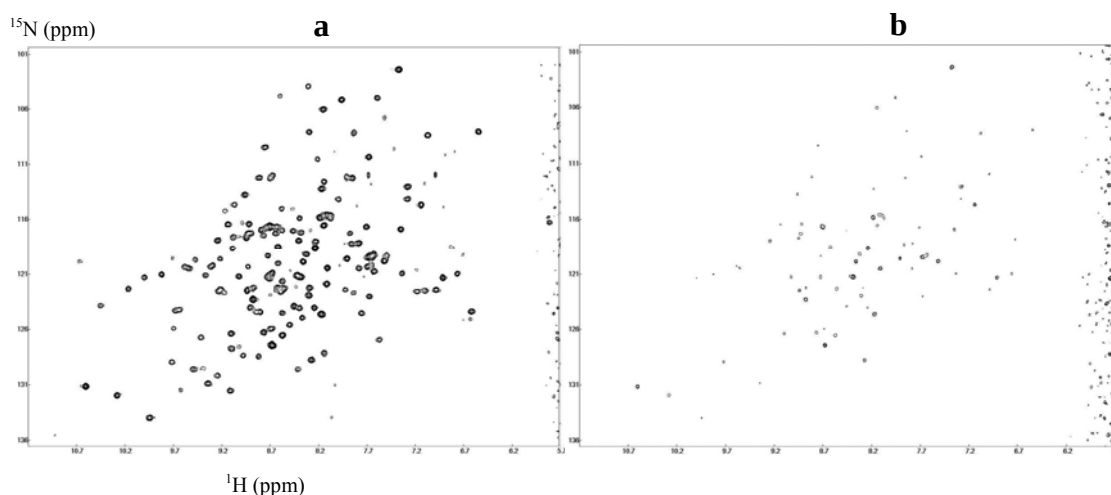


Figure VI-30: Exemples de spectres réalisés pour déterminer les temps de relaxation T_1 . Les délais tr correspondant aux spectres a et b ont pour valeurs respectives 0 et 3000 ms. L'analyse de la diminution de l'intensité de chaque pic en fonction du délai tr conduit à la détermination du temps de relaxation T_1 de chaque résidu.

Les valeurs des temps de relaxation T_1 et T_2 , ont été extraites en réalisant un ajustement de l'intensité des pics croisés à l'aide d'une fonction mono-exponentielle décroissante, comme nous l'avons décrit dans le chapitre Matériels et Méthodes. La Figure VI-31 illustre l'ajustement des intensités des pics croisés extraites des spectres de mesure de T_2 pour le résidu Leu58.

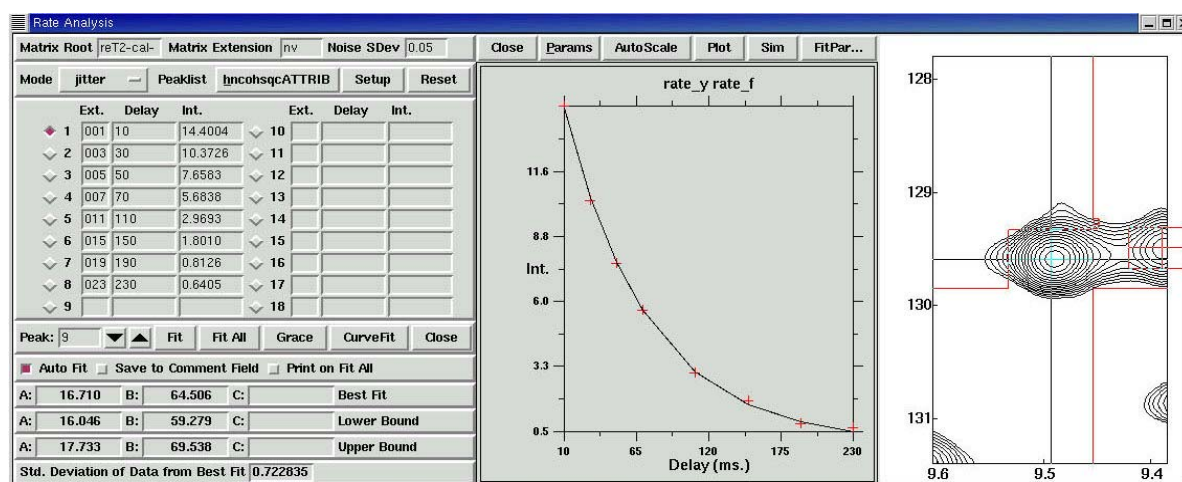
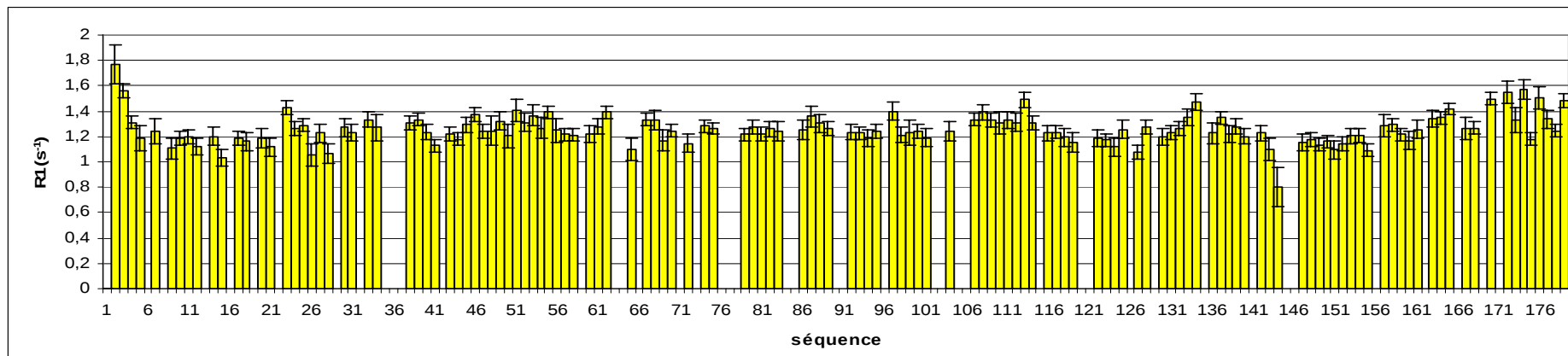


Figure VI-31 : Utilisation du module rate analysis de Nmrview pour réaliser l'ajustement monoexponentiel des intensités des pics croisés extraites de la série de spectres de mesure de T_2 de la leucine 58.

La détermination des valeurs des vitesses de relaxation longitudinale R_1 et transversale R_2 n'a été possible que pour 141 des 165 résidus non proline pour des raisons de superpositions de signaux. Les résultats sont présentés sur la Figure VI-32.

a/



b/

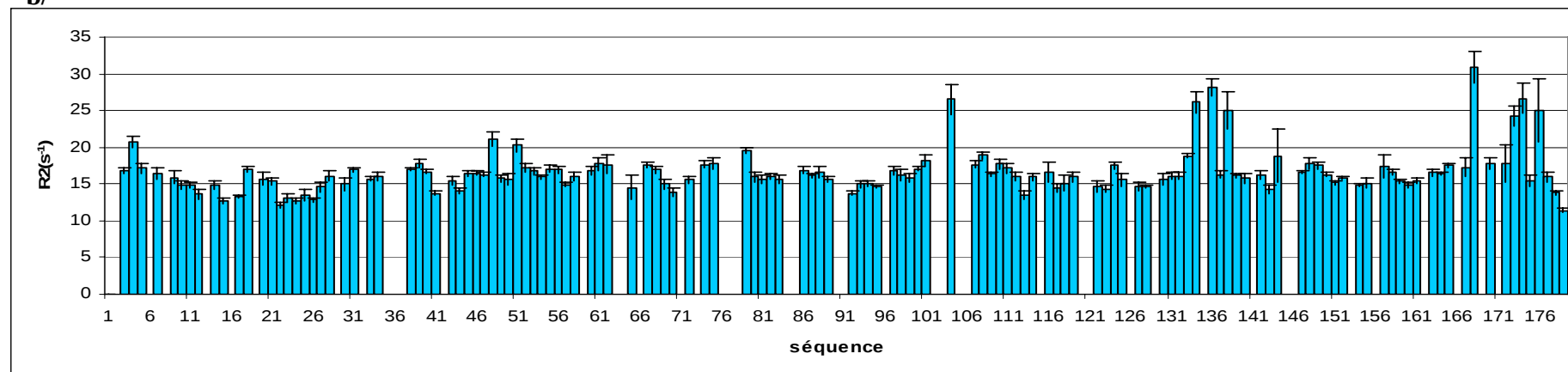
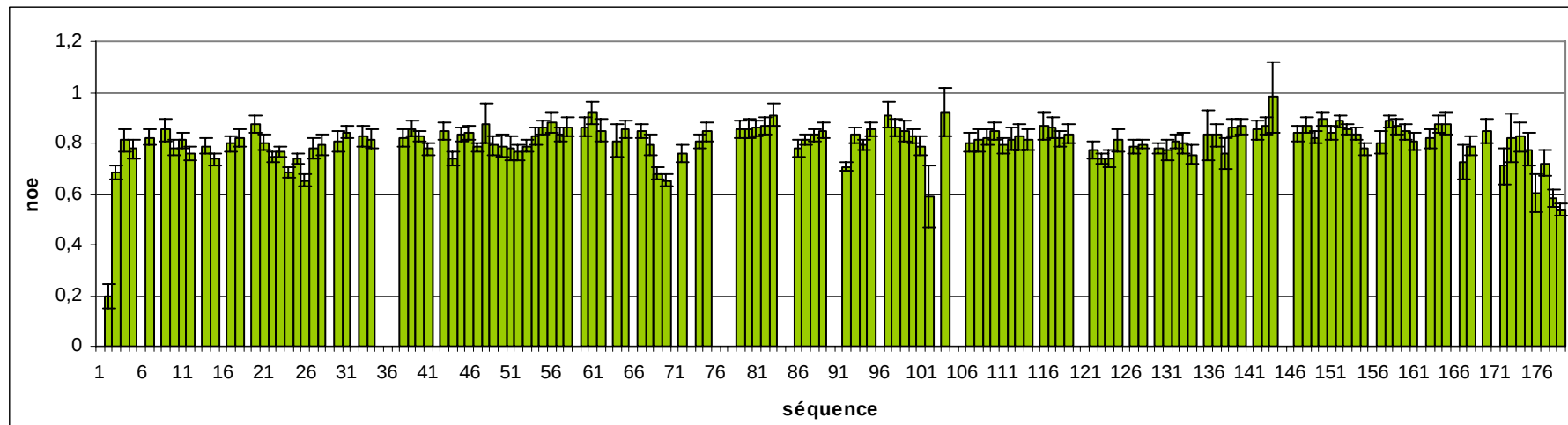


Figure VI-32 : Valeurs des vitesses de relaxation longitudinale R_1 et transversale R_2 le long de la séquence de la protéine CG7054.

a/



b/

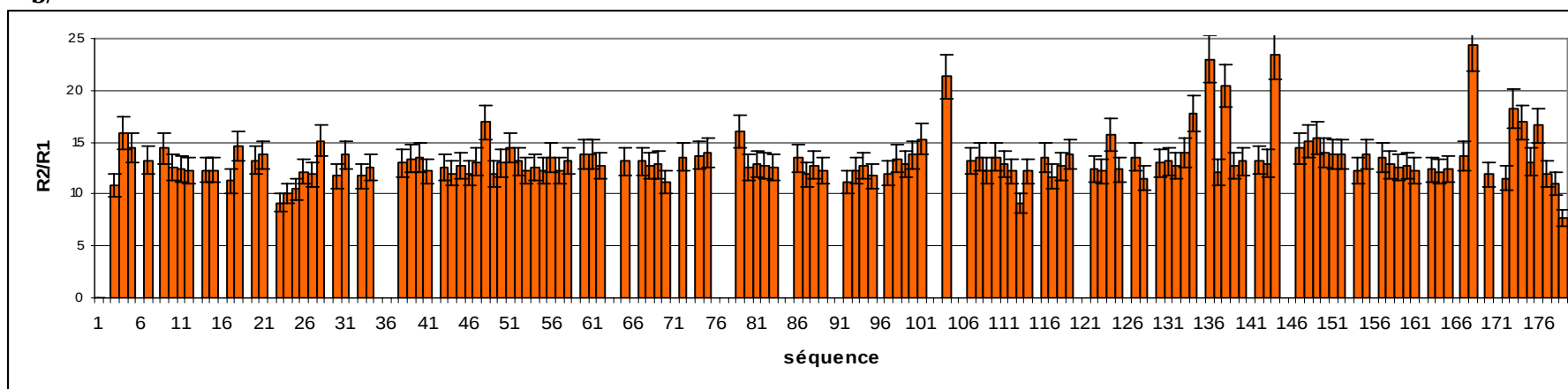


Figure VI-33 : Valeurs des NOE hétéronucléaire ^1H - ^{15}N (a) et du rapport R_2/R_1 (b) le long de la séquence de la protéine CG7054

La valeur moyenne des vitesses de relaxation R_1 des noyaux ^{15}N de la protéine CG7054 est de $1,26 \pm 0,15 \text{ s}^{-1}$. Elle est de $16,65 \pm 0,51 \text{ s}^{-1}$ pour la relaxation transversale R_2 . La valeur des R_1 n'est pas homogène tout au long de la séquence et varie de $0,81 \text{ s}^{-1}$ pour la l'alanine 144 à $1,77 \text{ s}^{-1}$ pour l'aspartate 2, alors que les valeurs de R_2 se situent entre $11,43 \text{ s}^{-1}$ pour la glycine 179 et $30,86 \text{ s}^{-1}$ pour la l'aspartate 168. La dispersion des R_2 est donc beaucoup plus importante que les R_1 . Certains résidus ont des valeurs de R_2 se distinguant particulièrement de leurs voisins : ce sont les résidus I4, G48, G51, G104, T134, R136, G138, D168, R173, N174 et T176 avec un R_2 très élevé (supérieur à 20 Hz). Cette série de résidus (R_2 élevés) est susceptible de présenter de l'échange chimique du fait d'un terme de relaxation supplémentaire. Nous n'observons pas de valeurs de R_2 significativement faibles qui traduisent généralement une dynamique interne importante dans la gamme pico-nanoseconde. Il est étonnant de ne pas observer de corrélation particulière entre les valeurs de R_1 ou R_2 et les structures secondaires.

3.6.2. Mesure de l'effet NOE hétéronucléaire ^1H - ^{15}N

Pour la mesure de l'effet NOE hétéronucléaire ^1H - ^{15}N , le délai d'irradiation des ^1H a été fixé à 5 s, avec un délai de 10 s entre 2 expériences pour assurer un bon retour à l'équilibre du système. Comme pour la détermination des R_1 et R_2 , les mesures ont été exploitées pour 141 résidus et sont récapitulées sur la Figure VI-33. Les valeurs de NOE hétéronucléaires sont toutes inférieures à la limite maximale autorisée par la théorie de 0,83, excepté pour les résidus I20, L56, V61, N83, L97, L140, N143, A144, A148, K150, G152, A158 et A164 dont l'intervalle de confiance ne contient pas la valeur 0,83. On peut expliquer ces valeurs soit par une incertitude sous estimée, soit par une saturation partielle de l'eau durant l'expérience de référence, menant à une sous-estimation de l'intensité de référence (Grzesiek et Bax, 1993). Pour les autres résidus, le NOE est compris entre 0,193 et 0,83. La valeur moyenne des NOE hétéronucléaires ^1H - ^{15}N de $0,80 \pm 0,04$ sur l'ensemble de la séquence. Cette moyenne est élevée, très proche de la limite théorique fixée à 0,83. Cela démontre que la protéine est globalement peu flexible.

3.6.3. Détermination du temps de corrélation τ_c

Il est intéressant de calculer le rapport R_2/R_1 . Pour une majorité des résidus ce rapport est situé entre 12 et 14. La région C-terminale se démarque avec des valeurs globalement inférieures. Par contre, les valeurs particulièrement élevées du R_2 des résidus I4, G48, G51, G104, T134, R136, G138, D168, R173, N174 et T176 se traduisent par un rapport R_2/R_1 remarquablement fort. Le résidu A144 présente une valeur de R_2/R_1 élevé qui provient d'une valeur de R_1 faible. Le rapport R_2/R_1 est sensible à l'anisotropie de réorientation globale de la protéine (Barbato *et al.*, 1992). Il semble donc que le tenseur de diffusion rotationnelle de la protéine CG7054 est plutôt isotrope, correspondant à une protéine relativement globulaire.

Pour déterminer le temps de corrélation de la protéine CG7054 nous avons calculé la valeur moyenne des rapports R_2/R_1 des résidus situés dans des structures secondaires et qui ne présentent pas de facteur d'échange (R_2 inférieur à 18 s^{-1}) (Figure VI-32). La valeur correspondante, $12,6 \pm 0,5 \text{ ns}$ est légèrement supérieure à la valeur attendue (de l'ordre de 11 ns) pour une protéine globulaire de 20 kDa à 20°C.

3.6.4. Interprétation des paramètres de relaxation par le formalisme de Lipari-Szabo.

Les données de relaxation obtenues expérimentalement ont été interprétées par le formalisme de Lipari-Szabo à l'aide du logiciel Modelfree en interface graphique avec Fast-Modelfree, comme décrit dans le chapitre Matériels et Méthodes. Cependant les résultats ne sont pas satisfaisants. La valeur moyenne des paramètres d'ordre S^2 : $0,96 \pm 0,02$ n'est pas réaliste et 42 résidus présentent un S^2 égal à 1,00. La valeur des S^2 de 18 résidus n'a pu être déterminée. Plusieurs hypothèses peuvent expliquer cette situation, elles concernent les données expérimentales :

- Le modèle isotrope n'est pas applicable à la protéine CG7054. La contribution de l'anisotropie se fait d'autant plus sentir que la protéine présente une taille importante, il faut donc en tenir compte dans les calculs. Dans l'état actuel des connaissances, la précision de la structure 3D disponible ne permet pas d'estimer avec précision le

tenseur anisotrope de diffusion dans la mesure où la structure présente encore des incertitudes, par exemple au niveau de l'extrémité N-terminale qui contribue à former un système asymétrique.

