



**HAL**  
open science

# étude structurale et fonctionnelle de la protéine KIN17 humaine

Albane Le Maire

► **To cite this version:**

Albane Le Maire. étude structurale et fonctionnelle de la protéine KIN17 humaine. Chimie. Université Pierre et Marie Curie - Paris VI, 2008. Français. NNT : . tel-00361216

**HAL Id: tel-00361216**

**<https://theses.hal.science/tel-00361216>**

Submitted on 13 Feb 2009

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

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

**UNIVERSITE PARIS VI – PIERRE ET MARIE CURIE**

ECOLE DOCTORALE Interfaces de la Chimie, de la Physique et de l'Informatique avec la  
Biologie

**THESE**

pour obtenir le titre de DOCTEUR de l'UNIVERSITÉ PIERRE ET MARIE CURIE  
Spécialité Biologie Structurale

présentée par **Albane le Maire**

**ETUDE STRUCTURALE ET FONCTIONNELLE DE LA PROTEINE KIN17 HUMAINE**

Soutenue le 15 janvier 2008

**Jury**

Dr B. Gilquin	Directeur de thèse
Dr JP. Samama	Rapporteur
Dr M. Kochoyan	Rapporteur
Dr. B. Seraphin	Examinateur
Pr. O. Lequin	Professeur Paris VI



J'exprime ma gratitude aux Dr **JP. Samama** et **M. Kochoyan** qui ont accepté de juger ce travail en tant que rapporteurs, ainsi qu'au Dr **B. Seraphin** et au Pr **O. Lequin** qui ont accepté de faire partie du jury. Je les remercie tous pour le temps qu'ils ont consacré à la lecture et à la critique de ce manuscrit.

Je remercie le Pr **A. Menez**, chef du Département d'Ingénierie et d'Etudes des Protéines du CEA Saclay de m'avoir accueillie et de m'avoir permis de réaliser cette thèse.

Je remercie le Dr **B. Gilquin**, mon directeur de thèse. Par sa disponibilité et son écoute, il a dirigé cette thèse tout en me laissant libre dans mes choix.

Je remercie le Dr **S. Zinn-Justin** qui a participé très activement à mon encadrement durant cette thèse. Son enthousiasme et son énergie ont été dynamisants tout au long de ma thèse.

Je remercie le Dr **J. Couprie** qui a également participé à mon encadrement. Il n'a pas hésité à me consacrer du temps et à donner des conseils qui ont été précieux pour ma thèse.

Je remercie les Dr. **E. Stura**, **JB. Charbonnier** et **MH. Ledu**, les cristallographes du laboratoire, pour leur aide aux différentes étapes des études de cristallographie et pour l'intérêt qu'ils ont porté à mon travail de thèse.

Je suis également très reconnaissante envers les personnes avec lesquelles j'ai collaboré :

Les Dr. **J. Angulo-Mora**, **L. Miccoli** et **G. Pinon-Lataillade** (LGR/DRR/CEA Fontenay-aux-Roses) ont initié l'étude de la protéine KIN17. Ils ont toujours été accueillants pour m'aider à faire des expériences ou discuter de mes résultats.

L'équipe dirigée par le Dr **M. Gondry** (SIMOPRO/CEA Saclay) a participé très activement à la production des domaines de la protéine hKIN17.

Le Pr **M. Schiltz** (LCR, EPFL, Lausanne) m'a énormément appris lors de la résolution des dérivés tungstate. Il a toujours été très accueillant pour les maintes fois où je l'ai sollicité par des questions ou pour des expériences.

Le Dr **P. Vachette** (IBBMC, Orsay) m'a énormément appris lors des études par SAXS. Il a toujours été disponible pour répondre à mes questions et discuter de l'analyse des données de SAXS, avec beaucoup de rigueur.

Le Dr **A. Galat** (SIMOPRO/CEA Saclay) a consacré beaucoup d'énergie pour rechercher les partenaires associés à la protéine hKIN17. J'ai compris beaucoup de choses grâce à son regard critique et ses nombreuses connaissances en biochimie.

Le Pr. **R. Lührmann** (MPI, Göttingen) m'accueillie à plusieurs reprises dans son laboratoire. Une rencontre à un congrès a permis de mettre en place rapidement cette collaboration. Dans ce laboratoire, j'ai travaillé essentiellement avec **J. Deckert** qui a pris le temps de bien m'expliquer les différentes expériences.

Le Pr. **M. Carmo-Fonseca** (IMM, Lisbonne) a accepté de faire des expériences sur KIN17 avant même que nous nous rencontrions.

Les Dr. **J. Chamot-Rooke**, **JP. le Caer** et **C. Malosse** (DCMR, Ecole Polytechnique) ont réalisé les analyses par spectrométrie de masse. Ils sont toujours partants pour de nouvelles analyses. Par leur intermédiaire, l'équipe du Dr. B. Monsarrat a accepté la tâche difficile d'analyses des histones modifiées.

Le Dr. **J. Acker** m'a accueilli pour des analyses de Southwestern.

Le Dr. **L. Carlier** a réalisé la structure par RMN d'un domaine de hKIN17. Je lui souhaite plein de bonnes choses pour son post-doc et pour la suite.

Je remercie S. Gras, que j'ai encadrée pour son stage de Master, pour sa contribution à ce travail. Je remercie toutes les autres personnes du laboratoire que j'ai côtoyées pendant ma thèse : R. Guerois, F. Oschenbein, S. Gasparini, S. Caputo, G. Charier, P. Llinas, C. Laguri, C. Tellier, G. Letellier, N. Lancelot, JN. Polonowski, M. Gallopin, E. Kondé, B. Matot.

Je remercie les personnes que j'ai rencontrées au cours de ma Formation Supérieure Biomédicale à l'IFSBM, Villejuif. En particulier, je suis très reconnaissante envers le Dr. **MC. Labastie** qui m'a donné de nombreux très bons conseils tout au long de ma thèse ainsi qu'envers le Dr. **A. Sarasin** qui a été un tuteur de thèse attentif.

Un grand merci à mes amis : ceux qui se sont également lancés dans l'aventure de la thèse et ceux qui sont dans d'autres sphères. Tous sont précieux.

Un dernier très grand merci à ma famille :

à mes parents pour m'avoir transmis l'essentiel et le sens du travail. Merci pour leur présence discrète.

à mes frère, sœurs, beaux-frères, belle-sœur, neveux et nièces qui sont supers.

à ma belle-famille, toujours accueillante et encourageante.

à mon mari, Gueric, pour tout ce que nous vivons et à nos jumelles, Alix et Camille, nées le 08 mars 2007, tout simplement merveilleuses !

INTRODUCTION.....	8
<b>1. LA PROTEINE KIN17.....</b>	<b>9</b>
1.1. L'IDENTIFICATION ET LES CARACTERISTIQUES DE LA PROTEINE KIN17 .....	9
1.2. LES PROPRIETES DECRITES POUR LA PROTEINE KIN17.....	11
<b>2. LA PROTEINE KIN17 ET LES DIFFERENTS MECANISMES CELLULAIRES.....</b>	<b>13</b>
2.1. LE ROLE EVENTUEL DE KIN17 DANS LA REPLICATION DE L'ADN.....	13
2.2. LE ROLE EVENTUEL DE KIN17 DANS LA REPARATION DE L'ADN.....	14
2.3. LE ROLE EVENTUEL DE KIN17 DANS LE METABOLISME DE L'ARN .....	16
<b>3. L'APPROCHE UTILISEE POUR L'ETUDE STRUCTURALE ET FONCTIONNELLE DE LA PROTEINE HKIN17 .....</b>	<b>19</b>
MATERIEL ET METHODES.....	20
<b>1. LA PREPARATION DES DOMAINES DE LA PROTEINE HKIN17 POUR LES ETUDES STRUCTURALE ET FONCTIONNELLE .....</b>	<b>22</b>
1.1. DOMAINES D'ETUDE DE LA PROTEINE .....	22
1.2. RECHERCHE DES CONDITIONS D'EXPRESSION POUR PRODUIRE LES DOMAINES DE LA PROTEINE .....	22
1.3. PREPARATION DES DOMAINES DE LA PROTEINE.....	24
1.4. L'EXPERIENCE DE PULL-DOWN ENTRE LES DOMAINES DE HKIN17 .....	26
<b>2. LES METHODES UTILISEES POUR L'ETUDE STRUCTURALE DE LA PROTEINE HKIN17</b>	<b>27</b>
2.1. METHODES D'ANALYSE DE SEQUENCES ET DE PREDICTIONS DE STRUCTURE.....	27
2.2. CRISTALLOGRAPHIE DU DOMAINE K1 .....	27
2.3. EXPERIENCES RMN SUR LES DIFFERENTS DOMAINES .....	32
2.4. MESURES DE DICHROÏSME CIRCULAIRE.....	34
2.5. MESURES DE DLS.....	34
2.6. MESURES DE SAXS.....	35
2.7. COMBINAISON DES INFORMATIONS A HAUTE ET BASSE RESOLUTION .....	44
<b>3. LES EXPERIENCES POUR L'ETUDE FONCTIONNELLE DE LA PROTEINE HKIN17 .....</b>	<b>48</b>
3.1. ETUDES D'INTERACTION DE LA PROTEINE HKIN17 AVEC LES ACIDES NUCLEIQUES.....	48
3.2. ETUDES D'INTERACTION DE LA PROTEINE HKIN17 AVEC LES HISTONES.....	49
3.3. EXPERIENCES DE PULL-DOWN POUR LA RECHERCHE DES PARTENAIRES PROTEIQUES ASSOCIES AUX DOMAINES K1 ET K2 .....	51
3.4. EXPERIENCES D'ISOLEMENT DE COMPLEXES ASSOCIES AU DOMAINE C-TERMINAL DE HKIN17 .....	53
3.5. IDENTIFICATION DES PROTEINES ISOLEES PAR SPECTROMETRIE DE MASSE.....	57
3.6. EXPERIENCES CONCERNANT LA RECHERCHE D'UN ROLE EVENTUEL DE HKIN17 DANS L'EPISSAGE....	61
RESULTATS DE L'ANALYSE STRUCTURALE.....	65
<b>1. LA PROTEINE HKIN17 EST UNE PROTEINE MODULAIRE .....</b>	<b>66</b>
1.1. PREDICTION DES REGIONS DESTRUCTUREES .....	66
1.2. IDENTIFICATION DE DOMAINES CONNUS DANS LES REGIONS GLOBULAIRES .....	67
1.3. VALIDATION EXPERIMENTALE DES PREDICTIONS DE REGIONS GLOBULAIRES .....	67
1.4. DELIMITATION DES DOMAINES POUR L'ANALYSE STRUCTURALE.....	68
<b>2. LA CARACTERISATION STRUCTURALE DU DOMAINE C-TERMINAL.....</b>	<b>69</b>
2.1. CRISTALLISATION DU DOMAINE C-TERMINAL.....	69
2.2. DETERMINATION DES PHASES .....	80
2.3. STRUCTURE A HAUTE RESOLUTION DU DOMAINE C-TERMINAL PAR RADIO-CRISTALLOGRAPHIE .....	85
2.4. ETUDE PAR RMN DU DOMAINE C-TERMINAL .....	99

