



**HAL**  
open science

**Identification d'une nouvelle famille de GTPase de fonction inconnue et Bases structurales de la reconnaissance antigénique par les lymphocytes T humains, Deux approches de biologie structurale par cristallographie.**

Stéphanie Gras

► **To cite this version:**

Stéphanie Gras. Identification d'une nouvelle famille de GTPase de fonction inconnue et Bases structurales de la reconnaissance antigénique par les lymphocytes T humains, Deux approches de biologie structurale par cristallographie.. Biophysique [physics.bio-ph]. Université Joseph-Fourier - Grenoble I, 2006. Français. NNT : . tel-00140708

**HAL Id: tel-00140708**

**<https://theses.hal.science/tel-00140708>**

Submitted on 10 Apr 2007

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

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

**THESE**

Présentée par

**Stéphanie Gras**

*Pour obtenir le titre de docteur  
De l'Université Joseph Fourier – Grenoble I  
(arrêtés ministériels du 5 juillet 1984 et du 30 mars 1992)*

Spécialité : Cristallographie et RMN Biologiques

**Identification d'une nouvelle famille de GTPase  
de fonction biologique inconnue  
et  
Bases structurales de la reconnaissance  
antigénique par les lymphocytes T humains  
Deux approches de biologie structurale par  
cristallographie**

Membres du jury :

<b>Rapporteurs</b>	<b>Pedro ALZARI</b> <b>Jacqueline CHERFILS</b>
<b>Examineurs</b>	<b>Marc BONNEVILLE</b> <b>Marie-Paule LEFRANC</b> <b>Jean VICAT</b>
<b>Directeur de thèse</b>	<b>Dominique HOUSSET</b>

*Thèse préparée au  
Laboratoire de Cristallographie et Cristallogénèse des Protéines  
Institut de Biologie Structurale Jean-Pierre Ebel  
CEA-CNRS-UJF*



Je tiens à remercier l'ensemble des personnes qui ont contribué à faire aboutir ce grand projet que fut ma thèse.

Merci aux membres du jury d'avoir accepté d'évaluer et de juger mon travail.

Je tiens à remercier Eva Pebay-Peyroula de m'avoir accueilli dans le DEA Cristallographie et RMN biologiques et de m'avoir orienter vers Dominique pour mon stage de DEA. Merci également à l'ensemble des personnes du LCCP de m'avoir accueilli pendant ces quatre années. Plus particulièrement je tiens à remercier Philippe, Eric T, Anne et Jean-Baptiste avec qui ce fut un plaisir de collaborer et de discuter science et autre ... Je tiens également à remercier ceux qui ont sus être là pour le boulot et pour les autres moments : Christine, Kiki, Lydie, Claudine, Fred (Junior), Constance (Potence), Eric P. Merci également à ceux qui sont passés par le labo au cours de ces années : Nicolas B., merci d'avoir partager le bureau pendant quelques mois et merci pour les soirées foot, Noémi merci pour tous les bons moments que nous avons partager ensemble et pour le Nabaztag ;-), merci également à Sonia, Nadège et Céline pour leur soutien et leur présence. Je remercie également l'ensemble des membres de l'IBS qui m'ont offert un cadre privilégié pour ma thèse, et plus particulièrement à Patricia (quelle voie et quel soutien), Laurence B, Laurence S, Delphine, Fred.M, Katy et Anne-Marie.

L'ensemble de mes résultats n'auraient pu être obtenus sans le travail de mes collaborateurs. Tout d'abord je remercie l'équipe de Jean Armengaud du CEA de Marcoule pour l'ensemble de leur travail, leur motivation et l'ensemble de nos discussions. Sur le projet PACE j'ai également travaillé avec l'équipe de Patrick Forterre de l'institut Pasteur à Paris qui est à l'initiative de ce projet et grâce à qui j'ai pu découvrir les archae. Merci également à Didier Flament et son équipe, de l'IFREMER de Brest, qui nous ont rejoint plus récemment sur ce projet. Et j'ai également collaboré pour le projet immuno avec l'équipe de Marc Bonneville de Nantes, un grand merci à toute l'équipe et plus particulièrement à Xavier, Bérangère, Klara et Marc, pour votre travail et votre écoute et aussi pour les langoustines ;-). Merci également à l'équipe de Bernard Malissen de Marseille pour l'ensemble des manip qui nous ont permis de débiter avec de bonnes bases.

Un très grand merci à Dominique pour m'avoir encadré en DEA et surtout de m'avoir gardé en thèse ;-). Merci pour ta patience, ton enthousiasme, ta passion pour la science, ta motivation pour les projets, ton encadrement très pédagogique, ta disponibilité à tout moment. Merci de m'avoir guider vers la force ... , et d'avoir tant partager avec moi aussi bien la science que le reste. Un grand merci également à Sophie, Félix, Hippolyte et Philemon pour leur soutien, leur amitié, merci d'avoir retrouver chaque page de mon manuscrit lors des corrections ...

Et mes derniers remerciements, mais non les moindres, vont à ma famille. Tout d'abord à mes parents et à mon frère merci pour votre soutien pendant l'ensemble de mes études, merci de croire en moi aussi fort et d'être là à tous moments. Merci pour les eucaryotes et les car en sac ;-). Et merci à Doudou (mon mari) qui m'accompagne dans la vie, merci d'avoir eu la patience de relire mon manuscrit !!!.

*Merci à vous d'avoir fait de ma thèse un moment si agréable et si riche, que la force vous guide ...*





# SOMMAIRE



# Projet PAB0955

## **INTRODUCTION 1**

---

1. Présentation du projet PAB0955.....	1
1.1. Les protéines PACE .....	1
1.2. Le contexte génomique des protéines de la famille PACE12.....	2
1.3. Les homologues eucaryotes des protéines PACE12.....	4
1.4. La protéine PAB0955.....	7
2. Les GTPases .....	7
3. Une classification des GTPases.....	9
3.1. La famille des interrupteurs moléculaires : domaine G avec un feuillet- $\beta$ mixte .....	9
3.1.1. Les petites protéines G .....	10
3.1.2. Les facteurs d'élongation et d'initiation (EF-Tu, EF-G, IF) .....	13
3.1.3. Les protéines G hétérotrimériques, sous unité alpha.....	15
3.1.4. Les GBP (guanylate-binding protein) .....	17
3.1.5. Les Dynamines.....	18
3.1.6. Les GTPases non classiques.....	19
3.2. La famille des SRP : domaine G avec un feuillet- $\beta$ parallèle.....	20
3.3. La famille des tubulines : sans domaine G .....	23

## **CHAPITRE I**

### **BIOCHIMIE ET CRISTALLISATION 26**

---

I.1. Biochimie de PAB0955.....	26
I.1.1. Expression et purification de PAB0955.....	26
I.1.2. Degré d'oligomérisation de PAB0955.....	27
I.1.3. Activité GTPase de PAB0955 .....	27
I.1.4. Repliement de PAB0955 .....	28
I.2. Cristallisation de PAB0955 .....	29
I.2.1. Principe de la Cristallisation .....	30
I.2.1.1. Diagramme de phase .....	30
I.2.1.2. Technique de la goutte suspendue .....	31

I.2.2. Cristallisation de PAB0955 .....	32
I.2.2.1. Cristallisation de la protéine Native et Apo .....	32
I.2.2.2. Cristallisation de la protéine avec des sels d'atomes lourds .....	34
I.2.2.3. Cristallisation de la protéine en complexe avec des nucléotides .....	36

## CHAPITRE II

### RESOLUTION DES STRUCTURES

**40**


---

II.1. L'expérience de diffraction .....	40
II.1.1. Le phénomène de diffraction.....	40
II.1.2. Le traitement des données cristallographiques .....	39
II.1.2.1. Indexation des taches de diffraction .....	41
II.1.2.2. Intégration des intensités mesurées.....	42
II.1.2.3. Mise à l'échelle.....	42
II.1.2.4. Quelques facteurs de qualité.....	43
II.2. La détermination de la structure de PAB0955.....	44
II.2.1. Utilisation du signal de différence isomorphe.....	45
II.2.1.1. Enregistrement d'un jeu de données sur un cristal natif .....	45
II.2.1.2. Enregistrement d'un jeu de données sur un cristal dérivé.....	46
II.2.2. Utilisation du signal anomal .....	48
II.2.2.1. Principe du signal anomal .....	48
II.2.2.2. Tentative de phasage par la méthode MAD .....	50
II.2.2.3. Réussite du phasage par la méthode SIRAS .....	53
II.2.3. Construction de la structure de PAB0955 .....	54
II.2.4. Le remplacement moléculaire .....	55
II.3. Les structures de PAB0955 .....	58
II.3.1. La structure de la protéine sans nucléotide .....	58
II.3.1.1. Description de la structure .....	58
II.3.1.2. Description de l'interface dimérique .....	60
II.3.2. Les structures de PAB0955 avec des nucléotides tri-phosphate .....	61
II.3.2.1. PAB0955 co-cristallisée avec du GTP $\gamma$ S .....	61
II.3.2.2. PAB0955 co-cristallisée avec du GTP à 18°C .....	62
II.3.2.3. PAB0955 co-cristallisée avec du GTP à 4°C.....	63
II.3.2.4. PAB0955 co-cristallisée avec du GTP et de l'EDTA.....	64
II.3.3. Les structures de PAB0955 avec des nucléotides di-phosphate .....	64
II.3.3.1. PAB0955 co-cristallisée avec du GDP .....	64
II.3.3.2. PAB0955 co-cristallisée avec du GDP et additifs .....	65

## CHAPITRE III

## ANALYSE STRUCTURALE

67

III.1. Comparaison des structures de PAB0955 .....	67
III.1.1. Superposition des différentes structures de PAB0955 .....	67
III.1.1.1. Méthode de superposition .....	67
III.1.1.2. Résultat de la superposition .....	68
III.1.2. Analyse des changements structuraux.....	69
III.1.2.1. Le motif G1 : la P-loop .....	69
III.1.2.2. Le motif G3 : <sup>101</sup> DTPGQ <sup>105</sup> .....	72
III.1.2.3. Le motif G5 : <sup>223</sup> SAK <sup>225</sup> .....	74
III.1.2.4. La lysine 40 .....	75
III.1.2.5. Le motif <sup>65</sup> GPN <sup>67</sup> .....	76
III.1.3. Les différentes conformations de l'insertion I2.....	78
III.1.4. La rigidité du site nucléotidique .....	80
III.1.4.1. La conformation arquée du GTPγS .....	80
III.1.4.2. La position particulière du GTP par rapport au GDP .....	81
III.1.5. Le film de l'hydrolyse du GTP .....	82
III.2. Recherche d'homologues structuraux et analyse.....	86
III.2.1. Recherche d'homologues structuraux par le serveur DALI.....	86
III.2.1.1. La déthiobiotine synthétase d' <i>Escherichia coli</i> .....	87
III.2.1.2. La SRP de <i>Thermus aquaticus</i> .....	90
III.2.1.3. L'arsenite d' <i>Escherichia coli</i> .....	92
III.2.1.4. La N10-Formyltetrahydrofolate Synthétase de <i>M. thermoacetica</i> .....	92
III.2.1.5. La petite protéine G Ras de l'homme .....	92
III.2.1.6. Informations issues de DALI .....	94
III.2.2. Recherche d'homologues structuraux dans la PDB.....	96
III.2.3. Recherche d'homologue pour l'insertion I2 .....	97
III.3. Identification d'une nouvelle famille .....	98
III.3.1. L'interface dimérique .....	99
III.3.2. Conservation des motifs G .....	100
III.3.3. Conservation du motif spécifique GPN.....	101

## CHAPITRE IV

## LES MUTANTS DE PAB0955

103

IV.1. Le choix des différents mutants.....	103
IV.1.1. La glutamine 105.....	104
IV.1.2. La lysine 40.....	104
IV.1.3. L'asparagine 67.....	104
IV.1.4. L'aspartate 36.....	105

IV.2. Le comportement des mutants en solution.....	105
IV.2.1. Le mutant Q105A en solution.....	106
IV.2.2. Les mutants K40A et K40R en solution.....	106
IV.2.3. Les mutants N67A et N67D en solution.....	106
IV.2.4. Les mutants D36A et D36N en solution.....	107
IV.3. Structure et analyse des mutants.....	107
IV.3.1. La cristallisation des mutants .....	107
IV.3.2. Enregistrement cristallographique des mutants.....	109
IV.3.3. Analyse structurale des mutants .....	110
IV.3.3.1. Analyse structurale des mutants D36A et D36N .....	112
<i>La structure du mutant D36A avec du GTP</i> .....	112
<i>La structure du mutant D36N avec du GTP</i> .....	113
<i>La structure du mutant D36A avec du GDP</i> .....	115
IV.3.3.2. Analyse structurale des mutants N67A, N67D et K40R.....	116
<i>Les structures du mutant N67A avec du GTP et du GDP</i> .....	116
<i>La structure du mutant N67D avec du GDP</i> .....	117
<i>La structure du mutant K40R avec du GDP</i> .....	118

## CHAPITRE V

### RECHERCHE DE PARTENAIRES

119

---

V.1. Les protéines de <i>Pyrococcus abyssi</i> .....	119
V.1.1. Expérience de pull-down.....	119
V.1.2. Résonance de plasmon de surface.....	120
V.1.3. Le rôle de ces partenaires.....	122
V.1.3.1. La topoisomérase VI.....	122
V.1.3.2. La DnaG primase.....	123
V.2. Les protéines impliquées dans la réplication de l'ADN.....	124
V.3. Test de phosphorylation.....	

125

### CONCLUSION ET PERSPECTIVES

127

---

# Projet Immunologie

## INTRODUCTION 131

---

1. Le système immunitaire .....	131
2. Les récepteurs des lymphocytes T (TCR) .....	132
2.1. Les réarrangements des gènes des TCR .....	133
2.2. La structure du TCR .....	135
3. Les protéines CMH .....	137
3.1. Les gènes du CMH.....	137
3.2. La structure des protéines CMH .....	138
4. Les peptides antigéniques .....	140
4.1. Les peptides présentés par les molécules CMH de classe I.....	140
4.2. Les peptides présentés par les molécules CMH de classe II .....	141
5. La rencontre des acteurs de la réponse immunitaire.....	141
5.1. Comment les $L_T$ différencient les peptides du soi et ceux du non-soi ?.....	141
5.2. Le complexe ternaire : TCRpCMH.....	142
6. Quelques définitions.....	143
7. But du projet.....	145

## CHAPITRE I

### LE HLA-A2 ET LES DIFFERENTS PEPTIDES 147

---

I.1. Le choix des partenaires moléculaires .....	147
I.1.1. La molécule CMH de classe I : HLA-A2m .....	147
I.1.2. Le choix des peptides .....	148
I.2. Cristallisation des complexes pCMH .....	149
I.2.1. Les premiers essais de cristallisation .....	150
I.2.2. Notre premier cristal de HLA-A2m .....	151
I.2.3. La technique d'ensemencement .....	151



**CHAPITRE II**  
**LES STRUCTURES DE COMPLEXES**  
**BINAIRES** **155**

---

II.1. La résolution des structures de complexes binaires .....	155
II.2. La structure des complexes peptide/HLA-A2m .....	156
II.3. L'analyse des structures .....	160
II.3.1. Le postulat de départ est-il vérifié ? .....	160
II.3.2. Une nouvelle observation : la stabilité du peptide .....	162
II.3.2.1. La densité électronique de pp65 .....	163
II.3.2.2. La densité électronique de bmlf1 .....	163
II.3.2.3. La densité électronique des peptides hcv1, hcv2 et hcv3 .....	164
II.3.3. Comparaison avec les autres structures de peptide/HLA-A2m .....	166
II.3.3.1. Comparaison de bmlf1 et de melanA Ag3 .....	167
II.3.3.2. Comparaison de pp65 et SL9 .....	168
II.3.3.3. Comparaison de hcv1 et pp65 .....	170
II.3.3.4. Comparaison de hcv2 et melanA (décapeptide) .....	170
II.3.3.5. Comparaison de hcv3 et melanA (nonapeptide) .....	171

**CHAPITRE III**  
**LES COMPLEXES TERNAIRES** **172**

---

III.1. Le choix des TCR : les tris multimériques pCMH .....	172
III.2. La production de TCR .....	173
III.3. Les premiers essais de cristallisation .....	174
III.3.1. Les premiers cristaux .....	175
III.3.2. Que contiennent ces cristaux ? .....	176
III.3.3. La structure de l'homodimère de V $\alpha$ .....	177

**CONCLUSIONS ET PERSPECTIVES** **178**

---

**BIBLIOGRAPHIE** **180**

---

**ARTICLE** **193**

---





## Abréviations

ADN	Acide désoxyribonucléique
ARNm	Acide ribonucléique messenger
ARNt	Acide ribonucléique de transfert
ATP	Adenosine triphosphate
ATPase	ATP hydrolase
CDR	Complementarity determining regions
CMH	Complexe Majeur d'Histocompatibilité
DTT	1,4-Dithiothreitol
E.S.R.F	European Synchrotron Radiation Facilities
EDTA	ethylenediaminetetraacetic acid
GAP	GTPase Activating Protein
GBP	Guanylate Binding Protein
DTPA-BMA	acide diéthylènetriamine-pentaacétique-bis(méthylamide)
GEF	Guanin nucleotide Exchange factor
GMPppp	Guanosine Mono Phosphate phosphoaminophosphoric acid
GDP	Guanosine di-phosphate
GTP	Guanosine tri-phosphate
GTPase	GTP hydrolase
GTP- $\gamma$ -S	Guanosine 5'-(3-O-thio)triphosphate
HLA	Human Leukocyte Antigen
MAD	Multiple Wavelength Anomalous Dispersion
NER	Nucleotide Exision Repair
PACE	Protéines d'archaea conservées chez les eucaryotes
pCMH	complexe entre le peptide et la molécule CMH
PCNA	Proliferating Cell Nuclear Antigen
PEG	Poly éthylène glycol
RF-C	Replication factor Complexe
RMN	Résonance Magnétique nucléaire
rmsd	root mean square difference
SAD	Single wavelength anomalous diffraction
SRP	Signal Recognition Particule
TCR	T Cell Receptor
V <sub>m</sub>	Coefficient de Matthews



INTRODUCTION

DU PROJET

PAB0955



J'ai rejoint le projet PACE lors de mon stage de DEA. À mon arrivée au laboratoire plusieurs essais de cristallisation avaient été effectués sur la protéine PAB0955, sans succès. La mise en évidence de l'hétérogénéité en solution de PAB0955 venait d'être fait lorsque j'ai débuté le projet. La résolution de ce problème nous a permis de cristalliser cette protéine.

## 1. Présentation du projet PAB0955

### 1.1. Les protéines PACE

Les protéines PACE (protéines d'archaea conservées chez les eucaryotes) ont été mises en évidence par une approche génomique effectuée par l'équipe de Patrick Forterre, de l'Institut Pasteur à Paris (Matte-Tailliez et al., 2000) (<http://www-archbac.u-psud.fr>). Ces protéines de fonction biologique inconnue sont particulièrement intéressantes car, d'une part elles sont conservées chez les archaea et les eucaryotes malgré 3 milliards d'années de divergence, et d'autre part, elles n'ont pas d'homologue bactérien. Cette conservation entre les deux règnes archaea et eucaryotes, est souvent retrouvée pour des protéines impliquées dans des processus fondamentaux comme la réplication, la transcription ou la traduction de l'ADN (Olsen and Woese, 1997).

Ces processus fondamentaux chez les archaea ressemblent beaucoup à ceux retrouvés chez les eucaryotes, mais ils en sont une copie simplifiée, pour exemple le complexe RF-C et le protéasome.

Le complexe du facteur de réplication RF-C sert à charger l'anneau de processivité PCNA (Proliferating Cell Nuclear Antigen) sur l'ADN. Ceci permet aux ADN polymérase de répliquer l'ADN. Le complexe RF-C est de 5 sous-unités (une grande et 4 petites) arrangées sous forme d'anneau.

Chez les eucaryotes les petites sous-unités sont formées par 4 protéines différentes nommées RFC-A à RFC-D (Bowman et al., 2004). Chez les archaea les petites sous-unités sont formées par la même protéine, elles conservent la conformation en anneau comme chez les eucaryotes (Henneke et al., 2002).

Un autre exemple connu de cette simplification chez les archaea est celui de la machinerie de dégradation des protéines : le protéasome. Ce complexe est composé de 2 anneaux- $\alpha$  externes et deux anneaux- $\beta$  internes, chaque anneau étant composé de 7 sous-unités. Chez les eucaryotes, les 7 sous-unités de chaque anneau sont constituées par des protéines différentes



soit 14 protéines au total (Coux et al., 1996). Chez les archaea, chaque anneau est constitué de 7 protéines identiques, soit seulement deux protéines différentes répétées chacune 14 fois (Dahlmann et al., 1991).

Ces ressemblances d'organisation, entre les deux règnes, nous laissent supposer que nos résultats structuraux et fonctionnels seront directement applicables aux homologues eucaryotes et plus précisément à l'homme. Nous nous sommes donc intéressés aux protéines d'archaea qui ayant une organisation simplifiée facilite la compréhension du mécanisme. Les résultats pourront par la suite servir à l'étude du mécanisme chez l'homme, facilitant la mise en évidence des différences entre les deux règnes. Sur le plan expérimental, les protéines d'archaea que nous avons étudiées présentent un autre intérêt, elles sont issues de *Pyrococcus abyssi* un organisme hyperthermophile. Elles sont donc très stables. Ceci facilite les essais de cristallisation car la solution de protéine reste homogène.

L'équipe de Patrick Forterre a identifié à ce jour 36 protéines PACE. Nous nous sommes intéressés à la protéine de *Pyrococcus abyssi* produit du gène *PAB0955*, membre de la famille de protéines nommée PACE12.

## [1.2. Le contexte génomique des protéines de la famille PACE12](#)

La protéine PAB0955 possède un contexte génomique très intéressant au sein des archaea, d'où notre intérêt particulier pour cette protéine parmi les PACE. En effet chez les archaea les gènes des protéines PACE12 sont situés à proximité de gènes codant pour des protéines impliquées dans la réplication de l'ADN ou dans la division cellulaire (fig.1.).

La conservation de l'ordre des gènes peut mener à une interaction physique de leur produit. Cette observation décrite par Thomas Dandekar (Dandekar et al., 1998) permet de prédire la fonction, l'interaction avec des partenaires pour le produit de ces gènes.

Puisque l'organisation des gènes est conservée chez les archaea, nous pouvons supposer qu'une interaction physique existe entre ces protéines. D'un côté, la protéine PAB0955 de *Pyrococcus abyssi* est proche des gènes des protéines MinD (division cellulaire) (Sakai et al., 2001) et de l'autre de GinS 2-3 (Marinsek et al., 2006) et du Mini Chromosome Maintenance MCM (réplication de l'ADN) (Pape et al., 2003). Sur la base de l'hypothèse ci-dessus nous

Introduction du projet PAB0955

supposons que PAB0955 pourrait être impliquée à l'interface entre la réplication de l'ADN et la division cellulaire.

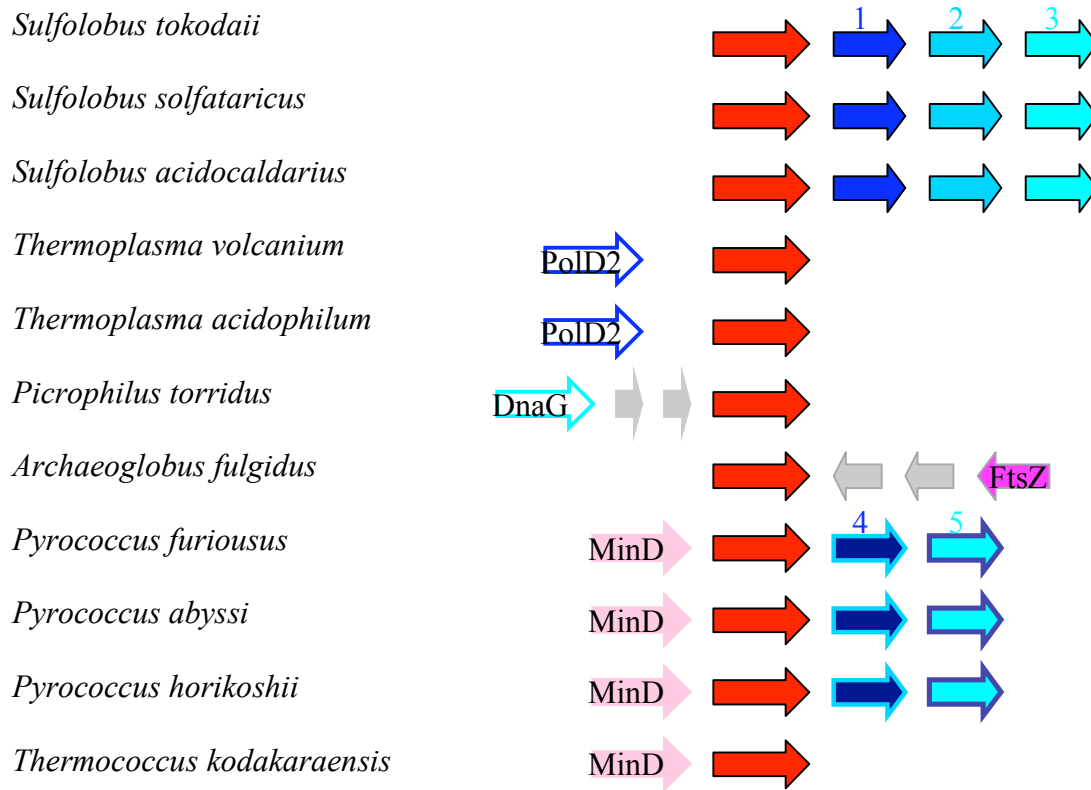


Figure.1. Contexte génomique des protéines de la famille PACE12.

Les gènes codant pour les protéines PACE12 sont en rouge, en bleu les gènes dont les protéines sont impliquées dans la réplication, en rose ceux impliqués dans la division cellulaire, et en gris pour d'autres protéines.

Les gènes notés 1 codent pour le PCNA, ceux notés 2 pour la primase PriS, ceux notés 3 pour Gins1-5, ceux notés 4 pour Gins-2-3 et les gènes notés 5 codent pour le MCM.

### 1.3. Les homologues eucaryotes des protéines PACE12

Il existe trois homologues des PACE12 chez les eucaryotes, et aucun homologue dans les 311 génomes bactériens séquencés à ce jour. Les trois homologues humains sont de taille différente : 374 résidus (Q9HCN4), 310 résidus (Q9NUE1) et 284 résidus (Q8NEI2). Parmi ces trois protéines, Q9HCN4 est la seule qui a été étudiée.

En 2000, Masahiko Nitta et ses collaborateurs ont nommé cette protéine Q9HCN4, XAB1, pour XPA binding protein 1. La protéine XAB1 interagit XPA (Xeroderma pigmentosum group A), cette dernière joue un rôle central dans la réparation de l'ADN par excision de nucléotide (NER) (Nitta et al., 2000). Par des expériences de double-hybride, les auteurs ont montré que XAB1 se lie à la partie N<sub>ter</sub> de XPA, et entraîne la localisation nucléaire de XPA. La protéine XAB1 qui est cytoplasmique (déterminé par immuno-fluorescence), servirait donc de transporteur à XPA. De plus, XAB1 présente une activité spécifique d'hydrolyse du GTP.

En 2003, Francesca Lembo et ses collaborateurs ont identifié une protéine interagissant avec MBD2 (mCpG-binding domain 2) qu'ils ont nommée MBDin (MBD interacting protein). Cette protéine s'est révélée être identique à XAB1 (Lembo et al., 2003). C'est également par des techniques de double hybride que cette interaction MBD2-MBDin a été identifiée. MBD2 est une des 5 protéines qui forme le complexe MeCP1 (methyls-CpG-binding proteins) capable de détecter les méthylations sur l'ADN. La protéine MBD2 dans ce complexe est responsable de la répression de la transcription des gènes fortement méthylés. Ces auteurs ont montré que XAB1/MBDin, XPA et MBD2 interagissaient ensemble. Ils proposent donc que ces trois protéines appartiennent à un même complexe, il y aurait donc un lien entre la méthylation des gènes et la réparation NER.

L'identité de séquence varie entre les différents homologues de la famille des protéines PACE12, elle est de 80% entre les génomes de *Pyrococcus*, de 30% avec les autres archaëa et de 20% avec les eucaryotes. Nous avons pu remarquer dans l'alignement de séquence que la conservation était parfaite pour certains motifs (fig.2). Ces motifs sont au nombre de 5 et sont séquentiellement conservés dans les 66 protéines PACE12 identifiées aujourd'hui. Nous verrons par la suite qu'ils sont importants pour la fonction enzymatique de PAB0955, et qu'ils forment également la signature de cette famille.

Introduction du projet PAB0955

									10	20				
Q9UYR9_PAB0955	-----	-----	-----	-----	-----	-----	-----	-----	MIVV	FV	GTAGSGKT	TLTGE	FGRYLE	D-
Q9HCN4_XAB1_Hu	-----	-----	-----	-----	-----	-----	-----	-----	MAAS	AAAAEL	QASGG	PRHPV	CLLV	LG
Q8VCE2_XAB1_Mo	-----	-----	-----	-----	-----	-----	-----	-----	MAAP	VAPSEP	QASRAP	QPPV	CLLV	LG
Q6FNF1_CANGA	-----	-----	-----	-----	-----	-----	-----	-----	MAL	STVICI	GMAGSGKT	TFMQ	RLNS	HIRSK
Q9NUE1_HUMAN	-----	-----	-----	-----	-----	-----	-----	-----	MAGA	APT	TAFGQ	AVI	GPPG	SGKT
Q8VEJ1_MOUSE	-----	-----	-----	-----	-----	-----	-----	-----	MAGA	APT	TAFGQ	AVI	GPPG	SGKT
Q9FMR4_ARATH	-----	-----	-----	-----	-----	-----	-----	-----	MV	FQVVI	GPPG	SGKT	TYC	NGMSQ
Q6BKW4_DEBHA	-----	-----	-----	-----	-----	-----	-----	-----	M	FQIV	I	GPPG	SGKT	TYC
Q8NEI2_HUMAN	-----	-----	-----	-----	-----	-----	-----	-----	M	PRYA	QLVM	G	PAGS	GKST
Q6CQA6_KLULA	MLDD	ILRH	CLQE	CQA	AKFIS	PNTK	STTIN	QMSR	VGV	LV	LPAGAGK	STFC	NAI	IISHMQ
Consensus											v	GpaGsGKt	T	c
														l
		30	40	50	60	70	80							
Q9UYR9_PAB0955		NYKV	VAV	VNLD	TGV	KEL	PY	-EPS	IDV	REF	VTV	VEE	IM	RE---
Q9HCN4_XAB1_Hu		GT	PPY	VINLD	PAV	HEV	PF	-PAN	IDIR	D	TV	KY	KE	VM
Q8VCE2_XAB1_Mo		GC	PPY	VINLD	PAV	HEV	PF	-PAN	IDIR	D	TV	KY	KE	VM
Q6FNF1_CANGA		KE	VPY	VINLD	PAV	LR	VPY	-GAN	IDIR	D	S	I	KY	KV
Q9NUE1_HUMAN		G	RRV	VNLD	PAN	E	GLPY	-E	CAV	DV	G	EL	V	G
Q8VEJ1_MOUSE		G	RRV	VNLD	PAN	D	GLPY	-E	CAV	DV	G	EL	V	G
Q9FMR4_ARATH		G	RK	V	VNLD	PAN	D	ALPY	-E	CGV	N	E	E	L
Q6BKW4_DEBHA		G	R	K	L	C	I	VNLD	PAN	D	ALPY	P	D	C
Q8NEI2_HUMAN		N	R	S	V	Q	V	VNLD	PAA	E	H	F	N	Y
Q6CQA6_KLULA		G	R	R	A	H	V	VNLD	PAA	E	A	T	K	Y
Consensus		gr	v	v	NLD	p	a	l	py	e	i	d	i	r
		90	100	110	120	130								
Q9UYR9_PAB0955		NEY	L	N	K	I	L	R	L	E	K	E	N	D
Q9HCN4_XAB1_Hu		D	Q	V	M	F	I	E	K	A	Q	N	S	K
Q8VCE2_XAB1_Mo		D	Q	V	M	F	I	E	K	A	Q	N	T	F
Q6FNF1_CANGA		D	Q	V	I	K	L	V	E	K	R	D	H	D
Q9NUE1_HUMAN		D	W	L	R	A	K	L	D	P	L	-R	G	-
Q8VEJ1_MOUSE		D	W	L	R	A	K	L	E	P	L	-R	G	-
Q9FMR4_ARATH		D	W	L	E	S	K	L	K	P	L	L	K	D
Q6BKW4_DEBHA		D	A	F	I	S	K	I	N	K	L	V	E	R
Q8NEI2_HUMAN		D	W	L	E	N	C	L	G	H	V	E	D	-
Q6CQA6_KLULA		D	W	L	D	E	E	I	G	D	Y	N	D	-
Consensus		d	ki	l	y	l	f	D	P	G	Q	E	l	f
		140	150	160	170	180	190							
Q9UYR9_PAB0955		K	K	P	N	D	Y	C	F	V	R	F	A	L
Q9HCN4_XAB1_Hu		T	N	P	V	T	F	M	S	N	M	L	Y	A
Q8VCE2_XAB1_Mo		T	N	P	V	T	F	M	S	N	M	L	Y	A
Q6FNF1_CANGA		S	S	P	T	T	F	M	S	N	M	L	Y	A
Q9NUE1_HUMAN		T	D	P	A	F	I	S	V	L	C	T	S	L
Q8VEJ1_MOUSE		T	D	P	A	F	I	S	V	L	C	T	S	L
Q9FMR4_ARATH		C	D	P	G	N	Y	V	S	L	L	S	L	S
Q6BKW4_DEBHA		T	S	P	S	Q	Y	I	S	I	L	L	S	L
Q8NEI2_HUMAN		V	E	S	F	K	F	I	S	G	I	L	A	L
Q6CQA6_KLULA		I	D	T	S	K	F	F	S	G	A	L	S	A
Consensus		p	f	s	l	a	l	s	m	l	l	l	p	v

Introduction du projet PAB0955

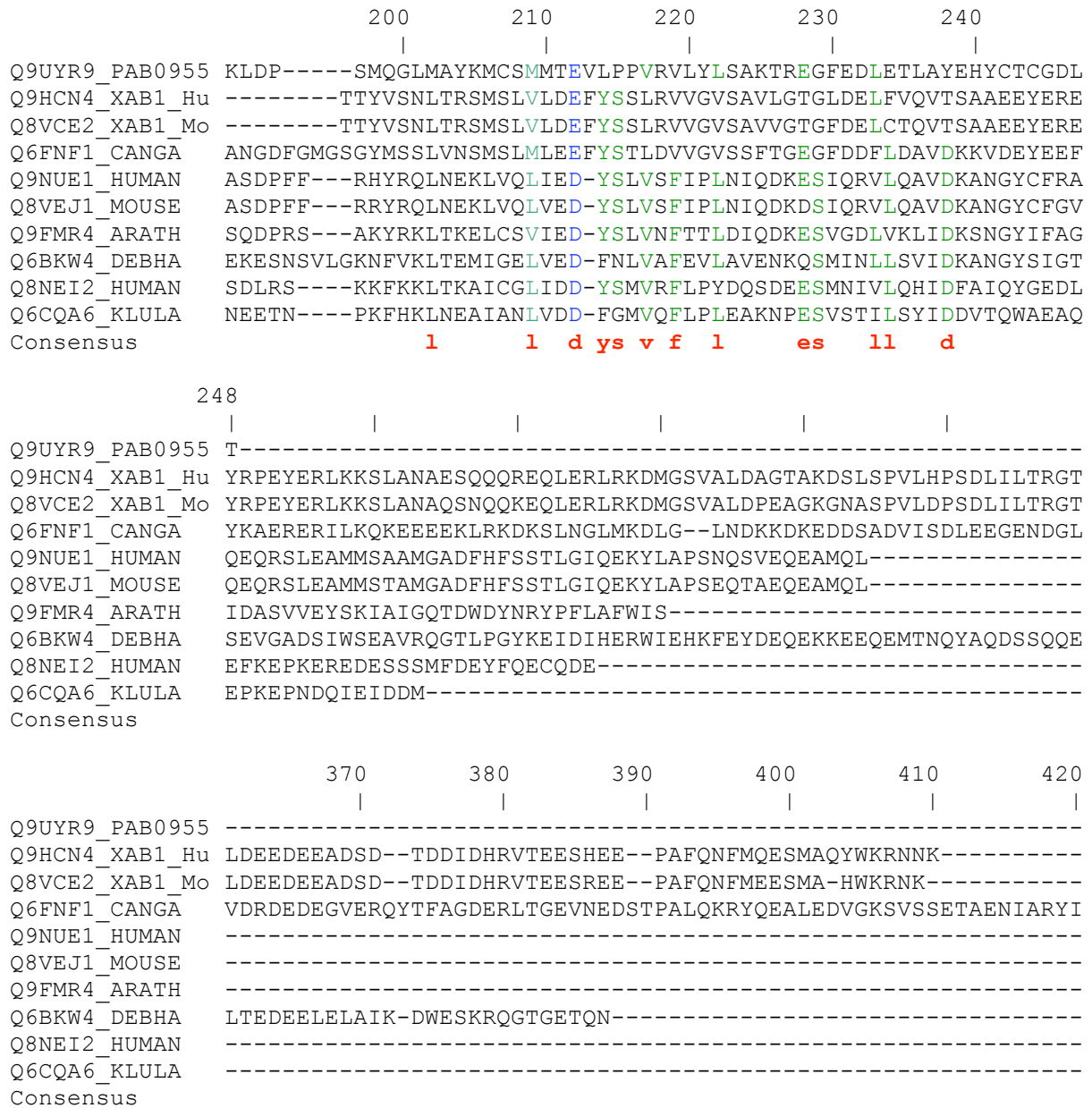


Figure.2. Alignement de séquence de 10 protéines de la famille PACE12.

En rouge sont représentés les acides aminés strictement conservés, en bleu et vert les résidus homologues conservant la même nature. En jaune sont surlignés les motifs conservés dans cette famille de protéine. PAB0955, que nous avons étudiée, est ici alignée avec 9 autres protéines eucaryotes de la famille PACE12, les trois humaines, deux de souris, une de *Candida glabrata* (CANGA), une d'*Arabidopsis thaliana* (ARATH), une de *Debaryomyces hansenii* (DEBHA) et une de *Kluyveromyces lactis* (KLULA).

De plus, la délétion systématique de chacun des trois gènes codant pour les protéines de la famille PACE12 chez la levure *Saccharomyces cerevisiae* a été testée. Leur délétion est létale, soulignant l'importance de ces protéines au niveau cellulaire (<http://mips.gsf.de/genre/proj/yeast>). Chez cette levure, les gènes sont Yjr072c homologue de Q6FNF1, Yor262w homologue de Q6BKW4 et YLR243w homologue de Q6CQA6.

#### 1.4. La protéine PAB0955

La protéine PAB0955 contient une séquence reconnue pour fixer les nucléotides tri-phosphates tel que l'ATP (adénosine tri-phosphate) ou le GTP (guanosine tri-phosphate) en N<sub>ter</sub>. Cette séquence est un motif de Walker de type A GX<sub>1</sub>X<sub>2</sub>X<sub>3</sub>X<sub>4</sub>GK[T/S], nommée aussi P-loop (phosphate binding loop), qui sert à fixer les groupements phosphates des nucléotides (Walker et al., 1982).

La similitude de séquence (environ 10% d'identité) entre les protéines de la famille PACE12 et les petites protéines G a fait que ces protéines ont été classées dans le COG1100 (Natale et al., 2000). Le COG 1100 (cluster of orthologous groups of proteins) regroupe différentes protéines hydrolysant le GTP (GTPases) dont Sar1, une petite protéine G. L'activité d'hydrolyse spécifique du GTP par PAB0955 a été vérifiée en solution, confirmant sa qualité de GTPase. Afin de mieux situer cette nouvelle GTPase, qui potentiellement peut être une petite protéine G, l'introduction est consacrée aux différentes GTPases dont la structure est connue.

## 2. Les GTPases

Les GTPases sont des enzymes capables de fixer et d'hydrolyser le GTP. La plupart des GTPases (exception faite de la tubuline et de son homologue bactérien FtsZ) conservent une organisation tridimensionnelle semblable, avec un cœur formé par un domaine G qui fixe les nucléotides et d'éventuelles insertions greffées sur ce cœur.

Le domaine G, qui caractérise la plupart des GTPases, forme le site de fixation et d'hydrolyse des nucléotides (Paduch et al., 2001). Il s'arrange en repliement de type Rossmann, c'est-à-dire avec un feuillet- $\beta$  entouré d'hélices- $\alpha$ . Ce domaine G est une alternance de brins- $\beta$ , classiquement au nombre de 6, et d'hélices- $\alpha$  qui sont généralement au nombre de 5. En plus

de sa conservation structurale, le domaine G possède également une conservation de séquence pour 5 motifs nommés G1 à G5 qui forment des boucles. Le motif G1 forme la P-loop qui fixe les phosphates du nucléotide, sa séquence est celle d'un motif de Walker de type A, elle est située entre le brin- $\beta$ 1 et l'hélice- $\alpha$ 1 c'est-à-dire en N<sub>ter</sub> du domaine G. Le motif G2 sert à fixer le Mg<sup>2+</sup> nécessaire à l'hydrolyse du GTP et selon la famille de GTPases, c'est un résidu thréonine ou acide aspartique conservé. Le motif G3 fixe le phosphate- $\gamma$  du GTP et le Mg<sup>2+</sup>, sa séquence, DX<sub>1</sub>X<sub>2</sub>G, correspond à un motif de Walker de type B (Walker et al., 1982). Ces trois premiers motifs G sont situés du même côté que les groupements phosphates du nucléotide. Alors que les motifs G4 et G5 servent à fixer la base guanine, et donnent la spécificité pour le GTP et le GDP aux protéines qui les portent. Le motif G4 possède une séquence consensus NKXD. Le motif G5 n'est pas présent dans toutes les GTPases, il est plus difficile à repérer en séquence car il peut être formé uniquement d'une alanine, qui est le plus souvent précédé d'une cystéine ou d'une sérine.

Les GTPases sont impliquées dans diverses fonctions : transduction de signal, biosynthèse de protéines, contrôle et différenciation cellulaire, translocation de protéines à travers la membrane et le transport de vésicules. Elles jouent un rôle vital pour les cellules.

Afin de mieux comprendre et d'analyser la structure de PAB0955, il était important de faire un point sur les différentes GTPases et leur structure. En effet, la relation qui existe entre la conservation de la structure et celle de la fonction peut permettre de comprendre la fonction ou, du moins, le type de voie biologique dans laquelle est impliquée la protéine d'intérêt. Nous espérons donc qu'en déterminant la structure tridimensionnelle de PAB0955 et en l'analysant, nous obtiendrons des informations sur sa fonction biologique. Lors de cette étude, nous avons classé les GTPases en trois sous-familles en fonction de la conformation de leur domaine G. Dans la littérature, les revues traitant des GTPases concernent le plus souvent un seul type de sous-famille, les petites protéines G (Paduch et al., 2001), les interrupteurs moléculaires (Vetter and Wittinghofer, 2001, Kaziro et al., 1991, Sprang, 1997) ou les SRP (Keenan et al., 2001), mais jamais l'ensemble des GTPases en détail.

Dans une étude publiée en 2002, un découpage des GTPases en deux sous-familles a été fait par Detlet D. Leipe et al. (Leipe et al., 2002), sur la base de la séquence dans le but de comprendre l'évolution de ces protéines. La première classe de protéines nommée TRAFAC (after translation factor) inclut les facteurs de translation, les petites protéines G, les dynamines, les GBP (Guanylate Binding Protein), les protéines G hétérotrimériques et

d'autres GTPases dont la structure n'est pas connue. La seconde classe nommée SIMIBI (after signal recognition particle, MinD, and BioD) regroupe des ATPases comme MinD et la déthiobiotine syntéthase (dont nous reparlerons dans l'analyse structurale), et également des GTPases comme les SRP (Signal Recognition Particle). Dans cette classification nous avons remarqué que la famille PACE12 était représentée par un de ses homologues humain XAB1. La protéine XAB1 est classée dans la famille SIMIBI. Ceci est en désaccord avec la classification des protéines de la famille PACE12 dans le COG1100 (les petites protéines G étant classées dans TRAFAC).

L'étude de Detlet D. Leipe a été réalisée sur la base de la séquence et ne considère pas toutes les GTPases, puisque les tubulines n'apparaissent pas dans ces deux sous-familles. C'est pourquoi, nous avons choisi un autre découpage en 3 sous-familles de GTPases sur la base de la structure. La première sous-famille inclut les protéines possédant un domaine G de type petite protéine G, avec un feuillet- $\beta$  mixte, que nous nommerons la famille des interrupteurs moléculaires. La seconde regroupe les protéines possédant un domaine G de type SRP, avec un feuillet- $\beta$  parallèle, et regroupe uniquement les SRP et leurs récepteurs. Elle se nommera donc la famille des SRP. La troisième famille sera celle des tubulines, elle regroupe les tubulines et les protéines FtsZ (homologue bactérien) qui ne possèdent pas de domaine G.

Cette classification structurale est souvent liée à une fonction ou à un mode de fonctionnement dans des voies métaboliques précises. Ceci peut éventuellement nous guider sur la fonction de PAB0955.

### 3. Une classification des GTPases

#### 3.1. La famille des interrupteurs moléculaires : domaine G avec un feuillet- $\beta$ mixte

Les GTPases de cette première sous-famille sont les plus nombreuses, elles regroupent entre autres les petites protéines G (environ une centaine de protéines). Elles possèdent un domaine G dont le feuillet- $\beta$  est mixte, avec le brin- $\beta$ 2 anti-parallèle aux autres brins. Elles sont qualifiées d'interrupteurs moléculaires car elles ont la particularité de pouvoir osciller entre une forme active, liant le GTP, et une forme inactive, liant le GDP. Au cours de leur cycle



d'hydrolyse du GTP, de part leur changement de conformation en fonction de la nature du nucléotide, les interrupteurs moléculaires sont capables de fixer différents partenaires et donc d'induire différents types de signaux.

### 3.1.1. Les petites protéines G

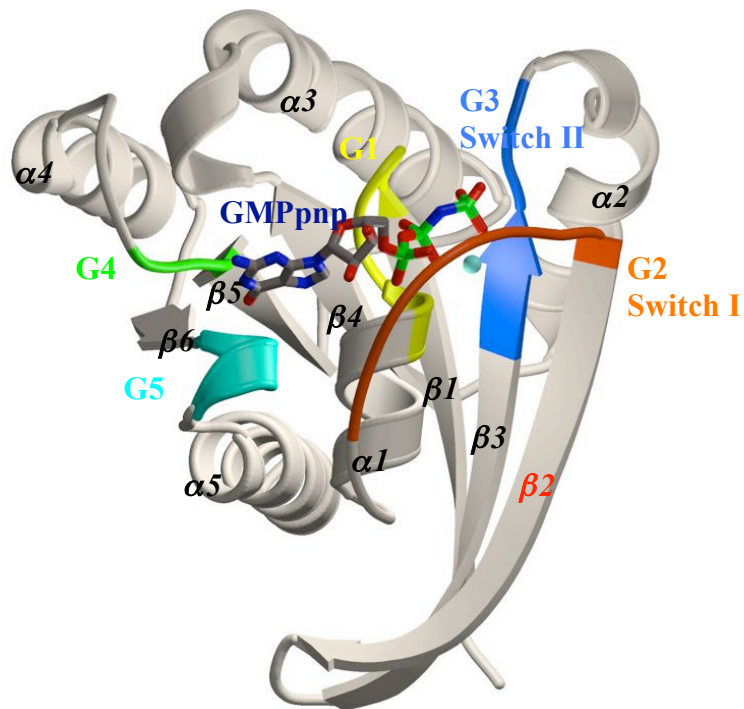
Les GTPases les plus étudiées sont celles de la super famille des petites protéines G, car elles sont impliquées dans de nombreuses maladies chez l'homme, dont plusieurs cancers.

Les petites protéines G sont divisées en 5 familles, selon leurs fonctions biologiques :

- \_ RAS (différenciation et prolifération cellulaire)
- \_ RAB (reconnaissance et la fusion vésiculaires)
- \_ RHO (cytosquelette)
- \_ RAN (transport nucléoplasmique et le cycle cellulaire)
- \_ ARF (formation vésiculaire).

Le repliement global de ces protéines est de type Rossmann comme pour toutes les GTPases ou ATPases. Elles sont nommées petites protéines G car elles sont généralement constituées d'environ 180 acides aminés, donc sont de petit poids moléculaire (20 kDa). Leur structure forme le domaine G et représentent le motif minimum pour fixer et hydrolyser le GTP.

Le domaine G, contenant les 5 motifs G (Bourne et al., 1991) (fig.3), est constitué de 5 à 6 brins- $\beta$ , formant un feuillet- $\beta$  mixte (le brin- $\beta_2$  est antiparallèle aux autres brins- $\beta$ ), pris en sandwich par des hélices- $\alpha$  dans les interrupteurs moléculaires.



*Figure.3. Structure de Ras humaine en complexe avec un analogue du GTP (GMPppnp). (code PDB : 1CTQ (Scheidig et al., 1999))*

Le motif G1, la P-loop, sert à la fixation des groupements phosphates du nucléotide en formant une boucle grâce à ses glycines conservées. Sa séquence consensus  $GX_1X_2X_3X_4GK[T/S]$  est un motif de Walker de type A (Walker et al., 1982), contient une lysine qui stabilise les charges négatives des phosphates. Ce motif se localise, pour les petites protéines G, entre le premier brin- $\beta$  et la première hélice- $\alpha$ , c'est le cas pour la plupart des GTPases.

Le motif G2 qui relie l'hélice- $\alpha_1$  au brin- $\beta_2$ , n'a pas de séquence consensus mais possède une thréonine conservée impliquée dans la fixation du magnésium ( $T^{35}$  dans hRas). Ce motif est également nommé « switch I » pour les interrupteurs moléculaires, car il change de conformation en fonction de la nature du nucléotide fixé.

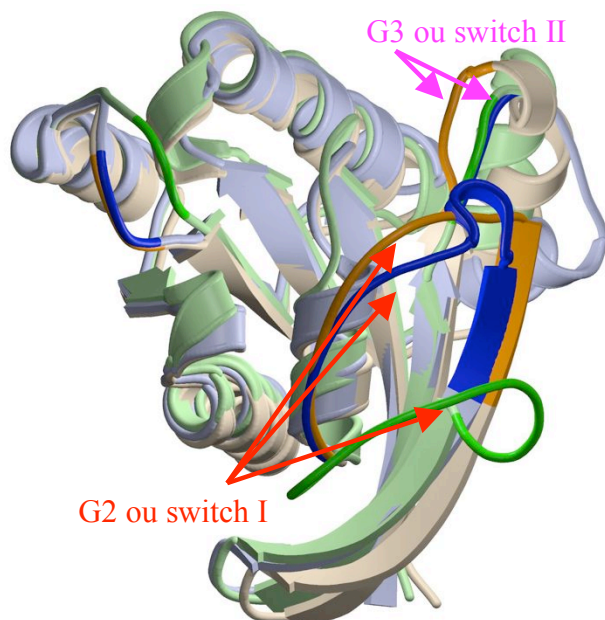
Le motif G3 est un motif de Walker de type B, dont la séquence consensus est  $DX_1X_2G$ . Il est situé en  $N_{ter}$  de la seconde hélice- $\alpha$ , après le brin- $\beta_2$  qui est anti-parallèle aux autres brins- $\beta$ . Ce motif sert à la fixation du magnésium et du phosphate- $\gamma$ . Il est aussi appelé « switch II » pour les interrupteurs moléculaires car il est flexible et change de conformation en fonction de la nature du nucléotide fixé. Dans beaucoup de petites protéines G, ce motif contient en plus une glutamine,  $DX_1X_2GQ$ , impliquée dans le mécanisme d'hydrolyse du GTP par stabilisation de l'état intermédiaire, (Q61 dans hRas).

Le motif G4, NKXD, reliant le brin- $\beta_5$  à l'hélice- $\alpha_4$ , sert à la fixation de la base guanine. Les acides aminés asparagine et acide aspartique forment des liaisons hydrogènes avec la base, la lysine forme des contacts de van der Waals avec le noyau aromatique de la base.

Le motif G5 stabilise la base guanine en formant une liaison hydrogène avec l'atome d'oxygène O6 spécifique de la guanine (absent dans l'adénine). Ce motif est responsable, avec le motif G4, de la sélectivité des protéines vis-à-vis de la guanine. La séquence du motif G5 est SAK le plus souvent, avec une alanine conservée qui fixe la base par l'atome d'azote de sa chaîne principale. Ce motif G5 n'est pas présent dans toutes les GTPases.

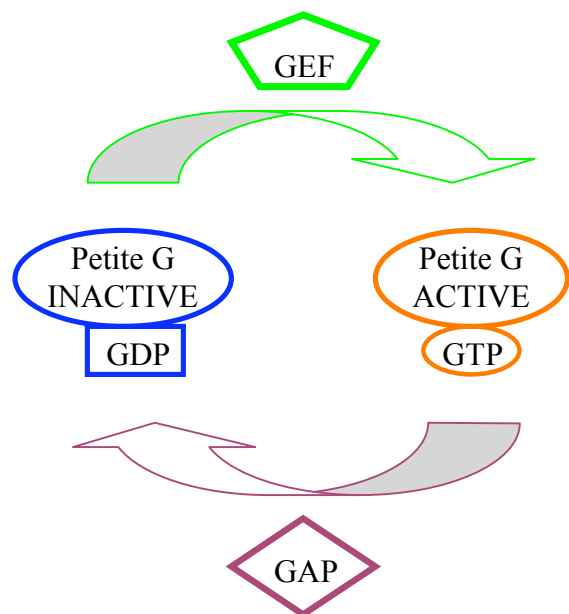
La désignation des motifs G sera conservée tout au long du manuscrit, et les différences de structure ou de séquence de ces motifs selon les protéines seront signalées.

La particularité de ces petites protéines G réside dans leur capacité à changer de structure en fonction du nucléotide fixé. Ce changement est important et se constate assez facilement sur les structures de la figure 4.a avec comme exemple la protéine Ras humaine. Les régions nommées « switch I et II » (motifs G2 et G3) étant mobiles, en absence de nucléotide, il est fréquent qu'elles ne soient pas modélisées dans les structures cristallographiques car leur densité électronique est trop faible.



*Figure.4.a. Superposition de trois structures de hRas.*

*En orange hRas en complexe avec le GMPppp (1CTQ), en bleu hRas en complexe avec du GDP (4Q21(Milburn et al., 1990)) et en vert la protéine dans sa forme complexée avec un facteur d'échange (GEF) qui n'est pas représenté (1BKD (Boriack-Sjodin et al., 1998)). Les flèches indiquent les déplacements des motifs G2 et G3.*



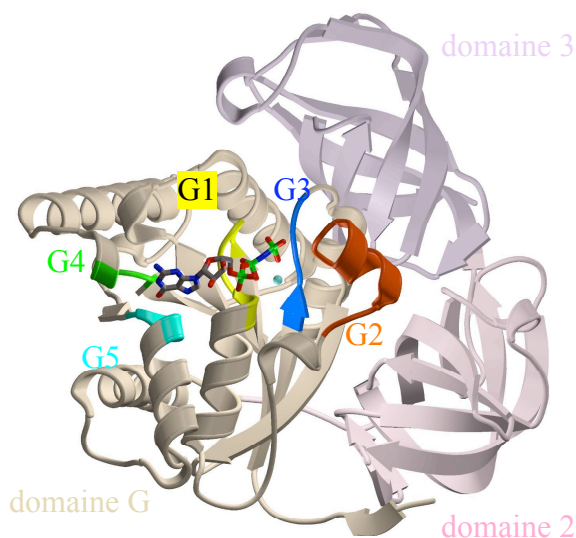
*Figure.4.b. Cycle d'hydrolyse du GTP par des petites protéines G.*

*Le facteur d'échange, GEF, charge la protéine en GTP et le facteur d'activation, GAP, active l'hydrolyse du GTP.*

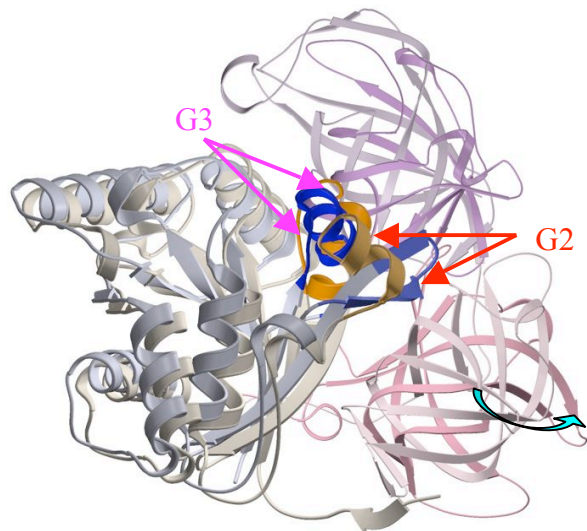
L'état des petites protéines G est contrôlé par la nature du nucléotide fixé. Lorsque la protéine fixe du GDP, elle est inactive et ne peut transmettre de signal. Comme son activité d'échange de nucléotide est faible, elle est régulée par d'autres facteurs protéiques (fig.4.b.) (Sprang, 1997). Ainsi une protéine GEF (Guanine nucleotide Exchange Factor) vient catalyser l'activation de la petite protéine G en favorisant l'échange du GDP par du GTP. Dans sa conformation liant le GTP la petite protéine G est active, elle peut donc transmettre un signal cellulaire. L'hydrolyse du GTP en GDP est catalysée par une protéine GAP (GTPase Activating Protein) et permet le retour à l'état inactif de la petite protéine G qui se retrouve alors avec du GDP dans son site nucléotidique.

### 3.1.2. Les facteurs d'élongation et d'initiation (EF-Tu, EF-G, IF)

Chaque étape de la biosynthèse des protéines (traduction) est contrôlée par de nombreuses macromolécules dont les facteurs d'initiation (IF ou eIF) et d'élongation (EF-Tu et EF-G). Ces facteurs sont composés de 3 domaines, le domaine 1 est un domaine G, les domaines 2 et 3 sont structurés en tonneau- $\beta$  antiparallèle (fig.5.a.). Ces facteurs appartiennent à la famille des interrupteurs moléculaires et changent de conformation selon le type de nucléotide fixé (fig.5.b.) (Sprang, 1997).



*Figure.5.a.* Structure de EF-Tu de *Thermus aquaticus* en complexe avec du GMPppp (1EFT (Kjeldgaard et al., 1993)). Le domaine G est en beige, le domaine 2 en rose et le domaine 3 en violet.

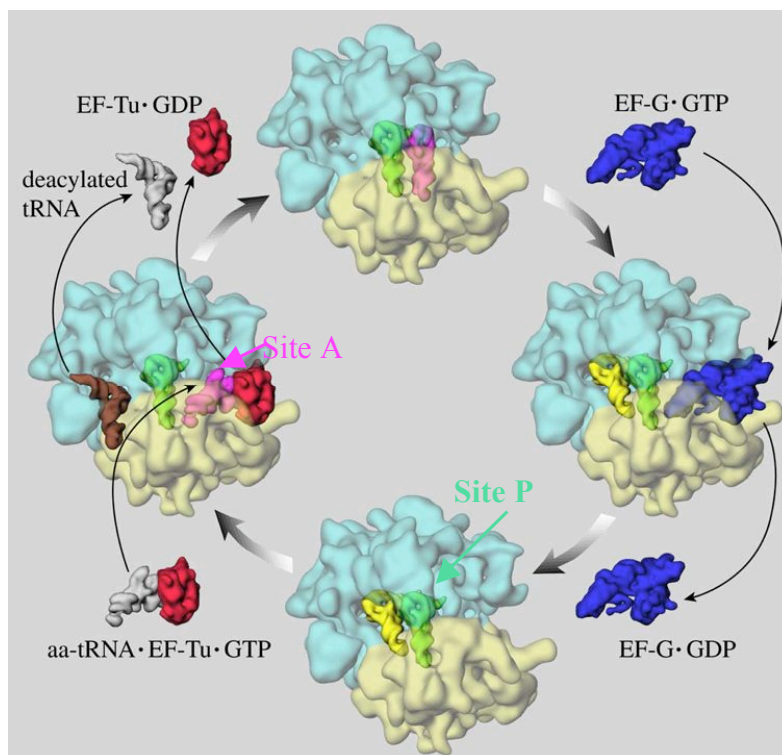


*Figure.5.b.* Superposition des complexes EF-Tu/GMPppp et EF-Tu/GDP (1TUI (Polekhina et al., 1996)). En beige et orange le domaine G de EF-Tu/GMPppp, en bleu celui de EF-Tu/GDP. En rose et violet foncé sont représentés les domaines 2 et 3 de EF-Tu/GDP.

Sur la figure 5.b. les changements structuraux entre les structures liant un analogue du GTP et celle liant le GDP sont signalés par des flèches rouges pour le motif G2 (switch I) et rose pour le motif G3 (switch II). La position du domaine G par rapport à celle des deux autres domaines tourne de presque 90° entre les deux structures, en plus des mouvements dans le domaine G. Ce basculement se produit au moment de l'hydrolyse du GTP ; cela sépare le facteur d'élongation de l'ARN de transfert (ARNt).

La traduction a lieu au niveau du ribosome qui reçoit les ARNt transportant les acides aminés pour former la chaîne polypeptidique et l'ARN messenger (ARNm) qui contient l'information de séquence à traduire. La traduction est un processus qui se divise en trois étapes, l'initiation, l'élongation et la terminaison.

Les facteurs d'initiation, comme leur nom l'indique, débutent la traduction en permettant aux différentes molécules mises en jeu d'être bien positionnées. Ils permettent la formation d'un complexe contenant : le ribosome, l'ARNm au niveau de son codon d'initiation et l'ARNt d'initiation lié à une méthionine. À ce moment, les facteurs d'initiation sont capables de fixer du GTP, qui par son hydrolyse va permettre un changement de conformation rendant le complexe d'initiation fonctionnel pour l'élongation de la chaîne polypeptidique, et libère des facteurs d'initiation.



*Figure.6. Cycle d'élongation de la chaîne polypeptidique. En cyan la grande sous unité du ribosome, en beige la petite, en jaune le site de terminaison de la traduction.*

Les facteurs d'élongation EF-Tu et EF-G servent à l'élongation de la chaîne polypeptidique au cours de la traduction des protéines (fig.6).

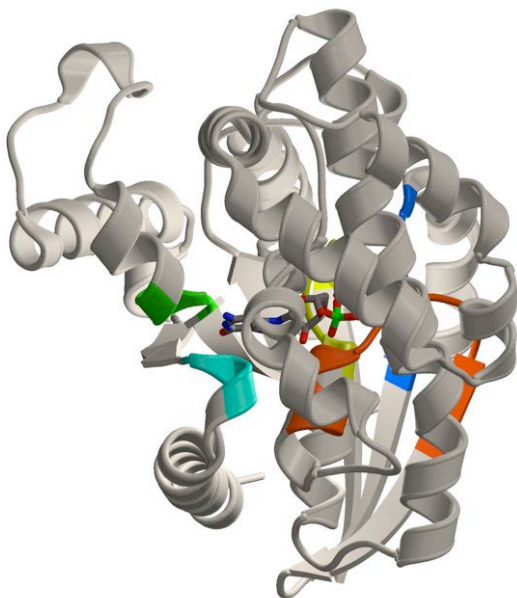
Les EF-Tu fixent et délivrent l'ARNt lié avec un acide aminé au site A du ribosome dans sa forme liant le GTP. Le ribosome agit alors comme activateur de l'hydrolyse du GTP (comme les GAP des petites protéines G), ce qui



libère le facteur d'élongation EF-Tu. La liaison peptidique peut alors se former, au site A du ribosome ; il y a donc élongation de la chaîne polypeptidique. Le complexe ARNt-polypeptide est alors transféré du site A au site P sur le ribosome. Ceci se fait grâce au facteur EF-G qui, par hydrolyse du GTP, peut se lier au ribosome. Il y a alors séparation de l'ARNt non chargé du site P ; le ribosome glisse alors sur l'ARNm, ce qui entraîne un transfert de l'ARNt-polypeptide du site A au site P. Ainsi le polypeptide est en place pour accepter la formation d'une nouvelle liaison peptidique avec un autre acide aminé, qui grâce au déplacement de l'ARNm correspondra au codon suivant dans la séquence. L'étape de terminaison intervient lorsqu'il y a un codon stop sur la séquence d'ARNm.

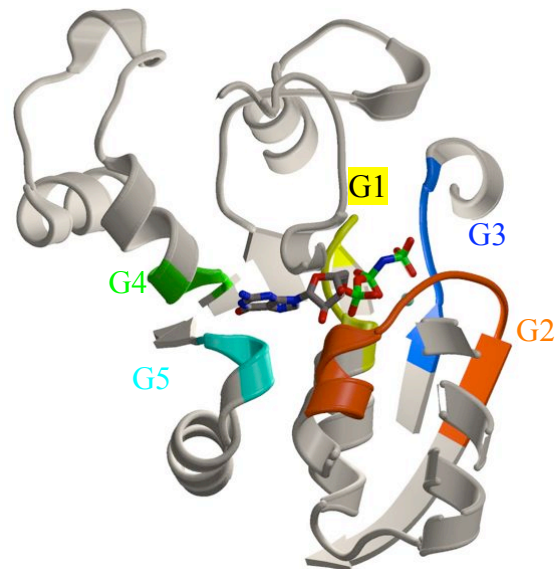
### 3.1.3. Les protéines G hétérotrimériques, sous unité alpha

Les signaux extracellulaires sont reconnus par des récepteurs à la surface de la membrane cytoplasmique et transmis à l'intérieur des cellules par l'intermédiaire de récepteurs. Les récepteurs appartenant à la famille des récepteurs à 7 hélices transmembranaires, sont couplés aux protéines G hétérotrimériques. Ces protéines possèdent 3 sous-unités  $G\alpha$ ,  $G\beta$  et  $G\gamma$ . La sous-unité  $G\alpha$  porte la fonction d'hydrolyse du GTP (fig.7.a. et 7.b.).



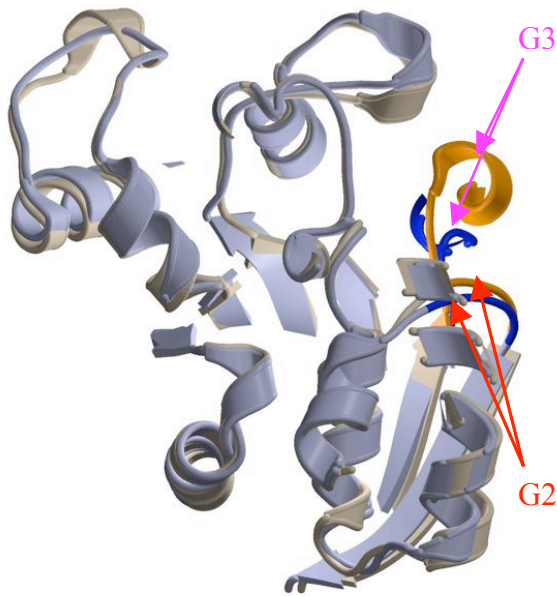
*Figure.7.a. Structure de la sous-unité  $G\alpha$  de Rat en complexe avec du GMPpnp (1CIP (Coleman and Sprang, 1999)).*

*La structure est ici complète, mais des hélices- $\alpha$  recouvrent le site nucléotidique.*



*Figure.7.b. Structure de la sous-unité  $G\alpha$  de Rat en complexe avec du GMPpnp (1CIP).*

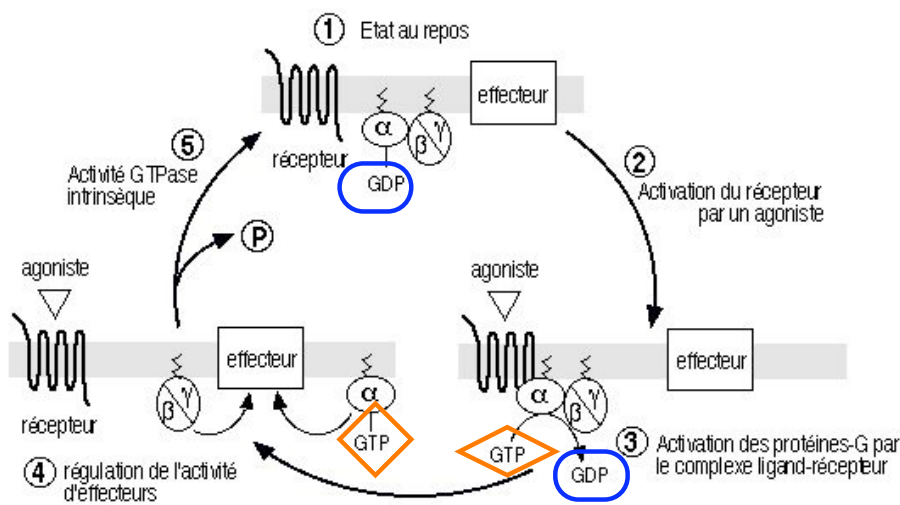
*La vue est dans la même orientation que celle de la fig.7.a. mais sans les hélices- $\alpha$  qui recouvrent le site nucléotidique.*



*Figure.7.c. Superposition des structures de Gα en complexe avec du GMPppn (orange) et celle en complexe avec du GDP et du PO<sub>4</sub><sup>2-</sup> (bleu) (1GIT (Berghuis et al., 1996)).  
Les flèches indiquent les zones mobiles.*

Cette GTPase oscille entre une conformation active liant le GTP (en orange) et une conformation inactive liant le GDP (en bleu) (fig.7.c.). Comme les petites protéines G, les protéines G hétérotrimériques sont aussi des interrupteurs moléculaires qui changent de structure en fonction du nucléotide. La sous-unité α change de conformation selon le type de nucléotide fixé. Les sous-unités β et γ, toujours liées ensemble, forment le domaine de reconnaissance du récepteur ainsi que le domaine d'interaction et de régulation des effecteurs (Kaziro et al., 1991).

Il existe différentes classes de sous-unités α : α<sub>s(olf)</sub>, α<sub>i(o,t,g,z)</sub>, α<sub>q</sub> et α<sub>12</sub>, elles sont impliquées dans différentes fonctions biologiques. Lorsque la sous-unité α fixe du GTP, elle peut réguler des effecteurs intracellulaires comme l'adényl-cyclase, la phospholipase Cβ, les canaux K<sup>+</sup> et Ca<sup>2+</sup>, la phosphodiesterase GMP cyclique. Lorsqu'elle fixe du GDP, la sous-unité Gα est liée

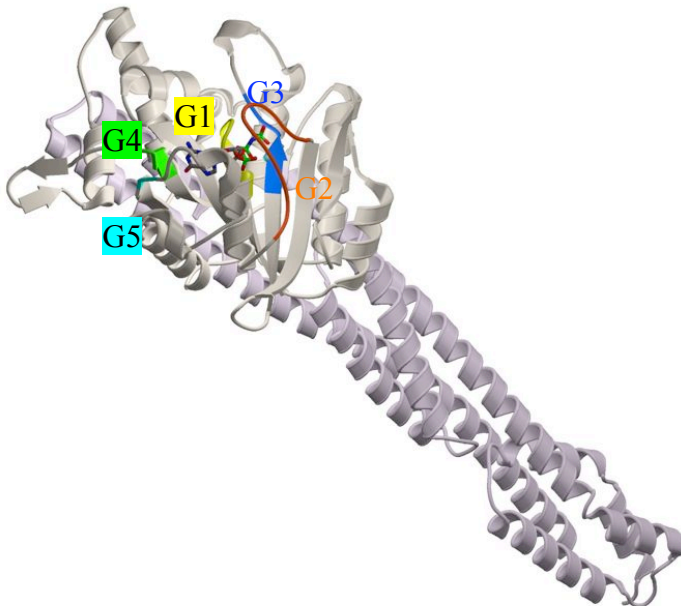


*Figure.8. Cycle des protéines G hétérotrimériques*

avec les deux autres sous-unités βγ et ne peut donc fixer d'effecteur (Sprang, 1997). Seule la sous-unité α liant du GTP, dissociée des sous-unités βγ, peut interagir avec des effecteurs (fig.8).

### 3.1.4. Les GBP (guanylate-binding protein)

Les GBP sont de grandes GTPases (60 à 100 kDa) impliquées dans la réponse cellulaire anti-virale. Elles constituent la classe la plus abondante de protéines induites par l'interféron- $\gamma$ , une substance immunomodulatoire induisant l'expression de gènes impliqués dans la réponse anti-virale (Prakash et al., 2000b). Ces protéines GBP sont multidomaines (fig.9.a). Elles possèdent un grand domaine G (300 résidus) structuralement similaire à celui des petites protéines G. Ces protéines ont en plus un autre domaine en hélice- $\alpha$  allongé (violet), ce qui fait qu'elles sont de plus haut poids moléculaire (60-100 kDa) que les petites protéines G (20 kDa).



*Figure.9.a. Structure de la GBP1 humaine en complexe avec du GMPppp (1F5N (Prakash et al., 2000b)).*

*Le domaine G est en beige, et le domaine en hélice- $\alpha$  est en violet.*

est de 0.001 à 0.1 min<sup>-1</sup> (Prakash et al., 2000a). Cette vitesse d'hydrolyse plus rapide serait due à la présence d'un motif conservé dans les GBP (<sup>103</sup>DXEKGD<sup>108</sup> dans la GBP1 humaine). Ce motif riche en résidus chargés pourrait avoir un rôle catalytique similaire à celui d'une GAP (activation de l'hydrolyse du GTP).

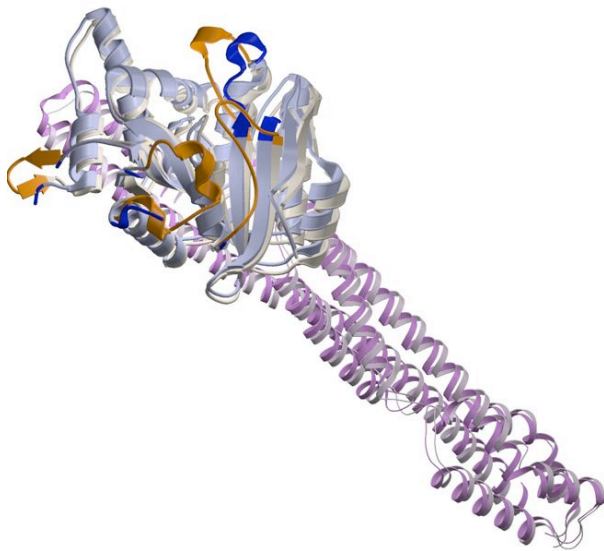
Les GBP sont capables de se dimériser de façon nucléotide-dépendante (Ghosh et al, 2006). En absence de nucléotide, en complexe avec du GMP ou du GDP, la GBP est monomérique.