- La présence de mouvements internes caractérisés par des temps de l'ordre du τ_c . Le logiciel Modelfree détecte pour la protéine CG7054 des temps d'échange de plusieurs nanosecondes. Ces contributions ne sont alors plus négligeables et le formalisme de Lipari-Szabo n'est plus adapté pour analyser et extraire des paramètres dynamiques à partir des valeurs expérimentales. Cette situation est parfois rencontrée lors des mécanismes de catalyse, de transition allostériques, ou plus généralement lorsque des changements de structures sont induits par une interaction avec un ligand (Wand, 2001). Il est possible que la protéine CG7054 puisse présenter de telles caractéristiques.
- La présence en solution d'une quantité faible de dimères en échange rapide fausse les mesures de l'ensemble des paramètres dynamiques mesurés. Cette situation conduit à la mesure de vitesses de relaxation ^{15}N apparentes R_1^{app} et R_2^{app} :

$$R_1^{app} = p_M R_1^M + (1 - p_M) R_1^D$$

$$R_2^{app} = p_M R_2^M + (1 - p_M) R_2^D + R_{ex}$$

Avec p_M la population du monomère et M ou D se réfère au monomère ou au dimère respectivement.

Dans l'état actuel des connaissances nous ne pouvons donc pas conclure sur la présence de phénomènes dynamiques importants dans la région du site de fixation putatif des ligands anioniques, comme l'analyse des spectres le laissait présager. Pour répondre à cette question il faudra recalculer les paramètres dynamiques dans l'hypothèse d'un modèle anisotrope lorsque la structure finale de la protéine CG7054 sera disponible. Il faudra ensuite étudier la présence éventuelle de dimères en solution à haute concentration, par exemple par diffusion dynamique de la lumière ou par des mesures de diffusion par RMN. Enfin, il serait intéressant de réaliser un échantillonnage plus complet de la

fonction de densité spectrale en intégrant les valeurs expérimentales de R_1 , R_2 , et NOE hétéronucléaires obtenues à plusieurs valeurs de champs magnétique B_0 .

3.7. Interaction de la protéine CG7054 avec la phosphoryléthanolamine

Par définition, la famille des PEBP est caractérisée par la capacité de fixer la phosphatidyléthanolamine. Nous avons voulu déterminer si cette propriété était vérifiée pour la protéine CG7054. Pour des raisons de solubilité lors de ce premier essai d'interaction avec un ligand, nous avons utilisé la phosphoryléthanolamine dont l'interaction avec les PEBP, notamment la PEBP-1 bovine, est bien documentée (Serre *et al.*, 1998) : la phosphoryléthanolamine vient se placer au fond de la cavité considérée comme le site de fixation des ligands anioniques. La phosphoryléthanolamine correspond à la tête polaire de la phosphatidyléthanolamine :

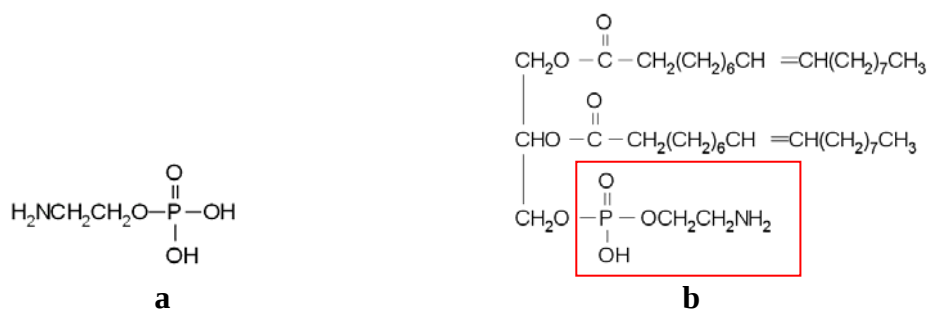


Figure VI-34 : formule de la phosphoryléthanolamine (a) et de la phosphatidyléthanolamine (b). L'absence des chaînes d'acides gras rend la phosphoryléthanolamine particulièrement soluble.

La technique classique pour détecter par RMN la fixation d'un ligand sur une protéine consiste à réaliser des expériences dites "empreintes". Cette stratégie permet de détecter les modifications de l'environnement magnétique et donc du déplacement chimique des atomes de la cible situés au voisinage du site de liaison. Le principe de ces méthodes repose sur l'observation sélective du signal provenant de la macromolécule cible marquée, dans des expériences où le ou les ligands, non marqués, sont ainsi « invisibles ». Cet effacement sélectif est possible, même en présence de concentrations très importantes de ligand, ce qui permet la détermination de constantes d'affinité très faibles.

Nous avons donc enregistré des spectres ^1H - ^{15}N HSQC en présence de quantités croissantes de phosphoryléthanolamine.

La Figure VI-35 représente une superposition des ^1H - ^{15}N HSQC obtenus.

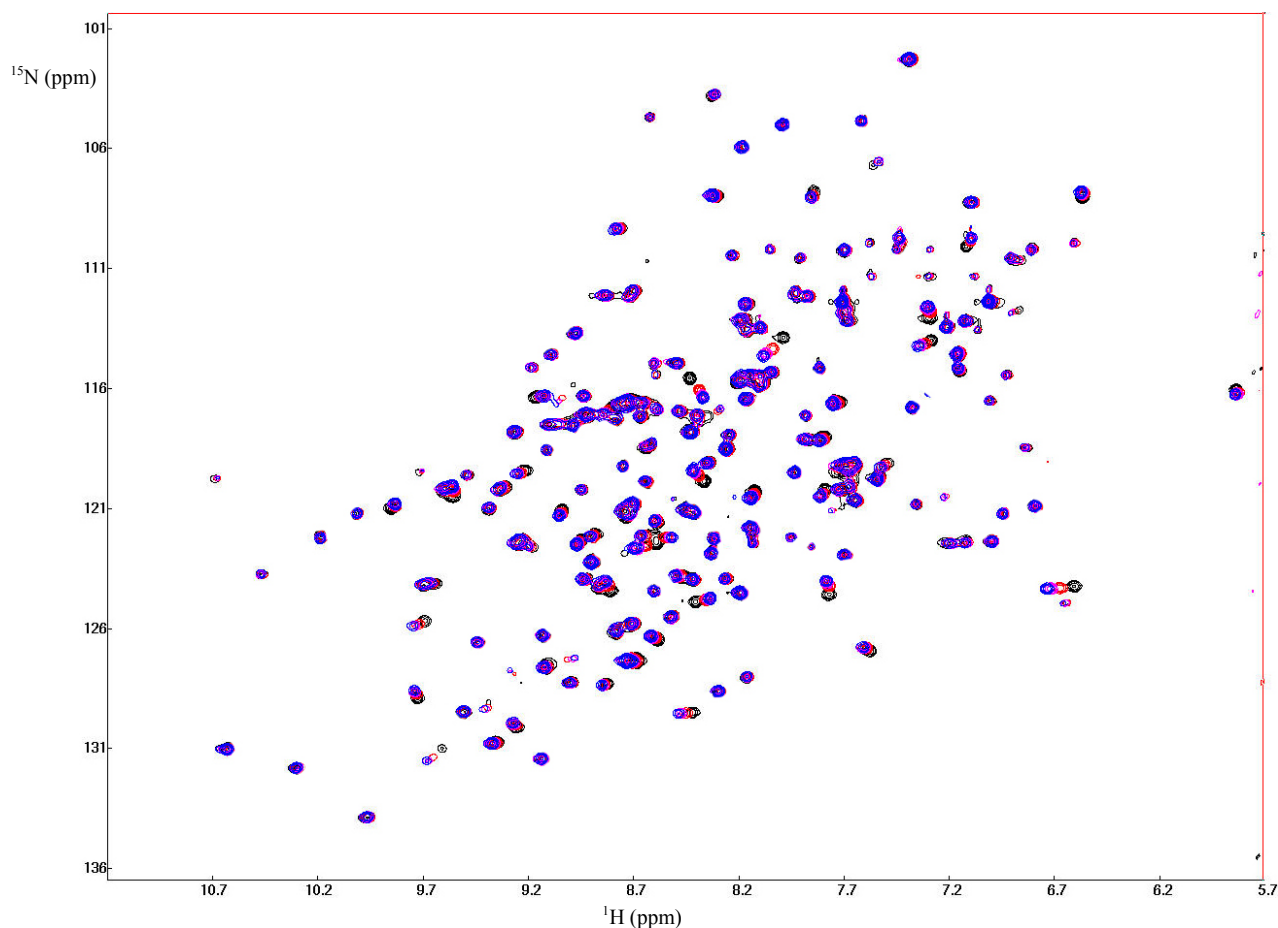


Figure VI-35 : superposition des spectres ^1H - ^{15}N HSQC acquis en présence de phosphoryléthanolamine 0mM (noir), 8 mM (rouge), 16 mM (magenta) et 32 mM (bleu)

L'observation des spectres ^1H - ^{15}N HSQC superposés permet de déterminer les acides aminés qui subissent une modification de leur environnement en présence de phosphoryléthanolamine. La variation observée des déplacements chimiques ^1H et ^{15}N peut être évaluée et normalisée en utilisant la notion de déplacement quadratique moyen :

$$\Gamma = \sqrt{\frac{(\Delta^1\text{H})^2 + (\Delta^{15}\text{N})^2}{2}}$$

où $\Delta^1\text{H}$ et $\Delta^{15}\text{N}$ sont les variations des déplacements chimiques observées ^1H et ^{15}N .

Il faut remarquer que le déplacement quadratique moyen n'est relié ni qualitativement ni quantitativement à l'interaction directe avec le ligand. En effet, nous observons les variations d'environnement à proximité des protons amide du squelette peptidique, or il est probable que l'interaction directe se réalise avec les chaînes latérales exposées à la surface du site de fixation de la protéine. Les variations de déplacement chimiques observées sont donc des conséquences secondaires de l'interaction et peuvent traduire un effet « boule de neige » sur la structure. De plus un changement minime de la position d'un cycle aromatique peut entraîner une modification plus importante de l'environnement magnétique d'un proton amide qu'une chaîne aliphatique et ainsi engendrer un déplacement plus important du pic de corrélation sur le spectre ^1H - ^{15}N HSQC. Les valeurs de Γ doivent donc être évaluées globalement.

Des essais préliminaires ont montré que la protéine précipitait si la concentration en phosphoryléthanolamine dépasse 50 mM. Si on dilue l'échantillon, il redevient soluble : la précipitation est réversible. Le pH 6,5 a été maintenu constant et nous avons montré par ailleurs que la force ionique de la solution liée à la présence de phosphoryléthanolamine ne permet pas d'expliquer la précipitation. L'effet observé semble donc spécifique : la phosphoryléthanolamine induit un changement dans la structure de la protéine CG7054 qui entraîne son agrégation puis sa précipitation réversible. Il s'agit peut être d'un effet similaire à ce qui a été observé pour la PEBP bovine qui oligomérisé en présence de fortes concentrations de phosphatidyléthanolamine et forme un gel (Bernier *et al.*, 1986).

Il faut rappeler que les PEBP présentent une affinité pour les groupements phosphates. La présence du tampon phosphate de potassium utilisé pour l'enregistrement des spectres nécessaires à l'attribution et à la détermination de la structure est donc susceptible de gêner l'interaction en provoquant un phénomène de compétition. Nous avons souhaité pour ce premier essai d'interaction conserver les conditions expérimentales de l'étude structurale. Pour la suite nous envisageons d'utiliser un tampon sans groupement phosphate.

Dans notre cas, nous avons observé que 18 corrélations étaient modifiées en présence de phosphoryléthanolamine (Figure VI-36):

résidu	déplacement quadratique moyen Γ observé en présence de 32 mM de phosphoryléthanolamine
3	0,619
4	0,547
5	0,401
7	0,11
9	0,17
61	0,094
75	0,225
97	0,063
101	0,117
104	0,368
107	0,06
116	0,172
125	0,122
143	0,114
144	0,58
145	0,12
147	0,281
148	0,144
149	0,189
150	0,319

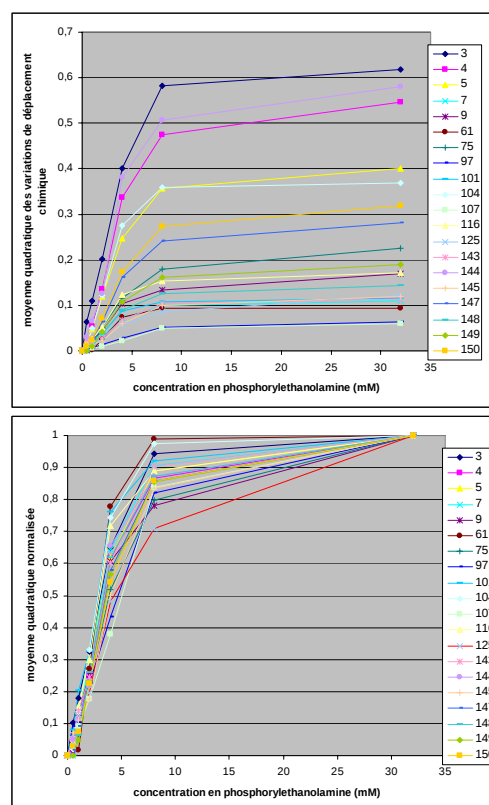


Figure VI-36 : Evaluation du déplacement des pics sur le spectre ^1H - ^{15}N HSQC en fonction de la quantité de phosphoryléthanolamine. (a) Moyenne quadratique observée pour chaque pic. (b) moyenne quadratique normalisée par rapport au déplacement maximum du pic.

Ces résidus sont situés :

- à l'extrémité N-terminale (D3, I4, V5, D7 et L9).
- au niveau du site d'interaction présumé (V101, G104, K107) ou à proximité (V61, E75).

- au niveau d'une hélice située en surface de la protéine (N143, A144, R145, F147, A148, A149, K150 et H151) ou à proximité (I125).

Les résidus restant (L97, F116) sont plus dispersés dans la structure.

Si on observe la variation des déplacements chimiques des pics en fonction de la concentration en phosphoryléthanolamine, on constate que l'ensemble des pics réagit de la même manière, avec un demi équilibre atteint pour des concentrations entre 5 et 10 mM de phosphoryléthanolamine. Si cette concentration est forte, il ne faut pas oublier la présence d'un tampon phosphate 50 mM qui entre probablement en compétition avec la phosphoryléthanolamine pour se fixer sur la protéine CG7054. Il serait donc intéressant de mesurer l'affinité en absence de phosphate. Les courbes obtenues sont en première approximation monophasiques (Figure VI-36).

L'effet de l'ajout de phosphoryléthanolamine sur la structure de la protéine CG7054 n'est pas dû à une variation de pH ni à un effet salin. En effet lorsqu'on ajoute du sel dans la même gamme de concentration que la phosphoryléthanolamine, on n'observe pas de variation de la position des pics, sauf au niveau de l'extrémité N-terminale (résidus 2 à 9) non structurée. L'effet observé avec la phosphoryléthanolamine est donc spécifique sauf pour les résidus N-terminaux qui subissent un effet salin non spécifique.

Les résidus impliqués dans la zone de fixation putative des ligands anioniques des autres PEBP sont concernés par l'interaction (V101, G104, K107). La valine 61 se situe à proximité de l'aspartate 62, et le glutamate 75 à côté de l'isoleucine 76 et de l'histidine 78. Ceci confirme que le site de liaison des ligands anioniques est conservé et situé au niveau de la cavité observée sur la protéine CG7054. Nous avons vu sur la structure de la protéine CG7054 que la géométrie de ce site est moins conservée par rapport aux autres PEBP de structure 3D connue. Toutefois, comme nous l'avons déjà remarqué, les chaînes latérales de ces résidus peuvent se réorienter lors de l'interaction avec un ligand.

On observe par ailleurs une deuxième zone impliquée dans l'interaction : il s'agit de l'hélice H1 (A144 à K150) située en surface de la protéine CG7054. La présence d'un second site de fixation de ligand a été présumée depuis longtemps, mais jamais confirmée au niveau structural (Schoentgen *et al.*, 1992). L'analyse de la structure de la protéine CG7054 met en évidence l'existence d'une seconde cavité située au dessous de la

première, le long de cette hélice (Figure VI-37). Cette cavité est bordée par des résidus cationiques (Arg139, Lys150 et His 151) qui peuvent servir de point d'ancrage pour le groupement phosphate.

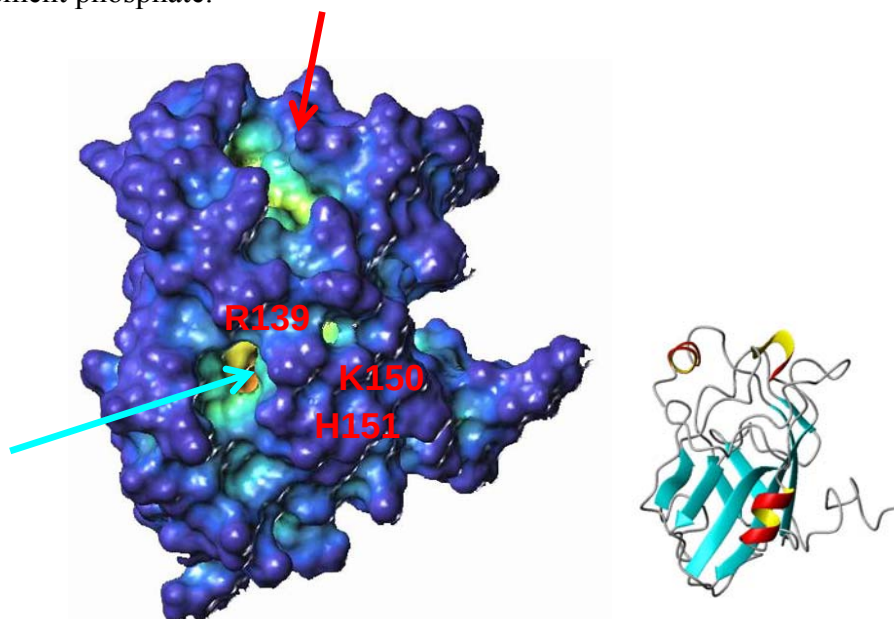


Figure VI-37 : Surface de la protéine CG7054. La flèche rouge indique la cavité conservée des PEBP. La flèche bleue indique une seconde cavité, située sous l'hélice H1. Les résidus susceptibles d'interagir avec le groupement phosphate sont indiqués. L'orientation de la surface est indiquée avec la figure en ruban. La surface a été réalisée avec Sybyl, la figure en ruban avec Molmol.