<b>3.</b>	<b>LA CARACTERISATION STRUCTURALE DE LA REGION N-TERMINALE .....</b>	<b>103</b>
3.1.	DETERMINATION DE LA STRUCTURE TRIDIMENSIONNELLE DU DOMAINE 51-160 PAR RMN .....	103
3.2.	ESSAI DE DETERMINATION DE LA STRUCTURE DE LA REGION 1-160 PAR RMN .....	117
3.3.	ETUDE STRUCTURALE DE LA REGION 1-160 PAR SAXS .....	117
3.4.	DONNEES SAXS ET RMN COMPATIBLES POUR PROPOSER UN MODELE DE LA REGION 1-160 .....	126
<b>4.</b>	<b>VERS LA STRUCTURE DE LA PROTEINE ENTIERE.....</b>	<b>130</b>
4.1.	ABSENCE D'INTERACTION ENTRE LES REGIONS N- ET C-TERMINALES .....	130
4.2.	PREPARATION DE LA PROTEINE ENTIERE POUR LES ETUDES STRUCTURALES .....	131
4.3.	ENREGISTREMENT DE DONNEES DE SAXS SUR LA PROTEINE ENTIERE.....	133
 <b>RESULTATS DE L'ANALYSE FONCTIONNELLE.....</b>		<b>136</b>
<b>1.</b>	<b>LES DOMAINES IMPLIQUES DANS L'INTERACTION DE LA PROTEINE AVEC LES ACIDES NUCLEIQUES .....</b>	<b>137</b>
1.1.	INTERACTION DE LA PROTEINE AVEC L'ADN .....	137
1.2.	INTERACTION DE LA PROTEINE AVEC L'ARN.....	142
<b>2.</b>	<b>L'INTERACTION DU DOMAINE C-TERMINAL AVEC LES HISTONES .....</b>	<b>144</b>
2.1.	LIAISON DU DOMAINE C-TERMINAL AUX HISTONES H3 ET H4 .....	144
2.2.	LIAISON DU DOMAINE C-TERMINAL A L'HISTONE H3 MODIFIEE.....	147
<b>3.</b>	<b>LA RECHERCHE DE PARTENAIRES PROTEIQUES ASSOCIES AUX DOMAINES N- ET C-TERMINAUX .....</b>	<b>150</b>
3.1.	MISE AU POINT DU GST-PULLDOWN AVEC DES EXTRAITS MARQUES AU <sup>35</sup> S .....	150
3.2.	PULL-DOWN A GRANDE ECHELLE ET COUPLE A L'IDENTIFICATION DES PROTEINES PAR SPECTROMETRIE DE MASSE .....	151
<b>4.</b>	<b>L'ISOLEMENT DE COMPLEXES ASSOCIES AU DOMAINE C-TERMINAL .....</b>	<b>155</b>
4.1.	ASSOCIATION DU DOMAINE C-TERMINAL A UN COMPLEXE DE HAUT POIDS MOLECULAIRE.....	155
4.2.	ASSOCIATION DU DOMAINE C-TERMINAL A UN COMPLEXE DE P <sub>I</sub> GLOBAL ACIDE .....	157
4.3.	COMBINAISON DE LA FILTRATION MOLECULAIRE ET DU GEL ND-IEF.....	159
4.4.	IDENTIFICATION DES PROTEINES PRESENTES DANS LE COMPLEXE ASSOCIE AU DOMAINE C-TER.....	159
<b>5.</b>	<b>EST-CE QUE LA PROTEINE JOUE UN ROLE DANS L'EPISSAGE DES ARN PRE-MESSAGERS ? .....</b>	<b>164</b>
5.1.	PULL-DOWN DE REACTION D'EPISSAGE.....	164
5.2.	PULL-DOWN DES SNRNPs.....	165
5.3.	TEST D'EPISSAGE PAR IMMUNODEPLETION-COMPLEMENTATION .....	166
 <b>DISCUSSION ET PERSPECTIVES.....</b>		<b>172</b>
<b>1.</b>	<b>DISCUSSION ET PERSPECTIVES .....</b>	<b>173</b>
1.1.	L'ETUDE STRUCTURALE DE LA PROTEINE hKIN17 .....	173
1.2.	CE QU'APPORTE LA DETERMINATION DE LA STRUCTURE TRIDIMENSIONNELLE DES DOMAINES DE LA PROTEINE hKIN17 SUR LEUR FONCTION DANS LA PROTEINE.....	174
<b>2.</b>	<b>CONCLUSION.....</b>	<b>183</b>
 <b>BIBLIOGRAPHIE.....</b>		<b>184</b>
<b>ANNEXES.....</b>		<b>193</b>

Ici est donné le sens des **abréviations** qui ne sont pas expliquées dans le texte

**ADN** : Acide DésoxyriboNucléique  
**ARN** : Acide RiboNucléique  
**DTT** : 1,4-DiThioThreitol  
**EDTA** : Ethylene Diamine Tetra Acetic Acid  
**GSH** : Glutathione  
**GST** : Glutathion-S-Transferase  
**HPLC** : High Pressure Liquid Chromatography  
**HSQC** : Heteronuclear Single Quantum Coherence spectroscopy  
**NOESY** : Nuclear Overhauser Effect Spectroscopy  
**PBS** : Phosphate Buffer Saline  
**PEG** : Poly Ethylene Glycol  
**PMSF** : PhenylMethylSulfonyl Fluoride  
**PVDF** : PolyVinylidene DiFluoride  
**RMSD** : Root Mean Square Deviation  
**SDS-PAGE** : Sodium Dodecyl Sulfate-Poly Acrylamide Gel Electrophoresis  
**SH3** : Src-Homology 3  
**TAP(-Tag)**: Tandem Affinity Purification  
**TCEP** : Tris(2-CarboxyEthyl)Phosphine hydrochloride  
**TEV** : Tobacco Etch Virus  
**TRIS** : Tris(hydroxymethyl)amino methane

Ici sont données des **définitions** succinctes des termes anglais utilisés dans le texte

**Cluster** : regroupement de plusieurs éléments  
**Cross-linking** : création d'un lien (lien covalent entre deux protéines)  
**Dot-blot** : dépôt circulaire d'une protéine sur une membrane pour des études d'interaction  
**Far Western** : méthode pour mettre en évidence une interaction entre une protéine sur une membrane et une autre protéine en solution  
**Immunoblotting=Western blot** : détection d'une protéine sur membrane avec des anticorps dirigée contre cette protéine  
**Linker** : chaîne d'acides aminés entre deux domaines d'une protéine  
**Northwestern** : méthode pour mettre en évidence une interaction entre une protéine sur une membrane et un ARN en solution  
**Pull-down** : capture par affinité de protéines  
**Slurry** : suspension  
**Southwestern** : méthode pour mettre en évidence une interaction entre une protéine sur une membrane et un ADN en solution  
**Spliceosome** : ensemble des complexes ribonucléoprotéiques assurant l'épissage  
**Winged-helix** : hélice ailée, sous-classe de la superfamille des domaines hélice-coude-hélice de liaison à l'ADN  
**ZZ** : domaine construit à partir de deux domaines de liaison IgG de la protéine A de Staphylocoque



# **PREAMBULE**



## Le cadre scientifique de la thèse

Les protéines sont les principaux acteurs du vivant, non seulement par leur quantité (environ 70 % en masse des molécules constituant une cellule), mais surtout par la diversité des fonctions qu'elles exercent, allant de la catalyse de réactions chimiques jusqu'à la structuration de la matière vivante, en passant par des rôles de moteurs moléculaires ou encore de communication entre cellules. Elles sont formées de longues chaînes d'acides aminés, qui se replient dans l'espace. La structure d'une protéine correspond à une conformation stable mais modifiable en fonction de l'environnement puisqu'elle repose sur des interactions faibles. Elle est à la base des fonctions assurées par la protéine. **Connaître la structure aux détails atomiques près, comprendre ses propriétés dynamiques, suivre les changements de conformation d'une protéine en action, intégrer ces connaissances aux données biochimiques et fonctionnelles** constituent le cœur de la biologie structurale et permet d'avancer considérablement dans la compréhension de la fonction des protéines.

Plusieurs techniques biophysiques peuvent être utilisées lors de la détermination d'une structure et chacune de ces techniques apporte une information différente : architecture globale, détails atomiques, réarrangements structuraux, interactions avec les partenaires, mécanismes enzymatiques. La cristallographie par rayons X, la RMN, la diffusion de neutrons ou de rayons X et la microscopie électronique font partie de ces techniques. **La caractérisation structurale des protéines modulaires illustre bien la nécessité d'utiliser des approches complémentaires.** Ces protéines sont constituées de plusieurs modules adoptant chacun une structure de manière autonome. Cependant, les différents modules sont reliés par des séquences flexibles, ce qui crée une hétérogénéité structurale à l'échelle de la protéine entière. Les protéines modulaires ressemblent donc à des « lego » moléculaires qui peuvent se restructurer en cours de réaction.