Même si leur domaine G est structuralement proche de celui des petites protéines G, il existe plusieurs caractéristiques spécifiques aux GBP. Elles contiennent des insertions par rapport au domaine G canonique. Le motif G4, NKXD, est remplacé par une boucle RD où l'acide aspartique joue le même rôle (<sup>181</sup>TLRD<sup>184</sup> dans hGBP1). De plus les GBP ont une faible affinité pour les nucléotides à base guanine, mais ont une forte activité d'hydrolyse du GTP. Leur vitesse d'hydrolyse est de 80 min<sup>-1</sup> alors que celles des petites protéines G



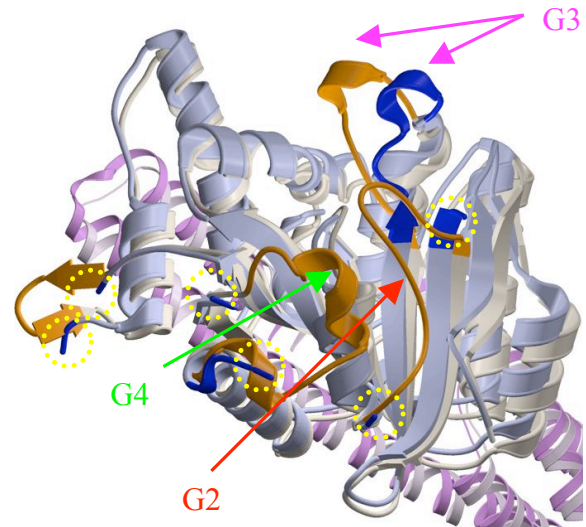
Elle devient dimérique en fixant un nucléotide tri-phosphate (GTP ou GMPppnp) (Prakash et al., 2000b).

Comme les protéines de la superfamille Ras, les GBP changent de conformation selon le nucléotide fixé (fig.9. b et c), notamment au niveau des régions « switch » (motifs G2 et G3).



*Figure.9.b.* Superposition de la structure de la hGBP1 seule (bleu) et en complexe avec du GMPppnp (orange).

En violet clair le domaine en hélice- $\alpha$  de la structure en complexe avec du GMPppnp (1F5N) et en violet foncé celui de la structure seule (1DG3 (Prakash et al., 2000a)).



*Figure.9.c.* Zoom de la figure.9.b. sur le domaine G.

Les cercles jaunes indiquent les extrémités des parties non modélisées dans la structure sans nucléotide.

La structure de hGBP1 sans nucléotide contient 4 régions qui n'ont pu être modélisées dans la densité électronique : le motif G2, le motif G3, le motif G4 et une région située après le motif G5. En absence de nucléotide, les interrupteurs moléculaires sont moins stables (voir instables pour certains) et certains motifs G sont alors désordonnés. La présence et la nature du nucléotide affectent la conformation et la stabilité de la protéine.

### 3.1.5. Les Dynamines

Les dynamines sont impliquées dans l'endocytose, le transport de vésicules et le maintien de la morphologie des mitochondries. Ce sont de grandes GTPases (100 kDa) composées de 5

domaines : un domaine G en N<sub>ter</sub>, un domaine GED (GTPase effector domain) qui joue le rôle de GAP, un domaine qui interagit avec le domaine GED, et deux autres domaines qui interagissent avec la membrane des vésicules (Sever et al., 2000). Comme les GBP, elles possèdent une faible affinité pour les nucléotides et une forte activité d'hydrolyse (supérieure à celle des petites protéines G); ceci étant probablement dû à leur domaine GED. Le domaine G ressemble à celui des petites protéines G avec des insertions supplémentaires. Les structures résolues pour la dynamine sont celles du domaine G de la dynamine A de *Dictyostelium discoideum* avec du GDP et sans nucléotide. Dans les deux structures, qui sont similaires, 3 régions n'ont pu être modélisées dont les deux régions switch. Il est donc difficile de comparer les structures puisque les régions qui sont susceptibles de se réarranger ne sont pas modélisées. Toutefois, nous pouvons penser que la dynamine change de conformation selon la nature des nucléotides. En effet l'absence du domaine GED pourrait expliquer l'agitation des régions qui n'ont pu être modélisées. Les similitudes structurales avec les autres interrupteurs moléculaires confirment que la fixation des nucléotides doit entraîner des réarrangements structuraux. Ces changements structuraux auraient pour but d'activer la fusion des membranes (Niemann et al., 2001).

### **3.1.6. Les GTPases non classiques**

Il existe en plus de ces familles de protéines G bien identifiées, différentes GTPases dont la structure est connue et diffère de celle des autres familles. Parmi ces protéines, il y a les protéines Eng A (1MKY, (Robinson et al., 2002)), B (1PUI non publié) et C (1U0L, (Shin et al., 2004) et 1TH9 (Levdikov et al., 2004) qui sont impliquées dans le cycle cellulaire et sont donc des cibles pour produire des molécules anti-bactériennes (1SUL, (Ruzheinikov et al., 2004)). Ces protéines contiennent deux domaines en plus de leur domaine G. Leur domaine G est similaire à celui des petites protéines G avec un feuillet- $\beta$ 2 antiparallèle et des régions non modélisées en absence de nucléotide triphosphate.

Il y a également les protéines : trmE (1RFL, non publié), Obg (1UDX, (Kukimoto-Niino et al., 2004)) et ERA (1EGA, (Chen et al., 1999)) qui sont des GTPases avec un domaine G dont les régions switch sont stables et un feuillet- $\beta$  mixte avec le second brin- $\beta$  anti-parallèle. Mais ces protéines n'ont pu être cristallisées avec des nucléotides, donc on ne peut étudier leur mode de fixation. Et enfin il y a la protéine TOC34 (1H65 (Sun et al., 2002)) qui est également une

GTPase capable de se dimériser. La dimérisation permet à la protéine de mimer l'action d'une protéine GAP de petite protéine G.

### 3.2. La famille des SRP : domaine G avec un feuillet- $\beta$ parallèle

Le complexe ribonucléoprotéique : SRP (Signal Recognition Particle) (Kuglstatter et al., 2002) intervient dans la sécrétion des polypeptides, en cours de traduction, à travers la membrane du réticulum endoplasmique (RE). Le complexe est formé par l'ARN et la protéine SRP. Cette protéine est une GTPase, multidomaine avec un domaine G.

Dans une première étape, la protéine SRP liant le GTP (bleu) reconnaît la séquence signal du polypeptide (rouge) en cours de synthèse. Puis, elle transporte le ribosome avec le polypeptide à la membrane, du RE chez les eucaryotes et la membrane plasmique chez les procaryotes, par l'interaction avec son récepteur liant le GTP à proximité de l'appareil de translocation (étape 2). Les deux complexes ainsi chargés en GTP peuvent interagir entre eux par leurs domaines G, ce qui auto-active l'hydrolyse de GTP. L'hydrolyse sépare la SRP du polypeptide fixé au ribosome, le récepteur de la SRP se détache du translocon (étape 3). Le ribosome est alors fixé sur le translocon et cela assure la sécrétion du polypeptide à travers la membrane (fig.10) (Keenan et al., 2001).

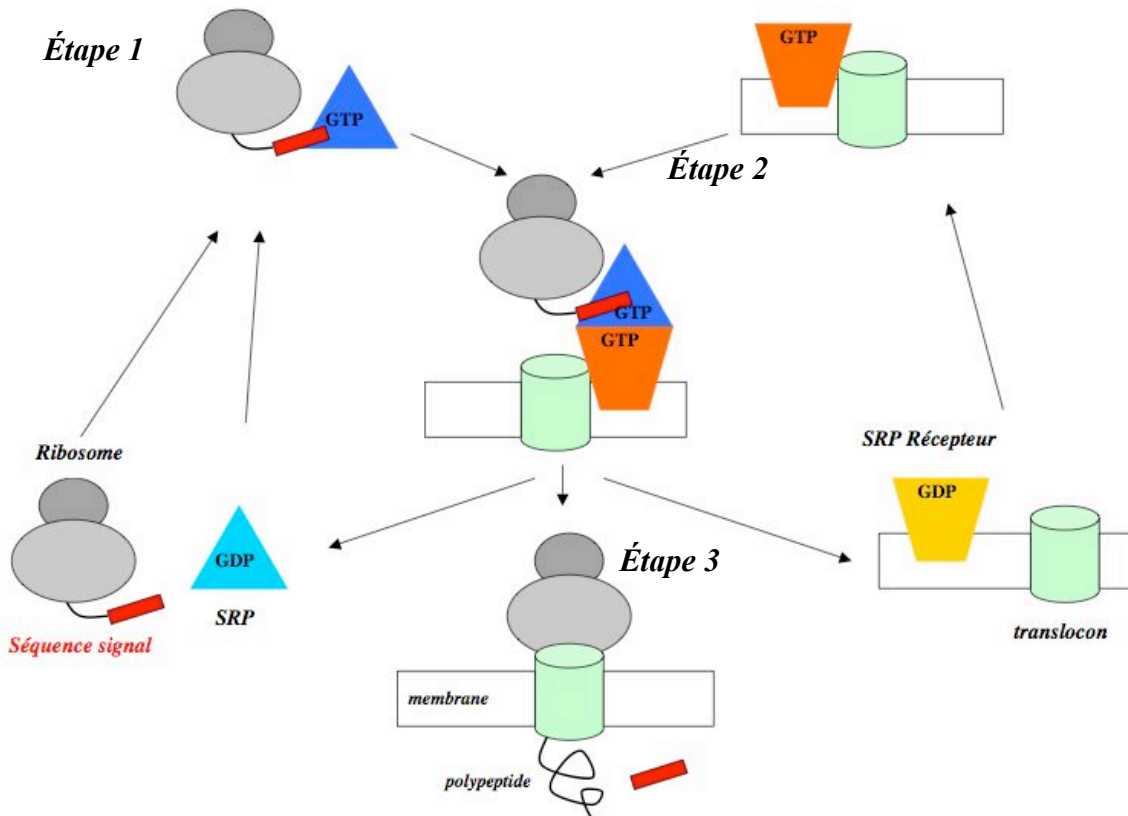
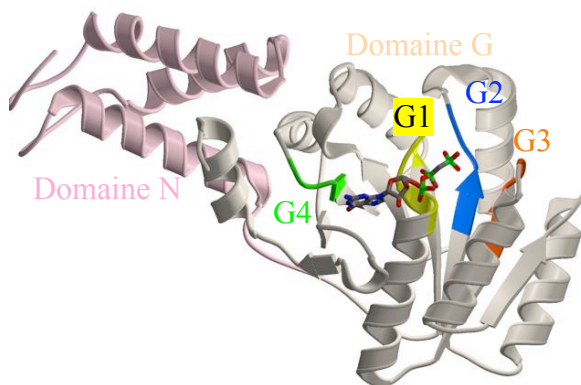


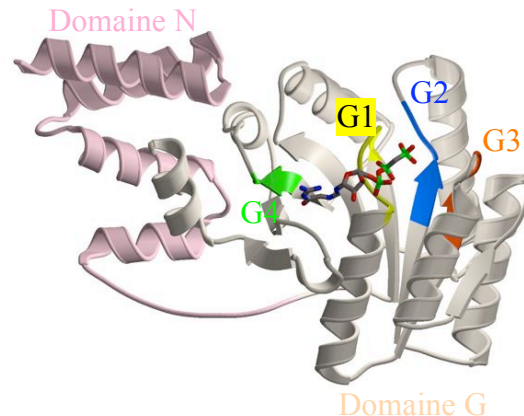
Figure.10. Translocation du polypeptide en cours de traduction à travers la membrane.

La protéine SRP et son récepteur sont des GTPases multidomaines, qui forment un hétérodimère pour activer l'hydrolyse du GTP au moment de la translocation du polypeptide. Chez *Thermus aquaticus*, la protéine SRP est nommée Ffh (fifty-four homolog), par analogie avec la SRP54 humaine qui est une des 6 protéines formant le complexe SRP (Keenan et al., 2001). Nous parlerons par la suite de Ffh et de son récepteur FtsY car leurs structures avec différents nucléotides ont été résolues.

La protéine Ffh est formée de trois domaines : un domaine N de 100 résidus en N<sub>ter</sub>, un domaine G (200 résidus) qui possède le site de fixation des nucléotides et interagit avec le récepteur et un domaine M en C<sub>ter</sub> (140 résidus) riche en méthionine qui fixe l'ARN et la chaîne polypeptidique en cours de traduction (fig.11.a.). Le récepteur de la SRP, lui contient deux domaines : un domaine N en hélices- $\alpha$ , et un domaine G qui fixe le nucléotide et interagit avec la SRP (fig.11.b.).



*Figure.11.a. Structure des domaines N et G de la Ffh en complexe avec du GMPppcp (1RJ9 (Egea et al., 2004)).  
Le domaine N est en rose, le domaine G en beige, le GMPppcp est représenté en bâton.*

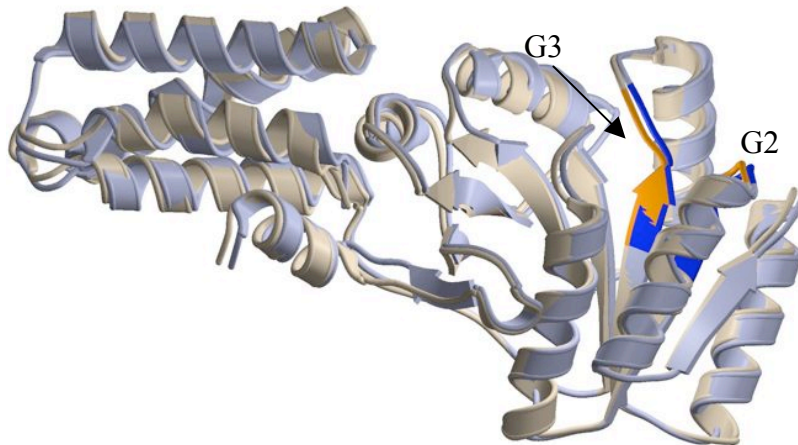


*Figure.11.b. Structure de FtsY en complexe avec du GMPppcp (1RJ9 (Egea et al., 2004)).*

Les domaines G de la Ffh et de son récepteur FtsY sont structurellement similaires. Même si le repliement est globalement proche de celui d'un domaine G canonique de petite protéine G, il s'en distingue significativement. Tout d'abord le motif G2, qui fixe le magnésium, contient un acide aspartique conservé, D<sup>135</sup> dans Ffh et D<sup>139</sup> dans FtsY, au lieu d'une thréonine dans les interrupteurs moléculaires. Ce motif G2 est situé après le brin- $\beta$ 2 dans Ffh et FtsY, alors qu'il se situe avant le brin- $\beta$ 2 dans les petites protéines G. Cette localisation différente fait qu'il

n'existe pas de grande boucle (switch I) avant le brin- $\beta$ 2 dans Ffh et FtsY. Ainsi le brin- $\beta$ 2 est parallèle aux autres brins et le feuillet- $\beta$  est donc totalement parallèle.

De plus, les structures de Ffh en complexe avec du GDP, du GMPpnp et en absence de FtsY, ne montrent pas de réarrangements structuraux (fig.11.c). L'acide aspartique D<sup>135</sup> de Ffh se déplace de 0.39 Å entre la structure liant le GDP et celle liant le GMPpnp, alors que pour hRas par exemple le déplacement de la T<sup>35</sup> est de 3.12 Å. Il en est de même pour le motif G3

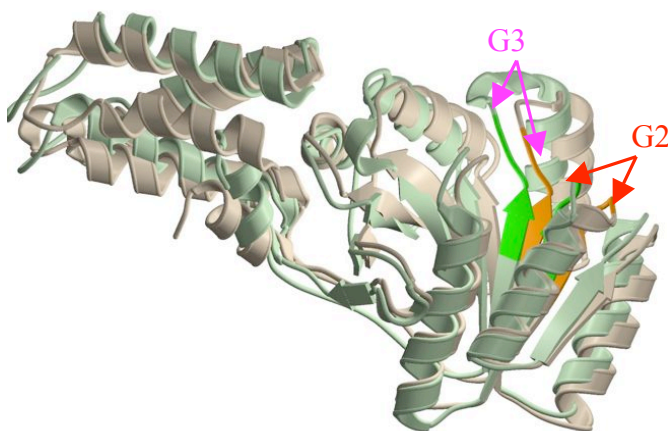


*Figure.11.c. Superposition des structures de Ffh en complexe avec du GDP en bleu (1NG1 (Freyermann et al., 1999) et en complexe avec du GMPpnp en orange (1JPJ (Padmanabhan and Freyermann, 2001)).*

*Les zones plus foncées indiquent les motifs G2 et G3 qui n'ont pas changé de conformations entre les deux structures.*

(switch II) qui ne subit pas dans Ffh de réarrangements structuraux aussi important que dans les petites protéines G. Le déplacement de la G<sup>190</sup> de Ffh est de 2 Å alors que celui de la G<sup>60</sup> de hRas est de 5.19 Å entre les structures en complexes avec le GDP et le GMPpnp.

La protéine Ffh seule ne subit pas de réarrangements structuraux selon la nature du nucléotide fixé. Lorsqu'elle interagit avec son récepteur, FtsY, des changements structuraux sont



*Figure.11.d. Superposition de Ffh (sans FtsY) en complexe avec du GMPpnp (en orange 1JPJ) et en complexe avec FtsY (en vert 1RJ9) (Egea et al., 2004).*

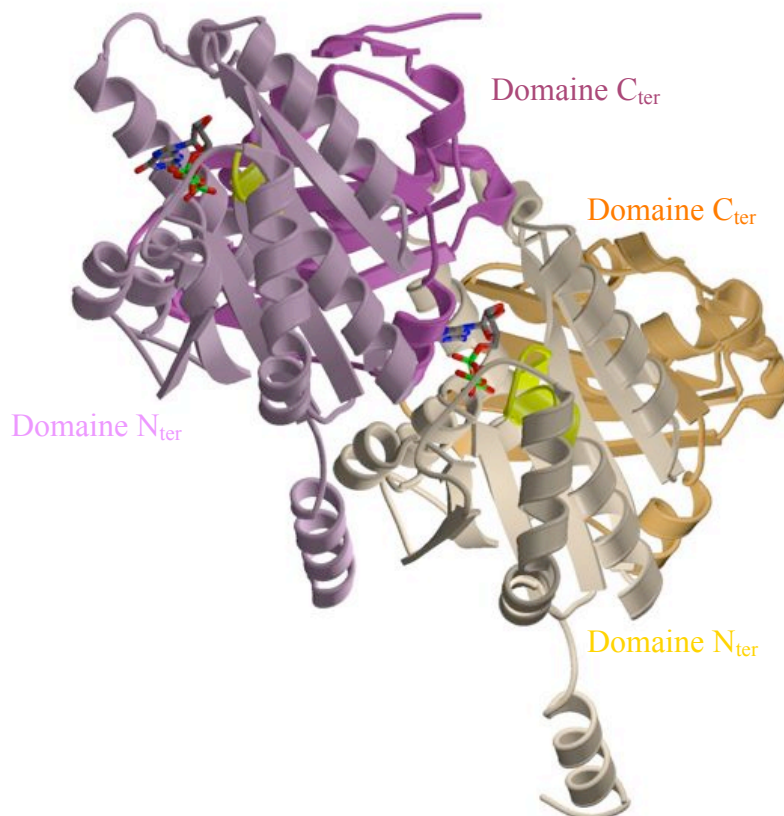
observés notamment au niveau des motifs G2 et G3 (figure.11.d.). Ceci montre que ces motifs G2 et G3 sont également flexibles dans les SRP et leurs récepteurs. Leurs conformations dépendent de l'état d'oligomérisation de ces protéines et non de la nature du nucléotide fixé. La formation du complexe Ffh/FtsY déclenche



l'hydrolyse du GTP, avec l'apport d'une arginine dans le site actif de chaque protéine. Ainsi, les protéines Ffh et FtsY s'auto-activent.

### 3.3. La famille des tubulines : sans domaine G

La formation des microtubules joue un rôle important dans les mécanismes liés à la croissance, la division, la différenciation, le transport et la sécrétion cellulaire. La tubuline est une protéine constitutive majeure des microtubules. C'est une GTPase qui est notamment régulée par des protéines de la famille de la stathmine. La structure de la tubuline a été déterminée en complexe avec la stathmine. La stathmine est une phosphoprotéine qui module la dynamique des microtubules en réponse à des stimuli externes. Le complexe tubuline/stathmine est composé d'un hétérodimère de tubuline  $\alpha\beta$  et d'une molécule de stathmine (Gigant et al., 2000).



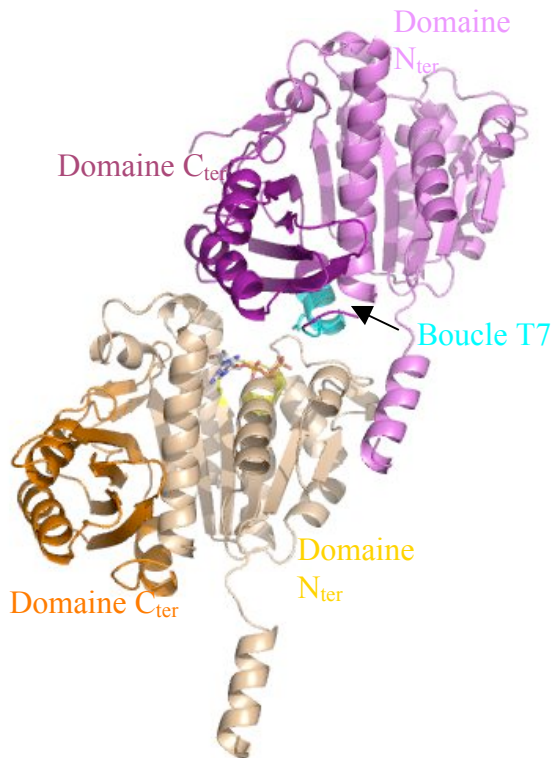
*Figure.12.a. Structure du dimère de FtsZ en complexe avec du GTP (1W5A (Oliva et al., 2004)).*

*En beige et violet clair sont représentés les domaines N<sub>ter</sub> des deux monomères de FtsY, et en orange et violet foncé sont représentés les domaines C<sub>ter</sub> de ces monomères, et en jaune la P-loop.*

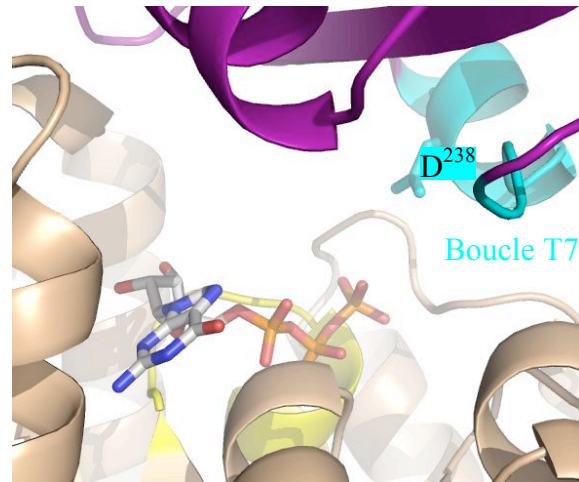
La protéine FtsZ est la tubuline procaryote. Elle polymérise également pour former les microtubules essentiels à la division cellulaire. Chez les procaryotes, FtsZ adopte la même structure que la tubuline (Oliva et al., 2004). FtsZ est formée de deux domaines : un N<sub>ter</sub> qui contient le site de fixation du nucléotide et un C<sub>ter</sub> qui sert à l'hydrolyse du GTP. Le domaine N est différent des domaines G des autres GTPases. Il ne contient pas de motifs G. Ce domaine adopte un repliement de type

Rossmann, avec un feuillet- $\beta$  parallèle composé de 6 brins entouré d'hélices- $\alpha$ . Le site nucléotidique n'est complet que lorsque la protéine est dimérique. La dimérisation se fait par l'interaction d'un domaine  $N_{ter}$  d'un monomère avec le domaine  $C_{ter}$  de l'autre monomère (fig.12.a.).

Le domaine  $C_{ter}$ , composé d'hélices- $\alpha$  et de brins- $\beta$ , contient la boucle T7. Cette boucle contient l'acide aspartique ( $D^{238}$  de *Methanococcus jannaschii*) qui est essentiel à l'hydrolyse du GTP (Oliva et al., 2004) (fig.12.b. et c).



*Figure.12.b.* Structure de FtsZ en complexe avec du GTP (1W5A (Oliva et al., 2004)).



*Figure.12.c.* Zoom de la figure.12.b. au niveau du site nucléotidique.

La boucle T7 du domaine  $C_{ter}$  d'un monomère apporte l'acide aminé D238 qui vient compléter le site nucléotidique du domaine  $N_{ter}$  de l'autre monomère.

Sous sa forme liant le GTP la protéine FtsZ dimérise. Après l'hydrolyse du GTP, les monomères se séparent.

La tubuline et la protéine FtsZ adoptent un repliement de type Rossmann mais sont bien différentes des GTPases décrites précédemment. Elles ne contiennent pas de motif G et leur P-loop n'est pas un motif de Walker de type A. En effet, la séquence qui fixe les groupements phosphates du nucléotide est une séquence modifiée riche en glycine ( $^{130}\text{GLGGGTGTG}^{138}$  pour FtsZ de *Methanococcus jannaschii*). La structure du monomère ne semble pas subir de

réarrangement structural lors de l'hydrolyse du GTP, cependant ces protéines oscillent entre une forme active et inactive selon le type de nucléotide fixé et leur degré d'oligomérisation.

En résumé la majorité des GTPases sont des interrupteurs moléculaires qui changent de conformation en fonction du nucléotide fixé. D'un point de vue structural, elles sont monomériques, leur feuillet- $\beta$  est mixte et leurs motifs G2 et G3 changent de conformation selon le nucléotide fixé. L'oscillation entre leur état actif et inactif est contrôlée par les protéines GAP et GEF.

Les protéines de la famille des SRP possèdent un feuillet- $\beta$  parallèle, leur structure ne varie pas sensiblement en fonction du nucléotide fixé (pas de GEF ou de GAP). Cependant l'hydrolyse du GTP est activée par l'hétérodimérisation entre la SRP et son récepteur.

Les propriétés biochimiques de ces domaines G (affinité pour le nucléotide, vitesse d'hydrolyse) dépendent visiblement du rôle biologique tenu. L'ensemble de ces données structurales et biochimiques sur les domaines G va nous permettre de situer PAB0955 dans ce contexte et nous aider à identifier son rôle et son mode de fonctionnement en mettant en évidence les différences et les similitudes avec les structures de domaines G déjà connues.

À titre d'exemple, aucune GTPase homodimérique avec un domaine G n'a été identifiée à ce jour.





CHAPITRE I

BIOCHIMIE

ET

CRISTALLISATION



## I.1. Biochimie de PAB0955

### I.1.1. Expression et purification de PAB0955

La protéine codée par le gène *PAB0955* de *Pyrococcus abyssi* a été produite et purifiée par nos collaborateurs de l'équipe de Jean Armengaud du CEA de Marcoule. La présence de codons rares chez *Pyrococcus abyssi* a nécessité l'utilisation de la souche Rosetta(DE3)pLys qui contient des gènes codant pour des ARNt supplémentaires pour les acides aminés arginine, isoleucine, proline et leucine. La première étape de purification se fait par chauffage de l'extrait cellulaire à 80°C, car la majorité des protéines d'*Escherichia coli* sont dénaturées alors que les protéines de *Pyrococcus abyssi* qui est une archaea hyperthermophile résistent à des hautes températures (environ 100°C). Puis, la protéine PAB0955 est purifiée sur une colonne d'affinité grâce à son étiquette constituée de 14 acides aminés dont six histidines, située en N<sub>ter</sub>. Notre protéine ayant un point isoélectrique théorique de 5.6, la dernière étape de purification se fait sur une colonne échangeuse d'anion (Resource-Q). La purification permet d'obtenir 6,6 mg de protéine recombinante pure par litre de culture. La protéine est ensuite analysée par spectrométrie de masse et par séquençage N<sub>ter</sub>. Ceci a confirmé que c'était bien la protéine PAB0955 recombinante qui a été purifiée, avec une masse moléculaire attendue de 30298 Da au vue de sa séquence primaire (fig.I.1.).

```
MRGSHHHHHH GMASMIVVV GTAGSGKTTL TGEFGRYLED NYKVAYVNLD
TGVKELPYEP SIDVREFVTV EEIMREGYGP NGAIVESYDR LMEKFNEYLN
KILRLEKEND YVLIDTPGQM ETFLFHEFGV RLMENLPYPL VVYISDPEIL
KKPNDYCFVR FFALLIDLRL GATTIPALNK VDLLSEEEKE RHRKYFEDID
YLTARLKLDP SMOGLMAYKM CSMTEVLPP VRVLYLSAKT REGFEDLETL
AYEHYCTCGD LT
```

Figure.I.1. Séquence de PAB0955 recombinante

Surligné en vert l'étiquette 6 histidines, surligné en jaune le motif de Walker de type A qui fixe les groupements phosphates du nucléotide.

### I.1.2. Degré d'oligomérisation de PAB0955

La protéine recombinante PAB0955 contient 262 acides aminés. Lors de mon arrivée au laboratoire, différents états d'oligomérisation (fig.I.2.) avaient été révélés par une analyse de chromatographie d'exclusion. Les différents états d'oligomérisation (30% de dimère, 50% de tétramère et 20% d'octamère) sont dus aux deux cystéines situées en C<sub>ter</sub> de la protéine qui forment un pont disulfure inter-chaîne. Une étude par électrophorèse sur gel d'acrylamide en conditions dénaturantes, a permis de constater que cette oligomérisation était réversible par ajout de DTT (fig.I.3.). En présence de DTT la protéine est homogène, sa masse correspond à celle d'un dimère. Les tests de cristallisation qui suivent ont tous étaient réalisés en présence de DTT afin de conserver une solution protéique homogène monodisperse.

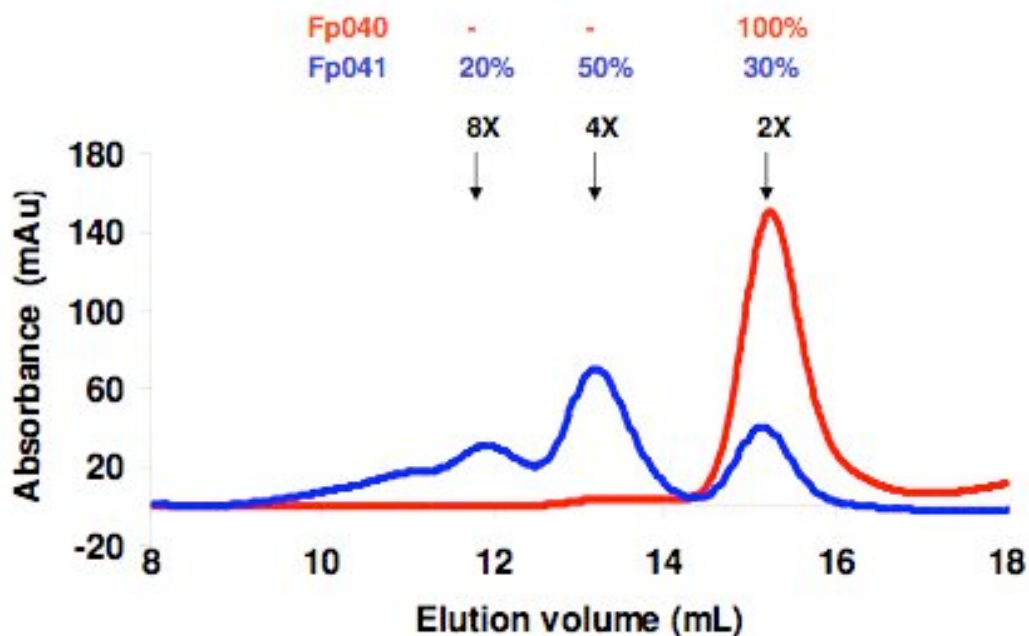


Figure.I.2. Profil d'élution de PAB0955 par chromatographie d'exclusion.

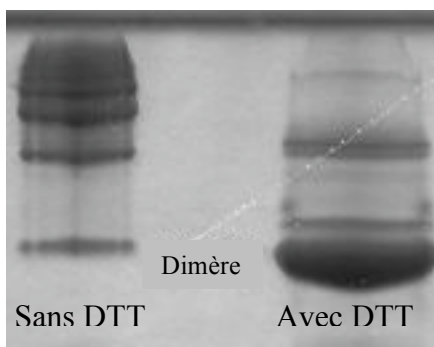


Figure.I.3. Gel d'électrophorèse natif

Figure.I.2. En bleu la fraction FP041 ne contient pas de DTT, il y a plusieurs multimères (8X : octamère, 4X : tétramère, 2X dimère). En rouge la fraction Fp040 qui contient du DTT, la protéine est sous forme dimérique exclusivement.

Figure.I.3. Gel d'électrophorèse à 12.5% d'acrylamide en condition native. L'ajout de DTT permet d'obtenir majoritairement la forme dimérique.

### I.1.3. Activité GTPase de PAB0955

La protéine PAB0955 possède un motif de Walker de type A en N<sub>ter</sub> (<sup>7</sup>GTAGSGKT<sup>14</sup>), permettant la fixation et l'hydrolyse de nucléotide tri-phosphate (fig.I.1.). L'activité hydrolase de PAB0955 a donc été testée en solution montrant que PAB0955 est une GTPase. Ces tests d'activité ont été réalisés à la température optimale de croissance de *Pyrococcus abyssi* (82°C).

Des quantités croissantes de protéines (0.2 à 1µM) sont incubées avec une quantité fixe de GTP (200µM) dans un tampon EPPS à 50mM à pH 7 avec du MgCl<sub>2</sub> à 5mM. Le tampon EPPS ne varie pas de pH à haute température. L'ensemble protéine et GTP à 80°C est analysé à différents temps (0, 5, 10, 15, 30 et 60 min) grâce à la séparation des différents nucléotides. L'hydrophobicité du GTP (substrat) est différente de celle du GDP (produit) cela permet de les séparer (colonne C18).

À cette température, optimale pour l'archaea, l'activité spécifique d'hydrolyse du GTP de PAB0955 est de 0,012 µmol de GTP hydrolysé en GDP par minute et par milligramme de protéine. Cette vitesse d'hydrolyse est plus faible que celle d'une protéine de type kinase (de l'ordre du µmol), 10 fois plus rapide que celle d'une petite protéine G sans GAP (Sauvage et al., 1991).

L'activité GTPase de la protéine XAB1, homologue humain de PAB0955, a été observée par auto-radiographie sur gel (Nitta et al., 2000). Après 30 minutes, le GTP est hydrolysé partiellement par XAB1, environ 30 à 40% d'après l'intensité des bandes sur le gel. Les conditions expérimentales pour mesurer l'activité d'hydrolyse de XAB1 utilisent 10 fois moins de substrat que d'enzyme. Au contraire nous avons utilisé 100 fois plus de substrat que de protéine pour mesurer l'activité d'hydrolyse de PAB0955. Il est donc difficile de comparer leur activité à la nôtre, mais dans les deux cas, la vitesse d'hydrolyse n'est pas comparable à celle d'une kinase.

### I.1.4. Repliement de PAB0955

Avant d'entreprendre des tests de cristallisation, il est important de savoir si la protéine étudiée est repliée ; c'est-à-dire si elle adopte des structures secondaires de type hélice-α ou brin-β. Les protéines présentant des parties non repliées ou trop flexibles pour adopter toujours la même conformation ont peu de chance de cristalliser. Sur le serveur ExPASy

(<http://www.expasy.org/tools/#secondary>) il est possible de faire une prédiction de structure secondaire. Notre protéine devrait, selon la prédiction, présenter une alternance de brins- $\beta$  et d'hélices- $\alpha$ . Sachant que les GTPases adoptent un repliement de type Rossmann, ceci est en accord avec la prédiction de structure secondaire.

Nos collaborateurs du CEA de Marcoule ont analysé la solution de protéine pure par dichroïsme circulaire. Cette technique permet de connaître la proportion d'hélices- $\alpha$ , de brins- $\beta$  et de parties non structurées dans une protéine. Le dichroïsme circulaire repose sur le fait que les protéines sont des structures optiquement actives qui n'absorbent pas de la même façon la lumière polarisée circulairement à droite et à gauche. De cette analyse, ils ont pu confirmer que la protéine était repliée ; et qu'elle adoptait une structure avec environ 50% en hélice- $\alpha$  et 50% en brin- $\beta$ .

Une fois la protéine purifiée, elle nous est envoyée congelée. C'est pourquoi nous avons tout d'abord vérifié que le transport et la congélation n'avaient pas affecté son repliement. Pour cela, nous avons utilisé la résonance magnétique nucléaire à une dimension, pour savoir si la protéine était correctement repliée et pour tester sa thermostabilité. Les spectres ont montré que la protéine était repliée, et qu'elle le restait en augmentant la température (jusqu'à 48°C). Elle est donc bien thermostable. La propriété de thermostabilité vient du fait que *Pyrococcus abyssi* est une archaea hyperthermophile. La conservation de cette propriété nous indique que la protéine n'a pas été altérée au cours du transport ou par la congélation.

La protéine recombinante PAB0955 est donc pure, homogène en solution, dimérique, thermostable, elle présente des structures secondaires et possède une activité d'hydrolyse spécifique pour le GTP.

## I.2. Cristallisation de PAB0955

Afin de pouvoir déterminer la structure tridimensionnelle par rayons X de notre protéine, il faut tout d'abord obtenir des cristaux de celle-ci. Le cristal est un agencement périodique de molécules dans les trois dimensions de l'espace. Cette organisation tridimensionnelle spécifique fait que les rayons X diffusés par les nombreuses molécules du cristal vont interférer de façon constructive dans certaines directions privilégiées, permettant la mesure d'un spectre de diffraction.

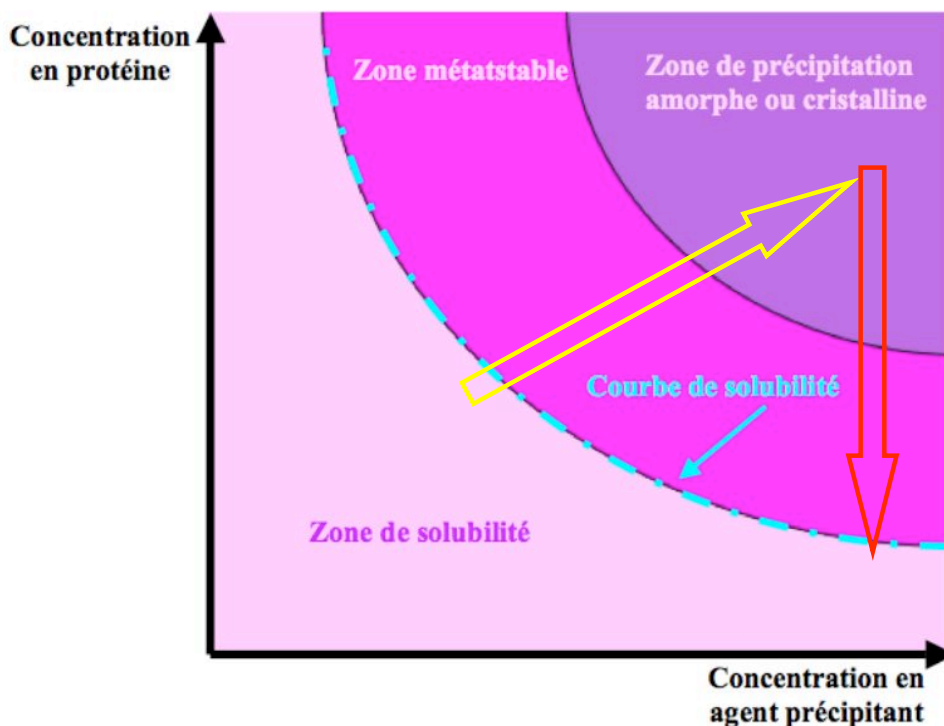
Les cristaux doivent être mono-cristallin afin d'obtenir un spectre de diffraction de bonne qualité permettant la résolution de la structure atomique de notre protéine.

## I.2.1. Principe de la Cristallisation

### I.2.1.1. Diagramme de phase

La solubilité d'une macromolécule est fonction de nombreux paramètres comme le pH, la température, la force ionique, l'effet d'additifs, etc. Lors de la cristallisation, la concentration de la macromolécule augmente jusqu'à atteindre l'état de sur-saturation favorable à l'obtention de cristaux (fig.I.4.).

Les cristaux sont obtenus en formant des gouttes au-dessus d'un réservoir. L'équilibre s'établit entre la goutte et le réservoir par diffusion des espèces volatiles jusqu'à ce que la tension de vapeur soit la même dans la goutte et dans le réservoir. Dans la plupart des cas, la goutte est initialement plus diluée que le réservoir ce qui entraîne une évaporation partielle de la goutte. La protéine qui ne s'échange pas avec le réservoir reste dans la goutte et donc se concentre.



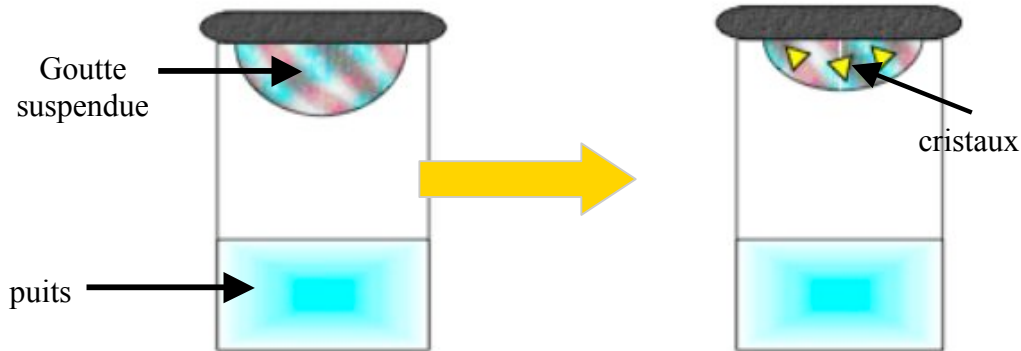
*Figure.I.4. Diagramme de phase d'une protéine*

*La flèche jaune indique l'évolution du système lors de l'équilibration par diffusion de vapeur jusqu'à la nucléation. La flèche rouge indique l'évolution de la goutte après nucléation, les cristaux se forment et la concentration en protéine diminue dans la goutte.*



### I.2.1.2. Technique de la goutte suspendue

La méthode la plus utilisée pour cristalliser les macromolécules est la diffusion de vapeur en goutte suspendue (fig.I.5.) (McPherson, 1990).



*Figure.I.5. Système de cristallisation par diffusion de vapeur en goutte suspendue. La flèche orange indique l'évolution du système lors de l'équilibration par diffusion de vapeur.*

Nous avons utilisé deux types de boîtes pour la cristallisation en goutte suspendue. Les boîtes Limbro et Nextal. Dans une boîte Limbro, la goutte est formée sur une lamelle de verre siliconée, qui ensuite est scellée par de la graisse pour avoir un système hermétique. Dans une boîte Nextal, la goutte est formée sur un bouchon de plastique, qui est visé sur le puit, et qui possède un joint plastique rendant le système aussi hermétique que la graisse (fig.I.6.).



*Figure.I.6. Boîte Nextal de cristallisation*

La cinétique de la diffusion dépend des différences de concentration dans la goutte et dans le réservoir mais également de la géométrie du système, du volume de la goutte et de sa distance au réservoir (DeTitta and Luft, 1995). Les volumes de réservoir sont classiquement de 0,5 ml ou 1 ml. Les gouttes sont généralement formées du mélange de la protéine avec la solution du réservoir dans un rapport 1 pour 1, la goutte a donc un volume initial de 2  $\mu$ l.

Lorsque l'on fait de larges criblages des différentes conditions de cristallisation, grâce à des kits commerciaux, les robots de cristallisation utilisent la technique de la goutte assise. Le principe d'obtention des cristaux reste le même, par diffusion de vapeur, mais la goutte est assise sur une cupule au-dessus du puits au lieu d'être suspendue.

## I.2.2. Cristallisation de PAB0955

### I.2.2.1. Cristallisation de la protéine Native et Apo

Il faut tout d'abord déterminer la concentration de protéine appropriée à utiliser, c'est à dire la gamme de concentration pour laquelle nos gouttes ne seront ni toutes limpides ni toutes précipitées. Le but étant de se placer, au moment de la formation des gouttes, au plus proche de la zone de solubilité de la protéine. Cette zone est estimée par l'observation de gouttes formées avec une gamme de concentration en protéine et en agent précipitant. Cette estimation sert de point de départ pour les premiers essais de cristallisation et peut être affinée par la suite.

Nous avons ainsi estimé qu'une concentration de l'ordre de  $8\text{mg.ml}^{-1}$  était appropriée pour un premier essai de cristallisation avec PAB0955.

Pour couvrir une large de gamme de différentes conditions de cristallisation nous avons utilisé des kits commerciaux. Avec les premiers essais de cristallisation que nous avons entrepris sur PAB0955, nous avons obtenu des oursins (amas de fines aiguilles) avec du sulfate d'ammonium comme agent précipitant et des précipités avec le PEG6000. Nous avons donc utilisé les kits nommés Crystal screen I et II (Jancarik and Kim, 1991), et un kit contenant du sulfate d'ammonium à différents pH de chez Hampton Research. Le kit crystal screen I et II permet de tester une gamme de pH de 4 à 9, avec ou sans sel et différentes natures d'agents précipitants. Ces premiers tests nous ont permis d'obtenir les premiers cristaux de PAB0955 ( $20\ \mu\text{m}$ ) après une journée (fig.I.7.a).

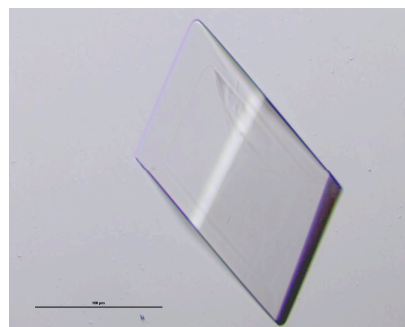
La solution de cristallisation contient du tri-Na-citrate pH 5.6 à  $0.1\ \text{M}$ ; 20% isopropanol (v/v) ; 20% PEG 4000 (w/v) et du DTT à  $20\ \text{mM}$ . L'isopropanol est un solvant volatil, qui s'évapore rapidement et crée des mouvements de convection dans les gouttes entraînant les cristaux. Afin d'éviter ces problèmes et dans l'objectif d'optimisation de nos cristaux (pour qu'ils soient plus volumineux) l'isopropanol a été remplacé par d'autres solvants. Parmi eux, le dioxane permet d'obtenir les cristaux les plus volumineux (fig.I.7.b). Ces cristaux obtenus

en 6 jours sont reproductibles. Ils ont permis l'enregistrement d'un jeu de données complet à l'E.S.R.F. (voir chapitre II résolution de structure).



*Figure I.7.a. Cristaux de PAB0955 native (20  $\mu\text{m}$ ).*

*Condition de cristallisation*  
*tri-Na citrate pH 5.6 à 0.1 M*  
*20% isopropanol*  
*20% PEG 4K*  
*DTT à 20 mM*  
*PAB0955 à 8mg.mL<sup>-1</sup>*



*Figure I.7.b. Cristaux de PAB0955 native (300  $\mu\text{m}$ ).*

*Condition de cristallisation*  
*tri-Na citrate pH 5.6 à 0.1 M*  
*3% dioxane*  
*15% PEG 4K*  
*DTT à 20 mM*  
*PAB0955 à 8mg.mL<sup>-1</sup>*

Les sources de rayons X puissantes comme le rayonnement synchrotron provoquent des dommages d'irradiation importants dans les cristaux (production de radicaux libres, rupture de liaison covalente ...). Une des solutions pour diminuer ces dommages d'irradiation, qui déforment et détruisent l'édifice cristallin, est de congeler les cristaux et de les maintenir pendant l'enregistrement des données sous un flux constant d'azote gazeux à environ  $-160^{\circ}\text{C}$ . La congélation permet de limiter la diffusion des radicaux libres et ralentit donc leur effet. Malheureusement, la congélation crée des tensions sur le cristal qui introduisent des défauts à différents degrés. Pour conserver la qualité des cristaux, s'ils se désorganisent au moment de la congélation, il est nécessaire de tester différents agents cryo-protectant (Garman and Schneider, 1997). Selon la protéine et ses conditions de cristallisation, il faut trouver le cryo-protectant adéquat pour ne pas fragiliser le cristal. Il existe par exemple un kit de 36 cryo-protectants différents, chez Hampton Research. Parmi ces cryo-protectants, il y a des huiles (paratone, paraffine), des sels à forte concentration (nitrate de lithium à 8 M, chlorure de sodium à 5 M), des sucres (glucose, tréhalose, raffinose), des polyols (butandiol, hexanediol ...), des solvants organiques (éthanol, propanol ...) et des polymères (éthylène glycol, des PEG de différentes tailles ...).

### I.2.2.2. Cristallisation de la protéine avec des sels d'atomes lourds

Il n'y a pas d'homologue proche en séquence (d'au moins 20% d'identité) de PAB0955, dont la structure ait été résolue. Par exemple il y a seulement 12% d'identité de séquence entre PAB0955 et la petite protéine Ras humaine. Nous n'avons donc pas tenté le remplacement moléculaire pour résoudre la structure de PAB0955. Nous avons décidé de déterminer les phases avec des cristaux de protéines contenant des atomes lourds pour pouvoir résoudre la structure de PAB0955. Des trempages de nos cristaux natifs et des co-cristallisations de la protéine avec différentes solutions de sel d'atomes lourds ont été réalisés.

En collaboration avec Meike Stelter et Richard Kahn du Laboratoire de Cristallographie des Macromolécules de l'IBS nous avons pu tester un dérivé cagé de gadolinium DTPA-BMA (fig.I.8.) (Girard et al., 2003).

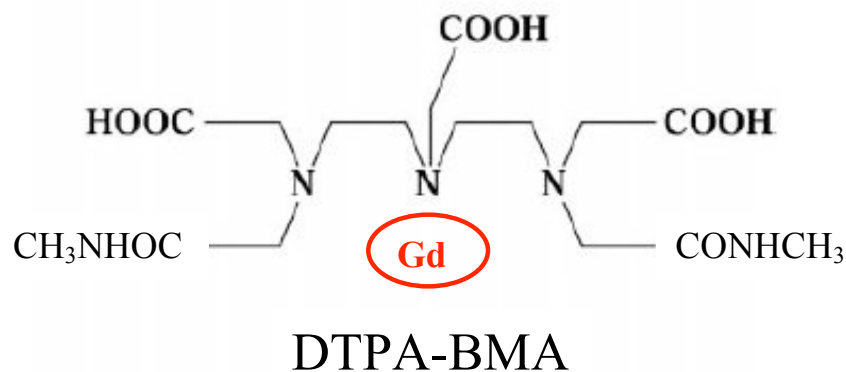


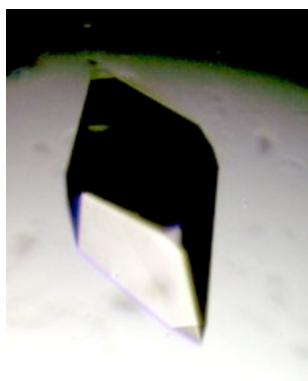
Figure.I.8. Complexe cagé de gadolinium DTPA-BMA.

L'avantage de cet ion  $Gd^{3+}$  est que son facteur  $f'$  est de  $28e^-$  à son seuil L3 ( $\lambda = 1.7117 \text{ \AA}$ ) et  $12e^-$  pour le rayonnement d'une anode tournante de laboratoire ( $\lambda = 1.54 \text{ \AA}$ ). Avec des cristaux dérivés de gadolinium, il est donc possible de déterminer les phases avec une source de rayons X de laboratoire. De plus, les complexes de gadolinium cagé sont non toxiques, ce qui facilite leur manipulation. D'après les études menées par Meike Stelter durant sa thèse, les conditions de cristallisation pour les protéines qu'elle a testé, restent les mêmes en présence de complexe de gadolinium. Cela facilite l'obtention de cristaux dérivés par co-cristallisation.

Les cristaux natifs ont été trempés dans des solutions de Gd-DTPA-BMA à différentes concentrations (10, 50 et 100 mM) et pendant différents temps (quelques heures à 2 jours). Après cryo-protection comme pour les cristaux natifs, ces cristaux dérivés ont été congelés.

Nous avons alors enregistré trois jeux de données à l'E.S.R.F. sur la ligne de lumière BM30-A à différentes longueurs d'ondes, au seuil du Gd ( $\lambda = 1.7117 \text{ \AA}$ ), à l'inflexion ( $\lambda = 1.70107 \text{ \AA}$ ) et une éloignée des deux premières ( $\lambda = 1.3001 \text{ \AA}$ ). Ceci afin de pouvoir déterminer les phases à l'aide du signal anomal enregistré à plusieurs longueurs d'onde. Les cristaux diffractaient jusqu'à  $3 \text{ \AA}$  avec un  $R_{\text{sym}}$  global de 8% pour le meilleur jeu, ce qui est déjà élevé pour extraire le signal anomal qui est faible. Les cristaux ont tout de même permis de déterminer la présence d'un site de Gadolinium par monomère de PAB0955. Le taux d'occupation du gadolinium n'était par contre pas suffisant pour pouvoir résoudre la structure de PAB0955 avec ces cristaux dérivés.

Afin d'augmenter le taux d'occupation du site de Gd-DTPA-BMA, des co-cristallisations de la protéine avec ce complexe ont été réalisées. Des cristaux ont été obtenus en 6 jours à 293K par la technique de la goutte suspendue en faisant une gamme de concentration en PEG4000 à partir des conditions de cristallisation de la protéine native, en ajoutant à chaque goutte  $1 \mu\text{l}$  de Gd-DTPA-BMA à différentes concentrations (10, 50 et 100 mM). Ceci nous a permis d'enregistrer un jeu de données complet à l'E.S.R.F. sur la ligne de lumière ID14-eh4 ( $\lambda = 1.2782 \text{ \AA}$ ), avec un cristal de  $100 \mu\text{m}$  obtenu par co-cristallisation avec du Gd-DTPA-BMA à 100 mM (fig.I.9).



*Figure I.9. Cristal de PAB0955 en complexe avec du Gd-DTPA-BMA ( $100 \mu\text{m}$ ).*

*Condition de cristallisation  
tri-Na citrate pH 5.6 à 0.1 M  
16% PEG 4K  
3% dioxane  
DTT à 20 mM  
PAB0955 à  $10.7 \text{ mg.ml}^{-1}$   
Gd-DTPA-BMA à 100 mM*

Après traitement de ce jeu à une seule longueur d'onde (S.A.D.) la présence du site de fixation de l'ion Gadolinium a été confirmée. Le taux d'occupation du Gadolinium et la qualité de diffraction de ce cristal ont permis d'obtenir des phases de très bonne qualité (voir le chapitre II résolution de structure).

### I.2.2.3. Cristallisation de la protéine en complexe avec des nucléotides

Après avoir résolu la structure de la protéine native et sans nucléotide, des essais de cristallisation ont été entrepris avec différents nucléotides à base guanine.

Des trempages de cristaux natifs ont été réalisés avec un analogue non-hydrolysable du GTP, le GTP $\gamma$ S qui contient un atome de soufre à la place d'un des atomes d'oxygène lié au phosphate- $\gamma$ . Les cristaux natifs ne supportent pas la présence de nucléotide, ils se fissurent et se dissolvent. Il est donc impossible d'obtenir les structures de la protéine en complexe avec des nucléotides par trempage. Ceci s'est vérifié pour l'ensemble des nucléotides testés.

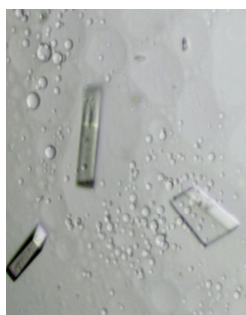
#### **Les nucléotides tri-phosphates**

La co-cristallisation de PAB0955 avec le GTP $\gamma$ S dans une large gamme autour des conditions de cristallisation de la protéine apo n'a pas eu plus de succès. Aucun cristal n'a pu être obtenu à partir de ces conditions.

La co-cristallisation de la protéine avec le GTP $\gamma$ S a donc été réalisée avec les kits commerciaux de Hampton en nanogoutte (robot de cristallisation de l'EMBL). Ceci dans le but de tester une large gamme de conditions de cristallisation avec peu de matériel biologique. Sur les 576 conditions différentes testées, 97 ont permis l'obtention de cristaux soit 17%. Ce nombre important de conditions variées de cristallisation (toujours avec du PEG) montre que la protéine est sans doute plus stable en présence d'un nucléotide (donc plus apte à cristalliser). Parmi toutes ces conditions, celles qui permettent d'obtenir les cristaux les plus volumineux sont proches de celle de la protéine apo, un sel (acétate d'ammonium) remplaçant le solvant (dioxane).

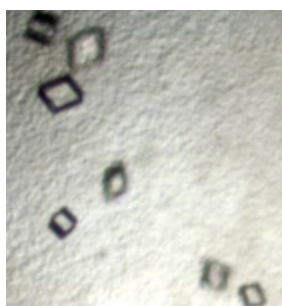
Les premières co-cristallisations avec le GTP $\gamma$ S ont été réalisées en utilisant un rapport d'une molécule de nucléotide pour un monomère de protéine. Le taux d'occupation du GTP $\gamma$ S observé dans ces structures est d'environ 50%, en conséquence le rapport nucléotide/protéine a été augmenté. Pour avoir un taux d'occupation de 100%, il fallut utiliser un ratio quatre fois plus important (fig.I.10.a.).

Deux autres analogues non-hydrolysable du GTP ont également été testés en co-cristallisation avec PAB0955 dans les mêmes conditions, le GMPp $\beta$ cp (fig.I.10.b.) et le GMPpnp (fig.I.10.c.) qui contiennent un atome de carbone ou d'azote entre les phosphates  $\beta$  et  $\gamma$  respectivement.



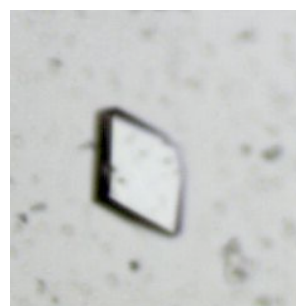
*Figure.I.10.a. Cristaux de PAB0955 avec GTP $\gamma$ S. (100  $\mu$ M)*

Condition de cristallisation  
tri-Na citrate pH 5.6 à 0.1 M  
13% PEG 4K  
NH<sub>4</sub> acétate à 0.2 M  
  
DTT à 20 mM  
PAB0955 à 8 mg.ml<sup>-1</sup>  
GTP $\gamma$ S à 0.65 mM  
18°C



*Figure.I.10.b. Cristaux de PAB0955 avec GMPpcp. (50  $\mu$ M)*

Condition de cristallisation  
tri-Na citrate pH 5.6 à 0.1 M  
17% PEG 6K  
NH<sub>4</sub> acétate à 0.2 M  
NaCl à 50 mM  
DTT à 20 mM  
PAB0955 à 5 mg.ml<sup>-1</sup>  
GMPpcp à 0.65mM  
18°C

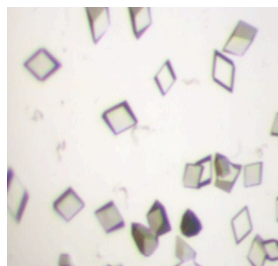


*Figure.I.10.c. Cristaux de PAB0955 avec GMPpnp. (80  $\mu$ M)*

Condition de cristallisation  
tri-Na citrate pH 5.6 à 0.1M  
16% PEG 6K  
NH<sub>4</sub> acétate à 0.2 M  
  
DTT à 20 mM  
PAB0955 à 7 mg.ml<sup>-1</sup>  
GMPpnp à 0.65mM  
18°C

Le GTP a également été testé, dans trois conditions différentes. La première dans des conditions similaires aux précédentes, afin de savoir si notre protéine était active dans les conditions de cristallisation. Des cristallisations sont faites en parallèle à 18°C où nous espérons observer une hydrolyse du GTP (fig.I.11.a.) et à 4°C, où nous espérons ralentir le mécanisme d'hydrolyse (fig.I.11.b.). Dans un second temps, des co-cristallisations avec un chélateur de magnésium, l'EDTA, ont été réalisées afin de savoir si, comme pour les autres GTPases, le Mg<sup>2+</sup> était essentiel à l'hydrolyse du GTP (fig.I.11.c.).

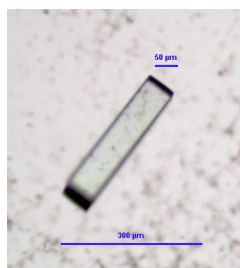




*Figure.I.11.a. Cristaux de PAB0955 avec GTP (18°C). (50  $\mu$ M)*

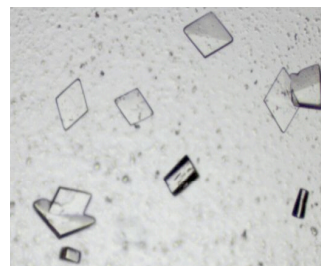
Condition de cristallisation  
tri-Na citrate pH 5.6 à 0.1 M  
15% PEG 4K  
NH<sub>4</sub> acétate à 0.2 M

DTT à 20 mM  
PAB0955 à 8 mg.ml<sup>-1</sup>  
GTP à 0.65 mM  
18°C



*Figure.I.11.b. Cristaux de PAB0955 avec GTP (4°C). (300  $\mu$ M)*

Condition de cristallisation  
Acide citrique pH 5 à 0.1 M  
8% PEG 6K et 10% glycérol  
Lithium Cl à 1 M  
NaCl à 50 mM  
DTT à 20 mM  
PAB0955 à 8 mg.ml<sup>-1</sup>  
GTP à 1mM  
4°C



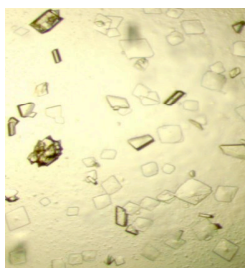
*Figure.I.11.c. Cristaux de PAB0955 avec GTP (EDTA). (80  $\mu$ M)*

Condition de cristallisation  
tri-Na citrate pH 5.6 à 0.1 M  
12% PEG 4K  
NH<sub>4</sub> acétate à 0.1 M  
DTT à 20 mM  
PAB0955 à 8 mg.ml<sup>-1</sup>  
GTP à 0.65mM  
18°C  
EDTA à 10mM

### **Le nucléotide di-phosphate avec et sans additifs**

La co-cristallisation en présence de GDP, produit de l'hydrolyse du GTP, a également été réalisée, pour savoir si la protéine avait de l'affinité pour son produit (fig.I.12.a.). Des cristaux ont été obtenus dans des conditions de cristallisation similaires à celles des autres nucléotides. Puis, afin de mimer un état intermédiaire de la réaction d'hydrolyse, plusieurs analogues du phosphate ont été co-cristallisés en présence de GDP, ainsi que du phosphate lui-même. Nous avons ainsi obtenu des cristaux en présence de sels de tungstate de sodium (fig.I.12.b.), de phosphate de sodium et de fluorure d'aluminium (fig.I.12.c.).





*Figure.I.12.a. Cristaux de PAB0955 avec GDP. (50 μM)*

Condition de cristallisation  
tri-Na citrate pH 5.6 à 0.1 M  
13% PEG 4K  
NH<sub>4</sub> acétate à 0.2 M

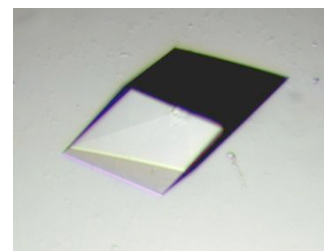
DTT à 20 mM  
PAB0955 à 8 mg.ml<sup>-1</sup>  
GDP à 0.65 mM  
18°C



*Figure.I.12.b. Cristaux de PAB0955 avec GDP (WO<sub>4</sub>). (200 μM)*

Condition de cristallisation  
tri-Na citrate pH 5.6 à 0.1 M  
13% PEG 4K, 10% glycérol  
NH<sub>4</sub> acétate à 0.2 M  
NaCl à 50 mM

DTT à 20 mM  
PAB0955 à 8 mg.ml<sup>-1</sup>  
GDP à 0.65 mM  
18°C



*Figure.I.12.c. Cristaux de PAB0955 avec GDP (PO<sub>4</sub> et AlF) (230 μM)*

Condition de cristallisation  
MES pH 6 à 0.1 M  
9% PEG 6K  
Lithium Cl à 1 M

DTT à 20 mM  
PAB0955 à 8 mg.ml<sup>-1</sup>  
GDP à 0.65 mM  
18°C  
NaF à 56 mM avec de l'AlCl<sub>3</sub> à 8 mM  
ou PO<sub>4</sub>Na à 20 mM

CHAPITRE II

RESOLUTION DES

STRUCTURES



## II.1. L'expérience de diffraction

### II.1.1. Le phénomène de diffraction

Pour observer une macromolécule à la résolution atomique, il faut utiliser un rayonnement dont la longueur d'onde de l'ordre de la taille d'un atome, c'est-à-dire de l'ordre de l'angström. Ceci explique l'utilisation des rayons X dont les longueurs d'ondes s'étalent de 0.1 à 3 Å. La cristallographie par rayons X utilise le phénomène de diffusion élastique des rayons X par la matière. Le caractère élastique de la diffusion fait que les longueurs d'ondes incidentes et diffusées sont les mêmes.

Les photons X sont absorbés par le nuage électronique des atomes de la macromolécule et réémis sans perte d'énergie avec un déphasage de 180°. La diffraction est un cas particulier de la diffusion, propre aux objets périodiques, avec une diffusion que dans certaines directions de l'espace. La diffraction est due aux interférences constructives qui existent dans des directions particulières, en fonction de la maille du cristal et de son orientation. Dans l'édifice cristallin, la périodicité tridimensionnelle crée des interférences constructives et le grand nombre de mailles ( $10^{11}$  à  $10^{15}$ ) amplifie le signal, le rendant ainsi détectable.

Le phénomène de diffraction est décrit mathématiquement par les équations de Laue, reliant le vecteur de diffusion  $\mathbf{S}$  aux paramètres de maille du cristal décrit par les vecteurs de base du réseau cristallin que sont  $\mathbf{a}$ ,  $\mathbf{b}$  et  $\mathbf{c}$ .

$$\mathbf{S} = \mathbf{s}_0 - \mathbf{s} \quad \left\{ \begin{array}{l} \mathbf{S} : \text{le vecteur de diffusion} \\ \mathbf{s}_0 : \text{la direction du faisceau incident} \\ \mathbf{s} : \text{la direction du faisceau diffracté} \end{array} \right.$$

Lorsque des taches de diffraction sont observées dans une direction  $\mathbf{s}$ , il y a interférence constructive entre les ondes diffractées par chacune des mailles du cristal et les équations de Laue ci-dessous sont vérifiées :

$$\begin{array}{l} \mathbf{S} \cdot \mathbf{a} = h \\ \mathbf{S} \cdot \mathbf{b} = k \\ \mathbf{S} \cdot \mathbf{c} = l \end{array} \quad \left\{ \begin{array}{l} \mathbf{S} : \text{le vecteur de diffusion} \\ \mathbf{a}, \mathbf{b} \text{ et } \mathbf{c} : \text{les vecteurs de base du réseau cristallin} \\ h, k \text{ et } l : \text{des nombres entiers} \end{array} \right.$$

Lors de l'expérience de diffraction d'un cristal, c'est l'intensité  $I_{hkl}$  de la réflexion hkl du faisceau diffracté qui est mesurée. L'intensité est proportionnelle au carré du module du facteur de structure. Ce facteur de structure,  $F_{hkl}$ , est proportionnel à la transformée de

Fourrier de la densité électronique  $\rho(x,y,z)$ . En retour la densité électronique peut être obtenue à partir de l'ensemble des facteurs de structures par la transformation de Fourier inverse.

$$\rho(x,y,z) = \frac{1}{V} \sum_h \sum_k \sum_l F_{hkl} e^{(-2\pi i (hx+ky+lz))}$$

$$\left\{ \begin{array}{l} \rho(x,y,z) : \text{la densité électronique} \\ V : \text{le volume de la maille directe} \\ F_{hkl} : \text{le facteur de structure de la réflexion } hkl \end{array} \right.$$

$$F_{hkl} = |F_{hkl}| e^{i\phi_{hkl}}$$

$$\left\{ \begin{array}{l} F_{hkl} : \text{le facteur de structure de la réflexion } hkl \\ \phi : \text{la phase de la réflexion } hkl \end{array} \right.$$

L'expérience de diffraction d'un cristal permet d'obtenir les intensités de chaque réflexion  $hkl$ , qui permettent de calculer le module des facteurs de structure  $|F_{hkl}|$ . Pour réussir à calculer la densité électronique du modèle, il nous faut encore connaître la phase  $\phi_{hkl}$  de chacune des réflexions. Cette partie est abordée au paragraphe II.2.

### II.1.2. Le traitement des données cristallographiques

Lors de l'enregistrement des clichés de diffraction d'un cristal nous observons sur le détecteur des « taches de diffraction » : il s'agit des pixels du détecteur qui ont été excités par des rayons diffractés. Le nombre de photons reçus par l'ensemble des pixels formant la tache permet de déterminer l'intensité de cette dernière.

À chaque tache de diffraction est attribué un indice  $hkl$  décrivant la direction de l'onde diffractée (indexation). Ensuite, l'amplitude de cette onde est mesurée à partir des intensités  $I_{hkl}$  (intégration).

#### II.1.2.1. Indexation des taches de diffraction

Le principe de l'indexation consiste à rechercher une maille et sa matrice d'orientation permettant d'attribuer à chaque tache de diffraction, repérée sur le détecteur, des coordonnées entières. Ces coordonnées sont les indices  $hkl$  des taches de diffraction. La détermination de la maille permet de sélectionner les groupes d'espace compatibles avec la géométrie de la maille.

### II.1.2.2. Intégration des intensités mesurées

La position de chaque tache étant connue grâce à la maille et la matrice d'orientation, nous pouvons déterminer l'intensité de ces taches. Pour cela, il existe deux façons de procéder : soit par sommation du contenu des pixels après soustraction du bruit de fond, soit par superposition d'un profil d'intégration (profile fitting), méthode que nous avons choisie. Deux logiciels étaient disponibles au laboratoire pour traiter les données cristallographiques : le programme MOSFLM développé par A. Leslie (Leslie, 1991) et le programme XDS développé par W. Kabsch (Kabsch, 1993). Ils diffèrent par leur façon de calculer les profils de taches.

Le programme MOSFLM calcule un profil en deux dimensions des taches de diffraction, méthode optimale lorsque la mosaïcité est de l'ordre ou inférieur au pas d'oscillation (voir après les facteurs de qualité). Un des avantages de MOSFLM est qu'il possède une interface graphique permettant de bien suivre le déroulement du traitement de données.

Le programme XDS au contraire est une sorte de « boîte noire », qui calcule un profil tridimensionnel des taches de diffraction, permettant une meilleure intégration dans le cas de cristaux mosaïque (lorsque la mosaïcité est supérieure ou égale au pas d'oscillation). Dans la plupart des cas les deux programmes donnent des résultats similaires. L'ensemble de nos jeux de données ont été traités avec le programme XDS.

### II.1.2.3. Mise à l'échelle

Dans le cas d'un système parfait, les réflexions équivalentes par symétrie doivent avoir la même intensité. Plusieurs paramètres font que les réflexions équivalentes n'ont pas la même intensité. Par exemple, au fur et à mesure de l'enregistrement de données, le cristal subit des dégradations dues aux rayons X, qui diminuent l'intensité des taches diffractées. Des intensités différentes peuvent également être provoquées par la variation d'intensité du rayonnement synchrotron. Lors d'enregistrement de jeux de données à l'E.S.R.F., les réflexions équivalentes ne vont pas recevoir exactement la même intensité de rayons X, leurs taches de diffraction n'auront donc pas les mêmes intensités. Il est donc important de mettre les données à l'échelle afin de minimiser globalement l'écart entre les intensités des réflexions équivalentes par symétrie. Après mise à l'échelle, les intensités moyennes peuvent être calculées à partir des mesures d'intensités des taches équivalentes par symétrie.

#### II.1.2.4. Quelques facteurs de qualité

Au cours du traitement des clichés de diffraction, les programmes utilisés donnent différents facteurs statistiques permettant de contrôler la qualité de diffraction du cristal.

##### La résolution

Selon la résolution du jeu de données, nous aurons une structure plus ou moins détaillée à l'échelle atomique. Plus la résolution est haute plus la structure obtenue est précise. La résolution dépend de la qualité du cristal et de la configuration de l'expérience. La limite de résolution généralement utilisée est celle pour laquelle le rapport signal sur bruit est supérieur ou égal à 2 dans la dernière coquille de résolution.

##### Le rapport signal sur bruit : $I/\sigma(I)$

Plus le signal est élevé par rapport au bruit de fond, meilleures sont les estimations des intensités et donc des facteurs de structure. Ce facteur rend compte de la précision des mesures et permet de déterminer le seuil de coupure en résolution d'un jeu de données.

##### Le facteur $R_{\text{sym}}$ ( $R_{\text{merge}}$ lorsqu'il y a plusieurs cristaux)

Le  $R_{\text{sym}}$  représente le désaccord entre les intensités des réflexions équivalentes par symétrie. Plus il est bas, meilleur est l'accord. Classiquement, au-dessus de 10% pour la valeur globale du facteur  $R_{\text{sym}}$ , la qualité des données est discutable. Dans la dernière coquille de résolution, nous nous sommes limités à un facteur  $R_{\text{sym}}$  d'une valeur maximale de 40%.

$$R_{\text{sym}} = \frac{\sum_{hkl} \sum_{i=1, n_{hkl}} |I_{hkl}^i - \langle I_{hkl} \rangle|}{\sum_{hkl} \sum_{i=1, n_{hkl}} I_{hkl}^i}$$

$$\begin{cases} hkl & : \text{les indices de la réflexion } hkl \\ I_{hkl} & : \text{l'intensité de la réflexion } hkl \\ \langle I_{hkl} \rangle & : \text{la moyen de la réflexion unique } hkl \end{cases}$$

Si un jeu de données est indexé dans le mauvais groupe d'espace, les opérateurs de symétrie appliqués ne sont pas les bons, le facteur  $R_{\text{sym}}$  augmente très sensiblement car les réflexions considérées comme équivalentes par symétrie ne le sont pas. Au cours de l'enregistrement, lorsque le cristal se dégrade, l'intensité diffractée diminue et la valeur du facteur le  $R_{\text{sym}}$  augmente.

### La mosaïcité

Le cristal de protéine n'est pas un édifice cristallin parfait ; il faut plutôt le considérer comme l'association de plusieurs petits blocs cristallins légèrement désorientés les uns par rapport aux autres. La mosaïcité représente la différence moyenne, exprimée en degré, qui existe entre les différents blocs. Plus la mosaïcité est grande, plus la tache de diffraction est étalée sur le détecteur (x, y) et apparaît sur un domaine étendu en  $\varphi$  (angle d'oscillation) et, en conséquence plus le signal recueilli par chaque pixel est faible. La mosaïcité provient de différentes origines. Au cours de la croissance cristalline, les molécules constituant le cristal peut ne pas s'empiler comme elle le devraient. Au moment de la congélation des cristaux, les tensions subies peuvent également augmenter la mosaïcité. Lorsque la mosaïcité est trop élevée, les taches de diffraction peuvent s'élargir au point de se superposer avec leur voisine, empêchant alors une estimation correcte de leur intensité.

### La complétude

La complétude est le rapport entre le nombre de réflexions uniques effectivement mesurées et le nombre théorique de réflexions uniques qui pourraient être mesurées à une résolution donnée. Ce facteur, exprimé en pourcentage, doit être le plus élevé possible, typiquement au-dessus de 90%.