Cette hélice est particulièrement importante pour les PEBP-1 humaine et bovine car la sérine 153, qui assure la régulation du rôle de ces protéines dans les voies de transduction du signal, est située sur cette hélice. Cependant cette serine n'est pas conservée sur la protéine CG7054, ce qui laisse présumer des différences fonctionnelles avec les PEBP-1 humaine et bovine.

Cette étude devra bien sûr être complétée par l'étude d'autres ligands et en particulier la phosphatidyléthanolamine, ligand 'naturel' de la PEBP bovine. Toutefois sans préjuger du résultat, l'examen des potentiels de lipophilie à la surface de la protéine CG7054 montre que cette surface est essentiellement hydrophile alors que la PEBP bovine fait apparaître une large cavité hydrophobe juste en dessous du site d'interaction qui pourrait servir de zone d'interaction avec une chaîne lipidique (Figure VI-38).

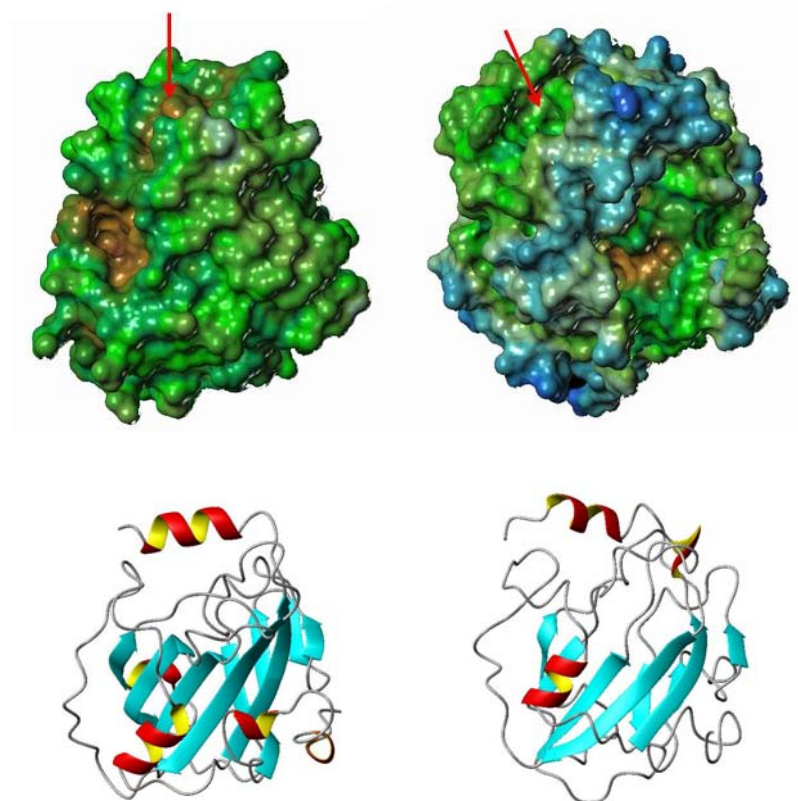


Figure VI-38 : Surfaces colorées en fonction de la lipophilie (en brun : les zones hydrophobes, en bleu : les zones hydrophiles et en vert : les zones intermédiaires). A gauche : PEBP-1 bovine, à droite : CG7054. Les flèches rouges indiquent la cavité conservée des PEBP. L'orientation des surfaces est indiquée avec les figures en ruban
Figure obtenue avec le module MOLCAD de SYBYL.

4. Conclusion

Dans ce travail, nous avons exprimé et purifié la protéine CG7054 (marquée ou non) en quantité importante. L'étude de sa stabilité par RMN a montré que la protéine produite est soluble, repliée et parfaitement stable. Il s'agit donc d'un excellent candidat que nous avons retenu pour une étude structurale. Après acquisition des spectres RMN hétéronucléaires, nous avons réalisé l'attribution des noyaux ^1H , ^{15}N et ^{13}C du squelette peptidique et des chaînes latérales de la protéine CG7054. Sa structure a ensuite été déterminée, elle présente les caractéristiques globales des PEBP : elle est constituée de feuillet β organisés en deux clés grecques et d'une boucle de type $\alpha\beta\alpha$. Cependant la topologie de plusieurs boucles n'est pas conservée. La cavité constituant le « site actif » conservé des PEBP est présente mais la configuration des chaînes latérales des résidus qui

la composent varie sensiblement par rapport aux autres PEBP. Par ailleurs l'étude de la dynamique du squelette peptidique de la protéine CG7054 a été initiée mais l'utilisation du formalisme de Lipari-Szabo n'a pas donné de résultats satisfaisants pour décrire cette dynamique. Enfin des essais d'interaction avec la phosphoryléthanolamine ont été réalisés et se sont avérés positifs. Nous montrons l'implication du « site actif » dans cette interaction mais également la présence d'une seconde zone de fixation de la phosphoryléthanolamine.

Conclusions et perspectives

VII. Conclusions, perspectives

L'objectif de ma thèse a été de réaliser l'expression, la purification et l'étude structurale par RMN de PEBP de drosophile. Ce travail s'intègre dans le cadre d'un projet réunissant plusieurs équipes du Centre de Biophysique Moléculaire et vise à approfondir les connaissances sur les relations structure / fonction au sein de cette famille de protéine. Ce travail constitue la première étude réalisée par RMN sur la famille des PEBP, il ouvre la porte à l'étude de la structure, de la dynamique, mais aussi des interactions de ces protéines avec des ligands anioniques ou des protéines.

Durant cette thèse trois PEBP de drosophile (CG18594, CG10298 et CG7054) ont été exprimées chez *Escherichia coli* et purifiées en grande quantité dans le but de réaliser des études biophysiques et structurales et d'obtenir des anticorps spécifiques. Le rôle biologique de ces protéines est en cours d'étude par nos collaborateurs qui disposent de mutants pour CG18594 et CG7054. Parmi les trois protéines produites et purifiées, une seule (CG7054) s'est avérée compatible avec l'étude de sa structure 3D par RMN hétéronucléaire.

La protéine CG18594 n'est pas stable à haute concentration et à 20°C : elle se dénature et a tendance à s'agréger. La protéine est stable à 5°C mais les spectres enregistrés ne sont pas utilisables dans le cadre de la résolution d'une structure 3D. L'étude RMN est donc compromise. Cependant, elle est en cours d'essai de cristallisation et les premiers cristaux ont été obtenus. Les études par dichroïsme circulaire ont par ailleurs montré un processus de dénaturation non conventionnel de la protéine CG18594 avec persistance de structures secondaires à haute température ou à pH très acide. L'étude par RMN à pH 2,3 de la protéine montre l'absence de structure tertiaire : nous avons affaire à une forme « molten globule ». Il serait intéressant pour la suite des travaux de déterminer si cette propriété présente une pertinence au niveau biologique ou s'il s'agit d'une curiosité de physicochimiste.

L'étude de la stabilité de la protéine CG10298 par RMN a montré que sa structure évoluait lentement au cours du temps. Dans la mesure où nous avons obtenu la protéine CG7054 à la même période et que cette dernière s'est avérée parfaitement stable, nous

l'avons choisie pour réaliser l'étude structurale. Néanmoins les spectres obtenus pour la protéine CG10298 sont prometteurs et la durée de stabilité de cette protéine (environ 10 jours sans aucune évolution) n'est pas incompatible avec une étude structurale. En effet, il est possible de réaliser rapidement l'acquisition de l'ensemble des spectres nécessaires à l'attribution et à la résolution de la structure sur un échantillon concentré (la protéine est soluble jusqu'à 2 mM) avec un spectromètre 800MHz équipé d'une cryosonde. Le rendement de production de la protéine étant très bon, il est possible d'envisager l'utilisation de plusieurs échantillons. Enfin, il est probable que l'optimisation des conditions de l'échantillon (pH, sel, etc.) puisse résoudre ce problème de stabilité et permette une étude sereine par RMN. Par ailleurs, des anticorps ont été produits à partir de la protéine purifiée, ils ont été utilisés par l'équipe de Daniel Locker pour établir le mode d'expression de cette protéine. Enfin, nous avons montré que cette PEBP de drosophile est la plus proche de la protéine bovine. Ceci nous a permis de construire un modèle par homologie. Même si ce modèle n'est pas aussi précis qu'une structure expérimentale, il permettra une étude des propriétés de surface (potentiels électrostatiques et de lipophilie) qui pourront donner des pistes quant aux modes d'interaction de la protéine avec des partenaires, une fois que ceux-ci seront connus, en attendant de disposer d'une structure RMN.

La solubilité et la stabilité de la protéine CG7054 en font un bon candidat pour une étude par RMN. L'attribution de ses noyaux ^1H , ^{13}C et ^{15}N a été réalisée et sa structure 3D a été déterminée mais comporte encore des incertitudes. Cette structure présente les caractéristiques globales des PEBP : elle est constituée de feuillets β organisés en deux clés grecques et d'une extrémité C-terminale de type $\alpha\beta\alpha$. Cependant la structure de plusieurs boucles n'est pas conservée. La cavité constituant le « site actif » conservé des PEBP est présente mais la configuration des chaînes latérales des résidus qui la composent varie sensiblement par rapport aux autres PEBP. Par ailleurs l'étude de la dynamique du squelette peptidique de la protéine CG7054 a été initiée, mais avec les connaissances actuelles l'utilisation du formalisme de Lipari-Szabo n'a pas donné de résultats satisfaisants. Des études d'interaction avec la phosphoryléthanolamine ont été réalisées et se sont avérées positives. Nous montrons l'implication du « site actif » dans cette interaction mais également la présence d'un second site de fixation de la phosphoryléthanolamine. Concernant la suite des travaux, nous allons prochainement

affiner la structure de la protéine CG7054 en intégrant de nouvelles contraintes expérimentales provenant de l'acquisition d'un nouveau NOESY édité ^{13}C enregistré à 800 MHz. Par ailleurs, nous voulons réaliser rapidement de nouvelles études d'interactions avec des ligands anioniques en présence d'un tampon ne comportant pas de groupement phosphate. Dans ces conditions, nous souhaitons tester de nouveau la phosphoryléthanolamine, afin d'observer un éventuel changement de l'affinité de la protéine, mais aussi la phosphatidyléthanolamine, le GTP, ainsi que la morphine, ligands 'naturels' de la PEBP bovine.

A plus long terme, l'objectif est de continuer la caractérisation biologique, physicochimique et structurale pour l'ensemble des PEBP de drosophile. La comparaison de ces 7 protéines permettra d'entrevoir les relations structure / fonction fines pour cette famille de protéines et, nous l'espérons, permettra de mieux cerner le rôle et la fonction des PEBP chez l'Homme. Dans ce cadre, le clonage des sections codantes des 4 gènes restants est en cours de réalisation. Concernant les perspectives au niveau structural, nous souhaitons déterminer la structure de ces PEBP seules ou en interaction avec des ligands anioniques ou des partenaires protéiques, une fois que ceux-ci auront été déterminés. Nous envisageons une étude par RMN ou par cristallographie, voire les deux, afin de tirer de chaque technique le maximum d'informations.

La connaissance d'une structure sans fonction qui lui est associée ne présente que peu d'intérêt. Aussi, nos collaborateurs qui disposent de lignées mutantes pour les protéines CG7054 et CG18594, souhaitent disposer de mutants de l'ensemble des PEBP de drosophile. S'il est difficile d'obtenir des mutants, l'équipe de Daniel Locker envisage d'avoir recours à la méthode alternative des petits ARN interférents pour bloquer l'expression des gènes codant les PEBP. Le phénotype lié à l'absence d'expression d'un gène est en effet souvent indicateur du rôle biologique de la protéine correspondante. Par ailleurs, la possibilité d'étudier les interactions entre les lignées mutantes permet de caractériser les phénomènes de compensation qui interviennent au sein d'une même famille de protéines. Un autre aspect développé par nos collaborateurs concerne les phénotypes associés à la surexpression des PEBP, par exemple au niveau des voies de signalisation.

Concernant la recherche de partenaires, nous souhaitons d'une part identifier les ligands anioniques qui peuvent être fixés par les PEBP de drosophile, et d'autre part identifier les ligands protéiques qui peuvent interagir avec ces protéines. Concernant les ligands anioniques, la bonne caractérisation de la PEBP-1 bovine nous oriente vers les phospholipides, les nucléotides, les hormones et les dérivés de morphine. Un projet visant à caractériser ces interactions par Biacore avant une étude RMN est en cours de réflexion. Concernant la recherche de partenaires protéiques, l'étude des protéines de drosophile est facilitée par la présence de plusieurs bases de données d'interaction (comme <http://www.thebiogrid.org/>) qui regroupent l'ensemble des résultats réalisés à grande échelle d'expériences comme le double hybride bactérien. La disponibilité d'anticorps spécifiques est un point clé pour mettre en œuvre la recherche de partenaires protéiques au laboratoire. En effet, cela ouvre la voie à un grand nombre de techniques comme l'immunoprécipitation, la formation de colonnes d'affinité etc. Nous possédons actuellement des anticorps pour les 3 protéines étudiées durant ma thèse.

Références bibliographiques

Références Bibliographiques

- Abragam, A. (1961). "Principles of Nuclear Magnetism." Oxford University Press.
- Ahn, J. H., Miller, D., Winter, V. J., Banfield, M. J., Lee, J. H., Yoo, S. Y., Brady, R. L. and Weigel, D. (2005). "An external loop confers functional specificity of FT and TFL1." To be Published.
- Ahn, J. H., Miller, D., Winter, V. J., Banfield, M. J., Lee, J. H., Yoo, S. Y., Henz, S. R., Brady, R. L. and Weigel, D. (2006). "A divergent external loop confers antagonistic activity on floral regulators FT and TFL1." Embo J **25**(3): 605-14.
- Banfield, M. J., Barker, J. J., Perry, A. C. and Brady, R. L. (1998). "Function from structure? The crystal structure of human phosphatidylethanolamine-binding protein suggests a role in membrane signal transduction." Structure **6**(10): 1245-54.
- Banfield, M. J. and Brady, R. L. (2000). "The structure of Antirrhinum centroradialis protein (CEN) suggests a role as a kinase regulator." J Mol Biol **297**(5): 1159-70.
- Barbato, G., Ikura, M., Kay, L. E., Pastor, R. W. and Bax, A. (1992). "Backbone dynamics of calmodulin studied by ¹⁵N relaxation using inverse detected two-dimensional NMR spectroscopy: the central helix is flexible." Biochemistry **31**(23): 5269-78.
- Bax, A. and Ikura, M. (1991). "An efficient 3D NMR technique for correlating the proton and ¹⁵N backbone amide resonances with the alpha-carbon of the preceding residue in uniformly ¹⁵N/¹³C enriched proteins." J Biomol NMR **1**(1): 99-104.
- Bernier, I. and Jolles, P. (1984). "Purification and characterization of a basic 23 kDa cytosolic protein from bovine brain." Biochim Biophys Acta **790**(2): 174-81.
- Bernier, I., Tresca, J. P. and Jolles, P. (1986). "Ligand-binding studies with a 23 kDa protein purified from bovine brain cytosol." Biochim Biophys Acta **871**(1): 19-23.
- Bradley, D., Carpenter, R., Copsey, L., Vincent, C., Rothstein, S. and Coen, E. (1996). "Control of inflorescence architecture in Antirrhinum." Nature **379**(6568): 791-7.
- Brooks, Bruccoleri, R. E., Olafson, B. D., States, D. J., Swaminathan, S. and Karplus, M. (1983). "CHARMM: A program for macromolecular energy, minimization, and dynamics calculations." J.Comp.Chem. **4**: 187-217.
- Brunger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W., Jiang, J. S., Kuszewski, J., Nilges, M., Pannu, N. S., Read, R. J., Rice, L. M., Simonson, T. and Warren, G. L. (1998). "Crystallography & NMR system: A new software suite for macromolecular structure determination." Acta Crystallogr D Biol Crystallogr **54**(Pt 5): 905-21.
- Brunger, A. T. and Nilges, M. (1992). "Xplor manual version 3.0."
- Bucquoy, S., Jolles, P. and Schoentgen, F. (1994). "Relationships between molecular interactions (nucleotides, lipids and proteins) and structural features of the bovine brain 21-kDa protein." Eur J Biochem **225**(3): 1203-10.
- Caesar, R. and Blomberg, A. (2004). "The stress-induced Tfs1p requires NatB-mediated acetylation to inhibit carboxypeptidase Y and to regulate the protein kinase A pathway." J Biol Chem **279**(37): 38532-43.
- Chatterjee, D., Bai, Y., Wang, Z., Beach, S., Mott, S., Roy, R., Braastad, C., Sun, Y., Mukhopadhyay, A., Aggarwal, B. B., Darnowski, J., Pantazis, P., Wyche, J., Fu, Z., Kitagawa, Y., Keller, E. T., Sedivy, J. M. and Yeung, K. C. (2004). "RKIP sensitizes prostate and breast cancer cells to drug-induced apoptosis." J Biol Chem **279**(17): 17515-23.