La compréhension des bases moléculaires de leurs fonctions biologiques repose au départ sur une approche réductionniste, dans laquelle chaque fonction est disséquée au niveau moléculaire et la structure de chaque domaine fonctionnel ainsi identifié est analysée. **L'enjeu est aussi d'aller plus loin, c'est-à-dire de remonter aux architectures supramoléculaires présentes dans le contexte cellulaire**, de comprendre les propriétés structurales et dynamiques permettant l'assemblage des différents complexes, et de relier l'évolution de ces assemblages à leurs fonctions biologiques. Cela implique de pouvoir aborder des structures de complexes de plus en plus larges et souvent transitoires dans le temps. De plus, les propriétés dynamiques permettant les réarrangements structuraux nécessaires aux fonctions doivent être comprises.

## La protéine KIN17 en quelques mots

Très conservée, plus abondante dans les cellules en prolifération que dans les cellules normales, stimulée par les rayonnements, corrélée aux dommages de l'ADN, autant de caractéristiques qui ont suscité l'intérêt pour l'analyse structurale et fonctionnelle de la protéine KIN17 afin de caractériser sa ou ses fonction(s) dans la cellule. La protéine KIN17 a été particulièrement étudiée *in vitro* et *in vivo* dans le laboratoire de J. Angulo (LGR/DRR/CEA Fontenay-aux-Roses) (Masson et al. 2003; Miccoli et al. 2005) dans le cadre

du projet du laboratoire visant à évaluer les effets biologiques produits par les génotoxiques physiques et chimiques sur les cellules de mammifères *in vitro* et *in vivo* et à caractériser la réponse des cellules de mammifères aux radiations ionisantes au niveau moléculaire. Les données de ce laboratoire nous invitent à **préciser les rôles éventuels de la protéine KIN17 dans les mécanismes de réplication de l'ADN, réparation de l'ADN et métabolisme de l'ARN.**

## Le plan du manuscrit

La première partie (**Introduction**) présente les propriétés connues de la protéine KIN17 et les implications possibles de cette protéine dans les différents mécanismes cellulaires. A partir de là, nous présentons les objectifs du travail de thèse mené sur la protéine humaine KIN17.

La seconde partie (**Matériel et Méthodes**) présente en détails les méthodes utilisées pendant la thèse ainsi que les procédures expérimentales.

La troisième partie (**Résultats**) décrit (1) les résultats obtenus lors de la caractérisation structurale de la protéine humaine KIN17, par cristallographie aux rayons X, RMN et SAXS et (2) les résultats obtenus lors de la caractérisation fonctionnelle de la protéine, par des études d'interaction avec des partenaires connus et par la recherche de partenaires associés à l'un ou l'autre des domaines de la protéine.

La quatrième partie (**Discussion et Perspectives**) présente une synthèse des travaux réalisés au cours de la thèse et propose des perspectives à ce travail.

# **INTRODUCTION**



---

<b>1.</b>	<b>LA PROTEINE KIN17 .....</b>	<b>9</b>
1.1.	L'IDENTIFICATION ET LES CARACTERISTIQUES DE LA PROTEINE KIN17 .....	9
1.2.	LES PROPRIETES DECRITES POUR LA PROTEINE KIN17.....	11
<b>2.</b>	<b>LA PROTEINE KIN17 ET LES DIFFERENTS MECANISMES CELLULAIRES.....</b>	<b>13</b>
2.1.	LE ROLE EVENTUEL DE KIN17 DANS LA REPLICATION DE L'ADN.....	13
2.1.1.	<i>Le niveau d'expression du gène KIN17 et la prolifération cellulaire.....</i>	<i>13</i>
2.1.2.	<i>L'implication de la protéine KIN17 dans la réplication de l'ADN.....</i>	<i>13</i>
2.2.	LE ROLE EVENTUEL DE KIN17 DANS LA REPARATION DE L'ADN.....	14
2.2.1.	<i>Les mécanismes de réparation de l'ADN .....</i>	<i>14</i>
2.2.2.	<i>La protéine KIN17 est surexprimée après des dommages de l'ADN.....</i>	<i>15</i>
2.2.3.	<i>La protéine KIN17 est en partie associée à la chromatine.....</i>	<i>15</i>
2.2.4.	<i>Le lien entre les mécanismes de réparation de l'ADN et la chromatine .....</i>	<i>15</i>
2.3.	LE ROLE EVENTUEL DE KIN17 DANS LE METABOLISME DE L'ARN .....	16
2.3.1.	<i>Le rôle éventuel de la protéine KIN17 dans le contrôle de la transcription.....</i>	<i>16</i>
2.3.2.	<i>Les mécanismes d'épissage des ARN pré-messagers .....</i>	<i>17</i>
2.3.3.	<i>Le rôle éventuel de la protéine KIN17 dans l'épissage .....</i>	<i>18</i>
<b>3.</b>	<b>L'APPROCHE UTILISEE POUR L'ETUDE STRUCTURALE ET FONCTIONNELLE DE LA PROTEINE HKIN17 .....</b>	<b>19</b>



# 1. LA PROTEINE KIN17

## 1.1. L'identification et les caractéristiques de la protéine KIN17

En 1991, le gène *KIN17* de souris a été identifié sur la base de la réactivité croisée de la protéine pour laquelle il code avec des anticorps dirigés contre la protéine RecA (Angulo et al. 1991a). En effet, la protéine KIN17 de souris a une région limitée à 39 résidus qui est homologue à 47% à une région située dans la partie C-terminale de la protéine RecA d'*E. coli* (Angulo et al. 1991b) et impliquée dans la liaison de la protéine RecA à l'ADN (Kurumizaka et al. 1996). La protéine bactérienne RecA est impliquée dans la recombinaison homologue, la réponse SOS et la réponse aux dommages de l'ADN induits par mutagenèse (Cox 1999). Le gène *KIN17* humain a ensuite été identifié sur le chromosome 10 (Kannouche et al. 2000). Ce gène est exprimé de façon ubiquitaire dans les tissus humains (pancréas, rein, muscle squelettique, foie, poumon, placenta, cerveau, coeur, rate, thymus, prostate, testicule, ovaire, petit intestin, colon et leucocytes) bien que chaque tissu ait un niveau d'expression du gène spécifique (Figure 1-1). Le coeur, le muscle squelettique et les testicules présentent les niveaux d'ARNm de *KIN17* les plus élevés (Kannouche et al. 2000).

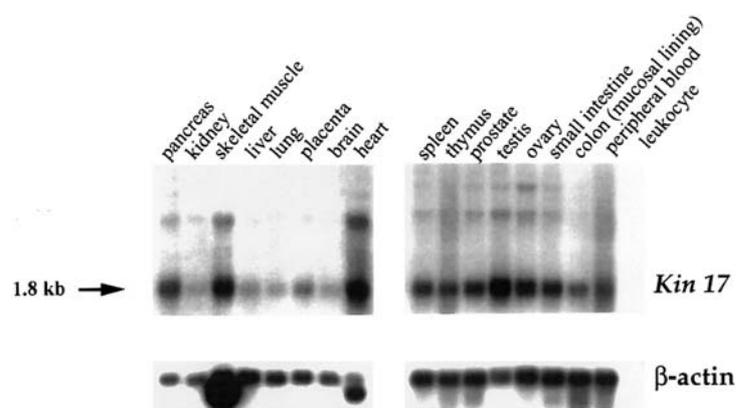


Figure 1-1 : Expression de l'ARNm de *KIN17* dans différents tissus humains. La bande à 1.8kb correspond à l'ARN de *KIN17* humain (Kannouche et al. 2000).

La protéine KIN17 humaine (hKIN17) codée par ce gène est une protéine de 45kDa, de nature basique. La protéine est présente dans une grande diversité d'organismes eucaryotes (animaux, plantes, champignons et protistes). Un alignement de toutes les séquences de KIN17 décrites à ce jour est présenté en annexe 1. Sur la Figure 1-2, un alignement de quelques séquences représentatives des séquences de KIN17 d'animaux, de plantes, de parasites et de levures est donné.