### La redondance

La redondance est le nombre de fois où l'intensité d'une réflexion unique est mesurée. Plus une réflexion est mesurée un grand nombre de fois, meilleure sera l'estimation de son intensité moyenne  $I$ , plus petite sera celle de son écart type  $\sigma(I)$  et donc meilleure sera la qualité du jeu de données.

## II.2. La détermination de la structure de PAB0955

Pour résoudre une structure par cristallographie aux rayons X, il faut calculer la carte de densité électronique de la macromolécule. Dans ce calcul, il y a deux éléments essentiels : les modules des facteurs de structures des réflexions hkl et les phases. Seules les intensités (reliées aux amplitudes des facteurs de structure  $|F_{hkl}|$ ) des faisceaux diffractés sont enregistrés ; l'information de phase des ondes est perdue.

$$F_{hkl} = |F_{hkl}| e^{i\phi_{hkl}} \quad \phi_{hkl} \text{ la phase}$$



Il existe plusieurs méthodes pour déterminer les phases des facteurs de structure. Dans notre cas, nous avons utilisé le signal de différence isomorphe et le signal anomal.

### II.2.1. Utilisation du signal de différence isomorphe

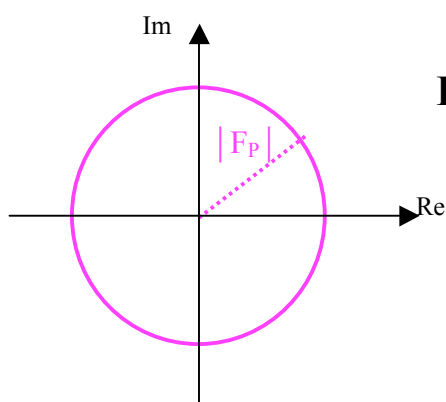
La détermination des phases par remplacement isomorphe multiple, MIR, est la méthode la plus ancienne mise au point par Perutz en 1956 (Perutz, 1956). Cette technique a notamment permis de résoudre les deux premières structures cristallographiques de protéine : l'hémoglobine et la myoglobine en 1960 (Perutz, 1960). Le but du remplacement isomorphe est d'introduire un atome lourd dans un cristal natif pour perturber les intensités et les phases des réflexions sans perturber la maille et l'empilement cristallin. Ensuite les signaux enregistrés avec et sans atome lourd sont comparés, afin de déterminer la contribution du (ou des) atome(s) lourd(s) au signal de diffraction.

#### II.2.1.1. Enregistrement d'un jeu de données sur un cristal natif

Les cristaux natifs de PAB0955 obtenus après optimisation des conditions de cristallisation, ont permis d'enregistrer un jeu complet de données à 2.15 Å (fig.II.6.a.). Les cristaux ont été congelés dans l'azote liquide après avoir été plongés dans une solution cryo-protectrice contenant la solution mère enrichie à 30% de PEG4000 (w/v).

Le traitement de ce jeu de données a été effectué avec le programme XDS. Il nous a permis de définir que les cristaux appartenant à un groupe d'espace trigonal, sans toutefois pouvoir discriminer entre les groupes  $P3_221$  et  $P3_121$ . Seules les phases finales obtenues permettront de discriminer entre ces groupes d'espace.

Les données natives permettent de calculer  $|F_P|$ , et de tracer un cercle de rayon  $|F_P|$  dans le plan complexe (diagramme d'Argan). L'objectif est de pouvoir déterminer la contribution des atomes lourds soit  $F_H$ . Si l'isomorphisme est vérifié entre le cristal natif et le cristal dérivé la relation ci-dessous est vérifiée :



$$\mathbf{F}_{PH} = \mathbf{F}_P + \mathbf{F}_H$$

$F_{PH}$  : facteur de structure du dérivé  
 $F_P$  : facteur de structure de la protéine  
 $F_H$  : facteur de structure de l'atome lourd

II.2.1.2. Enregistrement d'un jeu de données sur un cristal dérivé

Les cristaux trempés dans une solution de complexe de gadolinium ou ceux obtenus par co-cristallisation, peuvent servir de cristaux dérivés pour enregistrer le signal isomorphe. Dans notre cas ces cristaux sont isomorphes aux cristaux natifs. Des jeux de données ont été enregistrés sur les deux types de cristaux dérivés (fig.II.4. et fig.II.5.b.). Après avoir traité les jeux de données, il faut déterminer la présence d'atome(s) lourd(s) dans les cristaux dérivés. Pour cela, on utilise la fonction de Patterson qui permet de déterminer la position et le nombre d'atomes lourds fixés.

La fonction de Patterson  $P(x,y,z)$  est la fonction d'auto-corrélation de la densité électronique, elle se calcule facilement par la formule suivante :

$$P(x,y,z) = \frac{1}{V} \sum_{hkl} (F_{hkl})^2 e^{2\pi i(hx + ky + lz)}$$

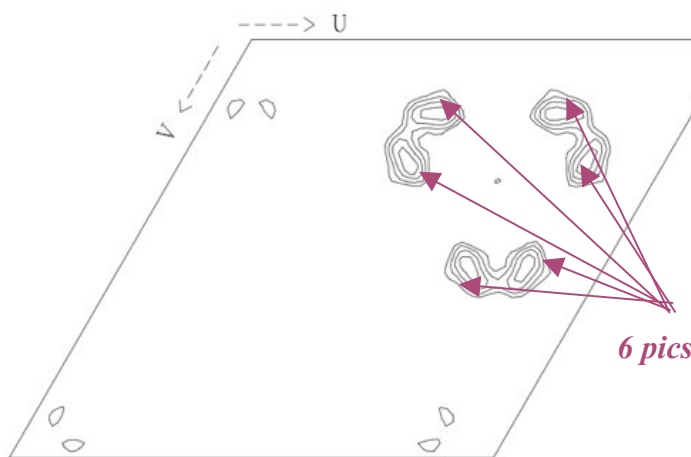
Cette fonction représente en quelque sorte les vecteurs inter-atomiques d'un cristal, chacun d'entre eux correspondant à un pic dans la fonction de Patterson. La hauteur des pics, issus de ces vecteurs inter-atomiques est proportionnelle au produit des numéros atomiques  $Z$  des atomes concernés. Les pics de la fonction de Patterson sont très nombreux en raison du grand nombre d'atomes contenus dans la protéine, ils se superposent et ne sont donc pas interprétables directement. Le calcul de la fonction de différence de Patterson peut être considérée comme la fonction de Patterson de l'atome lourd seul.

$$P(x,y,z) = \frac{1}{V} \sum_{hkl} (|F_{PHkl}| - |F_{Pkl}|)^2 e^{2\pi i(hx + ky + lz)}$$

En effet dans l'hypothèse où  $F_H \ll F_P$ , la différence  $(|F_{PH}| - |F_P|)^2$  est approximativement proportionnelle à  $|F_H|^2$ . Cette fonction de Patterson va donc nous permettre de localiser les

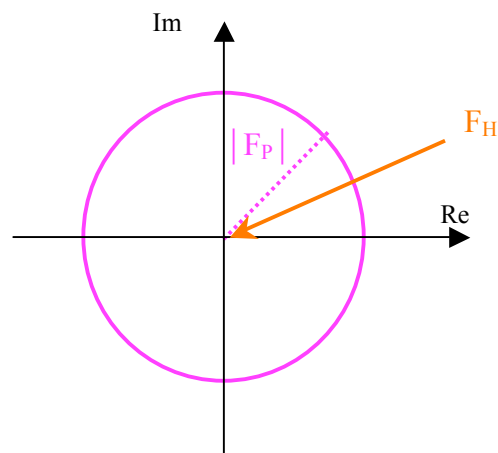
atomes lourds dans la maille du cristal.

Il existe des sections particulières dans la fonction de Patterson qui sont les sections de Harker. Dans ces sections, se trouvent les pics correspondant aux vecteurs inter-atomiques entre un atome et son symétrique par l'un des opérateur

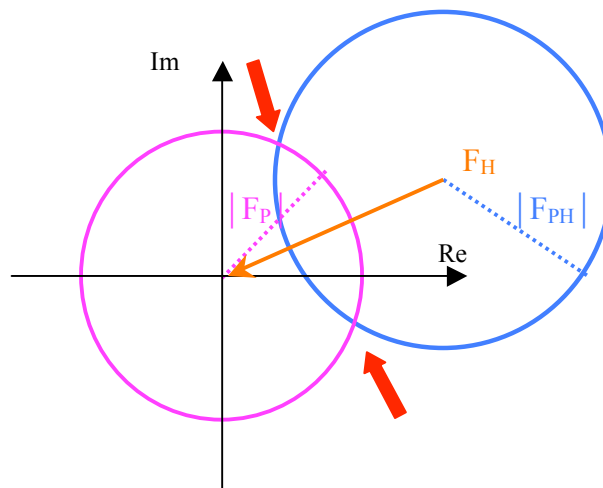


*Figure.II.1. Carte de Patterson différence isomorphe. Section de Harker  $w=1/3$*

de symétrie du cristal. Nous pouvons savoir, grâce au nombre de pics, combien d'atomes lourds sont fixés par unité asymétrique. Il y a 6 pics dans la section de Harker, qui correspondent aux plus hauts pics. PAB0955 cristallise dans un groupe d'espace ( $P3_221$  ou  $P3_121$ ) contenant 6 positions équivalentes, il y a donc un ion gadolinium fixé sur notre protéine par unité asymétrique. Les pics étant distincts, les coordonnées de l'atome lourd peuvent être calculées à partir des coordonnées des pics. Connaissant la position de l'atome lourd, le vecteur  $F_H$  est connu en amplitude et en phase ; nous pouvons donc le tracer en pointant son extrémité à l'origine de notre diagramme, ceci afin de respecter la fermeture du triangle formé entre les trois vecteurs ( $F_{PH} = F_P + F_H$ ).



Avec le cristal dérivé,  $|F_{PH}|$  est connu. Nous pouvons donc tracer un cercle de rayon  $|F_{PH}|$  dont l'origine est la même que celle du vecteur  $F_H$ .



Les flèches rouges indiquent les deux points d'intersections des deux cercles ; ceci signifie qu'il y a deux solutions possibles pour les phases. En utilisant un second dérivé isomorphe et en appliquant la même construction, les trois cercles ont un seul point d'intersection et l'indétermination est alors levée. Cette ambiguïté peut aussi être levée en utilisant la diffusion anormale du gadolinium, qui en plus d'être un atome lourd est un diffuseur anormal.

## II.2.2. Utilisation du signal anomal

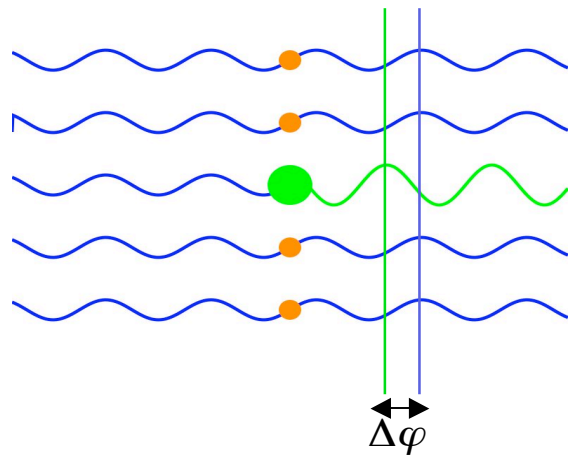
### II.2.2.1. Principe du signal anomal

Le phénomène de diffusion anormale est dû au fait que le facteur de diffusion atomique ( $f$ ) d'un atome varie fortement, en module et en phase, dans la région du seuil d'absorption de l'atome. Avec le signal anomal, l'isomorphisme entre les cristaux n'est plus un problème puisque les données peuvent être enregistrés, à des longueurs d'onde différentes sur un cristal unique. La détermination des phases par le signal anomal se fait par une technique d'enregistrement à 3 longueurs qui correspondent : au maximum de  $f''$ , au minimum de  $f''$  et une longueur d'onde loin du seuil d'absorption ; c'est la méthode MAD (multiple wavelength anomalous diffraction) (Pahler et al., 1990). Les cristaux utilisés pour déterminer les phases par la méthode MAD ne nous ont pas permis d'obtenir une qualité de phase suffisante pour construire un modèle. Dans certain cas une seule longueur d'onde, pour laquelle le facteur de diffusion présente une partie imaginaire ( $f''$ ) élevée suffit : c'est la méthode SAD (single wavelength anomalous diffraction). Nous avons utilisé la méthode SAD sur un cristal co-cristallisé avec un diffuseur anomal (le gadolinium) et obtenus des phases exploitables pour la construction du modèle de PAB0955.

#### *Le déphasage*

Dans le modèle de diffusion des rayons X par des atomes diffusant normalement, le rayonnement diffusé est déphasé de  $180^\circ$  par rapport au rayonnement incident. Dans ce cas, le facteur de diffusion  $f$  est réel. Toutefois lorsque la longueur d'onde utilisée correspond au seuil d'absorption d'un atome lourd dans le cristal, il y a alors absorption du rayonnement incident par les électrons des couches profondes de l'atome lourd.

Ces électrons sont alors excités ce qui les amène dans le continuum au delà de l'énergie d'ionisation. Le rayonnement diffracté par le diffuseur anomal (en vert) est déphasé,  $\Delta\phi$ , par rapport aux autres rayonnements diffractés par les atomes « légers » (en bleu) (fig.II.2.).



*Figure.II.2. Représentation schématique du déphasage*

Le facteur de diffusion

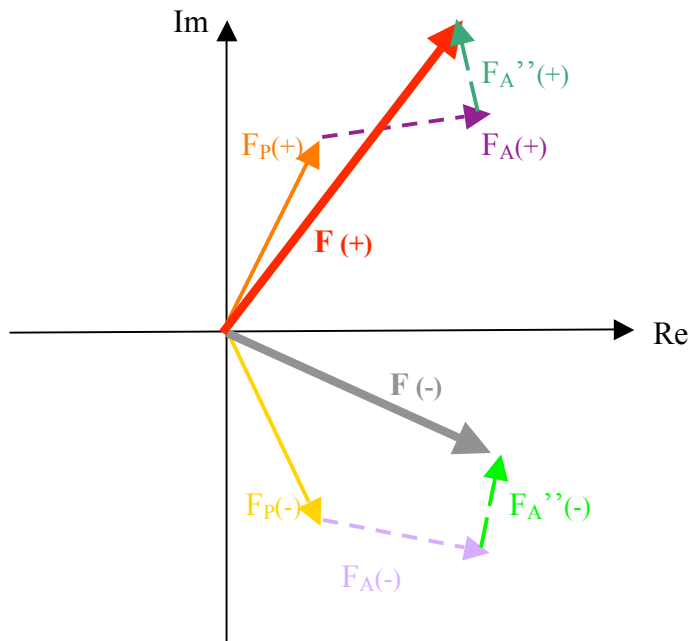
Pour tenir compte du déphasage le facteur de diffusion  $f$  est représenté par un nombre complexe qui dépend de la longueur d'onde  $\lambda$  du rayonnement incident :

$$f = f^o + f'(\lambda) + if''(\lambda)$$

Lorsque le rayonnement incident est loin du seuil d'absorption de l'atome  $f'$  et  $f''$  sont négligeables. Si la longueur d'onde utilisée est proche du seuil d'absorption d'un atome,  $f'$  et  $f''$  prennent des valeurs significatives.

Le déphasage, dû au diffuseur anormal, brise la loi de Friedel selon laquelle, les modules des facteurs de structure  $|F(hkl)|$  et  $|F(-h-k-l)|$  sont égaux. Lorsque la loi de Friedel est brisée nous obtenons :

$$\begin{aligned} F_{PA}(+) &= F_P + F_A + iF_A'' \\ F_{PA}^*(-) &= F_P + F_A - iF_A'' \end{aligned} \quad \left\{ \begin{array}{l} F_{PA}^*(-) : \text{complexe conjugué de } F_{PA}(+) \\ F_A'' \text{ dépend de } f'' \\ F_A \text{ dépend de } f^o \text{ et de } f' \end{array} \right.$$



Graphiquement, cela entraîne la représentation ci-contre, avec  $F_A''$  décalé de  $+90^\circ$  par rapport à  $F_A$ , donc  $F(+)$  et  $F^*(-)$  ne sont pas symétriques et  $|F(+)|$  est différent de  $|F(-)|$ .

Avec le même type de diagramme d'Argand que pour celui utilisé dans le cas du signal isomorphe, il est possible de tracer les différents vecteurs des facteurs de structures et de déterminer  $F_p$ . Dans le cas d'un enregistrement à plusieurs longueurs d'ondes,  $f'$  et  $f''$  dépendant de la longueur d'onde, nous avons différentes valeurs de  $F_{PA}(+)$  et de  $F_{PA}^*(-)$ , ce qui lève l'indétermination de phase.

### II.2.2.2. Tentative de phasage par la méthode MAD

L'atome de gadolinium a la particularité d'être à la fois un atome lourd (64 électrons) et un bon diffuseur anormal dont le seuil d'absorption est accessible en cristallographie. À son seuil d'absorption  $L_{III}$  ( $\lambda = 1.711 \text{ \AA}$ ) la composante imaginaire du facteur de diffusion du gadolinium,  $f''$ , vaut  $28e^-$ ; elle vaut  $12 e^-$  à la longueur d'onde  $K\alpha$  du cuivre ( $\lambda = 1,54 \text{ \AA}$ ) qui est celle du rayonnement produit par les générateurs de laboratoire. Il est donc possible de déterminer les phases avec une source de rayons X de laboratoire si un atome de gadolinium est fixé dans le cristal. Le complexe de gadolinium que nous avons testé, en collaboration avec Meike Stelter et Richard Kahn du Laboratoire de Cristallographie des Macromolécules de l'IBS, contient un ion  $Gd^{3+}$  (en cyan) lié à un acide diéthylènetriamine-pentaacétique-

*bis*(méthylamide) DTPA-BMA (fig.II.3.) (Ehnebom and Pedersen, 1992).

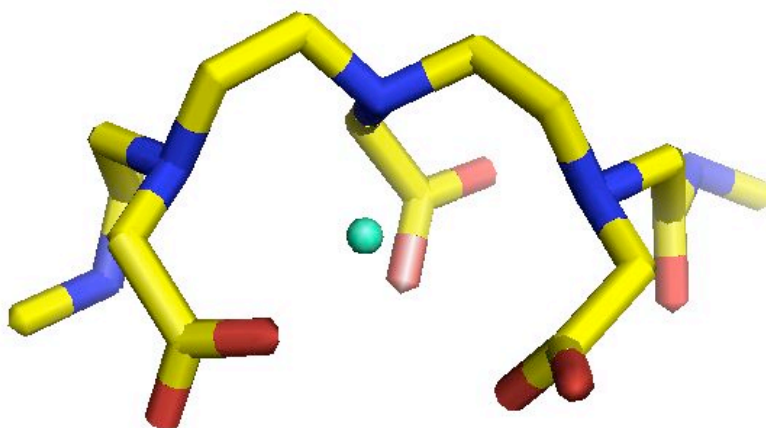


Figure.II.3. Structure du complexe gadolinium-DTPA-BMA

compatible avec celle des canaux de solvant des cristaux de protéines. En plus de leurs propriétés intéressantes pour déterminer les phases, ces complexes ont l'avantage de ne pas être toxique et donc d'être facilement manipulables.

La fixation du complexe dépend de la nature de la cage, non de l'interaction entre l'ion gadolinium et la protéine. Il existe plusieurs types de cages (Girard et al., 2003) qui ont été étudiées pour que leur taille soit

Les premiers tests de fixation du Gd-DTPA-BMA ont été fait par trempage de cristaux natifs de PAB0955 dans la solution de gadolinium. Étant donné que nous avons des cristaux de la protéine native disponibles, le trempage était la technique la plus rapide et la plus facile à tester.

Sur un cristal natif de  $170 \mu\text{m}$  trempé pendant 2 jours dans une solution de Gd-DTPA-BMA à  $100 \text{ mM}$ , nous avons enregistré des jeux de données à 3 longueurs d'ondes (fig.II.4.).

	Jeu 1 (max f')	Jeu 2	Jeu 3 (mini f')
Résolution (Å)	50-3.20 (3.3-3.2)	50-3.00 (3.1-3.0)	50-3.30 (3.4-3.3)
Rsym (%)	8.3 (28.4)	10.1 (27.5)	7.9 (36.8)
Complétude (%)	97.9 (93.9)	97.5 (97.3)	98.2 (95.4)
I/σI	<b>13.57 (5.38)</b>	12.83 (6.06)	<b>12.05 (3.76)</b>
Redondance	4.1	6.5	3.9
Facteur d'agitation thermique B (Å <sup>2</sup> )	67	80	87
Paramètres de maille (Å)	a=b=61.06 c=117.38	a=b=60.99 c= 117.43	a=b=61.25 c= 118.07
Longueur d'onde (Å)	1.7107	1.3000	1.7117
Temps d'exposition par degré	180 s	120 s	180 s

*Figure.II.4. Statistiques cristallographiques sur le cristal dérivé obtenu par trempage.*

Le premier jeu à avoir été enregistré est celui qui correspond au seuil  $L_{III}$  du gadolinium. Nous avons remarqué qu'avec le même temps d'exposition (180 secondes), l'intensité totale et celle de la dernière coquille de résolution sont moins élevées pour le jeu 3 (en bleu) que pour le jeu 1 (en rouge). Le cristal semble donc s'altérer au cours des enregistrements successifs. Le changement des paramètres de maille indique un changement de l'empilement cristallin. Le troisième jeu de données a ses paramètres de maille a et b qui augmentent de 0.2 Å ; c augmente de 0.6 Å. Ce changement de maille montre que le cristal se dilate avec l'irradiation de l'échantillon.

Pour vérifier l'isomorphisme de ces trois jeux de données, il faut les fusionner. Après la fusion des trois jeux, le facteur  $R_{sym}$  total est de 15% à 3 Å. Entre 50 et 10 Å, le facteur  $R_{sym}$  est de 11% pour les jeux fusionnés et d'environ 5% dans les jeux séparés. Il y a donc un défaut d'isomorphisme entre ces jeux, lié à l'irradiation du cristal pendant l'enregistrement.

Le signal anomal est un signal faible qui provient de la différence  $|F(+)| - |F(-)|$ . La valeur du facteur  $R_{sym}$  élevé indique un rapport  $I/\sigma(I)$  plus faible qui risque de masquer le signal anomal. La résolution peu élevée (3 Å) et la mosaïcité de l'ordre de 1° sont probablement dues au trempage, et la dégradation du cristal au cours des enregistrements, font que les phases calculées à partir de ces jeux de données n'ont pas suffi à fournir une carte de densité électronique interprétable.

Les essais de détermination des phases ont été faits avec le programme SHARP (De la Fortelle and Bricogne, 1997). Le taux d'occupation du gadolinium était faible et son facteur d'agitation thermique B était supérieur à 200 Å. La présence de signal anomal prouve la fixation de gadolinium.

Le site de fixation de l'ion gadolinium a pu être déterminé par la carte de différence de Patterson anormale (fig.II.5.a.) calculée avec le programme SHELX (Uson and Sheldrick, 1999).

Comme avec le signal de différence isomorphe, un site de gadolinium a été observé par monomère de protéine. Le gadolinium est donc capable de se fixer sur la protéine PAB0955. Ce résultat était donc encourageant pour tester une co-cristallisation du gadolinium avec notre protéine.

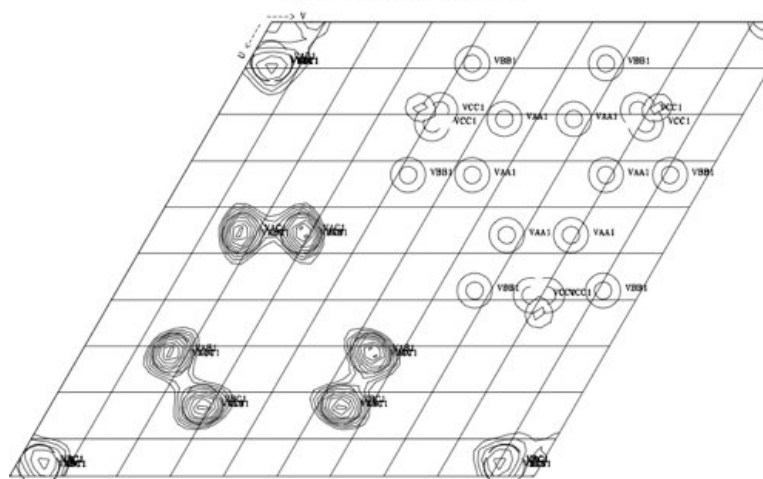


Figure.II.5.a. Carte de différence de Patterson anormale. Section de Harker  $w=1/3$ .

Résolution (Å)	50-2.14
$R_{\text{sym}}$ (%)	5.5
Complétude (%)	99.2
$I/\sigma(I)$	36.39
Redondance	25.6
Paramètres de maille (Å)	a=b=60.47 c=117.22
Longueur d'onde (Å)	1.278

Figure.II.5.b. Statistiques cristallographiques du cristal dérivé obtenu par co-cristallisation.

### II.2.2.3. Réussite du phasage par la méthode SIRAS

Afin d'augmenter le taux d'occupation du gadolinium dans nos cristaux et stabiliser sa fixation, la co-cristallisation de PAB0955 en présence de Gd-DTPA-BMA a été réalisée. Les différentes expériences effectuées par Meike Stelter sur différentes protéines avec des complexes de gadolinium, ont montré que la présence du complexe ne changeait pas les conditions de cristallisation de la protéine. Nous avons obtenu par co-cristallisation des cristaux dérivés dans les mêmes conditions que celle de la protéine native auxquelles s'ajoutent du Gd-DTPA-BMA à 100 mM. Un jeu de données à une seule longueur d'onde



( $\lambda = 1,278 \text{ \AA}$ ) a été enregistré sur un cristal dérivé. Ce cristal diffracte jusqu'à  $2.14 \text{ \AA}$  (fig.II.6). Le même site de fixation de l'ion gadolinium a été déterminé dans les cristaux dérivés co-cristallisés que dans ceux trempés.

Le pouvoir de phasage dû au signal anomal de ce jeu de données, déterminé par le programme SHARP (De la Fortelle and Bricogne, 1997), est de 0.827 (signal anomal sur les réflexions acentriques). Ce jeu de données aurait suffi à déterminer les phases puisqu'il est de très bonne qualité. Toutefois nous disposons d'un bon jeu de données natives et des jeux à 3 longueurs d'ondes ; nous avons tout de même combiné l'information de phase de tous les jeux. Ceci a permis de déterminer les phases avec plus de précision.

Les phases étant de très bonne qualité à une résolution de  $2.14 \text{ \AA}$ , nous avons pu débiter la construction d'un premier modèle de façon automatique. Deux programmes différents ont été utilisés pour cette construction, d'un côté ARP/WARP (Morris et al., 2002) après aplatissement de solvant par SOLOMON-DM (1994) et d'autre part le programme RESOLVE (Terwilliger, 2003). Le modèle construit par ARP/WARP contenait 125 acides aminés sur 262, et celui construit par REOLVE en contenait 157. En combinant ces deux modèles, nous avons pu obtenir 184 acides aminés construits, ce qui représente 70% de notre protéine, dont 64% était construit avec les chaînes latérales. Après quelques cycles d'affinement de ce modèle avec le programme REFMAC (Murshudov et al., 1997), les statistiques ont montré que la solution était correcte (facteur  $R = 0.372$  et  $R_{\text{free}} = 0.389$ ). Le reste du modèle, soit 78 acides aminés a été construit manuellement.

### II.2.3. Construction de la structure de PAB0955

Après quelques cycles d'affinement avec le programme REFMAC (Murshudov et al., 1997) sur le modèle obtenu automatiquement, l'accord entre nos données de diffraction et celles calculées à partir de notre structure est bon (facteur  $R$  de 37.2%, facteur  $R_{\text{free}}$  de 38.9%).

Le reste du modèle fut construit manuellement en utilisant le programme de visualisation O (Jones et al., 1991) pour construire et le programme REFMAC pour affiner les différents modèles intermédiaires. Les cartes de densité électronique utilisées pour la construction du modèle sont les cartes différences pondérées  $2mF_{\text{obs}} - DF_{\text{calc}}$  et  $mF_{\text{obs}} - F_{\text{calc}}$ .

Parmi les parties construites manuellement dans PAB0955, il y avait une grande région (résidus 195 à 217) avec une densité électronique très discontinue. Au vu du peu d'information dans cette région et de la difficulté que j'ai eu à construire un modèle dans cette région, il n'est pas étonnant que les deux programmes automatiques utilisés n'aient pu nous donner d'information sur cette région.

En fin d'affinement le modèle de la protéine native contenait 248 acides aminés sur 262 et 61 molécules d'eau (code PDB : 1YR6). Les acides aminés manquant dans le modèle sont les dix premiers appartenant à l'étiquette histidine et les quatre derniers pour lesquels il n'y a pas de densité électronique. En fin d'affinement, la région contenant les acides aminés 193 à 217 possède une densité électronique uniquement autour de la chaîne principale. Le facteur d'agitation thermique B est élevé, 100 Å, le double du facteur B moyen de la structure.

La structure contenant l'atome de gadolinium et sa cage est la même que celle de la protéine native, mise à part la présence du Gd-DTPA-BMA qui nous a permis de déterminer les phases.

#### *Validation du modèle*

Les facteurs R et Rfree sont deux facteurs qui permettent de contrôler l'accord entre les facteurs de structures observés et calculés.

$$R = \frac{\sum_{hkl} ||F_{obs}| - |F_{calc}||}{\sum_{hkl} |F_{obs}|}$$

Le facteur R diminue nécessairement au cours de l'affinement car on optimise l'accord entre les  $|F_{obs}|$  et  $|F_{calc}|$ . Aux résolutions classiques (2 à 3 Å), le nombre d'observation n'est pas largement supérieur au nombre de paramètres, il est important de contrôler que le modèle représente une interprétation objective des données cristallographiques. Un autre facteur moins biaisé est le facteur Rfree calculé sur un ensemble de réflexions exclu de la procédure d'affinement. Ces réflexions, choisies de manière aléatoire, représentent en général 5 à 10% des réflexions totales (Brunger, 1993).

Les statistiques sur les grandeurs stéréochimiques nous indiquent si le modèle est géométriquement acceptable. Les contraintes stéréochimiques lors de l'affinement sont pondérées afin d'obtenir des écarts moyens inférieurs à 1.5° sur les angles, à 0.015Å sur les

liaisons inter-atomiques. Pour vérifier la stéréochimie de la chaîne principale, on utilise le diagramme de Ramachandran (Ramakrishnan and Ramachandran, 1965). Ce diagramme contient l'ensemble des valeurs des angles  $\phi$  et  $\psi$  autorisées pour la liaison peptidique.

### II.2.3. Le remplacement moléculaire

Une autre technique pour résoudre le problème des phases est d'avoir un modèle dont la structure proche de celle de notre protéine : c'est le remplacement moléculaire. Nous avons utilisé le remplacement moléculaire pour résoudre les structures de PAB0955 en complexe avec différents nucléotides. La structure de notre protéine native sans nucléotide a servi de modèle pour déterminer les autres structures.

Le remplacement moléculaire a été effectué avec le programme AMoRe (Automatic Molecular Replacement) développé par Jorge Navaza (Navaza, 1994). Le but du remplacement moléculaire est de placer le modèle dans la maille de sorte à obtenir le meilleur accord entre les facteurs de structure calculés à partir du modèle et ceux observés pour la protéine d'intérêt. Le programme AMoRe teste de façon automatique et systématique différents paramètres de rotation puis de translation du modèle choisi. La recherche de solution se déroule en six étapes successives, exécutées par les sous-programme suivants:

*SORTING* : trie les modules des facteurs de structures mesurés,

*TABLING* : oriente et positionne le modèle à l'origine d'une maille triclinique optimale et calcule les facteurs de structure correspondant,

*ROTING* : calcule la fonction de rotation pour différentes orientations et sélectionne les meilleures solutions,

*TRAINING* : calcule la fonction de translation pour les orientations sélectionnées et sélectionne les meilleures solutions,

*FITING*: affine par la méthode des moindres carrés, les solutions obtenues en considérant le modèle comme un ou plusieurs corps rigides,

*MR2IC* : calcule la matrice de rotation et le vecteur de translation à appliquer au modèle initial.

Le fichier MR2IC contient également deux facteurs statistiques permettant d'apprécier la qualité de la solution de remplacement moléculaire. Un facteur de corrélation entre les

facteurs de structure calculés ( $F_{\text{calc}}$ ) et les facteurs de structures observés ( $F_{\text{obs}}$ ), qui doit être le plus élevé possible et le facteur R qui doit être le plus faible possible.

Après avoir obtenu notre structure de la protéine apo, une fois l'analyse structurale débutée, nous nous sommes demandé si tout de même le remplacement moléculaire aurait pu être utilisé pour résoudre notre structure. En se fondant sur l'homologie de séquence pourtant faible et la similitude structurale des GTPases à P-loop, comme nous l'avons testé avec hRas. Le remplacement moléculaire avec hRas, par le programme AMoRe, nous a donné un facteur de corrélation de 25,8% et un facteur R de 51,4%. Avec la structure la plus proche obtenue par le programme DALI, la déthiobiotine synthase le facteur de corrélation est de 28% et le facteur R de 51,6%. Dans les deux cas, la corrélation est relativement faible, la reconstruction du modèle que nous aurions obtenu aurait sans doute pris beaucoup plus de temps.

	Native	GTP $\gamma$ S	GTP <sup>EDTA</sup>	GTP <sup>4°C</sup>	GTP <sup>18°C</sup>	GDP	MgGDP
Groupe d'espace	<b><i>P</i><sub>3</sub><i>2</i><sub>1</sub></b>	<b><i>P</i><sub>3</sub><i>2</i><sub>1</sub></b>	<b><i>P</i><sub>3</sub><i>2</i><sub>1</sub></b>	<b><i>P</i><sub>2</sub><i>1</i><sub>2</sub><i>1</i><sub>2</sub></b>	<b><i>P</i><sub>3</sub><i>2</i><sub>1</sub></b>	<b><i>P</i><sub>2</sub><i>1</i></b>	<b><i>P</i><sub>2</sub><i>1</i></b>
Paramètres de maille (Å)	a=b=60.3 c=117.1	a=b=60.2 c=115.9	a=b=60.7 c=116.8	a=58.8 b=84.0 c=53.1	a=b=60.8 c=117.2	a=59.1 b=84.9 c=60.2 $\beta=95^\circ$	a=59.4 b=85.2 c=60.5 $\beta=94.6^\circ$
Résolution (Å)	50-2.15 (2.23-2.15)	50-2.08 (2.15-2.08)	20-2.40 (2.46-2.40)	100-2.80 (2.90-2.80)	100-2.40 (2.45-2.40)	50-2.30 (2.38-2.30)	50-1.75 (1.80-1.75)
R <sub>sym</sub> (%)	5.7 (34.8)	6.7 (28.4)	7.5 (40)	9.3 (45.5)	6.7 (40.4)	6.8 (36.1)	9.8 (37.2)
Complétude (%)	100 (100)	87.6 (94)	98.5 (98.2)	97.8 (93.8)	99 (98.8)	99.9 (100)	98.6 (93.5)
$\langle I \rangle / \langle \sigma(I) \rangle$	8.2 (2.2)	6.6 (2.3)	20.7 (7)	15.19 (4.6)	19.55 (5.76)	5.2 (2.1)	12.47 (5.91)
Ligne de lumière (ESRF)	ID14-eh2	BM30-A	ID14-eh1	BM30-A	BM30-A	générateur du laboratoire	ID14-eh2
Longueur d'onde (Å)	0.933	0.979	0.934	0.979	0.979	1.548	0.934
Code PDB	1YR6	1YR7	1YR8	1YR9	Non déposé	1YRA	1YRB

Figure.II.6.a. Tableau des statistiques cristallographiques.

Facteur de Corrélation (%)		73.5	57.5	67.9	63	69.4	66.5
Facteur R (%)		33.4	40.8	39.5	40.5	33.7	35.8

Figure.II.6.b. Tableau des statistiques de remplacement moléculaire.

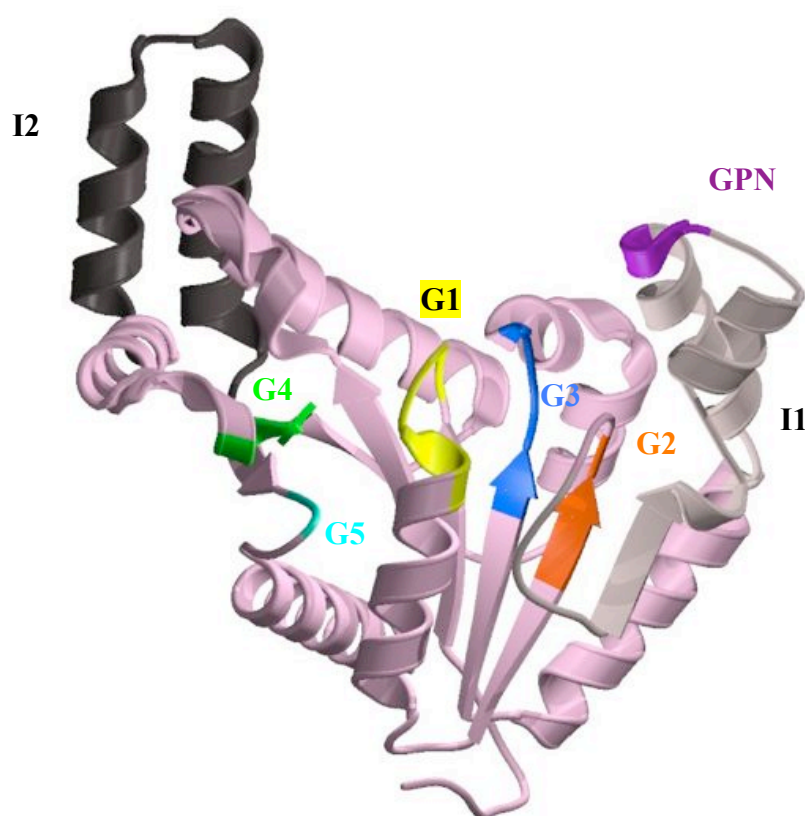
## II.3. Les structures de PAB0955

### II.3.1. La structure de la protéine sans nucléotide

#### II.3.1.1. Description de la structure

Les cristaux de la protéine apo avec ou sans Gd-DTPA-BMA appartiennent au groupe d'espace  $P3_221$ . Le Vm indique la présence d'un monomère par unité asymétrique. Le dimère biologique est formé dans le cristal grâce à la symétrie cristallographique d'ordre 2.

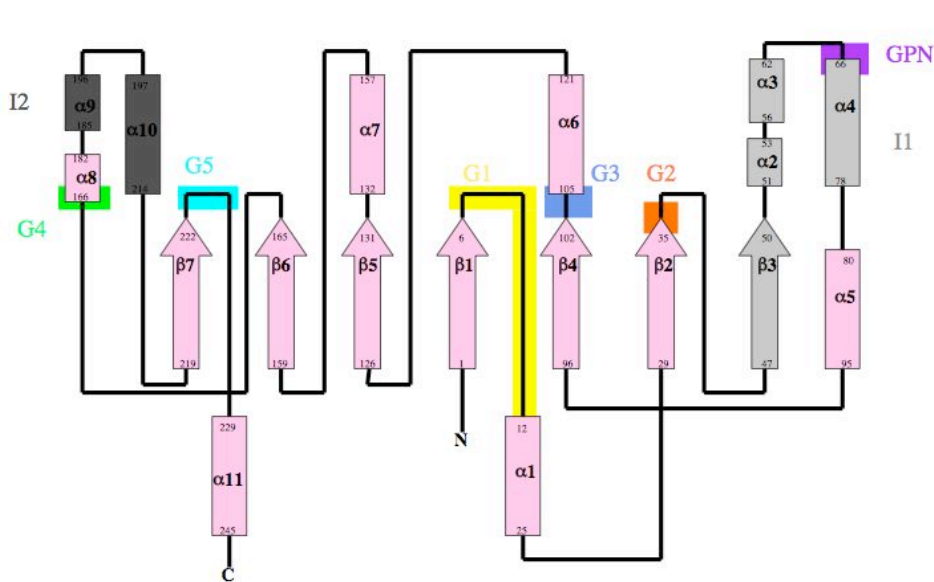
La structure de PAB0955 adopte un repliement de type Rossmann, classique pour les protéines liant et hydrolysant le GTP. Le monomère de PAB0955 est composé d'un feuillet central parallèle constitué de 7 brins- $\beta$  pris en « sandwich » par 11 hélices- $\alpha$  (fig.II.7.).



*Figure.II.7. Structure du monomère de PAB0955*

*Les 5 motifs G sont indiqués de la même couleur que dans l'introduction, un motif spécifique de PAB0955 est signalé en violet nommé motif GPN. Le domaine G est en rose, les deux insertions I1 et I2 sont signalées en gris clair et foncé respectivement.*

Dans la séquence de la protéine, nous avons identifié un motif de Walker de type A connu pour lier les groupements phosphates des nucléotides. Ce motif adopte bien une conformation en boucle dans PAB0955 et correspond à  $^7\text{GTAGSGKT}^{14}$ . Le cœur de la protéine, constituant le domaine G est constitué de 6 brins- $\beta$  (exception du brin- $\beta_3$ ), et de 6 hélices- $\alpha$  ( $\alpha_1$ ,  $\alpha_5$ ,  $\alpha_6$ ,  $\alpha_7$ ,  $\alpha_8$  et  $\alpha_{11}$ ) (fig.II.8.).



La structure de PAB0955 présente un domaine G, contenant les 5 motifs G, avec deux insertions supplémentaires (I1 et I2).

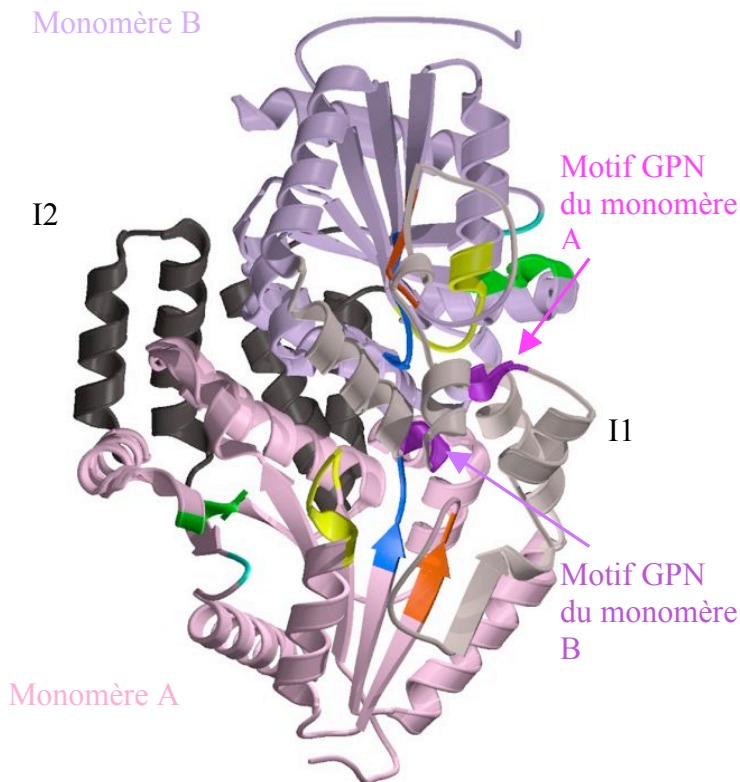
Figure.II.8. Topologie de PAB0955.

La première insertion I1 est constituée du brin- $\beta_3$ , des hélices  $\alpha_2$ ,  $\alpha_3$ , et  $\alpha_4$ . Elle comprend les acides aminés 41 à 80 situés entre les motifs G2 et G3 du site nucléotidique. La seconde insertion, I2, est constituée uniquement d'hélices ( $\alpha_9$  et  $\alpha_{10}$ ), c'est la partie construite manuellement dont la densité électronique est discontinue et l'agitation thermique élevée. Elle se situe entre les motifs G4 et G5 du site nucléotidique, elle contient les acides aminés 183 à 217. De part son agitation thermique élevée, nous avons tout d'abord supposé que cette seconde insertion était flexible d'où sa densité électronique partielle. En comparant les structures de PAB0955 en complexe avec des nucléotides obtenues dans d'autres groupes d'espace, notamment en  $P2_1$ , nous avons identifié deux conformations pour cette boucle (voir chapitre III).

Dans la structure de la protéine apo des molécules d'eau sont présentes dans la P-loop, mimant la position des groupements phosphate. Il n'y a pas de nucléotide dans cette structure.

### II.3.1.2. Description de l'interface dimérique

En solution, nous avons observé la protéine sous forme de dimère. Dans les cristaux, nous observons le même comportement. Le dimère est construit grâce à un axe 2, qui peut être cristallographique (l'unité asymétrique contient alors 1 monomère, groupe d'espace  $P3_221$  ou  $P2_12_12$ ) ou non cristallographique (dimère dans l'unité asymétrique, groupe d'espace  $P2_1$ )



*Figure.II.9. Structure du dimère biologique de PAB0955. Le domaine G d'un monomère est représenté en rose, l'autre en violet. Les insertions des deux monomères sont en gris.*

(fig.II.9.). Le site nucléotidique est présent à l'interface dimérique. Les sites nucléotidiques ne partagent pas de motif G avec l'autre monomère. Par contre, nous avons observé lors de l'analyse structurale un nouveau motif (GPN) impliquant des acides aminés d'un monomère dans le site nucléotidique de l'autre monomère.

L'interface dimérique est très grande, chaque monomère enfoui  $2210 \text{ \AA}^2$ . Elle est constituée à 64% d'acides aminés hydrophobes.

Plusieurs acides aminés aromatiques de chaque monomère entrelacent leur cycle au cœur du dimère. Ceci explique pourquoi nous observons toujours un dimère. Le monomère correctement replié ne doit pas exister seul en solution. Les deux insertions I1 et I2 représentent 39% de l'interface dimérique, jouant un rôle dans la stabilité du dimère.



## II.3.2. Les structures de PAB0955 avec des nucléotides tri-phosphate

### II.3.2.1. PAB0955 co-cristallisée avec du GTP $\gamma$ S

Les cristaux obtenus par co-cristallisation de PAB0955 avec du GTP $\gamma$ S appartiennent au groupe d'espace  $P3_221$ , avec un monomère par unité asymétrique.

Le site nucléotide est occupé à 100% par du GTP- $\gamma$ -S (fig.II.10.a.). Dans cette structure nous n'avons pas observé de magnésium et le nucléotide présente une conformation atypique. Les trois groupements phosphates sont normalement alignés. Dans notre structure, ils forment un coude (fig.II.10.b.).

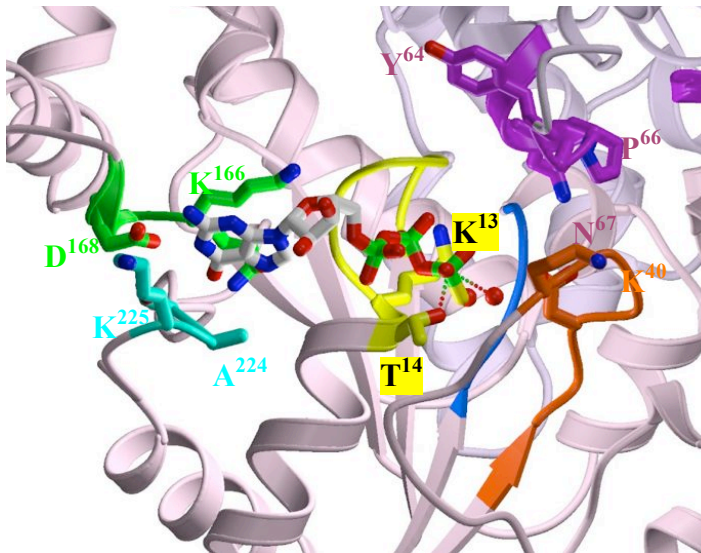


Figure.II.10.a. Structure du site nucléotidique de PAB0955 en complexe avec du GTP $\gamma$ S.

Le GTP $\gamma$ S est représenté en bâton, les molécules d'eau sont en rouge. Un monomère est représenté en rose et l'autre en violet.

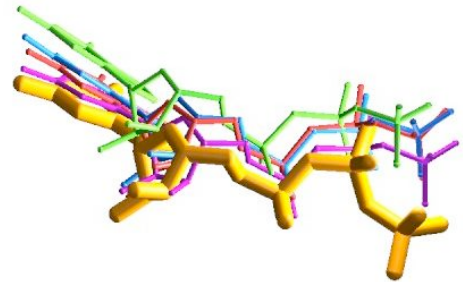


Figure.II.10.b. Superposition de différentes GTPases en complexe avec du GTP ou un analogue.

En ocre notre protéine, en bleu la SRP (1NRJ (Schwartz and Blobel, 2003), en rouge RhoA (1A2B (Ihara et al., 1998), en violet Rac (1E96 (Lapouge et al., 2000) et en vert Arf1 (1J2J (Shiba et al., 2003).

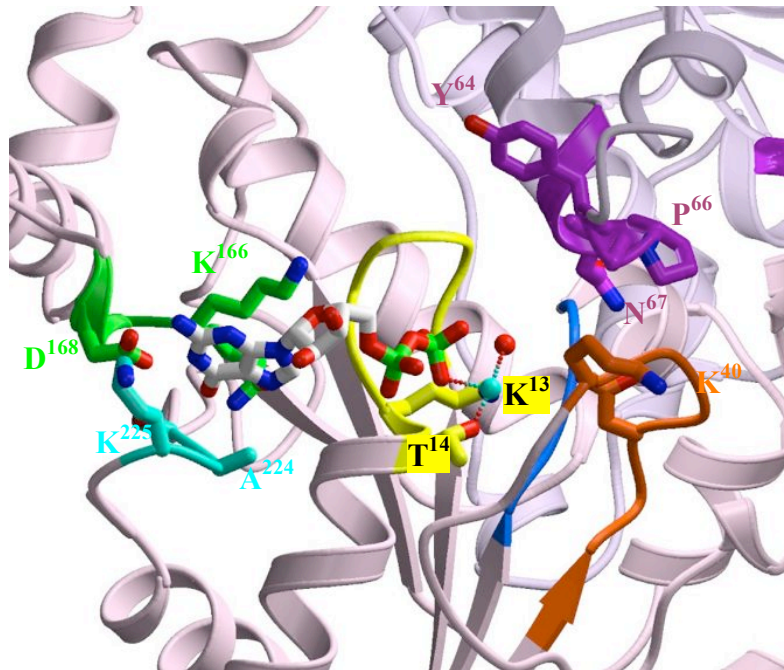
Les cristaux obtenus en présence de GMPPnp étaient de moins bonne qualité de diffraction. Cependant une structure a été obtenue à 2.4 Å. Le site nucléotidique contient du GDP, ce qui montre que l'hydrolyse a eu lieu pour cet analogue du GTP faiblement hydrolysable. La structure (non déposée) est identique à celle obtenue par hydrolyse du GTP.

En présence de GMPpcp, les cristaux sont multiples et s'indexent dans au moins deux groupes d'espace différents ( $P2_1$  et  $P3_221$ ). Les taches de diffraction sont étalées, le traitement n'a pas été possible. Nous n'avons donc pas pu obtenir de structure de PAB0955 en complexe avec le GMPpcp. La présence de deux groupes d'espace suggère plusieurs conformations du GMPpcp dans le site nucléotidique ou qu'il y a eu une hydrolyse partielle de GMPpcp.

### II.3.2.2. PAB0955 co-cristallisée avec du GTP à 18°C

Le but de cette expérience était de savoir si notre protéine était active dans nos conditions de cristallisation. La P-loop est occupée à 100% par du GDP dans cette structure (fig.II.11.). Il y a également un ion magnésium lié avec l'hydroxyle de la thréonine 14, avec un atome d'oxygène du phosphate- $\beta$  du GDP et une molécule d'eau.

La protéine est donc bien active dans nos conditions de cristallisation puisqu'il y a eu hydrolyse du GTP.



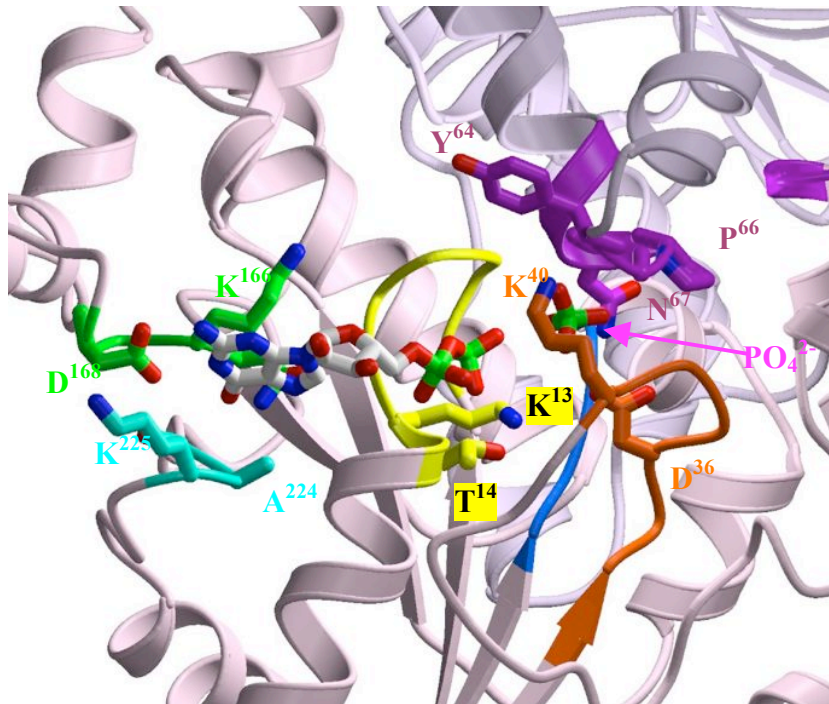
*Figure.II.11.* Structure du site nucléotidique de PAB0955 en complexe avec du GTP co-cristallisé à 18°C.

*Le magnésium est représenté en cyan, la molécule d'eau en rouge. Un monomère est représenté en rose et l'autre en violet. Le GDP est représenté en bâton.*

II.3.2.3. PAB0955 co-cristallisée avec du GTP à 4°C

Puisque PAB0955 est issue d'un archaéa dont la température optimale est de 82°C, nous espérons bloquer ou ralentir son hydrolyse à 4°C. Les cristaux obtenus avec du GTP à 4°C appartiennent au groupe d'espace  $P2_12_12$  (un monomère par unité asymétrique).

Dans cette structure  $GTP^{4^\circ C}$  la P-loop contient du GDP et un  $PO_4^{2-}$ , l'hydrolyse a donc bien eu lieu (fig.II.12.).



*Figure.II.12. Structure de PAB0955 co-cristallisée avec du GTP à 4°C ( $GTP^{4^\circ C}$ ).*

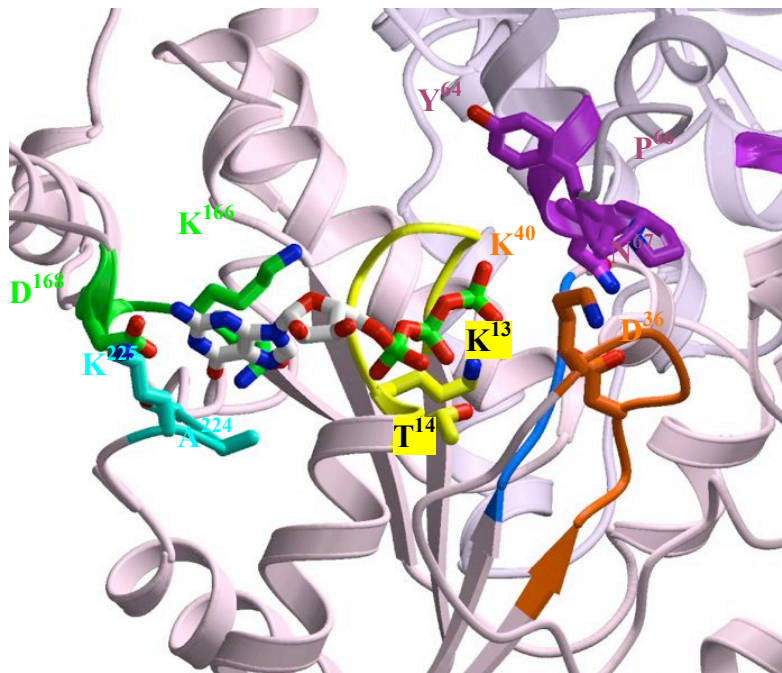
*La protéine est en complexe avec du GDP et du  $PO_4^{2-}$ , représentés en bâton. Un monomère est représenté en rose et l'autre en violet.*

Même si l'hydrolyse n'est pas bloquée, la réaction est affectée puisqu'il y a un ion phosphate dans le site nucléotidique. À 4°C la protéine semble ne pas pouvoir libérer son deuxième produit d'hydrolyse, le  $PO_4^{2-}$ . Cette structure nous révèle sans doute une étape intermédiaire de l'hydrolyse du GTP.

Dans cette structure, il existe une différence concernant la seconde insertion I2. La densité électronique d'I2, discontinue dans les cristaux en  $P3_221$ , est continue dans ces cristaux. Le changement de groupe d'espace,  $P2_12_12$ , a permis la formation de contacts cristallins stabilisant cette insertion. Ainsi, nous avons pu confirmer le repliement en hélices- $\alpha$  que nous avons difficilement construit dans la forme apo de la protéine.

### II.3.2.4. PAB0955 co-cristallisée avec du GTP et de l'EDTA

Le magnésium est essentiel pour la plupart des GTPases, aussi bien pour la fixation des nucléotides que pour l'hydrolyse du GTP (Sprang, 1997). L'EDTA (Acide éthylènediaminetétracétique) est un chélateur de cation bivalent et trivalent ; il chélate le magnésium avec une constante de formation de  $4.9.10^8 \text{ M}^{-1}$



*Figure.II.13.* Structure de PAB0955 co-cristallisée avec du GTP et de l'EDTA ( $\text{GTP}^{\text{EDTA}}$ ).

Un monomère est représenté en rose et l'autre en violet. Le GTP est représenté en bâton.

([www.chm.bris.ac.uk/motm/edta/edtah.htm](http://www.chm.bris.ac.uk/motm/edta/edtah.htm)).

Nous espérons ainsi pouvoir bloquer l'hydrolyse en ajoutant de l'EDTA.

Dans cette structure  $\text{GTP}^{\text{EDTA}}$  la P-loop contient du GTP avec un taux d'occupation de 70% pour le phosphate- $\gamma$  (fig.II.13.).

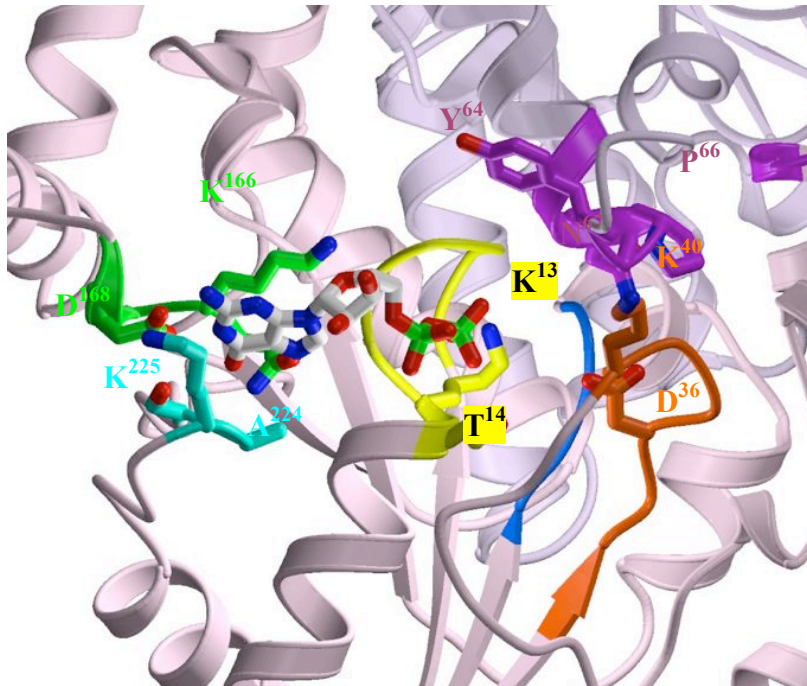
L'hydrolyse du GTP est donc partielle. Le magnésium est donc essentiel à l'hydrolyse du GTP dans PAB0955.

### II.3.3. Les structures de PAB0955 avec des nucléotides di-phosphate

#### II.3.3.1. PAB0955 co-cristallisée avec du GDP

Afin de savoir si notre protéine était capable de fixer le produit d'hydrolyse du GTP, des co-cristallisations ont été réalisées en présence de GDP. Les cristaux obtenus appartiennent au groupe d'espace  $P2_1$ , avec un dimère par unité asymétrique.





*Figure.II.14.* Structure de PAB0955 co-cristallisée avec du GDP.

*Un monomère est représenté en rose et l'autre en violet. Le GDP est représenté en bâton.*

structure.

La P-loop est occupée par du GDP avec un taux d'occupation de 100% dans les deux monomères. Le nucléotide présente la même conformation dans les deux monomères. La protéine est donc bien capable de fixer le produit de son hydrolyse. Nous n'avons pas observé de magnésium ou de phosphate.

La densité électronique de la seconde insertion est continue pour les deux monomères dans cette

### II.3.3.2. PAB0955 co-cristallisée avec du GDP et des additifs

Afin d'observer des états intermédiaires de l'hydrolyse du GTP, nous avons co-cristallisé la protéine en présence de GDP avec du fluorure d'aluminium (Wittinghofer, 1997), un sel de tungstate de sodium (Stavriddi et al., 2002) et un sel de phosphate de sodium. Les essais de co-cristallisation avec de l' $\text{AlF}_x$  et du  $\text{PO}_4\text{Na}$  en présence de GDP, ne nous ont pas permis d'observer ces molécules dans notre structure, seul le GDP est fixé dans la P-loop. Les différentes structures obtenues à partir de ces essais étaient à chaque fois identiques à celle obtenue avec le GDP seul.

Le fait que notre protéine ne fixe pas le fluorure d'aluminium peut avoir au moins deux explications : un problème expérimental ou l'absence de partenaire.

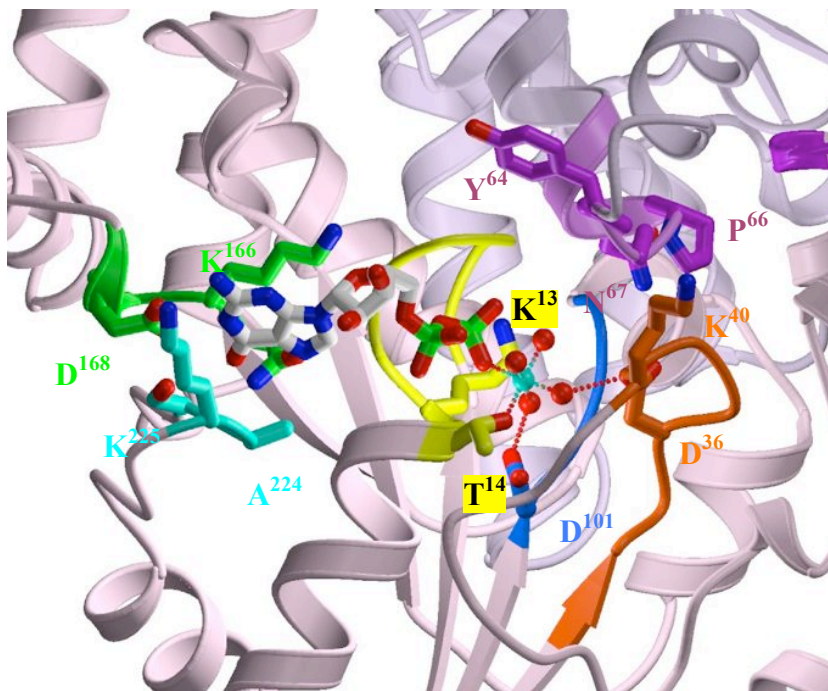
La formation du fluorure d'aluminium se fait en variant le rapport stœchiométrique entre le fluorure de sodium et le chlorure d'aluminium. Les différents rapports que nous avons utilisés n'étaient peut-être pas ceux permettant d'obtenir du fluorure d'aluminium. Les conditions de

cristallisation en présence de fluorure d'aluminium ( $AlF_X$ ) changent peut-être de celles que nous avons testées.

Pour ce qui est de l'absence de partenaire, cette observation fait suite à l'analyse des structures de petites protéines G en complexe avec l' $AlF_X$ . La fixation de ce composé se fait en présence de la protéine activatrice GAP (Scheffzek et al., 1998), avec un résidu arginine conservée dans les GAP qui vient stabiliser l' $AlF_X$ . La protéine PAB0955 est cristallisée seule, peut-être qu'un partenaire lui permettrait de fixer l' $AlF_X$ .

En présence de  $WO_4Na$  et de GDP, les cristaux appartiennent au groupe d'espace  $P2_1$ , avec un dimère par unité asymétrique. Le rmsd est de 0.47 Å avec la structure de la protéine apo.

Dans la structure, il n'y a pas de tungstate dans la P-loop, mais un ion magnésium (cyan) hexacoordonné par 4 molécules d'eau (rouge), le phosphate- $\beta$  du GDP et le groupement hydroxyle de la thréonine 14 (fig.II.15.).



*Figure.II.15. Structure de PAB0955 co-cristallisée avec du GDP et du  $WO_4Na$ .*

*Un monomère est représenté en rose et l'autre en violet. Le GDP est représenté en bâton, le magnésium est en cyan et les molécules d'eau sont en rouge.*

Le magnésium est également lié via les molécules d'eau, par l'aspartique 36 du motif G2 (orange) et par l'aspartique 101 du motif G3 (bleu). Lorsque nous avons co-cristallisé PAB0955 avec du GDP et du magnésium nous n'avons jamais observé de magnésium dans la P-loop. Seule la structure du GTP hydrolysé à 18°C contient du magnésium.

Conditions de cristallisation	Native	GTP $\gamma$ S	GTP <sup>EDTA</sup>	GTP4°C	GTP18°C	GDP	GDP WO <sub>4</sub> <sup>2-</sup>
Contenu de la P-loop dans les structures	Molécules d'eau	GTP $\gamma$ S	70% GTP 30% GDP	GDP PO <sub>4</sub> <sup>2-</sup>	GDP Mg 1 molécule d'eau	GDP	GDP Mg 4 molécules d'eau
Contenu de l'unité asymétrique	monomère	monomère	monomère	monomère	monomère	dimère	dimère
Rwork	0.232	0.208	0.236	0.225	0.211	0.186	0.195
Rfree	0.299	0.295	0.325	0.346	0.324	0.265	0.247
Rcryst	0.235	0.209	0.242	0.232	0.218	0.190	0.196
liaisons (Å)	0.010	0.009	0.009	0.008	0.009	0.011	0.008
Angles (°)	1.246	1.297	1.192	1.175	1.185	1.373	1.215
Facteur B (Å <sup>2</sup> )	53.6	45.4	48.9	46.8	50.1	34.9	33.8
Code PDB	1YR6	1YR7	1YR8	1YR9	Non déposé	1YRA	1YRB

*Figure.II.16. Tableau des statistiques d'affinement et récapitulatif du contenu de chaque structure de PAB0955.*

CHAPITRE III

ANALYSE

STRUCTURALE



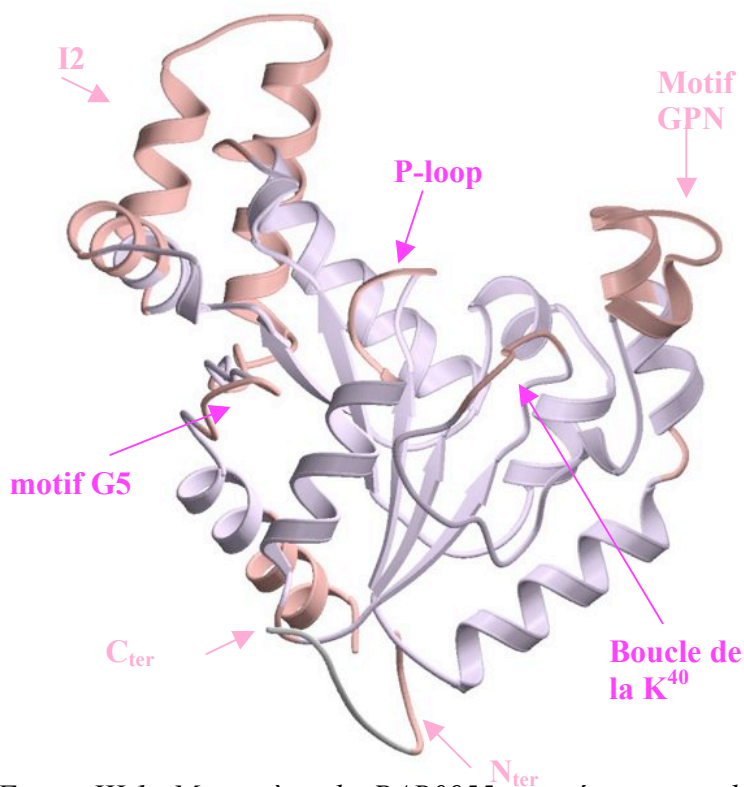


## III.1. Comparaison des structures de PAB0955

### III.1.1. Superposition des différentes structures de PAB0955

#### III.1.1.1. Méthode de superposition

Afin de comparer les différentes structures de PAB0955 entre elles, nous avons superposé toutes les structures avec la même méthode. Les superpositions ont été réalisées avec le programme ALIGN (Satow et al., 1986) sur les atomes C $\alpha$  uniquement. La superposition est réalisée sur un monomère de la protéine (violet) dont plusieurs régions ont été exclues (rose)



*Figure.III.1.* Monomère de PAB0955, représentation des résidus utilisés pour l'alignement (violet) en vue de l'analyse structurale.

*En rose sont représentés les résidus exclus de l'alignement structural.*

Par conséquent la superposition se fait sur 154 C $\alpha$  correspondant au cœur du repliement Rossmann et sur un seul monomère car, selon le groupe d'espace, le dimère est ou non-cristallographique.

(fig.III.1.). Nous avons exclu de l'alignement les régions les moins conservées au niveau structural, pour ne conserver que le cœur de la protéine. Les régions exclues de l'alignement sont donc: la P-loop, la boucle de la lysine40, le motif G5, la boucle GPN de l'insertion I1, la boucle charnière entre les hélices  $\alpha$ 3 et  $\alpha$ 4 et l'insertion I2. Les deux extrémités n'étant pas modélisées entièrement pour toutes nos structures, elles ont également été exclues.

### III.1.1.2. Résultat de la superposition

La comparaison des différentes structures que nous avons obtenues de PAB0955 a révélé que les changements structuraux étaient modestes, par rapport à ceux observés dans les petites protéines G. Pour exemple, entre la structure de hRas sans nucléotide (1BKD) (Boriack-Sjodin et al., 1998) et celle liant du GTP (1LFD) (Huang et al., 1998) le rmsd est de 0.82 Å sur 141 C $\alpha$ . Les motifs impliqués dans la liaison du nucléotide subissent des mouvements de grandes amplitudes (jusqu'à 15 Å pour le motif G3). Pour les mêmes structures de PAB0955 (apo et GTP<sup>EDTA</sup>), nous avons calculé un rmsd de 0.27 Å sur 154 C $\alpha$  (fig.III.2.). Parmi les régions impliquées dans la fixation du nucléotide, c'est le motif G5 qui subit la plus grande variation (un peu plus de 2 Å).

En revanche dans PAB0955, l'insertion I2 (résidus 183-217) subit des mouvements de grandes amplitudes entre les différentes structures, jusqu'à 10 Å pour certains acides aminés. Les mouvements les plus importants sont observés dans l'hélice- $\alpha$ 10 (résidus 197-217). Cette hélice est en contact avec l'hélice- $\alpha$ 10 de l'autre monomère.

	GTP $\gamma$ S	GTP <sup>EDTA</sup>	GTP <sup>4°C</sup>	GTP <sup>18°C</sup>	GDP	MgGDP
Apo protéine	0.38 Å	<b>0.27 Å</b>	0.49 Å	0.25 Å	0.44 Å	0.48 Å
GTP $\gamma$ S		0.34 Å	<b>0.58 Å</b>	0.30 Å	0.42 Å	0.42 Å
GTP <sup>EDTA</sup>			0.52 Å	0.25 Å	0.46 Å	0.48 Å
GTP <sup>4°C</sup>				0.51 Å	0.48 Å	0.55 Å
GTP <sup>18°C</sup>					0.42 Å	0.46 Å
GDP						<b>0.20 Å</b>

*Figure.III.2. Tableau des rmsd entre les différentes structures de PAB0955, superposition sur le coeur d'un monomère (154 résidus).*

Les structures les plus proches sont celles de PAB0955 en complexe avec le GDP et le MgGDP avec un rmsd de 0.2 Å (bleu fig.III.2.). Ces deux complexes fixent le même nucléotide (GDP) seule la présence du magnésium diffère entre ces deux structures. Sa présence ne change pas la structure de PAB0955, il est donc logique que ce soient les structures les plus proches.

Les complexes de PAB0955 les plus différents, avec un rmsd de 0.58 Å, sont ceux avec du GTP $\gamma$ S et avec MgGDP (fig.III.2.). Une des structure fixe un analogue du substrat de la

protéine et l'autre le produit de l'hydrolyse du substrat. Leurs structures sont les plus éloignées probablement parce qu'elles montrent le début et la fin de l'hydrolyse du GTP.

Si la conformation globale des complexes reste identique, il existe des parties de la protéine qui se déplacent en fonction de la nature du nucléotide fixé dans la P-loop. Ce sont les motifs du site nucléotidique qui subissent les plus grands changements, ainsi que l'insertion I2.

### III.1.2. Analyse des changements structuraux

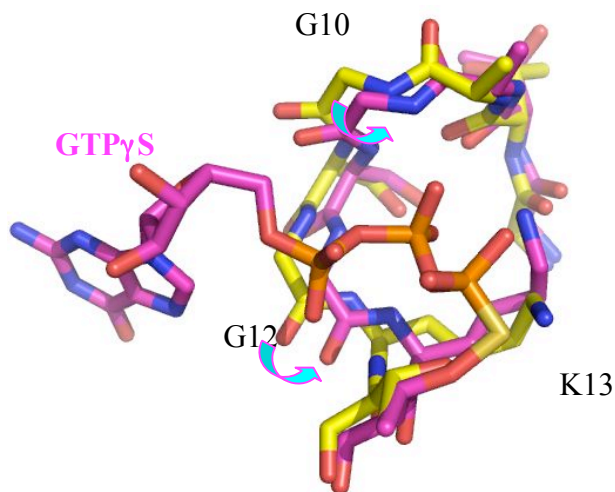
#### III.1.2.1. Le motif G1 : la P-loop

Dans la structure de la protéine apo, la P-loop (résidus 7 à 14) contient des molécules d'eau à la place des groupements phosphates. Cette boucle agit un peu comme une « mâchoire » qui se referme sur les groupements phosphates pour les stabiliser et s'ouvre pour permettre la libération du GDP après hydrolyse.

Les différentes structures ont été superposées sur la structure de la protéine apo, selon la technique expliquée dans le paragraphe III.1.1.1. Nous avons pris comme référence la protéine apo pour observer l'influence des nucléotides sur la structure de PAB0955.