- Chautard, H., Jacquet, M., Schoentgen, F., Bureau, N. and Benedetti, H. (2004). "Tfs1p, a member of the PEBP family, inhibits the Ira2p but not the Ira1p Ras GTPase-activating protein in *Saccharomyces cerevisiae*." *Eukaryot Cell* **3**(2): 459-70.
- Coggins, B. E. and Zhou, P. (2003). "PACES: Protein sequential assignment by computer-assisted exhaustive search." *J Biomol NMR* **26**(2): 93-111.
- Corbit, K. C., Trakul, N., Eves, E. M., Diaz, B., Marshall, M. and Rosner, M. R. (2003). "Activation of Raf-1 signaling by protein kinase C through a mechanism involving Raf kinase inhibitory protein." *J Biol Chem* **278**(15): 13061-8.
- Cornilescu, G., Delaglio, F. and Bax, A. (1999). "Protein backbone angle restraints from searching a database for chemical shift and sequence homology." *J Biomol NMR* **13**(3): 289-302.
- Delaglio, F., Grzesiek, S., Vuister, G. W., Zhu, G., Pfeifer, J. and Bax, A. (1995). "NMRPipe: a multidimensional spectral processing system based on UNIX pipes." *J Biomol NMR* **6**(3): 277-93.
- Eddy, S. R. (1995). "Multiple alignment using hidden Markov models." *Proc Int Conf Intell Syst Mol Biol* **3**: 114-20.
- Eghbalnia, H. R., Bahrami, A., Wang, L., Assadi, A. and Markley, J. L. (2005a). "Probabilistic Identification of Spin Systems and their Assignments including Coil-Helix Inference as Output (PISTACHIO)." *J Biomol NMR* **32**(3): 219-33.
- Eghbalnia, H. R., Wang, L., Bahrami, A., Assadi, A. and Markley, J. L. (2005b). "Protein energetic conformational analysis from NMR chemical shifts (PECAN) and its use in determining secondary structural elements." *J Biomol NMR* **32**(1): 71-81.
- Elion, E. A. (1998). "Routing MAP kinase cascades." *Science* **281**(5383): 1625-6.
- Farrow, N. A., Muhandiram, R., Singer, A. U., Pascal, S. M., Kay, C. M., Gish, G., Shoelson, S. E., Pawson, T., Forman-Kay, J. D. and Kay, L. E. (1994). "Backbone dynamics of a free and phosphopeptide-complexed Src homology 2 domain studied by ¹⁵N NMR relaxation." *Biochemistry* **33**(19): 5984-6003.
- Fernandez, C. and Wider, G. (2003). "TROSY in NMR studies of the structure and function of large biological macromolecules." *Curr Opin Struct Biol* **13**(5): 570-80.
- Flatmark, T. (1967). "Multiple molecular forms of bovine heart cytochrome c. V. A comparative study of their physicochemical properties and their reactions in biological systems." *J Biol Chem* **242**(10): 2454-9.
- Frayne, J., Ingram, C., Love, S. and Hall, L. (1999). "Localisation of phosphatidylethanolamine-binding protein in the brain and other tissues of the rat." *Cell Tissue Res* **298**(3): 415-23.
- Fu, Z., Kitagawa, Y., Shen, R., Shah, R., Mehra, R., Rhodes, D., Keller, P. J., Mizokami, A., Dunn, R., Chinnaiyan, A. M., Yao, Z. and Keller, E. T. (2006). "Metastasis suppressor gene Raf kinase inhibitor protein (RKIP) is a novel prognostic marker in prostate cancer." *Prostate* **66**(3): 248-56.
- Fu, Z., Smith, P. C., Zhang, L., Rubin, M. A., Dunn, R. L., Yao, Z. and Keller, E. T. (2003). "Effects of raf kinase inhibitor protein expression on suppression of prostate cancer metastasis." *J Natl Cancer Inst* **95**(12): 878-89.
- Gasteiger, E., Gattiker, A., Hoogland, C., Ivanyi, I., Appel, R. D. and Bairoch, A. (2003). "ExPASy: The proteomics server for in-depth protein knowledge and analysis." *Nucleic Acids Res* **31**(13): 3784-8.
- Gibbons, R., Adeoya-Osiguwa, S. A. and Fraser, L. R. (2005). "A mouse sperm decapacitation factor receptor is phosphatidylethanolamine-binding protein 1." *Reproduction* **130**(4): 497-508.

- Glattard, E., Muller, A., Aunis, D., Metz-Boutigue, M. H., Stefano, G. B. and Goumon, Y. (2006). "Rethinking the opiate system? Morphine and morphine-6-glucuronide as new endocrine and neuroendocrine mediators." Med Sci Monit **12**(6): SR25-7.
- Golovanov, A. P., Hautbergue, G. M., Wilson, S. A. and Lian, L. Y. (2004). "A simple method for improving protein solubility and long-term stability." J Am Chem Soc **126**(29): 8933-9.
- Goumon, Y., Angelone, T., Schoentgen, F., Chasserot-Golaz, S., Almas, B., Fukami, M. M., Langley, K., Welters, I. D., Tota, B., Aunis, D. and Metz-Boutigue, M. H. (2004). "The hippocampal cholinergic neurostimulating peptide, the N-terminal fragment of the secreted phosphatidylethanolamine-binding protein, possesses a new biological activity on cardiac physiology." J Biol Chem **279**(13): 13054-64.
- Goumon, Y., Muller, A., Glattard, E., Marban, C., Gasnier, C., Strub, J. M., Chasserot-Golaz, S., Rohr, O., Stefano, G. B., Welters, I. D., Van Dorsselaer, A., Schoentgen, F., Aunis, D. and Metz-Boutigue, M. H. (2006). "Identification of morphine-6-glucuronide in chromaffin cell secretory granules." J Biol Chem **281**(12): 8082-9.
- Goumon, Y., Stefano, G. B., Aunis, D. and Metz-Boutigue, M. H. (2002). "Implication of endogenous morphine in the communication between neuroendocrine and immune systems." Ann N Y Acad Sci **971**: 542-3.
- Grandy, D. K., Hanneman, E., Bunzow, J., Shih, M., Machida, C. A., Bidlack, J. M. and Civelli, O. (1990). "Purification, cloning, and tissue distribution of a 23-kDa rat protein isolated by morphine affinity chromatography." Mol Endocrinol **4**(9): 1370-6.
- Grzesiek, S. and Bax, A. (1993). "The importance of not saturating H₂O in protein NMR. application to sensitivity enhancement and NOE measurements." J. Am. Chem. Soc. **115**: 12593–12594.
- Guex, N. and Peitsch, M. C. (1997). "SWISS-MODEL and the Swiss-PdbViewer: an environment for comparative protein modeling." Electrophoresis **18**(15): 2714-23.
- Guntert, P. (2004). "Automated NMR structure calculation with CYANA." Methods Mol Biol **278**: 353-78.
- Guntert, P., Braun, W. and Wuthrich, K. (1991). "Efficient computation of three-dimensional protein structures in solution from nuclear magnetic resonance data using the program DIANA and the supporting programs CALIBA, HABAS and GLOMSA." J Mol Biol **217**(3): 517-30.
- Guntert, P., Mumenthaler, C. and Wuthrich, K. (1997). "Torsion angle dynamics for NMR structure calculation with the new program DYANA." J Mol Biol **273**(1): 283-98.
- Guntert, P. and Wuthrich, K. (1991). "Improved efficiency of protein structure calculations from NMR data using the program DIANA with redundant dihedral angle constraints." J Biomol NMR **1**(4): 447-56.
- Hengst, U., Albrecht, H., Hess, D. and Monard, D. (2001). "The phosphatidylethanolamine-binding protein is the prototype of a novel family of serine protease inhibitors." J Biol Chem **276**(1): 535-40.
- Herrmann, T., Guntert, P. and Wuthrich, K. (2002). "Protein NMR structure determination with automated NOE assignment using the new software CANDID and the torsion angle dynamics algorithm DYANA." J Mol Biol **319**(1): 209-27.
- Hickox, D. M., Gibbs, G., Morrison, J. R., Sebire, K., Edgar, K., Keah, H. H., Alter, K., Loveland, K. L., Hearn, M. T., de Kretser, D. M. and O'Bryan, M. K. (2002).

- "Identification of a novel testis-specific member of the phosphatidylethanolamine binding protein family, pebp-2." Biol Reprod **67**(3): 917-27.
- Hirtz, C., Chevalier, F., Centeno, D., Egea, J. C., Rossignol, M., Sommerer, N. and de Periere, D. (2005). "Complexity of the human whole saliva proteome." J Physiol Biochem **61**(3): 469-80.
- Houben, R., Michel, B., Vetter-Kauczok, C. S., Pfohler, C., Laetsch, B., Wolter, M. D., Leonard, J. H., Trefzer, U., Ugurel, S., Schrama, D. and Becker, J. C. (2006). "Absence of classical MAP kinase pathway signalling in Merkel cell carcinoma." J Invest Dermatol **126**(5): 1135-42.
- Huhn, D., von Schilling, C., Wilhelm, M., Ho, A. D., Hallek, M., Kuse, R., Knauf, W., Riedel, U., Hinke, A., Srock, S., Serke, S., Peschel, C. and Emmerich, B. (2001). "Rituximab therapy of patients with B-cell chronic lymphocytic leukemia." Blood **98**(5): 1326-31.
- Jazirehi, A. R., Vega, M. I., Chatterjee, D., Goodglick, L. and Bonavida, B. (2004). "Inhibition of the Raf-MEK1/2-ERK1/2 signaling pathway, Bcl-xL down-regulation, and chemosensitization of non-Hodgkin's lymphoma B cells by Rituximab." Cancer Res **64**(19): 7117-26.
- Johnson, B. A. (2004). "Using NMRView to visualize and analyze the NMR spectra of macromolecules." Methods Mol Biol **278**: 313-52.
- Kay, L. E., Torchia, D. A. and Bax, A. (1989). "Backbone dynamics of proteins as studied by ¹⁵N inverse detected heteronuclear NMR spectroscopy: application to staphylococcal nuclease." Biochemistry **28**: 8972-8979.
- Keller, E. T. (2004). "Metastasis suppressor genes: a role for raf kinase inhibitor protein (RKIP)." Anticancer Drugs **15**(7): 663-9.
- Keller, E. T., Fu, Z. and Brennan, M. (2005). "The biology of a prostate cancer metastasis suppressor protein: Raf kinase inhibitor protein." J Cell Biochem **94**(2): 273-8.
- Kendrew, J. C., Bodo, G., Dintzis, H. M., Parrish, R. G., Wyckoff, H. and Phillips, D. C. (1958). "A three-dimensional model of the myoglobin molecule obtained by x-ray analysis." Nature **181**(4610): 662-6.
- Kikkawa, R., Koya, D. and Haneda, M. (2003). "[Diabetic nephropathy--progress in its preclinical and clinical studies]." Nippon Naika Gakkai Zasshi **92**(9): 1650-61.
- Koradi, R., Billeter, M. and Wuthrich, K. (1996). "MOLMOL: a program for display and analysis of macromolecular structures." J Mol Graph **14**(1): 51-5, 29-32.
- Kroslak, T., Koch, T., Kahl, E. and Holtt, V. (2001). "Human phosphatidylethanolamine-binding protein facilitates heterotrimeric G protein-dependent signaling." J Biol Chem **276**(43): 39772-8.
- Krupnick, J. G. and Benovic, J. L. (1998). "The role of receptor kinases and arrestins in G protein-coupled receptor regulation." Annu Rev Pharmacol Toxicol **38**: 289-319.
- Laskowski, R. A., Chistyakov, V. V. and Thornton, J. M. (2005). "PDBsum more: new summaries and analyses of the known 3D structures of proteins and nucleic acids." Nucleic Acids Res **33**(Database issue): D266-8.
- Laskowski, R. A., Rullmann, J. A., MacArthur, M. W., Kaptein, R. and Thornton, J. M. (1996). "AQUA and PROCHECK-NMR: programs for checking the quality of protein structures solved by NMR." J Biomol NMR **8**(4): 477-86.
- Levy, F., Bulet, P. and Ehret-Sabatier, L. (2004). "Proteomic analysis of the systemic immune response of *Drosophila*." Mol Cell Proteomics **3**(2): 156-66.
- Linge, J. P., O'Donoghue, S. I. and Nilges, M. (2001). "Automated assignment of ambiguous nuclear overhauser effects with ARIA." Methods Enzymol **339**: 71-90.

- Lipari, G. and Szabo, A. (1982a). "Model-free approach to the interpretation of nuclear magnetic resonance relaxation in macromolecules. 1. Theory and range of validity." J. Am. Chem. Soc. **104**: 4546–4559.
- Lipari, G. and Szabo, A. (1982b). "Model-free approach to the interpretation of nuclear magnetic resonance relaxation in macromolecules. 2. Analysis of experimental results." J. Am. Chem. Soc. **104**: 4559–4570.
- Ludvigsen, S., Andersen, K. V. and Poulsen, F. M. (1991). "Accurate measurements of coupling constants from two-dimensional nuclear magnetic resonance spectra of proteins and determination of phi-angles." J Mol Biol **217**(4): 731-6.
- Maki, M., Matsukawa, N., Yuasa, H., Otsuka, Y., Yamamoto, T., Akatsu, H., Okamoto, T., Ueda, R. and Ojika, K. (2002). "Decreased expression of hippocampal cholinergic neurostimulating peptide precursor protein mRNA in the hippocampus in Alzheimer disease." J Neuropathol Exp Neurol **61**(2): 176-85.
- Marais, R. and Marshall, C. J. (1996). "Control of the ERK MAP kinase cascade by Ras and Raf." Cancer Surv **27**: 101-25.
- Marley, J., Lu, M. and Bracken, C. (2001). "A method for efficient isotopic labeling of recombinant proteins." J Biomol NMR **20**(1): 71-5.
- Matsukawa, N., Tooyama, I., Kimura, H., Yamamoto, T., Tsugu, Y., Oomura, Y. and Ojika, K. (1999). "Increased expression of hippocampal cholinergic neurostimulating peptide-related components and their messenger RNAs in the hippocampus of aged senescence-accelerated mice." Neuroscience **88**(1): 79-92.
- Meylaers, K., Clynen, E., Daloze, D., DeLoof, A. and Schoofs, L. (2004). "Identification of 1-lysophosphatidylethanolamine (C(16:1)) as an antimicrobial compound in the housefly, *Musca domestica*." Insect Biochem Mol Biol **34**(1): 43-9.
- Mima, J., Hayashida, M., Fujii, T., Narita, Y., Hayashi, R., Ueda, M. and Hata, Y. (2005). "Structure of the carboxypeptidase Y inhibitor IC in complex with the cognate proteinase reveals a novel mode of the proteinase-protein inhibitor interaction." J Mol Biol **346**(5): 1323-34.
- Mitake, S., Ojika, K., Katada, E., Otsuka, Y., Matsukawa, N. and Fujimori, O. (1996). "Distribution of hippocampal cholinergic neurostimulating peptide (HCNP) immunoreactivity in the central nervous system of the rat." Brain Res **706**(1): 57-70.
- Moodie, S. A., Willumsen, B. M., Weber, M. J. and Wolfman, A. (1993). "Complexes of Ras.GTP with Raf-1 and mitogen-activated protein kinase kinase." Science **260**(5114): 1658-61.
- Moore, C., Perry, A. C., Love, S. and Hall, L. (1996). "Sequence analysis and immunolocalisation of phosphatidylethanolamine binding protein (PBP) in human brain tissue." Brain Res Mol Brain Res **37**(1-2): 74-8.
- Ojika, K. and Appel, S. H. (1984). "Neurotrophic effects of hippocampal extracts on medial septal nucleus in vitro." Proc Natl Acad Sci U S A **81**(8): 2567-71.
- Oldfield, E. (1995). "Chemical shifts and three-dimensional protein structures." J Biomol NMR **5**(3): 217-25.
- Otsuka, Y. and Ojika, K. (1996). "Demonstration and characterization of hippocampal cholinergic neurostimulating peptide (HCNP) processing enzyme activity in rat hippocampus." Neurochem Res **21**(3): 369-76.
- Palmer III, A. G., Rance, M. and Wright, P. E. (1991). "Intramolecular motions of a zinc finger DNA-binding domain from Xfin characterized by proton-detected natural abundance ¹³C heteronuclear NMR spectroscopy." J. Am. Chem. Soc. **113**(12): 4371–4380.