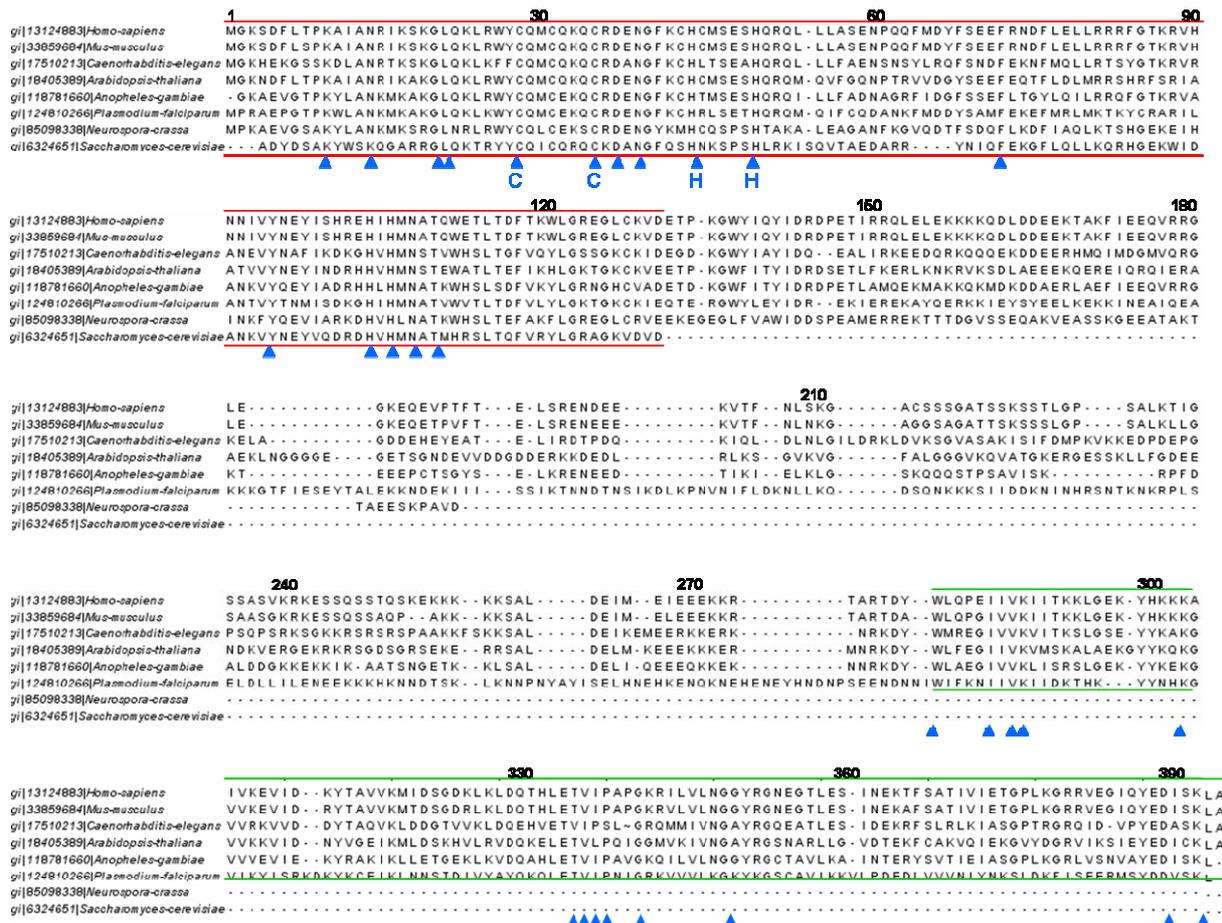


Figure 1-2 : Conservation des protéines KIN17 au cours de l'évolution. Alignement multiple de quelques séquences de protéines KIN17 réalisé avec CLUSTAL W et optimisé avec MUSCLE. La numérotation des acides aminés correspond à la séquence de la protéine KIN17 humaine. La région encadrée en rouge correspond à une région conservée dans toutes les séquences de KIN17 et la région encadrée en vert correspond à une région conservée chez les Eucaryotes Supérieurs seulement. Les flèches bleues pointent vers les résidus strictement conservés entre toutes les espèces.

La région N-terminale de la protéine (jusqu'au résidu 125) est très conservée entre toutes les espèces. La région centrale (région 126 à 267 environ) présente plus de variétés de séquence et de longueur. La région C-terminale, encore très conservée, est néanmoins absente des séquences des KIN17 des levures. La très forte conservation de la protéine dans l'évolution suggère que KIN17 est une protéine essentielle et que certaines de ses fonctions doivent également être conservées entre les espèces.

La suppression du gène homologue chez la levure, *rts2*, rend les cellules non viables (Giaever et al. 2002). Des interactions génétiques ont été observées entre *rts2* et *cdc40*, et *rts2* et *rps17a* (Davierwala et al. 2005) : *cdc40* code pour un facteur d'épissage des ARN pré-messagers qui interagit lui-même avec de nombreux facteurs d'épissage, comme cela a été montré par GST-pull-down (Hazbun et al. 2003) et TAP-Tag (Gavin et al. 2002), et qui joue un rôle dans la progression du cycle cellulaire (Kaplan et al. 2007) ; *rps17a* code pour la protéine ribosomale S17-A de la sous unité 40S. De plus, des expériences de TAP-Tag menées sur environ 2000 protéines de levure ont montré que Rts2 appartient à un complexe

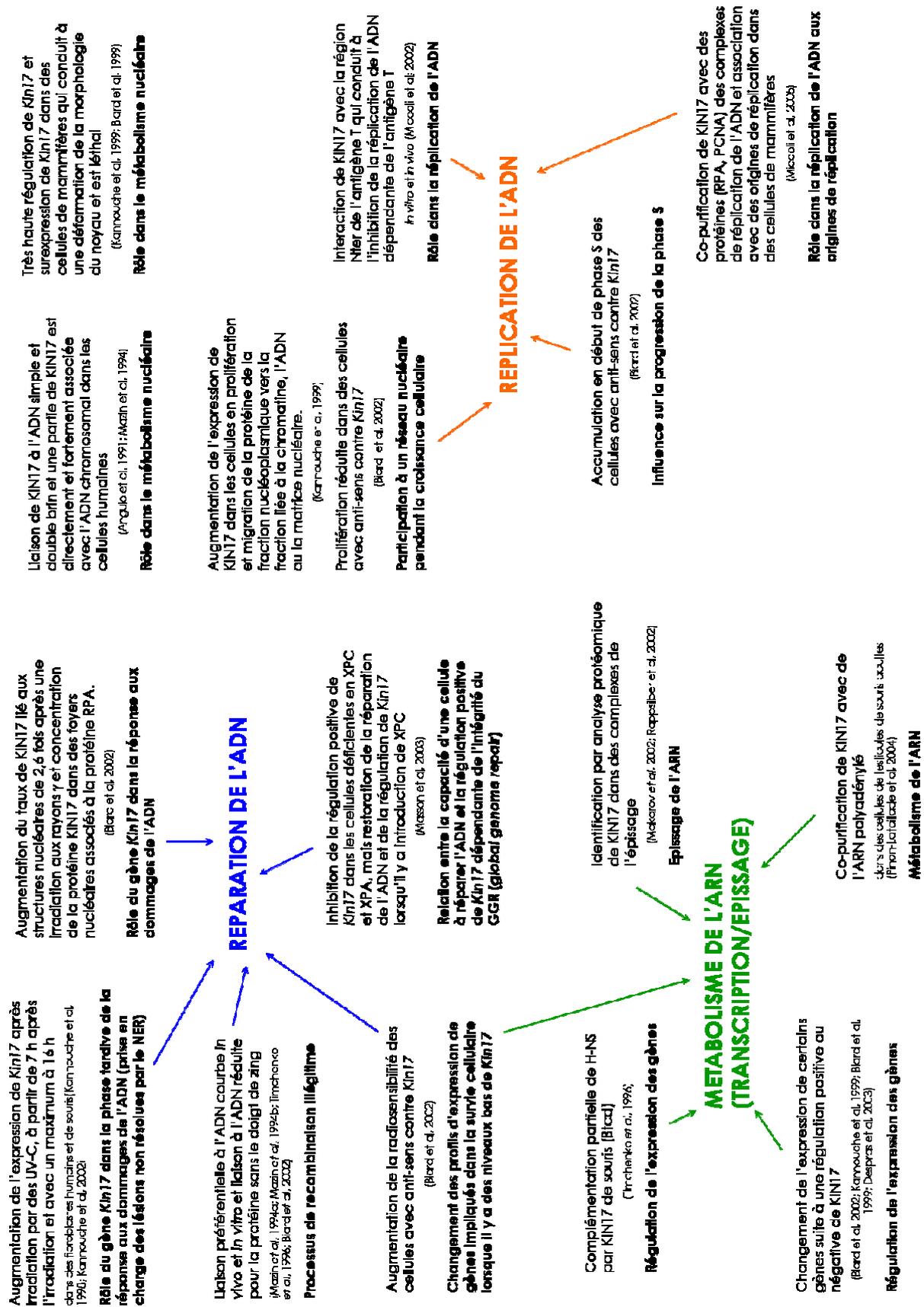
incluant la protéine ribosomale acide P0 appartenant à la sous unité 60S, le facteur d'élongation 1 alpha et la sous-unité B de l'ATP synthase de vacuole (Gavin et al. 2002).

La protéine hKIN17 est une protéine essentiellement nucléaire qui forme des foyers intranucléaires dans les cellules en prolifération (Kannouche et al. 2000). Elle a également été observée dans le cytoplasme des spermatides allongés (Pinon-Lataillade et al. 2004). Le gène *KIN17* est de manière générale très faiblement exprimé dans les cellules mais les cellules en division rapide contiennent 3 fois plus d'ARNm KIN17 que les cellules au repos (Kannouche et al. 1998). La concentration cellulaire de KIN17 est faible mais des plus hauts niveaux de KIN17 sont observés dans certaines cellules tumorales. Des quantifications par ELISA du niveau de protéine KIN17 endogène dans 16 lignées de cellules humaines ont montré une grande gamme d'expression allant de 100 000 molécules par noyau dans les cellules H1299 (cellules cancéreuses du poumon) à 4 600 molécules par noyau dans les cellules MeWo (mélanome) (Biard et al. 2003).

## 1.2. Les propriétés décrites pour la protéine KIN17

Comme cela est représenté schématiquement sur la page suivante, les propriétés de la protéine KIN17 peuvent être rattachées à trois grandes fonctions cellulaires : réplication de l'ADN, réparation de l'ADN et métabolisme de l'ARN (transcription/épissage). Nous détaillerons chacun des aspects dans le paragraphe suivant.

L'observation de structures discrètes appelées "foyers intra-nucléaires" contenant la protéine KIN17 dans les noyaux de cellules de mammifères reflète la compartimentation de transactions complexes impliquant les acides nucléiques comme la réplication, la réparation, la transcription ou l'épissage. Ce type de répartition en foyers a également été observé pour d'autres protéines comme par exemple la cycline A, PCNA, RPA, CAF-1, Rad51 et Brca1 (Newport et al. 1996). De manière générale, l'ensemble des propriétés décrites pour la protéine KIN17 suggère qu'elle est une protéine de la maintenance nucléaire.



---

## 2. LA PROTEINE KIN17 ET LES DIFFERENTS MECANISMES CELLULAIRES

---

### 2.1. Le rôle éventuel de KIN17 dans la réplication de l'ADN

#### 2.1.1. Le niveau d'expression du gène *KIN17* et la prolifération cellulaire

Le niveau d'expression du gène *KIN17* est plus faible dans les cellules quiescentes, confluentes ou différenciées par rapport à des cellules en prolifération (Biard et al. 1997; Kannouche et al. 1998; Pinon-Lataillade et al. 2004). En particulier, une augmentation du transcrit du gène *KIN17* dans des cellules de souris en culture est observée au niveau de la transition G1/S (Kannouche et al. 1998).