Structures	GTP <sup>EDTA</sup>	GTP $\gamma$ S	GTP <sup>4°C</sup>	GTP <sup>18°C</sup>	GDP	MgGDP
Résidus						
<b>G7</b>						0.64 Å
T8		0.75 Å	0.6 Å		0.71 Å	0.81 Å
A9	0.86 Å		0.97 Å	1.52 Å	0.81 Å	0.93 Å
G10		0.76 Å		1.11 Å	0.94 Å	1.24 Å
S11						0.6 Å
<b>G12</b>		1.06 Å			1.01 Å	1.31 Å
<b>K13</b>		1.2 Å			1.26 Å	1.41 Å
<b>T14</b>		0.65 Å			0.73 Å	0.73 Å

*Figure.III.3. Tableau des déplacements supérieurs à 0.5 Å sur les C $\alpha$  des résidus de la P-loop calculés par rapport à la structure de la protéine apo.*

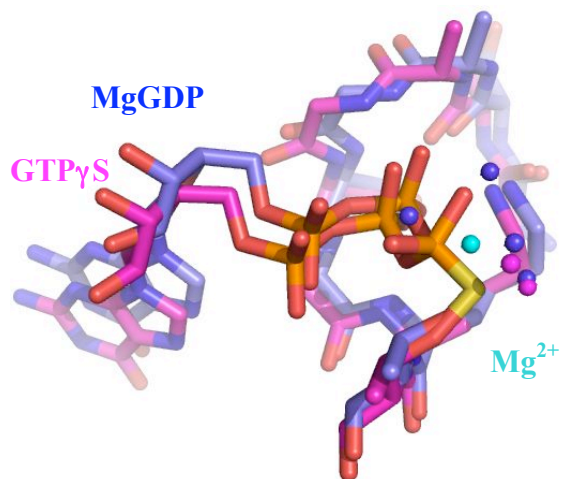


*Figure.III.4.a. Superposition des P-loop de PAB0955 apo (jaune) et en complexe avec du GTP $\gamma$ S (rose).*

*Les flèches en cyan indiquent le mouvement de fermeture de la P-loop sur le GTP $\gamma$ S.*

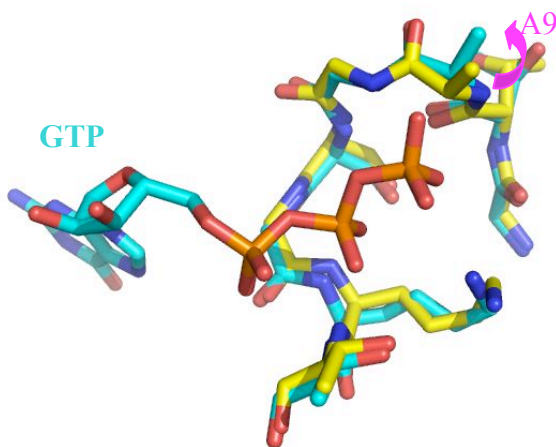
Dans la structure en complexe avec du GTP $\gamma$ S, il n'y a pas de magnésium dans le site nucléotidique. En superposant les structures GTP $\gamma$ S et MgGDP, nous nous sommes aperçus que le phosphate- $\gamma$  du GTP $\gamma$ S occupait la position du magnésium (fig.III.4.b.). Le magnésium et les atomes d'oxygène du phosphate- $\gamma$  établissent des liaisons avec les mêmes atomes de la protéine et par des molécules d'eau conservées. Par conséquent la position très particulière du GTP $\gamma$ S empêche le magnésium de se fixer.

En présence de nucléotides tri-phosphates nous avons observé deux comportements différents de la P-loop de PAB0955. Avec le GTP $\gamma$ S, les phosphates  $\alpha$  et  $\beta$  sont stabilisés par un effet de fermeture de cette boucle (fig.III.4.a.). Les acides aminés G10 (déplacement de 0.76 Å), G12 (déplacement de 1.06 Å) et K13 (déplacement de 1.2 Å) pointent le NH de leur chaîne principale vers les phosphates en formant des liaisons hydrogènes. Le phosphate- $\gamma$  quant à lui, en raison de son atome de soufre ne peut se plonger dans le site nucléotidique. Ce phosphate est un peu exclu du site et forme des liaisons avec la thréonine 14. Dans les structures d'autre GTPases en complexe avec du GTP $\gamma$ S, cette thréonine est liée avec le Mg $^{2+}$ .



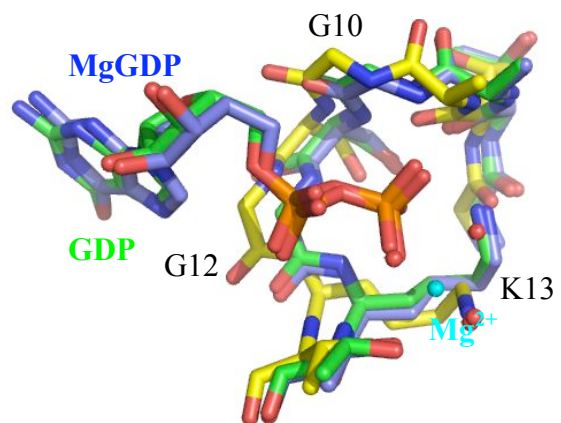
*Figure.III.4.b. Superposition des P-loop de PAB0955 en complexe avec du GTP $\gamma$ S (rose) et celle avec Mg et GDP (bleu).*

*Les molécules d'eau bleues sont celles de la structure MgGDP, en cyan le magnésium et en rose les molécules d'eau liant le GTP $\gamma$ S.*



*Figure.III.5.* Superposition des P-loop de PAB0955 apo (jaune) et en complexe avec du GTP<sup>EDTA</sup> (cyan).

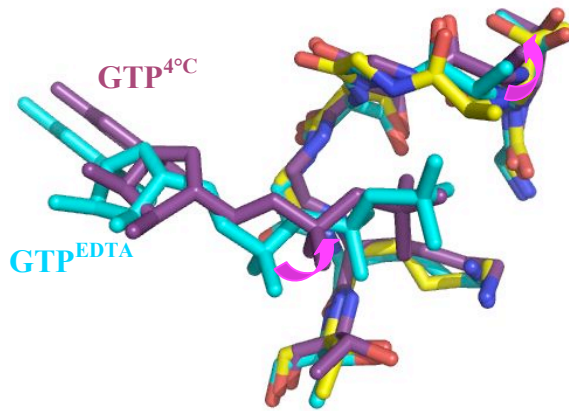
Les structures en complexe avec le GDP, avec et sans Mg<sup>2+</sup>, sont très similaires. La présence des groupements phosphates provoque un effet de fermeture de la P-loop. Comme pour la structure avec GTP $\gamma$ S, ce sont les acides aminés G10, G12 et K13 qui subissent les changements les plus importants, avec des déplacements de l'ordre de 1.2 Å.



*Figure.III.6.* Superposition des P-loop de PAB0955 apo (jaune), en complexe avec du GDP (vert) et avec GDP et magnésium (bleu).

Pour les structures obtenues après hydrolyse du GTP, selon la température, les mouvements de la P-loop ne vont pas dans le même sens, suggérant un moment différent dans le mécanisme d'hydrolyse du GTP.

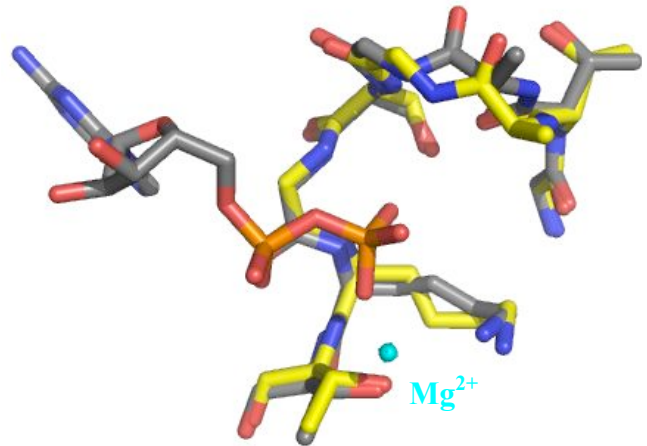
La structure de la P-loop du GTP<sup>4°C</sup> est très proche de celle avec GTP<sup>EDTA</sup>, ces deux structures correspondent peut-être à des étapes successives dans l'hydrolyse du GTP (fig.III.7.). La position des groupements phosphates diffère entre ces deux structures. Le phosphate- $\beta$  dans la structure GTP<sup>4°C</sup> est décalé, vers l'intérieur du site nucléotidique, de 1.1 Å par rapport à celui du GTP<sup>EDTA</sup>. Les groupements phosphates de la structure GTP<sup>4°C</sup> adoptent la même conformation que ceux de la structure avec GDP, c'est à dire plus enfouis



*Figure.III.7.* Superposition des P-loop de PAB0955 apo (jaune), GTP<sup>4°C</sup> (violet) et GTP<sup>EDTA</sup> (cyan). Les flèches roses indiquent les mouvements de la P-loop et du nucléotide.

La structure GTP<sup>18°C</sup> contient du GDP avec un ion magnésium. Elle présente un mouvement d'ouverture de la P-loop par rapport à celle de la protéine apo. La boucle s'ouvre, peut-être pour libérer le GDP produit de l'hydrolyse, car la structure GTP<sup>18°C</sup> peut correspondre à un moment de l'hydrolyse successif à celui de la structure GTP<sup>4°C</sup>. De plus les groupements phosphates de la structure GTP<sup>18°C</sup> se superposent avec ceux de la structure GTP<sup>EDTA</sup>, indiquant qu'il existe un mouvement de recule du nucléotide au cours de l'hydrolyse.

dans le site nucléotidique que ceux du GTP<sup>EDTA</sup>. Le nucléotide, plus flexible que la protéine, se relaxe plus rapidement. C'est pourquoi les P-loop sont similaires entre GTP<sup>EDTA</sup> et GTP<sup>4°C</sup>, alors que le nucléotide s'est déjà déplacé.

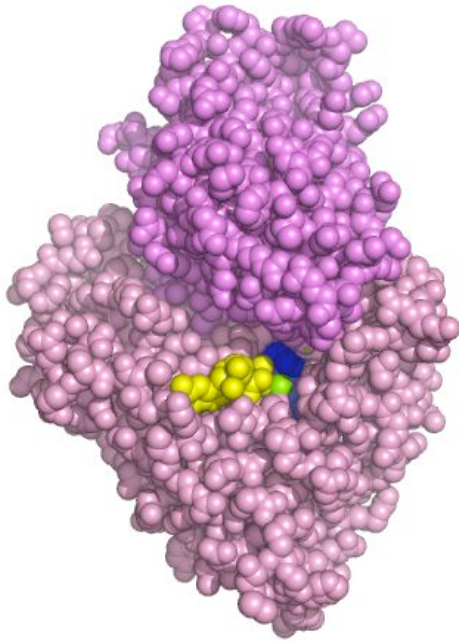


*Figure.III.8.* Superposition des P-loop de PAB0955 apo (jaune) avec celle de GTP<sup>18°C</sup> (gris).

### III.1.2.2. Le motif G3 : <sup>101</sup>DTPGQ<sup>105</sup>

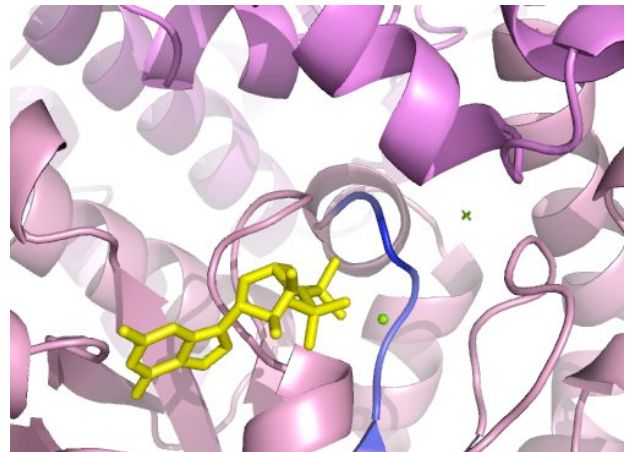
Le motif G3, <sup>101</sup>DTPGQ<sup>105</sup> dans PAB0955, correspond à la région switch II des interrupteurs moléculaires. Dans ces derniers, ce motif subit des réarrangements structuraux importants, avec des déplacements de l'ordre de 15 Å (voir III.1.1.2.). Dans PAB055, ce motif ne subit que de petits déplacements (1.2 Å au maximum). Le motif G3 dans notre protéine est situé à l'interface dimérique (fig.III.8. a et b). La localisation de ce motif à l'interface sans doute pourquoi celui-ci ne subit pas de mouvement de grande amplitude.





*Figure.III.8.a.* Structure de PAB0955 en complexe avec GDP et magnésium.

Le GDP est représenté en jaune, le magnésium en vert, le motif G3 en bleu, un monomère en rose et l'autre en violet. Les atomes sont représentés par leur rayon de van der Waals.



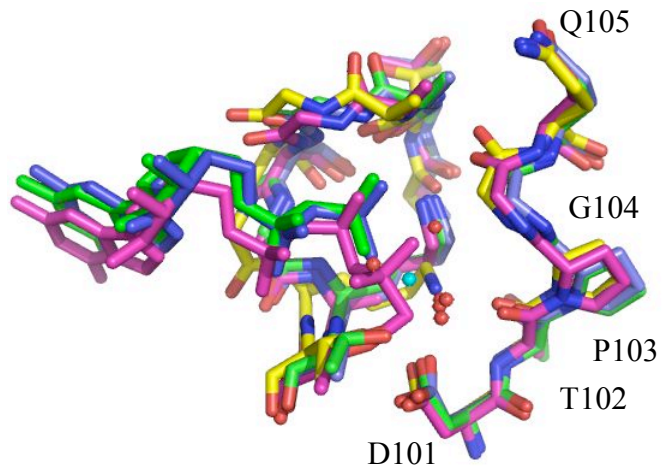
*Figure.III.8.b.* Structure de PAB0955 en complexe avec GDP et magnésium.

Les différents éléments sont représentés de la même couleur que sur la figure III.8.a. Les atomes sont modélisés par leur structure secondaire.

Les résidus du motif G3 subissent des déplacements supérieurs à 0.5 Å dans 3 structures : GTP $\gamma$ S, GDP et MgGDP (fig.III.9. a et b).

Structures	GTP $\gamma$ S	GDP	MgGDP
Résidus			
D101			
T102			
P103		0.83 Å	0.68 Å
G104	0.96 Å	1.21 Å	1.06 Å
Q105	0.67 Å	0.69 Å	

*Figure.III.9.a.* Tableau des déplacements supérieurs à 0.5 Å sur les C $\alpha$  des résidus du motif G3 calculés par rapport à la structure de la protéine apo.



*Figure.III.9.b.* Superposition des P-loop et des motifs G3 de la protéine apo (jaune), GTP $\gamma$ S (rose), GDP (vert) et MgGDP (bleu).



Dans ces trois structures, le mouvement de G3 va dans le même sens ; il s'éloigne du nucléotide. Ce mouvement du motif G3 permet au  $Mg^{2+}$  avec ses molécules d'eau et au phosphate- $\gamma$ S d'avoir suffisamment d'espace dans le site nucléotidique. Le  $Mg^{2+}$  avec ces 4 molécules d'eau dans la structure MgGDP et le phosphate- $\gamma$ S du complexe avec le GTP $\gamma$ S, occupent le même espace (fig.III.10. a et b).

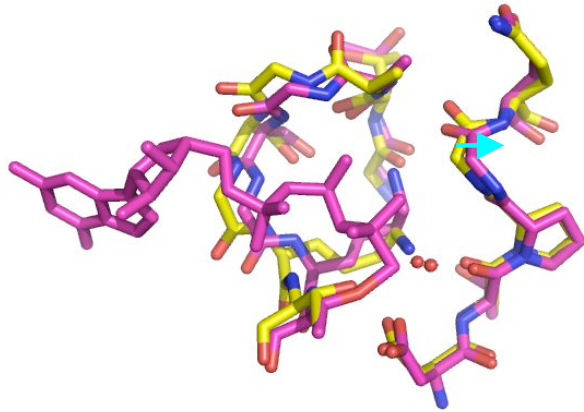


Figure.III.10.a. Superposition des P-loop et des motifs G3 de PAB0955 apo (jaune) et avec GTP $\gamma$ S (rose).

La flèche cyan indique le mouvement du motif G3.

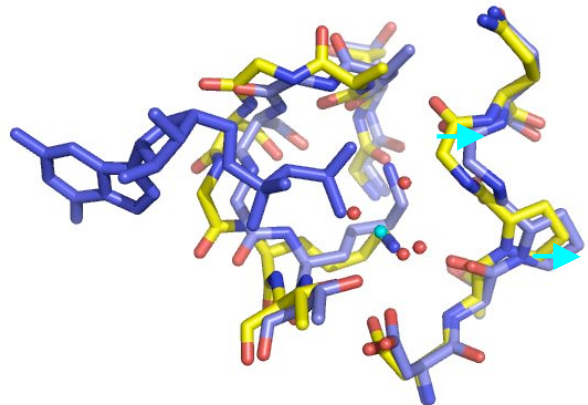


Figure.III.10.b. Superposition des P-loop et des motifs G3 de PAB0955 apo (jaune) et avec MgGDP (bleu).

Les flèches cyan indiquent le mouvement du motif G3.

### III.1.2.3. Le motif G5 : <sup>223</sup>SAK<sup>225</sup>

Le motif G5 qui fixe la base guanine du nucléotide se déplace très peu dans les structures GTP<sup>EDTA</sup> et GTP<sup>18°C</sup>. Dans ces deux structures, le motif G5 est légèrement plus proche du nucléotide que dans la forme apo. Ce mouvement est un peu plus accentué dans les structures GTP $\gamma$ S et GTP<sup>4°C</sup> de PAB0955 avec des déplacements sur les C $\alpha$  supérieurs à 0.5 Å. Les structures obtenues en complexe avec du GDP, possèdent des déplacements plus importants pour les résidus du motif G5, ces déplacements suivent ceux du nucléotide (fig.III.11).

Structures	GTP $\gamma$ S	GTP <sup>4°C</sup>	GDP	MgGDP
Résidus				
S223				0.77 Å
A224	0.61 Å	0.82 Å	1.6 Å	1.81 Å
K225	0.63 Å	1.25 Å	1.89 Å	2.14 Å

Figure.III.11. Tableau des déplacements supérieurs à 0.5 Å sur les C $\alpha$  des résidus du motif G5 calculés par rapport à la structure de la protéine apo.

Ce motif G5 adopte quasiment la même conformation dans les structures en complexe avec un nucléotide. Son mouvement est concerté avec celui du motif G4 fixant également la base guanine.

### III.1.2.4. La lysine 40

La lysine 40 appartient à une boucle qui fait suite au motif G2 fixant le magnésium (D36). Dans la plupart des structures, la densité électronique du groupement  $\text{NH}_3^+$  de la chaîne

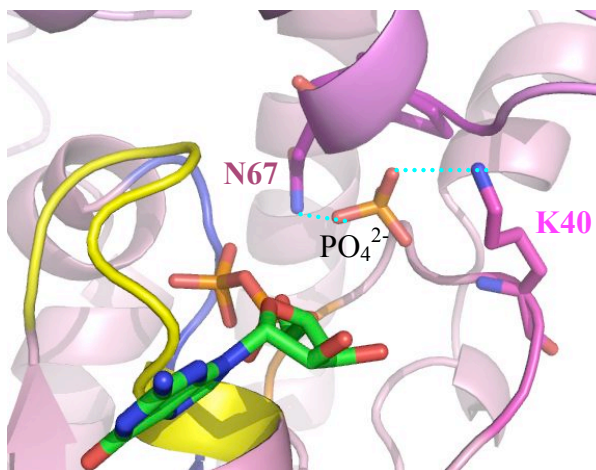


Figure.III.12.a. Site de fixation du  $\text{PO}_4^{2-}$  dans la structure  $\text{GTP}^{4^\circ\text{C}}$ .

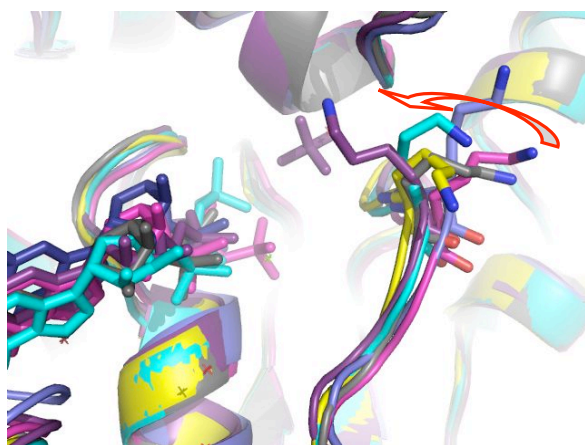


Figure.III.12.b. Superposition des lysines 40 de la protéine apo (jaune), le  $\text{GTP}^{\text{EDTA}}$  (cyan),  $\text{GTP}\gamma\text{S}$  (rose),  $\text{GTP}^{4^\circ\text{C}}$  (violet),  $\text{GTP}^{18^\circ\text{C}}$  (gris) et  $\text{MgGDP}$  (bleu).

latérale de cette lysine n'est pas visible. Dans la structure  $\text{GTP}^{4^\circ\text{C}}$ , la densité électronique de la chaîne latérale de la lysine 40 est continue car elle est en interaction avec le  $\text{PO}_4^{2-}$  (fig.III.12.a.). Sa chaîne latérale est alors stabilisée, la boucle qui la contient se rapproche du site nucléotidique. La lysine 40 peut jouer un rôle de stabilisation des charges négatives produites au cours de l'hydrolyse du GTP, facilitant ainsi la réaction. La flexibilité de cette boucle permet un basculement de la chaîne latérale de la lysine qui pourrait ainsi positionner son NZ à proximité du phosphate- $\gamma$  au cours de l'hydrolyse (fig.III.12.b.). Dans les structures obtenues par co-cristallisation en présence de GDP et de  $\text{GTP}\gamma\text{S}$ , cette boucle est éloignée du nucléotide (fig.III.12.c.). Nous pouvons donc supposer que ces structures ne sont pas en conformation active puisqu'elles ne favorisent pas la réaction d'hydrolyse.

Structures	$\text{GTP}^{\text{EDTA}}$	$\text{GTP}\gamma\text{S}$	$\text{GTP}^{4^\circ\text{C}}$	$\text{GTP}^{18^\circ\text{C}}$	GDP	MgGDP
Résidus						
G38		0.94 Å	0.6 Å		0.67 Å	0.89 Å
V39		1.46 Å	1.2 Å	0.65 Å	0.97 Å	1.45 Å
<b>K40</b>	0.65 Å	1.51 Å	1.31 Å		1.14 Å	1.98 Å
E41	0.61 Å	1.16 Å	1.05 Å		1.11 Å	1.54 Å

Figure.III.12.c. Tableau des déplacements supérieurs à 0.5 Å sur les  $\text{C}\alpha$  des résidus de la boucle contenant la lysine 40 calculés par rapport à la structure de la protéine apo.

### III.1.2.5. Le motif <sup>65</sup>GPN<sup>67</sup>

Un autre motif qui se déplace lorsqu'un nucléotide est fixé est le motif <sup>65</sup>GPN<sup>67</sup>, qui appartient à la première insertion I1. Ce motif se situe au début de l'hélice- $\alpha_4$ . Le motif GPN d'un monomère vient se placer dans le site actif de l'autre monomère. L'asparagine 67 est placée de telle sorte, qu'elle peut se tourner vers le phosphate- $\gamma$  et peut donc stabiliser le  $\text{PO}_4^{2-}$  produit au cours de l'hydrolyse. Dans la description qui suit (pour plus de clarté) nous décrirons le site nucléotidique du monomère A et le motif GPN du monomère B. Ces observations sont également valables pour le site nucléotidique du monomère B et le motif GPN du monomère A.

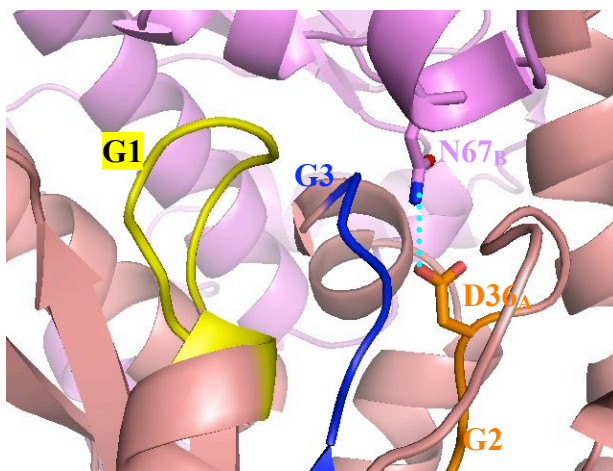


Figure.III.13.a. Structure de la protéine apo.

Le monomère A est en rose et le B en violet.

Dans la forme apo, le D36<sub>A</sub> appartenant au motif G2 est en interaction par sa chaîne latérale avec le N67<sub>B</sub>, ce qui montre la proximité de ce motif GPN avec le site nucléotidique (fig.III.13.a.).

Par rapport à cette position nous avons observé des changements dans deux sens opposés, tout comme pour la lysine 40 (fig.III.13.b.).

Structures	GTP $\gamma$ S	GTP <sup>EDTA</sup>	GTP <sup>18°C</sup>	GDP	MgGDP	GTP <sup>4°C</sup>
Résidus						
G63	0.78 Å			1.33 Å	1.36 Å	0.71 Å
Y64	0.66 Å			1.18 Å	1.55 Å	
G65	0.65 Å	0.85 Å	0.6 Å	1.3 Å	1.72 Å	1.25 Å
P66	0.78 Å	0.58 Å	0.65 Å	1.62 Å	1.78 Å	0.96 Å
N67	1.06 Å			1.99 Å	1.98 Å	0.89 Å
G68	0.83 Å			1.41 Å	1.46 Å	0.65 Å

Figure.III.13.b. Tableau des déplacements supérieurs à 0.5 Å sur les C $\alpha$  des résidus de la boucle contenant le motif GPN calculés par rapport à la structure de la protéine apo.

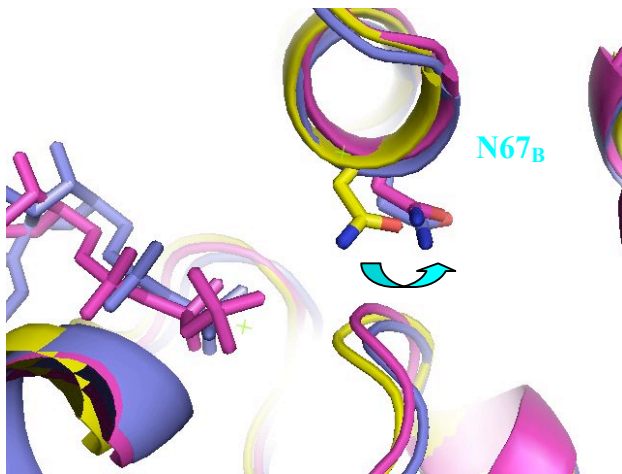


Figure.III.13.c. Superposition des structures apo (jaune), MgGDP (bleu) et GTP $\gamma$ S (rose).

Dans les structures avec le GTP $\gamma$ S, le GDP et le MgGDP, le motif GPN recule par rapport au nucléotide (fig.III.13.c.). Le N67<sub>B</sub> se trouve éloigné des groupements phosphates ; et donc l'hydrolyse n'est pas favorisée. Ces trois structures représentent probablement des conformations inactives de la protéine.

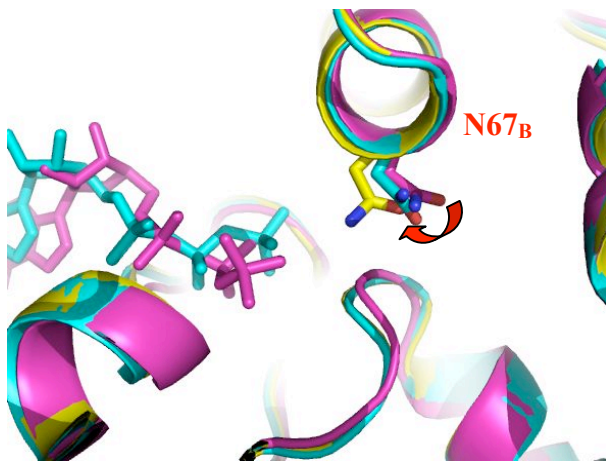


Figure.III.13.d. Superposition des structures apo (jaune), GTP<sup>EDTA</sup> (cyan) et GTP $\gamma$ S (rose).

La boucle GPN dans la structure GTP<sup>EDTA</sup> est plus proche du nucléotide que celle observée avec le GTP $\gamma$ S (fig.III.13.d.). Dans la structure GTP<sup>EDTA</sup>, le phosphate- $\gamma$  a un taux d'occupation de 70% indiquant qu'il y a eu hydrolyse partielle. Par conséquent la conformation du motif GPN est active pour l'hydrolyse du GTP.

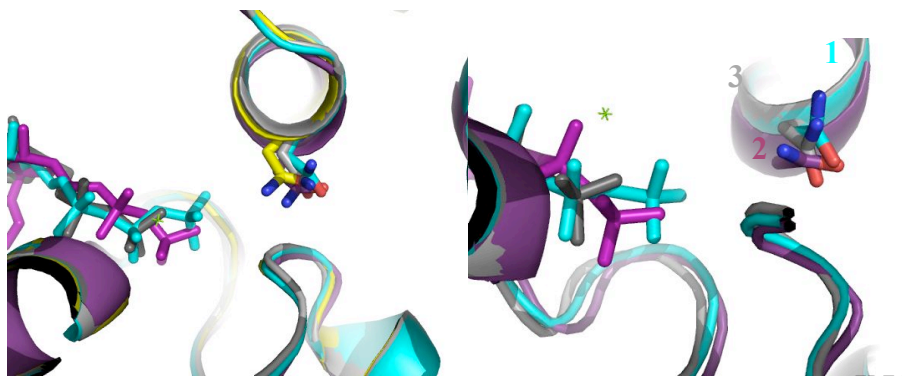


Figure.III.13.e. Superposition des structures apo (jaune), GTP<sup>EDTA</sup> (cyan), GTP<sup>18°C</sup> (gris) et GTP<sup>4°C</sup> (violet).

En comparant les conformations de la boucle GPN des structures GTP<sup>EDTA</sup>, GTP<sup>18°C</sup> et GTP<sup>4°C</sup> nous pouvons les classer par ordre

chronologique de l'hydrolyse du GTP (fig.III.13.e.). L'hydrolyse débute avec la position du GPN de la structure GTP<sup>EDTA</sup> (1). Puis, le GTP hydrolysé en GDP et PO<sub>4</sub><sup>2-</sup>, les charges négatives sont compensées par le N67 du monomère B dans la structure GTP<sup>4°C</sup> (2). Dans cette structure, le PO<sub>4</sub><sup>2-</sup> est lié avec l'asparagine 67 tournée vers le site nucléotidique. Le



$\text{PO}_4^{2-}$  est ensuite libéré et l'asparagine 67 se relaxe comme nous l'avons observé dans la structure  $\text{GTP}^{18^\circ\text{C}}$  (3).

Ce motif GPN, comme la lysine 40, peut stabiliser les charges négatives produites au cours de l'hydrolyse. Ce motif semble donc important pour l'hydrolyse du GTP. Son rôle permet d'expliquer la raison d'être de la forme dimérique de la protéine. En effet, c'est le motif GPN d'un monomère qui stabilise les charges négatives dans le site nucléotidique de l'autre monomère.

### III.1.3. Les différentes conformations de l'insertion I2

La seconde insertion I2, formée par les hélices  $\alpha 9$  et  $\alpha 10$ , est éloignée du site nucléotidique. Les mouvements de cette région sont les plus importants, surtout dans l'hélice  $\alpha 10$  entre les différentes structures de PAB0955.

Les premiers cristaux obtenus appartiennent au groupe d'espace  $P3_221$ . Les structures issues de ces cristaux possèdent une densité électronique discontinue pour l'insertion I2. Étant donné que dans ce groupe d'espace, I2 ne forme pas de contact avec des molécules symétriques, nous avons tout d'abord supposé que cette région était mobile. L'agitation thermique de celle-ci est deux fois plus élevée que celle de la protéine (I2 a un B moyen de  $100 \text{ \AA}^2$  et la protéine a un B moyen de  $50 \text{ \AA}^2$ ).

Nous avons obtenu ensuite des cristaux appartenant au groupe d'espace  $P2_12_12$ , dans lequel la densité électronique d'I2 est continue. Dans ce groupe d'espace, I2 est stabilisée par des contacts avec des molécules symétriques. Ces observations confortaient l'hypothèse d'une mobilité de cette région, qui en l'absence de contact cristallin possède une densité électronique discontinue.

Cependant, dans le groupe d'espace  $P2_1$  la densité électronique d'I2 est également continue alors qu'il n'y a pas de contact cristallin significatif. La différence est le contenu de l'unité asymétrique : un monomère pour les cristaux appartenant au groupe d'espace  $P2_12_12$  et un dimère pour ceux en  $P2_1$ . Le facteur B d'agitation thermique moyen d'I2 est de  $40 \text{ \AA}^2$  celui de la structure est de  $34 \text{ \AA}^2$  dans les cristaux appartenant au groupe d'espace  $P2_1$ . Cette région n'est donc pas agitée il n'y a pourtant pas de contact cristallin.

Lorsque nous avons superposé le monomère A (en rose) avec le monomère B (en vert), nous avons pu observer que toute la structure était superposée à l'exception de cette région. La plus

grande variation entre I2<sub>A</sub> et I2<sub>B</sub> est observée pour la proline 196 dont le déplacement est de 3.86 Å (fig.III.14. a et b).

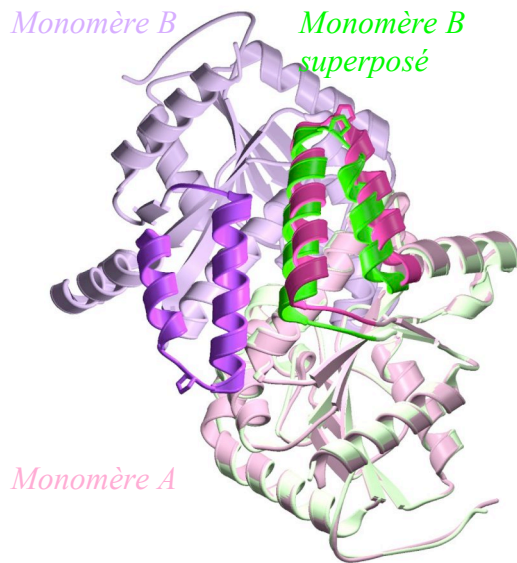


Figure.III.14.a. Superposition des monomères A (rose) et B (vert) des cristaux en  $P2_1$ .

Un monomère est en rose, l'autre en violet pour la structure du dimère.

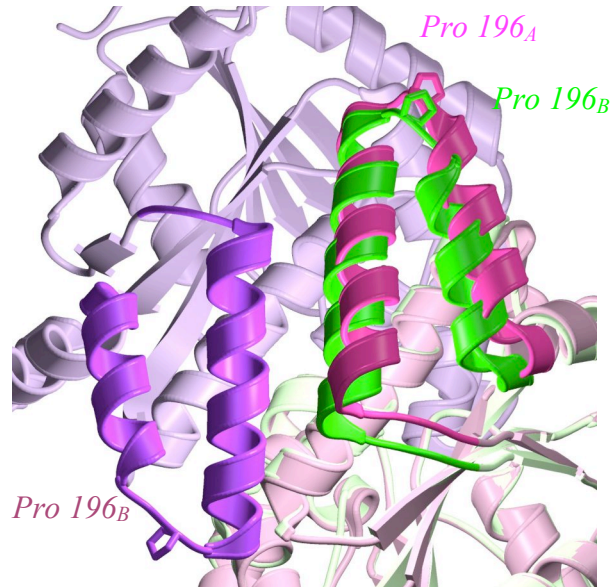


Figure.III.14.b. Superposition des monomères A (rose) et B (vert) des cristaux en  $P2_1$ .

En bâton est représentée la proline 196.

Les figures III.14. a et b montrent que les deux monomères sont différents quant au positionnement d'I2 par rapport au cœur de la protéine. Ceci présuppose qu'il existe au moins deux conformations différentes de cette insertion I2.

Dans les cristaux appartenant au groupe d'espace  $P3_221$ , il y a un monomère par unité asymétrique. La densité électronique observée est peut-être la moyenne de celle des deux conformations de I2 observées dans le groupe d'espace  $P2_1$ . La moyenne ne correspondant à aucune des deux structures, la densité électronique est difficilement interprétable.

Les cristaux contenant le dimère dans l'unité asymétrique sont ceux obtenus par co-cristallisation avec du GDP uniquement. La nature du nucléotide ne change pas la conformation de PAB0955 comme c'est le cas dans les petites protéines G. Cependant, l'insertion I2 et le groupe d'espace auquel appartiennent les cristaux semblent être affecté par le nucléotide co-cristallisé avec la protéine. Il est difficile de savoir si la structure de l'insertion I2 est affectée par l'empilement cristallin ou par le nucléotide. Si I2 est capable de changer de conformation en fonction du nucléotide, elle pourrait alors interagir avec un partenaire de façon nucléotide dépendante.

### III.1.4. La rigidité du site nucléotidique

Le site nucléotidique est situé à l'interface dimérique. Il est formé par les 5 motifs D d'un monomère et par le motif GPN apporté par l'autre monomère.

Nous avons pu observer avec la description des différents déplacements, que la protéine PAB0955 ne subit pas de changements structuraux importants avec la fixation du nucléotide. De plus nous avons obtenu une structure avec un GTP $\gamma$ S dans une conformation atypique, une structure avec GTP<sup>EDTA</sup> où les groupements phosphates reculent au lieu de s'ancrer dans le site nucléotidique. Tout ceci nous a amené à mieux analyser le site nucléotidique dans la zone de fixation des groupements phosphates.

#### III.1.4.1. La conformation arquée du GTP $\gamma$ S

Nous avons vu dans le paragraphe III.1.2.2. sur les mouvements du motif G3, que ceux-ci étaient restreints par l'interface dimérique. Dans les petites protéines G, lors de la fixation du GTP $\gamma$ S, il y a un mouvement de recul du motif G3 et également du G2 par rapport au site nucléotidique. Ces deux motifs, dans les interrupteurs moléculaires, sont capable de changer de conformation pour s'adapter à l'atome de soufre plus encombrant qu'un atome d'oxygène. Dans PAB0955, le motif G3 ne peut pas se reculer suffisamment pour laisser l'espace

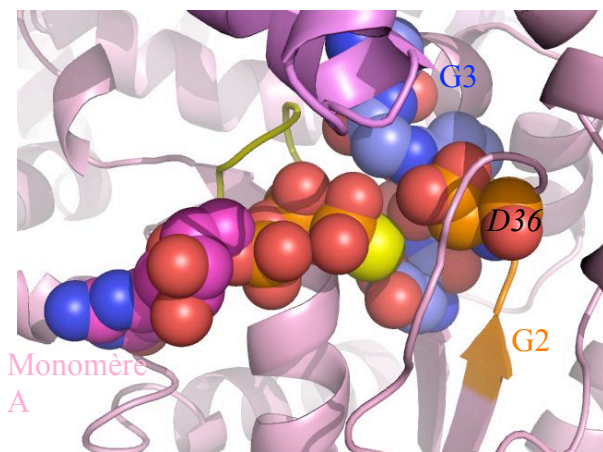


Figure.III.15.a. Structure du GTP $\gamma$ S de PAB0955.

*Le motif G3 est en bleu, le G2 en orange, l'atome de soufre du GTP $\gamma$ S en jaune.*

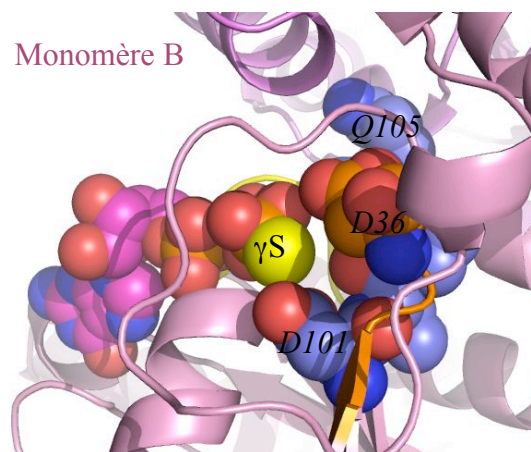


Figure.III.15.b. Structure du GTP $\gamma$ S de PAB0955.

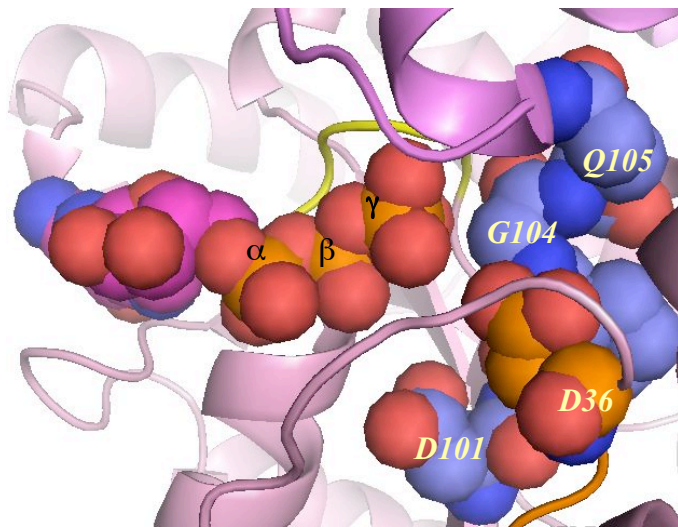
*L'atome de soufre ne peut pas s'enfouir dans le site, si le G3 ne se recule pas plus.*

nécessaire au positionnement de l'atome de soufre dans le site nucléotidique. Par conséquent c'est le GTP $\gamma$ S qui s'adapte au site nucléotidique et non l'inverse (fig.III.15. a et b).

C'est pourquoi nous obtenons cette conformation particulière du GTP $\gamma$ S dans notre protéine.

#### III.1.4.2. La position particulière du GTP par rapport au GDP

Dans la structure de PAB0955 avec du GTP<sup>EDTA</sup>, le phosphate- $\gamma$  remplit pleinement la poche nucléotidique. Comme nous l'avons décrit dans le paragraphe III.1.2.1. sur les mouvements de la P-loop, les groupements phosphates du GTP<sup>EDTA</sup> reculent de 1.1 Å par rapport à ceux du GDP (GTP<sup>4°C</sup>, GDP et MgGDP). Ayant chélaté le magnésium avec l'EDTA, nous avons supposé que ce mouvement de recul était dû à l'absence du magnésium (Vetter and Wittinghofer, 2001).



*Figure.III.16. Structure de PAB0955 GTP<sup>EDTA</sup>.*

*Le motif G3 est en bleu, le G2 en orange et la P-loop en jaune.*

La représentation des atomes par leur rayon de van der Waals montre que le phosphate- $\gamma$  est au contact du C $\alpha$  de la glycine 104 du motif G3 (fig.III.16.). Le GTP<sup>EDTA</sup> ne recule donc pas, mais il ne peut pas s'ancrer davantage dans le site nucléotidique. Ceci laisse supposer que la conformation du GTP dans la structure GTP<sup>EDTA</sup> est catalytique. La position plus enfouie des phosphates des GDP co-cristallisés peut avoir deux explications. La première serait que cette position « trop » enfouie bloque la protéine dans un état inactif, car le GDP est plus stable que le GTP. La seconde suppose qu'au cours de l'hydrolyse, le GDP rentre plus profondément dans le site nucléotidique, peut être pour aider le PO<sub>4</sub><sup>2-</sup> à sortir du site, puis le GDP se reculerait pour sortir également du site.

La représentation des atomes par leur rayon de van der Waals montre que le phosphate- $\gamma$  est au contact du C $\alpha$  de la glycine 104 du motif G3 (fig.III.16.). Le GTP<sup>EDTA</sup> ne recule donc pas, mais il ne peut pas s'ancrer davantage dans le site nucléotidique. Ceci laisse supposer que la conformation du GTP dans la structure GTP<sup>EDTA</sup> est catalytique. La position plus enfouie des phosphates des GDP co-cristallisés peut avoir deux explications. La première serait que cette position « trop » enfouie



### III.1.5. Le film de l'hydrolyse du GTP

Les différentes structures de PAB0955 obtenues permettent de proposer un modèle des différentes étapes du mécanisme d'hydrolyse du GTP en GDP et  $\text{PO}_4^{2-}$ . En comparant les structures en complexe avec les nucléotides, nous pouvons proposer un classement de celles-ci par ordre chronologique de l'hydrolyse du GTP.

Nous n'avons pu déterminer avec certitude le facteur initiant la réaction d'hydrolyse du GTP. Il existe différents modes d'hydrolyse (associatif ou dissociatif), et plusieurs facteurs initiateurs possibles (Maegley et al., 1996). Parmi les différentes hypothèses proposées pour l'initiation de l'hydrolyse du GTP, il y a l'attaque du GTP par une molécule d'eau nucléophile, le comportement du GTP comme une base captant des hydrogènes ou la P-loop qui donnerait un hydrogène au GTP. Nous n'avons pas observé de molécule proche ( $3.5\text{\AA}$  maximum) du nucléotide et conservée pour expliquer l'hydrolyse. Par contre dans la structure

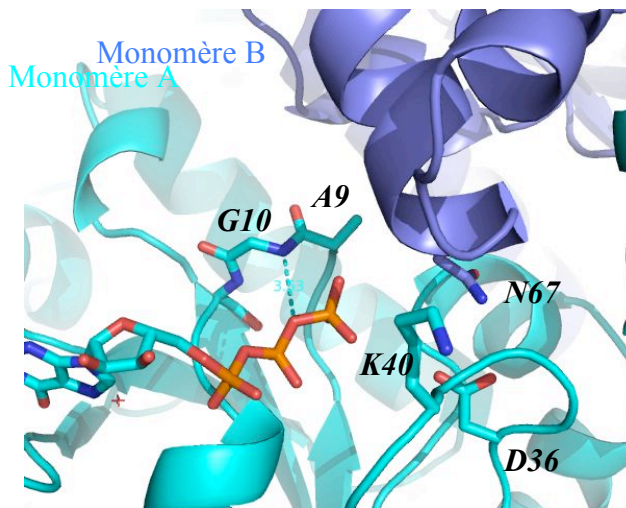


Figure.III.17.a. Structure de PAB0955  $\text{GTP}^{\text{EDTA}}$ .

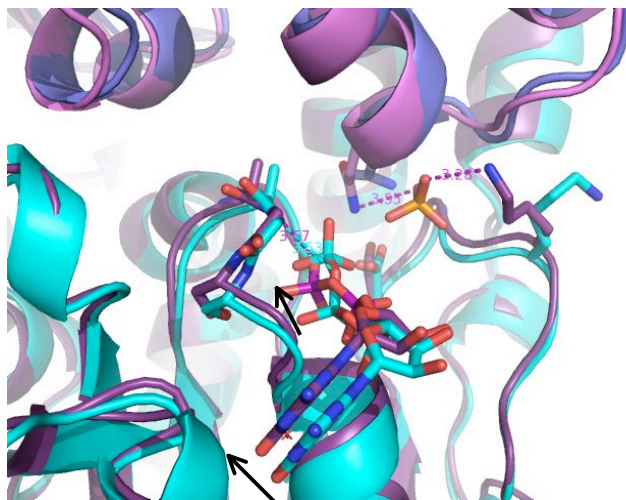


Figure.III.17.b. Superposition des structures de PAB0955  $\text{GTP}^{\text{EDTA}}$  (cyan) et  $\text{GTP}^{4^\circ\text{C}}$  (violet).

$\text{GTP}^{\text{EDTA}}$  de PAB0955 la glycine 10 (équivalente de G13 de hRas) pointe son NH vers la liaison entre les phosphates  $\beta$  et  $\gamma$ . La distance correspond à ce qui a été observé dans hRas (Maegley et al., 1996). Nous pourrions avoir aussi une liaison hydrogène catalytique entre la P-loop et le GTP (fig.III.17.a). La structure  $\text{GTP}^{\text{EDTA}}$  montre ainsi une image du début de l'hydrolyse. Puis, le GTP s'hydrolyse totalement en GDP et  $\text{PO}_4^{2-}$ , c'est la structure  $\text{GTP}^{4^\circ\text{C}}$  (fig.III.17.b.). Les phosphates plongent dans le site nucléotidique (flèches noires), certainement pour aider le  $\text{PO}_4^{2-}$  à en sortir. L'asparagine 67<sub>B</sub> se rapproche du site ainsi que la lysine 40<sub>A</sub>, et stabilisent le  $\text{PO}_4^{2-}$ .

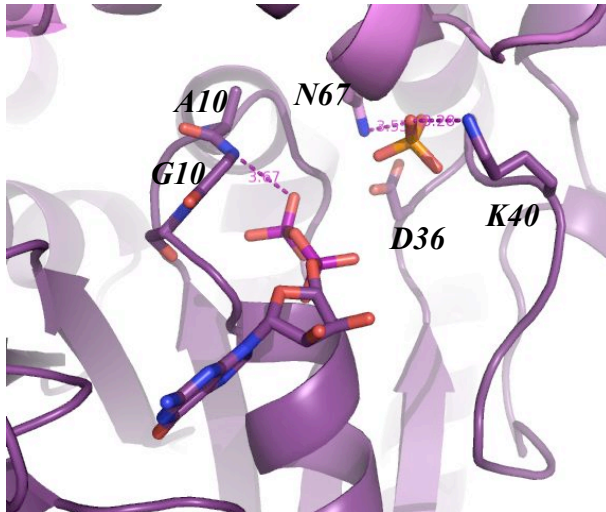


Figure.III.17.c. Structure de PAB0955 GTP<sup>4°C</sup>.

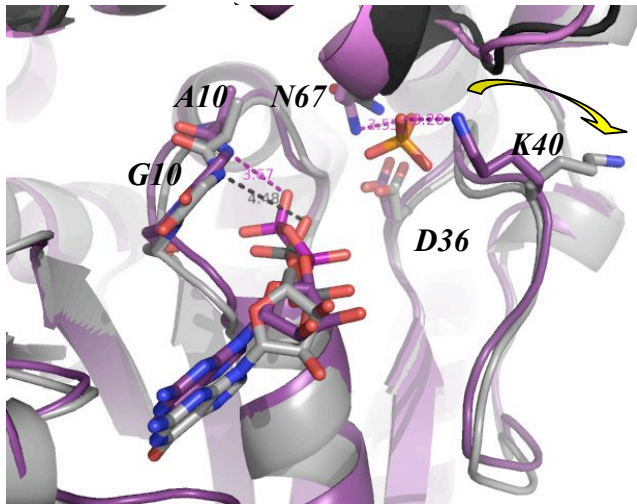


Figure.III.17.d. Superposition des structures de PAB0955 GTP<sup>4°C</sup> (violet) et de GTP<sup>18°C</sup> (gris).

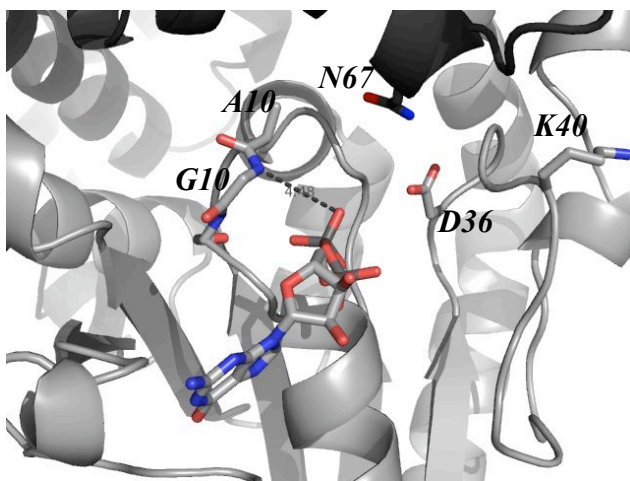


Figure.III.17.e. Structure de PAB0955 GTP<sup>18°C</sup>.

Le  $\text{PO}_4^{2-}$  est stabilisé par la lysine 40 et par l'asparagine 67, ces deux résidus semblent importants pour l'hydrolyse du GTP. Le GDP est lié avec le NH de la glycine 10, la P-loop suit le déplacement des groupements phosphates qui s'enfouissent dans le site nucléotidique dans la structure GTP<sup>4°C</sup> (fig.III.17.c).

Ensuite le  $\text{PO}_4^{2-}$  est libéré du site nucléotidique, le GDP se recule légèrement du site, c'est la structure GTP<sup>18°C</sup> (fig.III.17.d). La lysine 40, l'asparagine 67 ainsi que la P-loop se relaxent après le départ du  $\text{PO}_4^{2-}$ . Ces différentes régions reprennent une position proche de celle observée dans la protéine apo. Le GDP se relaxe également et se décale vers l'extérieur du site. Ce mouvement de va et vient du GDP dans le site, permet de supposer que notre hypothèse d'un accompagnement du  $\text{PO}_4^{2-}$  par le GDP au cours de l'hydrolyse, est probable. La structure GTP<sup>18°C</sup> montre que la P-loop s'ouvre. En effet la distance entre le NH de la glycine 10 et le phosphate- $\beta$  varie de 3.7 Å à 4.5 Å entre les structures GTP<sup>4°C</sup> et GTP<sup>18°C</sup>. Ainsi, le GDP est moins stabilisé et peut sortir de la P-loop, puis la structure revient dans la

conformation de la protéine apo prête à effectuer un nouveau cycle d'hydrolyse.

L'affinité de la protéine pour le GDP semble assez forte puisque le GDP reste fixé dans le site. Nous ne pouvons savoir, avec les structures, si la protéine, dans les gouttes de cristallisation, est capable de libérer le GDP issu de l'hydrolyse et de le fixer à nouveau, ou si elle le conserve après hydrolyse. Dans le cas où elle conserverait le GDP : soit la protéine possède un partenaire, de type GEF, capable de remplacer le GDP par du GTP soit l'affinité de la protéine pour le GTP étant plus grande que celle pour le GDP, l'échange se fait naturellement. Nous avons voulu tester cette dernière hypothèse en essayant d'échanger le GDP par du GTP. Les cristaux obtenus par co-cristallisation avec du GDP et trempés dans une solution de GTP se dissolvent, nous n'avons donc pas pu tester notre hypothèse.

En conclusion notre protéine est capable d'hydrolyser le GTP sans l'aide de partenaire. Nous avons pu remarquer qu'en solution et dans les cristaux, le magnésium est essentiel pour l'hydrolyse du GTP, puisqu'en présence d'EDTA la réaction ne se fait plus. En comparant les différentes structures, nous avons observé que l'asparagine 67 et la lysine 40 stabilisent un intermédiaire réactionnel. En conséquence ces deux acides aminés doivent jouer un rôle important dans l'hydrolyse du GTP.

## III.2. Recherche d'homologues structuraux et analyse

Afin de progresser dans la détermination de la fonction biologique de notre protéine et mieux comprendre les caractéristiques de sa fonction enzymatique, nous avons comparé notre structure avec celles de GTPases déjà connues.

Dans un premier temps, nous avons entrepris une analyse structurale comparative grâce au serveur DALI. Puis, nous avons étendu notre étude à l'ensemble des protéines à P-loop hydrolysant le GTP dont la structure est connue dans la PDB.

### III.2.1. Recherche d'homologues structuraux par le serveur DALI

Le programme DALI (Holm and Sander, 1993) est un serveur internet permettant de comparer une structure à l'ensemble de la PDB. Celui-ci a pour but d'identifier les protéines qui sont structurellement proches de la notre. La recherche d'homologues, à l'aide du serveur DALI, a été réalisée avec la structure du monomère de PAB0955.

Le résultat de cette recherche nous a donné une liste de 65 protéines classées par Z score donnant une idée du degré d'homologie structurale avec notre protéine. L'ensemble des protéines issues de cette recherche par DALI sont des protéines fixant des nucléotides triphosphates (ATP, GTP, UDP ...) et présentant un repliement de type Rossmann.

Ci-dessous, un tableau répertoriant les 5 premières protéines définies par DALI comme étant les plus proches structurellement de PAB0955 (fig.III.18), qui seront détaillées par la suite.

Z score	Protéine	Nucléotide	Code PDB	rmsd
9.8	Déthiobiotine synthétase	ATP	1BYI	2.1 Å pour 150 C $\alpha$
9.3	SRP	GTP	1NG1	1.94 Å pour 136 C $\alpha$
8.5	ArsA	ATP	1F48	1.97 Å pour 121 C $\alpha$
7.1	FTHFS	ATP	1EG7	1.75 Å pour 100 C $\alpha$
6.9	Petite protéine G (Ras)	GTP	1CTQ	2.9 Å pour 138 C $\alpha$

Figure.III.18. Tableau récapitulatif du résultat de la recherche d'homologues structuraux par DALI.

### III.2.1.1. La déthiobiotine synthétase d'*Escherichia coli*

La structure la plus proche de PAB0955 est celle de la protéine déthiobiotine synthétase d'*Escherichia coli* (1BYI) (Sandalova et al., 1999), avec un rmsd de 2.1 Å sur 150 C $\alpha$  superposés.

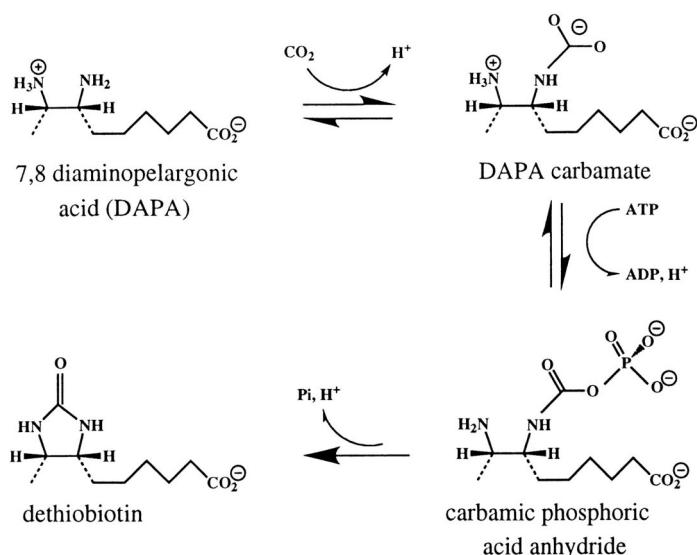
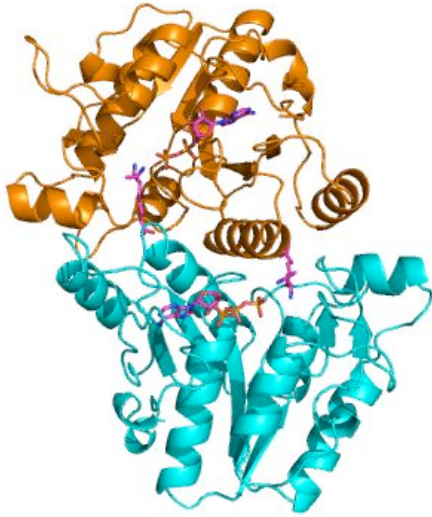


Figure.III.19.a. Cycle de la déthiobiotine synthétase.

Cette protéine intervient dans la voie de biosynthèse de la biotine en catalysant l'avant dernière étape consistant à produire de la déthiobiotine (fig.III.19.a.).

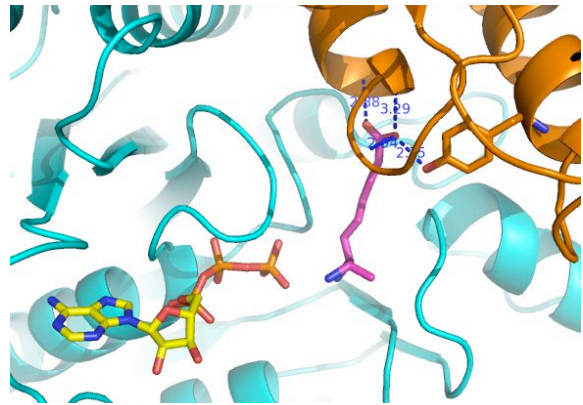




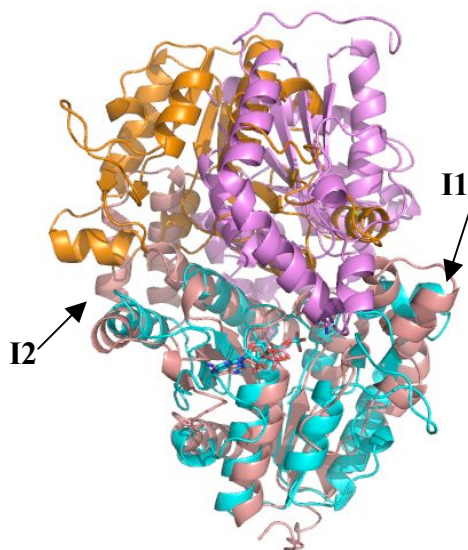
*Figure.III.19.b. Structure de la DTBS en complexe avec l'ATP et l'acide 7-8-Diaminopelargonic DAPA (1A82) (Kack et al., 1998).*

Chaque monomère fixe une molécule d'ATP sans avoir besoin de l'autre monomère. Par contre le substrat DAPA (rose) fixe sa partie  $\text{CO}^{2-}$  sur un monomère (en orange) et présente sa partie  $\text{NH}^{3+}$ , qui sera phosphorylée, au niveau du site nucléotidique de l'autre monomère. La fonction de phosphorylation du DAPA pour obtenir de la déthiobiotine ne peut donc se faire que si la protéine est dimérique.

C'est une ATPase de 24 kDa qui s'organise en homodimère avec un axe d'ordre 2 moléculaire qui coïncide avec l'axe 2 cristallographique (un monomère par unité asymétrique). Elle présente un repliement de type Rossmann avec 7 brins- $\beta$  parallèles entourés de 6 hélices- $\alpha$  (fig.III.19.b.). Les sites nucléotidiques sont à l'interface dimérique ; la distance entre les P-loop de chaque monomère est de 10 Å et la distance entre les phosphates- $\gamma$  de chaque ATP est de 24 Å.



*Figure.III.19.c. Structure de la DTBS en complexe avec l'ATP (jaune) et du DAPA (rose).*



*Figure.III.19.d. Superposition de PAB0955 (rose et violet) avec la DTBS (cyan et orange) (1A82).*

La comparaison de la DTBS (monomère cyan) avec PAB0955 (monomère rose) montre que le cœur des structures se superpose bien représentant environ 150 résidus. La région de PAB0955 des résidus 35 à 81, contenant I1, ne se superpose pas avec la DTBS, qui elle aussi possède une insertion à ce niveau (résidus 38 à 95). L'insertion I2 de PAB0955 n'a pas de région similaire dans DTBS qui est plus petite que notre protéine.

Comme pour notre protéine, le monomère de DTBS n'existe pas (Krell and Eisenberg, 1970). L'interface

dimérique est différente de celle observée dans PAB0955. Les deux monomères de la DTBS se posent l'un sur l'autre, il n'y a pas d'entrelacement comme dans notre protéine et la surface enfouie est moins grande. Chaque monomère de la DTBS enfouit 17% de sa surface à l'interface dimérique, soit  $1530 \text{ \AA}^2$ . La surface enfouie du dimère de DTBS est ( $3060 \text{ \AA}^2$ ) moins importante que celle de PAB0955 ( $4400 \text{ \AA}^2$ ). L'interface dimérique est moins contrainte dans la déthiobiotine synthétase permettant au substrat DAPA de se positionner au niveau du phosphate- $\gamma$ . Du fait de la rigidité du site nucléotidique de PAB0955, il est difficile d'imaginer un tel mode d'action dans notre GTPase.

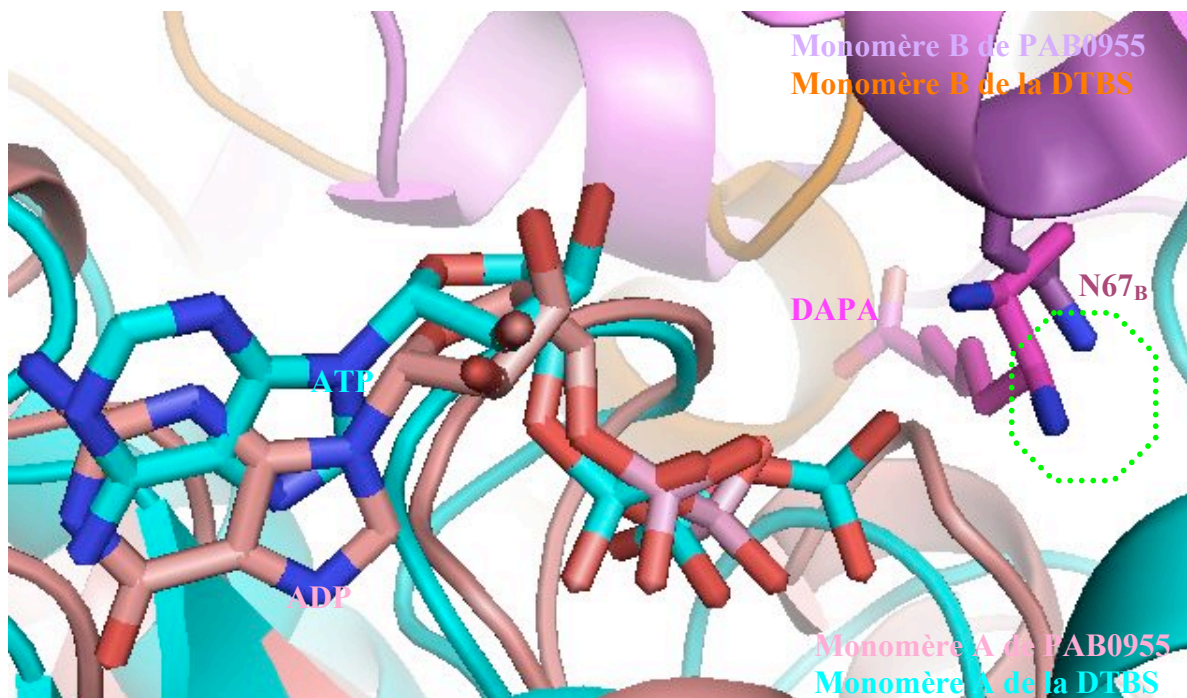


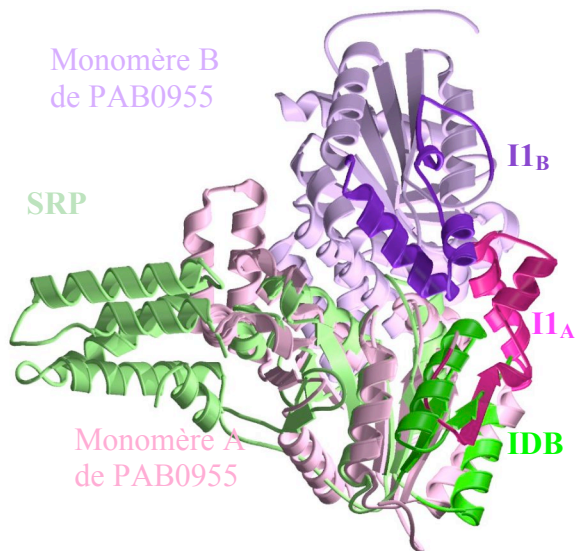
Figure.III.19.e. Superposition de PAB0955 (rose et violet) avec la DTBS (cyan et orange) (1A82).

Dans la DTBS, le substrat DAPA activé par le  $\text{CO}_2^{2-}$  (carbamate) vient par son oxygène du carbamate attaquer le phosphate- $\gamma$  de l'ATP. Dans la structure de la DTBS en complexe avec le substrat DAPA (rose) et l'ATP (cyan), le groupement NH qui va subir l'activation par le  $\text{CO}_2^{2-}$  puis attaquer l'ATP se superpose avec notre N67 (entouré en vert) (fig.III.19.e.). Ceci renforce l'idée que cet acide aminé joue un rôle dans l'hydrolyse du GTP, la formation du dimère est donc utile pour l'activité de la protéine PAB0955.

### III.2.1.2. La SRP de *Thermus aquaticus*

La deuxième protéine issue de la recherche par le serveur DALI, structurellement proche de PAB0955 est la SRP de *Thermus aquaticus* : Ffh (1NG1) (Freyman et al., 1999) avec un rmsd de 1.94 Å pour 136 C $\alpha$  superposés. La protéine Ffh et son récepteur (FtsY) sont impliqués dans la reconnaissance de séquence signal sur les peptides en cours de production et dans leur translocation à travers la membrane (voir introduction).

La protéine Ffh est une GTPase multidomaine qui possède un domaine G avec un repliement de type Rossmann, avec un feuillet- $\beta$  parallèle constitué de 7 brins- $\beta$  et entouré de 6 hélices- $\alpha$ . Contrairement à PAB0955 la SRP est monomérique et peut former un hétérodimère avec



*Figure.III.20. Superposition du monomère A de PAB0955 avec la SRP de *T. aquaticus*.*

son récepteur. Le domaine G la protéine Ffh se superpose avec celui de notre GTPase. Un domaine d'insertion nommé « Insertion Box Domain » (IBD) qui se superpose en partie avec notre insertion I1 a été identifié chez les SRP (Keenan et al., 2001) (fig.III.20.). Cette IBD est constituée par les résidus 129 à 181 dans Ffh de *Thermus aquaticus*, ce qui regroupe les brins- $\beta$  2 et 3 ainsi que les hélices- $\alpha$  2 et 3. L'insertion I1 (40 résidus) de PAB0955 est plus longue que l'IBD (33 résidus) de la protéine Ffh. Les résidus 51 à 77 de notre protéine n'ont

pas d'homologie structurale avec la protéine Ffh. La seconde insertion de PAB0955 ne se superpose pas avec la protéine Ffh et de même le domaine N<sub>ter</sub> de Ffh en hélice- $\alpha$  n'a pas d'homologie avec notre protéine qui est monodomaine.

La protéine Ffh en complexe avec le GTP peut lier son récepteur si lui-même est en complexe avec du GTP. L'hétérodimère ainsi formé s'auto-active pour permettre l'hydrolyse des 2 GTP. L'interaction entre Ffh et FtsY est forte puisque chaque protéine enfouie 1600 Å<sup>2</sup> à l'interface dimérique soit 14% de la surface totale de l'hétérodimère. La superposition des structures de Ffh en complexe avec du GDP-Mg (1NG1) (Freyman et al., 1999) avec le

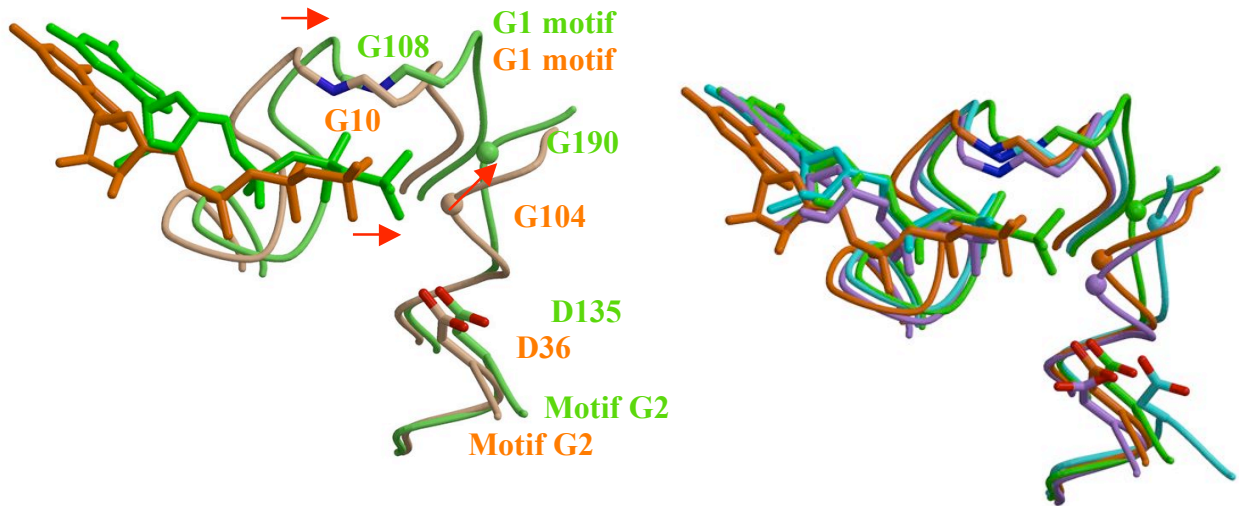
complexe Ffh/FtsY lié au GMPpnp (1RJ9) (Egea et al., 2004) montre que les motifs G2 et G3 sont affectées par la dimérisation.

Les motifs G2 de Ffh et FtsY sont comme celui de PAB0955, composés d'un acide aspartique qui fixe le magnésium. Dans Ffh de *Thermus aquaticus*, cet aspartate est le résidu 135. L'Asp135 est situé dans une boucle (134-141) qui subit un basculement vers le site nucléotidique dans le complexe actif avec le récepteur. La plus grande amplitude de mouvement dans cette boucle est observée pour le résidu 139 avec un déplacement de 5.2 Å. Ainsi, la distance entre l'oxygène Oδ1 de la chaîne latérale du D135 et du magnésium passe de 7.1 Å dans la structure monomérique à 4.3 Å dans la structure de l'hétérodimère Ffh/FtsY. La distance observée dans le complexe actif Ffh/FtsY entre le D135 et le Mg<sup>2+</sup> est proche de celle observée dans la structure de PAB0955 en complexe avec le GDP et le Mg<sup>2+</sup> (4.6 Å).

Le motif G3 se rapproche des groupements phosphates dans la structure du complexe Ffh/FtsY par rapport à celle de Ffh seule. Ce motif G3 correspond aux résidus <sup>187</sup>DTAGR<sup>191</sup> dans *Thermus aquaticus* ; le mouvement de plus grande amplitude est observé pour le résidu 191 avec un rmsd de 3.08 Å. Dans la structure Ffh-GDP la glycine du motif G3 est à 7.4 Å du phosphate-β, cette distance est de 6.2 Å dans le complexe avec le récepteur. Nous observons donc un mouvement de rapprochement du motif G3 vers les phosphates lorsque le complexe est formé. Dans PAB0955 en complexe avec le GDP et le Mg<sup>2+</sup>, la distance entre la glycine du motif G3 et le phosphate-β est de 5.5 Å ne permettant pas l'ajout d'un troisième phosphate dans cette configuration. Dans la structure de notre GTPase en complexe avec le GTP, nous avons observé un recul de 1.1 Å des phosphates α et β par rapport à ceux de la structure en complexe avec du GDP. Notre structure étant peu flexible, le motif G3 ne peut s'adapter à la nature du nucléotide qui s'adapte donc pour se fixer dans le site. La structure de la protéine Ffh étant flexible, elle s'adapte à la nature du nucléotide présent dans sa P-loop contrairement à PAB0955.

Ces différentes observations, ainsi que la conservation des positions relatives du phosphate-γ, des motifs G1 et G3 entre la structure GTP<sup>EDTA</sup> de PAB0955 et celle du complexe Ffh/FtsY indiquent que notre structure est plus proche de la conformation de Ffh activée par son récepteur que de Ffh seule (fig.III.20. b et c). En conclusion ceci nous laisse supposer que notre protéine est toujours en conformation active et qu'elle n'a sans doute pas besoin d'une autre protéine pour activer l'hydrolyse du GTP.