- Pantazis, P., Han, Z., Balan, K., Wang, Y. and Wyche, J. H. (2003). "Camptothecin and 9-nitrocamptothecin (9NC) as anti-cancer, anti-HIV and cell-differentiation agents. Development of resistance, enhancement of 9NC-induced activities and combination treatments in cell and animal models." Anticancer Res **23**(5A): 3623-38.
- Pardi, A., Billeter, M. and Wuthrich, K. (1984). "Calibration of the angular dependence of the amide proton-C alpha proton coupling constants, $3J_{HN\alpha}$, in a globular protein. Use of $3J_{HN\alpha}$ for identification of helical secondary structure." J Mol Biol **180**(3): 741-51.
- Park, S., Yeung, M. L., Beach, S., Shields, J. M. and Yeung, K. C. (2005). "RKIP downregulates B-Raf kinase activity in melanoma cancer cells." Oncogene **24**(21): 3535-40.
- Peng, J. W. and Wagner, G. (1992). "Mapping of the spectral densities of N-H bond motions in eglin c using heteronuclear relaxation experiments." Biochemistry **31**(36): 8571-86.
- Perry, A. C., Hall, L., Bell, A. E. and Jones, R. (1994). "Sequence analysis of a mammalian phospholipid-binding protein from testis and epididymis and its distribution between spermatozoa and extracellular secretions." Biochem J **301** (Pt 1): 235-42.
- Pervushin, K., Riek, R., Wider, G. and Wuthrich, K. (1997). "Attenuated T2 relaxation by mutual cancellation of dipole-dipole coupling and chemical shift anisotropy indicates an avenue to NMR structures of very large biological macromolecules in solution." Proc Natl Acad Sci U S A **94**(23): 12366-71.
- Pnueli, L., Carmel-Goren, L., Hareven, D., Gutfinger, T., Alvarez, J., Ganai, M., Zamir, D. and Lifschitz, E. (1998). "The SELF-PRUNING gene of tomato regulates vegetative to reproductive switching of sympodial meristems and is the ortholog of CEN and TFL1." Development **125**(11): 1979-89.
- Press, W. H., Flannery, B. P., Teukolsky, S. A. and Vetterling, W. T. (1986). Numerical recipes, Cambridge: 523-526.
- Reff, M. E., Carner, K., Chambers, K. S., Chinn, P. C., Leonard, J. E., Raab, R., Newman, R. A., Hanna, N. and Anderson, D. R. (1994). "Depletion of B cells in vivo by a chimeric mouse human monoclonal antibody to CD20." Blood **83**(2): 435-45.
- Robinson, N. E. and Robinson, A. B. (2004). "Prediction of primary structure deamidation rates of asparaginyl and glutaminyl peptides through steric and catalytic effects." J Pept Res **63**(5): 437-48.
- Roussel, G., Nussbaum, F., Schoentgen, F., Jolles, P. and Nussbaum, J. L. (1988). "Immunological investigation of a 21-kilodalton cytosolic basic protein in rat brain." Dev Neurosci **10**(2): 65-74.
- Saunders, P. T., McKinnell, C., Millar, M. R., Gaughan, J., Turner, K. J., Jegou, B., Syed, V. and Sharpe, R. M. (1995). "Phosphatidylethanolamine binding protein is an abundant secretory product of haploid testicular germ cells in the rat." Mol Cell Endocrinol **107**(2): 221-30.
- Schoentgen, F., Bucquoy, S., Seddiqi, N. and Jolles, P. (1993). "Two cytosolic protein families implicated in lipid-binding: main structural and functional features." Int J Biochem **25**(12): 1699-704.
- Schoentgen, F. and Jolles, P. (1995). "From structure to function: possible biological roles of a new widespread protein family binding hydrophobic ligands and displaying a nucleotide binding site." FEBS Lett **369**(1): 22-6.

- Schoentgen, F., Saccoccio, F., Jolles, J., Bernier, I. and Jolles, P. (1987). "Complete amino acid sequence of a basic 21-kDa protein from bovine brain cytosol." Eur J Biochem **166**(2): 333-8.
- Schoentgen, F., Seddiqi, N., Bucquoy, S., Jolles, P., Lemesle-Varloot, L., Provost, K. and Mornon, J. P. (1992). "Main structural and functional features of the basic cytosolic bovine 21 kDa protein delineated through hydrophobic cluster analysis and molecular modelling." Protein Eng **5**(4): 295-303.
- Schonwasser, D. C., Marais, R. M., Marshall, C. J. and Parker, P. J. (1998). "Activation of the mitogen-activated protein kinase/extracellular signal-regulated kinase pathway by conventional, novel, and atypical protein kinase C isotypes." Mol Cell Biol **18**(2): 790-8.
- Schubert, M., Labudde, D., Oschkinat, H. and Schmieder, P. (2002). "A software tool for the prediction of Xaa-Pro peptide bond conformations in proteins based on ¹³C chemical shift statistics." J Biomol NMR **24**(2): 149-54.
- Schuieler, M. M., Bataille, F., Weiss, T. S., Hellerbrand, C. and Bosserhoff, A. K. (2006). "Raf kinase inhibitor protein is downregulated in hepatocellular carcinoma." Oncol Rep **16**(3): 451-6.
- Serre, L., Pereira de Jesus, K., Zelwer, C., Bureau, N., Schoentgen, F. and Benedetti, H. (2001). "Crystal structures of YBHB and YBCL from *Escherichia coli*, two bacterial homologues to a Raf kinase inhibitor protein." J Mol Biol **310**(3): 617-34.
- Serre, L., Vallee, B., Bureau, N., Schoentgen, F. and Zelwer, C. (1998). "Crystal structure of the phosphatidylethanolamine-binding protein from bovine brain: a novel structural class of phospholipid-binding proteins." Structure **6**(10): 1255-65.
- Shaulian, E. and Karin, M. (2002). "AP-1 as a regulator of cell life and death." Nat Cell Biol **4**(5): E131-6.
- Simister, P. C., Banfield, M. J. and Brady, R. L. (2002). "The crystal structure of PEBP-2, a homologue of the PEBP/RKIP family." Acta Crystallogr D Biol Crystallogr **58**(Pt 6 Pt 2): 1077-80.
- Slupsky, C. M., Boyko, R. F., Booth, V. K. and Sykes, B. D. (2003). "Smartnotebook: a semi-automated approach to protein sequential NMR resonance assignments." J Biomol NMR **27**(4): 313-21.
- Sreerama, N., Venyaminov, S. Y. and Woody, R. W. (2000). "Estimation of protein secondary structure from circular dichroism spectra: inclusion of denatured proteins with native proteins in the analysis." Anal Biochem **287**(2): 243-51.
- Stapleton, M., Carlson, J., Brokstein, P., Yu, C., Champe, M., George, R., Guarin, H., Kronmiller, B., Pacleb, J., Park, S., Wan, K., Rubin, G. M. and Celniker, S. E. (2002). "A *Drosophila* full-length cDNA resource." Genome Biol **3**(12): RESEARCH0080.
- Stein, E. G., Rice, L. M. and Brünger, A. T. (1997). "Torsion-angle molecular dynamics as a new efficient tool for NMR structure calculation." J. Magn. Reson. **124**: 154-164.
- Thongboonkerd, V. (2004). "Proteomics in nephrology: current status and future directions." Am J Nephrol **24**(3): 360-78.
- Tohdoh, N., Tojo, S., Agui, H. and Ojika, K. (1995). "Sequence homology of rat and human HCNP precursor proteins, bovine phosphatidylethanolamine-binding protein and rat 23-kDa protein associated with the opioid-binding protein." Brain Res Mol Brain Res **30**(2): 381-4.

- Trakul, N. and Rosner, M. R. (2005). "Modulation of the MAP kinase signaling cascade by Raf kinase inhibitory protein." Cell Res **15**(1): 19-23.
- Tsugu, Y., Ojika, K., Matsukawa, N., Iwase, T., Otsuka, Y., Katada, E. and Mitake, S. (1998). "High levels of hippocampal cholinergic neurostimulating peptide (HCNP) in the CSF of some patients with Alzheimer's disease." Eur J Neurol **5**(6): 561-569.
- Tugarinov, V., Choy, W. Y., Orekhov, V. Y. and Kay, L. E. (2005). "Solution NMR-derived global fold of a monomeric 82-kDa enzyme." Proc Natl Acad Sci U S A **102**(3): 622-7.
- Turner, K. J., McKinnell, C., McLaren, T. T., Qureshi, S. J., Saunders, P. T., Foster, P. M. and Sharpe, R. M. (1996). "Detection of germ cell-derived proteins in testicular interstitial fluid: potential for monitoring spermatogenesis in vivo." J Androl **17**(2): 127-36.
- Tuttle, K. R. and Anderson, P. W. (2003). "A novel potential therapy for diabetic nephropathy and vascular complications: protein kinase C beta inhibition." Am J Kidney Dis **42**(3): 456-65.
- Ueda, Y., Hirai, S., Osada, S., Suzuki, A., Mizuno, K. and Ohno, S. (1996). "Protein kinase C activates the MEK-ERK pathway in a manner independent of Ras and dependent on Raf." J Biol Chem **271**(38): 23512-9.
- Vallee, B., Teyssier, C., Maget-Dana, R., Ramstein, J., Bureaud, N. and Schoentgen, F. (1999). "Stability and physicochemical properties of the bovine brain phosphatidylethanolamine-binding protein." Eur J Biochem **266**(1): 40-52.
- Vallee, B. S., Tauc, P., Brochon, J. C., Maget-Dana, R., Lelievre, D., Metz-Boutigue, M. H., Bureaud, N. and Schoentgen, F. (2001). "Behaviour of bovine phosphatidylethanolamine-binding protein with model membranes. Evidence of affinity for negatively charged membranes." Eur J Biochem **268**(22): 5831-41.
- Vierstraete, E., Verleyen, P., Baggerman, G., D'Hertog, W., Van den Bergh, G., Arckens, L., De Loof, A. and Schoofs, L. (2004a). "A proteomic approach for the analysis of instantly released wound and immune proteins in *Drosophila melanogaster* hemolymph." Proc Natl Acad Sci U S A **101**(2): 470-5.
- Vierstraete, E., Verleyen, P., Sas, F., Van den Bergh, G., De Loof, A., Arckens, L. and Schoofs, L. (2004b). "The instantly released *Drosophila* immune proteome is infection-specific." Biochem Biophys Res Commun **317**(4): 1052-60.
- Vuister, G. W. and Bax, A. (1993). "Quantitative J Correlation: A New Approach for Measuring Homonuclear Three-Bond J(¹H NHa) Coupling Constants in ¹⁵N-Enriched Proteins." JACS **115**: 7772-7777.
- Vuister, G. W., Delaglio, F. and Bax, A. (1993). "The use of ¹J C alpha H alpha coupling constants as a probe for protein backbone conformation." J Biomol NMR **3**(1): 67-80.
- Wand, A. J. (2001). "Dynamic activation of protein function: a view emerging from NMR spectroscopy." Nat Struct Biol **8**(11): 926-31.
- Wang, A. C. and Bax, A. (1996). "Determination of the Backbone Dihedral Angles ϕ in Human Ubiquitin from Reparametrized Empirical Karplus Equations." J. Am. Chem. Soc. **118**: 2483-2494.
- Weintraub, S. J. and Manson, S. R. (2004). "Asparagine deamidation: a regulatory hourglass." Mech Ageing Dev **125**(4): 255-7.
- Williamson, M. P., Havel, T. F. and Wuthrich, K. (1985). "Solution conformation of proteinase inhibitor IIA from bull seminal plasma by ¹H nuclear magnetic resonance and distance geometry." J Mol Biol **182**(2): 295-315.

- Wishart, D. S. and Sykes, B. D. (1994). "Chemical shifts as a tool for structure determination." Methods Enzymol **239**: 363-92.
- Wishart, D. S., Sykes, B. D. and Richards, F. M. (1991). "Relationship between nuclear magnetic resonance chemical shift and protein secondary structure." J Mol Biol **222**(2): 311-33.
- Yeung, K., Janosch, P., McFerran, B., Rose, D. W., Mischak, H., Sedivy, J. M. and Kolch, W. (2000). "Mechanism of suppression of the Raf/MEK/extracellular signal-regulated kinase pathway by the raf kinase inhibitor protein." Mol Cell Biol **20**(9): 3079-85.
- Yeung, K., Seitz, T., Li, S., Janosch, P., McFerran, B., Kaiser, C., Fee, F., Katsanakis, K. D., Rose, D. W., Mischak, H., Sedivy, J. M. and Kolch, W. (1999). "Suppression of Raf-1 kinase activity and MAP kinase signalling by RKIP." Nature **401**(6749): 173-7.
- Yeung, K. C., Rose, D. W., Dhillon, A. S., Yaros, D., Gustafsson, M., Chatterjee, D., McFerran, B., Wyche, J., Kolch, W. and Sedivy, J. M. (2001). "Raf kinase inhibitor protein interacts with NF-kappaB-inducing kinase and TAK1 and inhibits NF-kappaB activation." Mol Cell Biol **21**(21): 7207-17.
- Yuasa, H., Ojika, K., Mitake, S., Katada, E., Matsukawa, N., Otsuka, Y., Fujimori, O. and Hirano, A. (2001). "Age-dependent changes in HCNP-related antigen expression in the human hippocampus." Brain Res Dev Brain Res **127**(1): 1-7.
- Zhu, W., Bilfinger, T. V., Baggerman, G., Goumon, Y. and Stefano, G. B. (2001). "Presence of endogenous morphine and morphine 6 glucuronide in human heart tissue." Int J Mol Med **7**(4): 419-22.

Annexes

Annexe 1 : Clonage de la région codante du gène *cg7054*.

1. Matériel initial

La région codante correspondant au gène *cg7054* est fournie par le consortium international « Berkeley Drosophila Genome Project » (clone GH14779). Ce consortium met à disposition de la communauté scientifique une banque de plus de 14000 ADNc (Stapleton *et al.*, 2002). L'ADNc du gène *cg7054* est complet, il est fourni inclus dans le plasmide pOT2, entre le site 5' EcoRI et 3' XhoI.

2. Amplification par PCR du gène CG7054

La PCR est réalisée sur un thermocycleur MWG Biotech Primus. Les oligonucléotides amorces ont été synthétisés par MWG :

Forward CG7054: 5' – CAT ATG GAT GAC ATA GTA CCC GAT G 3'

Reverse CG7054 : 5' –CTC GAG CTA CCC TAC GAT CGT TTT ATT CC –3'

L'amorce « forward CG7054 » permet d'amplifier la portion codante du gène *cg7054* et d'intégrer en 5' du gène le site de coupure à l'endonucléase NdeI. L'amorce « reverse CG7054 » permet d'amplifier la portion codante du gène *cg7054* et d'intégrer le site de coupure à l'endonucléase XhoI en 3' du fragment codant.

Les conditions pour la PCR sont la *Taq* polymérase (1U ; *dynazyme II* produite par Ozyme), dNTPs (0,2mM), MgCl₂ (2mM), les amorces (0,2μM) et 1 μL d'ADN matrice, pour un volume total de 60 μL.

Le programme utilisé est le suivant:

- 94°C, 4 minutes
- 60°C, 1 minute

- 72°C, 1 minute
- 30 cycles: (94°C, 4 minutes - 60°C, 1 minute - 72°C, 1 minute)
- 72°C, 7 minutes

Les fragments amplifiés par PCR sont ensuite purifiés par électrophorèse sur gel d'agarose 1% et extraits du gel grâce au kit « Nucleospin Extract » de chez Macherey-Nagel.

3. Clonage dans le vecteur pCR 2.1 – TOPO

Le vecteur pCR2.1-TOPO est un vecteur navette utilisé pour permettre un meilleur rendement de transformation des produits de PCR par rapport à une transformation classique.

Le clonage dans le plasmide pCR2.1-TOPO s'effectue grâce à un kit fourni par *Invitrogen* comportant le vecteur, les bactéries « Top10 » compétentes et les différentes solutions requises. Le vecteur est vendu déjà ouvert avec un système de topoisomérases greffées de part et d'autre du site d'ouverture du plasmide. Les extrémités T débordantes en 3' permettent une hybridation avec un fragment d'ADN amplifié par PCR ayant une extrémité A débordante en 3' (apportée par la Taq polymérase *dynazyme II®*). Les topoisomérases permettent alors la formation d'une liaison phosphodiester avec le vecteur et le produit de PCR. De plus, le site d'insertion est situé dans le gène de la β -galactosidase pour permettre une sélection facile des vecteurs contenant l'insert sur milieu LB gélosé contenant IPTG et X-Gal.

3.1. Insertion dans le vecteur pCR2.1-TOPO

Le mélange réactionnel est le suivant :

- 2 μ L de produit de PCR purifié
- 1 μ L de solution saline du kit

- 2 μL de vecteur pCR2.1-TOPO
- 2 μL d'eau PPI

La réaction d'intégration du fragment dans le vecteur se déroule en 15 minutes à température ambiante.

3.2. Transformation dans les bactéries Top10:

Le produit {insert cg7054 - pCR2.1-TOPO} est mis en présence des bactéries compétentes Top10. Ce mélange est laissé 15 minutes sur la glace. Un choc thermique de 30 secondes à 42°C est ensuite réalisé afin de faciliter la pénétration du plasmide dans les bactéries. 250 μL de milieu SOC sont ajoutés et le tube est placé sous agitation, en position horizontale à 200 rpm pendant une heure à 37°C. Les bactéries sont ensuite étalées sur deux boîtes LB agar (20mL) contenant 80 μL de X-Gal 2%, 4 μL d'IPTG 1M et 20 μL d'ampicilline 100 mg.mL⁻¹. Les boîtes sont ensuite incubées une nuit à 37°C.

Les colonies blanches obtenues sont repiquées stérilement en milieu liquide : la colonie est prélevée à l'aide d'un cure dent et placée dans un tube stérile contenant 2,5 mL de LB et 2,5 μL d'ampicilline (100 mg.mL⁻¹). Le tube est placé sous agitation à 37°C pendant 10 heures.

3.3. Extraction de plasmide

La transformation des bactéries nécessite des contrôles afin de déterminer les clones positifs (ayant reçu le fragment d'intérêt). Deux protocoles existent : il s'agit des minipréparations classiques et des minipréparations par le kit «Nucleospin Plasmid » de chez Macherey-Nagel qui permet l'obtention d'une préparation plus pure.

Extraction classique (miniprep classique) : 1,5 mL des différentes cultures bactériennes sont placés stérilement dans des tubes eppendorf et centrifugés 5 minutes à 4000 rpm. Le surnageant est éliminé et le culot est repris dans 100 μL de solution I (Tris-HCl 25 mM pH 8, EDTA 10 mM, glucose 50 mM). Les tubes sont placés 5 minutes dans la glace. Ensuite 200 μL de solution II (SDS 1 %, NaOH 0,2 M) sont ajoutés. On retourne les tubes

pour mélanger délicatement avant de les placer 5 minutes dans la glace. 150 μL de solution III (acétate de potassium 5M pH 4,8) sont ajoutés. Le mélange est homogénéisé puis laissé 5 minutes dans la glace. Suite à une centrifugation de 15 minutes à 14000 rpm à 4°C, les surnageants sont prélevés et mis dans une nouvelle série de tubes eppendorf contenant 400 μL de mélange phénol / chloroforme (1/1 ; V/V). Les tubes sont alors vortexés 20 secondes pour émulsionner les 2 phases. Une centrifugation de 5 minutes à 14000 rpm est réalisée puis la phase aqueuse est prélevée et placée dans une nouvelle série de tubes contenant 700 μL d'éthanol 100%. Après agitation, les tubes sont laissés 10 minutes à 4°C. On centrifuge à 14000 rpm pendant 15 minutes. Le surnageant est éliminé et le culot qui contient le plasmide est séché au speed vac avant d'être repris dans 20 μL d'eau PPI contenant 0,1 mg.mL^{-1} de RNase.