La sur-expression du gène *KIN17* a un effet antiprolifératif (Biard et al. 1999; Kannouche et al. 1999). Par ailleurs, l'extinction du gène humain *KIN17* par interférence ARN dans des cellules HeLa réduit considérablement leur croissance dans les deux mois qui suivent la transfection (Despras et al. 2007). La croissance cellulaire est également affectée dans les cellules RKO exprimant un transcrit *antisens* de l'ADNc du gène *KIN17* (Biard et al. 2002). Ces travaux montrent que l'expression du gène *KIN17* doit être strictement régulée pour assurer la prolifération cellulaire.

#### 2.1.2. L'implication de la protéine *KIN17* dans la réplication de l'ADN

La sur-expression de la protéine *KIN17* murine dans des cellules HeLa ayant un vecteur doté d'une origine de réplication SV-40 (*simian virus 40*) intégré dans leur génome, inhibe la réplication dépendant de l'antigène T de l'origine SV-40 (Kannouche et al. 1999). La protéine *KIN17* pourrait donc avoir un rôle au sein du cycle cellulaire, au niveau du point de contrôle de la transition G1/S ou pendant la phase de synthèse de l'ADN.

Par ailleurs, la protéine hKIN17 est localisée aux sites de réplication de l'ADN (Biard et al. 2003). La majorité des foyers nucléaires de hKIN17 co-localisent avec les sites d'incorporation de BrdU visualisés en microscopie électronique. Une étude récente a montré que la protéine hKIN17 s'associe avec les origines de réplication (Miccoli et al. 2005). Elle semble appartenir au complexe de réplication de l'ADN et la déplétion de cette protéine réduit l'activité du complexe *in vitro* (Miccoli et al. 2005).

## 2.2. Le rôle éventuel de KIN17 dans la réparation de l'ADN

### 2.2.1. Les mécanismes de réparation de l'ADN

Les mécanismes de réparation de l'ADN sont déclenchés pour faire face aux dommages de l'ADN. En effet, l'ADN est intrinsèquement instable et continuellement soumis à des dégradations causées par des déficiences internes aux cellules ou par des perturbations externes (radicaux libres, lumière UV, agents alkylants ou radiations ionisantes). L'accumulation de mutations génétiques dues à des défauts de réparation de l'ADN est la cause principale du développement de tumeurs (Motoyama et al. 2004). La réponse cellulaire aux dommages de l'ADN inclut les mécanismes de réparation de l'ADN et les mécanismes de contrôle du cycle cellulaire. La réparation de l'ADN implique des protéines qui détectent le dommage de l'ADN (senseurs), des protéines qui transmettent le signal du dommage et des protéines capables de réparer le dommage. Chaque type de dommage de l'ADN est pris en charge par une voie de réparation spécifique (Figure 2-1) (Hoeijmakers 2001). Le contrôle du cycle cellulaire implique également des senseurs des dommages de l'ADN qui recrutent des médiateurs et des enzymes chargés de transmettre le signal jusqu'aux régulateurs du cycle cellulaire (Melo et al. 2001; Nyberg et al. 2002). Ainsi, le cycle cellulaire est interrompu pendant la réparation et reprend lorsque le dommage est réparé (Figure 2-1).

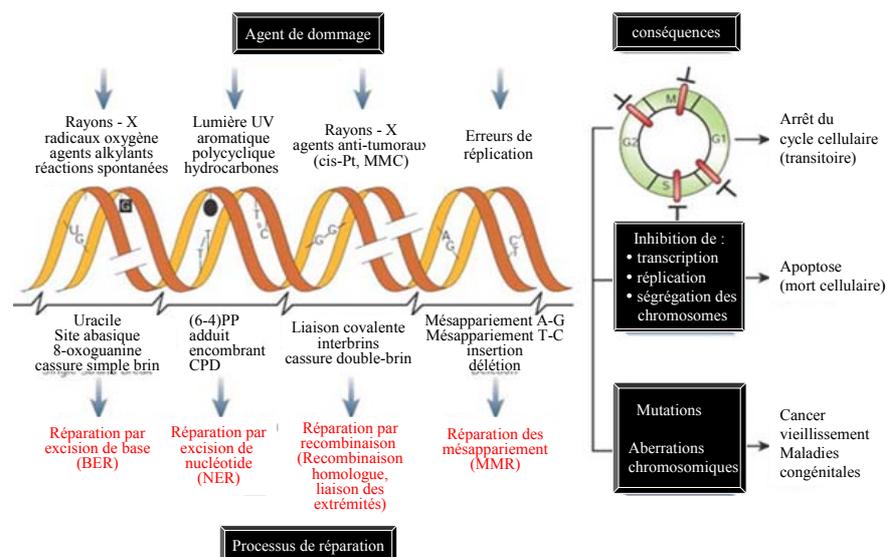


Figure 2-1 : Dommages de l'ADN, mécanismes de réparation et conséquences au niveau cellulaire (Hoeijmakers 2001).

Les protéines impliquées dans la surveillance du génome sont souvent de grande taille, possèdent de larges régions non structurées et sont impliquées dans un grand nombre d'interactions d'affinité généralement faibles. Des dysfonctionnements moléculaires dans ces interactions provoquent des déficits dans les voies de réparation correspondantes et peuvent ainsi conduire à des événements de carcinogenèse.

### **2.2.2. La protéine KIN17 est surexprimée après des dommages de l'ADN**

La propriété de la protéine KIN17 qui a suscité l'intérêt initial du laboratoire du Dr J. Angulo pour une étude plus approfondie de cette protéine est qu'elle est activée par les radiations ionisantes, les ultraviolets ou autres stress génotoxiques (Biard et al. 1997; Kannouche et al. 1998; Masson et al. 2001) et semble ainsi constituer un facteur de réponse aux stress. L'irradiation aux UVC de fibroblastes humains ou murins conduit à l'augmentation du transcrit du gène *KIN17* 7 heures après irradiation, avec un maximum atteint à 16 heures (Kannouche et al. 1998; Kannouche et al. 2000). La protéine KIN17 pourrait participer à un mécanisme, conservé chez les mammifères, permettant de prendre en charge les lésions résiduelles de l'ADN.

En revanche, dans des cellules déficientes en protéines XPA et XPC, la régulation positive de la protéine KIN17 après irradiation aux UVC est inhibée (Masson et al. 2003). Dans la lignée XP44RO issue d'une métastase testiculaire de mélanome d'un patient XP-C, l'induction du gène *KIN17* après UVC est restaurée par complémentation de la lignée avec un vecteur rétroviral portant l'ADNc du gène *XPC* (Masson et al. 2003). Des résultats similaires ont été observés chez la souris en comparant des fibroblastes normaux et XPA<sup>-/-</sup> (Blattner et al. 2000). Les protéines XPA et XPC, en dehors de leur rôle dans le système de réparation NER (Nucléotide Excision Repair), participent à une voie de signalisation des dommages aboutissant notamment à l'augmentation de l'expression de certains gènes. L'activité des protéines XPA et XPC est donc nécessaire à la réponse du gène *KIN17* après irradiation aux UVC.

### **2.2.3. La protéine KIN17 est en partie associée à la chromatine**

Une fraction de la protéine KIN17 entière et endogène des cellules HeLa se fixe directement à la chromatine (Biard et al. 2002). Elle est associée aux structures chromatiniennes à la fois dans les cellules en prolifération et dans les cellules arrêtées en phase G<sub>2</sub>-M, c'est-à-dire indépendamment du cycle cellulaire.

Suite à l'apparition de dommages de l'ADN causés par l'ajout de MNNG (agent hautement mutagène et monoalkylant), la fixation de la protéine KIN17 à la chromatine est considérablement augmentée (Miccoli et al. 2003). Cette observation indique que la régulation positive du gène *KIN17* observée après la formation de dommages de l'ADN conduit à une augmentation de la protéine liée à l'ADN chromatinien.

### **2.2.4. Le lien entre les mécanismes de réparation de l'ADN et la chromatine**

Une question majeure posée par l'étude des mécanismes de la réponse cellulaire aux dommages de l'ADN est comment l'environnement structural complexe de la chromatine est altéré en présence de lésions de l'ADN (Bakkenist et al. 2004; Peterson et al. 2004). L'unité fondamentale de la chromatine est le nucléosome composé de 146 pb d'ADN enroulées autour

d'un octamère protéique comprenant deux exemplaires de chacune des histones H3, H4, H2A et H2B (Luger et al. 1997). L'ADN nucléosomal est organisé de manière dynamique avec différents niveaux de compaction allant de fibres de 10nm à des structures telles que les chromosomes mitotiques condensés. L'extraordinaire degré de compaction de l'ADN au sein de la chromatine restreint l'accès à l'ADN de protéines de régulation impliquées dans des processus biologiques tels que la réplication, la transcription et la réparation de l'ADN. Pour relâcher cette compaction et rendre l'ADN accessible, la cellule utilise différents mécanismes de remaniement et de modification de la structure de la chromatine : modifications des histones, remodelage dépendant de l'ATP et incorporation de variants d'histones. En ce sens, des études biochimiques suggèrent que la structure de la chromatine est remodelée en présence de cassures double-brin (DSBs) (Smerdon et al. 1978; Jaberabansari et al. 1988; Sidik et al. 1990). De même, l'exposition des cellules à une irradiation UV semble relâcher la structure de la chromatine dans tout le noyau (Hittelman et al. 1984; Ljungman 1989; Rubbi et al. 2003). En plus de ces changements globaux, des perturbations locales de la chromatine et des modifications post-traductionnelles d'histones (acétylation, phosphorylation, méthylation, ubiquitination, ADP-ribosylation, sumoylation, glycosylation ou biotinylation) induites par des dommages de l'ADN peuvent créer des sites d'amarrage pour les protéines de réponse à ces dommages (Peterson et al. 2004). Finalement, une combinaison de modifications traduit un état particulier de la chromatine (de la Cruz et al. 2005) et la reconnaissance de ces combinaisons par des facteurs spécifiques permet le recrutement de protéines ou de complexes de protéines impliqués dans la régulation des différents processus biologiques autour de l'ADN.