*Figure.III.20.b.* Superposition des structures  $GTP^{EDTA}$  de PAB0955 (orange) et GMPPcp du complexe Ffh/FtsY (vert) (1RJ9).

*Figure.III.20.c.* Superposition des structures MgGDP (violet) et  $GTP^{EDTA}$  (orange) de PAB0955 ainsi que les structures GMPPcp de Ffh/FtsY (vert) et GDP de Ffh (cyan).

### III.2.1.3. L'ATPase ArsA d'*Escherichia coli*

Dans *E. coli*, il existe un complexe de deux protéines Ars A et B qui forment une pompe pour détoxifier la cellule de l'arsenic (As) et de l'antimoine (Sb), deux métalloïdes toxiques. La protéine ArsA est la troisième protéine homologue de PAB0955 dans la recherche par DALI, avec un rmsd de 1.97 Å pour 121 C $\alpha$  superposés. Cette protéine de 583 acides aminés possède deux domaines : A1 et A2, qui se positionnent de sorte à mimer un (fig.III.21.a.). L'ArsA est capable de fixer des métaux (As et Sb) ce qui stimule son activité ATPase. Les deux sites sont requis pour l'activité ATPase et pour le transport des métaux. Les deux sites nucléotidiques sont à l'interface des deux domaines séparés par 8 Å (fig.III.21.b.). La surface de contact représente 13% de la surface totale avec une large cavité centrale au niveau des sites nucléotidiques. Le domaine A1 est constitué d'un feuillet- $\beta$  parallèle composé de 7 brins- $\beta$  et contient un brin- $\beta$  anti-parallèle à l'extérieur du feuillet, l'ensemble est entouré de 11 hélices- $\alpha$ . Le domaine A2 ne contient pas de brin- $\beta$  supplémentaire des 7 qui constituent le feuillet. Une des autres différences entre les deux domaines réside dans leur mode de fixation du magnésium. Le domaine A1 lie le  $Mg^{2+}$  par 3 molécules d'eau, l'hydroxyle de la Thr 22 et l'Asp 45 directement, alors que l'Asp 45 du domaine A2 fixe le magnésium via une molécule d'eau.

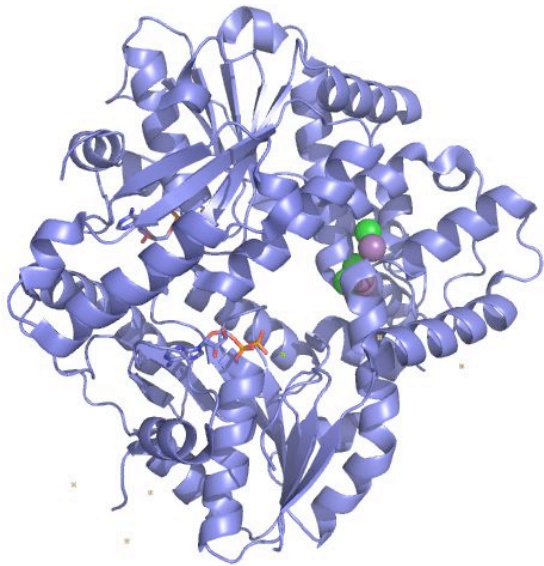


Figure.III.21.a. Structure de la protéine Arsenite (1F48) (Zhou et al., 2000) en complexe avec de l'ADP, des ions antimoine (Sb) et chlorure (Cl).

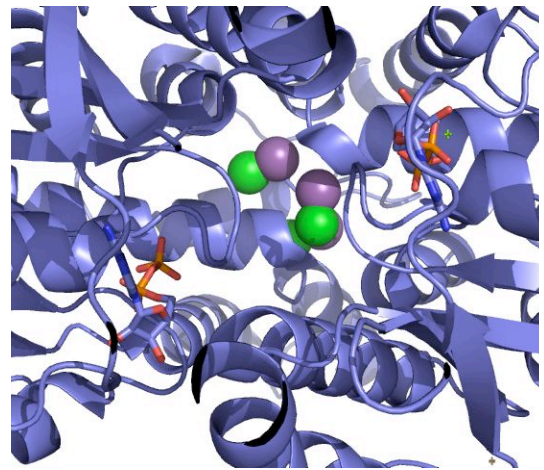


Figure.III.21.b. Structure du cœur de la protéine Arsenite.

En vert les ions chlorure, en violet les ions antimoine.

PAB0955 est superposé avec un des domaines de la protéine ArsA (A1). Les insertions de notre GTPase n'ont pas (ou peu) d'homologie structurale avec cette protéine. L'ArsA possède une insertion qui se superpose avec I1, mais qui est plus grande et de structure différente. Elle va interagir avec une insertion supplémentaire (147-191) de l'ArsA qui se situe au niveau de notre hélice- $\alpha$ 7. La seconde insertion de l'ArsA n'existe pas dans PAB0955 ; elle interagit avec son homologue sur le second domaine A2. La seconde insertion de PAB0955 n'a pas d'homologie structurale avec l'ArsA. Les insertions de PAB0955 et de l'ArsA ne sont pas homologues. La fixation du nucléotide pour la partie des groupements phosphates se fait de la même façon dans les deux protéines. Le D45 de l'ArsA et le D36 de PAB0955, fixant le magnésium, se superposent.

Les interfaces dimériques sont très différentes entre ces deux protéines, puisque PAB0955 est compacte alors que l'ArsA possède une grande cavité centrale et l'interaction entre les domaines A1 et A2 se fait principalement par les insertions extérieures. La cavité de l'ArsA s'étend sur environ 23 Å entre les nucléotides et les atomes d'antimoine pouvant accepter de possible partenaire contrairement à PAB0955 qui possède une structure compacte.

#### III.2.1.4. La N10-Formyltetrahydrofolate Synthétase de *M. thermoacetica*

Avec un rmsd de 1.75 Å pour 100 C $\alpha$  superposés la N10-Formyltetrahydrofolate Synthétase (FTHFS) est la quatrième structure la plus homologue à notre GTPase selon DALI (Radfar et al., 2000). La FTHFS est une ATPase de 559 acides aminés qui s'organise en homotétramère. Cette protéine est impliquée dans la voie de biosynthèse de purine, thymidylate, méthionine, d'histidine, pantothénate et de Met-ARNt formyl. La FTHFS catalyse la réaction de formylation du Tétrahydrofolate par hydrolyse d'une molécule d'ATP. Chaque monomère est constitué de trois domaines, un large et deux petits. Le domaine 1 contient les résidus 7 à 120 et 254 à 431 ; c'est le grand domaine constitué par un feuillet parallèle avec 7 brins- $\beta$  entourés de 8 hélices- $\alpha$ . Ce premier domaine fixe probablement l'ATP par une P-loop non canonique puisque sa séquence n'est pas un motif de Walker de type A et est plus longue que les autres P-loop (<sup>64</sup>AITPTAGEGKT<sup>75</sup>).

La superposition du domaine 1 de la FTHFS, qui adopte un repliement de type Rossmann supposé fixer l'ATP, avec PAB0955 présente moins d'homologie que les autres structures. Ce domaine 1 est composé de 2 régions de la protéine éloignées en séquence. Entre les résidus 120 et 254 vient se placer le domaine 2 de la FTHFS qui s'insère au niveau de la jonction entre les brins- $\beta$  2 et 3 de PAB0955. Le troisième domaine de cette protéine se situe en C<sub>ter</sub> du domaine 1 et ne se superpose pas avec notre GTPase. Dans cette structure, il n'y a pas de régions qui se superposent avec l'insertion I2 de PAB0955.

Il n'y a pas de structure en complexe avec un nucléotide disponible pour la FTHFS, il est donc difficile de comparer son mode de fixation des nucléotides avec notre protéine. Cette protéine est très différente de notre structure ; multidomaine, tétramérique, avec une P-loop particulière. En conséquence peu d'information fonctionnelle peuvent être apportées par cette structure.

#### III.2.1.5. La petite protéine G Ras de l'homme

Alors que les quatre premiers homologues obtenus par la recherche avec DALI, contiennent un feuillet- $\beta$  parallèle comme notre protéine, celui du cinquième homologue (hRas) est mixte. Comme indiqué dans l'introduction, la famille des protéines comme hRas contient un feuillet mixte avec le brin- $\beta_2$  antiparallèle aux autres brins.

Le cœur de PAB0955 se superpose avec celui de hRas, avec un rmsd de 2.9 Å sur 138 C $\alpha$  superposés. Les brins- $\beta_2$  se superposent, mais sont de sens opposé, de sorte que le N<sub>ter</sub> du brin- $\beta_2$  de hRas se superpose avec le C<sub>ter</sub> du brin- $\beta_2$  de PAB0955. La région switch I (acides aminés 24 à 36) de hRas n'a pas d'homologue dans PAB0955. Dans notre GTPase, ce sont les insertions I1, I2 ainsi que l'hélice- $\alpha_5$  qui n'ont pas d'homologie structurale avec hRas.

Les motifs G1, G4 et G5 sont bien superposés entre les deux protéines. Le motif G2 de hRas (résidus 24-36) et celui de PAB0955 (D36) n'ont pas d'homologie structurale (leurs tailles sont très différentes). Cependant, le D36 de PAB0955 se retrouve en superposition à proximité de la T35 de hRas. Ces deux résidus sont impliqués dans la fixation du magnésium, soit directement dans hRas, soit via une molécule d'eau dans notre GTPase.

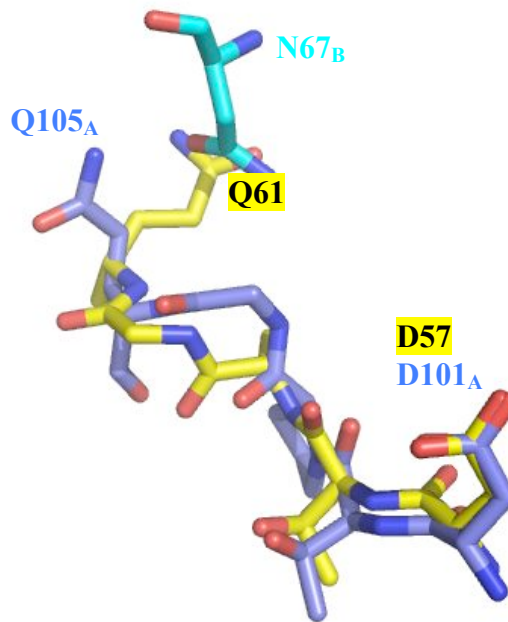
Le motif G3 est dans ces deux protéines un motif de Walker de type B, DXXG, avec deux résidus conservés. Ce motif G3 dans hRas correspond à la séquence <sup>57</sup>DTAGQ<sup>61</sup> et dans PAB0955 à <sup>101</sup>DTPGQ<sup>105</sup>. Le remplacement d'une alanine en proline dans ce motif change sa conformation structurale.

Le premier résidu de ce motif, D, est impliqué dans la fixation du magnésium, se superpose entre les deux protéines.

Le second, la thréonine, se superpose un peu moins bien avec une distance de 1.4 Å sur les C $\alpha$ , mais les chaînes latérales se trouvent quasiment au même endroit.

Par contre les C $\alpha$  entre la Proline 103 de PAB0955 et l'Alanine 59 de hRas sont distants de 2.9 Å, avec un éloignement du motif G3 par rapport au nucléotide dans notre protéine. La conformation du motif G3 imposé par la présence de la proline, fait que les deux résidus qui suivent, la glycine et la glutamine, sont éloignés entre les deux protéines.

Ainsi la chaîne latérale de la Q105 de PAB0955 ne se superpose pas avec celle de la Q61 de hRas. Dans notre protéine la Q105 est éloignée du nucléotide et ne peut jouer le même rôle de stabilisation de l'état intermédiaire que dans hRas. Lors de la superposition de ces motifs G3, nous nous sommes aperçus que, même si notre Gln105 ne se superposait pas avec la Gln61 de hRas, au contraire l'Asn67, elle se superpose avec la Gln61 de Ras (fig.III.22.). Ce résidu asparagine 67 qui appartient à la boucle GPN, interagit avec le PO<sub>4</sub><sup>2-</sup> dans le site nucléotidique de l'autre monomère de PAB0955. Cette observation renforce notre idée que l'interface dimérique est utile à la réaction d'hydrolyse du GTP.



*Figure.III.22. Superposition de hRas (jaune ICTQ) avec MgGDP de PAB0955 (bleu pour le monomère A et cyan pour le B).*

#### III.2.1.6. Informations issues de la recherche avec le serveur DALI

En comparant PAB0955 avec l'ensemble des protéines issues de la recherche par DALI, il apparaît que notre protéine possède plus de points communs avec les ATPases qu'avec les GTPases. Beaucoup de protéines hydrolysant l'ATP le font sans changer de conformation en fonction du type de nucléotide qu'elles fixent et plusieurs sont homodimériques. Les GTPases au contraire sont majoritairement des monomères et nécessitent la présence d'un partenaire pour hydrolyser le GTP.

Les valeurs de rmsd, toutes supérieures à 1.75 Å, obtenus pour les 5 premiers homologues structuraux, montrent qu'aucune de ces protéines n'est vraiment proche de PAB0955. Le repliement de type Rossmann est conservé puisque ces homologues possèdent une P-loop et hydrolyse des nucléotides tri-phosphate. De plus, les quatre premières protéines les plus homologues à PAB0955 possèdent toutes un feuillet-β parallèle, ce qui est plus rare dans les GTPases.



La conformation de notre protéine est différente de celles des autres protéines analysées. La plus proche GTPase de PAB0955 au niveau structural, est la protéine Ffh pour ce qui est du domaine G. La protéine Ffh est une protéine multi-domaine et monomérique contrairement à PAB0955. L'hétérodimère Ffh/FtsY est la forme active de la Ffh. La formation du complexe

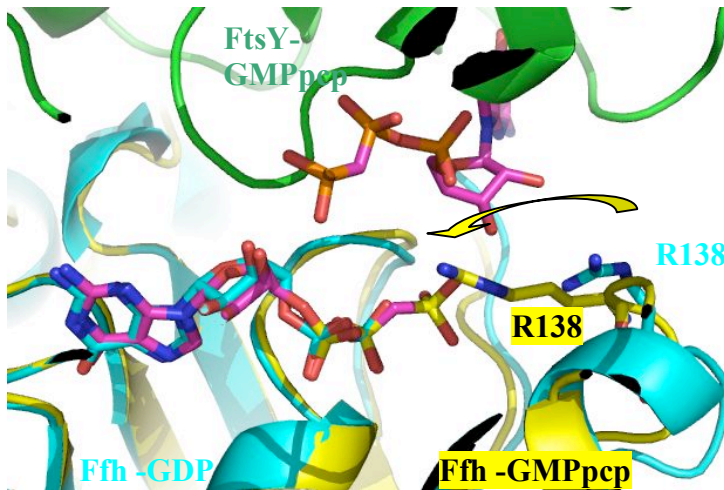


Figure.III.23.a. Superposition des structures de SRP seule (ING1 cyan) et en complexe avec son récepteur (1RJ9 jaune pour la SRP et vert pour SR).

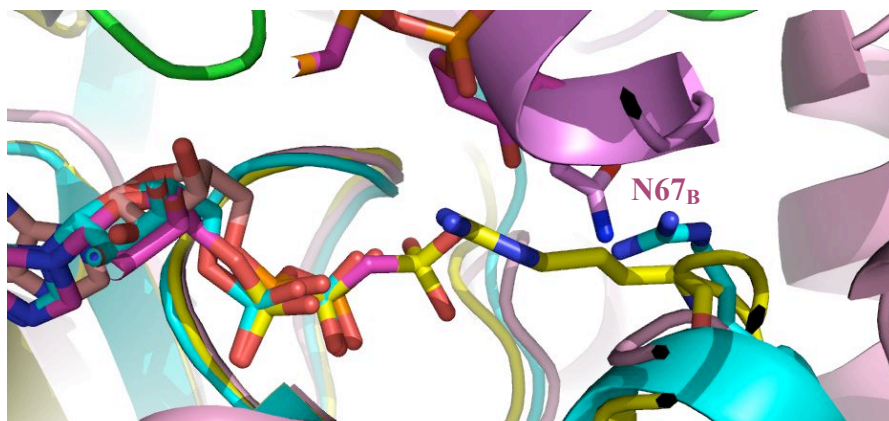


Figure.III.23.b. Superposition des structures GDP-SRP (cyan), GMPPcp-SRP-SR (jaune) et MgGDP-PAB0955 (rose pour le monomère A et violet pour le monomère B).

provoque des changements dans les boucles contenant les motifs G2 et G3. Dans le motif G3, il y a l'arginine 138 de la Ffh, lors de la formation du complexe cette arginine se rapproche du nucléotide et vient se lier avec le phosphate- $\gamma$  du GTP (fig.III.23.a.). Cette hétérodimérisation active l'hydrolyse du GTP.

L'arginine 138 de la Ffh, qui semble aider à l'hydrolyse du GTP, se superpose avec notre asparagine 67. Dans la structure MgGDP notre N67 se superpose avec la R138 de la Ffh seule (inactive).

Notre N67 peut prendre, par simple rotation, la même position que l'arginine 138 de la Ffh lorsqu'elle est en complexe avec son récepteur. Notre structure étant plus proche de la Ffh en complexe avec son récepteur que du monomère de Ffh, cela suppose que la structure de PAB0955 est active pour l'hydrolyse du GTP et que l'asparagine 67 joue un rôle important dans l'hydrolyse.

Cette analyse ne nous a pas donné d'élément sur la fonction biologique de notre protéine. Cependant, il nous est apparu évident que le classement de PAB0955 dans le COG1100 avec

la petite protéine G Sar1 est une erreur, il faudrait créer un sous-classement dans ce COG1100. La structure de PAB0955 est très différente de celle des petites protéines G, en raison notamment de son feuillet- $\beta$  parallèle et de sa conservation structurale indépendante de la nature du nucléotide fixé. Sur le plan fonctionnel, il est probable que notre protéine ne soit pas impliquée dans des voies de transduction du signal comme le sont les petites protéines G, en effet PAB0955 n'est pas un interrupteur moléculaire. Au niveau de l'hydrolyse du GTP, grâce à la comparaison avec Ffh et FtsY, il apparaît que notre protéine est en conformation active permanente et qu'elle ne nécessite pas la présence d'un partenaire pour hydrolyser le GTP.

### [III.2.2. Recherche d'homologues structuraux dans la PDB](#)

Pour essayer de déterminer la fonction de PAB0955, nous avons décidé d'analyser les différentes structures de protéines ATPase ou GTPase contenues dans la PDB. Ceci dans le but d'avoir une vision globale et complète de toutes les protéines qui pourraient nous apporter des informations fonctionnelles sur PAB0955.

Dans la PDB, 255 protéines adoptent un repliement de type Rossmann, contiennent une P-loop et ont une activité d'hydrolyse du GTP ou de l'ATP. La majorité de ces protéines sont des ATPases (174), dont la plupart possèdent un feuillet- $\beta$  parallèle (120) comme les kinases nucléotide monophosphate (NMPK), les protéines de la famille MinD (impliquées dans la division cellulaire), des ATPases liant l'ARN ou l'ADN et également des kinases.

Les 81 autres protéines identifiées dans cette large étude sont des GTPases, dont 90% contiennent un feuillet- $\beta$  mixte avec le brin- $\beta_2$  anti-parallèle aux autres brins- $\beta$ . Parmi ces protéines à feuillet- $\beta$  mixte, il y a les petites protéines G et les autres interrupteurs moléculaires.

Les GTPases avec un feuillet- $\beta$  parallèle sont seulement 14. Parmi elles il y a 4 protéines SRP, 2 récepteurs de SRP, 3 tubulines et 5 FtsZ (homologue bactérien de la tubuline). Cependant, la tubuline et FtsZ adoptent un repliement de type Rossmann mais ne possèdent pas de domaine G. En effet, dans la tubuline et FtsZ, le site nucléotidique est formé la dimérisation des protéines au cours de la formation du protofilament et ce site ne contient pas les motifs G (voir introduction).

Par conséquent comme pour l'analyse structurale par DALI, la recherche dans la PDB ne nous a pas apporté plus d'informations sur la fonction de notre GTPase. En analysant la

structure des différents interrupteurs moléculaires, cela a renforcé la conclusion que notre protéine n'appartenait pas à cette famille de GTPases. Les plus proches GTPases au niveau de la structure du domaine G sont les SRP et leurs récepteurs. Toutefois, ces protéines sont multidomaines, monomériques, et leur site nucléotidique change de conformation lors de la dimérisation Ffh/FtsY pour devenir actif. Même si nous n'avons pas d'information sur la fonction de PAB0955, cette analyse a exclu quelques pistes fonctionnelles. PAB0955 n'est pas un interrupteur moléculaire, elle ne doit pas être impliquée dans la transduction de signal. Ressemblant davantage aux SRP, elle est peut-être capable de transporter un partenaire comme le fait la SRP avec le ribosome pour l'amener au translocon.

De cette analyse structurale, nous pouvons conclure que notre protéine constitue une nouvelle famille de GTPases. Les 64 protéines de cette famille, contenues chez les Archaea et les Eucaryotes, ont une homologie de séquence qui nous laisse supposer que la structure sera conservée. La fonction biologique de cette nouvelle famille de protéines peut également être nouvelle pour des GTPases, expliquant pourquoi il n'y a pas de structure proche de notre protéine.

### [III.2.3. Recherche d'homologue pour l'insertion I2](#)

Si le cœur de la protéine, constitué par le repliement de type Rossmann contenant environ 150 acides aminés, se superpose avec les autres protéines ; il n'en est pas de même pour l'ensemble de la structure de PAB0955. Les deux insertions ont notamment peu d'homologie structurale pour I1 et aucune pour I2.

L'insertion I2 semble très caractéristique de notre GTPase, que ce soit par sa séquence ou par sa conformation repliée sur l'autre monomère (interaction par l'hélice- $\alpha$ 10). Pour savoir s'il existait des homologues structuraux de cette insertion, nous avons décidé d'effectuer une nouvelle recherche par DALI, uniquement sur cette partie I2 (183-217).

Le résultat de cette recherche montre que cette insertion est structurellement homologue à la famille des "Heat repeat" (Huntingtin-Elongation-A subunit-TOR). Un motif HEAT contient 32 à 61 résidus formant deux hélices anti-parallèles en hairpin, avec la première hélice plus courte que la seconde. Une boucle contenant une proline (196 dans PAB0955) facilite le changement de direction entre les deux hélices. L'interface entre les deux hélices est hydrophobe grâce à des résidus conservés, et les deux hélices forment un angle de 32°.



Dans les protéines, ces motifs HEAT sont organisés en arches ou en anneaux. Ils sont caractérisés par une répétition de repliements de 3 à 36 paires d'hélices- $\alpha$  antiparallèles formant des supers-hélices appelées "HEAT", comme dans Ro (Macrae and Doudna, 2005) ou dans l'importine- $\beta$  (Cingolani et al., 1999). Les protéines possédant ces nombreuses répétitions de motifs HEAT ont une fonction commune de reconnaissance protéine-protéine dans plusieurs processus cellulaires (Andrade et al., 2001). Le domaine de liaison de la protéine Arf de l'adaptin- $\gamma$  contient un seul motif HEAT et sa fonction est également le recrutement de la petite protéine G Arf pour la localiser au Golgi (Shiba et al., 2003). C'est pourquoi même sans répétition, ce motif peut tout de même servir d'interaction protéine-protéine. Dans PAB0955 il y a un motif HEAT par monomère. Les deux motifs HEAT du dimère sont tête-bêche, et pourraient servir ensemble à lier un partenaire protéique.

L'insertion I2, comme nous l'avons indiqué dans le paragraphe III.1.3., est la région qui subit les mouvements de plus grande amplitude entre les structures de PAB0955. Avec les cristaux obtenus par co-cristallisation avec du GDP, qui appartiennent au groupe d'espace  $P2_1$ , nous avons observé deux conformations différentes de cette I2. Le nucléotide semble influencer le groupe d'espace des cristaux, peut être car la région I2 a une conformation nucléotide dépendante. Cette hypothèse permet de supposer que si I2 sert à l'interaction de PAB0955 avec un partenaire, cette interaction pourrait se faire de manière nucléotide dépendante.

La protéine PAB0955 pourrait donc interagir avec un partenaire pour lui transférer l'énergie d'hydrolyse ou pour le transporter.

### III.3. Identification d'une nouvelle famille

Le repliement de PAB0955 est celui d'un domaine G pour le cœur de la protéine. Toutefois plusieurs spécificités structurales font que cette protéine est unique.

Premièrement son feuillet- $\beta$  parallèle la distingue des interrupteurs moléculaires. Sa conservation structurale quelle que soit la nature du nucléotide fixé dans sa P-loop. Ceci la différencie de toutes les GTPases car PAB0955 n'adapte pas sa structure au nucléotide, au contraire c'est le nucléotide qui s'adapte. Sa conformation dimérique biologique est également unique pour une GTPase.

L'analyse structurale de nos différents complexes nous a permis d'identifier des résidus supposés importants pour le mécanisme d'hydrolyse du GTP. De plus, par l'analyse des

séquences des 64 protéines constituant la famille de PAB0955, il s'avère que ces résidus supposés importants sont conservés (voir l'alignement de séquence dans l'introduction).

### III.3.1. L'interface dimérique

L'analyse structurale a mis en évidence l'importance de l'interface dimérique, avec la participation du N67 d'un monomère au site nucléotidique de l'autre monomère.

L'interface dimérique est constituée de 45 acides aminés, dont 27% sont homologues et 20% sont strictement conservés (G7, D36, M60, G65, P66, N67, G104 et E107) dans les protéines homologues à PAB0955 (archaea et eucaryote) (fig.III.24.).

Un axe de symétrie d'ordre 2, cristallographique ou non selon les groupes d'espace, passe au niveau de l'interface dimérique, chaque monomère engage donc les mêmes résidus à l'interface.

Cette interface dimérique doit être présente dans les autres protéines homologues de PAB0955 puisque les résidus engagés sont conservés. Ceci signifie que le motif GPN sera également présent à l'interface dans les autres protéines de cette famille, par conséquent il pourra remplir la même fonction que dans PAB0955.

Enfouis	-----***-----*-*-----**--*-
Q9HCN4-Hu	22 LLVL <b>GMAGSGKT</b> TFVQRLTGHHLHAQTPPY <b>VINLD</b> PAVHEVPFPANIDIRDTVKYKEVMK
Q8VCE2-Mo	22 LLVL <b>GMAGSGKT</b> TFVQRLTGHHLNKGCPY <b>VINLD</b> PAVHEVPFPANIDIRDTVKYKEVMK
P47122-Sc	6 IICI <b>GMAGSGKT</b> TFMQRLNSHLRAEKTPPY <b>VINLD</b> PAVLRVVPYGANIDIRDSIKYKKVME
Q9UYR9-PA	3 VVfV <b>GTAGSGKT</b> TLTGFEGRYLEDNYKVAY <b>VN-LD</b> TGVKELPYEPSIDVREFVTVEEIMR
Q8U3I2-PF	3 IAFL <b>GTAGSGKT</b> TLTGAFGRYLEENYRVAY <b>VN-LD</b> TGVENLPYKPDIDVRDFITVRDLMK
Q97Z85-SS	8 VFIL <b>GTAGSGKT</b> TLTKNLQDYLLDQEMDTA <b>VINLD</b> PAVEHLPYTPDFDVRDYVDAYEVMQ
Consensus	---- <b>GxAGsGKt</b> t-----l----- <b>v-NLD</b> --v--p-----d-r-----M-
	G1 <span style="margin-left: 250px;">G2</span>

Enfouis	-*--***--**-----***--*-----*
Q9HCN4-Hu	82 QYGL <b>GPN</b> GGIVTSLNLFATRFDQVMKFIKAQNMSKYVLI <b>DTPGQ</b> IEVFTWSASGTIITE
Q8VCE2-Mo	82 QYGL <b>GPN</b> GGIVTSLNLFATRFDQVMKFIKAQNTFRYVLI <b>DTPGQ</b> IEVFTWSASGTIITE
P47122-Sc	66 NYQL <b>GPN</b> GAIIVTSLNLFSTKIDQVIRLVEQKKDKFQNCII <b>DTPGQ</b> IECFVWSASGAIITE
Q9UYR9-PA	62 EG-Y <b>GPN</b> GAIVESYDRLMEKFNEYLNKILRLEKENDYVLI <b>DTPGQ</b> METFLFHEFGVRLME
Q8U3I2-PF	62 EG-Y <b>GPN</b> GAIVESYDRLLRVEHYINKILELEKSNDYVLI <b>DTPGQ</b> METFFIFHEFGVKLMN
Q97Z85-SS	68 NYHL <b>GPN</b> SSLIASIDLILTKASEIKSEIDQIEAN--YVLV <b>DTPGQ</b> IELFAYRDTGRLISQ
Consensus	---- <b>GPN</b> -----s----- <b>DXPGQ</b> -E-f-----g-----
	GPN <span style="margin-left: 250px;">G3</span>

Enfouis		-----*-----*-----**-*--**-*--***--*-----	
Q9HCN4-Hu	142	ALASSFPTVVIYVMDTSRSTNPVTFMSNMLYACSI LYKTKLPFIVVM	NKTD IIDHS---F
Q8VCE2-Mo	142	ALASSFPTVVIYVMDTSRSTNPVTFMSNMLYACSI LYKTKLPFIVVM	NKTD IIDHS---F
P47122-Sc	126	SFASSFPTVIAYIVDTPRNSPTTFMSNMLYACSI LYKTKLPMIVVF	NKTD VCKAD---F
Q9UYR9-PA	121	NLP---YPLVVYISDPEILKPKNDYCFVRRFFALLIDLRLGATTIPAL	NKVD LLSEEKER
Q8U3I2-PF	121	NLP---YPLVVYLSDPEILKRPHDYCFVRRFFSLLIDLRLGATTVPAL	NKVD LLSEKEFER
Q97Z85-SS	126	LIRGNKALGLFLLDSFLAKEARSFISLLLLSSSIKFRDLPLIINIL	NKID LLTEKELEQ
Consensus		-----d-----i-----	NKXD
			G4
Enfouis		-----*--**--**--*--**-----	
Q9HCN4-Hu	199	AVEWMQDFEAFQDALNQET-----TYVSNLTRSMSLVLDEFYSSLRVVG	SAV LG
Q8VCE2-Mo	199	AVEWMQDFEAFQDALNQET-----TYVSNLTRSMSLVLDEFYSSLRVVG	SAV VG
P47122-Sc	183	AKEWMTDFESFQAAIKEDQDLNGDNGLSGYMSLVNSMSLMLEEFYSQLDVVG	SSF TG
Q9UYR9-PA	178	HRKYFEDIDYLTARLKLDP-----SMQGLMAYKMCMMTEVLPPVRVLYL	SAK TR
Q8U3I2-PF	178	HKKYFEDIEYLTTKLKFDV-----STQGLLAYKMCQVFPELSPVRVLYL	SAK TG
Q97Z85-SS	186	ILAWGDNAENLIDELGRLD-----EYSLELVNLLIESLS-YNLIPM	SSE EG
Consensus		-----e-----	S--g
			G5

Figure.III.24. Alignement multiple de plusieurs protéines de la famille de PAB0955 (Q9UYR9), XAB1 Humaine (Q9HCN4), XAB1 de souris (Q8VCE2), Saccaromyces Cerevisiae (P47122), Pyrococcus furiosus (Q8U3I2) et Sulfolobus solfataricus (Q97Z85).

Les étoiles indiquent les acides aminés enfouis à l'interface dimérique, les rouges pour les acides aminés conservés et les bleus pour les résidus homologues.

### III.3.2. Conservation des motifs G

Dans les 64 protéines homologues en séquence à PAB0955, nous avons identifié que les motifs G étaient conservés (fig.III.24). Seul le G5 est présent dans seulement 50% de ces protéines, ce qui est également le cas dans les autres GTPases.

Le motif G1 est toujours un motif de Walker de type A, la séquence consensus est GXXGXGK(T/S), avec une glycine en position quatre strictement conservée. La P-loop de ces 64 protéines adoptera, sûrement, la même conformation et fixera les groupements phosphates.

Le motif G2 spécifique contenant un acide aspartique liant le magnésium via une molécule d'eau est également conservé. Son positionnement sera sans doute conservé, il est situé à la fin du brin-β<sub>2</sub> après une séquence d'acide aminé hydrophobe conservé. Nous avons nommé ce motif <sup>33</sup>VNLD<sup>36</sup> dans PAB0955. La valine 33 de ce motif VNLD est en interaction hydrophobe avec deux résidus dont la nature hydrophobe est conservée : les résidus Ile48 et Val98. La leucine 35 de ce motif est également en interaction avec des résidus hydrophobes

de nature conservée ; la Val50 et l'Ile100. La conservation de ces résidus hydrophobes nous laisse supposer que leur interaction sera également conservée. Ainsi l'aspartate 36 aura la même position en interaction indirecte avec le magnésium.

Le motif G3 sert à la fixation du magnésium par le D101 conservé, à l'interaction avec le phosphate- $\gamma$  par la G104 aussi conservée. Ce motif G3, DXPGQ, est très bien conservé dans ces 64 protéines, avec la présence de la Proline strictement conservée. Cette proline qui donne une conformation particulière au G3 de PAB0955, situé à l'interface dimérique, fait que la glutamine de notre protéine est éloignée et ne peut donc pas stabiliser l'état intermédiaire de la réaction (comme dans les petites protéines G). Nous pouvons donc supposer que dans toutes les protéines homologues à PAB0955 cette glutamine n'interviendra pas dans l'hydrolyse du GTP.

Le motif G4, (N/S/T)KXD fixe la base du nucléotide et fait la spécificité de PAB0955 pour la guanine, donc pour le GTP. Les protéines de la même famille que notre protéine hydrolyseront très certainement le GTP (comme le fait d'ailleurs XAB1).

La conservation de ces motifs nous indique que les protéines de la famille de PAB0955 fixeront de la même façon et sûrement avec la même affinité le GTP et le GDP que le fait notre GTPase.

### [III.3.3. Conservation du motif spécifique GPN](#)

Comme indiqué dans l'analyse entre hRas et notre GTPase, l'Asn67 de PAB0955 semble jouer le même rôle que la Gln61 de hRas : stabiliser le  $\text{PO}_4^{2-}$  issue de l'hydrolyse du GTP. De part sa position dans le site nucléotidique de l'autre monomère, cette N67 expliquerait, par son action dans le mécanisme d'hydrolyse, l'importance de la structure dimérique.

L'asparagine 67 se situe en  $N_{\text{ter}}$  de l'hélice- $\alpha 4$ , dans l'insertion I1, dans une séquence consensus GPN strictement conservée dans les protéines homologues de PAB0955. Ce motif GPN est une signature très spécifique de cette nouvelle famille. Nous l'avons donc nommée la famille des GTPases à boucle GPN (GPN loop GTPase). Ce motif très conservé va de pair avec la conservation de la structure quaternaire en homodimère des protéines de cette famille

de GTPase à boucle GPN. Ce motif GPN devrait avoir la même localisation, à l'interface et la même fonction dans les protéines de cette nouvelle famille de GTPases.

Grâce à l'analyse structurale de PAB0955 nous avons pu mettre en évidence que nous avons découvert une nouvelle famille de GTPase. Il est donc logique que nous n'ayons pu trouver de structure, suffisamment proche de la notre, pour nous aider à comprendre la fonction biologique de PAB0955. Cependant, ses particularités nous ont permis d'émettre des hypothèses sur le mécanisme d'hydrolyse du GTP. La compréhension de l'hydrolyse du nucléotide pourra nous aider à comprendre les différences d'activité des mutants de ces protéines.

CHAPITRE IV

LES MUTANTS DE

PAB0955

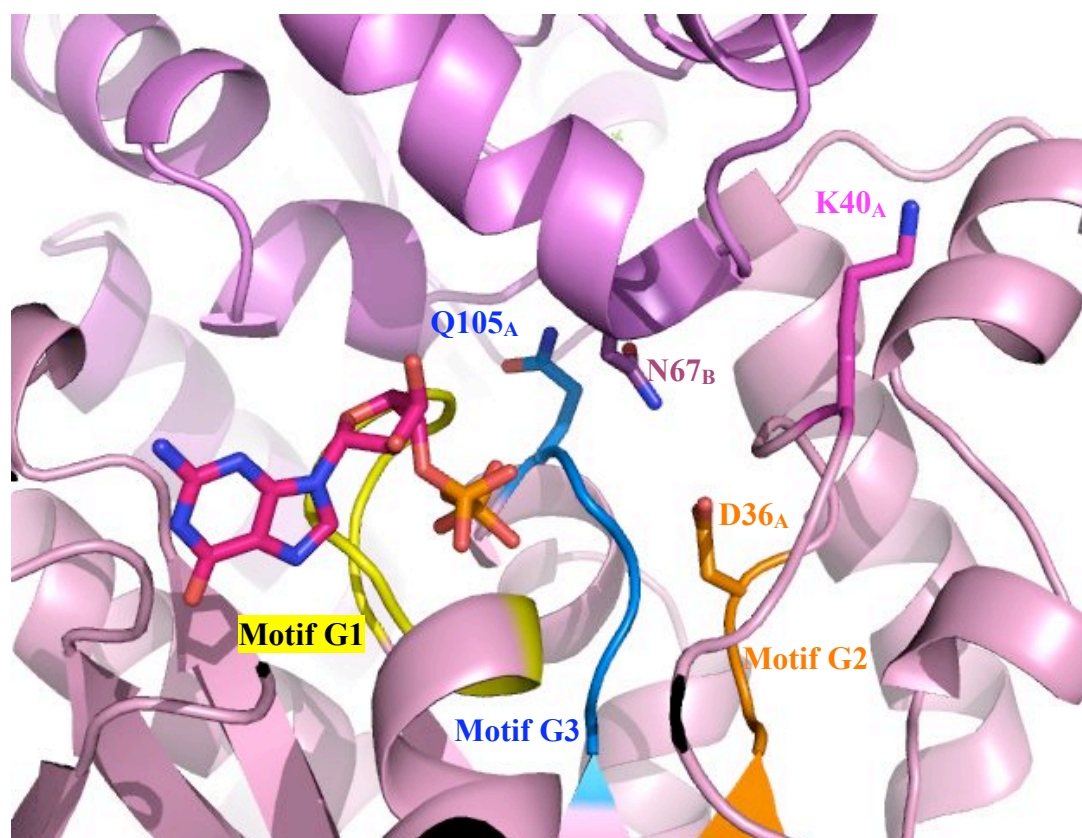




## IV.1. Le choix des différents mutants

Avec nos différents complexes, nous avons obtenu des structures à différents instants de la réaction d'hydrolyse du GTP. En analysant ces structures avec celles disponibles dans la PDB pour des GTPases, dont le mécanisme d'hydrolyse a été étudié, nous avons déterminé le rôle potentiel de plusieurs résidus.

Les acides aminés de la P-loop n'ont pas été mutés car ils sont conservés dans les autres GTPases et ne sont donc pas spécifiques de notre protéine. De plus, dans les petites protéines G, beaucoup de mutants sur les résidus de la P-loop ont été étudiés et nous pouvons supposer que les effets observés sur ces protéines seront similaires dans PAB0955. Nous nous sommes donc intéressé aux acides aminés spécifiques et conservés dans la famille des GTPases à boucle GPN (fig.IV.1.).



*Figure.IV.1. Structure du site nucléotidique de PAB0955.*

### IV.1.1. La glutamine 105

Lors de l'analyse structurale, nous avons pu observer que la glutamine 105 du motif G3 de PAB0955 était éloignée par rapport à celle des petites protéines G. Ceci est dû à la présence de la proline (au lieu d'une alanine) et à la localisation de ce motif à l'interface dimérique. Cette différence de position des glutamines du motif G3, fait que dans PAB0955 l'azote ND2 de ce résidu Q105 est trop éloigné pour stabiliser le  $\text{PO}_4^{2-}$  issu de l'hydrolyse du GTP. Nous avons donc supposé que contrairement aux petites protéines G (Marshall, 1993), la mutation de la Q105 en alanine n'affecterait pas l'hydrolyse du GTP.

### IV.1.2. La lysine 40

La position particulière de la lysine 40 dans la structure du GTP hydrolysé à 4°C, stabilisant le  $\text{PO}_4^{2-}$ , laisse penser que cet acide est important pour l'hydrolyse. Ce résidu basique, peut stabiliser les charges négatives et est conservé dans 50% des protéines de la famille des GTPases à boucle GPN. Nous pouvons donc supposer que ce résidu peut favoriser la réaction sans toutefois être essentiel. Deux mutations ont été faites sur la lysine 40 : K40A supposée diminuer l'activité de la GTPase et K40R pour conserver les charges positives. La chaîne latérale de l'arginine pourrait davantage compenser les charges négatives, si elle peut s'insérer dans le site nucléotidique.

### IV.1.3. L'asparagine 67

Lors de l'analyse structurale, nous avons observé que le N67 de PAB0955 se superposait à la Q61 de hRas. Nous supposons donc qu'elle joue le même rôle de stabilisation du  $\text{PO}_4^{2-}$  produit au cours de l'hydrolyse. Plusieurs fonctions ont été attribuées à cet glutamine 61 de hRas.

Tout d'abord l'activation d'une molécule d'eau qui activerait l'hydrolyse du GTP (Paduch et al., 2001) (Frech et al., 1994). Cette fonction a été ensuite écartée car le pKa de la glutamine n'est pas suffisant pour activer une molécule d'eau (Prive et al., 1992) (Langen et al., 1992). Cette glutamine aurait plutôt un rôle stabilisateur de l'état intermédiaire, en formant une liaison hydrogène par son ND2 avec le  $\text{PO}_4^{2-}$  (Prive et al., 1992).

La mutation de l'asparagine 67 va empêcher la stabilisation des produits de l'hydrolyse, ce qui va ralentir la réaction. Nous avons testé deux mutants de cette asparagine 67 : N67A et N67D. L'alanine ne va pas compenser les charges négatives, mais l'acide aspartique va apporter une charge négative qui devrait défavoriser davantage la réaction d'hydrolyse du GTP.

#### IV.1.4. L'aspartate 36

L'acide aminé D36 fixe le magnésium via une molécule d'eau. Sa mutation devrait déstabiliser la position du magnésium essentiel à l'hydrolyse du GTP. Cela devrait donc inhiber la réaction. Cette mutation devrait avoir le même effet que la présence d'EDTA. Pour être sûr du rôle clé de l'aspartate 36, nous l'avons muté en alanine et en asparagine. L'alanine étant plus petite, elle peut poser un problème au niveau de la structure de la protéine avec un site moins compact. Pour éviter ce problème le mutant D36N est aussi testé. Ces deux mutants ne devraient pas fixer le magnésium et le GTP ne devrait donc pas être hydrolysé.

### IV.2. Le comportement en solution des mutants

Nos collaborateurs du CEA de Marcoule ont produit les différents mutants proposés. Ils ont vérifié que les protéines mutantes soient correctement repliées, par dichroïsme circulaire. Les spectres de dichroïsme circulaire des protéines mutées sont identiques à ceux de la protéine sauvage. Cette analyse préliminaire est importante, car si les mutations affectent le repliement de la protéine, il est impossible de conclure sur le rôle des acides aminés mutés dans le mécanisme d'hydrolyse.

L'activité d'hydrolyse du GTP en solution, a été mesurée pour chaque mutant, en mesurant la quantité de GDP produit à 80°C en 30 minutes en solution (fig.IV.2).

#### IV.2.1. Le mutant Q105A en solution

La mutation Q105A n'a pas d'effet en solution sur l'hydrolyse du GTP, après 60 minutes le mutant produit autant de GDP que la protéine sauvage. Comme nous le supposions, ce résidu n'intervient pas dans l'hydrolyse du GTP. Cette confirmation est une différence supplémentaire avec les petites protéines G, dans lesquelles la mutation de ce résidu entraîne

une diminution de l'activité d'hydrolyse et peut conduire à l'observation d'état intermédiaire (Pasqualato and Cherfils, 2005).

Puisque ce mutant, dans PAB0955 n'affecte pas l'activité d'hydrolyse du GTP, il ne nous apportera pas d'élément sur le mécanisme d'hydrolyse, par conséquent nous n'avons pas déterminé sa structure.

#### [IV.2.2. Les mutants K40A et K40R en solution](#)

La mutation K40R ne change pas l'activité d'hydrolyse de la protéine en solution. En revanche le mutant K40A ne possède que 65% d'activité par rapport à la protéine sauvage. La lysine 40 intervient donc dans le mécanisme d'hydrolyse. La mutation de ce résidu n'abolit pas complètement l'activité, il n'est donc pas essentiel à l'hydrolyse. Seule la structure du mutant K40R a été étudiée déterminée, afin d'essayer de favoriser un intermédiaire réactionnel.

#### [IV.2.3. Les mutants N67A et N67D en solution](#)

La mutation N67A réduit de moitié l'activité de la protéine mutée par rapport à la protéine sauvage. Ceci confirme que ce résidu est important mais non essentiel pour l'hydrolyse du GTP, comme la lysine 40. La mutation de ce résidu en acide aspartique diminue de plus de 80% l'activité de la protéine. L'apport de charge négative dans le site nucléotidique défavorise la réaction d'hydrolyse, dont les produits sont eux aussi chargés négativement.

Les structures des deux mutants, N67A et N67D, ont été déterminées.

#### [IV.2.4. Les mutants D36A et D36N en solution](#)

Le résidu D36 a été muté en alanine et en asparagine, dans les deux cas l'activité est drastiquement réduite en solution. Une activité résiduelle de 8% par rapport à la protéine sauvage est observée pour le mutant D36A et de 17% pour le D36N en solution. Ce résidu D36 semble donc essentiel pour l'hydrolyse du GTP. L'effet de la mutation du D36 est comparable, en solution, à celui de l'EDTA, c'est-à-dire l'inhibition de l'hydrolyse du GTP.

Les structures des deux mutants, D36A et D36N ont été déterminées.

Mutant	% Activité résiduelle / activité de PAB0955 sauvage	Étude structurale
Q105A	102%	non
K40R	102%	oui
K40A	65%	non
N67D	18%	oui
N67A	51%	oui
D36N	17%	oui
D36A	8%	oui

Figure.IV.2. Tableau de l'activité spécifique d'hydrolyse du GTP des mutants de PAB095.

### IV.3. Structure et analyse des mutants

#### IV.3.1. La cristallisation des mutants

Pour les essais de cristallisation, les différents mutants ne se sont pas comportés de la même façon que la protéine sauvage. Certains nous ont permis d'obtenir rapidement des cristaux, D36A et D36N, un autre a cristallisé en un peu plus de temps que ne le fait la protéine sauvage, N67A, enfin les mutants N67D et K40R ont nécessité plus de recherche au niveau des conditions de cristallisation. Les essais de cristallisation ont tous été effectués en co-cristallisant du GDP et du GTP avec chaque protéine mutée à 18°C.

Les protéines mutées sur le résidu D36 ont été co-cristallisées avec du GTP et du GDP dans les mêmes conditions que la protéine sauvage (fig.IV.3.). Avec le mutant D36A, nous avons pu obtenir les structures de la protéine en complexe avec du GDP et du GTP. Pour le variant D36N, la structure en complexe avec du GTP a été obtenue. Par contre les cristaux obtenus en présence de GDP sont de forme irrégulière et diffractent à basse résolution (4 Å). L'intérêt de ces mutants est d'observer l'effet de la mutation à la fois sur la structure globale mais surtout les détails du site nucléotidique. Nous avons besoin d'une résolution plus haute que 3 Å. De ce fait, la structure du complexe D36N-GDP n'est pas utilisée pour notre étude.

<b>D36A</b> GTP	<b>D36A</b> GDP	<b>D36N</b> GTP
16% PEG 4K tri-Na citrate pH 5.6 à 0.1 M Acétate NH <sub>4</sub> à 0.2 M DTT à 20 mM D36A à 6.3 mg.ml <sup>-1</sup> GTP à 0.65 mM	12% PEG 4K tri-Na citrate pH 5.6 à 0.1 M Acétate NH <sub>4</sub> à 0.2 M DTT à 20 mM D36A à 6.3 mg.ml <sup>-1</sup> GDP à 0.65 mM	16% PEG 4K tri-Na citrate pH 5.6 à 0.1 M Acétate NH <sub>4</sub> à 0.2 M DTT à 20 mM D36N à 3.2 mg.ml <sup>-1</sup> GTP à 0.65 mM

*Figure.IV.3. Conditions de cristallisation des mutants de l'aspartate 36.*

Les protéines mutées sur le N67 et la K40 ont cristallisé plus difficilement. Les premiers essais en complexe avec les nucléotides, dans les mêmes conditions de cristallisation que celles utilisées pour la protéine sauvage n'ont, au départ, pas permis d'obtenir des cristaux. Il a fallu ajouter des additifs comme du NaCl ou du dioxane. Au final les conditions de cristallisations utilisées sont un mélange entre celles utilisées pour la protéine sauvage seule (dioxane) et en complexe avec les nucléotides (acétate d'ammonium) (figIV.4.). Nous n'avons pas pu obtenir de structure de la protéine mutée N67D co-cristallisée avec du GTP. Les cristaux obtenus contenaient au moins deux groupes d'espace différent et par conséquent ne pouvaient s'indexer.

<b>N67A</b> GTP	<b>N67A</b> GDP-WO <sub>4</sub>	<b>N67D</b> GDP	<b>K40R</b> GDP
19% PEG 4K tri-Na citrate pH 5.6 à 0.1 M Acétate NH <sub>4</sub> à 0.2 M <b>NaCl à 50 mM</b> DTT à 20 mM N67A à 5.3 mg.ml <sup>-1</sup> GTP à 0.65 mM	22% PEG 4K tri-Na citrate pH 5.6 à 0.1 M Acétate NH <sub>4</sub> à 0.2 M DTT à 20 mM N67A à 5.3 mg.ml <sup>-1</sup> GDP à 0.65 mM WO <sub>4</sub> Na à 1 mM	14% PEG 4K tri-Na citrate pH 5.6 à 0.1 M Acétate NH <sub>4</sub> à 0.2 M <b>NaCl à 100 mM</b> <b>3% dioxane</b> DTT à 20 mM N67D à 4.3 mg.ml <sup>-1</sup> GDP à 0.65 mM	18% PEG 4K tri-Na citrate pH 5.6 à 0.1 M Acétate NH <sub>4</sub> à 0.2 M <b>10% dioxane</b> DTT à 20 mM K40R à 4.8 mg.ml <sup>-1</sup> GDP à 0.65 mM

*Figure.IV.4. Conditions de cristallisation des mutants de l'asparagine 67 et de la lysine 40.*

### IV.3.2. Enregistrement cristallographique des mutants

Chaque cristal obtenu avec les différents mutants est cryo-protégé avec de l'huile de paratone avant d'être congelé dans l'azote liquide. Ce changement de cryo-protectant est dû à différents tests effectués au cours de nos manipulations sur les cristaux de la protéine sauvage. Les tâches de diffraction étant parfois légèrement ovales, la mosaïcité en général élevée pour nos cristaux, nous avons testé différents cryo-protectants en espérant améliorer leur qualité de diffraction. Parmi ceux testés, l'huile de paratone a permis d'obtenir des tâches plus fines qu'avec une concentration augmentée en PEG. Les jeux de données sur les différents cristaux ont été enregistrés à l'E.S.R.F., traités avec XDS et les différentes données cristallographiques sont notées dans le tableau de la figure.IV.5. pour les cristaux ayant permis d'obtenir une structure.

Nous avons obtenu au minimum une structure pour chacun des 5 mutants en complexe avec un nucléotide.



	<b>N67A</b>	<b>N67A</b>	<b>N67D</b>	<b>D36A</b>	<b>D36A</b>	<b>D36N</b>	<b>K40R</b>
	GTP	GDPWO <sub>4</sub>	GDP	GTP	GDP	GTP	GDP
Résolution (Å)	20-2.30 (2.4-2.3)	50-2.14 (2.27-2.14)	20-2.10 (2.2-2.1)	20-2.5 (2.6-2.5)	20-2.6 (2.7-2.6)	20-2.30 (2.4-2.3)	50-2.60 (2.7-2.6)
R <sub>sym</sub> (%)	7 (30)	5.3 (34.8)	8.5 (26.6)	6 (29.2)	6.8 (36.3)	9 (31.4)	6.7 (33.2)
Complétude (%)	95.8 (89.4)	98.7 (96.5)	97.7 (92.5)	96.4 (86.1)	95.9 (91.3)	98.2 (99.7)	97.2 (92.7)
I/σ(I)	13.89 (4.41)	31.37 (8.26)	17.57 (3.90)	17.97 (6.41)	11.72 (3.95)	14.71 (7.43)	19.28 (5.25)
Redondance	10	11.8	10	4.9	3.5	7.3	5.3
B (Å <sup>2</sup> )	47	35	42	61	64	37	50
paramètres de maille (Å)	a=b=60.1 c=116	a=b=60.3 c= 116.6	a=b=61 c= 117	a=b=61 c= 117	a=b=60 c= 117	a=b=60 c= 117	a=b=60.3 c= 116.6
λ (Å)	0.931	0.979	0.933	0.979	0.979	0.979	0.979
Ligne de lumière	ID14-eh3	BM30A	ID14-eh2	ID14-eh4	ID14-eh4	ID14-eh4	BM30A
Contenu de l'unité asymétrique	monomère	monomère	monomère	monomère	monomère	monomère	monomère
R	0.212	0.210	0.211	0.199	0.201	0.195	0.212
R <sub>free</sub>	0.272	0.294	0.288	0.297	0.331	0.276	0.323
liaisons (Å)	0.014	0.012	0.011	0.010	0.011	0.011	0.011
Angles (°)	1.491	1.478	1.373	1.405	1.466	1.402	1.416
Contenu de la P-loop	GDP Mg <sup>2+</sup> 1 H <sub>2</sub> O	GDP Mg <sup>2+</sup> 1 H <sub>2</sub> O	GDP Mg <sup>2+</sup> 2 H <sub>2</sub> O	GTP	GDP	GTP	GDP

*Figure.IV.5. Tableau des statistiques cristallographiques des mutants de PAB0955. Entre parenthèses les valeurs dans la dernière coquille de résolution.*

### IV.3.3. Analyse structurale des mutants

Tous les cristaux appartiennent au groupe d'espace  $P3_221$ , l'unité asymétrique contient donc un monomère. Chaque structure a été résolue par remplacement moléculaire avec le logiciel AMoRe, avec comme modèle la structure de MgGDP de la protéine sauvage (facteur de corrélation de l'ordre de 60%, facteur R de 40%). Les rmsd moyens obtenus par superposition

des structures des mutants entre elles et avec les structures de la protéine sauvage varient entre 0.2 et 0.6 Å (fig.IV.6.a.). Les structures des mutants entre elles sont également proches (fig.IV.6.b.). Les faibles rmsd montrent que le repliement global reste le même, les mutations effectuées n'affectent pas la structure de la protéine. Ceci signifie donc que les effets observés sur l'activité d'hydrolyse du GTP pour les protéines mutées ne sont pas dus à un changement de la structure de la protéine, mais sûrement à des modifications locales.

	<b>Native</b>	<b>GTP<math>\gamma</math>S</b>	<b>GTP<sup>EDTA</sup></b>	<b>MgGDP</b>	<b>GTP<sup>4°C</sup></b>	<b>GTP<sup>18°C</sup></b>
plus proche	D36A GDP 0,34 Å	N67D GDP 0,23 Å	D36A GDP 0,33 Å	N67D GDP 0,41 Å	D36A GDP 0,51 Å	N67D GDP 0,27 Å
plus loin	K40R GDP 0,41 Å	D36N GTP 0,36 Å	K40R GDP 0,38 Å	D36N GTP 0,56 Å	D36N GTP 0,62 Å	K40R GDP 0,38 Å

*Figure.IV.6.a. Tableau des statistiques de superposition des différentes structures des mutants avec celles de la protéine sauvage.*

	<b>N67A GTP</b>	<b>N67A GDP</b>	<b>N67D GDP</b>	<b>D36A GTP</b>	<b>D36A GDP</b>	<b>D36N GTP</b>	<b>K40R GDP</b>
<b>N67A GTP</b>		0,18 Å	0,25 Å	0,26 Å	0,26 Å	0,39 Å	0,26 Å
<b>N67A GDP</b>			0,25 Å	0,24 Å	0,26 Å	0,43 Å	0,24 Å
<b>N67D GDP</b>				0,27 Å	0,36 Å	0,37 Å	0,30 Å
<b>D36A GTP</b>					0,28 Å	0,40 Å	0,30 Å
<b>D36A GDP</b>						0,36 Å	0,36 Å
<b>D36N GTP</b>							0,46 Å

*Figure.IV.6.b. Tableau des statistiques de superposition des différentes structures des mutants.*

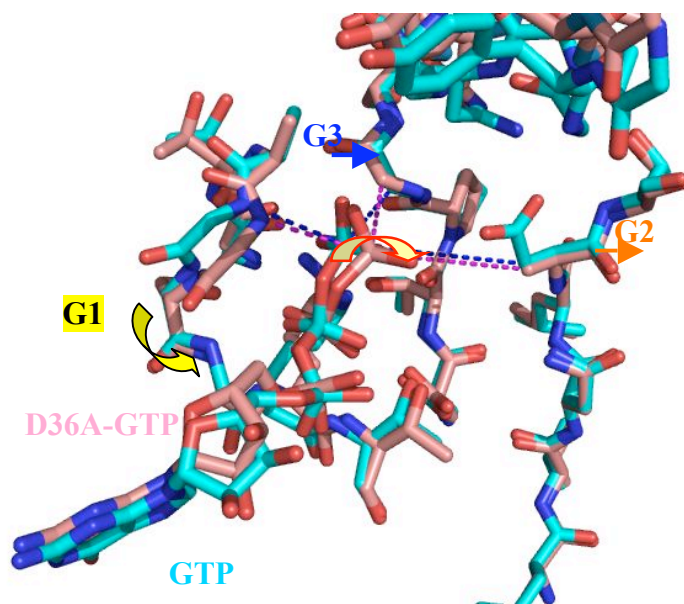
La superposition des structures a été faite sur le cœur d'un monomère de la protéine (154 C $\alpha$ ), comme pour les structures de la protéine sauvage. Parmi ces structures, nous avons pu observer des mutants qui n'hydrolysent pas le GTP (D36A et D36N) et d'autre l'hydrolysant, avec une vitesse plus faible que la protéine sauvage en solution (N67A, N67D et K40R).

#### IV.3.3.1. Analyse structurale des mutants D36A et D36N

Les protéines mutées sur l'aspartate 36 n'hydrolysent plus le GTP, ce qui nous a permis d'obtenir la structure des deux mutants D36A et D36N avec GTP.

##### *La structure du mutant D36A avec du GTP (D36A-GTP)*

Le C $\alpha$  de l'alanine 36 est situé au même endroit que celui de l'aspartate 36 dans la protéine sauvage. La position du phosphate- $\gamma$  dans la structure du mutant D36A en complexe avec du GTP (saumon) est proche de celle observée dans la structure GTP<sup>EDTA</sup> (cyan) de la protéine sauvage (fig.IV.7.a.). Les groupements phosphates dans la structure du mutant sont moins



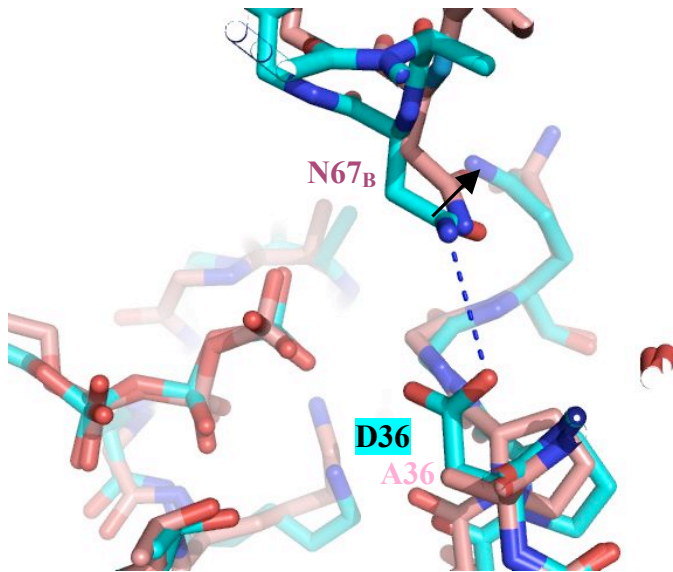
enfouit, et montrent un décalage de 1 Å (flèche) avec ceux de la structure GTP<sup>EDTA</sup> de la protéine sauvage. Ce déplacement des groupements phosphates est accompagné par celui des motifs G1, G2 et G3 (fig.IV.7.a.).

Ces mouvements concertés permettent de conserver les distances entre les différents motifs et le phosphate- $\gamma$  entre les deux structures (fig.IV.7.b.).

Figure.IV.7.a. Superposition de la structure GTP<sup>EDTA</sup> de la protéine sauvage (cyan) et de la structure GTP du mutant D36A (saumon).

Motifs	Liaison	GTP <sup>EDTA</sup>	D36A-GTP
Motif G1	G10 NH ---- P $\gamma$ GTP	3.6 Å	3.7 Å
Motif G2	G104 C $\alpha$ ---- P $\gamma$ GTP	4.3 Å	4.3 Å
Motif G3	D36 C $\alpha$ ----- P $\gamma$ GTP	7.5 Å	7.2 Å

Figure.IV.7.b. Tableau récapitulatif des liaisons conservées entre les structures GTP<sup>EDTA</sup> de la protéine sauvage (cyan) et de la structure GTP du mutant D36A (rose).



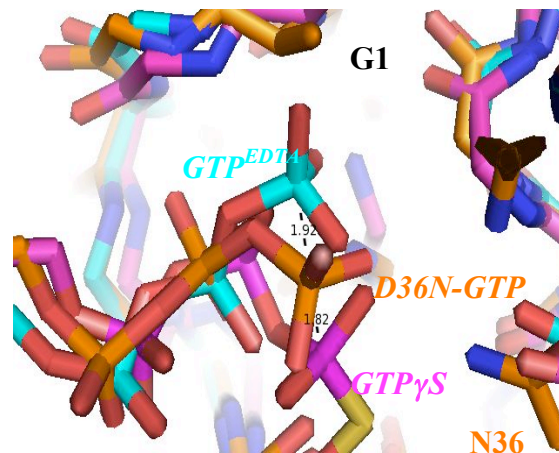
*Figure.IV.7.c.* Superposition de la structure  $GTP^{EDTA}$  de la protéine sauvage (cyan) et de la structure GTP du mutant D36A (saumon).

Le motif GPN est plus éloigné de 1.2 Å dans la structure du mutant D36A-GTP que dans celle  $GTP^{EDTA}$  de la protéine sauvage, laissant ainsi plus de place au phosphate- $\gamma$ , qui est plus exposé au solvant. L'absence de magnésium dû à la mutation de l'aspartique 36 en alanine, provoque une situation comparable à celle observée dans la structure de  $GTP^{EDTA}$ . Il n'y a pas d'hydrolyse du GTP. De plus l'absence de l'acide aspartique change les interactions avec l'asparagine 67 de la boucle GPN qui recule (fig.IV.7.c.) par

rapport au site nucléotidique.

#### **La structure du mutant D36N avec du GTP (D36N-GTP)**

La position du phosphate- $\gamma$  est surprenante dans la structure du mutant D36N complexée avec du GTP. Elle est intermédiaire entre celle observée dans les structures de la protéine sauvage  $GTP^{EDTA}$  et  $GTP\gamma S$ , avec une distance de presque 2Å entre les phosphates- $\gamma$  (fig.IV.8.a.). La P-loop ne suit pas le déplacement du phosphate- $\gamma$  dans le mutant D36N. Le motif G1 du mutant D36N a la même conformation que celui de la structure  $GTP^{EDTA}$ .



*Figure.IV.8.a.* Superposition des structures  $GTP^{EDTA}$  (cyan) et  $GTP\gamma S$  (rose) de la protéine sauvage et de la structure GTP du mutant D36N (orange).

De ce fait la glycine 10 engage une liaison avec un des oxygènes du phosphate- $\beta$  (2.88 Å) au lieu de faire une liaison avec le phosphate- $\gamma$  trop éloigné (5.33 Å). Le motif G3 ne suit pas non plus le déplacement du phosphate- $\gamma$ . La distance entre le C $\alpha$  de la glycine 104 et le phosphate- $\gamma$  est de 4.7 Å.

Le motif GPN lui se rapproche de 1 Å et en ferme le site nucléotidique par une liaison

entre l'Asn67<sub>B</sub> et l'Asn36<sub>A</sub>. De plus, l'azote ND2 de la chaîne latérale de l'asparagine 36 (acide aminé muté) forme une liaison hydrogène avec un oxygène du phosphate- $\gamma$ , comme la fait l'asparagine 67 (fig.IV.8.c.).

La distance entre le NH de la glycine 10 de la P-loop et l'atome d'oxygène liant les groupements phosphates  $\beta$  et  $\gamma$  est de 3.8 Å. Cette distance est proche de celle observée dans la structure GTP<sup>EDTA</sup> de la protéine native. Comme nous l'avons indiqué dans le paragraphe III.1.5. sur le mécanisme

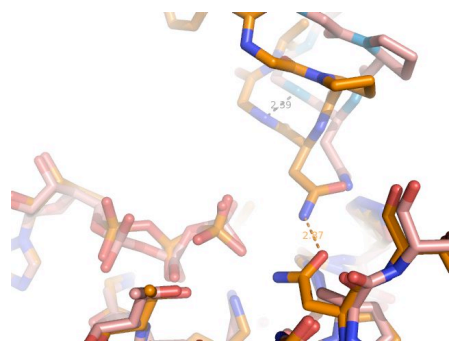


Figure.IV.8.d. Superposition des structures des mutants D36A (saumon) et D36N (orange) en complexe avec du GTP.

enzymatique, cette liaison pourrait déclencher l'hydrolyse du GTP. La position du GTP dans le mutant D36N nous montre peut-être une conformation possible du GTP. Cette position peut également montrer le chemin pris par le phosphate- $\gamma$  après hydrolyse du GTP, avec les acides aminés impliqués dans sa stabilisation, comme l'asparagine 67. Entre les structures D36N-GTP et D36A-GTP, la présence de l'asparagine (N36) provoque de nouvelles liaisons fermant le site nucléotidique et attire le phosphate- $\gamma$ , ce qu'il lui donne une position moins enfouie.

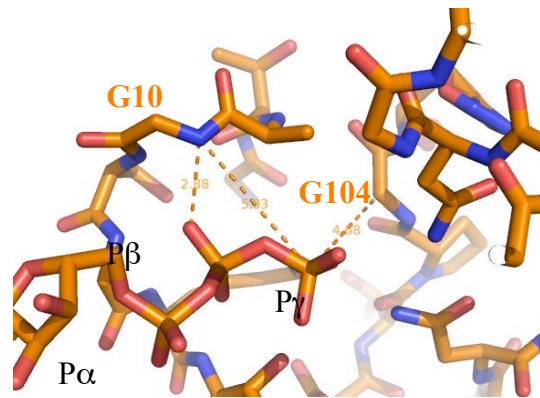


Figure.IV.8.b. Structure du mutant D36N en complexe avec du GTP.

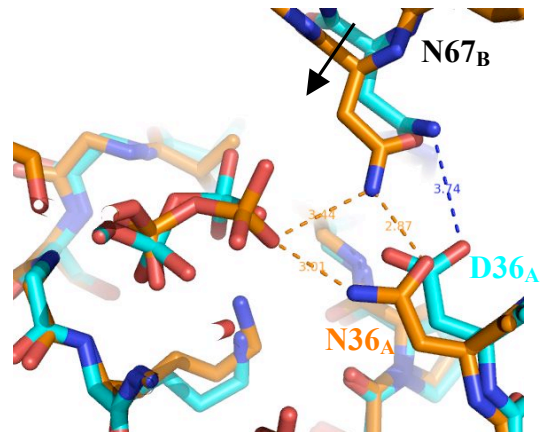
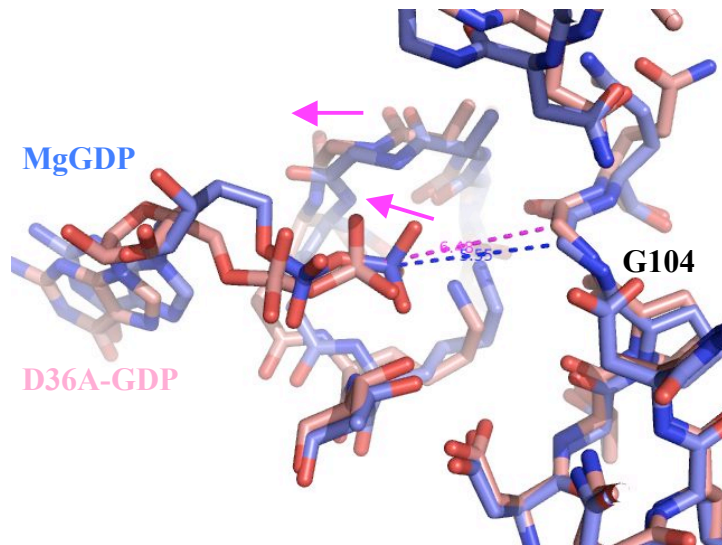


Figure.IV.8.c. Superposition de la structure GTP<sup>EDTA</sup> (cyan) de la protéine sauvage et de la structure GTP du mutant D36N (orange).



### La structure du mutant D36A avec du GDP (D36A-GDP)

La comparaison de la structure D36A en complexe avec du GDP par rapport à celle de la protéine sauvage avec MgGDP montre que les phosphates sont moins enfouis dans le mutant



*Figure.IV.9.* Superposition des structures MgGDP de la protéine sauvage (bleue) et GDP du mutant D36A (saumon).

que dans la sauvage. Il y a un décalage de 1 Å (flèche) entre les phosphates-β de ces deux structures.

Les groupements phosphates du mutant D36A se superposent avec ceux du GTP de la structure  $GTP^{EDTA}$ . Le motif G3 ne se déplace pas entre les structures D36A-GDP et MgGDP. Par conséquent la distance entre le Cα de la glycine 104 et le phosphate-β varie donc de 5.55

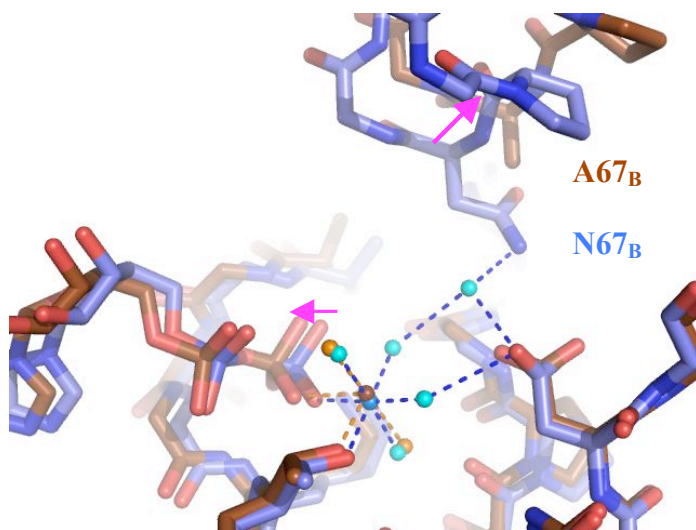
Å à 6.48 Å dans la structure D36A-GDP (décalage de 1 Å) (fig.IV.9). Le motif G2 contenant la mutation reste dans la même conformation que dans la protéine sauvage en complexe avec MgGDP. Les motifs G1, G4 et G5 du mutant D36A suivent le mouvement de recul du GDP par rapport au site nucléotidique.

La structure du mutant D36A-GDP montre qu'en absence de magnésium les groupements phosphates ne sont pas enfouis comme ils le sont dans la structure MgGDP de la protéine sauvage. Dans le paragraphe sur les différentes étapes du mécanisme enzymatique (III.1.5.), nous avons émis l'hypothèse d'un mouvement du nucléotide au cours de l'hydrolyse. La position du GDP, dans le mutant D36A, reflète peut-être ce qu'il se produit après l'hydrolyse du GTP. Lorsque le magnésium se retire du site, une fois le GTP hydrolysé, le GDP reculerait pour sortir également du site.

### IV.3.3.2. Analyse structurale des mutants N67A, N67D et K40R

#### **Les structures du mutant N67A avec du GTP et du GDP (N67A-GTP et N67A-GDP)**

Avec le mutant N67A, nous avons pu obtenir des cristaux en présence de GDP et de GTP. Les structures sont très similaires avec un rmsd de 0,18 Å (fig.IV.6.b). Nous observons du GDP avec magnésium dans la P-loop dans les structures des deux types de cristaux N67A-GTP et N67A-GDP. Ces résultats reflètent bien l'activité en solution de ce mutant, qui est de 51% par rapport à l'activité de la protéine sauvage. L'activité est plus lente, mais ces cristaux se forment plus lentement, le mutant a donc le temps d'hydrolyser complètement le GTP dans les gouttes. Par comparaison avec la structure MgGDP (bleu), les groupements phosphates



*Figure.IV.10. Superposition des structures MgGDP de la protéine sauvage (bleue) avec celle du mutant N67A co-cristallisé avec du GTP (marron).*

*Les molécules d'eau cyan et le magnésium bleu appartiennent à la structure MgGDP. Les molécules d'eau orange et le magnésium marron appartiennent à la structure N67A-GTP.*

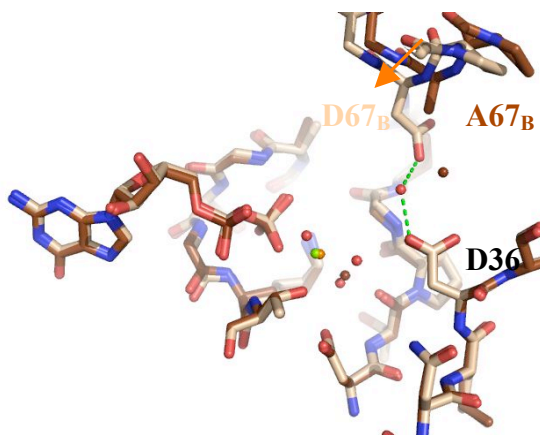
(fig.IV.10.). Dans la structure du mutant N67A obtenue avec du GTP, le magnésium est coordonné par deux molécules d'eau, dans la structure obtenue avec du GDP il n'y en a qu'une. La perte de molécules d'eau autour du magnésium peut le déstabiliser, ce qui expliquerait également la baisse d'activité de ce mutant, en plus du manque de stabilité de l'état intermédiaire dû à la mutation de l'asparagine 67.

dans le mutant N67A (marron) sont moins enfouis, avec un décalage de 0.62 Å. La variation la plus grande est observée pour la boucle GPN, qui dans le mutant N67A recule de 2.29 Å par rapport à celle de la protéine sauvage en complexe avec MgGDP. De plus, la couche de solvation autour du magnésium change dans les structures de ces mutants. La molécule d'eau liant l'Asn67<sub>B</sub>, l'Asp36<sub>A</sub> et une autre molécule d'eau (en interaction avec le magnésium) n'est pas présente dans ces structures de mutant



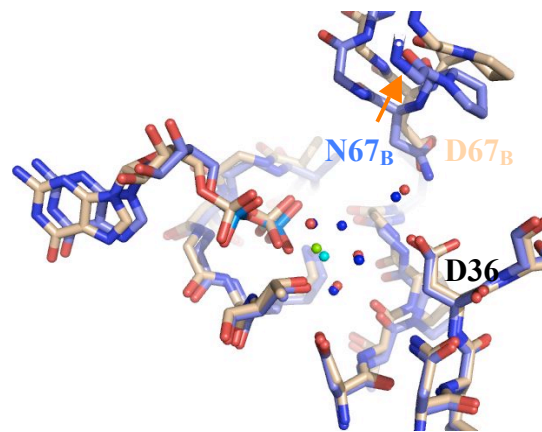
### La structure du mutant N67D avec du GDP (N67D-GDP)

Les cristaux obtenus avec le mutant N67D en présence de GTP ne nous ont pas permis d'obtenir de structure. La structure obtenue en présence de GDP, contient dans le site nucléotidique du GDP et un magnésium coordonné par 3 molécules d'eau. La structure ressemble beaucoup, pour les motifs G, à celles obtenues avec le mutant N67A (fig.IV.11.a.). La boucle GPN est cependant plus proche du site nucléotidique dans le mutant N67D. La boucle GPN est décalée de 1.4 Å par rapport à celle de la structure MgGDP de la protéine sauvage (fig.IV.11.b.). L'acide aspartique 67<sub>B</sub> est lié via une molécule à l'aspartique 36<sub>A</sub>, rapprochant la boucle GPN par à celle observée dans le mutant N67A ne contenant pas cette molécule d'eau. Cette liaison entre les deux acides aspartiques (D36 et D67) change la position du D36, décalant légèrement la position du magnésium. Ce mouvement peut également déstabiliser le magnésium et diminuer l'activité de la protéine.