Une fraction des plasmides extraits est ensuite analysée pour déterminer s'ils possèdent l'insert *cg7054* grâce à une digestion contrôlée par les enzymes NdeI et XhoI.

La réaction de coupure enzymatique se déroule durant 2 heures à 37°C. Les enzymes de restriction utilisées ainsi que leur tampon de réaction sont fournis par BioLabs. Pour chaque digestion, les enzymes sont utilisées dans un rapport 0,1 U/ μg d'ADN. Le résultat est observé par électrophorèse sur gel d'agarose 1% en présence de bromure d'éthidium. Le cas échéant, on visualise une bande de 540 pb correspondant à l'insert *cg7054*.

4. Préparation de l'insert CG7054

Après avoir détecté les plasmides contenant l'insert *cg7054*, nous réalisons une digestion préparative de ces plasmides afin de libérer le fragment 5'NdeI-*cg7054*-3'XhoI.

La digestion enzymatique est réalisée pendant 2 heures à 37°C puis les fragments sont purifiés sur gel d'agarose 1% en présence de bromure d'éthidium. La bande d'intérêt de 540 pb correspondant à l'insert *cg7054* est découpée du gel et purifiée grâce au kit « nucleospin extract » de chez Macherey-Nagel.

5. Transfert dans le vecteur pET31b ouvert, ligation

Le vecteur pET31b est un vecteur d'expression communément utilisé pour produire les protéines PEBP recombinantes.

L'ouverture du vecteur pET31b est réalisée de la même manière que la préparation de l'insert *cg7054*, mais la bande purifiée correspond à un ADN de 5700 pb. Le vecteur ouvert et l'insert sont mis en présence et une étape de ligation est réalisée avec le kit « Rapid DNA ligation kit » fourni par Roche selon les recommandations du fabricant.

Le produit de ligation est ensuite transformé dans des bactéries C600, facilement transformables et permettant un séquençage d'ADN aisé, puis dans les bactéries *E. Coli* BL21DE3, dont le rendement de transformation est plus faible mais qui permettent l'expression de la protéine.

6. Etape de transformation

Le produit de ligation {pET31b – *cg7054*} doit ensuite être intégré dans les bactéries C600 pour amplifier l'ADN. Il faut pour cela des bactéries C600 compétentes.

6.1. Préparation de cellules compétentes

Une pré-culture de bactéries C600 est relancée à partir d'un stock glycérol dans 2 mL de LB et 2 μ L d'ampicilline (100 mg.mL⁻¹). On incube à 37°C sous agitation pour la nuit. Le lendemain les bactéries sont diluées à une DO à 600 nm de 0,1, dans 10 mL de milieu LB ampicilline. La culture est incubée à 37°C sous agitation jusqu'à obtenir une DO à 600 nm de 0,6 à 0,8. La culture est enfin centrifugée stérilement 5 minutes à 4000 rpm, puis le culot de cellules est repris dans 500 μ L de TSB froid.

Les bactéries compétentes sont aliquotées et conservées dans la glace ou à 4°C pour une durée ne dépassant pas 24h, ou sous forme de stocks glycérols à -80°C pour des durées plus longues.

6.2. Transformation

50 µl de cellules compétentes C600 sont mises en présence de 4µL du produit de ligation vecteur-insert {pET31b – cg7054}. Le mélange est laissé dans la glace 30 minutes puis on réalise un choc thermique de 30 secondes à 42°C. Ensuite 250 µL de TSB et 8 µL de glucose 50% sont ajoutés. L'ensemble est incubé une heure à 37°C sous agitation avant d'être déposé sur boîte LB agar ampicilline. Les clones obtenus sont repiqués en milieu liquide dans 2,5 mL de LB ampicilline.

Les clones de bactéries C600 sont ensuite analysés comme précédemment par digestion contrôle afin de vérifier que l'insert *cg7054* est bien intégré dans le plasmide pET31b. Si c'est le cas, le plasmide est alors transformé dans des bactéries compétentes de la lignée BL21 DE3 (protocole identique à la transformation du plasmide dans les cellules C600).

7. Analyse finale des clones de bactéries BL21DE3 ayant intégré le plasmide pET31b-CG7054

Pour vérifier la présence de l'insert *cg7054*, son absence de mutation et la bonne expression de la protéine correspondante, nous réalisons les contrôles suivants :

- Digestion contrôle par les enzymes NdeI et XhoI sur le plasmide purifié.
- Séquençage de l'ADN du plasmide purifié (MWG Biotech) à partir des cellules C600.
- Test d'induction : une culture de bactérie en milieu liquide LB en présence d'ampicilline 100 µg.mL⁻¹ est incubée jusqu'à une DO à 600 nm de 0,6 à 0,8 puis induite avec 1mM d'IPTG. Après 2 heures d'induction, 1 mL de culture est centrifugé et les bactéries sont reprises dans 100 µL de tampon de charge pour SDS-PAGE. Le mélange est soniqué dans un bain à ultrasons pendant 15 minutes puis centrifugé à 14400 rpm durant 10 minutes. Le surnageant est chargé sur gel SDS-PAGE 15%. Après migration les protéines sont visualisées par coloration au bleu de Coomassie.

- Séquençage des 12 acides aminés N-terminaux par séquençage d'Edman. Ce séquençage est réalisé avec 2 µg de protéine purifiée sur un séquenceur Procise 492 de chez Applied Biosystems.

Annexe 2 : Paramètres d'acquisition des expériences 2D et 3D mises en œuvre pour l'étude structurale de la protéine CG7054

expérience	HSQC	HNCA	HNCOCA	HNCO	HNCACO	HNCACB
champs (MHz)	600	600	600	600	600	800
temps de mélange (ms)						
noyau en dimension 1	¹ H	¹ H	¹ H	¹ H	¹ H	¹ H
noyau en dimension 2	¹⁵ N	¹³ C	¹³ C	¹³ C	¹³ C	¹³ C
noyau en dimension 3		¹⁵ N	¹⁵ N	¹⁵ N	¹⁵ N	¹⁵ N
np - dimension 1	1472	2048	2048	2048	2048	2048
ni - dimension 2	128	64	63	64	64	100
ni2 - dimension 3		32	32	32	32	60
sw - dimension 1 (Hz)	8000	8000	8000	8000	8000	13000
sw - dimension 2 (Hz)	2200	4526	4526	3017	3017	16094
sw2 - dimension 3 (Hz)		3000	3000	2200	2200	2850
nombre de scans	8	16	16	8	32	8

expérience	CBCACONH	15N TOCSY-HSQC	CCONH	HCCONH	HBHACONH
champs (MHz)	600	600	600	600	600
temps de mélange (ms)		40			
noyau en dimension 1	¹ H	¹ H	¹ H	¹ H	¹ H
noyau en dimension 2	¹³ C	¹ H	¹³ C		
noyau en dimension 3	¹⁵ N	¹⁵ N	¹⁵ N	¹⁵ N	¹⁵ N
np - dimension 1	2048	1216	1472	1472	1472
ni - dimension 2	64	128	128	250	130
ni2 - dimension 3	48	40	45	32	32
sw - dimension 1 (Hz)	8000	9000	8000	8000	8000
sw - dimension 2 (Hz)	12070	9000	12068	8000	8000
sw2 - dimension 3 (Hz)	3000	2200	2200	2200	2200
nombre de scans	16	8	8	8	8

expérience	HNHA	HCCH-COSY	HCCH-TOCSY	hbCBcgcdceHE	hbCBcgcdHD
champs (MHz)	600	600	800	600	600
temps de mélange (ms)		140	14		
noyau en dimension 1	^1H	^1H	^1H	^1H	^1H
noyau en dimension 2		^1H	^1H	^{13}C	^{13}C
noyau en dimension 3	^{15}N	^{13}C	^{13}C		
np - dimension 1	1472	1472	756	1472	1024
ni - dimension 2	128	200	212	37	37
ni2 - dimension 3	32	32	96		
sw - dimension 1 (Hz)	8000	8000	9612	8000	8000
sw - dimension 2 (Hz)	8000	8000	5600	4530	4526
sw2 - dimension 3 (Hz)	2200	12070	14079		
nombre de scans	8	16	8	600	304

expérience	15N-NOESY-HSQC	13C-NOESY-HSQC	13C-ARO-NOESY-HSQC
champs (MHz)	800	600	600
temps de mélange (ms)	150	150	120
noyau en dimension 1	^1H	^1H	^1H
noyau en dimension 2	^1H	^1H	^1H
noyau en dimension 3	^{15}N	^{13}C	^{13}C
np - dimension 1	1024	1472	1472
ni - dimension 2	122	128	128
ni2 - dimension 3	38	106	58
sw - dimension 1 (Hz)	12000	8000	8000
sw - dimension 2 (Hz)	10000	8000	8000
sw2 - dimension 3 (Hz)	3000	12070	7542
nombre de scans	8	8	8

Annexe 3 : Résultats obtenus pour les indices de déplacement chimique de la protéine CG7054.

ND : non déterminé ; B : brin β ; H : hélice α

	HA	CA	CO	CB	Consensus	
résidu	CSI	CSI	CSI	CSI	CSI	Interprétation
M1	0	0	0	0	0	ND
D2	-1	1	-1	-1	0	ND
D3	-1	1	-1	0	0	ND
I4	-1	0	-1	0	0	ND
V5	1	-1	-1	1	0	ND
P6	1	0	0	0	0	ND
D7	-1	1	-1	-1	0	ND
V8	0	1	-1	1	0	ND
L9	1	-1	-1	1	0	ND
D10	-1	1	-1	0	0	ND
A11	-1	-1	-1	1	0	ND
V12	-1	-1	-1	0	0	ND
P13	-1	0	0	0	0	ND
A14	-1	1	-1	-1	0	ND
G15	0	0	0	0	0	ND
T16	1	-1	-1	1	0	ND
I17	0	-1	-1	1	1	B
K18	1	0	-1	-1	1	B
V19	1	-1	-1	1	1	B
I20	1	-1	-1	1	1	B
Y21	-1	0	0	1	0	ND
G22	-1	1	0	-1	0	ND
D23	-1	1	0	-1	0	ND
D24	-1	1	-1	-1	0	ND
L25	1	-1	-1	-1	0	ND
E26	1	0	0	1	0	ND
V27	-1	1	-1	0	0	ND
K28	1	-1	-1	0	0	ND
Q29	-1	1	0	-1	0	ND
G30	-1	0	1	-1	0	ND
N31	-1	0	-1	-1	0	ND
E32	1	-1	1	0	1	B
L33	1	-1	0	1	1	B
T34	1	-1	-1	0	1	B
P35	-1	-1	0	1	0	ND
T36	1	-1	1	0	0	ND
Q37	-1	1	1	-1	0	ND
V38	1	-1	-1	-1	0	ND
K39	-1	1	-1	-1	0	ND
D40	0	-1	-1	1	0	ND
Q41	-1	-1	-1	-1	0	ND

P42	-1	0	0	0	0	ND
I43	-1	-1	-1	0	0	ND
V44	1	-1	-1	1	0	ND
S45	0	0	0	1	0	ND
W46	1	-1	-1	-1	0	ND
S47	-1	1	1	-1	0	ND
G48	0	0	1	0	0	ND
L49	1	0	-1	0	0	ND
E50	0	0	0	-1	0	ND
G51	0	1	1	0	0	ND
K52	0	-1	-1	0	0	ND
S53	1	-1	0	1	0	ND
N54	-1	1	1	-1	0	ND
L55	1	-1	-1	1	1	B
L56	1	0	-1	1	1	B
T57	1	0	-1	1	1	B
L58	1	-1	-1	1	1	B
L59	1	-1	-1	1	1	B
M60	1	-1	1	1	1	B
V61	1	-1	-1	1	1	B
D62	1	-1	-1	1	1	B
P63	0	0	0	0	0	ND
D64	-1	1	-1	0	0	ND
A65	0	-1	-1	1	0	ND
P66	1	0	0	0	0	ND
T67	-1	-1	-1	1	0	ND
R68	-1	1	1	0	0	ND
Q69	-1	1	0	-1	0	ND
D70	-1	-1	-1	-1	0	ND
P71	0	0	0	0	0	ND
K72	-1	1	1	0	0	ND
Y73	1	-1	-1	-1	0	ND
R74	-1	-1	1	1	0	ND
E75	1	1	0	1	0	ND
I76	1	-1	1	0	0	ND
L77	1	1	-1	-1	0	ND
H78	-1	1	-1	-1	0	ND
W79	1	-1	-1	1	1	B
S80	1	-1	0	0	1	B
V81	1	-1	-1	1	1	B
V82	1	-1	-1	1	1	B
N83	-1	1	0	-1	0	ND
I84	-1	-1	-1	1	0	ND

P85	-1	0	0	0	0	ND
G86	-1	1	1	-1	0	ND
S87	-1	0	0	0	0	ND
N88	-1	0	0	-1	0	ND
E89	-1	-1	-1	0	0	ND
N90	-1	-1	-1	0	0	ND
P91	1	0	0	0	0	ND
S92	-1	1	1	0	0	ND
G93	-1	1	1	-1	0	ND
G94	-1	0	-1	-1	0	ND
H95	0	0	-1	0	0	ND
S96	1	-1	1	-1	0	ND
L97	0	1	1	-1	0	ND
A98	1	0	-1	0	0	ND
D99	-1	1	-1	0	0	ND
Y100	-1	1	-1	-1	0	ND
V101	-1	-1	-1	1	0	ND
G102	-1	0	1	-1	0	ND
S103	-1	1	1	-1	0	ND
G104	0	0	0	0	0	ND
P105	0	0	0	0	0	ND
P106	-1	0	0	0	0	ND
K107	-1	1	0	-1	0	ND
D108	-1	1	-1	-1	0	ND
T109	1	-1	0	1	0	ND
G110	0	0	0	0	0	ND
L111	0	0	1	0	0	ND
H112	1	-1	-1	-1	0	ND
R113	0	0	-1	1	0	ND
Y114	-1	0	-1	-1	0	ND
I115	1	-1	-1	1	1	B
F116	1	-1	-1	1	1	B
L117	1	-1	-1	1	1	B
L118	1	-1	-1	1	1	B
Y119	1	-1	1	1	1	B
R120	-1	1	-1	0	0	ND
Q121	0	-1	-1	0	0	ND
E122	1	1	0	1	0	ND
N123	0	-1	0	1	0	ND
K124	-1	1	-1	0	0	ND
I125	-1	-1	-1	1	0	ND
E126	-1	1	0	-1	0	ND
E127	-1	-1	0	-1	0	ND
T128	-1	-1	-1	-1	0	ND
P129	-1	0	0	0	0	ND
T130	-1	1	0	1	0	ND
I131	0	0	-1	0	0	ND
S132	-1	0	1	1	0	ND
N133	-1	1	0	-1	0	ND
T134	0	-1	-1	0	0	ND
T135	0	0	-1	1	0	ND

R136	-1	1	-1	0	0	ND
T137	-1	1	-1	0	0	ND
G138	0	0	1	0	0	ND
R139	-1	0	-1	-1	0	ND
L140	0	0	1	-1	0	ND
N141	-1	1	0	-1	-1	H
F142	-1	1	-1	0	-1	H
N143	-1	0	-1	1	-1	H
A144	-1	1	1	-1	-1	H
R145	-1	1	1	-1	-1	H
D146	-1	1	1	-1	-1	H
F147	-1	1	1	0	-1	H
A148	-1	1	1	-1	-1	H
A149	-1	1	1	-1	-1	H
K150	-1	1	1	-1	-1	H
H151	-1	0	0	-1	-1	H
G152	-1	1	1	-1	-1	H
L153	0	0	-1	-1	0	ND
G154	-1	1	1	-1	0	ND
E155	1	-1	-1	-1	0	ND
P156	-1	0	0	0	0	ND
I157	1	-1	-1	0	1	B
A158	0	-1	-1	1	1	B
A159	1	-1	0	1	1	B
N160	1	-1	-1	1	1	B
Y161	1	-1	-1	1	1	B
Y162	1	-1	-1	1	1	B
Q163	1	-1	-1	1	1	B
A164	1	0	-1	1	1	B
Q165	1	-1	-1	1	1	B
Y166	-1	1	1	0	0	ND
D167	-1	-1	-1	0	0	ND
D168	-1	1	0	-1	0	ND
Y169	-1	1	1	0	0	ND
V170	0	0	0	0	0	ND
P171	-1	0	0	0	0	ND
I172	-1	1	1	-1	-1	H
R173	-1	1	1	-1	-1	H
N174	-1	1	1	-1	-1	H
K175	-1	1	1	0	-1	H
T176	-1	-1	-1	1	0	ND
I177	0	-1	-1	-1	0	ND
V178	1	-1	-1	1	0	ND
G179	0	0	0	0	0	ND