Dans ce cadre là, le laboratoire s'intéresse à la caractérisation structurale et dynamique de protéines de la réparation et de la signalisation des dommages de l'ADN capables de reconnaître les modifications d'histones, en identifiant les repliements de domaines proposés comme interagissant avec la chromatine et en recherchant leurs motifs d'interaction avec les histones. Le laboratoire s'est focalisé plus spécifiquement sur différentes protéines modulaires contenant un domaine appartenant à la « royal family », une famille de domaines eucaryotes dérivant probablement d'un ancêtre commun unique (Maurer-Stroh et al. 2003). Les domaines Tudor, Chromo, MBT et PWWP appartiennent à cette famille, ce qui suggère qu'ils partagent certaines fonctions biologiques. Il a été montré récemment et plus particulièrement pendant ma thèse que des domaines Tudor, Chromo et MBT reconnaissent des peptides modifiés d'histones (Kim et al. 2006) ou d'autres protéines (Huang et al. 2007). Au cours de ma thèse, nous avons cherché à savoir si l'un des domaines de la protéine KIN17 humaine dont la distribution des amas hydrophobes rappelle celle des domaines de la Royal Family avait une fonction de reconnaissance de la chromatine modifiée.

## **2.3. Le rôle éventuel de KIN17 dans le métabolisme de l'ARN**

### **2.3.1. Le rôle éventuel de la protéine KIN17 dans le contrôle de la transcription**

La protéine KIN17 possède plusieurs caractéristiques d'un facteur de transcription. En effet, l'analyse de la séquence de la protéine KIN17, très conservée entre les espèces, montre que la protéine possède un domaine de liaison aux acides nucléiques (doigt de zinc en N-

terminal) et un signal de localisation nucléaire dans sa région centrale qui peut permettre le déplacement de la protéine vers le noyau après sa synthèse. Mais surtout, la protéine a la propriété de lier l'ADN simple et double brin et elle a été observée associée à l'ADN chromosomique dans des cellules humaines (Angulo et al. 1991b; Mazin et al. 1994a). De plus, la protéine KIN17 de souris est capable de se substituer au facteur de transcription H-NS (*histone-like nucleoid-structuring*) dans des bactéries déficientes pour ce facteur et d'activer ou de réprimer la transcription de gènes appartenant à l'opéron H-NS (Timchenko et al. 1996).

### 2.3.2. Les mécanismes d'épissage des ARN pré-messagers

La plupart des gènes eucaryotes sont exprimés comme précurseurs d'ARNm (ARN pré-messagers) qui sont convertis en ARNm lors de l'épissage, une étape essentielle de l'expression des gènes pendant laquelle les séquences non-codantes (introns) sont enlevées et les séquences codantes (exons) sont liées ensemble (Yong et al. 2004; Will et al. 2005). L'épissage des ARN pré-messagers nucléaires est réalisé par une machinerie macromoléculaire très dynamique appelée le complexe d'épissage (*spliceosome*). En taille (plusieurs megadaltons) et en complexité, le spliceosome est comparable au ribosome. Les snRNP (*small nuclear ribonucleoprotein particles*) sont les composants majeurs des complexes de l'épissage des ARN pré-messagers dans le noyau des cellules eucaryotes. Elles sont abondantes (par exemple, plus de  $10^6$  U1 snRNP dans les cellules HeLa). Elles ont été purifiées à partir d'extraits de cellules humaines et leur composition en protéines a été déterminée. Chacune des snRNPs (U1, U2, U4, U5 et U6) contient un petit ARN nucléaire riche en uridine (snRNA, U1, U2, U5 ou U4/U6) en association forte avec de multiples protéines : les sept protéines Sm (B/B', D1, D2, D3, E, F et G) et de nombreuses protéines spécifiques à chaque snRNA. Plus de 50 protéines sont recrutées dans le spliceosome en tant que composants stables des snRNPs. Un grand nombre de protéines non associées aux snRNPs sont aussi associées au spliceosome.

La reconnaissance et l'enlèvement de chaque intron *in vitro* nécessitent un réassemblage complet du spliceosome. Cet assemblage a lieu en une suite de différentes étapes distinctes, définies par l'addition et le relargage des snRNPs (Figure 2-2). L'assemblage du spliceosome est initié par la reconnaissance du site d'épissage en 5' de l'ARN pré-messager par U1snRNP pour former le complexe E. A cette étape, U2 snRNP est associé faiblement avec l'ARN pré-messager. Ensuite, l'interaction de U2 snRNP avec l'ARN est stabilisée et conduit à la formation du complexe A, aussi appelé prespliceosome. L'interaction de U4, U5 et U6 préassemblés en un tri-snRNP avec le complexe A conduit au complexe B. Ce complexe contient ainsi tous les U snRNA dans un état précatalytique. Il subit ensuite des réarrangements structuraux majeurs, incluant la déstabilisation et la perte de U1 et U4 snRNPs et le recrutement de facteurs protéiques additionnels conduisant à son activation catalytique et à la formation du complexe B\* (spliceosome activé). Ainsi la formation du spliceosome activé est permise par des stabilisations et déstabilisations de multiples interactions ARN-ARN, ARN-protéine et protéine-protéine. L'épissage est un mécanisme catalytique en deux étapes. Pendant la première étape, le site 5' d'épissage est clivé et le complexe C est formé. Dans la seconde étape, le site 3' d'épissage est clivé, l'intron est libéré et les exons 5' et 3' sont liés pour former l'ARN messenger. A la dissociation du spliceosome, les produits de l'épissage de l'ARN pré-messager sont libérés et les différentes sous-unités du spliceosome sont intégrées dans de nouveaux cycles d'épissage.

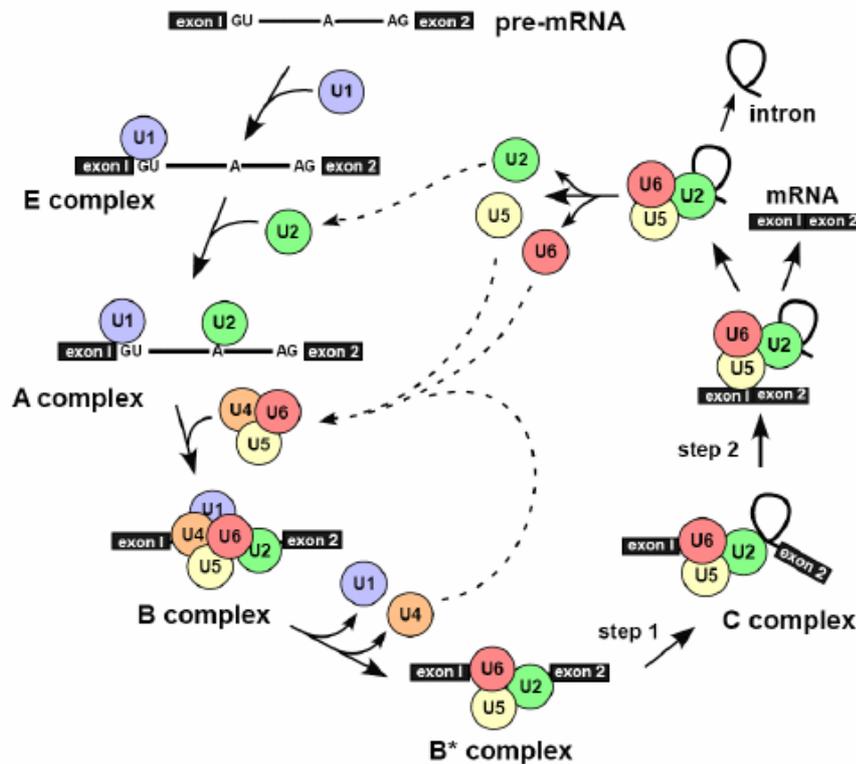


Figure 2-2 : Différentes étapes de l'assemblage de l'épissage dépendant de U2. Chez les eucaryotes, il existe un autre épissage moins abondant (épissage dépendant de U12). Par souci de simplification, ne sont représentées que les interactions des snRNPs. Les séquences d'exon et d'intron sont représentées par des rectangles et des lignes respectivement (Will et al. 2005).

L'épissage des ARN pré-messager est spatialement et temporellement lié à d'autres processus cellulaires, incluant la transcription, la polyadénylation et l'export des ARNm.

### 2.3.3. Le rôle éventuel de la protéine KIN17 dans l'épissage

La protéine KIN17 de souris interagit avec l'ARN *in vivo* et les protéines KIN17 humaine et murine interagissent directement avec différents types d'ARN *in vitro* (Pinon-Lataillade et al. 2004).

De plus, trois analyses par protéomique du spliceosome humain (Makarov et al. 2002; Rappsilber et al. 2002; Deckert et al. 2006) ont permis d'identifier la protéine hKIN17 dans des complexes de l'épissage. En particulier, la présence de la protéine a été détectée dans le complexe B activé (B\*) (Makarov et al. 2002). Dans une analyse en parallèle de quatre purifications du complexe B (Deckert et al. 2006), la protéine hKIN17 a été détectée une seule fois dans ce complexe B, suggérant qu'elle pourrait y être faiblement associée ou que puisqu'associée à d'autres complexes du spliceosome (complexe B\* en particulier), elle pourrait être présente en tant que contaminant dans les préparations de complexe B. En revanche, la protéine KIN17 n'est pas présente dans les complexes A (Hartmuth et al. 2002) et C (Jurica et al. 2002). En outre, KIN17 n'a pas été retrouvée dans d'autres analyses à grande échelle du spliceosome (Neubauer et al. 1998; Zhou et al. 2002; Chen et al. 2007a).

Ghislaine Pinon-Lataillade du laboratoire LGR/DRR/CEA Fontenay-aux-Roses a montré que plusieurs protéines impliquées dans l'épissage des ARN pré-messagers coimmunoprécipitent avec KIN17 dans des cellules HeLa et dans des testicules de souris : il s'agit des protéines U1-70K, ASF/SF2, hnRNPA1 et l'ARN polymérase II, des protéines Sm et des ARN hélicases DDX1 et p68 (communication personnelle). De plus, elle a montré la colocalisation partielle de KIN17 avec les protéines Sm, principaux composants des snRNPs, et avec les protéines SC35 et ASF/SF2, protéines du compartiment nucléaire *speckles*, suggérant qu'une fraction de la protéine KIN17 appartient à ces structures qui sont considérées comme des sites de régulation du pool de facteurs protéiques nécessaires à la transcription et à la maturation des ARN pre-messagers.