*Figure.IV.11.a.* Superposition des structures de N67D-GDP (beige) avec N67A-GDP (marron).

*Les molécules d'eau et les ions magnésium sont représentés par des sphères rouge et vert respectivement pour la structure N67D-GDP, puis marron et orange respectivement pour N67A-GDP.*



*Figure.IV.11.b.* Superposition des structures de N67D-GDP (beige) avec MgGDP de la protéine sauvage (bleue).

*Les molécules d'eau et les ions magnésium sont représentés par des sphères rouge et vert respectivement pour la structure N67D-GDP, puis bleues et cyan respectivement pour MgGDP.*

### La structure du mutant K40R avec du GDP (K40R-GDP)

Avec le mutant K40R, nous espérons pouvoir piéger un état intermédiaire mieux stabilisé par le groupement guanidinium de l'arginine qu'il ne l'est par la lysine. Malheureusement, ce mutant cristallise difficilement et les quelques cristaux obtenus en présence de GTP diffractent très mal. Nous n'avons donc pas obtenu de structure d'état intermédiaire de la

réaction, ni pu observer l'hydrolyse du GTP avec le mutant K40R. Dans la structure du mutant K40R en présence de GDP, la chaîne latérale de l'arginine n'a pas pu être modélisée (densité électronique faible).

Ces mutants ont confirmé les hypothèses que nous avons émises sur les différents rôles des acides aminés conservés et spécifiques des GTPases à boucle GPN. En effet, il est évident que l'acide aspartique 36, qui stabilise le magnésium, est essentiel à l'activité hydrolase de PAB0955. Les acides aminés N67 et K40 ont un rôle de stabilisation de l'état intermédiaire produit au cours de l'hydrolyse du GTP. La présence d'acides aminés stabilisant l'état intermédiaire laisse penser que PAB0955 n'aurait pas besoin de partenaire pour activer l'hydrolyse du GTP (type GAP). Il est possible que son activité soit augmentée par la présence d'un facteur d'échange (type GEF) permettant de remplacement le GDP par du GTP.

CHAPITRE V

RECHERCHE DE

PARTENAIRES



La résolution et l'analyse des différentes structures des protéines PAB0955 sauvage et mutantes ne nous ont pas apporté d'informations précises sur la fonction biologique de cette protéine. Le fait que PAB0955 représente une nouvelle famille de GTPase explique sans doute le fait que nous n'ayons pas, par sa structure, d'information sur sa fonction biologique. La présence du motif HEAT nous incite tout de même à penser que cette protéine est impliquée dans une interaction protéine-protéine, peut-être dans une voie de transport de protéines partenaires. Afin de déterminer l'existence de partenaires de PAB0955, des essais de pull-down et de Résonance de plasmon de surface (SPR) ont été réalisées par l'équipe de Didier Flament de l'IFREMER de Brest.

## V.1. Les protéines de *Pyrococcus abyssi*

### V.1.1. Expérience de pull-down

La technique de pull-down est utilisée pour rechercher un partenaire d'une protéine d'intérêt avec à partir par exemple d'un extrait cellulaire total ou d'une solution de protéines purifiées. Lorsque l'expérience se fait à partir d'un extrait cellulaire, il faut que le ou les partenaires soient en quantité suffisante pour être détectés et analysés.

Le principe de la technique de pull-down consiste à fixer la protéine d'intérêt sur une colonne ou sur des billes, puis faire passer un extrait cellulaire (sur la colonne ou les billes) afin de piéger les partenaires de cette protéine.

Tout d'abord il faut préparer un extrait cellulaire de *Pyrococcus abyssi* afin d'isoler les protéines solubles contenues dans l'archaea. Parmi ces protéines il y a en a une ou plusieurs qui peuvent être les partenaires de PAB0955. Les cellules en culture de *Pyrococcus abyssi* sont cassées afin de libérer les protéines solubles contenues dans les cellules. Les protéines issues de cet extrait cellulaire ne contiennent pas de protéines membranaires.

Puis il faut préparer la protéine d'intérêt. La protéine PAB0955 recombinante contenant une étiquette 6 histidines en N<sub>ter</sub> ceci permet son accroche sur des billes magnétiques contenant du

Cobalt. La protéine est incubée 20 minutes à 25°C avec ces billes, puis lavée 3 fois afin d'éviter les interactions non-spécifiques.

Les billes ainsi couplées à PAB0955 sont incubées avec un extrait cellulaire de *Pyrococcus abyssi* pendant 2 heures à 4°C. Après plusieurs lavages des billes, les protéines sont éluées par un tampon Tris à pH 8.8 à 1 M, 2% SDS, 5%  $\beta$ -mercaptoéthanol, 8% sucrose et 0.1 % de bleu de bromophénol. Cet éluât protéique est ensuite séparé par électrophorèse sur gel d'acrylamide en condition dénaturante. Les différentes bandes issues de l'éluât sont ensuite analysées par spectrométrie de masse. Seules deux protéines ont pu être identifiées : la sous unité B de la topoisomérase VI et la protéine DnaG, d'autres partenaires potentiels sont visibles sur le gel (fig.V.1.) mais n'ont pu être analysés par spectrométrie de masse.

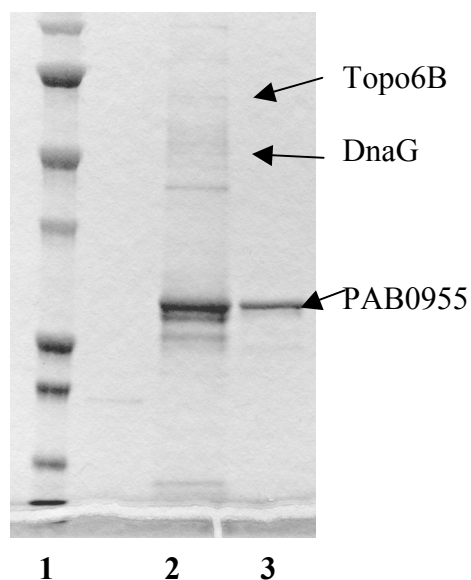


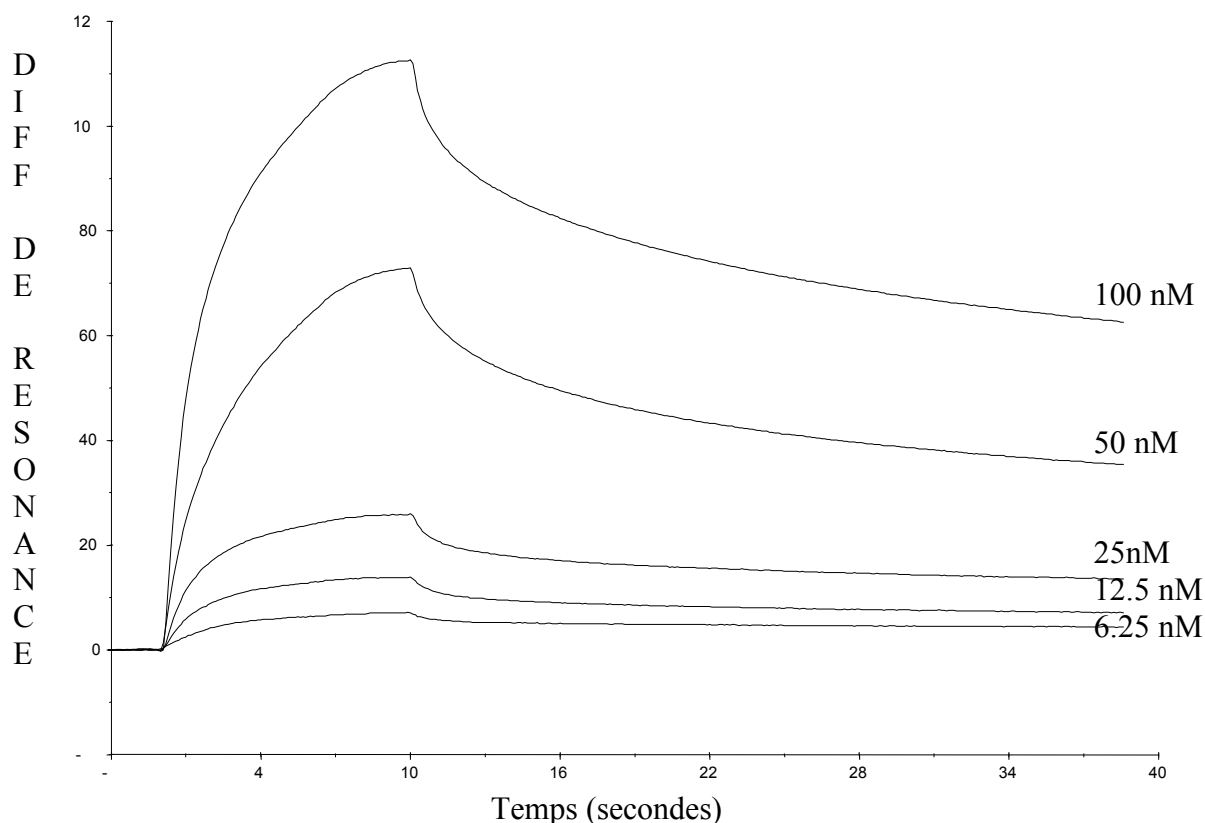
Figure.V.1. Gel d'électrophorèse à 12% d'acrylamide en condition dénaturante.

1 : marqueur de poids moléculaire  
2 : éluât du pull-down  
3 : PAB0955 seule

### [V.1.2. Résonance de plasmon de surface](#)

La topoisomérase VI de *Sulfolobus shibatae* étant disponible à l'IFREMER de Brest, et présentant 35% d'identité avec celle de *Pyrococcus abyssi*, son interaction avec PAB0955 a été testée par résonance de plasmon de surface (SPR). La protéine PAB0955 recombinante est immobilisée sur la membrane de la SPR et une solution contenant la topoisomérase VI passe sur cette membrane. La topoisomérase VI est un hétérotétramère de deux sous-unités A et deux sous-unités B. L'interaction entre la topoisomérase VI et PAB0955 est confirmée par

SPR (fig.V.2.). La constante de dissociation  $K_d$  entre les deux protéines qui est d'environ  $5.10^{-10}$  M. Seule la sous-unité B qui a été détectée par pull-down, mais en SPR la sous-unité B seule n'interagit pas avec PAB0955, la topoisomérase VI doit être entière pour interagir avec notre GTPase, suggérant une possible dégradation de la sous-unité A au moment de l'éluion.



*Figure.V.2. Courbe de SPR entre la topoisomérase VI et PAB0955 à différentes concentrations.*

La protéine DnaG n'étant pas disponible, elle n'a pas pu être testée par SPR.

### [V.1.3. Le rôle de ces partenaires](#)

#### [V.1.3.1. La topoisomérase VI](#)

Les topoisomérases sont des enzymes qui régulent le niveau de surenroulement des molécules d'ADN. Pour altérer le nombre de liaisons de l'ADN, ces enzymes doivent d'abord casser l'un des deux brins ou les deux à la fois. Cette cassure est réalisée grâce à un résidu tyrosine



de la topoisomérase qui attaque un phosphate du squelette de l'ADN, provoquant une liaison covalente (phosphotyrosine) temporaire entre l'enzyme et l'ADN. Il existe deux classes de topoisomérases, type I et type II. Les topoisomérases de type I cassent un brin d'ADN. Les topoisomérases de type II nécessitent l'hydrolyse de l'ATP pour casser l'ADN double brin, et transfèrent un autre segment de double brin à travers la cassure. Ces enzymes sont présentes dans tous les processus impliquant l'ADN et sont essentielles pour la fin de la réplication (Champoux, 2001).

Les topoisomérases de type II peuvent être divisées en deux classes : les IIA et les IIB (Bergerat et al., 1997). Les topoisomérases de type IIA sont les plus nombreuses, elles se retrouvent dans certains bactériophages, virus et archaées ainsi que chez tous les eucaryotes et toutes les bactéries. Elles sont homologues en séquence, mais leur organisation oligomérique peut varier. Les topoisomérases de type IIA organisent leur fonction catalytique de façon modulaire, avec un domaine ATPase situé dans la région B (en  $N_{\text{ter}}$ ), les activités de liaison et de cassure de l'ADN sont situées sur la région A (en  $C_{\text{ter}}$ ) (Corbett and Berger, 2003). Les topoisomérases de type IIB sont moins nombreuses que celles de type IIA, la topoisomérase VI est la seule représentante de cette famille (Champoux, 2001). Cette topoisomérase VI a été identifiée chez les archaées et existe également chez les plantes (Hartung and Puchta, 2001).

La topoisomérase VI utilisée pour les expériences de SPR est celle de *Sulfolobus shibatae*, c'est un hétérotétramère composé de deux sous-unités A (de 45 kDa chacune) et de deux sous-unités B (de 60 kDa chacune). La sous-unité B porte la fonction de fixation et d'hydrolyse de l'ATP. L'hydrolyse de l'ATP induit un changement conformationnel des sous-unités A qui sont liées avec les B. Les sous-unités A interagissent alors avec l'ADN, et leur résidu tyrosine se retrouve bien placé pour permettre la cassure du double brin d'ADN (fig.V.3.). La sous-unité A contient un domaine toprim (topoisomérase-primase) requis pour le clivage de l'ADN (Aravind et al., 1998). Ce domaine toprim contient une centaine de résidus avec deux motifs conservés, un central avec un glutamate conservé et un autre avec deux aspartates conservés (DXD).

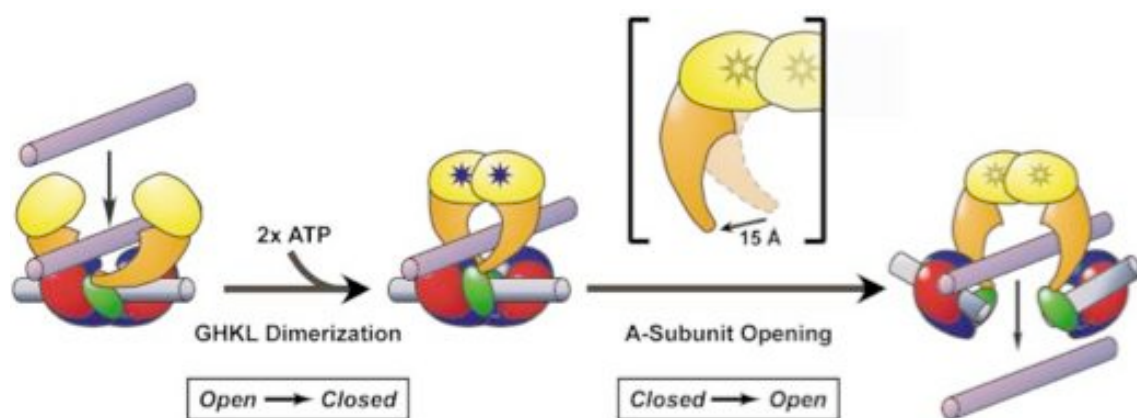


Figure.V.3. issue de l'article de K.D. Corbett and J.M.Berger (Corbett and Berger, 2003)  
En jaune sous-unité B, en rouge, bleu et vert les sous-unités A

### V.1.3.2. La DnaG primase

La primase initie la réplication de l'ADN par la synthèse des amorces d'oligonucléotides nécessaire à l'élongation de l'ADN par la polymérase. Chez les archaea, l'homologue de la DnaG primase bactérienne contient le domaine toprim très conservés dans les deux règnes, mais le domaine impliqué dans la fixation de l'ADN n'est pas identifié. La fonction cellulaire des DnaG-like protéines chez les archaea n'a pas encore été clairement identifiée. Seule la DnaG-like de *Sulfolobus solfataricus* a été retrouvée associée à l'exosome, qui est la machinerie de dégradation de l'ARN. Chez les bactéries, les DnaG-primases sont impliquées dans la réplication de l'ADN. Nous ne pouvons pas exclure que chez les archaea ces protéines jouent également un rôle au niveau de l'ADN.

Le résultat de l'interaction entre la topoisomérase VI et la DnaG primase, impliquées dans la réplication de l'ADN, et PAB0955, est en accord avec notre hypothèse de départ basée sur le contexte génomique de PAB0955. En effet comme indiqué dans l'introduction nous nous sommes intéressés à cette GTPase car son contexte génomique nous laisse supposer qu'elle est impliquée dans la réplication de l'ADN et/ou la division cellulaire. Cette hypothèse est en accord avec le résultat de l'expérience de pull-down.

Nous avons identifié une interaction avec ces deux protéines, mais pour l'instant l'implication au niveau fonctionnel reste inconnue.

Afin de poursuivre la recherche de partenaires de PAB0955, et comprendre sa fonction biologique, nous avons testé par SPR la possible interaction de notre GTPase avec des protéines connues pour être impliquées dans la réplication de l'ADN.

## V.2. Les protéines impliquées dans la réplication de l'ADN

Plusieurs protéines issues de *Pyrococcus abyssi* étaient disponibles dans le laboratoire de nos collaborateurs de l'IFREMER de Brest. Ces protéines sont homologues de protéines eucaryotes, connues pour être impliquées dans la réplication de l'ADN. Les protéines testées en SPR sont : les ADN polymérases des familles B et D (Gueguen et al., 2001), le complexe RF-C (Henneke et al., 2002), le PCNA (proliferating cell nuclear antigen) (Henneke et al., 2002), l'endonucléase Flap 1 et la RNase HII.

Comme pour la technique de SPR entre PAB0955 et la topoisomérase VI, notre GTPase est fixée à la membrane et un flux de solution contenant les différentes protéines est passé sur la membrane.

Seule la protéine RF-C présente une interaction spécifique avec PAB0955. Le facteur de réplication C est une ATPase qui a pour fonction de charger le PCNA sur l'ADN. Une fois le PCNA sur l'ADN des polymérases se fixent pour poursuivre la synthèse de l'ADN et le RF-C est libéré (Henneke et al., 2002).

La constante d'interaction entre PAB0955 et RF-C est de l'ordre du nanomolaire, mais la valeur exacte de cette constante n'a pu être déterminée. L'interaction entre ces deux protéines ne répond pas à un modèle mathématique simple de 1 pour 1, sûrement à cause des natures multimériques des deux protéines. En effet RF-C est composé par 4 petites sous-unités et une grande, et PAB0955 est dimérique. Cette interaction a été confirmée par pontage chimique au glutaraldéhyde par l'équipe de Jean Armengaud du CEA de Marcoule.

Après analyse sur gel d'acrylamide en condition dénaturante, ils ont pu observer une bande à 70 kDa. Après transfert sur membrane de nitrocellulose et immunorévélation par des anticorps anti-PAB0955 et anti-RF-C, ils ont pu déterminer que cette bande de 70 kDa correspondait à un hétérodimère PAB0955/RF-C.

Un monomère de PAB0955 s'associe donc avec un monomère de RF-C. L'interaction entre les deux protéines doit se faire près de l'interface dimérique de PAB0955, empêchant le pontage du dimère de PAB0955 en présence de RF-C.

Des essais de cristallisation ont été entrepris entre la petite sous-unité de RF-C et PAB0955. Malheureusement aucun cristal de protéine n'a été obtenu. La cristallisation étant sûrement gênée par le caractère hétérogène de la solution de RF-C. En effet en absence de la grande sous-unité de RF-C, les petites sous-unités sont capables de former des multimères dont le nombre de sous-unités engagées diffère.

L'interaction entre RF-C et PAB0955 ne change pas les fonctions connues des deux protéines. En effet pour RF-C ni son activité de charge du PCNA sur l'ADN, ni son activité ATPase ne sont perturbées par la présence de PAB0955. Réciproquement l'activité GTPase de PAB0955 reste inchangée en présence de RF-C.

### V.3. Test de phosphorylation

L'autophosphorylation de GTPases a déjà été observée pour des protéines comme Era (Sood et al., 1994), Obg (Welsh et al., 1994) et également sur le mutant A59T de la petite protéine G Ras humaine (John et al., 1988). Cependant aucune implication biologique de cette autophosphorylation n'a été démontrée. Nous avons voulu tester l'hypothèse de cette réaction avec PAB0955.

Le test d'autophosphorylation de PAB0955 a été effectué par Arnaud Hecker de l'équipe de Patrick Forterre. L'expérience a été réalisée avec des composés radioactifs afin d'identifier par autoradiographie l'éventuelle fixation du phosphate- $\gamma$ . Les molécules utilisées sont  $[\gamma^{32}\text{P}]\text{-GTP}$  ;  $[\gamma^{32}\text{P}]\text{-ATP}$  et  $^{32}\text{Pi}$ . Nous avons testé la présence d'autophosphorylation sur la protéine PAB0955 sauvage seule, en présence de RF-C, de topoisomérase VI et également le mutant D36A de PAB0955. Après incubation des différentes protéines avec chacune des molécules à 80°C, une analyse a été réalisée par électrophorèse à deux dimensions après traitement à la phosphatase alcaline.

Un faible signal de radioactivité a été observé pour la protéine sauvage seule ou complexée, et sur la protéine mutée D36A, ceci uniquement en présence de  $[\gamma^{32}\text{P}]\text{-GTP}$ . Aucun signal n'a été détecté en présence d'EDTA, de  $[\gamma^{32}\text{P}]\text{-ATP}$  et  $^{32}\text{Pi}$  sur PAB0955 sauvage ou mutante. Et aucun signal n'est détecté sur RF-C ou sur la topoisomérase VI.

La nature des résidus autophosphorylés a été identifiée, elle n'est pas spécifique, puisque trois type d'acides aminés sont concernés par l'autophosphorylation : sérine, thréonine et tyrosine.

De plus le mutant D36A qui présente une d'activité résiduelle en solution (8%), s'autophosphoryle autant que la protéine sauvage.

Comme pour les GTPases Obg, Era et le mutant A59T de hRas, une autophosphorylation est détectée, sans pour autant avoir une signification biologique. Dans notre cas, le manque de spécificité et le faible signal confirment le caractère probablement parasite de cette autophosphorylation.

CONCLUSION

ET

PERSPECTIVES





## *Conclusion et perspectives du projet PAB0955*

Grâce à nos travaux cristallographiques, nous avons pu atteindre certains des objectifs du projet. Le premier était bien évidemment de déterminer la structure de PAB0955, puis de l'analyser afin d'en déterminer la fonction biologique ou du moins de découvrir dans quel mécanisme biologique cette GTPase est impliquée.

La détermination structurale de cette protéine réalisée au début du projet, nous a rapidement permis de tester des complexes avec les différents nucléotides. L'analyse structurale de toutes ces structures nous a permis d'identifier qu'il s'agissait d'une nouvelle famille de GTPase, d'où le fait que nous n'ayons pas identifié de protéine avec une identité suffisante pour effectuer un remplacement moléculaire. Sur la base de notre analyse structurale et de la conservation des résidus à travers les différentes protéines identifiées comme étant des homologues de PAB0955, nous avons pu proposer un rôle pour certains acides aminés. L'analyse en solution, et la détermination structurale des protéines mutées sur ces résidus ont confirmé les différentes hypothèses quant à leur fonction. Ce qui a permis, entre autre de nommer cette nouvelle famille de protéine les GPN loop GTPases, rapport à la boucle GPN à l'interface dimérique strictement conservée en séquence et impliquée dans l'hydrolyse du GTP. Ces mutants ont permis de mettre en évidence les principaux acides aminés et motifs caractérisant cette nouvelle famille de GTPases. Les différents résultats structuraux obtenus avec PAB0955 seront sûrement transposables aux autres protéines de cette famille, notamment aux protéines humaines.

Ce projet fut l'occasion de mieux connaître les GTPases, dont la découverte d'une nouvelle famille nous a amené à étudier en détail les autres familles de GTPases. La découverte d'une nouvelle famille de GTPase ouvre la perspective de nouvelles fonctions pour ces protéines bien conservées au cours de l'évolution. L'implication potentielle de PAB0955 dans la réplication de l'ADN renforce l'idée que les GTPases sont des enzymes essentielles impliquées, comme le sont les petites protéines G, dans des mécanismes vitaux pour les cellules.

Les structures de GTPase avec un brin- $\beta$ 2 parallèle et un domaine G, mis à part PAB0955, se retrouvent uniquement dans les protéines SRP et leur récepteur. Mais ces protéines reste relativement éloignées de PAB0955 ; il est donc difficile d'en tirer des conclusions sur le plan biologique.

L'analyse structurale avec les SRP montre que PAB0955 est dans une conformation active pour l'hydrolyse du GTP. Ceci nous laisse supposer que PAB0955 n'a pas besoin de protéine

## *Conclusion et perspectives du projet PAB0955*

activatrice pour l'hydrolyse du GTP comme les GAP pour les petites protéines G. Notre GTPase semble se suffire pour hydrolyser seule le GTP.

Avec les structures, il nous est difficile d'apporter des éléments fonctionnels sur PAB0955, car c'est une nouvelle famille et donc il n'y a pas de point de comparaison avec d'autres protéines. L'apport structural nous a tout de même permis d'étudier et de dévoiler le mécanisme enzymatique de cette GTPase.

Les expériences de pull-down et de SPR ont été entreprises par nos collaborateurs, afin de comprendre la fonction de cette protéine. Les résultats obtenus montrent une possible implication de PAB0955 dans des mécanismes au niveau de l'ADN. Ceci confirme nos hypothèses de départ basées sur le contexte génomique de PAB0955. Cependant il faut encore comprendre les rôles biologiques de ces interactions.

Dans ce but, il faut confirmer l'interaction entre PAB0955 et DnaG, identifier les autres protéines éluées après pull-down. Il faudra tester par exemple l'influence de PAB0955 sur l'activité ATPase de la topoisomérase VI et sur sa fixation à l'ADN.

Pour ce qui est de l'étude structurale des complexes, les essais de co-cristallisation entre RF-C et PAB0955 n'ont pas permis d'obtenir de cristaux de protéines. La topoisomérase VI de *Pyrococcus abyssi* n'est pas disponible. De plus le laboratoire d'Herman Van Tilbeurg qui travaille sur la topoisomérase VI de *Sulfolobus shibatae* est intéressé par ce partenaire de la topoisomérase VI qu'est PAB0955.

Quelle que soit la fonction exacte de ces protéines GPN loop GTPases, il semble d'après les études chez la levure *Saccharomyces cerevisiae* qu'elles soient vitales. Les études génétiques sur les cellules d'archaea connaissent encore des difficultés, donc la délétion du gène de PAB0955 chez *Pyrococcus abyssi* n'a pu être faite, il est probable que le phénotype soit également létal. Il n'est donc pas étonnant que les partenaires de notre GTPase soient impliqués dans des mécanismes vitaux.

Ce qui est intéressant avec les trois partenaires identifiés actuellement pour PAB0955, c'est leur répartition dans les différents règnes. Ainsi RF-C est présent chez les archaea et les eucaryotes, la topoisomérase VI est seulement présente chez les archaea, les plantes et les

algues rouges, et DnaG est présente chez les bactéries et les archaea avec des fonctions différentes.

Les partenaires de XAB1, un des homologues humains de PAB0955, sont des protéines qui n'ont pas d'homologue chez les archaea. Il semble donc qu'au cours de l'évolution les partenaires moléculaires des GPN loop GTPases se soient différenciés. D'un côté ceci contraste avec la conservation de séquence, et probablement de structure, des homologues de PAB0955 dont aucun homologue n'a été identifié dans les 311 génomes de bactéries séquencé actuellement. D'un autre côté il n'est pas étonnant d'identifier plusieurs partenaires pour cette protéine. Car notre hypothèse de départ était que PAB0955 soit impliquée à l'interface entre division cellulaire et réplication de l'ADN. En jouant un rôle biologique à l'interface de deux mécanismes biologiques il est évident qu'il n'existe pas un seul partenaire pour cette GTPase. De plus les processus fondamentaux chez les archaea étant souvent un modèle simplifié de ceux des eucaryotes, il est probable que les partenaires soient plus nombreux chez les eucaryotes.

Un autre élément important qui peut permettre de comprendre pourquoi ces GPN loop GTPases pourraient avoir des partenaires différents chez les archaea et chez les eucaryotes est l'insertion I2. Cette région est la moins bien conservée en séquence dans ces GTPases. Le repliement de type HEAT motif et ses multiples conformations nous laisse supposer que l'insertion I2 pourrait constituer un site d'interaction avec un partenaire. Comme I2 n'est pas conservée en séquence à travers cette famille, elle pourrait fixer des partenaires différents.

Les résultats structuraux de PAB0955 pourront donc sans doute servir de modèles pour les autres GPN loop GTPases. Pour ce qui est des partenaires et donc de la fonction biologique de ces GTPases chez les archaea et chez les eucaryotes, ils risquent d'être différent. Ceci étant sûrement dû à une co-évolution de ces deux règnes, expliquant une diversification de partenaires, et donc une diversification de fonction.



INTRODUCTION

DU PROJET

IMMUNOLOGIE



## 1. Le système immunitaire

Le système immunitaire sert à protéger l'organisme contre une attaque possible par un agent infectieux. Il permet à tout organisme de se défendre contre une très grande variété de pathogènes. Classiquement, l'on distingue les réponses immunitaires innée et adaptative.

La réponse immunitaire innée est immédiate. Elle est présente dans toutes les espèces d'invertébrés et vertébrés. C'est la première ligne de défense contre une attaque de pathogènes. Elle fait intervenir des cellules phagocytaires capables d'internaliser et de détruire les pathogènes, des cellules cytotoxiques et des molécules telles que le complément.

La réponse immunitaire adaptative est spécifique des vertébrés gnathostomes. C'est une réponse plus longue à se mettre en place (quelques jours). Cette réponse est spécifique de l'agent infectieux et confère une mémoire immunitaire (sollicitée par les vaccins par exemple). Elle fait intervenir des cellules spécialisées, les lymphocytes qui expriment à leur surface des récepteurs d'antigènes membranaires, capables de détecter la présence de pathogènes. Ces récepteurs d'antigènes comprennent les immunoglobulines (IG), aussi appelées anticorps, et les récepteurs T (TCR) qui sont exprimés respectivement par deux types de lymphocytes : les lymphocytes B expriment les IG ou anticorps et les lymphocytes T expriment les récepteurs T (TCR). Après reconnaissance de l'agent infectieux par ces récepteurs spécifiques, les lymphocytes B prolifèrent et se différencient en plasmocytes lesquels sécrètent des anticorps solubles (réponse humorale) de même spécificité que les anticorps membranaires. Les lymphocytes T peuvent être classés selon leur fonction. Les lymphocytes T cytotoxiques (Tc) détruisent la cellule infectée par un virus ou la cellule devenue cancéreuse (réponse cellulaire). Les lymphocytes T auxiliaires ou helper (Th) aident à la différenciation des lymphocytes B en plasmocytes et ainsi contribuent à la réponse humorale.

Les IG reconnaissent l'antigène sous forme native, non modifiée tandis que les TCR reconnaissent l'antigène sous forme de peptides présentés par les protéines du complexe majeur d'histocompatibilité (CMH) (HLA pour *Human Leukocyte Antigen* chez l'homme). Ces peptides proviennent de la dégradation protéolytique (ou « processing ») des antigènes protéiques dans la cellule infectée ou tumorale. Les protéines du CMH appartiennent à deux classes. Les protéines CMH de classe I sont ubiquitaires et présentes à la surface de presque toutes les cellules de l'organisme, elles permettent de signaler aux lymphocytes T cytotoxiques (Tc) si une cellule a été infectée ou est devenue tumorale. Les protéines de



classe II ne sont présentes qu'à la surface de cellules présentatrices d'antigènes (CPA) dites professionnelles (macrophages, cellules dendritiques, lymphocytes B) (fig.6.), elles permettent de signaler la présence de pathogènes aux lymphocytes T auxiliaires (Th) qui, suite à cette reconnaissance, sont ainsi activés et aident les lymphocytes B à se différencier en plasmocytes.

Notre étude porte sur deux des acteurs principaux impliqués dans la réponse immunitaire adaptative, les protéines du CMH et les TCR. Les protéines du CMH présentent en permanence des peptides en surface de nos cellules, afin de maintenir la tolérance au soi et afin d'alerter le système immunitaire, en cas d'invasion intracellulaire. Les peptides issus de la dégradation de protéines (pathogènes ou non) sont pris en charge par le CMH, pour être présentés aux TCR. La fonction du TCR est de distinguer l'origine des peptides (soi ou non-soi) présentés par le CMH à la surface de la cellule que le lymphocyte vient de reconnaître.

## 2. Les récepteurs des lymphocytes T (TCR)

Les récepteurs T ou TCR (T cell Receptor), à la surface des lymphocytes T, est responsable de la reconnaissance spécifique d'un antigène, lorsque celui-ci est présenté sous forme de peptide en complexe avec une protéine CMH à la surface d'une autre cellule. L'interaction entre le TCR et le complexe peptide antigénique-CMH (pCMH) initie une cascade de signalisation (fig.1) (Hudrisier and Gairin, 1996) qui induit une activation spécifique : par exemple, l'élimination de la cellule infectée par un  $L_T$  cytotoxique (Bevan, 1984).

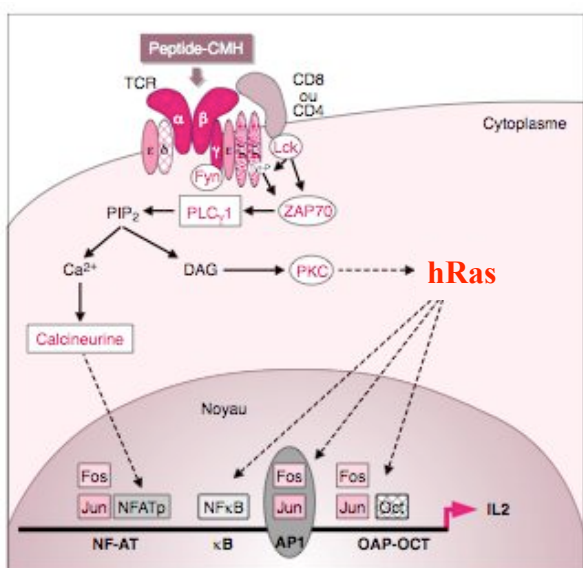


Figure.1. La transduction du signal par le TCR.

Cette cascade de signalisation active les kinases Fyn et Lck qui phosphorylent les motifs ITAM, puis ZAP-70 est recrutée. Il s'en suit une cascade d'activation impliquant plusieurs protéines dont hRas une petite protéine G. Les facteurs de transcription Fos, Jun, NFAT, NFκB et Oct se fixent sur les séquences promotrices du gène IL2 et entraîne sa transcription (fig.1.).

Les lymphocytes T qui expriment les récepteurs T  $\alpha\beta$  (TCR $\alpha\beta$ ) sont de deux types: les lymphocytes T cytotoxiques, pour la plupart CD8<sup>+</sup>, tuent les cellules dont le CMH présente un peptide viral ou tumoral, et les lymphocytes T auxiliaires (helper), pour la plupart CD4<sup>+</sup>, sont là pour activer, via un deuxième signal de reconnaissance, les CPA c'est-à-dire soit les macrophages (T<sub>H1</sub>), soit les lymphocytes B (T<sub>H2</sub>). Les TCR CD8<sup>+</sup>, qui font l'objet de notre étude, jouent un rôle important dans la lutte contre des infections virales comme la variole (Jackson et al., 1977), la grippe (Ennis et al., 1977), la rage (Wiktor et al., 1977), le SIDA (Chermann et al., 1983), etc. Les TCR CD4<sup>+</sup>, eux, jouent un rôle important dans la lutte contre des infections comme la tuberculose ou la lèpre (Bach et al., 1983).

## 2.1. Les réarrangements des gènes des TCR

Pour faire face à la multitude d'antigènes existant, notre organisme exprime une grande variété de récepteurs T capables de reconnaître les différents peptides antigéniques complexés au CMH (pCMH). Par exemple, il y a 20<sup>9</sup> possibilités de séquences différentes pour un nonapeptide, soit 5.10<sup>11</sup> Ag différents.

Notre génome contient 30 000 gènes alors qu'on estime que le nombre de récepteurs T différents que l'organisme est capable de synthétiser est de l'ordre de 10<sup>12</sup>. Un nombre équivalent d'IG est synthétisé par l'organisme. Ces nombres sont atteints grâce aux réarrangements que subissent les gènes des TCR et des IG, par un mécanisme propre aux vertébrés gnathostomes. Ainsi les différentes chaînes  $\alpha$  et  $\beta$  du TCR sont formées par un réarrangement de gènes V (variable) et J (jonction) pour la chaîne  $\alpha$  et de gènes V, D (diversité) et J pour la chaîne  $\beta$  à l'origine de la grande diversité des TCR (fig.2.). Chez l'homme, le locus TRA est localisé sur le chromosome 14, et le locus TRB est localisé sur le chromosome 7. Il existe 43-45 gènes TRAV, 50 gènes TRAJ en amont d'un seul gène TRAC. Les transcrits des gènes réarrangés TRAV-TRAJ (qui codent le domaine V-ALPHA) sont épissés avec le gène TRAC. Il existe 40-48 gènes TRBV, 2 gènes TRBD, 12-13 gènes TRBJ et 2 gènes TRBC. Les transcrits de gènes réarrangés TRBV-TRBD-TRBJ (qui codent le domaine V-BETA) sont épissés avec l'un des deux gènes TRBC (Lefranc and Lefranc).

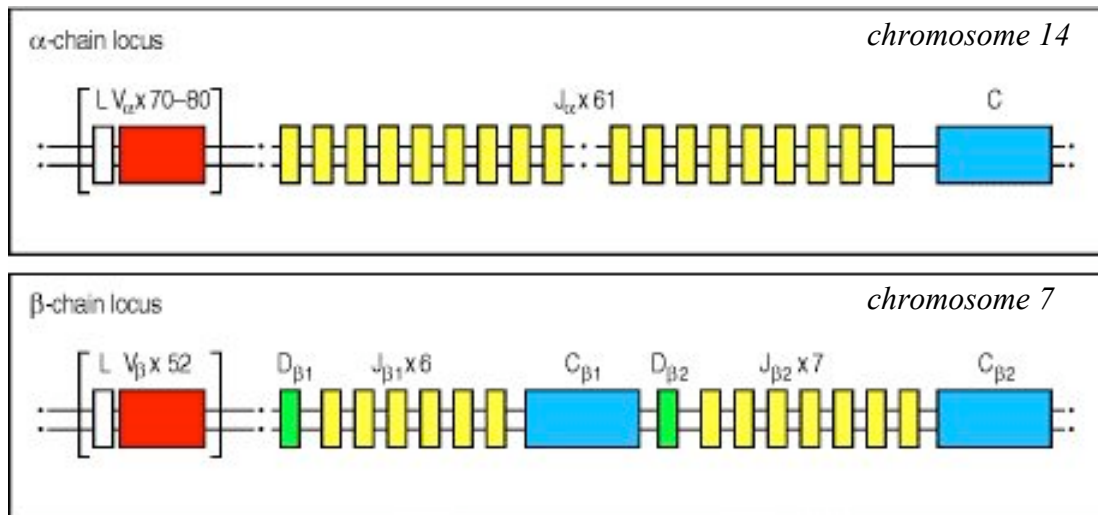


Figure.2. Les gènes codant les chaînes  $\alpha$  et  $\beta$  du TR.

Il existe également une diversité jonctionnelle V-J, V-D et D-J qui produit des insertions/délétions aléatoires dans les régions jonctionnelles (fig.3.). Les variations au niveau des jonctions sont la source essentielle de diversité ( $10^{13}$  contre  $10^6$  pour la recombinaison génétique). Cette diversité se localise au niveau des CDR3 (Complementarity Determining Region) de chaque chaîne. Ces régions CDR3 sont impliquées dans l'interaction du TCR avec le peptide/CMH.

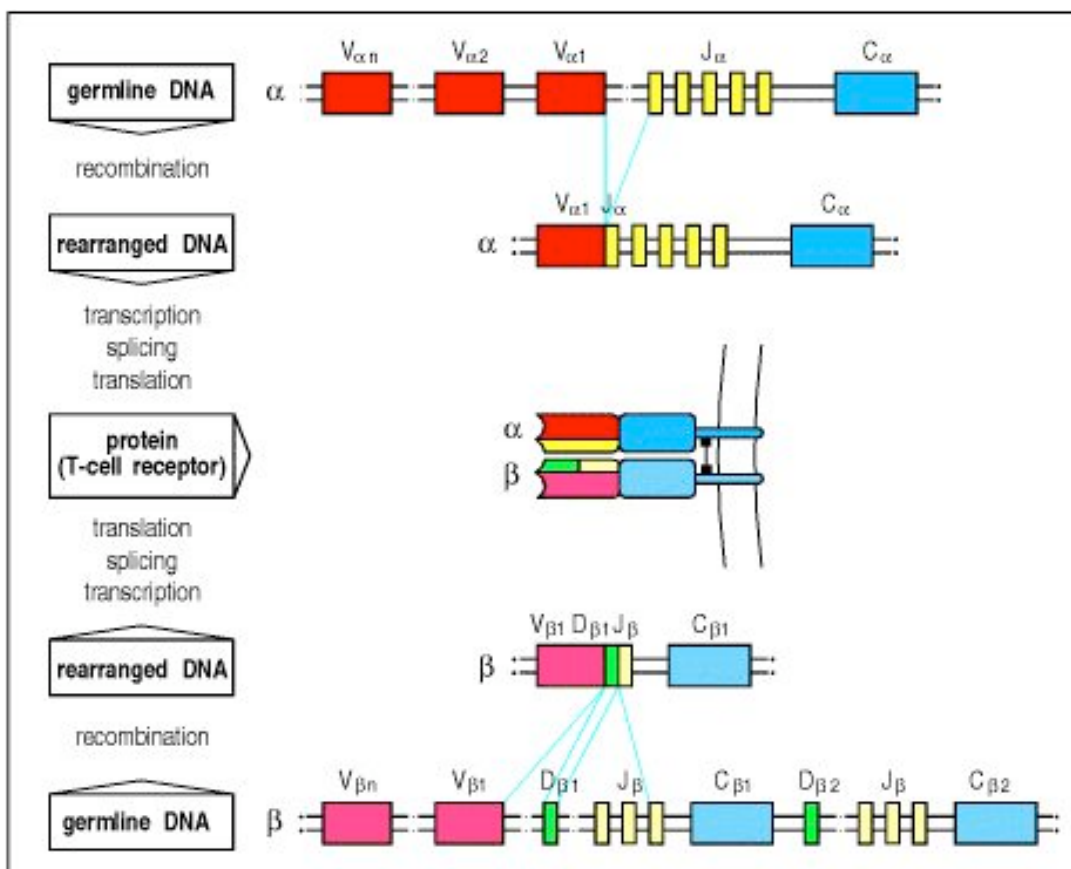


Figure.3. Les réarrangements des gènes codant les chaînes  $\alpha$  et  $\beta$  du TCR.

Cette variabilité permet de construire théoriquement environ  $10^{16}$  TCR différents. Un individu, lui, ne possède qu'environ  $10^8$  TCR différents à un instant donné. Le fait qu'un TCR donné puisse reconnaître plusieurs complexes pCMH permet sans doute d'assurer l'existence d'au moins un TCR pour chaque antigène. Cette dégénérescence de reconnaissance du TCR est sans doute cruciale pour faire face à la très grande variété d'antigène.

## 2.2. La structure du TCR

La première structure d'une chaîne  $\beta$  de TCR fut résolue en 1995 par l'équipe de G. Bentley (Bentley et al., 1995a) (code PDB :1BEC), et les deux premières structures de complexe ternaire TCR-pCMH furent résolues l'année suivante (Garboczi et al., 1996, Garcia et al., 1996). Depuis, 24 structures de complexes ternaires ont été résolues avec des TCR $\alpha\beta$  (Kaas. et Lefranc M.-P.).

Ces structures permettent d'établir des règles générales sur la reconnaissance du pCMH par le TCR, mais les détails de la spécificité de reconnaissance sont complexes. Les différents complexes TCR-pCMH ont permis de comprendre comment ces trois partenaires interagissaient. Pour autant ces complexes ne permettent pas d'expliquer comment et pourquoi un complexe pCMH déclenche une réponse immunitaire. Afin de comprendre les clefs qui font qu'une interaction devient productive, il faut sans doute d'avantage de structures de complexes ternaires afin de les comparer et d'en analyser les différences. Un des principaux obstacles à la détermination de nouvelles structures de complexe ternaire est l'obtention de TCR en quantité suffisante, stable et homogène pour permettre son étude par cristallographie. Nous verrons d'ailleurs dans le chapitre III sur les complexes ternaires, que nous rencontrons des difficultés pour produire le TCR.

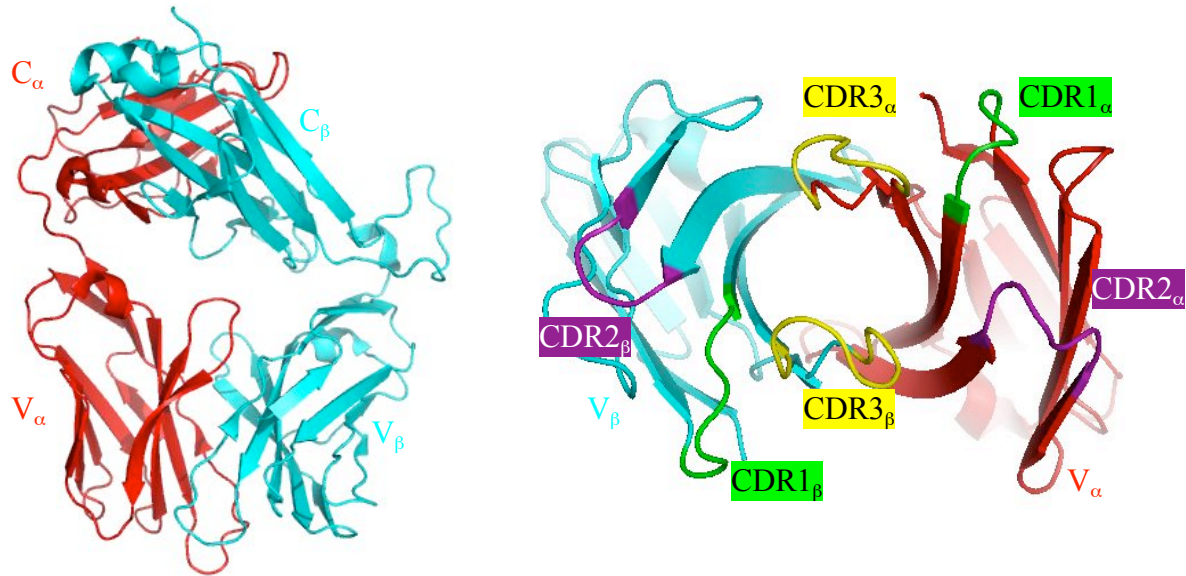


Figure.4.a. La structure du TCR 1G4

Figure.4.b. Représentation des régions CDR.

Pour les deux figures le TCR est 1G4 (code PDB : 2BNU (Chen et al., 2005)), avec la chaîne  $\alpha$  en rouge et la chaîne  $\beta$  en cyan.

Comme le révélait leur structure génétique et leur séquence, la structure tridimensionnelle des TCR $\alpha\beta$  est très similaire à celles des fragments Fab des anticorps (Garcia et al., 1999, Garcia et al., 1996, Bentley et al., 1995b). Les domaines des TCR et des fragments Fab sont de type Ig-like. Le domaine Ig-like est composé d'un sandwich de feuillets  $\beta$  antiparallèles. Les chaînes  $\alpha$  et  $\beta$  du TCR adoptent un repliement à deux domaines Ig-like, V-ALPHA et C-ALPHA, pour la chaîne  $\alpha$ , V-BETA et C-BETA pour la chaîne  $\beta$ , les domaines variables V-ALPHA (V-J-REGION) et V-BETA (V-D-J-REGION) étant respectivement issus du réarrangement des gènes TRAV-TRAJ et TRBV-TRBD-TRBJ (fig.4.a).

Les domaines variables de chaque chaîne, situés en  $N_{\text{ter}}$ , s'associent entre eux de façon non covalente. Les domaines constants, situés en  $C_{\text{ter}}$ , s'associent également entre eux et forment un pont disulfure inter-chaîne. Chaque chaîne possède un domaine transmembranaire à son extrémité carboxylique.

Les domaines variables portent les CDR qui sont les boucles formant le site de reconnaissance du complexe pCMH. Chaque domaine variable de chaque chaîne comporte 3 CDR, nommés CDR1, CDR2 et CDR3 (fig.4.b.). Le TCR est aidé lors de l'interaction avec le pCMH par des co-récepteurs spécifiques du type de  $L_T$ , le CD8 pour la majorité des  $L_T$  cytotoxiques ( $CD8^+$ ) et les co-récepteurs CD4 pour la majorité des  $L_T$  auxiliaires ( $CD4^+$ ).

### 3. Les protéines CMH

Dans la réponse adaptative immunitaire cellulaire, les peptides antigéniques sont présentés aux TCR par les protéines de CMH de classe I ou II, caractérisés par la présence de deux G-DOMAINS qui comportent chacun une hélice  $\alpha$  et quatre brins  $\beta$  antiparallèles et qui ensemble forment le site de liaison du peptide (ou « groove »). Pour les deux classes de CMH l'architecture d'ensemble est similaire, cependant dans le cas des CMH de classe I, les deux G-DOMAINS appartiennent à la même chaîne alpha, laquelle est associée de manière non covalente à la beta2-microglobuline, tandis que dans le cas des CMH de classe II, les deux G-DOMAINS appartiennent à deux chaînes, alpha et beta. La première structure de CMH qui a été résolue est celle du HLA-A2 (CMH de classe I) en 1987 par l'équipe de Don Wiley (Bjorkman et al., 1987) (code PDB :1HLA), depuis de nombreuses structures de complexes pCMH ont été résolues avec plus de 250 entrées dans PDB et dans IMGT/3Dstructure-DB (<http://imgt.cines.fr>).

#### 3.1. Les gènes du CMH

Les protéines de CMH sont présentes à la surface de la plupart des cellules de l'organisme. Ceci permet d'alerter le système immunitaire de la présence éventuelle de pathogènes dans les cellules.

Chez l'homme, les gènes du CMH (ou HLA) sont situés sur le chromosome 6 (fig.5.). Un individu possède une quinzaine de gènes CMH de classes I et II. Les gènes CMH sont les plus polymorphes de notre génome, il existe de nombreux allèles pour chaque gène. Par exemple pour le CMH de classe I de type B il existe 395 allèles différents. En conséquence la probabilité de trouver la même combinaison allélique pour deux individus est donc extrêmement faible. La sensibilité aux antigènes est donc différente d'un individu à l'autre. Une des premières conséquences est que la variété des CMH au sein d'un individu permet de présenter une large gamme de peptides. Cette diversité limite les possibilités qu'un agent infectieux échappe à la vigilance du système immunitaire. De même, un pathogène a peu de chance de pouvoir affecter l'espèce entière, car les individus ont une sensibilité différente grâce à cette variabilité allélique des protéines CMH. La spécificité du répertoire des  $L_T$  d'un individu est restreinte par la combinaison des protéines CMH qu'il possède, puisqu'elle joue



un rôle majeur dans la phase de sélection des  $L_T$ . Cette spécificité est aussi la principale cause des rejets de greffes.

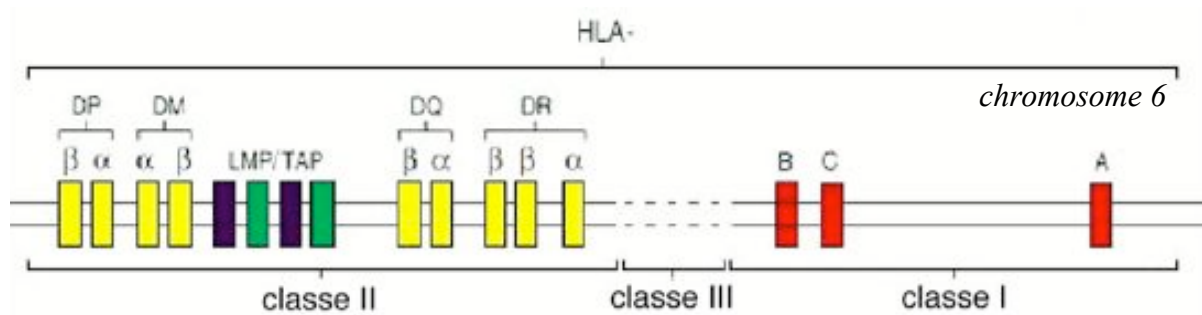


Figure.5. Les gènes du CMH humain (ou HLA).

### 3.2. La structure des protéines CMH

Les protéines CMH de classe I sont exprimées par toutes les cellules nucléées, alors que les protéines CMH de classe II sont présentes à la surface des cellules présentatrices d'Ag professionnelles (CPA) (macrophages, cellules dendritiques, lymphocytes B) (fig.6.).

	CMH de classe I	CMH de classe II
Macrophages, cellules dendritiques, lymphocytes B, lymphocytes T activés	+	+
Neutrophiles	+	+
Foie, reins, neurones	+	-
Globules rouges	-	-

Figure.6. Tableau récapitulatif de la localisation des protéines CMH sur les cellules de l'organisme.

Les peptides présentés par ces deux classes de CMH sont différents par leur longueur et leur origine (voir paragraphe D). Les CMH de classe I complexés avec un peptide, sont reconnus par les  $L_T$  cytotoxiques ( $CD8^+$ ), et les CMH de classe II le sont par les  $L_T$  auxiliaires ( $CD4^+$ ).

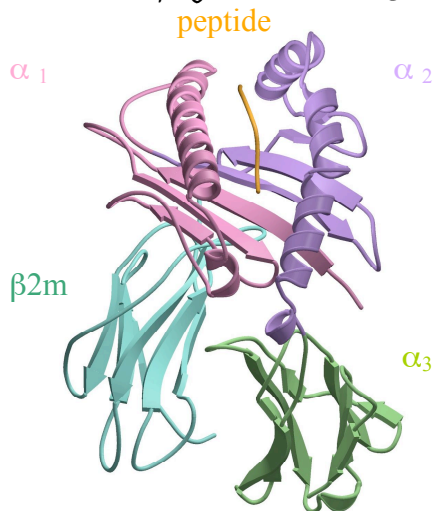


Figure.7. La structure de HLA-A2 (CMH de classe I).

Les protéines CMH de classe I sont composées de 2 chaînes polypeptidiques, associées de façon non covalente. La chaîne  $\alpha$ , constituée d'environ 300 résidus, est constituée de deux G-DOMAINS, ALPHA1 et G-ALPHA2 qui forment le site de fixation du peptide et d'un troisième domaine Ig-like (C-LIKE-DOMAIN). La partie  $C_{\text{ter}}$  de la chaîne s'ancre à la membrane des cellules. La chaîne  $\alpha$  est associée de manière non covalente à la  $\beta 2$ -microglobuline, d'environ 100 résidus. qui

forme un C-LIKE-DOMAIN et interagit de façon non-covalente avec les 3 domaines de la chaîne  $\alpha$ . Le site de fixation du peptide (fig.7.) dans les CMH de classe I est formé par les hélices de G-ALPHA1 et G-ALPHA2, qui entourent le peptide. Le « plancher » du sillon est formé par un feuillet  $\beta$  auquel contribuent les quatre brins anti-parallèles de G-ALPHA1 et G-ALPHA2. La particularité de ce site de fixation du peptide est sa fermeture aux extrémités. Cette conformation fermée permet la fixation, principalement, de peptides courts (en général de 8 à 10 résidus) dont les deux extrémités sont ancrées dans le sillon peptidique. Cependant il existe des exceptions avec des peptides plus longs (13 résidus) fixés par des CMH de classe I (Tynan et al., 2005a). Ces peptides plus longs ancrent également leur deux extrémités dans le sillon peptidique, avec la partie centrale très protubérante par rapport à un peptide de taille plus « classique ».

Les chaînes latérales des résidus du peptide peuvent être soit enfouies à l'intérieur du CMH, soit être exposées au TCR (ou au solvant dans les complexes binaires pCMH). Le polymorphisme (plusieurs allèles) des CMH affecte essentiellement les résidus impliqués dans les interactions avec le peptide (résidus du sillon peptidique). Ainsi chaque allèle permet de fixer un répertoire de peptides distincts, puisque les résidus du sillon peptidique sont différents d'une protéine CMH à l'autre. Dans le cas de notre étude nous sommes intéressés à la molécule HLA-A2, CMH de classe I.

Les protéines CMH de classe II sont formées de 2 chaînes  $\alpha$  et  $\beta$ , ces chaînes sont toutes deux codées par des gènes CMH et ancrées à la membrane. Chaque chaîne est formée de 2



domaines : un G-DOMAIN (G-ALPHA ou G-BETA) et un C-LIKE-DOMAIN. Les domaines G-ALPHA et G-BETA des CMH de classe II sont structurellement semblables aux domaines G-ALPHA1 et G-ALPHA2 des CMH de classe I. Le sillon peptidique formé par les domaines G-ALPHA et G-BETA, est semblable à celui des CMH de classe I mais les extrémités sont ouvertes, ce qui permet d'accueillir des peptides plus longs (typiquement 13 à 17 résidus). Le C-LIKE-DOMAIN des chaînes  $\alpha$  et  $\beta$  des CMH de classe II ressemble structurellement au C-LIKE-DOMAIN de la  $\beta$ 2-microglobuline et à celui des chaînes  $\alpha$  des CMH de classe I, à la différence qu'ils sont tous les deux ancrés à la membrane par leur extrémité  $C_{\text{ter}}$ .

Chaque protéine CMH peut présenter théoriquement de l'ordre de  $10^7$  à  $10^8$  peptides différents.

## 4. Les peptides antigéniques

Les peptides antigéniques présentés par les protéines du CMH sont issus d'une dégradation protéolytique : l'apprêtement peptidique. Les épitopes peuvent donc provenir de toutes les régions des protéines, contrairement aux épitopes reconnus par les anticorps et situés à la surface des Ag natifs. Le mécanisme d'apprêtement peptidique est différent pour les deux classes de protéines CMH, et correspond à des compartiments cellulaires différents.

### 4.1. Les peptides présentés par les protéines CMH de classe I

Dans le cytosol se trouvent des protéines d'origine différentes, produites par l'individu (protéines du soi) ou issues de pathogène (protéines du non-soi, virales par exemple). Toutes ces protéines sont dégradées dans le cytosol par le protéasome, en fragments de 8 à 10 acides aminés. Ces peptides sont ensuite amenés du cytosol au réticulum endoplasmique où ils seront fixés par les protéines CMH de classe I partiellement repliées.

Les peptides utilisés dans notre étude sont issus de la dégradation de différents virus dans le cytosol (virus Epstein Barr EBV, virus du CytoMégalo CMV, virus de l'Hépatite C VHC). Ces peptides présentés par les protéines CMH de classe I, sont donc reconnus par des  $L_T$  cytotoxiques ( $CD8^+$ ).

## 4.2. Les peptides présentés par les protéines CMH de classe II

Les CMH de classe II présentent des peptides issus de la dégradation protéique par des protéases à l'intérieure des vésicules des cellules CPA, cellules phagocytaires et lymphocytes B après internalisation. Ces peptides ont généralement une longueur d'environ 13 à 17 acides aminés. Une fois chargée avec son peptide, la protéine CMH de classe II migre vers la surface de la cellule et peut alors présenter son peptide aux  $L_T$  auxiliaires ( $CD4^+$ ).

La reconnaissance par les  $L_T CD4^+$  de ces pCMH de classe II conduit à l'activation de la cellule phagocytaire pour éliminer le pathogène, ou à l'activation des  $L_B$  qui vont se différencier en plasmocytes et sécréter des anticorps spécifiques contre le pathogène.

## 5. La rencontre des acteurs de la réponse immunitaire

### 5.1. Comment les $L_T$ différencient les peptides du soi de ceux du non-soi ?

La détection des pathogènes par les TCR est une interaction clé de la réponse cellulaire. Ce mécanisme de reconnaissance intervient à plusieurs étapes de la vie des  $L_T$ . Les cellules T doivent être capables d'interagir avec les protéines CMH sans être autoréactives, de discriminer les peptides du soi de ceux du non-soi présentés par les CMH, d'activer une réponse immunitaire contre les pathogènes et de conserver les cellules mémoire pour le maintien de l'immunité.

Les mécanismes de maturation et de sélection permettent aux TCR d'acquérir une spécificité pour les complexes pCMH. Tout d'abord les lymphocytes T subissent une double sélection dans le thymus : (i) une sélection positive, basée sur la capacité des TCR à interagir avec les pCMH, où seuls survivent les cellules T capables d'interagir suffisamment avec les CMH complexés aux peptides du soi (dégradation de protéines de l'organisme) (Bevan, 1977, Zinkernagel et al., 1978), et (ii) une sélection négative, qui a pour but d'éliminer les  $L_T$  autoréactifs, ou les  $L_T$  qui interagissent trop fortement avec les pCMH contenant des peptides du soi (Lo et al., 1986, Kappler et al., 1987).

La maturation du répertoire de  $L_T$  spécifiques s'effectue à partir des thymocytes (cellules progénitrices) qui migrent de la moelle osseuse au thymus. Dans le thymus, ces cellules se différencient et prolifèrent, durant cette phase a lieu les réarrangements des gènes dans les locus TRA et TRB qui permettent de synthétiser les TCR. Les TCR subissent le processus de

sélection par présentation d'Ag du soi complexés à des protéines CMH (pCMH) à la surface de cellules CPA, ce qui va en faire des  $L_T$  compétents. Ces  $L_T$  peuvent ensuite passer dans le système circulant (vaisseaux sanguins et lymphatiques) où ils pourront reconnaître des pCMH à la surface de cellules infectées ou tumorales.

## 5.2. Le complexe ternaire : TCR-pCMH

La rencontre du TCR à la surface des lymphocytes cytotoxiques avec le complexe pCMH à la surface de cellules infectées ou tumorales est appelée par les immunologistes le « baiser de la mort », car il entraîne la mort de la cellule présentatrice d'Ag (CPA) s'il y a activation du  $L_T$  cytotoxique. Cette rencontre est un événement rapide, la demi-vie du complexe ternaire est de 2 à 10 secondes, et de faible affinité ( $K_d = 1$  à  $100 \mu M$ ). Cette interaction a lieu dans le cadre d'une structure supramoléculaire à l'interface entre le  $L_T$  et la CPA nommée « synapse immunologique » ou SMAC (SupraMolecular Activation Cluster) par analogie à la synapse neuronale (Paul and Seder 1994).

La première structure de complexe ternaire a été déterminée par cristallographie en 1996 (Garboczi et al., 1996). Depuis, PDB et IMGT/3Dstructure-DB disposent de 24 complexes ternaires (dont 17 avec des CMH de classe I) (Kaas et Lefranc). Le nombre de structures de complexe ternaire est très restreint comparé à la diversité des molécules TCR et CMH. Ceci permet néanmoins de disposer d'un échantillon structural pour analyser cette interface.

Les complexes ternaires TCR-peptide-CMH de classe I enfouissent à l'interface entre 600 et  $1000 \text{ \AA}^2$ . Le mode global d'interaction est conservé entre les différentes structures, mais les interactions interatomiques ne le sont pas. Ainsi, parmi les différentes structures de complexes ternaires le domaine V-ALPHA contribue de 37 à 74%, le domaine V-BETA de 32 à 67% à l'interaction faite par le TCR (Rudolph et al., 2006). De même la contribution de chaque CDR varie d'un complexe à l'autre. Cette observation est surprenante car malgré une architecture et des régions conservées (en structure et en séquence) du côté TCR comme du côté CMH, il ne semble pas y avoir de règle de reconnaissance autre que l'efficacité de l'interaction.

## 6. Quelques définitions

Pour mieux comprendre la suite du manuscrit quelques définitions s'imposent.

### Affinité/Avidité

L'affinité est l'attraction spécifique entre deux molécules possédant chacune un site d'interaction. L'avidité traduit cette attraction spécifique en présence de plusieurs sites d'interaction. Ceci se traduit directement sur la demi-vie du complexe.

Par exemple un TCR a une affinité pour un complexe pCMH donné. Alors qu'un lymphocyte T, qui possède à sa surface plusieurs TCR, a une avidité pour une cellule présentatrice d'antigène, du fait que la synapse immunologique contient un grand nombre de TCR et de pCMH.

### Fréquence des cellules T

La fréquence désigne le nombre de cellules T réactives pour un complexe pCMH donné dans une population de cellules T n'ayant pas rencontré l'antigène avant (naïf ou à réactivité croisée). Par exemple pour un individu séronégatif pour l'hépatite C, la fréquence sera le nombre de  $L_T$  spécifiques activés par un peptide du virus de l'hépatite C (VHC) parmi un nombre de cellules données à la population lymphocytaires du sang périphérique.

### Réactivité croisée

La réactivité croisée du lymphocyte T vis-à-vis d'une protéine CMH, est la capacité du TCR à reconnaître plusieurs peptides présentés par la même protéine CMH (Tallquist et al., 1996, Bongrand and Malissen, 1998). Cette propriété permet au répertoire de cellules T sélectionné dans le thymus de reconnaître un grand nombre de peptides étrangers. Cette dégénérescence de la spécificité du TCR semble nécessaire pour une protection efficace contre les divers pathogènes existants (Mason, 1998).

La réactivité croisée peut provenir d'un mécanisme de mimétisme (par exemple : deux peptides différents présentés par des protéines CMH différentes présentent au TCR des surfaces quasi identiques) (Tynan et al., 2005b). Elle peut également avoir pour origine l'adaptabilité des CDR qui permettent au TCR d'interagir productivement avec des peptides très différents.

### CMH allogénique

Un individu possède un certain répertoire de protéines CMH qui, par suite du polymorphisme des différents gènes est différent d'un individu à l'autre. Une protéine de CMH d'un individu est allogénique pour un autre individu qui ne la possède pas. Dans le cas de notre étude par exemple, nous avons sélectionné des individus possédant la molécule HLA-A2 (dit A2<sup>+</sup>) et des individus ne possédant pas la molécule HLA-A2 (dit A2<sup>-</sup>). Donc pour un individu A2<sup>-</sup>, la molécule de HLA-A2 est une molécule de CMH allogénique.

### Alloréactivité

Un à 10% des L<sub>T</sub> circulants sont capables de reconnaître des variants alléliques des protéines CMH. Ces protéines CMH allogéniques (du non-soi) diffèrent essentiellement par les résidus du sillon peptidique des protéines CMH de l'individu. Elles présentent donc des répertoires de peptides du soi différents. Ainsi un grand nombre des peptides du soi deviennent des Ag potentiels, puisqu'ils n'ont jamais été présentés aux lymphocytes T (au moment de la sélection dans le thymus). Dans le cas d'une infection par un pathogène le nombre de L<sub>T</sub> activés s'élève au plus à quelques milliers. Alors que dans le cas d'une réaction alloréactive ce nombre est 100 à 1000 fois supérieur. L'alloréactivité entraîne donc une réponse immunitaire très forte. Deux modèles ont été proposés pour expliquer la grande fréquence des cellules T alloréactives : l'un indépendant de la reconnaissance du peptide (Bevan, 1984) et l'autre dépendant de la reconnaissance du peptide (Matzinger and Bevan, 1977).

Dans le cas des transplantations, le rejet de greffe est un exemple d'alloréactivité. Le greffon possédant des protéines CMH absentes chez le receveur (allogénique), il y a un risque de rejet par activation du système immunitaire du receveur qui reconnaît les CMH allogéniques comme pathogène.

### TR public et TR privé (Welsh, 2006)

Lors d'une infection par un pathogène les complexes pCMH activent les L<sub>T</sub>. Parmi les cellules T il y a des clones dit « publics » et d'autres dits « privés ». Les TCR « publics » présentent les mêmes caractéristiques chez un grand nombre d'individus, c'est-à-dire que la réponse immunitaire sera dominée par un TCR quasiment unique pour l'ensemble de la population. Un TCR « unique » sous-entend qu'il y a récurrence à divers degrés ; récurrence des chaînes  $\alpha$ , des chaînes  $\beta$ , des CDR, ou des domaines variables de chaque chaîne du TCR. Une autre caractéristique des populations de TCR « publics » est que ces populations sont

restreintes, comme s'il n'y avait qu'une seule solution de réponse face à l'antigène. Au contraire les TCR « privés » sont plus variés et sont spécifiques de chaque individu.

## 7. But du projet

La compréhension de la reconnaissance TCR-pCMH au niveau atomique est un défi mais l'intérêt au niveau thérapeutique devrait être important. Le but à long terme de ce projet est d'établir les connaissances fondamentales essentielles pour permettre la conception de nouveaux vaccins capables d'induire une réponse immunitaire forte, efficace et durable.

**L'objectif actuel du projet est de savoir s'il y a une relation entre les caractéristiques structurales des complexes pCMH et la fréquence et/ou la diversité des clones de cellules T (TCR « privé » ou « public ») spécifiques sélectionnés.**

Nous voulons, à l'aide de plusieurs structures de complexes binaires, analyser le mode de fixation de différents peptides par le HLA-A2, afin d'établir les liens entre la fréquence et/ou le type de répertoire de  $L_T$  sélectionnés.

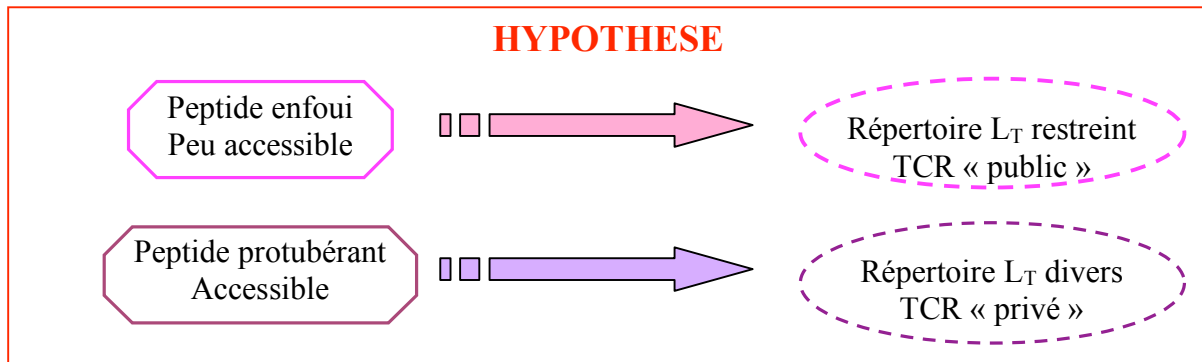
Si les liens existent, cela nous aidera à établir un modèle prédictif, dans le but de pouvoir concevoir des peptides capables d'induire une réponse immunitaire forte, efficace et durable.

L'hypothèse de départ, nous permettant de supposer qu'un lien existe entre les structures et la fréquence/diversité des cellules T, est celle avancée par l'équipe de Peter C. Doherty (Turner et al., 2005) de l'accessibilité du peptide au solvant. Dans leurs travaux, les auteurs ont mis en évidence un changement de répertoire de cellules T activées, induit par une mutation dans un peptide issu du virus de la grippe. Le peptide sauvage (PA224) sélectionne un répertoire divers de TCR « privé », alors que le peptide muté (NP366) sélectionne un répertoire plus restreint de TCR « public ». La mutation concerne l'arginine en position 7, qui est remplacée par une alanine, ceci induit un changement de conformation du peptide diminuant son accessibilité au solvant. L'hypothèse est donc qu'un pCMH possédant un peptide « enfoui », peu accessible au solvant, présente une surface composite relativement plane au TCR.

Dans ce cas de figure, il y a peu d'éléments spécifiques à reconnaître pour le TCR, et un nombre de modes d'interactions productives plus restreint. Ce qui peut expliquer le fait que les TCR sélectionnés soient peu divers, voir unique. La même « solution » moléculaire peut se retrouver pour un ensemble de la population.

## Introduction du projet immunologie

Dans le cas d'un peptide protubérant, plus accessible au solvant, au contraire il se pourrait que plusieurs modes d'interactions productives soient envisageables expliquant la mobilisation d'un répertoire plus divers de TCR, avec une solution différente pour chaque individu (TCR « privé »).



CHAPITRE 1

LE HLA-A2

ET LES

DIFFÉRENTS

PEPTIDES





## I.1. Le choix des partenaires moléculaires

### I.1.1. La molécule CMH de classe I : HLA-A2m

La molécule de CMH étudiée dans le cadre de notre projet est HLA-A\*0201, dont la chaîne lourde  $\alpha$ , constituée de 277 résidus, a été mutée sur le résidu 245 (alanine en valine) afin qu'elle ne fixe pas les co-facteurs CD8 des  $L_T$  cytotoxiques. Ceci ne change ni la structure ni les propriétés de fixation du peptide par le HLA-A2m. La mutation est faite pour ne pas avoir de signal CD8 dépendant au niveau des tris cellulaires effectués avec les complexes pCMH sur les cellules T (voir chapitre III). Cette chaîne- $\alpha$  est en interaction non covalente avec la  $\beta$ 2-microglobuline, qui possède 99 résidus. L'ensemble de ces deux molécules forme la molécule de HLA-A2 que nous nommons HLA-A2m pour indiquer le présence de la mutation dans la chaîne  $\alpha$ .

La molécule HLA-A2 est la molécule de CMH de classe I la plus répandue dans la population européenne, environ 50%, donc il existe de nombreuses données biologiques. C'est pourquoi pour notre projet nous avons choisi le HLA-A2 comme molécule de CMH pour l'étude de différents peptides.

La production de la chaîne- $\alpha$  et de la  $\beta$ 2-microglobuline se fait dans *E. coli*, laquelle exprime ces protéines sous forme de corps d'inclusion. Les peptides sont synthétisés par la société Eurogentec. La molécule de HLA-A2 se replie correctement uniquement en présence des 3 éléments qui la constituent : la chaîne- $\alpha$ , la  $\beta$ 2-microglobuline et le peptide. La production en corps d'inclusion nous oblige à procéder à une étape de repliement des protéines après solubilisation des corps d'inclusion. Le désavantage de cette étape de repliement est son faible rendement (moins de 10%). En contrepartie, ce protocole de production a l'avantage de purifier nos protéines, en large excès dans les corps d'inclusion, rendant la purification plus simple. De plus au moment de l'étape de repliement nous pouvons fixer le peptide de notre choix, point essentiel pour notre étude.