Annexe 4 : Angles prédits par le logiciel Talos pour la protéine CG7054

Résidu	Phi	Psi	W79	-123 ± 11	143 ± 17
D2	-69 ± 10	-20 ± 18	S80	-105 ± 18	123 ± 12
I4	-110 ± 37	143 ± 13	V81	-130 ± 22	152 ± 17
V5	-117 ± 12	144 ± 21	N83	54 ± 7	44 ± 12
V8	-148 ± 34	150 ± 15	E89	-129 ± 31	153 ± 18
A11	-110 ± 22	139 ± 19	G94	89 ± 5	3 ± 12
G15	84 ± 9	14 ± 1	H95	-101 ± 31	135 ± 14
T16	-116 ± 24	143 ± 17	D99	-100 ± 31	132 ± 12
I17	-134 ± 14	132 ± 12	Y100	-70 ± 9	135 ± 7
K18	-104 ± 15	123 ± 9	G104	9999 ± 0	9999 ± 0
V19	-125 ± 16	133 ± 11	D108	-73 ± 11	-32 ± 19
I20	-119 ± 14	127 ± 11	T109	-107 ± 18	-1 ± 16
Y21	-76 ± 17	147 ± 20	G110	28 ± 84	-4 ± 112
D24	-84 ± 19	-18 ± 20	L111	-72 ± 14	134 ± 16
L25	-104 ± 17	110 ± 19	H112	-112 ± 23	148 ± 28
E26	-106 ± 25	132 ± 21	R113	-122 ± 29	146 ± 21
V27	-105 ± 23	127 ± 6	Y114	-88 ± 20	139 ± 29
N31	-84 ± 31	145 ± 14	I115	-129 ± 12	142 ± 8
E32	-108 ± 12	126 ± 10	F116	-113 ± 13	136 ± 16
L33	-117 ± 12	143 ± 17	L118	-122 ± 17	137 ± 6
T36	-88 ± 11	162 ± 5	Y119	-118 ± 13	136 ± 15
Q37	-62 ± 9	-30 ± 15	R120	-107 ± 40	142 ± 15
K39	-86 ± 8	-36 ± 21	Q121	-110 ± 19	125 ± 12
D40	-124 ± 26	143 ± 10	E122	-103 ± 15	142 ± 10
Q41	-113 ± 25	119 ± 26	N123	-122 ± 19	144 ± 21
I43	-80 ± 14	136 ± 15	E126	-86 ± 31	130 ± 12
V44	-117 ± 19	141 ± 20	E127	-84 ± 22	114 ± 13
W46	-97 ± 16	148 ± 40	T130	-79 ± 8	132 ± 11
G51	83 ± 14	17 ± 17	I131	-92 ± 24	129 ± 17
K52	-92 ± 28	150 ± 20	S132	-92 ± 21	141 ± 25
N54	-69 ± 8	-34 ± 14	L140	-76 ± 19	-12 ± 29
L55	-141 ± 17	148 ± 15	N141	-79 ± 12	-15 ± 24
L56	-143 ± 16	141 ± 17	F142	-86 ± 22	-5 ± 22
T57	-133 ± 16	135 ± 10	N143	-73 ± 15	125 ± 16
L58	-128 ± 16	131 ± 9	A144	-58 ± 4	-32 ± 8
L59	-141 ± 13	154 ± 12	R145	-64 ± 5	-40 ± 12
M60	-129 ± 17	139 ± 16	D146	-67 ± 6	-38 ± 7
V61	-132 ± 11	147 ± 16	F147	-65 ± 3	-43 ± 5
D64	-97 ± 30	141 ± 22	A148	-60 ± 5	-42 ± 5
T67	-109 ± 28	158 ± 13	K150	-65 ± 7	-33 ± 14
R68	-59 ± 5	-30 ± 14	E155	-93 ± 21	119 ± 23
Y73	-109 ± 16	148 ± 13	I157	-103 ± 18	-3 ± 21
H78	-82 ± 14	-26 ± 7	A158	-141 ± 8	149 ± 14

A159	-148 ± 17	148 ± 19
N160	-139 ± 16	139 ± 18
Y161	-132 ± 15	146 ± 15
Y162	-145 ± 15	159 ± 15
Y166	-67 ± 18	132 ± 10
D167	-80 ± 14	-7 ± 25
D168	-69 ± 13	-21 ± 24
Y169	-75 ± 16	-36 ± 20
V170	-60 ± 6	-42 ± 7
I172	-66 ± 10	-41 ± 6
R173	-64 ± 7	-39 ± 5

Annexe 5 : Table de déplacements chimiques de la protéine CG7054.

résidu	N	C	C ^α	C ^β	other
M1		176.823	55.886	35.716 (2.095, 2.095)	C ^γ , 25.660 (*, *); C ^ε , 23.573 (*)
D2	122.344 (8.124)	176.223	56.028 (4.472)	41.118 (2.846, 2.707)	
D3	115.625 (8.424)	176.222	55.390 (4.675)	41.282 (2.686, 2.580)	
I4	113.930 (7.981)	174.532	62.980 (3.997)	38.009 (1.844)	C ^{γ1} , 26.115 (1.191, 1.191); C ^{γ2} , 17.209 (0.373); C ^{δ1} , 13.569 (0.528)
V5	124.588 (7.762)	176.158	57.886 (5.133)	34.322 (2.041)	C ^{γ1} , 20.888 (0.850); C ^{γ2} , 19.645 (0.642)
P6		178.294	63.899 (5.563)	33.982 (2.410, 2.149)	C ^γ , 24.275 (2.077, 1.775); C ^δ , 48.851 (3.620, 3.493);
D7	124.869 (8.393)	176.688	58.523 (4.283)	40.531 (2.565, 2.458)	
V8	115.839 (8.082)	172.993	63.698 (4.202)	34.428 (2.004)	C ^{γ1} , 22.132 (0.857); C ^{γ2} , 20.714 (0.992)
L9	114.010 (7.274)	176.081	51.992 (4.907)	47.627 (1.742, 1.381)	C ^γ , * (1.439); C ^{δ1} , 26.365 (0.862); C ^{δ2} , 23.051 (1.092)
D10	116.300 (8.938)	176.199	55.289 (4.609)	41.477 (2.846, 2.634)	
A11	119.670 (7.521)	175.779	51.211 (4.451)	21.645 (1.340)	
V12	123.201 (8.892)	174.977	61.182 (3.922)	32.842 (2.035)	C ^{γ1} , 20.888 (1.056, 1.000); C ^{γ2} , 22.752 (1.000)
P13		176.259	62.245 (4.559)	32.366 (2.168, 1.594)	C ^γ , 28.094 (1.739, 1.651); C ^δ , 50.528 (4.146, 3.447)
A14	122.390 (8.957)	176.235	53.953 (4.266)	19.087 (1.542)	
G15	102.379 (7.369)	173.581	44.886 (3.839, 4.636)		
T16	116.599 (8.704)	172.901	60.116 (5.319)	70.882 (4.169)	C ^{γ2} , 20.293(1.282);
I17	122.363 (6.992)	174.095	60.567 (4.265)	42.465 (1.400)	C ^{γ1} , 28.732 (0.672, 1.419); C ^{γ2} , 19.233(1.009); C ^{δ1} , 15.605(0.772)
K18	131.437 (9.121)	175.403	56.176 (4.823)	32.524 (2.058, 1.739)	C ^γ , 24.981 (1.428, 1.292); C ^δ , 29.517 (1.624, 1.624); C ^ε , 42.015 (2.926, 2.948)
V19	127.290 (8.684)	174.161	61.690 (4.595)	34.218 (1.689)	C ^{γ1} , 23.826 (0.370); C ^{γ2} , 18.927 (0.269)
I20	125.833 (8.689)	175.647	59.577 (4.658)	40.743 (1.733)	C ^{γ1} , 27.031 (1.405, 1.019); C ^{γ2} , 17.382 (0.720); C ^{δ1} , 12.586 (0.753)
Y21	127.690 (9.109)	175.646	58.832 (4.508)	40.743 (3.044, 2.590)	
G22	107.950 (8.302)	174.090	46.303 (3.870, 3.938)		
D23	120.780 (8.693)	176.989	56.038 (4.493)	40.691 (2.693, 2.595)	
D24	116.470 (8.151)	175.349	54.601 (4.475)	40.747 (2.841, 2.730)	
L25	121.770 (8.120)	174.831	54.006 (4.643)	41.650 (1.804, 1.350)	C ^γ , 26.289 (1.623); C ^{δ1} , 25.510(0.884, 0.825); C ^{δ2} , 23.976 (0.825)
E26	124.574 (8.182)	176.497	55.730 (4.944)	31.994 (1.921, 1.808)	C ^γ , 36.783 (2.407, 2.214)
V27	125.550 (8.504)	175.753	64.189 (3.996)	33.254 (2.291)	C ^{γ1} , 23.084 (1.278, 1.189); C ^{γ2} , 23.806 (1.196)
K28	133.910 (9.955)	174.504	54.465 (4.913)	33.291 (1.915, 1.444)	C ^γ , 24.331 (1.343, 1.343); C ^δ , 29.331 (1.768, 1.768); C ^ε , 42.946 (3.038, 3.001)
Q29	116.814 (8.788)	176.235	58.003 (3.838)	25.450 (2.146, 2.525)	C ^γ , 33.281 (2.661, 2.661); C ^δ , 180.872; N ^{ε2} , 110.117 (7.426, 7.112)
G30	103.800 (8.321)	174.783	45.261 (2.555, 4.216)		
N31	117.660 (8.953)	172.298	53.731 (4.591)	36.300 (3.060, 3.060)	N ^{δ2} , 113.593 (8.143, 7.055)
E32	121.080 (8.735)	176.696	55.171 (5.480)	30.458 (2.042, 1.955)	C ^γ , 36.730 (2.362, 2.132)
L33	128.250 (8.984)	176.897	53.213 (4.914)	45.658 (1.135, 1.000)	C ^γ , 27.894 (1.402); C ^{δ1} , 25.824 (0.064); C ^{δ2} , 24.500 (0.698)
T34	112.140 (8.824)	174.343	58.933 (5.024)	68.610	
P35		175.906	54.539 (4.397)	41.187 (2.616, 2.709)	C ^γ , 19.368 (1.428;1.428)
T36	121.299 (8.238)	176.149	61.043 (5.053)	69.331 (4.543)	C ^{γ2} , 21.672(1.276); H ^{γ1} , 0.985
Q37	119.560 (7.654)	178.771	58.333 (3.860)	28.438 (1.516, 1.516)	C ^γ , 34.944 (2.417, 2.417); C ^δ , 179.392 N ^{ε2} , 109.983 (7.568, 6.597)
V38	108.258 (7.082)	175.997	59.323 (4.403)	30.835 (1.851)	C ^{γ1} , 20.229 (0.761); C ^{γ2} , 18.525 (-0.755)
K39	120.874 (6.773)	175.475	58.712 (4.231)	31.857 (1.853, 1.430)	C ^γ , 22.648 (1.233, 1.306); C ^δ , 29.107 (1.146, 1.300); C ^ε , 42.015 (2.881, 2.868)
D40	121.010 (8.429)	173.455	52.717 (5.078)	43.089 (2.631, 2.466)	
Q41	119.541 (7.915)	174.114	53.334 (1.968)	30.194 (1.362, 1.362)	C ^γ , 33.513 (2.028, 2.028); C ^δ , 179.825 N ^{ε2} , 113.224 (7.662, 5.885)
P42		175.850	61.871 (4.160)	31.369 (0.996, -0.237)	C ^γ , 25.883 (1.658, 1.540); C ^δ , 50.093 (2.886, 2.381)
I43	119.100 (8.329)	175.962	61.293 (4.040)	37.935 (1.732)	C ^{γ1} , 27.429 (1.566, 1.156); C ^{γ2} , 17.339

V44	128.750 (8.290)	174.828	60.023 (5.278)	34.648 (1.833)	0.794); C ^{δ1} , 12.290 (0.814)
S45	117.850 (9.255)	173.571	57.316 (4.686)	66.304 (3.798, 3.720)	C ^{γ1} , 22.990 (1.089); C ^{γ2} , 22.752 (0.135);
W46	118.194 (7.859)	175.099	55.987 (5.311)	25.572 (3.462, 2.951)	C ^{δ1} , 126.120 (7.237); C ^{ε3} , * (7.238); N ^{ε1} , 131.852 (10.291)
S47	115.500 (8.132)	175.659	61.268 (4.332)	62.987 (3.935, 3.788)	
G48	119.777 (10.676)	175.850	45.072 (4.679, 3.829)		
L49	120.210 (8.939)	175.734	54.901 (4.663)	42.447 (2.161, 1.555)	C ^γ , 27.102 (1.780); C ^{δ1} , 27.137 (0.931); C ^{δ2} , 23.156 (0.929)
E50	123.750 (10.464)	176.358	56.741 (4.505)	29.502 (2.133, 2.225)	C ^γ , 35.046 (2.358, 2.461)
G51	104.655 (8.606)	175.227	45.959 (3.954, 4.372)		
K52	117.950 (8.235)	174.790	55.295 (4.715)	33.019 (2.125, 1.678)	C ^γ , 25.132 (1.393, 1.393); C ^δ , 28.753 (1.641, 1.641); C ^ε , 42.321 (3.043, 3.043)
S53	112.160 (7.853)	174.030	57.172 (4.869)	66.976 (4.270, 3.841)	
N54	120.205 (9.605)	177.094	56.314 (4.538)	38.785 (3.008, 3.097)	N ^{δ2} , 113.506 (8.091, 7.198)
L55	117.155 (8.393)	176.340	53.374 (5.421)	47.599 (1.382, 1.255)	C ^γ , 23.553 (0.789); C ^{δ1} , 26.088(0.669); C ^{δ2} , 26.088 (0.669)
L56	126.665 (9.422)	174.202	54.877 (5.321)	47.563 (1.500, 1.108)	C ^γ , * (1.423); C ^{δ2} , 23.681(0.790);
T57	116.595 (8.635)	171.668	62.863 (5.312)	75.020 (3.816)	C ^{γ2} , 21.551 (1.165);
L58	129.530 (9.502)	173.920	52.965 (5.536)	47.148 (1.502, 0.994)	C ^γ , 27.787 (1.158); C ^{δ1} , 26.809 (0.324); C ^{δ2} , 23.491 (1.006)
L59	122.450 (9.223)	176.418	53.009 (5.342)	48.476 (1.957, 1.399)	C ^γ , 27.102 (1.700); C ^{δ1} , 26.267 (1.087); C ^{δ2} , 23.427 (0.673)
M60	119.990 (8.619)	176.556	55.201 (5.976)	37.645 (2.213, 1.947)	C ^γ , 32.216 (1.699, 1.215)
V61	119.455 (9.210)	173.894	58.851 (5.812)	37.597 (2.126)	C ^{γ1} , 22.130 (1.109); C ^{γ2} , 17.159 (0.857)
D62	117.400 (8.777)	176.201	49.136 (6.078)	43.063 (3.080, 2.622)	
P63		175.873	63.212 (4.729)	32.694	C ^γ , 25.669 (1.566, 1.566)
D64	127.417 (9.073)	170.960	55.002 (4.663)	41.994 (2.501, 2.302)	
A65	116.134 (5.830)		48.075 (4.643)	22.677	
P66³		175.497	65.522 (5.158)	34.870 (2.721, 2.400)	C ^δ , 50.094 (4.085, 3.199); C ^γ , 25.391 (2.123, 2.014)
T67	107.900 (6.549)	173.932	59.288 (4.438)	71.290 (4.314)	C ^{γ2} , 20.964 (1.202);
R68	124.296 (8.852)	177.149	59.805 (3.772)	31.002 (1.533, 1.908)	C ^γ , 26.873 (1.723, 1.723); C ^δ , 44.438 (3.091, 3.170)
Q69	113.105 (8.181)	176.126	57.283 (4.171)	28.698 (2.161, 2.016)	C ^γ , 34.274 (2.466, 2.466); C ^δ , 180.485; N ^{ε2} , 112.430 (7.693, 6.999)
D70	116.800 (7.359)	173.250	52.181 (4.851)	40.341 (2.645, 2.409)	
P71		178.236	63.407 (4.641)	28.975 (2.512, 1.971)	C ^γ , 26.503 (1.990, 1.990); C ^δ , 48.851 (3.539, 3.459)
K72	123.980 (8.247)	177.421	58.998 (4.035)	33.216 (1.565, 1.391)	C ^γ , 24.389 (1.084, 0.923); C ^δ , 29.392 (1.512, 2.895); C ^ε , 44.897 (3.255, 3.110)
Y73	116.520 (8.711)	174.620	54.732 (5.301)	38.647 (3.482, 2.511)	
R74	120.260 (7.696)	177.546	52.128 (4.236)	32.190 (2.529, 2.151)	C ^γ , 26.984 (1.659, 2.151); C ^δ , 43.872 (3.715, 3.288)
E75	128.950 (9.720)	176.443	58.149 (4.684)	31.737 (2.400, 1.648)	C ^γ , 35.178 (2.620, 2.037)
I76	117.197 (8.926)	177.712	60.424 (5.263)	37.896 (2.204)	C ^{γ1} , 28.966 (1.932, 1.312); C ^{γ2} , 17.159(0.589); C ^{δ1} , 12.810(1.006)
L77	115.848 (8.072)	175.403	56.172 (4.835)	40.588 (1.728, 2.047)	C ^γ , 32.467 (1.433); C ^{δ1} , 24.963 (*)
H78	127.398 (8.682)	174.079	58.830 (4.041)	30.061 (3.658, 3.649)	
W79	120.257 (7.784)	174.065	54.925 (5.084)	34.722 (3.442, 3.442)	C ^{δ1} , 124.755 (8.346); C ^{ε3} , * (8.347); N ^{ε1} , 131.048 (10.610)
S80	122.080 (8.875)	173.626	57.079 (5.266)	64.006 (3.663, 3.441)	
V81	130.750 (9.348)	174.378	61.552 (4.910)	34.823 (1.851)	C ^{γ1} , 22.406 (0.971); C ^{γ2} , 20.131 (0.676)
V82	116.350 (9.144)	175.972	58.136 (5.572)	36.432 (2.188)	C ^{γ1} , 21.781 (1.150); C ^{γ2} , 17.733 (0.977)
N83	114.540 (9.074)	174.731	54.056 (4.103)	35.893 (3.622, 3.205)	N ^{δ2} , 117.471 (9.017, 8.374)
I84	122.230 (9.215)	174.978	60.307 (3.818)	39.547 (1.962)	C ^{γ1} , 29.587 (1.622, 0.954); C ^{γ2} , 16.538 (1.393); C ^{δ1} , 14.052 (0.799)
P85		178.739	64.950 (4.370)	32.149 (1.949, 2.510)	C ^γ , 28.080 (*, *)
G86	110.450 (8.215)	175.051	46.834 (4.077, 3.705)		
S87	112.510 (8.153)	173.756	58.395 (4.376)	63.374 (4.177, 4.046)	
N88	124.000 (8.405)	175.496	52.815 (4.887)	38.509 (3.595, 3.188)	N ^{δ2} , 111.378 (7.283, 7.068)
E89	117.150 (8.656)	171.421	55.524 (3.929)	30.763 (2.009, 1.731)	C ^γ , 36.140 (2.224, 2.098)
N90	114.520 (7.139)	173.133	50.926 (3.001)	39.631 (2.973, 2.541)	N ^{δ2} , 116.555 (7.742, 6.996)
P91		176.276	64.381 (4.817)	30.711 (2.357, 2.092)	C ^γ , 27.282 (2.161, 1.879); C ^δ , 50.715 (3.211, 3.739)
S92	118.550 (8.241)	175.953	60.253 (4.347)	63.628 (3.914, 3.914)	
G93	109.300 (8.763)	175.421	45.428 (3.913, 4.232)		
G94	105.031 (7.964)	172.370	44.579 (4.486, 3.466)		
H95	120.509 (9.550)	174.462	55.309 (4.937)	33.070 (3.189, 2.854)	C ^{δ2} , 116.583 (6.494)
S96	122.470 (8.590)	174.782	56.498 (4.946)	63.007 (3.831, 3.831)	