---

### **3. L'APPROCHE UTILISEE POUR L'ETUDE STRUCTURALE ET FONCTIONNELLE DE LA PROTEINE HKIN17**

---

Plusieurs propriétés ont été décrites pour la protéine KIN17, dans différents mécanismes cellulaires (réplication de l'ADN, réparation de l'ADN, métabolisme de l'ARN) et elles sont présentées de manière exhaustive dans le paragraphe précédent. Cependant la ou les activité(s) biochimique(s) de la protéine ne sont pas connue(s). Mon travail de thèse s'inscrit dans le cadre d'une collaboration avec le laboratoire LGR/DRR/CEA Fontenay-aux-Roses dont l'objectif vise à préciser le rôle de la protéine KIN17. Alors que ce laboratoire a plutôt une approche de biologie cellulaire, dans notre laboratoire nous abordons plutôt cette recherche d'un point de vue structural. Ainsi, cela permet de conjuguer des informations à l'échelle de la cellule avec des données moléculaires.

Dans ce cadre là, j'ai, au cours de ma thèse, cherché :

- à déterminer la structure de la protéine hKIN17 à l'aide de différentes méthodes biophysiques (cristallographie aux rayons X, RMN et SAXS) ;
- à caractériser les interactions de la protéine hKIN17 avec ses partenaires biologiques à l'aide de différentes méthodes biochimiques ;
- à localiser les surfaces fonctionnelles de la protéine en combinant des informations structurales et fonctionnelles.

Je me suis attachée à donner un éclairage nouveau à la caractérisation fonctionnelle de la protéine KIN17 humaine grâce à l'utilisation de méthodes structurales variées et complémentaires. En parallèle, la recherche de partenaires pour chacun des domaines structuraux identifiés a permis d'apporter des informations concernant le rôle global de la protéine, tout en précisant le rôle de chacun des domaines au sein de la protéine.



# **MATERIEL ET METHODES**



<b>1. LA PREPARATION DES DOMAINES DE LA PROTEINE HKIN17 POUR LES ETUDES STRUCTURALE ET FONCTIONNELLE .....</b>	<b>22</b>
1.1. DOMAINES D'ETUDE DE LA PROTEINE .....	22
1.2. RECHERCHE DES CONDITIONS D'EXPRESSION POUR PRODUIRE LES DOMAINES DE LA PROTEINE .....	22
1.3. PREPARATION DES DOMAINES DE LA PROTEINE.....	24
1.3.1. Purification du domaine K1 .....	24
1.3.2. Purification du domaine K2 .....	25
1.3.3. Purification du domaine K3 .....	25
1.3.4. Purification des autres régions utilisées dans les différentes expériences.....	26
1.4. L'EXPERIENCE DE PULL-DOWN ENTRE LES DOMAINES DE hKIN17 .....	26
<b>2. LES METHODES UTILISEES POUR L'ETUDE STRUCTURALE DE LA PROTEINE HKIN17</b>	<b>27</b>
2.1. METHODES D'ANALYSE DE SEQUENCES ET DE PREDICTIONS DE STRUCTURE .....	27
2.2. CRISTALLOGRAPHIE DU DOMAINE K1 .....	27
2.2.1. Présentation de la cristallographie aux rayons X.....	27
2.2.2. Résolution de structure.....	29
2.2.2.1. Cristallogenèse .....	29
2.2.2.2. Traitement des données .....	30
2.2.2.3. Résolution du problème de phase .....	30
2.2.2.4. Affinement des structures.....	31
2.3. EXPERIENCES RMN SUR LES DIFFERENTS DOMAINES .....	32
2.3.1. Présentation de la résonance magnétique nucléaire.....	32
2.3.2. Expériences RMN réalisées.....	33
2.3.2.1. Echantillons de protéine préparés pour les études RMN .....	33
2.3.2.2. Enregistrement des différentes expériences RMN.....	34
2.4. MESURES DE DICHROÏSME CIRCULAIRE.....	34
2.5. MESURES DE DLS.....	34
2.6. MESURES DE SAXS .....	35
2.6.1. Présentation de la diffusion aux petits angles.....	35
2.6.2. Enregistrement des données de SAXS .....	36
2.6.2.1. Mesures à l'EPFL, Lausanne.....	36
2.6.2.2. Mesures à l'IBBMC, Orsay .....	39
2.6.3. Paramètres structuraux globaux.....	41
2.6.3.1. Détermination de la masse moléculaire des particules en solution .....	41
2.6.3.2. Estimation du rayon de giration par la relation de Guinier .....	41
2.6.3.3. Fonction de distribution des distances P(r).....	42
2.6.4. Détermination de la forme des particules .....	43
2.6.4.1. Modélisation <i>ab initio</i> .....	43
2.6.4.2. Modélisation à partir de modèles existants.....	44
2.7. COMBINAISON DES INFORMATIONS A HAUTE ET BASSE RESOLUTION .....	44
2.7.1. Limites des méthodes à haute résolution pour l'étude des régions flexibles.....	45
2.7.2. Combinaison des informations à haute et basse résolution pour l'étude de protéines multi-domaines et de complexes.....	46
2.7.3. Combinaison des informations à haute et basse résolution pour la caractérisation de changements conformationnels .....	46
<b>3. LES EXPERIENCES POUR L'ETUDE FONCTIONNELLE DE LA PROTEINE HKIN17 .....</b>	<b>48</b>
3.1. ETUDES D'INTERACTION DE LA PROTEINE HKIN17 AVEC LES ACIDES NUCLEIQUES.....	48
3.1.1. Expérience de Southwestern.....	48
3.1.2. Expérience de dot-blot .....	48
3.1.3. Etudes d'interaction des domaines de hKIN17 avec l'ARN .....	48
3.2. ETUDES D'INTERACTION DE LA PROTEINE HKIN17 AVEC LES HISTONES.....	49
3.2.1. Expériences de GST pull-down entre le domaine C-terminal et les histones.....	49
3.2.2. Mise en évidence de l'interaction du domaine K1 avec l'histone H3 par far western.....	50
3.3. EXPERIENCES DE PULL-DOWN POUR LA RECHERCHE DES PARTENAIRES PROTEIQUES ASSOCIES AUX DOMAINES K1 ET K2 .....	51
3.3.1. Présentation de la méthode de pull down.....	51
3.3.2. Préparation d'extraits totaux et nucléaires de cellules HeLa marqués au <sup>35</sup> S.....	52
3.3.3. Essais préliminaires de GST-pulldown .....	52

3.3.4.	<i>Protocole de pulldown préparatif</i> .....	52
3.4.	EXPERIENCES D'ISOLEMENT DE COMPLEXES ASSOCIES AU DOMAINE C-TERMINAL DE hKIN17 .....	53
3.4.1.	<i>Expériences de filtration sur gel</i> .....	53
3.4.1.1.	Principe de la filtration sur gel .....	53
3.4.1.2.	Filtration sur gel des extraits nucléaires incubés avec le domaine C-ter.....	53
3.4.2.	<i>Expériences de gel ND-IEF</i> .....	54
3.4.2.1.	Principe du gel ND-IEF.....	54
3.4.2.2.	Gel ND-IEF des extraits nucléaires incubés avec le domaine C-ter .....	55
3.4.3.	<i>Immuno-précipitation</i> .....	55
3.4.4.	<i>Expérience préparative pour isoler les complexes</i> .....	56
3.5.	IDENTIFICATION DES PROTEINES ISOLEES PAR SPECTROMETRIE DE MASSE.....	57
3.5.1.	<i>Méthodes pour identifier les protéines isolées</i> .....	57
3.5.1.1.	Séparation des protéines .....	57
3.5.1.2.	Identification des protéines.....	57
3.5.2.	<i>Identification des protéines par spectrométrie de masse</i> .....	58
3.5.2.1.	Digestion à la trypsine des protéines dans le gel .....	58
3.5.2.2.	Enregistrement des spectres de masse .....	59
3.5.2.3.	Analyse des spectres de masse .....	60
3.6.	EXPERIENCES CONCERNANT LA RECHERCHE D'UN ROLE EVENTUEL DE hKIN17 DANS L'ÉPISSAGE....	61
3.6.1.	<i>Pull-down de réaction d'épissage</i> .....	61
3.6.2.	<i>Pull-down des snRNPs</i> .....	62
3.6.3.	<i>Recherche des conditions d'immunodéplétion</i> .....	62
3.6.4.	<i>Test d'épissage avec les extraits nucléaires déplétés</i> .....	63
3.6.5.	<i>Test de complémentation</i> .....	63

# 1. LA PREPARATION DES DOMAINES DE LA PROTEINE HKIN17 POUR LES ETUDES STRUCTURALE ET FONCTIONNELLE

## 1.1. Domaines d'étude de la protéine

Dans un premier temps, trois domaines d'étude ont été préparés à partir de la prédiction des régions globulaires de la protéine hKIN17. Le domaine K1 correspond aux résidus 268 à 393, le domaine K2 correspond aux résidus 50 à 160 et le domaine K3 correspond aux résidus 1 à 160 (Figure 1-1).