La production des complexes pCMH, ainsi que les tris cellulaires, ont été réalisés par nos collaborateurs de l'équipe de Marc Bonneville (INSERM Nantes).

### I.1.2. Le choix des peptides

La sélection des peptides a été faite par nos collaborateurs de Nantes suite à leurs travaux sur la réponse de cellules T cytotoxiques dirigées contre le virus de l'hépatite C (VHC). Les patients infectés par le VHC guérissent spontanément pour 20% d'entre eux, et 80% développent une hépatite chronique. Parmi les 80% d'individus malades, 20% développent en 10 à 20 ans une cirrhose, et 3% un carcinome hépatocellulaire. L'hépatite C étant un problème de santé publique, la compréhension de la réponse immunitaire face à cette infection est très importante. Plusieurs facteurs virologiques et immunologiques ont un impact sur la réponse au traitement et sur l'élimination du virus de l'hépatite C.

Parmi les facteurs immunologiques liés à l'individu, il est établi que la réponse immunitaire cellulaire contribue à l'élimination du virus et à la détermination de l'évolution clinique de la maladie. Dans leur étude, nos collaborateurs souhaitaient savoir s'il existait une corrélation entre la qualité et/ou l'intensité des réponses des cellules T anti-VHC et l'évolution clinique de la maladie.

Parmi les cellules T activées par présentation d'un peptide de VHC par une molécule CMH de classe I, ils ont identifié 3 peptides complexés à HLA-A2 (hcv1, hcv2 et hcv3). Les complexes pCMH avec les 3 peptides du VHC présentés par HLA-A2m sont capables d'activer des  $L_T$  cytotoxiques ( $CD8^+$ ) chez des individus séropositifs (ayant déjà rencontré l'Ag) mais aussi chez les individus séronégatifs et également chez les individus  $A2^-$  (n'ayant pas le HLA-A2).

Des expériences de tris cellulaires avec d'autres peptides (pp65 et bmlf1) avaient déjà été entreprises par nos collaborateurs. Ces peptides (pp65 et bmlf1) en complexe avec HLA-A2m n'activent pas de cellules T cytotoxiques chez les individus séronégatifs et  $A2^-$ .

Les populations de cellules T cytotoxiques activées par ces 5 peptides sont différentes. Pour les peptides du virus de l'hépatite C, les répertoires de cellules T sélectionnés sont diversifiés, alors que pour les deux autres peptides les répertoires sont restreints (fig.I.1.a. et b).

Ces résultats d'activation de cellules T  $CD8^+$  spécifiques pose la question de savoir pourquoi les complexes pCMH sélectionnent des répertoires différents ? Le but étant de savoir si cette sélection a une base structurale.

	BMLF1	PP65	HCV1, HCV2 et HCV3
Séro +	100%	100%	100%
Séro -	Difficile	0%	100%
A2 <sup>-</sup>	0%	0%	100%
Répertoire L <sub>T</sub>	Restreint Récurrence Va Récurrence CDR3	Restreint Récurrence Va Récurrence CDR3	Divers

*Figure.I.1.a. Résumé des données de tri cellulaire en fonction du peptide fixé par le HLA-A2m.*

	1	2	3	4	5	6	7	8	9	10
pp65 (CMV)	N	L	V	P	M	V	A	T	V	
bmlf1 (EBV)	G	L	C	T	L	V	A	M	L	
HCV1	C	I	N	G	V	C	W	T	V	
HCV2	K	L	V	A	L	G	I	N	A	V
HCV3	L	L	F	N	I	L	G	G	W	V

*Figure.I.1.b. Séquence des différents peptides utilisés.*

## I.2. Cristallisation des complexes pCMH

La première structure de HLA-A2 a été déterminée en 1987 par l'équipe de Don Wiley (Bjorkman et al., 1987). Le peptide n'a pu être modélisé dans le sillon peptidique dans cette structure, car le cristal était formé de HLA-A2 complexé avec différents peptides. En effet, les molécules de HLA-A2 avaient été purifiées à partir de cellules de tissus humain.

Depuis, 41 structures avec HLA-A2 complexé à différents peptides ont été résolues, dont 9 en complexe avec un TCR. Plus de la moitié de ces structures ont été résolues avec des cristaux obtenus avec du MES à pH 6.5 à 25 mM comme tampon, du PEG 6000, 8000 ou 3350 ou du sulfate d'ammonium comme précipitant, et une concentration de 10 mg.ml<sup>-1</sup> en protéine. Nos premiers essais de cristallisation avec HLA-A2m en complexe avec le peptide pp65, ont donc

été réalisés autour des conditions décrites dans la littérature (Sliz et al., 2001, Bouvier et al., 1998, Khan et al., 2000, Borbulevych et al., 2005).

### 1.2.1. Les premiers essais de cristallisation

Les premiers essais de cristallisation contiennent du MES à pH 6.5 à 25 mM comme tampon, une gamme de concentration d'ammonium sulfate de 1 à 2.5 M, ou 20 à 26% en PEG 3350, ou 13 à 20% en PEG 6000 et 10 mg.ml<sup>-1</sup> de protéine.

Dans la plupart des conditions la protéine précipitait rapidement. Nous avons alors décidé de diminuer la concentration en protéine à 5mg.ml<sup>-1</sup>. Les observations restant les mêmes, nous avons ensuite diminué la concentration en agent précipitant. La protéine précipitait également en quelques heures.

La protéine précipitait dans plus de la moitié des conditions, malgré la diminution des concentrations en agent précipitant et en protéine. Notre attention s'est donc portée sur d'autres facteurs : la température, la concentration de sel, le pH, etc.

Les conditions des différents auteurs utilisant une concentration de 25 mM pour le tampon est assez inhabituelle, en général la concentration utilisée pour un tampon est de 100 mM. Cette faible concentration en tampon pourrait être insuffisante pour maintenir le pH égal à celui du tampon. Afin de savoir s'il y avait une différence de pH, nous avons mesuré les pH dans plusieurs conditions.

Un mélange de MES à pH 6.5 à 25 mM avec 20% PEG 6000 donne un pH final de 5.5, soit une unité pH d'écart avec celle du tampon. L'acidité du PEG 6000 n'est pas compensée par une faible concentration en MES. Ceci explique peut-être pourquoi notre protéine précipitait indépendamment de sa concentration et de celle en agent précipitant.

La même expérience, avec du MES à pH 6.5 à 100 mM, donne un pH final est de 5.9. Ceci reste encore éloigné du pH initial de notre tampon à pH 6.5, mais en augmentant la concentration de MES l'acidité du PEG est diminuée.

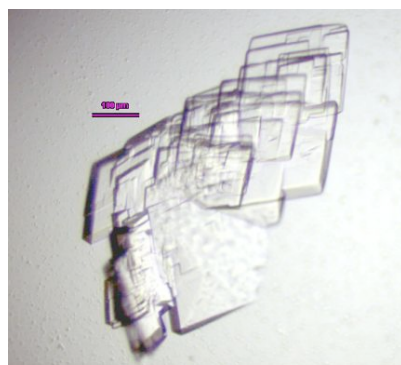
Pour conserver un pH de 6.5 après ajout de PEG il semble que le MES ne soit pas le meilleur tampon.

La même expérience avec différents tampons a donc été réalisée, plusieurs d'entre eux ont été capables de compenser l'acidité du PEG en maintenant un pH proche de 6.5.

Nous avons donc reproduit nos premiers essais de cristallisation (PEG et ammonium sulfate comme agents précipitant) avec ces différents tampons et une concentration en protéine de  $5\text{mg.ml}^{-1}$ .

### I.2.2. Notre premier cristal de HLA-A2m

Après un mois un amas de cristaux a poussé dans 20% PEG 6000 avec comme tampon du sodium citrate à pH 6.5 à 100 mM (fig.I.2.).



*Figure.I.2. Cristaux de HLA-A2m/pp65*

20% PEG 6000

Na Citrate pH 6.5 (100 mM)

$5\text{ mg.ml}^{-1}$  de protéine

$18^{\circ}\text{C}$

Nous avons découpé cet amas en fragments, et testé différents cryo-protectant (30% PEG 6000, huile de paratone, huile de paraffine). Un premier jeu de données a été enregistré au laboratoire avec le fragment cryo-protégé dans l'huile de paraffine.

Ces données complètes jusqu'à  $3\text{ \AA}$ , nous ont permis d'obtenir : le groupe d'espace  $P2_1$ , les paramètres de maille ( $a = 53\text{ \AA}$ ,  $b = 81\text{ \AA}$ ,  $c = 56\text{ \AA}$  et  $\beta = 113^{\circ}$ ), et la structure (voir chapitre II).

Tous les fragments, cryo-protégés différemment, ont été testés à l'E.S.R.F., et nous avons pu obtenir une structure à  $1.6\text{ \AA}$  de résolution (fig.I.5.). Nous n'avons pas observé de différence de qualité de diffraction des cristaux en fonction de la nature du cryo-protectant.

En conclusion le complexe peptide/HLA-A2m est sensible au pH, c'est un point très important pour la cristallisation.

### I.2.3. La technique d'ensemencement

À partir de ce fragment, nous avons utilisé la technique d'ensemencement pour obtenir de nouveaux cristaux de pp65/HLA-A2m. Le principe de l'ensemencement est d'introduire un

germe cristallin pour déclencher la nucléation. Les germes sont introduit 24 heures après la formation des gouttes. Ainsi en quelques heures les premiers cristaux apparaissent.

Il existe plusieurs techniques pour introduire des germes de cristaux, nous avons tout d'abord utilisé un cil pour toucher les cristaux légèrement broyés dans la goutte. Des petits germes se fixent au cil, puis le cil est passé dans une goutte pré-équilibrée afin d'y déposer les germes cristallins. Cela nous a permis d'obtenir de nouveaux cristaux de pp65/HLA-A2m, mais également sous forme d'amas de plaquettes.

Le problème de cette technique est le manque de contrôle de la quantité de germes introduit. Une autre façon d'obtenir les germes cristallins est de broyer les cristaux. Le micro-broyeur est un pilon en verre qui broie finement les cristaux. Ceci est réalisé dans la solution-mère enrichie de 5% en PEG pour que les cristaux ne fondent pas, car il est important de conserver l'édifice cristallin. Ainsi nous disposons d'une solution de germes cristallins, et des dilutions successives sont réalisées (1, 1/10, 1/100, 1/1000). Dans ce cas la quantité de germes introduit dans les gouttes est mieux contrôlée et plus reproductible.

Pour les complexes de HLA-A2m avec bmlf1, hcv1 et hcv2 cette technique a été utilisée dans des conditions de cristallisation proches de celles utilisées pour le peptide pp65 avec ou sans sel supplémentaire (fig.I.3. a, b et c).

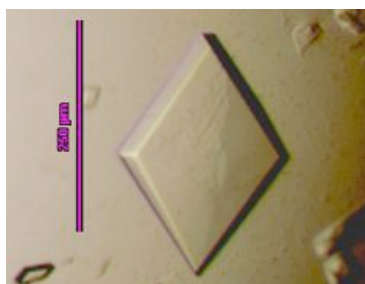


Figure.I.3.a

Cristal HLA-A2m/bmlf1

15% PEG 6000

Na citrate pH 6.5 à 0.1 M

NaCl à 0.1 M

5 mg.ml<sup>-1</sup>

dilution des germes 1/100

18°C

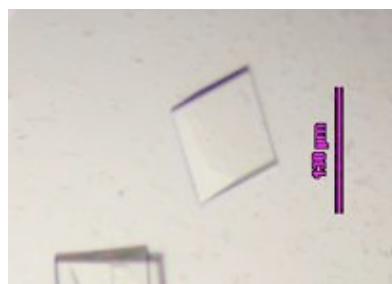


Figure.I.3.b

Cristal HLA-A2m/hcv1

17% PEG 6000

Na citrate pH 6.5 à 0.1 M

NaCl à 0.05 M

5 mg.ml<sup>-1</sup>

dilution des germes 1/100

18°C

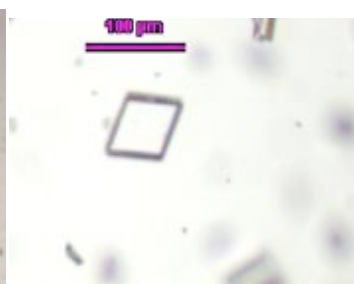


Figure.I.3.c

Cristal HLA-A2m/hcv2

18% PEG 6000

Na citrate pH 6.5 à 0.1 M

5 mg.ml<sup>-1</sup>

dilution des germes 1

4°C

La cristallisation du complexe HLA-A2m/hcv3 a été plus difficile. Après 10 jours, dans des conditions de cristallisation proche de celles des autres complexes pCMH, des amas de fines aiguilles sont apparues, appelées des oursins (fig.I.4.a). À partir de ces oursins, nous avons obtenu de nouveaux germes qui ont servi à ensemercer de nouvelles gouttes. Après l'ensemencement de germes issus des oursins, nous avons pu obtenir des baguettes séparées (fig.I.4.b). Un jeu de données complet a pu être enregistré, sur une baguette, à l'E.S.R.F. (fig.I.5.). Ce jeu de données nous a permis de comprendre pourquoi les cristaux avaient changé de forme, le groupe d'espace est différent ( $C2$  au lieu de  $P2_1$ ), donc l'empilement cristallin peut également différer.

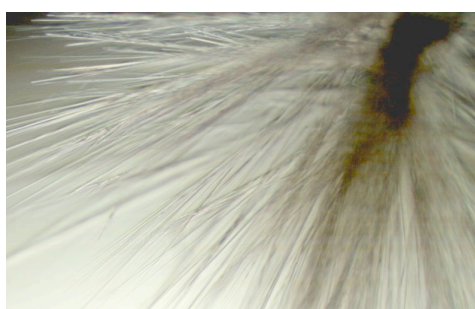


Figure.I.4.a. Oursin de HLA-A2m/hcv3

18% PEG 6000

Na citrate pH 6.5 à 0.1M

NaCl à 0.1M

5 mg.ml<sup>-1</sup>

dilution des germes de losange : 1

18°C

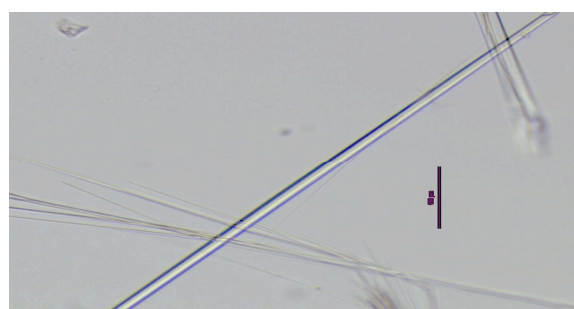


Figure.I.4.b. Baguette de HLA-A2m/hcv3

18% PEG 6000

Na citrate pH 6.5 à 0.1M

NaCl à 0.1M

5 mg.ml<sup>-1</sup>

dilution des germes d'oursin : 1/10

18°C



	<b>pp65</b>	<b>bmlf1</b>	<b>hcv1</b>	<b>hcv2</b>	<b>hcv3</b>
Résolution (Å)	50-1.60 (1.64-1.60)	20-1.10 (1.20-1.10)	20-1.30 (1.40-1.30)	20-1.95 (2.0-1.95)	20-2.30 (2.40-2.30)
Rsym (%)	7.7 (33)	9.1 (31.3)	12.5 (29.3)	10.4 (32.6)	13.3 (31.4)
Complétude (%)	92.6 (74)	86.1 (75.1)	94.4 (79.1)	91.8 (85.5)	88.7 (75)
I/σ(I)	6.4 (2.8)	10.17 (2.99)	10.47 (3.18)	9.95 (3.05)	8.05 (4.32)
Redondance	2.8	3.8	4.6	3.8	3.9
B (Å <sup>2</sup> )	16	10	12	18	32
Groupe d'espace	<b>P2<sub>1</sub></b>	<b>P2<sub>1</sub></b>	<b>P2<sub>1</sub></b>	<b>P2<sub>1</sub></b>	<b>C2</b>
Paramètres de maille	a=53.2 Å b=80.8 Å c=56.5 Å β=112.7°	a=51.5 Å b=79.4 Å c=55.6 Å β=112.3°	a=53.1 Å b=80.8 Å c=56.2 Å β=112.3°	a=54.7 Å b=82.1 Å c=57.5 Å β=113.2°	a=98.4 Å b=38 Å c=120.3 Å β=91.1°
Ligne de lumière	ID14-eh2	ID14-eh3	ID14-eh2	BM30-A	ID14-eh2
Longueur d'onde	0.933 Å	0.931 Å	0.918 Å	0.979 Å	0.933 Å
R	0.188	0.183	0.180	0.183	0.228
Rfree	0.233	0.207	0.206	0.250	0.313
Rmsd sur les liaisons (Å)	1.226	1.160	1.167	1.480	1.612
Rmsd sur les angles (°)	0.009	0.007	0.007	0.017	0.014

*Figure.I.5. Tableau des statistiques cristallographiques des différents complexes peptide/HLA-A2m, et des statistiques issues des affinements en cours.*

CHAPITRE II

LES STRUCTURES

DE COMPLEXES

BINAIRES



## II.1. La résolution des structures de complexes binaires

Les différentes structures de complexe peptide/HLA-A2m ont été résolues par remplacement moléculaire avec le programme AMoRe (Navaza, 1994). La première structure de complexe binaire résolue est celle avec le peptide pp65, pour laquelle nous avons pris comme modèle la structure de HLA-A2 de plus haute résolution dans la PDB (code PDB : 1I4F)(Hillig et al., 2001). Puis pour les 4 autres complexes, nous avons utilisé la structure de pp65/HLA-A2m comme modèle. Les structures sont affinées avec le programme REFMAC (Murshudov et al., 1997) et construites à l'aide du logiciel O (Jones et al., 1991).

Pour chaque structure, les peptides ont pu être modélisés dans la densité électronique sans ambiguïté. La conformation générale du HLA-A2m est similaire. Les domaines  $\alpha 1$  et  $\alpha 2$ , soit les 180 premiers résidus des différents complexes obtenus, ont été structurellement alignés sur leur C $\alpha$  avec le programme ALIGN (Satow et al., 1986) (fig.II.1.a.). Après avoir superposé les domaines  $\alpha 1$  et  $\alpha 2$  du HLA-A2m, les rmsd sur les C $\alpha$  des peptides ont été calculés avec le programme ALIGN (fig.II.1.b.).

	<b>bmlf1</b>	<b>hcv1</b>	<b>hcv2</b>	<b>hcv3</b>		<b>bmlf1</b>	<b>hcv1</b>	<b>hcv2</b>	<b>hcv3</b>
<b>pp65</b>	0.45 Å	<b>0.12 Å</b>	0.33 Å	0.41 Å	<b>pp65</b>	1.35 Å	<b>0.49 Å</b>	1.18 Å	0.65 Å
<b>bmlf1</b>		<b>0.42 Å</b>	<b>0.60 Å</b>	<b>0.67 Å</b>	<b>bmlf1</b>		<b>1.00 Å</b>	<b>1.49 Å</b>	<b>1.22 Å</b>
<b>hcv1</b>			0.33 Å	0.39 Å	<b>hcv1</b>			1.42 Å	0.83 Å
<b>hcv2</b>				0.43 Å	<b>hcv2</b>				1.22 Å

*Figure.II.1.a. Alignement structural des domaines  $\alpha 1$  et  $\alpha 2$  des différents complexes peptides/HLA-A2m.*

*Figure.II.1.b. Alignement structural des peptides des différents complexes peptides/HLA-A2m.*

Les rmsd obtenus par la superposition des domaines  $\alpha 1$  et  $\alpha 2$  montrent que les structures de HLA-A2m en complexe avec hcv1 et pp65 sont proches (fig.II.1.a. en rouge). Il en est de même pour la superposition des deux peptides (fig.II.1.b. en rouge). Nous avons également remarqué que le complexe bmlf1/HLA-A2m avait la structure la plus différente.

Les rmsd entre les domaines  $\alpha 1$  et  $\alpha 2$  sont faibles (de 0.12 Å à 0.67 Å) ce qui montre la relative rigidité du CMH au niveau du sillon peptidique. Les rmsd calculés pour les peptides sont plus élevés que pour le HLA-A2m, avec des valeurs comprises entre 0.49 Å et 1.49 Å, indiquant une variation de conformation entre les peptides. Les rmsd des peptides sont corrélés avec les petites différences observées sur les domaines  $\alpha 1$  et  $\alpha 2$  du HLA-A2m.

## II.2. La structure des complexes peptide/HLA-A2m

Nous avons étudié cinq complexes binaires, dans le but de tester l'hypothèse d'un lien entre l'accessibilité au solvant et la protubérance des peptides présentés par une molécule CMH avec la sélection d'un répertoire « privé » ou « public » de  $L_T$ .

L'accessibilité au solvant, dans les complexes binaires, reflète la surface accessible au TCR lors de l'interaction TCR-pCMH. De même la protubérance du peptide est liée à sa surface accessible, puisque plus le peptide est protubérant plus sa surface accessible augmente.

Afin de déterminer si le lien existe entre ces deux paramètres et leur implication dans la sélection des  $L_T$ , nous avons analysé différents paramètres qui sont répertoriés dans les tableaux des figures II.2. a à e.

Les paramètres analysés dans notre étude sont :

\_ **la surface accessible au solvant**, calculée en absence de molécules d'eau et de symétries cristallines dans le cas du complexe hcv3. La surface accessible au solvant (et donc au TCR) a été calculée avec le programme NACCESS (Hubbard et al., 1993) avec comme paramètre une sonde de 1.4 Å.

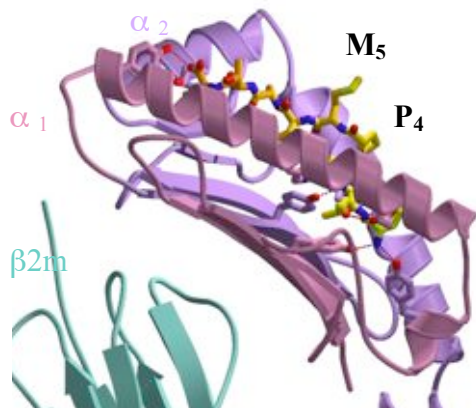
\_ **la conformation de chacun des acides aminés des peptides**, c'est-à-dire observer si les résidus généralement ancrés le sont (2, 3 et 9 pour les nonapeptides) et quels sont les résidus exposés au solvant.

\_ **la protubérance** de chaque résidu du peptide par rapport aux hélices- $\alpha$  des domaines  $\alpha 1$  et  $\alpha 2$  et également par rapport aux quatre autres peptides dont nous avons déterminé la structure.

\_ **le type d'interaction** formé entre le peptide et la molécule de HLA-A2m. En général les contacts sont de nature hydrophobes avec la molécule de CMH et les liaisons hydrogènes se font par la chaîne principale du peptide. Nous avons pu observer qu'un peptide formait des liaisons hydrogènes par ses chaînes latérales, nous avons nommé ces liaisons résidu-dépendante, puisqu'elles dépendent directement de la nature des résidus du peptide.

La structure des domaines  $\alpha 1$  et  $\alpha 2$  du HLA-A2m de nos structures présente un rmsd d'environ 0.5 Å avec les autres structures du HLA-A2m précédemment déterminées, ce qui est du même ordre que les rmsd présentés dans le tableau de la figure II.1.a. Les rmsd calculés sur les différentes structures de peptides complexés à la molécule HLA-A2m varie de 0.5 Å à 2 Å, ce qui également comparable avec les valeurs du tableau de la figure II.1.b.

Dans la représentation ci-dessous le sillon peptidique, formé par le domaine  $\alpha_1$  en rose (résidus 1 à 92) et le domaine  $\alpha_2$  en violet (résidus 93 à 180), contient le peptide représenté en bâton pour reconnaître les différents résidus, les atomes de carbones sont colorés graduellement de jaune à orange du N<sub>ter</sub> au C<sub>ter</sub> du peptide. Les résidus enfouis sont surlignés en jaune, ceux exposés en bleu.



Le peptide pp65

N	L	V	P	M	V	A	T	V
---	---	---	---	---	---	---	---	---

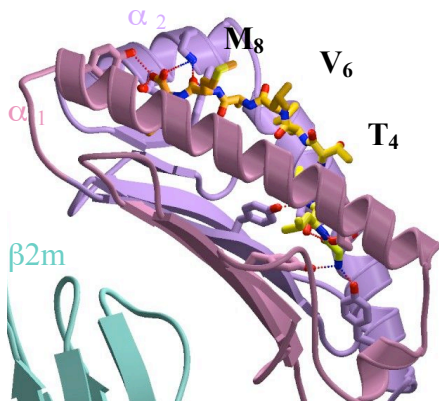
**Résidu protubérant** : M5

**Conformation du peptide** : enfouie

**Interaction résidu-dépendante** : aucune

**Surface accessible au solvant** : 359 Å<sup>2</sup>

Figure.II.2.a. Structure de pp65/HLA-A2m et récapitulatif des paramètres étudiés.



Le peptide bmlf1

G	L	C	T	L	V	A	M	L
---	---	---	---	---	---	---	---	---

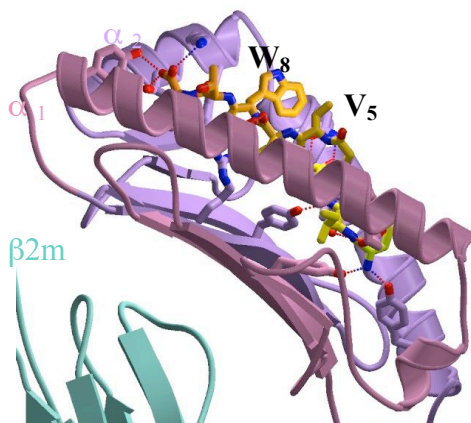
**Résidus protubérants** : T4, V6 et M8

**Conformation du peptide** : protubérant

**Interaction résidu-dépendante** : aucune

**Surface accessible au solvant** : 347 Å<sup>2</sup>

Figure.II.2.b. Structure de bmlf1/HLA-A2m et récapitulatif des paramètres étudiés.



Le peptide hcv1

C	I	N	G	V	C	W	T	V
---	---	---	---	---	---	---	---	---

**Résidus protubérants** : V5 et G4

**Conformation du peptide** : enfouie

**Interaction résidu-dépendante** : N3 avec Q155  
T8 avec D77

**Surface accessible au solvant** : 253 Å<sup>2</sup>

Figure.II.2.c. Structure de hcv1/HLA-A2m et récapitulatif des paramètres étudiés.

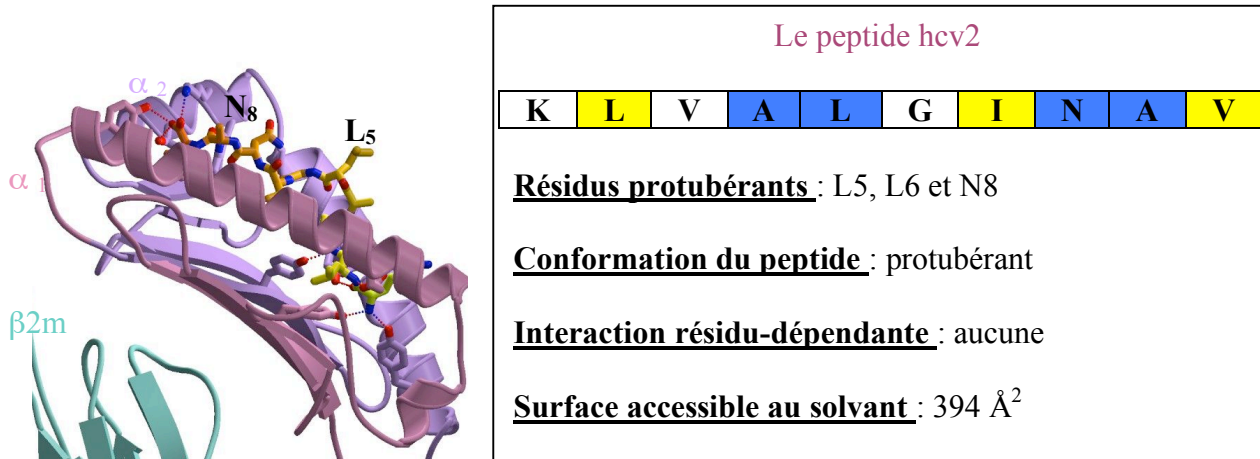


Figure.II.2.d. Structure de hcv2/HLA-A2m et récapitulatif des paramètres étudiés.

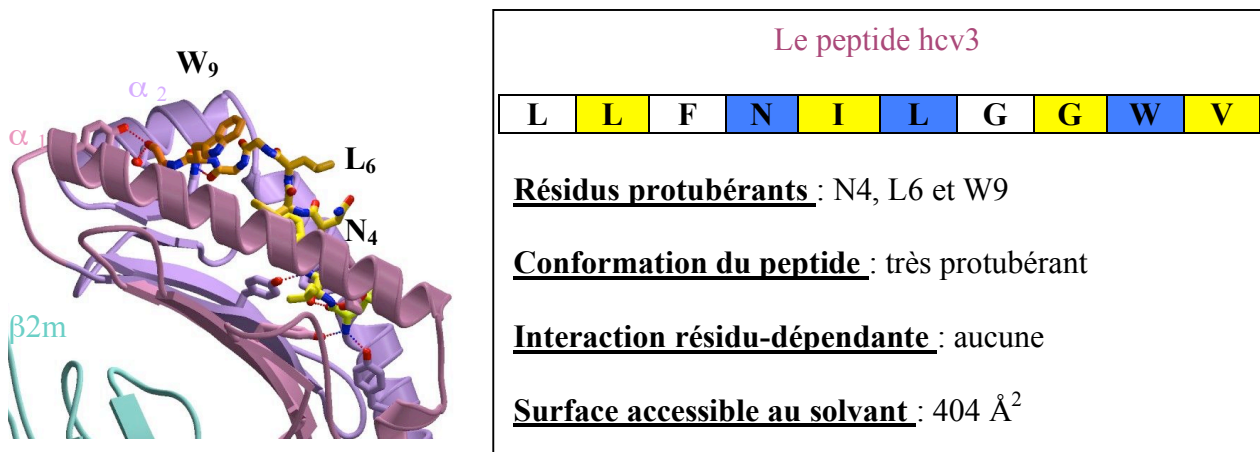


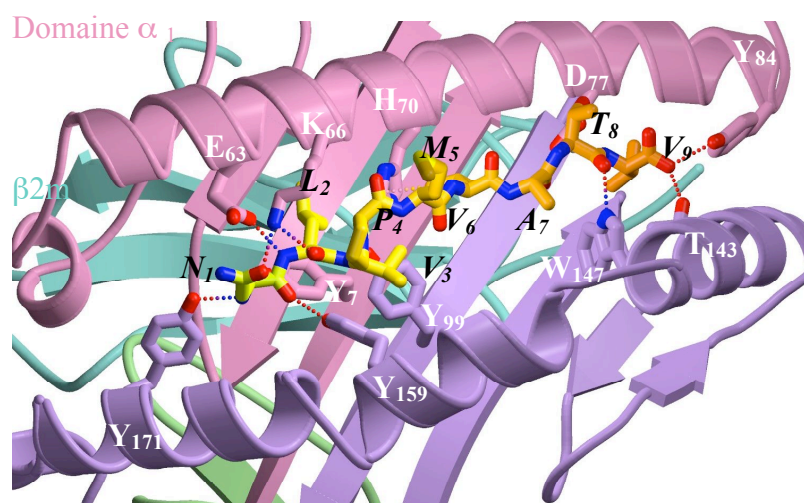
Figure.II.2.e. Structure de hcv3/HLA-A2m et récapitulatif des paramètres étudiés.

Le peptide le plus protubérant parmi les 5 dont nous disposons est hcv3, puis hcv2 et bmlf1 sont aussi protubérants l'un que l'autre, enfin pp65 et hcv1 sont enfouis. Une première remarque est que l'accessibilité au solvant du peptide et sa protubérance ne sont pas liées. Les peptides pp65 et bmlf1 exposent une surface comparable, alors que bmlf1 est plus protubérant que pp65.

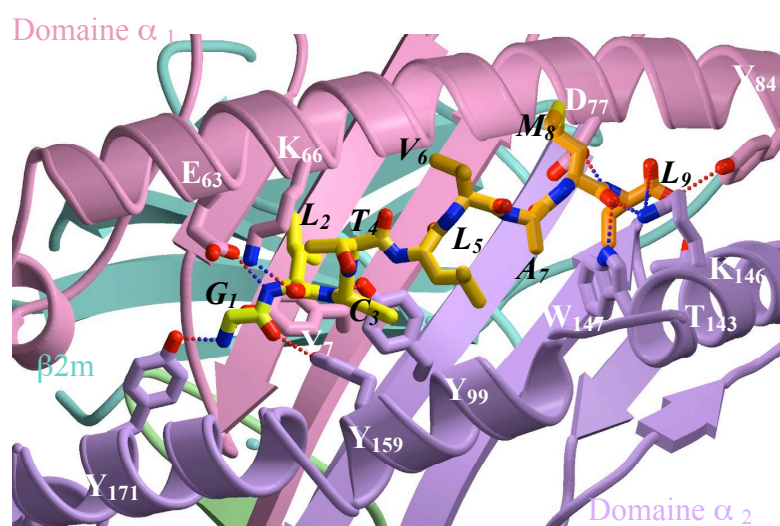
Les liaisons résidu-dépendantes n'existent que dans le peptide hcv1.

Les figures II.3. a à e suivantes montrent une vue du sillon peptidique pour chacune de nos structures.

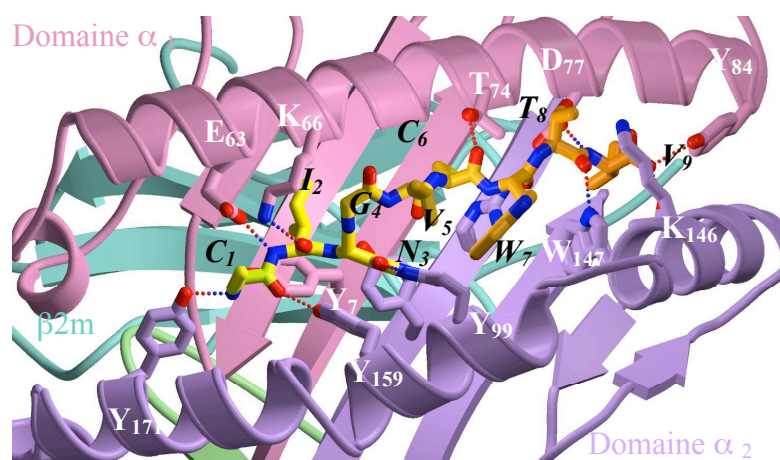




*Figure II.3.a. Structure de pp65/HLA-A2m vue de dessus de la poche peptidique*



*Figure II.3.b. Structure de bmlf1/HLA-A2m vue de dessus de la poche peptidique*



*Figure II.3.c. Structure de hcv1/HLA-A2m vue de dessus de la poche peptidique*



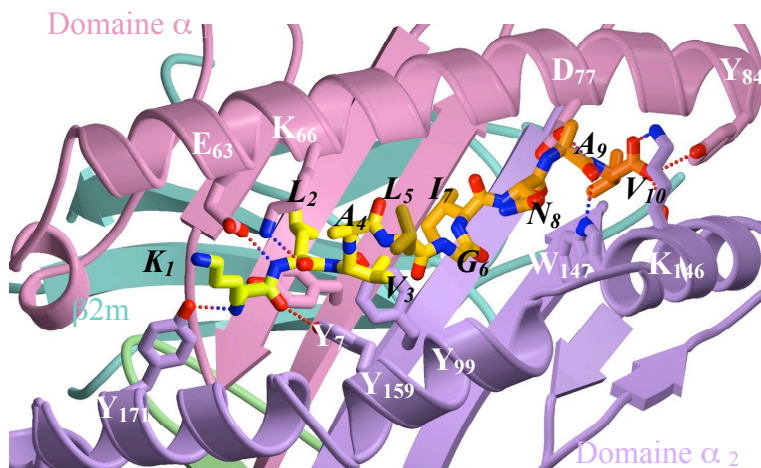


Figure II.3.d. Structure de hcv2/HLA-A2m vue de dessus de la poche peptidique.

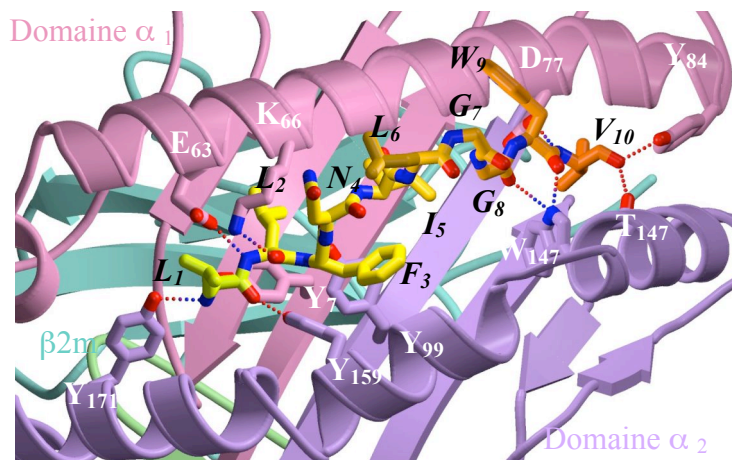


Figure II.3.e. Structure de hcv3/HLA-A2m vue de dessus de la poche peptidique

## II.3. L'analyse des structures

### II.3.1. Le postulat de départ est-il vérifié ?

L'hypothèse que nous nous proposons de tester est que : les complexes pCMH présentant des peptides enfouis sélectionnent un répertoire de  $L_T$  « public », ceux accessibles au solvant sélectionnent un répertoire de  $L_T$  « privé ».

Les complexes pCMH avec les peptides pp65 et bmlf1 sélectionnent un répertoire de  $L_T$  cytotoxiques restreint (fig.II.4.). Les peptides du virus de l'hépatite C complexés à la molécule HLA-A2m activent un répertoire de  $L_T$  divers.

Les individus séropositifs (séro+) sont ceux dont le système immunitaire a déjà été en contact avec le virus, les séronégatifs (séro-) n'ont pas rencontré le virus. Il est difficile de trouver des individus séronégatifs pour le virus d'Epstein-Barr, puisque 90% de la population possède des anti-corps anti-EBV.

Figure.II.4. Résumé de l'efficacité des tris multimériques réalisés avec nos 5 peptides

	BMLF1 (EBV)	PP65 (CMVH)	HCV1, HCV2 et HCV3
Individus Séro +	100%	100%	100%
Individus Séro -	Difficile	0%	100%
Individus A2 <sup>-</sup>	0%	0%	100%
Répertoire cytotoxiques (CD8 <sup>+</sup> )	Restreint Récurrence V $\alpha$ Récurrence CDR3	Restreint Récurrence V $\alpha$ Récurrence CDR3	Divers

Peu accessible au solvant ?

« peptide enfoui » ?

accessible au solvant ?

« peptide protubérant » ?

Avec les différentes structures que nous avons obtenues, nous avons observé que le peptide hcv1 déroge à l'hypothèse de départ. Il possède une surface accessible au solvant un tiers plus faible que les peptides pp65 et bmlf1 (fig.II.2. a à c). Pour les peptides hcv2 et hcv3 la surface accessible est plus élevée d'environ 10% par rapport à pp65 et bmlf1.

Le type de répertoire de L<sub>T</sub> CD8<sup>+</sup> activé par les 5 complexes pCMH étudiés ne dépend donc pas de la seule notion d'accessibilité au solvant du peptide. Le peptide hcv1 est un contre-exemple à l'hypothèse de départ. De même le caractère protubérant des peptides ne permet pas non plus d'expliquer la sélection du répertoire de L<sub>T</sub> activé. Le peptide le plus protubérant est hcv3 (cyan), puis hcv2 (rose) et bmlf1 (jaune) sont aussi protubérants l'un que l'autre, et enfin hcv1 (fushia) et pp65 (vert) sont relativement enfouis (fig.II.5. a à f).

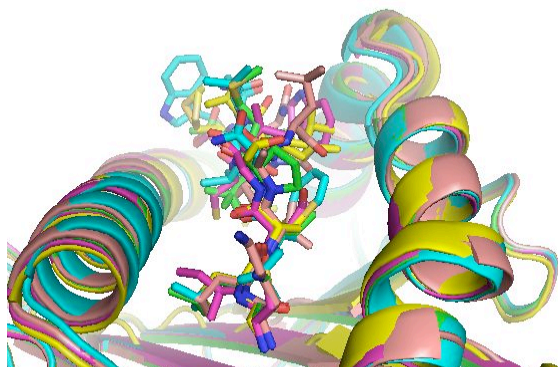


Figure.II.5.a. Superposition de nos 5 complexes peptide/HLA-A2m.

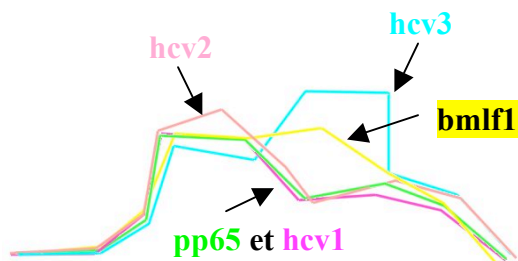
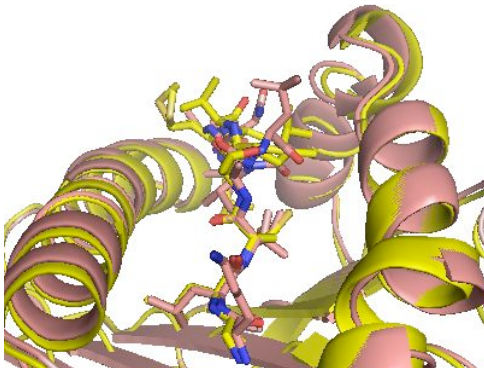
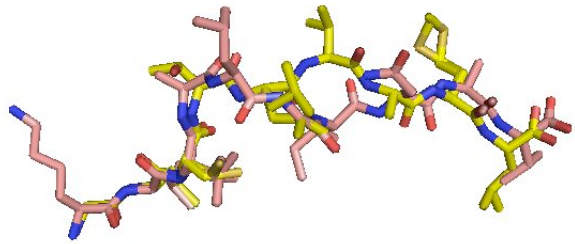


Figure.II.5.b. Superposition des structures de nos 5 peptides.

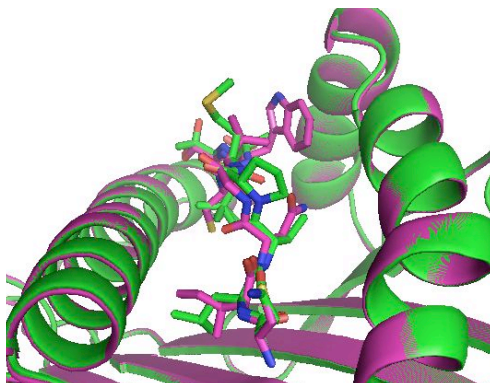
En cyan hcv3, en jaune bmlf1, en rose hcv2, en vert pp65 et en fushia hcv1.



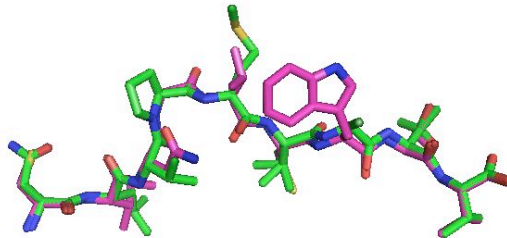
*Figure.II.5.c. Superposition des structures de HLA-A2m en complexe avec hcv2 (rose) et bmlf1 (jaune).*



*Figure.II.5.d. Superposition des structures de hcv2 (rose) et bmlf1 (jaune).*



*Figure.II.5.e. Superposition des structures de HLA-A2m en complexe avec pp65 (vert) et hcv1 (fushia).*



*Figure.II.5.f. Superposition des structures de pp65 (vert) et hcv1 (fushia).*

La protubérance et l'accessibilité au solvant des peptides ne nous permettent pas de prédire le type de répertoire de  $L_T$  cytotoxiques activés. La stabilité du peptide dans le sillon du CMH peut être un paramètre influençant le mécanisme de reconnaissance. Si le TCR interagit avec une surface pCMH dont le peptide est mobile le coût entropique va augmenter, défavorisant l'interaction. L'interaction entre le TCR et le pCMH pourrait également dépendre du type de liaison formée avec le peptide.

L'accessibilité au solvant et la protubérance du peptide peuvent influencer sa stabilité, mais ils ne peuvent à eux seuls expliquer la sélection du répertoire de  $L_T$ .

### [II.3.2. Une nouvelle observation : la stabilité du peptide](#)

Au cours de l'analyse de nos 5 structures de complexes binaires, nous nous sommes rendu compte que la densité de bmlf1 était faible malgré la très haute résolution de cette structure

(1.1 Å). Cette observation nous a mené à analyser plus en détail les cartes 2Fo-Fc de densité électronique. Nous nous sommes aperçus que les cartes de densité électronique pour les peptides pp65 et bmlf1 étaient faibles pour certains résidus, alors que celles des peptides de hcv décrivaient parfaitement le peptide. Ci-dessous la présentation des densités électroniques obtenues pour nos structures.

### II.3.2.1. La densité électronique de pp65

La densité électronique du peptide pp65, dont la structure a été obtenue à 1.60 Å, est discontinue pour la chaîne latérale de la méthionine 5 (fig.II.6.). Seules deux conformations avec un taux d'occupation de 30% pour chacune ont pu être modélisées. Donc le complexe pp65/HLA-A2m présente une surface mobile au TCR, avec la chaîne latérale de sa méthionine 5 non stabilisée.

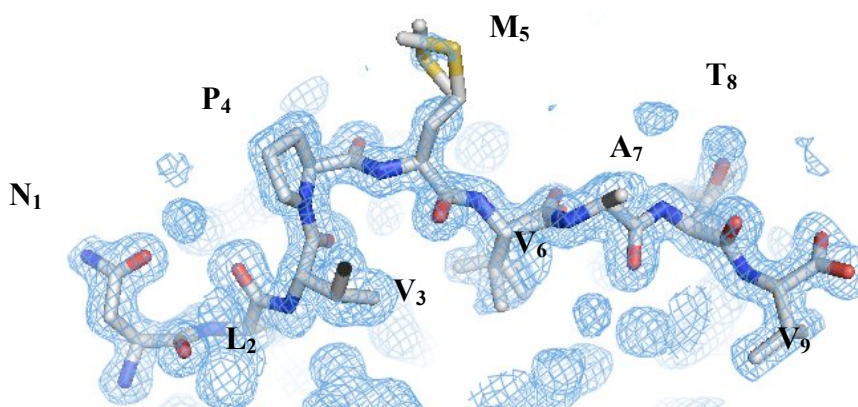
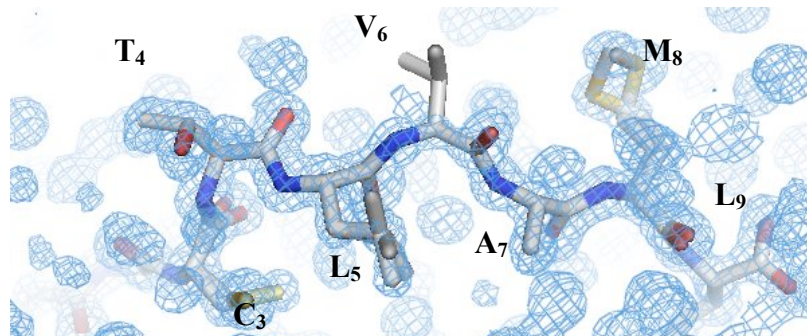


Figure.II.6. Densité électronique de pp65, carte 2Fo-Fc contourée à  $1\sigma$ .

### II.3.2.2. La densité électronique de bmlf1

Dans le cas du peptide bmlf1 en complexe avec HLA-A2m, la densité est discontinue sur la chaîne principale au niveau de la valine 6 (fig.II.7.). Après affinement il y a une densité résiduelle négative ( $-3\sigma$ ) autour du C $\alpha$  de la valine 6, indiquant probablement une seconde conformation, minoritaire, de la chaîne principale. Il n'y a pas de densité électronique autour de la chaîne latérale de la valine 6. De plus la carte 2Fo-Fc n'est pas continue pour le résidu leucine 5 malgré la résolution de 1.10 Å. Ce résidu présente deux conformations, une enfouie tournée vers le plancher du sillon peptidique et une exposée au solvant.





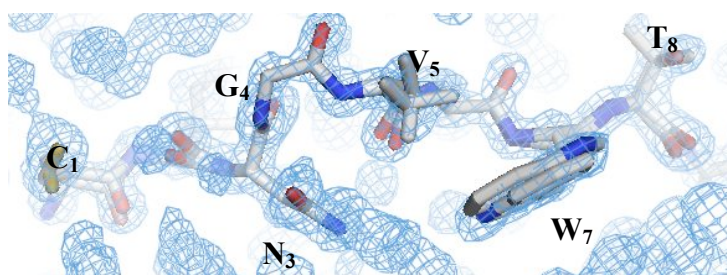
*Figure.II.7.* Densité électronique de *bmlf1*, carte *2Fo-Fc* contournée à  $1\sigma$ .

Le complexe *bmlf1*/HLA-A2m présente également une surface au TCR encore plus mobile que celle de *pp65*.

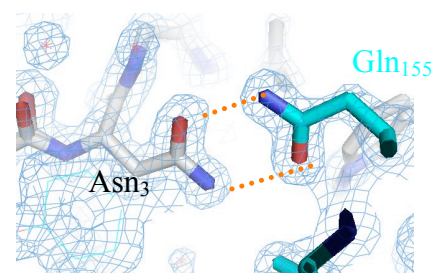
### II.3.2.3. La densité électronique des peptides *hcv1*, *hcv2* et *hcv3*

Dans le cas des trois peptides de VHC, les densités électroniques autour des peptides continues et ont permis de construire sans ambiguïté chaque peptide. Il semble que ces peptides soient plus stables que *pp65* et *bmlf1*.

Dans la structure de HLA-A2m avec *hcv1* le peptide possède de nombreuses doubles conformations (Cys1, Ile2, Val5, Trp7 et Thr8) qui ont pu être modélisées sans ambiguïté et sont en accord avec la densité électronique (fig.II.8.a.). Le peptide engage des liaisons hydrogènes entre la chaîne latérale de l'asparagine 3 et la glutamine 155 du HLA-A2m, favorisant la stabilité du peptide (fig.II.8.b.).

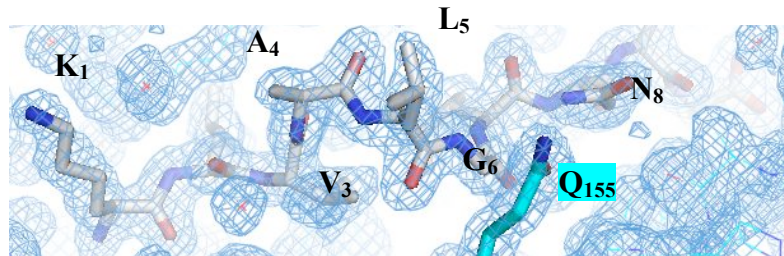


*Figure.II.8.a.* Densité électronique de *hcv1*, carte *2Fo-Fc* contournée à  $1\sigma$ .



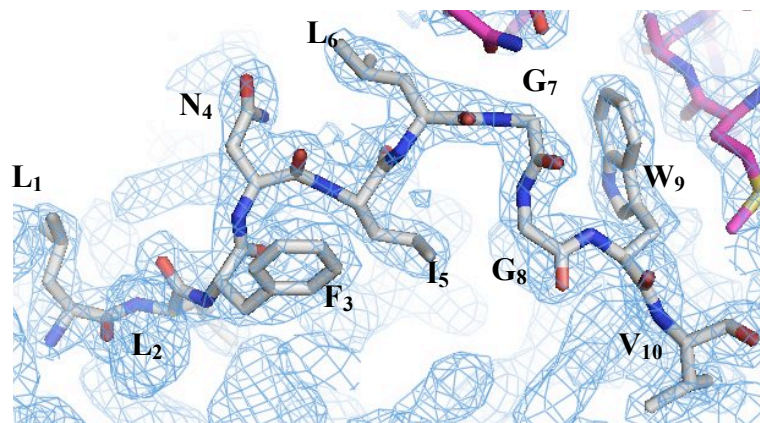
*Figure.II.8.b.* Agrandissement de la figure.II.12.a

Dans la structure du complexe hcv2/HLA-A2m, la densité électronique est continue autour du peptide. La leucine 5 même exposée au solvant est stable (fig.II.9.). Ce résidu hydrophobe est stabilisé par la glutamine 155 du HLA-A2m via des contacts de van der Waals.



*Figure.II.9. Densité électronique de hcv2, carte 2Fo-Fc contourée à 1  $\sigma$ .  
Le peptide est en blanc et le HLA-A2m en cyan.*

Dans la structure de HLA-A2m en complexe avec le peptide hcv3, obtenue à 2.3Å, la densité électronique autour du peptide est également continue. Comme indiqué en rose sur la figure.II.10. un symétrique cristallin (fushia) interagit avec le peptide à son extrémité C<sub>ter</sub>.



*Figure.II.10. Densité électronique de hcv3, carte 2Fo-Fc contourée à 1  $\sigma$ .  
Le peptide est en blanc et le symétrique cristallin en fushia.*

La présence de ce symétrique cristallographique pose problème dans l'interprétation de la stabilité du peptide hcv3. En effet nous ne pouvons pas savoir si la stabilité du peptide est intrinsèque ou si elle est dûe aux contacts cristallins. Un élément de réponse est apporté par la phénylalanine en position 3 du peptide hcv3. Ce résidu Phe3 est situé en N<sub>ter</sub> du peptide, où il n'y a pas de contact cristallin. La densité électronique de la Phe3 est continue dans hcv3. Au contraire dans le peptide SL9 du HIV (code PDB : 1T1W, pas de publication), qui possède également une Phe en position 3, elle a une densité électronique discontinue. La densité électronique est faible pour l'ensemble du peptide SL9 indiquant que sa structure est mobile dans le sillon du HLA-A2m. En comparant ces deux peptides nous pouvons donc supposer

que la Phe3 reflète la stabilité du peptide. Par conséquent le peptide hcv3 semble être stable et cela peut être même en absence de contacts cristallins, ceci en est une bonne indication.

Nous avons observé que les peptides de HCV sont plus stables que pp65 et bmlf1. Ceci se traduit au niveau de la continuité de la carte de densité électronique. Le niveau de stabilité n'a pu être apprécié que par l'observation des cartes de densité électronique.

Les facteurs d'agitation thermique B après affinement ne rendent pas compte de la flexibilité observée dans les cartes de densité électronique. Entre nos différentes structures, le facteur B du peptide est du même ordre que celui calculé sur les domaines  $\alpha 1$  et  $\alpha 2$  du HLA-A2m.

Ceci est dû à plusieurs facteurs: la méthode de résolution, les contraintes stéréochimiques imposées au facteur B, la résolution. Les chaînes latérales après affinement ne possèdent pas le même taux d'occupation, qui est corrélé avec le facteur B.

Lorsque nous avons comparé nos structures avec celles de pCMH disponibles dans la PDB, afin d'expliquer la conformation et la stabilité des peptides, nous voulions analyser le facteur B. Les logiciels d'affinement étant très différents, les facteurs B ne sont pas comparables.

L'analyse de nos structures avec celles des autres complexes peptide/HLA-A2m a donc été réalisée en observant les cartes de densité électronique.

### II.3.3. Comparaison avec les autres structures de peptide/HLA-A2m

Afin de comprendre le rôle de chaque résidu dans la conformation du peptide (ce qui est présenté au TCR) nous avons comparé nos structures avec celles disponibles dans la PDB. Pour analyser la conformation de chaque peptide il nous a fallu d'après les facteurs de structures disponibles calculer les cartes de densité électronique.

En tout il y existe 18 peptides différents en séquence et en structure, complexés à HLA-A2m dans la PDB. Les déviations standard entre ces peptides ont été calculées sur leur  $C\alpha$ , et les plus proches pour chacune de nos structures sont notées dans le tableau ci-dessous (fig.II.11.).

Cette analyse par superposition a été réalisée dans le but de comprendre ce qui peut stabiliser ou déstabiliser un peptide dans la gorge du HLA-A2m, c'est-à-dire ce qui lui donne sa conformation. Pour les peptides les plus proches les facteurs de structure étaient disponibles dans la PDB, nous avons donc pu calculer les cartes de densité électronique et les analyser pour chaque résidu.

	rmsd	
bmlf1 G L C T L V A M L	0.77Å	melanA Ag3 (1QEW) F L W G P R A L V
pp65 N L V P M V A T V	0.46Å	HIV SL9 (3A) (1S8D) S L A N T V A T L
hcv1 C I N G V C W T V	0.49Å	pp65 N L V P M V A T V
hcv2 K L V A L G I N A V	0.47Å	MelanA (decapeptide) (1JF1) E L A G I G I L T V
hcv3 L L F N I L G G W V	0.60Å	MelanA (nonapeptide) (1JHT) A L G I G I L T V

Figure.II.11. Complexes pCMH structurellement les plus proches de nos structures. En jaune sont surlignés les acides aminés conservés en séquence.

### II.3.3.1. Comparaison de bmlf1 et de melanA Ag3

Entre la structure de bmlf1 et celle du peptide Ag3 de melanA (1QEW, pas de publication), les C $\alpha$  des 4 premiers résidus et des trois derniers sont superposés (fig.II.12.a.). Dans le cas du peptide Ag3 de melanA la densité électronique est continue (ce peptide est stable) sauf pour la chaîne latérale de l'arginine 6 qui n'est pas stabilisée (fig.II.12.b.). La surface accessible au solvant de ce peptide (376Å<sup>2</sup>) est supérieure à celle de bmlf1 (347Å<sup>2</sup>). Il n'y a pas de liaison hydrogène entre les chaînes latérales du peptide de melanA et le HLA-A2m, seulement des contacts hydrophobes.

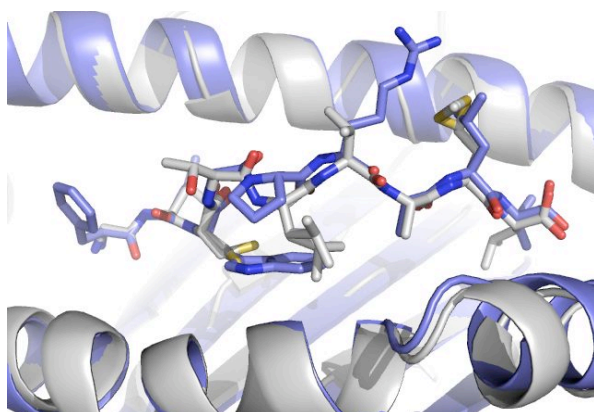


Figure.II.12.a. Superposition des structures de bmlf1 (blanc) avec melanA Ag3 (1QEW, bleu).

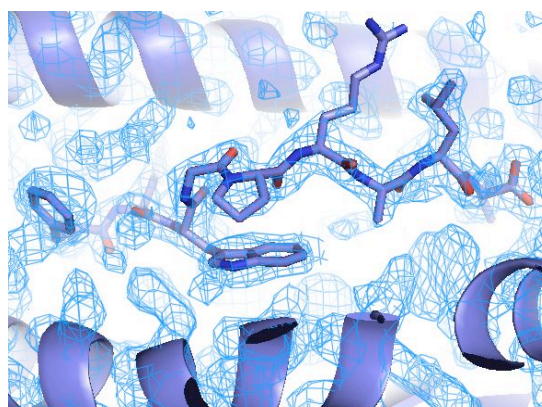
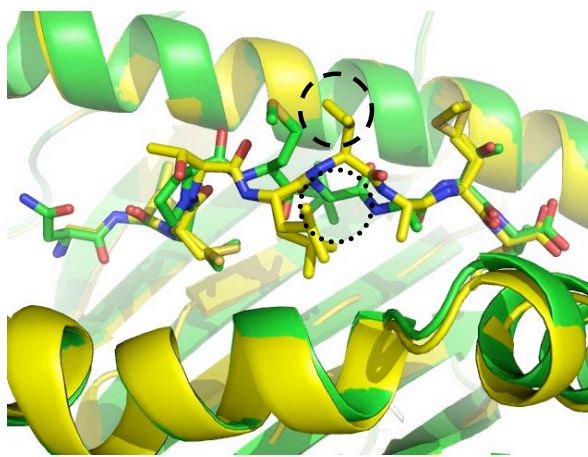


Figure.II.12.b. Densité électronique de melanA Ag3, carte 2Fo-Fc contourée à 1 $\sigma$ .



La stabilité du peptide de melanA pourrait être due à l'interaction entre les résidus Trp3 et Pro5 qui forment des contacts de van der Waals importants (fig.II.12.b.). Cette conformation étant certainement permise par la présence d'une glycine entre ces deux résidus. Alors que dans bmlf1 la leucine en position 5 est mobile, elle présente deux conformations dont la densité électronique est faible comme celle de la chaîne principale de la valine 6 (au niveau du C $\alpha$ ).

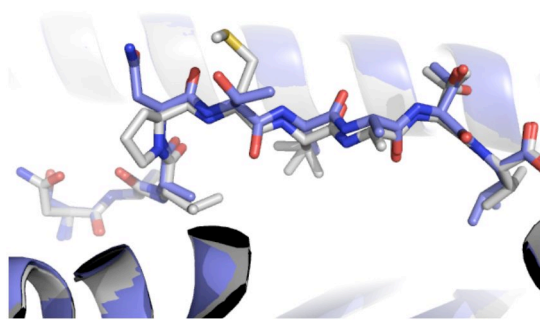


*Figure.II.13. Superposition des structures de HLA-A2m en complexe avec bmlf1 (jaune) et pp65 (vert).*

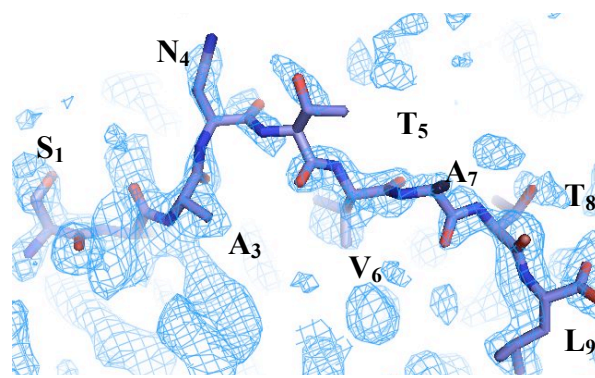
Le résidu en position 6 est soit enfoui et interagit avec le HLA-A2m soit exposé vers le TCR. Pour un résidu hydrophobe comme la valine de bmlf1, la conformation enfouie serait certainement plus favorable. Dans le peptide pp65 la valine 6 est enfouie et stable. Les extrémités des peptides pp65 et bmlf1 se superposent mais au niveau des résidus 5 et 6 il y a une grande différence de conformation (fig.II.13.). Au niveau du résidu 4 (P pour pp65 et T dans bmlf1) il y a un écart de 0.70 Å entre les C $\alpha$ . Nous avons émis l'hypothèse que la conformation enfouie de la valine 6 était peut-être due à la présence de la proline en position 4. Afin de stabiliser bmlf1 nous proposons un variant avec une proline en 4.

### II.3.3.2. Comparaison de pp65 et SL9

La plus proche structure de pp65 est celle d'un variant peptidique de SL9 du virus du SIDA (code PDB : 1S8D, pas de publication). Les séquences ont en commun 4 résidus et le chemin pris par les chaînes principales est le même



*Figure.II.14. Superposition des structures de HLA-A2m en complexe avec pp65 (blanc) et SL9 (bleu).*



*Figure.II.15. Densité électronique de SL9, carte 2Fo-Fc contournée à 1 $\sigma$ .*

(fig.II.14.). Le résidu SL9 a une densité électronique discontinue sur la chaîne principale, comme celles des 5 autres variants peptidiques de cet antigène (fig.II.15.).

Ce peptide SL9 étant instable ne nous donne pas beaucoup d'indication pour stabiliser pp65.

La seconde structure la plus proche de celle de pp65 est celle avec le peptide hcv1, qui lui est stable. Entre les structures de hcv1 et pp65 le rmsd est de 0.49 Å. Les peptides conservent en

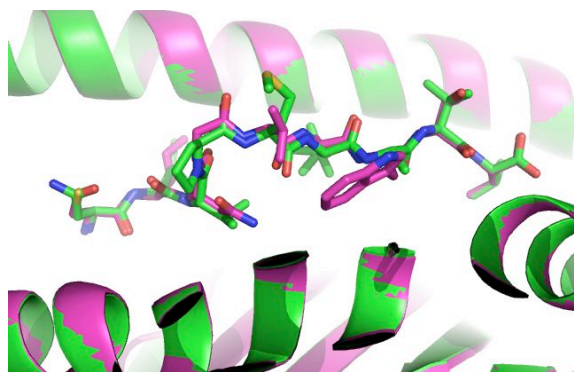


Figure.II.16.a. Superposition des structures de HLA-A2m en complexe avec pp65 (vert) et hcv1 (fushia).

séquence les deux résidus en C<sub>ter</sub> et suivent des chemins très proches au niveau des C $\alpha$  (fig.II.16.a.).

Nous avons analysé de plus près les structures, pour essayer de comprendre pourquoi hcv1 est plus stable que pp65, alors que les deux peptides prennent le même chemin dans le sillon peptidique. Il apparaît que le résidu cystéine 6 est plus enfoui dans hcv1 que dans pp65. La

poche qui accueille ce résidu ancre en position 6 des nonapeptides, est chargée positivement (2 His et 1 Arg) (fig.II.16.b.). Un résidu comme la valine, hydrophobe, n'est pas le plus favorable pour interagir avec les résidus basiques du HLA-A2m, au contraire de la cystéine qui est polaire.

Afin de stabiliser le peptide pp65 nous allons muter la valine 6 en cystéine (pour mimer hcv1), et en glycine qui prenant moins de place pourrait s'ancrer plus profondément et peut-être être plus stable.

Le peptide hcv1 interagit via les chaînes latérales de ses acides aminés avec le HLA-A2m, renforçant sûrement sa stabilité. Notamment le tryptophane en position 7 qui forme de nombreux contacts hydrophobes avec le HLA-A2m, doit donc être favorable pour la stabilité du peptide. Le peptide pp65 contient une alanine en position 7 dont le C $\alpha$  se

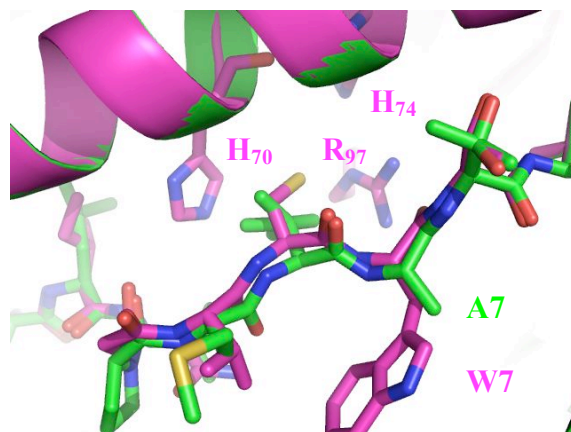


Figure.II.16.b. Agrandissement de la figure.II.16.a.

superpose avec celui de la Trp7 de hcv1. Si l'espace est rempli par un acide aminé avec une chaîne latérale plus encombrante, le peptide devrait être plus stable. Afin de vérifier cette hypothèse nous allons muter A7 en histidine dans pp65.

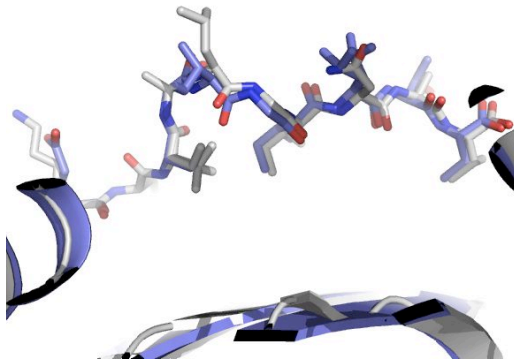
La mutation sera réalisée avec un résidu moins encombrant que le tryptophane, ceci afin d'essayer de stabiliser le peptide sans changer la surface reconnue par le TCR. Le but étant d'activer des populations de  $L_T$  à réactivité croisée, c'est-à-dire que les mêmes TCR reconnaissent à la fois pp65 sauvage et son variant. Si le variant est plus stable nous espérons observer une différence de fréquence dans la réponse immunitaire. Si la fréquence augmentait nous disposerions d'un variant peptidique déclenchant une meilleure réponse immunitaire que le peptide sauvage (très utile pour les vaccins).

#### II.3.3.3. Comparaison de hcv1 et pp65

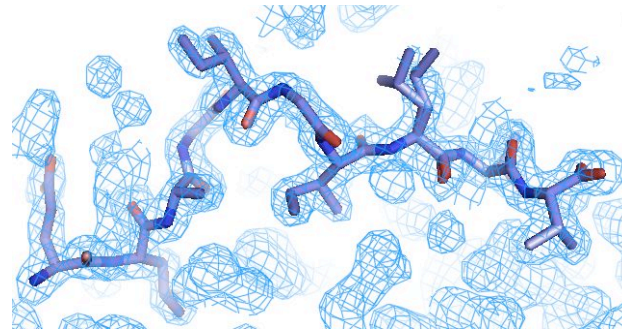
Comme nous l'avons indiqué dans le paragraphe III.3.3.2. le peptide hcv1 est stable. Cette stabilité étant peut être dû aux liaisons hydrogènes engagées par les chaînes latérales des acides aminés donc peptide-dépendantes. Notamment une double liaison hydrogène formée entre l'asparagine 3 du peptide hcv1 et la glutamine 155 du HLA-A2m. Dans l'hypothèse que ces liaisons résidu-dépendantes stabilisent le peptide, nous avons proposé de muter l'Asn3 en valine, pour déstabiliser le peptide hcv1.

#### II.3.3.4. Comparaison de hcv2 et melanA (décapeptide)

Les décapeptides complexés avec HLA-A2m sont moins nombreux dans la PDB (4 structures uniquement), il est donc plus difficile de comparer l'influence des différentes natures des résidus sur la conformation générale du peptide. Le peptide hcv2 prend le même chemin que le décapeptide de melanA (fig.II.17.a.). Le seul résidu à avoir une conformation différente est celui en position 5. Dans hcv2 c'est une leucine exposée au solvant et stable malgré sa nature hydrophobe. Dans le décapeptide de melanA c'est une isoleucine qui est plutôt enfouie contre l'hélice- $\alpha$  du domaine  $\alpha 2$ . Sa position peut être dû à la présence de la glycine en position 4 dans melanA. Ces deux peptides sont stables.

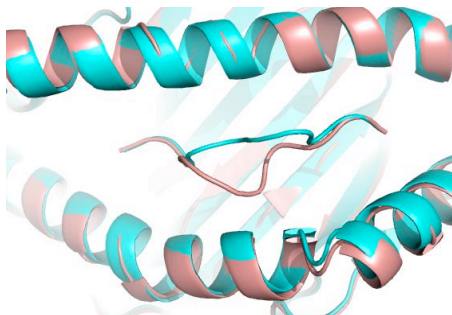


*Figure.II.17.a.* Superposition des structures de HLA-A2m en complexe avec hcv2 (blanc) et le décapeptide de melanA (bleu) (1JF1) (Sliz et al., 2001).

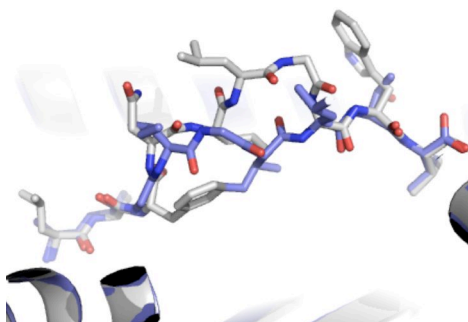


*Figure.II.17.b.* Densité électronique du décapeptide de melanA (1JF1), carte 2Fo-Fc contournée à  $1\sigma$ .

### II.3.3.5. Comparaison de hcv3 et melanA (nonapeptide)



*Figure.II.18.a.* Superposition des structures de HLA-A2m en complexe avec hcv2 (rose) et hcv3 (cyan).



*Figure.II.18.b.* Superposition des structures de HLA-A2m en complexe avec hcv3 (blanc) et melanA (bleu).

Bien que hcv3 soit un décapeptide, il prend un chemin proche de celui des nonapeptides. Par rapport au chemin pris par hcv2, celui de hcv3 est très différent (fig.II.18.a.). Celui de hcv2 est proche de l'hélice- $\alpha$  du domaine  $\alpha_2$ , ce qui laisse un vide entre le peptide et le domaine  $\alpha_1$ . Alors que dans le cas de hcv3, le chemin des C $\alpha$  du peptide est droit dans le sillon du HLA-A2m.

Le peptide le plus proche de hcv3 est celui du nonapeptide de melanA. Les trois résidus en N<sub>ter</sub> ainsi que les 3 en C<sub>ter</sub> sont parfaitement superposés entre ces deux peptides (fig.II.18.b.). Le peptide hcv3 contenant 10 acides aminés, la partie centrale de hcv3 est plus protubérante que celle du nonapeptide de melanA. Cela se traduit au niveau des surfaces accessibles au solvant qui est de 275Å<sup>2</sup> pour melanA est de 404 Å<sup>2</sup> pour hcv3.

Pour hcv2 et hcv3 nous n'avons, actuellement, pas proposé de variant peptidique pour les déstabiliser. Nous attendons de connaître la conséquence des 5 autres variants peptidiques, sur les nonapeptides pour savoir si nos hypothèses se confirment ou non.



CHAPITRE III

LES COMPLEXES

TERNAIRES



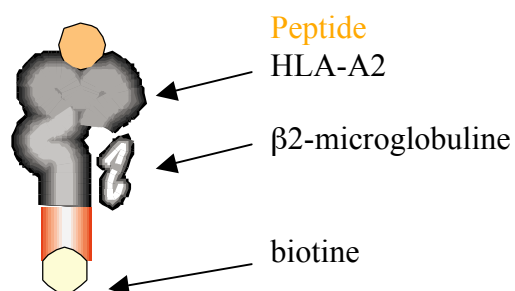


Afin de connaître les résidus des peptides impliqués dans la reconnaissance du TCR il est important de disposer des structures de complexes ternaires TCR-pCMH. Les structures de complexes ternaires nous permettront d'affiner notre analyse quant au rôle des conformations des peptides. Le TCR en interaction avec un complexe pCMH donné n'interagit pas forcément avec toute la surface accessible au solvant du peptide présenté par la molécule de CMH.

Dans l'objectif de déterminer la structure de complexes ternaires, il faut tout d'abord identifier les TCR activés par chacun de nos 5 complexes pCMH, puis il faut produire les TCR en quantité (quelques milligrammes) et pureté (supérieure à 90%) nécessaire à des essais de cristallisation.