L97	129.510 (8.413)	180.082	56.498 (4.496)	40.908 (1.935, 1.478)	C^{γ} , 27.102 (1.818); $C^{\delta 1}$, 27.071 (1.174); $C^{\delta 2}$, 23.683 (1.174)
A98	122.230 (8.300)	175.210	52.052 (4.868)	19.678	
D99	124.370 (8.582)	175.946	54.621 (4.771)	41.414 (2.840, 2.478)	
Y100	122.120 (8.635)	175.091	60.812 (4.154)	38.155 (2.369, 1.221)	$C^{\delta 1}$, * (6.743); $C^{\epsilon 1}$, 118.352 (6.795)
V101	126.930 (7.572)	175.579	60.637 (4.094)	35.796 (1.725)	$C^{\gamma 1}$, 20.267 (0.921); $C^{\gamma 2}$, 22.131 (0.921)
G102	112.775 (8.092)	176.178	54.574 (2.759, 2.759)		
S103	120.071 (8.466)	177.768			
G104	131.098 (9.610)				
P106		177.901	62.287 (4.615)	32.084 (2.548, 2.136)	C^{γ} , 28.232 (2.340, 2.340); C^{δ} , 51.334 (4.576, 4.025)
K107	128.345 (8.827)	176.118	58.581 (3.322)	32.145 (1.227, 0.460)	C^{γ} , 24.764 (1.127, 0.741); C^{δ} , 29.397 (1.340, 1.340); C^{ϵ} , 42.283 (2.907, 2.907)
D108	118.570 (9.100)	175.590	56.879 (4.593)	38.909 (3.066, 2.988)	
T109	104.880 (7.595)	175.450	61.910 (4.727)	71.149 (5.274)	$C^{\gamma 2}$, 21.800 (1.327);
G110	111.940 (8.688)	173.760	44.965 (4.054, 4.303)		
L111	119.195 (8.736)	179.215	55.728 (4.431)	43.113 (1.865, 1.340)	C^{γ} , 27.102 (1.674); $C^{\delta 1}$, 25.611 (0.951); $C^{\delta 2}$, 23.407 (0.819)
H112	124.150 (9.638)	173.920	53.049 (5.187)	30.506 (3.395, 3.206)	$C^{\delta 2}$, 117.439 (7.470)
R113	122.650 (9.172)	172.603	56.010 (4.579)	32.535 (1.399, 1.399)	C^{γ} , 26.099(*, *)
Y114	116.868 (8.586)	174.793	57.705 (4.573)	38.657 (2.012, 0.987)	
I115	121.360 (8.718)	174.764	60.289 (4.770)	42.010 (1.688)	$C^{\gamma 1}$, 27.863 (0.823, 0.823); $C^{\gamma 2}$, 17.764 (0.975); $C^{\delta 1}$, 17.160 (0.743)
F116	125.728 (9.688)	174.832	56.873 (5.576)	42.088 (3.060, 2.943)	$C^{\delta 1}$, 131.581 (7.451); $C^{\epsilon 1}$, 131.678 (7.957); C^{ζ} , 131.465 (8.397)
L117	121.030 (9.374)	173.550	53.915 (4.965)	48.256 (1.939, 1.322)	C^{γ} , 26.480 (1.638); $C^{\delta 1}$, 25.620 (1.021, 0.995)
L118	122.910 (7.677)	174.251	52.873 (5.422)	47.081 (1.449, 0.776)	C^{γ} , 26.076 (0.796); $C^{\delta 1}$, 25.859(-0.149); $C^{\delta 2}$, 20.888 (-0.279)
Y119	124.205 (9.686)	176.736	57.148 (4.771)	42.516 (3.253, 2.562)	$C^{\delta 1}$, 133.536 (6.868)
R120	120.258 (9.320)	175.081	57.957 (4.430)	31.285 (1.826, 1.826)	C^{γ} , 29.015 (1.800, 1.800); C^{δ} , 43.854 (3.352, 3.120)
Q121	122.390 (7.932)	173.774	54.166 (4.587)	31.291 (1.733, 1.544)	C^{γ} , 32.171 (2.337, 2.044)
E122	126.230 (8.777)	176.618	57.444 (4.928)	31.461 (2.070, 2.070)	C^{γ} , 35.542 (2.298, 2.386)
N123	115.760 (8.201)	175.095	51.596 (5.080)	41.505 (2.796, 2.796)	$N^{\delta 2}$, 115.281 (7.814, 7.144)
K124	123.960 (8.923)	175.850	58.087 (3.620)	33.456 (1.752, 1.652)	C^{γ} , 25.792 (1.348, 1.229); C^{δ} , 29.702 (1.625, 1.625); C^{ϵ} , 42.291 (2.898, 2.898)
I125	130.160 (9.255)	175.118	60.915 (4.048)	40.614 (1.361)	$C^{\gamma 1}$, 29.517 (1.733, 0.664); $C^{\gamma 2}$, 16.636 (0.817); $C^{\delta 1}$, 13.376 (0.548, 0.548)
E126	127.295 (8.685)	176.452	57.730 (3.869)	29.871 (2.061, 1.971)	C^{γ} , 35.927 (2.292, 2.211)
E127	126.500 (8.586)	176.117	55.241 (3.694)	26.529 (0.373, 1.642)	C^{γ} , 36.760 (1.517, 2.218)
T128	119.475 (7.692)	172.066	60.911 (4.496)	68.355 (2.340)	$C^{\gamma 2}$, 21.509 (0.987)
P129		177.751	64.335 (4.532)	31.867 (2.436, 1.964)	C^{γ} , 28.092 (2.247, 2.099); C^{δ} , 50.715 (3.975, 3.716)
T130	123.835 (8.464)	175.115	64.325 (3.718)	70.997 (3.718)	$C^{\gamma 2}$, 22.304 (0.958)
I131	128.061 (8.147)	175.295	62.611 (4.192)	38.596 (1.747)	$C^{\gamma 1}$, 27.476 (2.088, 1.201); $C^{\gamma 2}$, 18.443 (1.093); $C^{\delta 1}$, 13.446(0.949)
S132	126.340 (9.119)	176.786	58.318 (4.610)	64.822 (4.423, 4.142)	
N133	122.140 (10.180)	175.497	54.147 (4.122)	37.087 (2.700, 3.325)	
T134	107.890 (7.832)	173.683	60.890 (4.532)	69.143 (5.053)	$C^{\gamma 2}$, 21.671 (1.274)
T135	119.225 (7.640)	172.774	62.367 (4.528)	70.371 (4.290)	$C^{\gamma 2}$, 22.095 (1.474)
R136	129.270 (9.382)	175.579	57.177 (4.111)	31.984 (1.777, 1.485)	C^{γ} , 28.046 (1.485, 1.485); C^{δ} , 43.671 (3.164, 3.110)
T137	117.110 (7.867)	174.623	65.749 (3.838)	68.609 (3.984)	$C^{\gamma 2}$, 21.935 (1.340*);
G138	115.166 (9.179)	176.045	44.822 (4.513, 3.844)		
R139	117.000 (8.472)	175.716	56.606 (4.501)	30.515	C^{δ} , 44.046(*, *)
L140	116.600 (7.721)	177.728	55.780 (4.485)	42.135 (1.867, 1.596)	C^{γ} , 28.665 (1.596); $C^{\delta 1}$, 24.673(0.651); $C^{\delta 2}$, 23.605 (0.441)
N141	115.560 (8.084)	175.293	55.781 (4.495)	37.130 (3.070, 2.790)	$N^{\delta 2}$, 110.243 (7.431, 7.284)
F142	121.230 (10.013)	173.434	58.661 (4.369)	40.073 (3.246, 2.836)	$C^{\delta 1}$, 132.070 (7.469); C^{ζ} , 129.218 (6.295)
N143	124.300 (6.609)	174.086	53.252 (4.740)	41.204 (2.732, 2.569)	$N^{\delta 2}$, 112.800 (7.677, 6.854)
A144	135.524 (10.976)	179.348	52.977 (3.950)	17.043	
R145	118.490 (8.631)	180.613	60.839 (4.174)	30.580 (2.175, 1.828)	C^{γ} , 29.574 (*, *)
D146	122.220 (8.589)	178.496	57.189 (4.424)	40.673 (2.667, 2.667)	
F147	124.470 (8.809)	177.489	62.188 (4.101)	40.474 (3.174, 2.988)	$C^{\delta 1}$, 134.140 (7.350)
A148	121.110 (9.031)	180.421	55.078 (3.759)	19.090	
A149	120.288 (8.109)	181.172	54.857 (4.222)	18.174	
K150	119.900 (8.354)	178.171	58.826 (3.879)	32.253 (1.695, 1.419)	C^{γ} , 24.394 (1.053, 0.718); C^{δ} , 29.171 (1.502, 1.419); C^{ϵ} , 42.003 (2.892, 2.892)
H151	113.010 (7.285)	174.747	55.933 (4.720)	30.269 (3.223, 2.285)	$C^{\delta 2}$, 120.887 (7.231); $C^{\epsilon 1}$, * (5.464)

G152	110.260 (7.686)	176.435	47.474 (4.153, 4.031)		
L153	117.261 (8.946)	176.358	55.037 (4.435)	41.446 (1.803, 1.510)	C^{γ} , 26.977 (1.509); $C^{\delta 1}$, 24.632 (0.470); $C^{\delta 2}$, 22.172 (0.882)
G154	105.900 (8.160)	174.204	45.680 (3.764, 4.107)		
E155	121.165 (8.402)	174.957	53.552 (4.816)	29.349 (2.199, 1.855)	C^{γ} , 36.423 (2.459, 2.404)
P156		176.605	63.794 (3.993)	30.764 (1.546, 1.081)	C^{γ} , 27.780 (1.985, 1.897); C^{δ} , 49.472 (3.984, 3.984)
I157	112.220 (8.706)	176.436	60.995 (4.409)	39.086 (2.080)	$C^{\gamma 1}$, 25.893 (1.062, 0.763); $C^{\gamma 2}$, 19.163 (1.013); $C^{\delta 1}$, 14.610 (0.763)
A158	120.675 (7.631)	173.992	51.140 (4.576)	23.510	
A159	118.043 (7.786)	176.719	51.218 (5.695)	23.658	
N160	113.690 (8.968)	173.301	52.511 (5.392)	40.905 (3.640, 2.594)	$N^{\delta 2}$, 110.757 (8.628, 6.864)
Y161	120.975 (9.834)	171.795	57.379 (6.124)	42.913 (3.379, 3.071)	$C^{\delta 1}$, 133.000 (7.011); $C^{\epsilon 1}$, 120.704 (6.769)
Y162	116.731 (8.762)	172.945	55.300 (5.054)	41.181 (3.121, 3.707)	
Q163	117.475 (9.085)	174.641	54.344 (6.165)	35.253 (2.253, 1.815)	C^{γ} , 36.320 (2.339, 2.085); C^{δ} , 178.689; $N^{\delta 2}$, 110.637 (7.906, 6.899)
A164	120.080 (9.296)	175.406	52.508 (5.453)	23.491	
Q165	119.620 (9.475)	173.116	52.943 (5.157)	33.506 (2.329, 1.693)	C^{γ} , 31.498 (2.375, 2.375); C^{δ} , 180.655; $N^{\delta 2}$, 115.473 (8.586, 6.915)
Y166	118.471 (8.632)	176.432	61.387 (3.942)	39.332 (2.922, 3.151)	$C^{\epsilon 1}$, 118.285 (6.817)
D167	124.956 (6.634)	175.000	52.894 (4.301)	42.319 (3.038, 3.256)	
D168	112.104 (7.917)	176.744	56.357 (4.251)	40.426 (2.676, 2.856)	
Y169	121.821 (8.579)	177.029	59.722 (4.426)	39.478 (3.004, 3.278)	$C^{\epsilon 1}$, * (6.810)
V170	120.027 (7.649)				
P171		179.192	65.658 (4.396)	31.524 (2.431, 1.836)	C^{γ} , 29.150 (2.163, 2.163); C^{δ} , 50.094 (3.722, 3.722)
I172	118.501 (6.827)	178.923	64.075 (3.613)	36.422 (1.843)	$C^{\gamma 1}$, 28.457 (1.620, 1.222); $C^{\gamma 2}$, 17.159 (0.901); $C^{\delta 1}$, 12.188 (0.957)
R173	121.107 (7.716)	180.158			
N174	117.139 (8.312)	177.448	55.304 (4.464)	37.359 (2.810, 2.963)	$N^{\delta 2}$, 111.367 (7.564, 7.311)
K175	119.155 (7.490)	177.644	58.647 (4.221)	32.793 (2.010, 2.010)	C^{γ} , 25.106 (1.768, 1.622); C^{δ} , 29.996 (1.768, 1.768); C^{ϵ} , 42.015 (3.039, 3.039)
T176	106.555 (7.530)	174.251	61.998 (4.420)	70.112 (4.492)	$C^{\gamma 2}$, 22.021(1.326);
I177	122.511 (7.192)	175.861	61.136 (4.246)	37.183 (1.886)	$C^{\gamma 1}$, 27.241 (1.725, 1.725); $C^{\gamma 2}$, 12.896(0.754); $C^{\delta 1}$, 18.108 (0.754)
V178	122.790 (8.305)	175.830	61.352 (4.515)	33.822 (2.343)	$C^{\gamma 1}$, 21.391 (0.988); $C^{\gamma 2}$, 19.020 (0.991)
G179	117.841 (8.428)	179.564	46.182 (4.368, 3.967)		

Expression, purification et étude structurale par RMN de PEBP de drosophile

Résumé

Les Phosphatidylethanolamine Binding Proteins (PEBP) sont des protéines d'environ 20 kDa conservées depuis les bactéries jusqu'aux mammifères. Elles sont caractérisées par leur capacité de fixer des ligands anioniques. Les PEBP agissent dans un grand nombre de processus cellulaires (notamment au niveau des voies de transduction du signal), et de processus physiologiques (comme la spermatogénèse) de l'organisme sain. Par ailleurs des modifications de l'expression de ces protéines sont corrélées à des situations pathologiques comme les cancers primaires et métastatiques, le diabète, la maladie d'Alzheimer et l'infertilité.

Il existe 7 PEBP chez la drosophile. Durant ce travail de thèse nous avons exprimé 3 d'entre elles chez *Escherichia coli* et mis au point les protocoles pour les purifier en grande quantité. Parmi les protéines produites, la protéine CG7054 a été choisie pour une étude structurale par RMN. Nous avons réalisé l'attribution des signaux ^1H , ^{13}C et ^{15}N du squelette peptidique et des chaînes latérales de cette protéine. L'étude de la dynamique de son squelette peptidique a été réalisée et sa structure 3D a été déterminée et analysée.

Mots clés : Résonance Magnétique Nucléaire, structure, dynamique, modélisation moléculaire, PEBP, *Drosophila melanogaster*, expression, purification

Expression, purification and NMR structural study of PEBPs from *Drosophila*

Abstract :

Phosphatidylethanolamine Binding Proteins (PEBPs) are proteins of about 20 kDa highly conserved proteins in all three kingdoms from bacteria to mammals. They are characterized by their capacity to bind anionic ligands. In addition PEBPs seem to be involved in numerous cellular processes (mostly in signalling pathways), and physiological processes (such as spermatogenesis) of the healthy organism. However modifications of the level of PEBPs expression are correlated with pathological situations like primary and metastatic cancers, diabetes, Alzheimer disease and infertility.

Seven PEBPs have been isolated in *drosophila*. During this PhD thesis three of them have been expressed in *Escherichia coli* and purification protocols have been developed to produce the protein in large quantity. One of them, CG7054, was selected for a structural study by NMR. The assignment of ^1H , ^{13}C and ^{15}N resonances was performed for the backbone and side chain atoms. The backbone dynamics was studied and the 3D structure was determined and compared to other PEBPs.

Keywords : Nuclear magnetic resonance, structure, dynamics, molecular modelling, PEBP, *Drosophila melanogaster*, expression, purification.

Discipline: Biophysique Moléculaire
Centre de Biophysique Moléculaire – CNRS – Orléans
Université d'Orléans – UFR Sciences