### III.1. Le choix des TCR : les tris multimériques pCMH

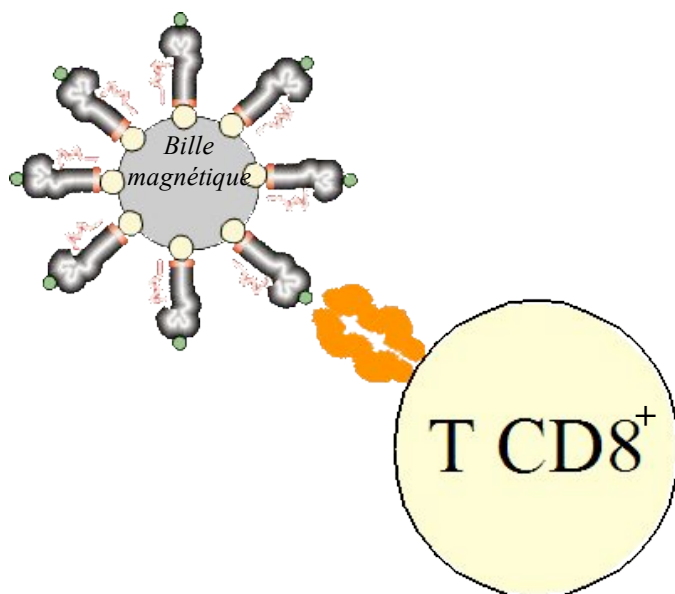
Afin de sélectionner les TCR qui reconnaissent un complexe pCMH donné, il faut « trier » dans le sang les  $L_T$   $CD8^+$  spécifiques de ce complexe pCMH. Pour cela nos collaborateurs de Nantes utilisent une méthode de tri utilisant des multimères de pCMH. La molécule de HLA-A2m contient au niveau de sa partie  $C_{ter}$  une étiquette permettant sa biotinylation (fig.III.1.).



*Figure.III.1. La molécule de HLA-A2m biotynilée.*

La biotine se fixe à la streptavidine, permettant ainsi d'accrocher plusieurs complexes pCMH (multimères) sur des billes magnétiques recouvertes de streptavidine (fig.III.2.).

Les  $L_T$   $CD8^+$  spécifiques sont récupérés par centrifugation grâce au poids des billes magnétiques. Ensuite les TCR sont séquencés ; ainsi nous pouvons savoir s'il y a récurrence des gènes V et J pour le domaine variable de la chaîne- $\alpha$ , et V, D et J pour le domaine variable de la chaîne- $\beta$ . Ceci nous permet de connaître le type de répertoire de  $L_T$  sélectionné pour un complexe pCMH donné.



*Figure.III.2. Schématisation du tri multimérique.*

S'il y a récurrence des domaines variables, la combinaison des différents CDR (régions du TCR interagissant avec le pCMH) sera restreinte, et le répertoire de  $L_T$  cytotoxique sélectionné sera plutôt de type « public ».

S'il n'y a pas récurrence des domaines variables des chaînes  $\alpha$  et  $\beta$ , alors les CDR seront divers et le répertoire de  $L_T$  « privé ».

### III.2. La production de TCR

Le premier TCR de notre étude est un TCR reconnaissant le complexe pp65/HLA-A2m. C'est le TCR immunodominant du répertoire « public » de  $L_T$  cytotoxique spécifique.

L'expression des chaînes du TCR se fait, comme pour celle du HLA-A2m, chez *E. coli*, donc en corps d'inclusion. Les deux chaînes sont exprimées séparément, puis elles sont réunies pour être correctement repliées.

Le rendement du repliement des chaînes est plus faible que celui du HLA, de l'ordre de 3% (contre 10% pour le HLA-A2m). Le faible rendement du repliement s'explique par le fait que normalement les chaînes du TCR sont glycosylées, ce qui semble augmenter la stabilité des domaines constants. Comme nous exprimons nos protéines dans *E. coli* il n'y a pas de glycosylation. Afin de favoriser l'interaction entre les deux chaînes du TCR, une étiquette de

complémentarité de charges est placée en C<sub>ter</sub> de chaque chaîne, avec une cystéine de chaque côté pour former un pont disulfure inter-chaîne.

EN DGGG CK (chaîne- $\alpha$ )



QD RGGG CD (chaîne- $\beta$ )

Dans notre cas, la chaîne- $\alpha$  contient à son extrémité C<sub>ter</sub> une étiquette 6 histidines. Ceci afin de séparer d'éventuels homodimères de chaîne  $\alpha\alpha$  ou  $\beta\beta$ . L'étiquette ne peut être située qu'au niveau de la partie C<sub>ter</sub>, car la partie N<sub>ter</sub> est le domaine variable et donc contient les domaines CDR qui interagissent avec le pCMH. Une l'étiquette est en N<sub>ter</sub> pourrait gêner l'interaction avec le pCMH

La chaîne- $\alpha$  contient 210 résidus (23350 Da) auxquels s'ajoutent 8 résidus pour l'étiquette histidine. La chaîne- $\beta$  contient 251 résidus (27886 Da). Soit un poids total pour le TCR correctement replié d'environ 52 kDa.

Différentes constructions de la chaîne- $\alpha$  ont été testées. La première construction contient l'étiquette histidine en C<sub>ter</sub>. L'étiquette 6 histidines pose parfois problème au moment de la cristallogénèse, car c'est une région située à l'extrémité de la protéine et qui est souvent flexible. Elle peut donc gêner la formation de l'empilement cristallin. Pour cela nous avons décidé de cliver cette étiquette, tout d'abord par une coupure ménagée à la carboxipeptidase, puis en introduisant un site de coupure à la thrombine pour obtenir une solution homogène.

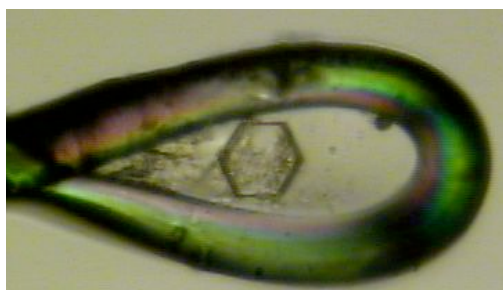
### III.3. Les premiers essais de cristallisation

Les premiers essais de cristallisation ont été réalisés autour des conditions de complexes ternaires avec HLA-A2m décrit dans la littérature (Stewart-Jones et al., 2003, Ding et al., 1998). Ils ont été réalisés avec un TCR contenant l'étiquette 6 histidines sans site de coupure (produit à Nantes). Dans ces premiers tests nous avons observé beaucoup de précipités, d'où l'idée par la suite d'introduire un site de coupure pour ne plus avoir d'étiquette 6 histidines qui peut poser problème pour la cristallogénèse.

Avant d'utiliser une nouvelle construction avec un site de coupure, nous avons voulu cliver ces histidines par la carboxypeptidase comme cela est décrit par P. Anton van der Merwe et ses collaborateurs (van der Merwe et al., 1997). La carboxypeptidase est une enzyme qui clive les acides aminés un par un en C<sub>ter</sub>. Cette coupure par la carboxypeptidase est suivie par spectrométrie de masse, afin de contrôler son avancement. Le problème de cette technique est le manque de reproductibilité et d'homogénéité. Avec cet échantillon de TCR sans étiquette 6 histidines, nous avons testé les 6 kits de cristallisation de Hampton Research disponibles au robot de cristallisation.

### III.3.1. Les premiers cristaux

Au bout de quelques jours de tout petits cristaux hexagonaux sont apparus. Ils ont été reproduits en goutte suspendue dans des conditions similaires à celles obtenues dans les kits de cristallisation (fig.III.3.). Nous avons pu enregistrer un jeu de données complet sur la ligne de lumière ID23-eh1 (fig.III.4.).



*Figure.III.3. Les premiers cristaux obtenus en présence de complexe ternaire (20µm).*

*Condition de cristallisation : sulfate d'ammonium à 1.6 M, acide citrique à pH 5 à 0.1M 3 mg.ml<sup>-1</sup> de TCRpCMH.*

Résolution (Å)	20-2.40 (2.50-2.40)
R <sub>sym</sub> (%)	12 (37.2)
Complétude (%)	96.1 (83.3)
I/σ(I)	10.13 (2.49)
Redondance	4.5
B (Å <sup>2</sup> )	32 Å <sup>2</sup>
Groupe d'espace	<b>P3<sub>2</sub>21</b>
Paramètres de maille	a=b=70.36 Å, c=96.52 Å
Ligne de lumière	ID23-eh1
Longueur d'onde	1 Å

*Figure.III.4. Statistiques cristallographiques du cristal de 20 µm obtenus en présence de complexe pCMH.*

### III.3.2. Que contiennent ces cristaux ?

Les dimensions de la maille donnent un  $V_m$  de  $0.7 \text{ \AA}^3 \cdot \text{Da}^{-1}$  pour le complexe TCRpCMH,  $1.3 \text{ \AA}^3 \cdot \text{Da}^{-1}$  pour le TCR et  $1.5 \text{ \AA}^3 \cdot \text{Da}^{-1}$  pour le complexe pCMH. Les limites de  $V_m$  pour les cristaux de macromolécules sont de  $1.7$  à  $4 \text{ \AA}^3 \cdot \text{Da}^{-1}$  (Matthews, 1968). Par conséquent il ne peut y avoir une molécule entière dans ces cristaux.

Nous avons donc testé le remplacement moléculaire, avec le logiciel AMoRe, avec différents domaines du HLA-A2m :  $\alpha 1$ ,  $\alpha 2$ , la  $\beta 2$ -microglobuline. Comme il n'y avait pas de solution contrastée en remplacement moléculaire, nous avons testé des domaines du TCR, avec comme modèle le TCR de séquence la plus proche (code PDB : 2BNU) (Chen et al., 2005).

La recherche avec un monomère de domaine variable de la chaîne- $\alpha$  donne un résultat en remplacement moléculaire, indiquant la présence de deux  $V\alpha$ . Avec deux monomères de  $V\alpha$  nous avons obtenu une solution de remplacement moléculaire avec un facteur de corrélation de 40.2% sur les facteurs de structure. Nous avons donc cristallisé un homodimère de domaine variable de la chaîne- $\alpha$ , dont la séquence a ensuite été confirmée par spectrométrie de masse.

L'origine de ce domaine  $V\alpha$  est encore incertaine, il semblerait que nous ayons un problème d'expression, puisque nous exprimons le  $V\alpha$  et la chaîne- $\alpha$  entière. Le problème au moment de la purification est que le domaine  $V\alpha$  interagit avec le TCR. Pour l'instant nous n'avons pas réussi à le séparer du TCR malgré plusieurs étapes de purification.

La chaîne- $\alpha$  contient l'étiquette 6 histidines au niveau de sa partie  $C_{\text{ter}}$ , donc sur colonne d'affinité pour l'étiquette 6 histidines, le  $V\alpha$  ne devrait pas être retenu. Cette étape de purification est faite après l'étape de repliement, le  $V\alpha$  peut donc s'accrocher sur le TCR ou sur des chaînes- $\alpha$  non appariées avec les chaînes- $\beta$ . Nous avons également essayé de séparer le  $V\alpha$  du TCR par un gradient de sel, mais le pI étant proche de celui du TCR cela a échoué. Puisque le  $V\alpha$  cristallise dans des conditions contenant du sulfate d'ammonium nous avons essayé de le séparer du TCR par précipitation au sulfate d'ammonium, mais en vain. Jusqu'à présent toutes nos expériences n'ont pas permis de séparer le domaine  $V\alpha$  du TCR.

Actuellement nous essayons de purifier la chaîne- $\alpha$  en conditions dénaturantes, en espérant que déplié le domaine  $V\alpha$  n'interagit pas avec la chaîne- $\alpha$ .

### III.3.3. La structure de l'homodimère de $V\alpha$

La structure montre un homodimère tête-bêche des domaines  $V\alpha$ , ce qui est différent de ce qui a été observé par Fields B.A. en 1995 pour un autre homodimère de  $V\alpha$  (code PDB : 1TVD) (Fields et al., 1995). La position tête-bêche est étonnante puisque cela ne mime pas le TCR en mettant du même côté leurs domaines CDR (fig.III.5). Cette structure est en cours d'analyse.

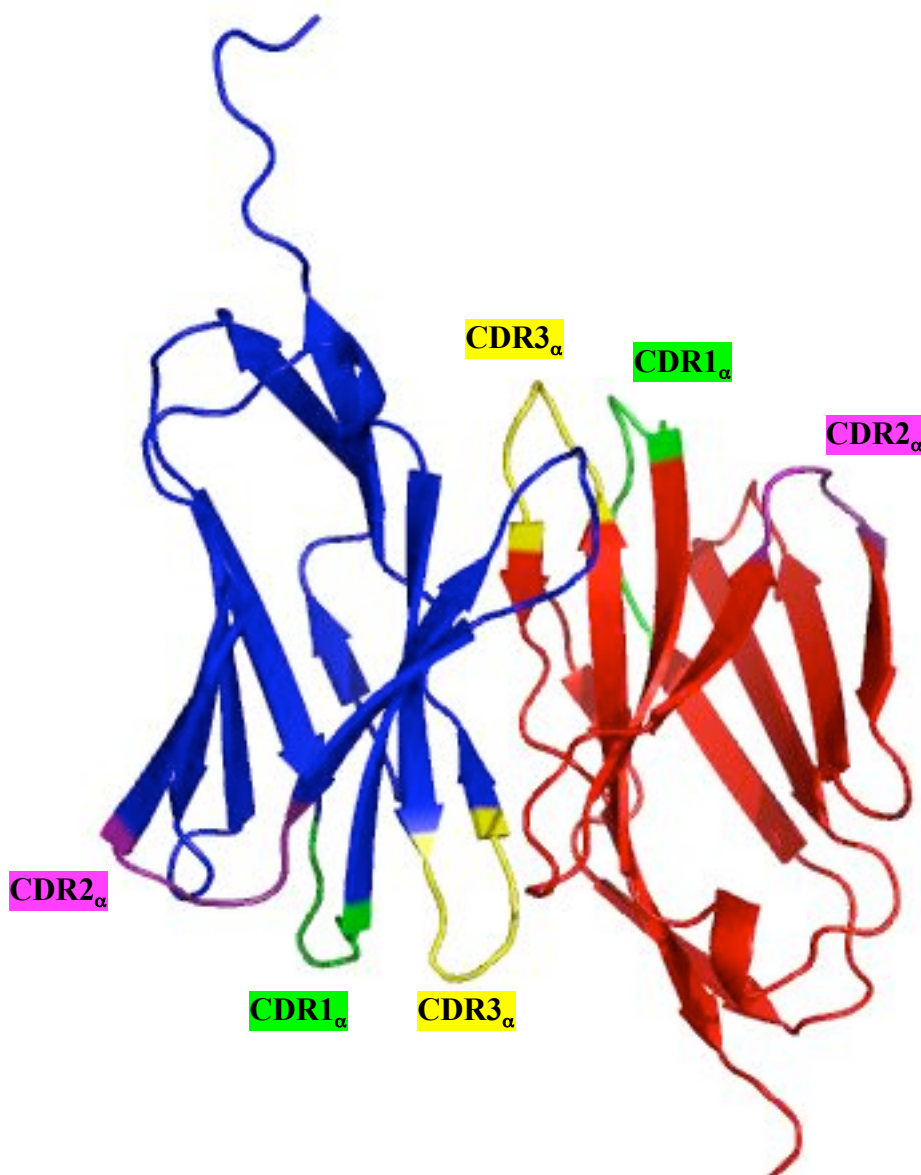


Figure.III.5. Structure de l'homodimère du domaine variable de la chaîne- $\alpha$  du TCR.

CONCLUSION

ET

PERSPECTIVES





## *Conclusion et perspectives du projet immunologie*

Les résultats structuraux sur les différents complexes pCMH que nous avons obtenu nous ont permis d'émettre une nouvelle hypothèse quant au mécanisme de sélection du type de répertoire de  $L_T$  activé. En effet il semblerait que la stabilité du peptide puisse guider la sélection de  $L_T$ .

Actuellement au sein du laboratoire nous avons trois objectifs principaux. Tout d'abord résoudre le problème de production et de purification de la chaîne- $\alpha$  du TCR, afin de réaliser de nouveaux essais de cristallisation. Nous cherchons de nouvelles conditions de cristallisation pour le complexe hcv3/HLA-A2m afin de ne plus avoir de contact cristallin au niveau du sillon peptidique. Ainsi nous pourrions analyser le peptide hcv3 et voir s'il est stable sans symétrie. Nous sommes également en train de produire la molécule de HLA-A2m en complexe avec les 5 variants peptidiques que nous avons proposés, afin de tester nos hypothèses sur la stabilité des peptides.

Sur la base de notre analyse structurale, nos collaborateurs ont commencé une étude de tri multimérique avec deux peptides, le premier possédant une densité électronique continue (la transcriptase inverse du HIV, code PDB 1I1F) (Kirksey et al., 1999) et le second a une densité discontinue (SL9). Ceci dans l'objectif de voir si sur la base de la qualité de la densité électronique et donc de la stabilité de l'antigène dans le sillon peptidique, la sélection de répertoire privé ou public de  $L_T$   $CD8^+$  est prévisible.

De plus les analyses de tris avec les peptides de HCV sont en cours. Il faudra séquencer les TCR immunodominant pour pouvoir les produire par la suite en quantité suffisante pour la cristallisation.

Il a été déterminé par M.J. Pittet en 2006 (Pittet et al., 2006) l'existence de peptides à réactivité croisée avec bmlf1 (Clute et al., 2005) et hcv1 (Wedemeyer et al., 2001). Ces deux peptides sont très intéressants pour nous sur le plan clinique mais également sur le plan structural. Les complexes pCMH avec ces deux peptides vont être produits et analysés parallèlement.

Il faut relier toutes les observations structurales avec celles de tris multimériques dans un premier temps. Puis essayer de voir si d'après ces résultats nous pouvons essayer de comprendre par quoi est guidée la réponse immunitaire lors d'une infection par le virus de l'hépatite C. Et comprendre le lien entre nos observations et le développement de la maladie chez les patients.

## *Conclusion et perspectives du projet immunologie*

À très long terme le but sera de pouvoir synthétiser des peptides efficaces dans le cadre d'une thérapie vaccinale. De nombreuses questions restent encore à résoudre, et comme nous l'avons vu le chemin vers les structures ternaires est bien plus complexe que celui des complexes binaires.

# BIBLIOGRAPHIE



A

- Andrade, M.A., Petosa, C., O'Donoghue, S.I., Muller, C.W. and Bork, P. (2001) Comparison of ARM and HEAT protein repeats. *J Mol Biol*, **309**, 1-18.
- Aravind, L., Leipe, D.D. and Koonin, E.V. (1998) Toprim--a conserved catalytic domain in type IA and II topoisomerases, DnaG-type primases, OLD family nucleases and RecR proteins. *Nucleic Acids Res*, **26**, 4205-4213.

B

- Bach, M.A., Hoffenbach, A., Lagrange, P.H., Wallach, D. and Cottenot, F. (1983) Mechanisms of T-cell unresponsiveness in leprosy. *Ann Immunol (Paris)*, **134D**, 75-84.
- Bentley, G.A., Boulot, G., Karjalainen, K. and Mariuzza, R.A. (1995a) Crystal structure of the beta chain of a T cell antigen receptor. *Science*, **267**, 1984-1987.
- Bentley, G.A., Boulot, G. and Mariuzza, R.A. (1995b) The structure of the antigen-binding site of immunoglobulins and T-cell receptors. *Res Immunol*, **146**, 277-290.
- Bergerat, A., Demassy, B., Gadelle, D., Varoutas, P.C., Nicolas, A. and Forterre, P. (1997) An atypical topoisomerase II from archaea with implications for meiotic recombination. *Nature*, **386**, 414-417.
- Berghuis, A.M., Lee, E., Raw, A.S., Gilman, A.G. and Sprang, S.R. (1996) Structure of the GDP-Pi complex of Gly203-->Ala g1alpha1: a mimic of the ternary product complex of gamma-catalyzed GTP hydrolysis. *Structure*, **4**, 1277-1290.
- Bevan, M.J. (1977) In a radiation chimaera, host H-2 antigens determine immune responsiveness of donor cytotoxic cells. *Nature*, **269**, 417-418.
- Bevan, M.J. (1984) High determinant density map may explain the phenomenon of alloreactivity. *Immunol. Today*, **5**, 128-130.
- Bjorkman, P.J., Saper, M.A., Samraoui, B., Bennett, W.S., Strominger, J.L. and Wiley, D.C. (1987) Structure of the human class I histocompatibility antigen, HLA-A2. *Nature*, **329**, 506-512.
- Bongrand, P. and Malissen, B. (1998) Quantitative aspects of T-cell recognition: from within the antigen-presenting cell to within the T cell. *Bioessays*, **20**, 412-422.
- Borbulevych, O.Y., Baxter, T.K., Yu, Z., Restifo, N.P. and Baker, B.M. (2005) Increased immunogenicity of an anchor-modified tumor-associated is due to the enhanced stability of the peptide/MHC complex : implications for vaccine design. *J. Immunology*, **174**, 4812-4820.

## Bibliographie

- Boriack-Sjodin, P.A., Margarit, S.M., Bar-Sagi, D. and Kuriyan, J. (1998) The structural basis of the activation of Ras by Sos. *Nature*, **394**, 337-343.
- Bourne, H.R., Sanders, D.A. and McCormick, F. (1991) The GTPase superfamily: conserved structure and molecular mechanism. *Nature*, **349**, 117-127.
- Bouvier, M., Guo, H.-C., Smith, K.J. and Wiley, D.C. (1998) Crystal structures of HLA-A\*0201 complexed with antigenic peptides with either the amino- or carboxyl-terminal group substituted by a methyl group. *PROTEINS : Structure, Function, and Genetics*, **33**, 97-106.
- Bowman, G.D., O'Donnell, M. and Kuriyan, J. (2004) Structural analysis of a eukaryotic sliding DNA clamp-clamp loader complex. *Nature*, **429**, 724-730.
- Brunger, A.T. (1993) Assessment of phase accuracy by cross validation: the free R value. Methods and applications. *Acta Crystallogr D Biol Crystallogr*, **49**, 24-36.
- C**
- (1994) The CCP4 suite: programs for protein crystallography. *Acta Crystallogr D Biol Crystallogr*, **50**, 760-763.
- Champoux, J.J. (2001) DNA topoisomerases: structure, function, and mechanism. *Annu Rev Biochem*, **70**, 369-413.
- Chen, J.-L., Stewart-Jones, G., Bossi, G., Lissin, N.M., Wooldridge, L., Choi, E.M.L., Held, G., Dunbar, P.R., Esnouf, R.M., Sami, M., Boultier, J.M., Rizkallah, P., Renner, C., Sewell, A., van der Merwe, P.A., Jackobsen, B.K., Griffiths, G. and Jones. (2005) Structural and kinetic basis for Heightened Immunogenicity of T Cell Vaccines. *J. Exp. Med.*, 201, 1243.
- Chen, X., Court, D.L. and Ji, X. (1999) Crystal structure of ERA: a GTPase-dependent cell cycle regulator containing an RNA binding motif. *Proc Natl Acad Sci U S A*, **96**, 8396-8401.
- Chermann, J.C., Barre-Sinoussi, F., Dauguet, C., Brun-Vezinet, F., Rouzioux, C., Rozenbaum, W. and Montagnier, L. (1983) Isolation of a new retrovirus in a patient at risk for acquired immunodeficiency syndrome. *Antibiot Chemother*, **32**, 48-53.
- Cingolani, G., Petosa, C., Weis, K. and Muller, C.W. (1999) Structure of importin-beta bound to the IBB domain of importin-alpha. *Nature*, **399**, 221-229.
- Clute, S.C., Watkin, L.B., Cornberg, M., Naumov, Y.N., Sullivan, J.L., Luzuriaga, K., Welsh, R.M. and Selin, L.K. (2005) Cross-reactive influenza virus-specific CD8+ T cells contribute to lymphoproliferation in Epstein-Barr virus-associated infectious mononucleosis. *J Clin Invest*, **115**, 3602-3612.
- Coleman, D.E. and Sprang, S.R. (1999) Structure of Galpha1.GppNHp, autoinhibition in a Galpha protein-substrate complex. *J Biol Chem*, **274**, 16669-16672.

## Bibliographie

Corbett, K.D. and Berger, J.M. (2003) Structure of the topoisomerase VI-B subunit: implications for type II topoisomerase mechanism and evolution. *Embo J*, **22**, 151-163.

Coux, O., Tanaka, K. and Goldberg, A.L. (1996) Structure and functions of the 20S and 26S proteasomes. *Annu Rev Biochem*, **65**, 801-847.

## D

Dahlmann, B., Kopp, F., Kuehn, L., Hegerl, R., Pfeifer, G. and Baumeister, W. (1991) The multicatalytic proteinase (prosome, proteasome): comparison of the eukaryotic and archaeobacterial enzyme. *Biomed Biochim Acta*, **50**, 465-469.

Dandekar, T., Snel, B., Huynen, M. and Bork, P. (1998) Conservation of gene order: a fingerprint of proteins that physically interact. *Trends Biochem Sci*, **23**, 324-328.

De la Fortelle, E.d.l. and Bricogne, G. (1997) Maximum-Likelihood Heavy-Atom Parameter Refinement for Multiple Isomorphous Replacement and Multiwavelength Anomalous Diffraction Methods. *Methods in Enzymology*, edited by Carter, J., C.W., and R.M. Sweet, Academic Press, New-York, **276**, 472-494.

DeTitta, G.T. and Luft, J.R. (1995) Rate of water equilibration in vapor-diffusion crystallization: dependence on the residual pressure of air in the vapor space. *Acta Crystallogr D Biol Crystallogr*, **51**, 786-791.

Ding, Y.H., Smith, K.J., Garboczi, D.N., Utz, U., Biddison, W.E. and Wiley, D.C. (1998) Two human T cell receptors bind in a similar diagonal mode to the HLA-A2/Tax peptide complex using different TCR amino acids. *Immunity*, **8**, 403-411.

## E

Egea, P.F., Shan, S.O., Napetschnig, J., Savage, D.F., Walter, P. and Stroud, R.M. (2004) Substrate twinning activates the signal recognition particle and its receptor. *Nature*, **427**, 215-221.

Ehnebom, L. and Pedersen, B.F. (1992) Molecular and crystal structure of a lanthanide complex, DyDTPA-BMA hydrate. *Acta chemica Scandinavica*, **46**, 126-130.

Ennis, F.A., Martin, W.J. and Verbonitz, M.W. (1977) Hemagglutinin-specific cytotoxic T-cell response during influenza infection. *J Exp Med*, **146**, 893-898.



**F**

- Fields, B.A., Ober, B., Malchiodi, E.L., Lebedeva, M.I., Braden, B.C., Ysern, X., Kim, J.K., Shao, X., Ward, E.S. and Mariuzza, R.A. (1995) Crystal structure of the V alpha domain of a T cell antigen receptor. *Science*, **270**, 1821-1824.
- Frech, M., Darden, T.A., Pedersen, L.G., Foley, C.K., Charifson, P.S., Anderson, M.W. and Wittinghofer, A. (1994) Role of glutamine-61 in the hydrolysis of GTP by p21H-ras: an experimental and theoretical study. *Biochemistry*, **33**, 3237-3244.
- Freymann, D.M., Keenan, R.J., Stroud, R.M. and Walter, P. (1999) Functional changes in the structure of the SRP GTPase on binding GDP and Mg<sup>2+</sup>+GDP. *Nat Struct Biol*, **6**, 793-801.

**G**

- Garboczi, D.N., Ghosh, P., Utz, U., Fan, Q.R., Biddison, W.E. and Wiley, D.C. (1996) Structure of the complex between human T-cell receptor, viral peptide and HLA-A2. *Nature*, **384**, 134-141.
- Garcia, K.C., Degano, M., Stanfield, R.L., Brunmark, A., Jackson, M.R., Peterson, P.A., Teyton, L. and Wilson, I.A. (1996) An alphabeta T cell receptor structure at 2.5 Å and its orientation in the TCR-MHC complex. *Science*, **274**, 209-219.
- Garcia, K.C., Teyton, L. and Wilson, I.A. (1999) Structural basis of T cell recognition. *Annu Rev Immunol*, **17**, 369-397. Garman, E.F. and Schneider, T.R. (1997) Macromolecular Cryocrystallography. *J. Appl. Cryst*, **30**, 211\_237.
- Garman, E.F. and Schneider, T.R. (1997) Macromolecular Cryocrystallography. *J. Appl. Cryst*, **30**, 211\_237.
- Gigant, B., Curmi, P.A., Martin-Barbey, C., Charbaut, E., Lachkar, S., Lebeau, L., Siavoshian, S., Sobel, A. and Knossow, M. (2000) The 4 Å X-ray structure of a tubulin:stathmin-like domain complex. *Cell*, **102**, 809-816.
- Girard, E., Stelter, M., Vicat, J. and Kahn, R. (2003) A new class of lanthanide complexes to obtain high-phasing-power heavy-atom derivatives for macromolecular crystallography. *Acta Crystallogr D Biol Crystallogr*, **59**, 1914-1922.
- Gueguen, Y., Rolland, J.L., Lecompte, O., Azam, P., Le Romancer, G., Flament, D., Raffin, J.P. and Dietrich, J. (2001) Characterization of two DNA polymerases from the hyperthermophilic euryarchaeon *Pyrococcus abyssi*. *European Journal of Biochemistry*, **268**, 5961-5969.

H

- Hartung, F. and Puchta, H. (2001) Molecular characterization of homologues of both subunits A (SPO11) and B of the archaeobacterial topoisomerase 6 in plants. *Gene*, **271**, 81-86.
- Henneke, G., Gueguen, Y., Flament, D., Azam, P., Querellou, J., Dietrich, J., Hubscher, U. and Raffin, J.P. (2002) Replication factor C from the hyperthermophilic archaeon *Pyrococcus abyssi* does not need ATP hydrolysis for clamp-loading and contains a functionally conserved RFC PCNA-binding domain. *J Mol Biol*, **323**, 795-810.
- Hillig, R.C., Coulie, P.G., Stroobant, V., Saenger, W., Ziegler, A. and Hulsmeijer, M. (2001) High-resolution structure of HLA-A\*0201 in complex with a tumour-specific antigenic peptide encoded by the MAGE-A4 gene. *J Mol Biol*, **310**, 1167-1176.
- Holm, L. and Sander, C. (1993) Protein structure comparison by alignment of distance matrices. *J Mol Biol*, **233**, 123-138.
- Huang, L., Hofer, F., Martin, G.S. and Kim, S.H. (1998) Structural basis for the interaction of Ras with RalGDS. *Nat Struct Biol*, **5**, 422-426.
- Hubbard, S J and Thornton, J M (1993) NACCESS, computer program, Department of Biochemistry and Molecular Biology, University College London.
- Hudrisier, D. and Gairin, J.E. (1996) La pharmacologie du récepteur des lymphocytes T et de ses ligands. *MS synthèse*, **12**, 1198.

I

- Ihara, K., Muraguchi, S., Kato, M., Shimizu, T., Shirakawa, M., Kuroda, S., Kaibuchi, K. and Hakoshima, T. (1998) Crystal structure of human RhoA in a dominantly active form complexed with a GTP analogue. *J Biol Chem*, **273**, 9656-9666.

J

- Jackson, T.M., Zaman, S.N. and Huq, F. (1977) T and B rosetting lymphocytes in the blood of smallpox patients. *Am J Trop Med Hyg*, **26**, 517-519.
- Jancarik, J. and Kim, S.-H. (1991) Sparse matrix sampling : a screening method for crystallization proteins. *J. Appl. Cryst*, **24**, 409-411.
- John, J., Frech, M. and Wittinghofer, A. (1988) Biochemical properties of Ha-ras encoded p21 mutants and mechanism of the autophosphorylation reaction. *J Biol Chem*, **263**, 11792-11799.

## Bibliographie

- Jones, T.A., Zou, J.Y., Cowan, S.W. and Kjeldgaard. (1991) Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallogr A*, **47** ( Pt 2), 110-119.
- K**
- Kaas. and Lefranc M.-P. (2005) T cell receptor/peptide/MHC molecular characterization and standardized pMHC contact sites in IMGT/3Dstructure-DB. *In Silico Biology*, **5**, 505-528
- Kabsch, W. (1993) Automatic processing of rotation diffraction data from crystals of initially unknown symmetry and cell constants. *J Appl Crystallogr*, **26**, 795-800.
- Kack, H., Gibson, K.J., Lindqvist, Y. and Schneider, G. (1998) Snapshot of a phosphorylated substrate intermediate by kinetic crystallography. *Proc Natl Acad Sci U S A*, **95**, 5495-5500.
- Kappler, J.W., Roehm, N. and Marrack, P. (1987) T cell tolerance by clonal elimination in the thymus. *Cell*, **49**, 273-280.
- Kaziro, Y., Itoh, H., Kozasa, T., Nakafuku, M. and Satoh, T. (1991) Structure and function of signal-transducing GTP-binding proteins. *Annu Rev Biochem*, **60**, 349-400.
- Keenan, R.J., Freymann, D.M., Stroud, R.M. and Walter, P. (2001) The signal recognition particle. *Annu Rev Biochem*, **70**, 755-775.
- Khan, A.R., Baker, B.M., Ghosh, P., Biddison, W.E. and Wiley, D.C. (2000) The structure and stability of an HLA-A\*0201/Octamerix Tax peptide complex with an empty conserved peptide-N-terminal binding site. *J. Immunology*, **164**, 6398-6405.
- Kirksey, T.J., Pogue-Caley, R.R., Frelinger, J.A. and Collins, E.J. (1999) The structural basis for the increased immunogenicity of two HIV-reverse transcriptase peptide variant/class I major histocompatibility complexes. *J Biol Chem*, **274**, 37259-37264.
- Kjeldgaard, M., Nissen, P., Thirup, S. and Nyborg, J. (1993) The crystal structure of elongation factor EF-Tu from *Thermus aquaticus* in the GTP conformation. *Structure*, **1**, 35-50.
- Krell, K. and Eisenberg, M.A. (1970) The purification and properties of dethiobiotin synthetase. *J Biol Chem*, **245**, 6558-6566.
- Kuglstatter, A., Oubridge, C. and Nagai, K. (2002) Induced structural changes of 7SL RNA during the assembly of human signal recognition particle. *Nat Struct Biol*, **9**, 740-744.
- Kukimoto-Niino, M., Murayama, K., Inoue, M., Terada, T., Tame, J.R., Kuramitsu, S., Shirouzu, M. and Yokoyama, S. (2004) Crystal structure of the GTP-binding protein Obg from *Thermus thermophilus* HB8. *J Mol Biol*, **337**, 761-770.

## Bibliographie

### L

- Langen, R., Schweins, T. and Warshel, A. (1992) On the mechanism of guanosine triphosphate hydrolysis in ras p21 proteins. *Biochemistry*, **31**, 8691-8696.
- Lapouge, K., Smith, S.J., Walker, P.A., Gamblin, S.J., Smerdon, S.J. and Rittinger, K. (2000) Structure of the TPR domain of p67phox in complex with Rac.GTP. *Mol Cell*, **6**, 899-907.
- Lefranc and Lefranc. (2001) The T cell receptor FactsBook, Academic Press.
- Leipe, D.D., Wolf, Y.I., Koonin, E.V. and Aravind, L. (2002) Classification and evolution of P-loop GTPases and related ATPases. *J Mol Biol*, **317**, 41-72.
- Lembo, F., Pero, R., Angrisano, T., Vitiello, C., Iuliano, R., Bruni, C.B. and Chiariotti, L. (2003) MBDin, a novel MBD2-interacting protein, relieves MBD2 repression potential and reactivates transcription from methylated promoters. *Mol Cell Biol*, **23**, 1656-1665.
- Leslie, A.G.W. (1991) in Crystallographic Computing (eds Moras, D., Podjany, A. D. & Thierry, J. C.). *Oxford Univ. Press*, 50-60.
- Levdikov, V.M., Blagova, E.V., Brannigan, J.A., Cladiere, L., Antson, A.A., Isupov, M.N., Seror, S.J., Wilkinson, A.J (2004) The Crystal Structure of YloQ, a Circularly Permuted GTPase Essential for Bacillus Subtilis Viability. *J.Mol.Biol*, **340**, 767-782.
- Lo, D., Ron, Y. and Sprent, T. (1986) Induction of MHC-restricted specificity and tolerance in the thymus. *Immunol Res*, **5**, 221-232.

### M

- Macrae, I.J. and Doudna, J.A. (2005) Ro's role in RNA reconnaissance. *Cell*, **121**, 495-496.
- Maegley, K.A., Admiraal, S.J. and Herschlag, D. (1996) Ras-catalyzed hydrolysis of GTP: a new perspective from model studies. *Proc Natl Acad Sci U S A*, **93**, 8160-8166.
- Marinsek, N., Barry, E.R., Makarova, K.S., Dionne, I., Koonin, E.V. and Bell, S.D. (2006) GINS, a central nexus in the archaeal DNA replication fork. *EMBO Rep*.
- Marshall, M.S. (1993) The effector interactions of p21ras. *Trends Biochem Sci*, **18**, 250-254.
- Mason, D. (1998) A very high level of crossreactivity is an essential feature of the T- cell receptor. *Immunol Today*, **19**, 395-404.
- Matte-Tailliez, O., Zivanovic, Y. and Forterre, P. (2000) Mining archaeal proteomes for eukaryotic proteins with novel functions: the PACE case. *Trends Genet*, **16**, 533-536.

## Bibliographie

- Matthews, B.W. (1968) Solvent content of protein crystals. *J Mol Biol*, **33**, 491-497.
- Matzinger, P. and Bevan, M.J. (1977) Hypothesis: why do so many lymphocytes respond to major histocompatibility antigens? *Cell Immunol*, **29**, 1-5.
- McPherson, A. (1990) Current approaches to macromolecular crystallization. *Eur J Biochem*, **189**, 1-23.
- Milburn, M.V., Tong, L., deVos, A.M., Brunger, A., Yamaizumi, Z., Nishimura, S. and Kim, S.H. (1990) Molecular switch for signal transduction: structural differences between active and inactive forms of protooncogenic ras proteins. *Science*, **247**, 939-945.
- Morris, R.J., Perrakis, A. and Lamzin, V.S. (2002) ARP/wARP's model-building algorithms. I. The main chain. *Acta Crystallogr D Biol Crystallogr*, **58**, 968-975.
- Murshudov, G.N., Vagin, A.A. and Dodson, E.J. (1997) Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallogr D Biol Crystallogr*, **53**, 240-255.

## N

- Natale, D.A., Shankavaram, U.T., Galperin, M.Y., Wolf, Y.I., Aravind, L. and Koonin, E.V. (2000) Towards understanding the first genome sequence of a crenarchaeon by genome annotation using clusters of orthologous groups of proteins (COGs). *Genome Biol*, **1**, RESEARCH0009.
- Navaza, J. (1994) AMoRE : an automated package for molecular replacement. *Acta Cryst A*, **50**, 157-163.
- Niemann, H.H., Knetsch, M.L., Scherer, A., Manstein, D.J. and Kull, F.J. (2001) Crystal structure of a dynamin GTPase domain in both nucleotide-free and GDP-bound forms. *Embo J*, **20**, 5813-5821.
- Nitta, M., Saijo, M., Kodo, N., Matsuda, T., Nakatsu, Y., Tamai, H. and Tanaka, K. (2000) A novel cytoplasmic GTPase XAB1 interacts with DNA repair protein XPA. *Nucleic Acids Res*, **28**, 4212-4218.

## O

- Oliva, M.A., Cordell, S.C. and Lowe, J. (2004) Structural insights into FtsZ protofilament formation. *Nat Struct Mol Biol*, **11**, 1243-1250.
- Olsen, G.J. and Woese, C.R. (1997) Archaeal genomics: an overview. *Cell*, **89**, 991-994.

P

- Padmanabhan, S. and Freymann, D.M. (2001) The conformation of bound GMPPNP suggests a mechanism for gating the active site of the SRP GTPase. *Structure*, **9**, 859-867.
- Paduch, M., Jelen, F. and Otlewski, J. (2001) Structure of small G proteins and their regulators. *Acta Biochim Pol*, **48**, 829-850.
- Pahler, A., Smith, J.L. and Hendrickson, W.A. (1990) A probability representation for phase information from multiwavelength anomalous dispersion. *Acta Crystallogr A*, **46 ( Pt 7)**, 537-540.
- Pape, T., Meka, H., Chen, S., Vicentini, G., van Heel, M. and Onesti, S. (2003) Hexameric ring structure of the full-length archaeal MCM protein complex. *EMBO Rep*, **4**, 1079-1083.
- Pasqualato, S. and Cherfils, J. (2005) Crystallographic evidence for substrate-assisted GTP hydrolysis by a small GTP binding protein. *Structure*, **13**, 533-540.
- Paul, W.E. and Seder, R.A. (1994) Lymphocyte responses and cytokines. *Cell*, **76**, 241-251.
- Perutz, M.F. (1956) Isomorphous replacement and phase determination in non centrosymmetric space groups. *Acta Cryst*, **9**, 867-873.
- Perutz, M.F. (1960) Structure of hemoglobin. *Brookhaven Symp Biol*, **13**, 165-183.
- Pittet, M.J., Gati, A., Le Gal, F.A., Bioley, G., Guillaume, P., de Smedt, M., Plum, J., Speiser, D.E., Cerottini, J.C., Dietrich, P.Y., Romero, P. and Zippelius, A. (2006) Ex vivo characterization of allo-MHC-restricted T cells specific for a single MHC-peptide complex. *J Immunol*, **176**, 2330-2336.
- Polekhina, G., Thirup, S., Kjeldgaard, M., Nissen, P., Lippmann, C. and Nyborg, J. (1996) Helix unwinding in the effector region of elongation factor EF-Tu-GDP. *Structure*, **4**, 1141-1151.
- Prakash, B., Praefcke, G.J., Renault, L., Wittinghofer, A. and Herrmann, C. (2000a) Structure of human guanylate-binding protein 1 representing a unique class of GTP-binding proteins. *Nature*, **403**, 567-571.
- Prakash, B., Renault, L., Praefcke, G.J., Herrmann, C. and Wittinghofer, A. (2000b) Triphosphate structure of guanylate-binding protein 1 and implications for nucleotide binding and GTPase mechanism. *Embo J*, **19**, 4555-4564.
- Prive, G.G., Milburn, M.V., Tong, L., de Vos, A.M., Yamaizumi, Z., Nishimura, S. and Kim, S.H. (1992) X-ray crystal structures of transforming p21 ras mutants suggest a transition-state stabilization mechanism for GTP hydrolysis. *Proc Natl Acad Sci U S A*, **89**, 3649-3653.

R

- Radfar, R., Leaphart, A., Brewer, J.M., Minor, W., Odom, J.D., Dunlap, R.B., Lovell, C.R. and Lebioda, L. (2000) Cation binding and thermostability of FTHFS monovalent cation binding sites and thermostability of N10-formyltetrahydrofolate synthetase from *Moorella thermoacetica*. *Biochemistry*, **39**, 14481-14486.
- Ramakrishnan, C. and Ramachandran, G.N. (1965) Stereochemical criteria for polypeptide and protein chain conformations. II. Allowed conformations for a pair of peptide units. *Biophys J*, **5**, 909-933.
- Robinson, V.L., Hwang, J., Fox, E., Inouye, M. and Stock, A.M. (2002) Domain arrangement of Der, a switch protein containing two GTPase domains. *Structure*, **10**, 1649-1658.
- Rudolph, M.G., Standfield, R.L. and Wilson, I.A. (2006) How TCRs bind MHCs, Peptides, and Coreceptors. *Annu. Rev. Immunol.*, **24**, 419-466.
- Ruzheinikov, S.N., Das, S.K., Sedelnikova, S.E., Baker, P.J., Artymiuk, P.J., Garcia-Lara, J., Foster, S.J. and Rice, D.W. (2004) Analysis of the open and closed conformations of the GTP-binding protein YsxC from *Bacillus subtilis*. *J Mol Biol*, **339**, 265-278.

S

- Sakai, N., Yao, M., Itou, H., Watanabe, N., Yumoto, F., Tanokura, M. and Tanaka, I. (2001) The three-dimensional structure of septum site-determining protein MinD from *Pyrococcus horikoshii* OT3 in complex with Mg-ADP. *Structure*, **9**, 817-826.
- Sandalova, T., Schneider, G., Kack, H. and Lindqvist, Y. (1999) Structure of dethiobiotin synthetase at 0.97 Å resolution. *Acta Crystallogr D Biol Crystallogr*, **55**, 610-624.
- Satow, Y., Cohen, G.H., Padlan, E.A. and Davies, D.R. (1986) Phosphocholine binding immunoglobulin Fab McPC603. An X-ray diffraction study at 2.7 Å. *J Mol Biol*, **190**, 593-604.
- Sauvage, C., Rumigny, J.F. and Maitre, M. (1991) Purification and characterization of G proteins from human brain: modification of GTPase activity upon phosphorylation. *Mol Cell Biochem*, **107**, 65-77.
- Scheidig, A.J., Burmester, C. and Goody, R.S. (1999) The pre-hydrolysis state of p21(ras) in complex with GTP: new insights into the role of water molecules in the GTP hydrolysis reaction of ras-like proteins. *Structure*, **7**, 1311-1324.
- Scheffzek, K., Ahmadian, M.R. and Wittinghofer, A. (1998) GTPase-activating proteins: helping hands to complement an active site. *Trends Biochem Sci*, **23**, 257-262.



## Bibliographie

- Schwartz, T. and Blobel, G. (2003) Structural basis for the function of the beta subunit of the eukaryotic signal recognition particle receptor. *Cell*, **112**, 793-803.
- Sever, S., Damke, H. and Schmid, S.L. (2000) Garrotes, springs, ratchets, and whips: putting dynamin models to the test. *Traffic*, **1**, 385-392.
- Shiba, T., Kawasaki, M., Takatsu, H., Nogi, T., Matsugaki, N., Igarashi, N., Suzuki, M., Kato, R., Nakayama, K. and Wakatsuki, S. (2003) Molecular mechanism of membrane recruitment of GGA by ARF in lysosomal protein transport. *Nat Struct Biol*, **10**, 386-393.
- Shin, D.H., Lou, Y., Jancarik, J., Yokota, H., Kim, R. and Kim, S.H. (2004) Crystal structure of YjeQ from *Thermotoga maritima* contains a circularly permuted GTPase domain. *Proc Natl Acad Sci U S A*, **101**, 13198-13203.
- Sliz, P., Michielin, O., Cerottini, J.C., Luescher, I., Romero, P., Karplus, M. and Wiley, D.C. (2001) Crystal structures of two closely related but antigenically distinct HLA-A2/melanocyte-melanoma tumor-antigen peptide complexes. *J Immunol*, **167**, 3276-3284.
- Sood, P., Lerner, C.G., Shimamoto, T., Lu, Q. and Inouye, M. (1994) Characterization of the autophosphorylation of Era, an essential *Escherichia coli* GTPase. *Mol Microbiol*, **12**, 201-208.
- Sprang, S.R. (1997) G protein mechanisms: insights from structural analysis. *Annu Rev Biochem*, **66**, 639-678.
- Stavridi, E.S., Huyen, Y., Loreto, I.R., Scolnick, D.M., Halazonetis, T.D., Pavletich, N.P. and Jeffrey, P.D. (2002) Crystal structure of the FHA domain of the Chfr mitotic checkpoint protein and its complex with tungstate. *Structure*, **10**, 891-899.
- Stewart-Jones, G.B., McMichael, A.J., Bell, J.I., Stuart, D.I. and Jones, E.Y. (2003) A structural basis for immunodominant human T cell receptor recognition. *Nat Immunol*, **4**, 657-663.
- Sun, Y.J., Forouhar, F., Li Hm, H.M., Tu, S.L., Yeh, Y.H., Kao, S., Shr, H.L., Chou, C.C., Chen, C. and Hsiao, C.D. (2002) Crystal structure of pea Toc34, a novel GTPase of the chloroplast protein translocon. *Nat Struct Biol*, **9**, 95-100.

## T

- Tallquist, M.D., Yun, T.J. and Pease, L.R. (1996) A single T cell receptor recognizes structurally distinct MHC/peptide complexes with high specificity. *J Exp Med*, **184**, 1017-1026.
- Terwilliger, T.C. (2003) Automated main-chain model building by template matching and iterative fragment extension. *Acta Crystallogr D Biol Crystallogr*, **59**, 38-44.



## *Bibliographie*

- Turner, S.J., Kedzierska, K., Komodromou, H., Gruta, N.L.I., Dunstone, M.A., Webb, A.I., Webby, R., Walden, H., Xie, W., McCluskey, J., Purcell, A.W., Rossjohn, J. and Doherty, P.C. (2005) Lack of prominent peptide-major histocompatibility complex features limits repertoire diversity in virus CD8<sup>+</sup> T cell populations. *Nature Immunology*, **6**, 382-389.
- Tynan, F.E., Burrows, S.R., Buckle, A.M., Clements, C.S., Borg, N.A., Miles, J.J., Beddoe, T., Whisstock, J.C., Wilce, M.C., Silins, S.L., Burrows, J.M., Kjer-Nielsen, L., Kostenko, L., Purcell, A.W., McCluskey, J. and Rossjohn, J. (2005a) T cell receptor recognition of a 'super-bulged' major histocompatibility complex class I-bound peptide. *Nature Immunology*, **6**, 1114-1122.
- Tynan, F.E., Elhassen, D., Purcell, A.W., Burrows, J.M., Borg, N.A., Miles, J.J., Williamson, N.A., Green, K.J., Tellam, J., Kjer-Nielsen, L., McCluskey, J., Rossjohn, J. and Burrows, S.R. (2005b) The immunogenicity of a viral cytotoxic T cell epitope is controlled by its MHC-bound conformation. *J. Exp. Med.*, **202**, 1249-1260.

## U

- Uson, I. and Sheldrick, G.M. (1999) Advances in direct methods for protein crystallography. *Curr Opin Struct Biol*, **9**, 643-648.

## V

- Van der Merwe, P.A., Bodian, D.L., Daenke, S., Linsley, P. and Davis, S.J. (1997) CD80 (B7-1) binds both CD28 and CTLA-4 with a low affinity and very fast kinetics. *J Exp Med*, **185**, 393-403.
- Vetter, I.R. and Wittinghofer, A. (2001) The guanine nucleotide-binding switch in three dimensions. *Science*, **294**, 1299-1304.

## W

- Walker, J.E., Saraste, M., Runswick, M.J. and Gay, N.J. (1982) Distantly related sequences in the alpha- and beta-subunits of ATP synthase, myosin, kinases and other ATP-requiring enzymes and a common nucleotide binding fold. *Embo J*, **1**, 945-951.
- Wedemeyer, H., Mizukoshi, E., Davis, A.R., Bennink, J.R. and Rehermann, B. (2001) Cross-reactivity between hepatitis C virus and Influenza A virus determinant-specific cytotoxic T cells. *J Virol*, **75**, 11392-11400.

## *Bibliographie*

Welsh, K.M., Trach, K.A., Folger, C. and Hoch, J.A. (1994) Biochemical characterization of the essential GTP-binding protein Opg of *Bacillus subtilis*. *J Bacteriol*, **176**, 7161-7168.

Welsh, R.M. (2006) Private specificities of heterologous immunity. *Current Opinion in Immunology*, **18**, 1-7.

Wiktor, T.J., Doherty, P.C. and Koprowski, H. (1977) In vitro evidence of cell-mediated immunity after exposure of mice to both live and inactivated rabies virus. *Proc Natl Acad Sci U S A*, **74**, 334-338.

Wittinghofer, A. (1997) Signaling mechanistics: aluminum fluoride for molecule of the year. *Curr Biol*, **7**, R682-685.

## Z

Zhou, T., Radaev, S., Rosen, B.P. and Gatti, D.L. (2000) Structure of the ArsA ATPase: the catalytic subunit of a heavy metal resistance pump. *Embo J*, **19**, 4838-4845.

Zinkernagel, R.M., Callahan, G.N., Althage, A., Cooper, S., Klein, P.A. and Klein, J. (1978) On the thymus in the differentiation of "H-2 self-recognition" by T cells: evidence for dual recognition? *J Exp Med*, **147**, 882-896.



ARTICLE



Stéphanie Gras,<sup>a</sup> Bernard  
Fernandez,<sup>b</sup> Valérie Chaumont,<sup>b</sup>  
Philippe Carpentier,<sup>a</sup> Jean  
Armengaud<sup>b</sup> and Dominique  
Housset<sup>a\*</sup>

<sup>a</sup>Laboratoire de Cristallographie et  
Cristallogénèse des Protéines, Institut de  
Biologie Structurale J.-P. Ebel, CEA-CNRS-UJF,  
41 Rue Jules Horowitz, F-38027 Grenoble,  
France, and <sup>b</sup>Service de Biochimie  
Post-génomique et Toxicologie Nucléaire,  
Département d'Ingénierie et d'Etudes des  
Protéines, CEA-VALRHO, Marcoule, BP 17171,  
F-30207 Bagnols sur Cèze, France

Correspondence e-mail:  
dominique.housset@ibs.fr

Received 26 October 2004  
Accepted 3 January 2005  
Online 20 January 2005

## Expression, purification, crystallization and preliminary crystallographic analysis of the PAB0955 gene product

PAB0955 from *Pyrococcus abyssi* is a prototype of a new Walker-type ATPase/GTPase conserved in archaea and eukaryota but not found in bacteria. PAB0955 has been expressed, purified and crystallized, and it has been shown that this thermostable protein is dimeric in reductive conditions. Crystals have been obtained either without nucleotide or in the presence of GDP or GTP $\gamma$ S. Preliminary X-ray crystallographic data up to 2.08 Å resolution have been collected from these crystals.

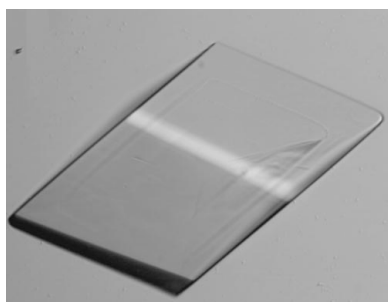
### 1. Introduction

The sequencing of entire genomes from organisms from the three domains of life has allowed the definition of clusters of orthologous groups of proteins (COGs) of unknown biological function. Among them, 32 COGs, named PACE (proteins of archaea conserved in eukaryotes), are common to archaea and eukaryota but are not found in bacteria (Matte-Tailliez *et al.*, 2000). Most of these proteins have been annotated as involved in translation, transcription, replication and DNA repair. We have started a systematic structural and functional analysis of these proteins in order to establish their biological function (Armengaud *et al.*, 2003, 2004). PAB0955, a protein encoded by the *Pyrococcus abyssi* genome, was our first target for crystallographic study. This 29 kDa protein (248 amino acids) has been annotated as an ATP/GTP-binding protein and a putative Walker-type ATPase/GTPase. Its function should be crucial as all eukaryotes and archaea have at least one homologue encoded in their genomes. The genetic environment of the *PAB0955* gene in the *P. abyssi* genome suggests a role for its product in the interplay between cell division and DNA replication (Matte-Tailliez *et al.*, 2000). Its closest known eukaryotic homologue is XPA-binding protein 1 found in mice and humans, with 28% sequence identity over 248 amino acids. The closest bacterial homologue has only 13% sequence identity. Moreover, GTPase activity (and the absence of significant ATPase activity) has been demonstrated *in vitro* with the purified recombinant protein (J. Armengaud, unpublished results). Since no remotely related proteins were detected in the Protein Data Bank, we decided to determine the crystal structure of PAB0955. This archaeal three-dimensional template should greatly help in unveiling the biological function of the whole eukaryotic and archaeal family. Here, we describe the expression, purification and crystallization of PAB0955. Preliminary crystallographic data collected on either crystals of uncomplexed native protein or crystals of protein grown in the presence of guanosine diphosphate (GDP) and guanosine diphosphate monothiophosphate (GTP $\gamma$ S) are presented.

### 2. Expression and purification

#### 2.1. Plasmid construction

The PAB0955 coding sequence was amplified from *P. abyssi* genomic DNA with two synthetic oligonucleotide primers: oAB18 (5'-GCTAGCATGATAGTCGTGTTT-3') and oAB19 (5'-TCAAGTCAGGTCACCGCACGT-3'). The resulting 753 bp PCR product was cloned into a pCRT7/NT-topo vector (Invitrogen). One of the

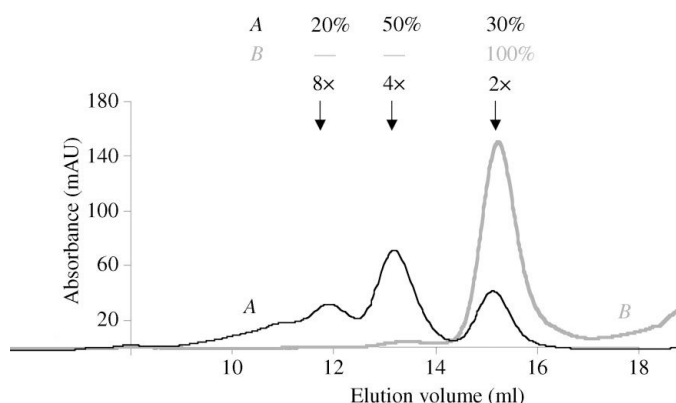


© 2005 International Union of Crystallography  
All rights reserved

recombinant plasmids, named pSBTN-AB34, was selected and its Xpress epitope removed by digestion with *NheI* endonuclease and religation with T4-DNA ligase. The resulting plasmid, named pSBTN-AB32, was sequenced on both strands in order to ascertain the integrity of the nucleotide sequence. This construction resulted in the addition of 14 extra residues at the N-terminus (MRG-SHHHHHHHGMS) of the recombinant protein. Prior to cloning, various prediction methods based on the amino-acid sequence (Odorico & Pellequer, 2003) were used to check that this N-terminal tag would not affect the secondary-structure prediction.

## 2.2. Expression and purification of His<sub>6</sub>-tagged PAB0955

Overexpression of the *PAB0955* construct was carried out essentially as previously described (Armengaud *et al.*, 2003) using *Escherichia coli* Rosetta(DE3)pLysS strain (Novagen) transformed with pSBTN-AB32. This strain carries extra tRNA genes for arginine, isoleucine, proline and leucine corresponding to codons commonly found in many archaea including the *Pyrococcus* genus but rarely used by *E. coli*. Indeed, the presence of 33 rare codons and six tandems of such rare codons in the *PAB0955* nucleotide sequence could have prevented expression in a regular strain. Bacterial growth was performed at 303 K to an OD<sub>600</sub> of 0.6. Expression was induced with 1 mM isopropyl- $\gamma$ -D-thiogalactopyranoside (IPTG). The soluble cell extract was prepared as previously described (Armengaud *et al.*, 2004), subjected to a 20 min heat treatment at 353 K and immediately centrifuged at 20 000g for 20 min at 277 K. Subsequent steps were performed at room temperature. NaCl was added to the resulting 300 ml clear supernatant to reach a final concentration of 250 mM. The sample was applied onto a Chelating Sepharose Fast Flow (50 ml, Amersham Biosciences) saturated with Ni<sup>2+</sup>. The major peak, which eluted from the column at about 200 mM imidazole, was concentrated and dialyzed overnight at 277 K against 50 mM KH<sub>2</sub>PO<sub>4</sub>/K<sub>2</sub>HPO<sub>4</sub> buffer pH 7.2 containing 10 mM NaCl. The sample was then applied onto a Resource-Q ion-exchange column (6 ml, Amersham Biosciences). Recombinant PAB0955 was eluted at approximately 400 mM NaCl and desalted by dialysis in Spectra/Por1 tubing (Spectrum Laboratories, molecular-weight cutoff 6–8 kDa) against 20 mM KH<sub>2</sub>PO<sub>4</sub>/K<sub>2</sub>HPO<sub>4</sub> buffer pH 7.2 containing 10 mM NaCl. A



**Figure 1**  
Elution profiles of PAB0955 samples on a Superdex200 gel-filtration column equilibrated with 50 mM Tris-HCl buffer pH 8.0 containing 400 mM potassium glutamate. The absorbance at 280 nm was recorded using an Äkta Purifier 10 FPLC system (Amersham Biosciences). Curve A corresponds to the elution of 240  $\mu$ g of PAB0955 without DTT. Curve B corresponds to the elution of 635  $\mu$ g of PAB0955 after incubation with 20 mM DDT. The different quaternary states indicated on the figure (2x, 4x and 8x) correspond to dimers, tetramers and octamers, respectively. They were estimated from the elution volume of calibrators resolved under the same conditions.

yield of approximately 6.6 mg of pure protein per litre of culture was obtained. Purified protein samples were flash-frozen in liquid nitrogen and stored at 193 K.

## 2.3. Gel-filtration analysis

Purified PAB0955 protein was analyzed by size-exclusion chromatography. The elution profile obtained using a Superdex 200 calibrated column (curve A in Fig. 1) clearly indicated the presence of at least three species corresponding to different oligomerization states (30% dimer, 50% tetramer, 20% octamer). The presence of four cysteines in the polypeptide sequence prompted us to test whether this heterogeneity could result from the formation of disulfide bonds. Indeed, after incubation with 10–20 mM dithiothreitol (DTT), the protein fraction appeared to be mainly dimeric as analysed by Superdex200 gel filtration (curve B in Fig. 1) and native electrophoresis gel (Fig. 2). Accordingly, crystallization assays were carried out in the presence of DTT.

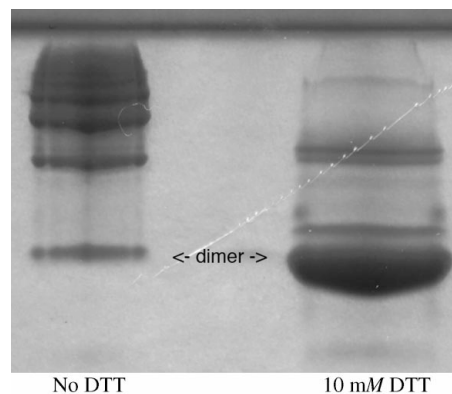
## 3. Crystallization

### 3.1. Crystals of native PAB0955

The first native protein crystals of about 20  $\mu$ m in size were obtained by the sitting-drop vapour-diffusion method using a Tecan robot and Hampton Research Crystal Screens I and II, Ammonium Sulfate, Malonate, Formate and Quick Screen crystallization kits. The initial crystallization condition [7.8 mg ml<sup>-1</sup> protein, 0.1 M trisodium citrate dihydrate pH 5.6, 20% (v/v) 2-propanol, 20% (w/v) PEG 4000 and 20 mM DTT] was optimized by substituting 2-propanol with different solvents provided by Additive Screen 1 (Hampton Research). Larger crystals (300  $\mu$ m) were obtained in 15% (w/v) PEG 4000, 3% (v/v) dioxane, 0.1 M trisodium citrate dihydrate pH 5.6 and 20 mM DTT. Hanging drops were prepared by mixing 1  $\mu$ l protein solution at a concentration of 7.8 mg ml<sup>-1</sup> with 1  $\mu$ l well solution at 291 K. Prismatic crystals (0.3  $\times$  0.1  $\times$  0.1 mm) were observed after 6 d (Fig. 3a).

### 3.2. Crystals of PAB0955 grown in the presence of nucleotides

A large-scale screening strategy was applied using a Cartesian automat and 15 crystallization screen kits (Crystal Screens I, II, Lite, PEG-Ion, MembFac, Natrix, Ammonium Sulfate, Screen Malonate, Formate, Quick Screen, PEG 6000, PEG LiCl, MPD, MME 5000 and

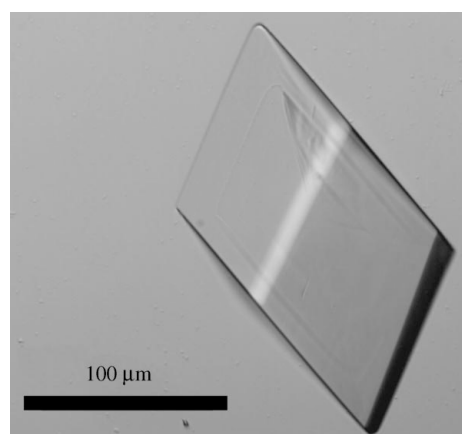


**Figure 2**  
Native 12.5% acrylamide gel electrophoresis of PAB0955. The concentration of PAB0955 sample was 4 mg ml<sup>-1</sup> in a buffer containing 20 mM NaH<sub>2</sub>PO<sub>4</sub> pH 7.2 and 50 mM NaCl. The migration was performed at 200 V. The migration of the protein without DTT and after incubation in 10 mM DTT are shown in the left and right column, respectively. The dimeric state clearly prevails in the presence of DTT.

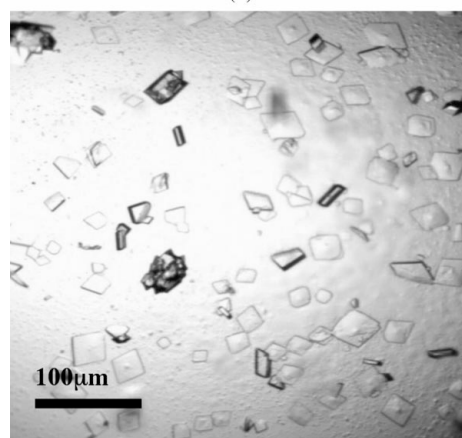
Index from Hampton Research). Systematically, 2  $\mu\text{l}$  of 1 M DTT solution was added to every 100  $\mu\text{l}$  of well solution prior to setting up the crystallization drops. Sitting drops were then prepared using 100 nl protein solution (final concentration of 7.8 mg ml<sup>-1</sup>) mixed with GTP $\gamma$ S (final concentration of 0.13 mM) and 100 nl well solution. After 1 d, 97 of the 596 crystallization conditions tested provided crystals. Conditions that led to larger crystals were reproduced by the hanging-drop technique using 1  $\mu\text{l}$  protein solution at 7.8 mg ml<sup>-1</sup>, 1  $\mu\text{l}$  GTP $\gamma$ S at 0.65 mM and 1  $\mu\text{l}$  well solution. The largest crystal (0.17  $\times$  0.16  $\times$  0.25 mm) grew from 13% (w/v) PEG 4000, 0.2 M ammonium acetate, 0.1 M trisodium citrate dihydrate pH 5.6 and 20 mM DTT after two weeks at 291 K. Attempts to crystallize PAB0955 in the presence of 0.65 mM GTP $\gamma$ S and under the same conditions as the native protein were unsuccessful. Crystals of PAB0955 grown in the presence of GDP were obtained under the same conditions after 2 d and had dimensions of 0.22  $\times$  0.12  $\times$  0.04 mm (Fig. 3b).

### 3.3. Data collection

Prior to being flash-cooled in liquid nitrogen, crystals were soaked in a cryoprotectant solution containing mother liquor with the PEG concentration increased to 30% (w/v). All data were collected at 100 K. The data set for a native protein crystal was collected on beamline ID14-eh2 of the European Synchrotron Radiation Facility



(a)



(b)

### Figure 3

(a) Native crystals of PAB0955 obtained in 15% (w/v) PEG 4000, 3% (v/v) dioxane, 0.1 M trisodium citrate dihydrate pH 5.6 and 20 mM DTT. (b) Crystals of PAB0955 in complex with GDP obtained in 13% (w/v) PEG 4000, 0.2 M ammonium acetate, 0.1 M trisodium citrate dihydrate pH 5.6 and 20 mM DTT.

**Table 1**

Crystallographic data statistics.

Values in parentheses are for the highest resolution shell.

	Native	GTP $\gamma$ S	GDP
Space group	$P3_121$ or $P3_221$	$P3_121$ or $P3_221$	$P2_1$
Unit-cell parameters (Å, °)	$a = b = 60.3$ , $c = 117.1$	$a = b = 60.2$ , $c = 115.9$	$a = 59.1$ , $b = 84.9$ , $c = 60.2$ , $\beta = 95$
Resolution (Å)	50–2.15 (2.23–2.15)	50–2.08 (2.15–2.08)	50–2.30 (2.38–2.30)
No. observations	140472 (10187)	83727 (5701)	88157 (8488)
No. unique reflections	13998 (1345)	12892 (1245)	26182 (2560)
$\langle I \rangle / \langle \sigma(I) \rangle$	8.2 (2.2)	6.6 (2.3)	5.2 (2.1)
Completeness (%)	100 (100)	87.6 (94.0)	99.9 (100)
$R_{\text{merge}}$ (%)	5.7 (34.8)	6.7 (28.4)	6.8 (36.1)

(ESRF) at a wavelength of 0.933 Å using an ADSC Q4 CCD detector. The data set of the PAB0955–GDP crystal was collected in-house with a Rigaku RU-200 rotating-anode X-ray generator at a wavelength of 1.542 Å (Cu  $K\alpha$ ) using an R-Axis IV<sup>++</sup> imaging-plate detector. The PAB0955–GTP $\gamma$ S crystallographic data set was collected on beamline BM-30A of the ESRF at a wavelength of 0.979 Å using a MAR CCD detector. Data processing was performed using XDS (Kabsch, 1993).

## 4. Results

### 4.1. The tagged PAB0955 product is hyperthermostable

To design a PAB0955 overexpression system, the Met-ATG codon located on the Watson strand at 1392033 (numbering of the whole *P. abyssi* genome sequence; Cohen *et al.*, 2003; <http://www-archbac.u-psud.fr/genomes/newpab/newpab.html>) was taken as the initiation codon. The upstream Met-TTG codon (at 1391946) that the genome annotators first mentioned in the database is rarely used as an initiator in archaea and the resulting N-terminal PAB0955 sequence aligns better with those of the other archaeal homologues. Rosetta (DE3)pLysS *E. coli* cells containing pSBTN-AB32 plasmid produced a large amount of the expected 30 kDa polypeptide after induction at 303 K. The PAB0955 protein remained soluble when the cell extract was heated at 358 K for 20 min, thus confirming the hyperthermostability of the tagged recombinant product.

### 4.2. Preliminary crystallographic data

Complete crystallographic data sets were collected from crystals of either native PAB0955 or PAB0955 co-crystallized with GTP $\gamma$ S or GDP. The crystals of native PAB0955 and PAB0955–GTP $\gamma$ S belong to the trigonal space group  $P3_121$  or  $P3_221$ . The unit-cell parameters are very similar (Table 1) and according to the Matthews coefficient ( $V_M = 2.35 \text{ \AA}^3 \text{ Da}^{-1}$ ) there is one molecule per asymmetric unit. The structure factors of the native and GTP $\gamma$ S co-crystallized PAB0955 were compared with the program SCALEIT from the CCP4 suite (Collaborative Computational Project, Number 4, 1994). The overall  $\Delta F/F$  ( $\sum_h |F_{\text{native}} - F_{\text{GTP}\gamma\text{S}}| / \sum_h |F_{\text{native}}|$ ) value of 0.32 indicates a significant difference in the unit-cell content and may arise from the binding of the nucleotide. The crystals of the PAB0955–GDP complex belong to the monoclinic space group  $P2_1$ . The Matthews coefficient indicates the presence of two molecules per asymmetric unit ( $V_M = 2.51 \text{ \AA}^3 \text{ Da}^{-1}$ ). Crystals belonging to the  $P2_1$  space group were only obtained in the presence of GDP in the crystallization solution. Experimental phases will be needed in order to check the specific binding of the nucleotide. The statistics of the different data sets are detailed in Table 1.



## 5. Conclusions

The presence of a nucleotide appears to facilitate crystallization, since only one condition of the 192 present in the Hampton Research Crystal Screens I and II, Ammonium Sulfate, Malonate, Formate and Quick Screen crystallization kits led to crystals for the native protein. In the presence of GTP $\gamma$ S, crystals of PAB0955 were obtained from 12 conditions from the same screening kits. These preliminary data are promising and searches for heavy-atom derivatives are under way. If useful derivatives are not readily obtained, recombinant SeMet-substituted proteins will be used for phasing. The three-dimensional structure should confirm the GTP specificity of PAB0955 and allow us to identify the residues involved in GTP hydrolysis. It will be also very useful in identifying putative partners. Along with other biochemical and genetic approaches, it may thus help to unveil the biological role of PAB0955 and by extension that of its human homologue.

We thank J. Dupuy (beamline FIP-BM30A) and S. Monaco (beamline ID14-eh2) for help with synchrotron data collection at the

ESRF (Grenoble). This work was supported in part by a grant from the Centre National de la Recherche Scientifique (Program Protéomique et Génie des Protéines, Appel à propositions 2001).

## References

- Armengaud, J., Fernandez, B., Chaumont, V., Rollin-Genetet, F., Finet, S., Marchetti, C., Myllykallio, H., Vidaud, C., Pellequer, J. L., Gribaldo, S., Forterre, P. & Gans, P. (2003). *J. Biol. Chem.* **278**, 31078–31087.
- Armengaud, J., Urbonavicius, J., Fernandez, B., Chaussinand, G., Bujnicki, J. M. & Grosjean, H. (2004). *J. Biol. Chem.* **279**, 37142–37152.
- Cohen, G. N., Barbe, V., Flament, D., Galperin, M., Heilig, R., Lecompte, O., Poch, O., Prieur, D., Querellou, J., Ripp, R., Thierry, J. C., Van der Oost, J., Weissenbach, J., Zivanovic, Y. & Forterre, P. (2003). *Mol. Microbiol.* **47**, 1495–1512.
- Collaborative Computational Project, Number 4 (1994). *Acta Cryst.* **D50**, 760–763.
- Kabsch, W. (1993). *J. Appl. Cryst.* **26**, 795–800.
- Matte-Tailliez, O., Zivanovic, Y. & Forterre, P. (2000). *Trends Genet.* **16**, 533–536.
- Odorico, M. & Pellequer, J. L. (2003). *J. Mol. Recognit.* **16**, 20–22.



Deux types de problématiques différentes peuvent motiver l'approche structurale d'une macromolécule. Soit la fonction de cette macromolécule est inconnue et l'on détermine sa structure afin de mieux connaître sa fonction ; cette démarche était l'un des défis du développement de la génomique structurale. Soit la fonction est connue et la structure va permettre de comprendre comment cette molécule remplit son rôle biologique.

Durant ma thèse, j'ai eu l'occasion de pouvoir aborder deux projets structuraux très différents, qui correspondent à ces deux approches de biologie structurale. C'est pourquoi mon manuscrit décrit ces deux projets dont la démarche scientifique est différente et complémentaire pour ma formation en biologie structurale.

Dans un premier temps, je décrirai les résultats obtenus sur le projet PAB0955, débuté en stage de DEA et poursuivi pendant les deux premières années de ma thèse. Ce projet avait pour objectif de déterminer la fonction de la protéine d'après sa structure. La protéine PAB0955 est de fonction biologique inconnue, elle hydrolyse le GTP et elle n'a pas de proche homologue dont la structure est connue. La démarche scientifique de ce projet a été de déterminer la structure de cette protéine, puis d'analyser sa structure en la comparant de manière systématique à l'ensemble des protéines ATPases et GTPases. Notre étude a été conduite dans l'objectif d'extraire un maximum d'information fonctionnelle à partir des données structurales.

Dans une deuxième partie, je décrirai notre étude sur des protéines du système immunitaire humain débuté en troisième année de thèse. Nous nous sommes intéressés à des molécules impliquées dans la reconnaissance d'antigène : le récepteur du lymphocyte T (TCR) et la molécule du complexe majeur d'histocompatibilité (CMH). L'objectif de ce projet est de comprendre comment un complexe antigène-CMH sélectionne un répertoire de lymphocyte T spécifique. L'objectif étant de déterminer le lien entre les caractéristiques structurales d'un complexe antigène-CMH et la diversité et la fréquence des lymphocytes T activés par ce complexe. En résumé corréler les observations structurales avec les données immunologiques dont nous disposons. En déterminant les bases structurales du caractère antigénique d'un peptide présenté par une molécule CMH nous pourrons, à plus long terme, produire des peptides efficaces dans le cadre d'une vaccination