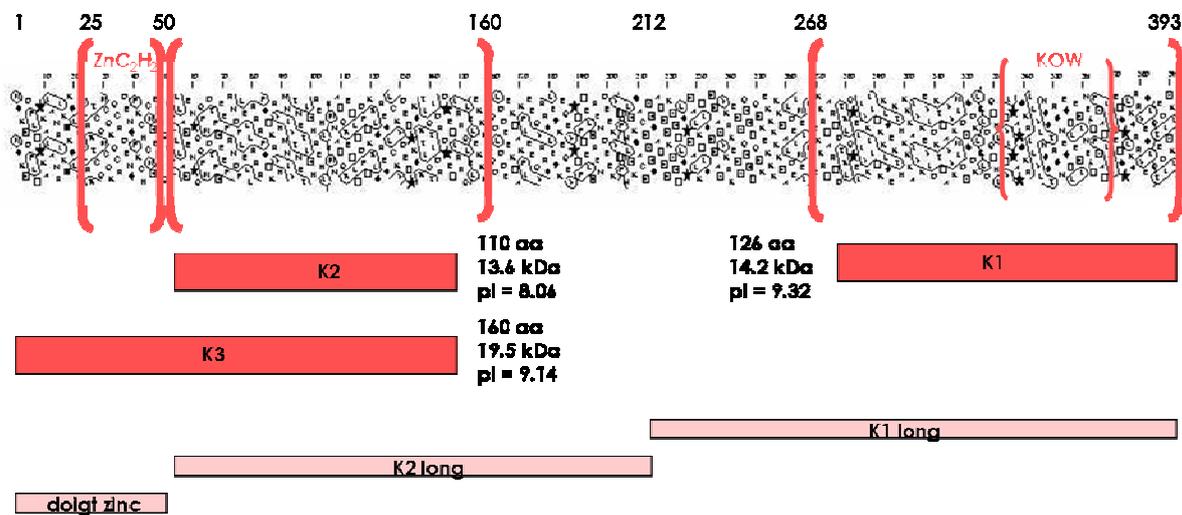


Figure 1-1 : Délimitations des principaux domaines d'étude de la protéine hKIN17 en rouge reportées sur la représentation HCA (Gaboriaud et al. 1987) de la protéine entière. Quelques caractéristiques (nombre de résidus, poids moléculaire et pI) sont données pour chacun de ces domaines. En rose sont représentés trois autres régions d'étude de la protéine hKIN17.

Au fur et à mesure de l'avancée des travaux sur la protéine hKIN17, d'autres régions d'étude ont été définies (Figure 1-1). Les domaines 'K1 long' et 'K2 long' ont été préparés pour ajouter une partie de la région centrale prédite comme étant déstructurée aux domaines K1 et K2 respectivement. Cette région flexible n'a pas été préparée seule car nous pouvions prévoir des problèmes de solubilité et de protéolyse lors de sa production et purification. Le domaine 'doigt zinc' correspond à la région N-terminale de la protéine (résidus 1 à 50).

## 1.2. Recherche des conditions d'expression pour produire les domaines de la protéine

Une étude de l'expression des domaines K1, K2 et K3 de la protéine hKIN17 en fusion avec sept partenaires différents (6His-GST, 6His, Trx, NusA, Gb1, ZZ, 6His-MBP), dans trois souches de bactérie d'expression (Rosetta (DE3) (Novagen), BL21 Star (DE3) et BL21-AI (Invitrogen)) et pour deux conditions d'expression (induction à 37°C pendant 3 heures à partir

de  $DO_{600nm}=0.6$ , induction à  $20^{\circ}C$  sur la nuit à partir de  $DO_{600nm}=1.2$ ) a été menée dans l'objectif de trouver la meilleure condition d'expression de chacun des domaines. Les tests ont été réalisés en microplaque au sein de la plate-forme d'expression des protéines de l'équipe du LMP/DIEP dirigée par M.Gondry (Braud et al. 2005). Une condition d'expression est retenue comme étant intéressante si la quantité de protéine obtenue sous forme soluble est grande par rapport à la quantité de protéine sous forme insoluble. Les meilleurs résultats ont été obtenus en utilisant des bactéries *E. coli* de souche Rosetta (DE3).

Tableau 1-1 : Résultats des tests d'expression en microplaque pour les domaines K1, K2 et K3 de la protéine hKIN17 réalisés en utilisant des bactéries *E. coli* de souche Rosetta. Pour chacun des domaines, 2 conditions d'induction (à  $37^{\circ}C$  et à  $20^{\circ}C$ ) et 5 partenaires de fusion (GST, His, Trx, GB1, ZZ) sont testés. Pour chaque résultat, une notation en nombre de + donne la quantité de protéines solubles (S) et insolubles (I) visibles sur gel. Un signe - signifie qu'il n'y a pas eu d'expression de la protéine de fusion.

<b>37°</b>	<b>GST S</b>	<b>GST I</b>	<b>Hls S</b>	<b>Hls I</b>	<b>Trx S</b>	<b>Trx I</b>	<b>GB1 S</b>	<b>GB1 I</b>	<b>ZZ S</b>	<b>ZZ I</b>
<b>K1</b>	+++	+	++	++	-	-	++(+)	+	+++(+)	+
<b>K2</b>	+	+++	+/-	+	+	+(+)	++(+)	++	++	+++
<b>K3</b>	+	+++	(+)	+(+)	++	++	++	+	++++	++

<b>20°</b>	<b>GST S</b>	<b>GST I</b>	<b>Hls S</b>	<b>Hls I</b>	<b>Trx S</b>	<b>Trx I</b>	<b>GB1 S</b>	<b>GB1 I</b>	<b>ZZ S</b>	<b>ZZ I</b>
<b>K1</b>	++++	+	+++	+	-	-	+	-	+++	+
<b>K2</b>	+++	++	++	+++	++++	+(+)	+	+	+++	+
<b>K3</b>	+	+++	++	+++	+++	+++(+)	+	-	+++	+

Il a fallu ensuite valider ces conditions d'expression obtenues en microplaque dans des conditions de culture en plus grand volume (100ml) avant de produire les protéines à grande échelle. D'après les résultats présentés dans le Tableau 1-1, nous avons cherché à valider en culture de 100ml les conditions suivantes :

<b>domaine K1</b>	<b>domaine K2</b>	<b>domaine K3</b>
<i>Induction à 20°C</i>	<i>Induction à 20°C</i>	<i>Induction à 20°C</i>
GST-K1	GST-K2	Trx-K3
ZZ-K1	Trx-K2	ZZ-K3
	ZZ-K2	<i>Induction à 37°C</i>
		ZZ-K3

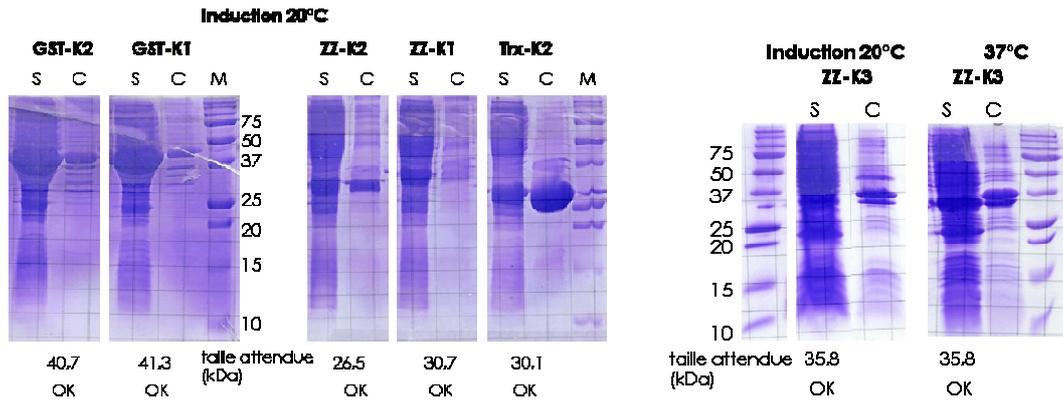


Figure 1-2 : Résultats des tests d'expression des domaines K1, K2 et K3 dans des cultures de 100ml présentés sur gel SDS-PAGE 15%, colorés au bleu de Coomassie. Pour chaque test, les protéines du surnageant (S) et du culot (C) de culture sont mises à migrer. Pour chaque construction, la taille théorique de la protéine de fusion est donnée ainsi que l'accord (OK) entre la taille théorique et la taille observée sur gel.

Les résultats des tests d'expression en bactéries *E. coli* de souche Rosetta DE3 et en culture de 100ml présentés en Figure 1-2 montrent que les meilleurs rapports de protéine soluble sur protéine insoluble sont obtenus pour les domaines K1 et K2 en fusion à la GST dans les conditions d'induction à 20°C et pour le domaine K3 en fusion avec une étiquette ZZ dans les conditions d'induction à 20°C et 37°C.

### 1.3. Préparation des domaines de la protéine

Les protocoles classiques de transformation des bactéries avec un plasmide d'intérêt, de culture bactérienne en milieu riche LB et en milieu minimum (marqué  $^{15}\text{N}$  et/ou  $^{13}\text{C}$ ), d'induction de l'expression d'une protéine d'intérêt et de lyse cellulaire ne sont pas décrits ici mais sont disponibles au laboratoire.

#### 1.3.1. Purification du domaine K1

L'ensemble de la purification est réalisée en chambre froide à 4°C selon le protocole classique. Une première étape de purification par affinité sur une colonne de *Glutathion Sepharose 4B* (Amersham Biosciences) est réalisée pour séparer le domaine K1 en fusion à la GST (GST-K1) de l'ensemble des protéines solubles du surnageant obtenu après lyse cellulaire. Ensuite, une coupure à la TEV-protéase de la protéine de fusion GST-K1 immobilisée sur colonne permet de récupérer le domaine K1, tout en laissant l'étiquette GST accrochée à la colonne. L'efficacité de la coupure de la TEV-protéase sur colonne est très bonne et proche de 100%. A l'issue de cette purification, la pureté de la protéine nous est apparue suffisante et la protéine K1 a été concentrée jusqu'à plusieurs mg/ml (concentration variable selon les lots de protéine).

Un premier échantillon de protéine K1 préparé en milieu riche a été concentré à 1mM et utilisé pour effectuer des mesures de dichroïsme circulaire et pour enregistrer des spectres RMN  $^1\text{H}$  dans le but de vérifier que le domaine K1 était replié en solution. Un échantillon de protéine K1 préparé à partir de 6 litres de milieu minimum marqué  $^{15}\text{N}$  a été concentré à

20mg/ml et utilisé pour l'enregistrement des spectres RMN hétéronucléaires  $^1\text{H}$ ,  $^{15}\text{N}$  et pour la cristallographie. Un échantillon de protéine K1 préparé à partir de 3 litres de milieu minimum marqué  $^{13}\text{C}$ ,  $^{15}\text{N}$  a été concentré à environ 5.5mg/ml (0.4mM) et utilisé pour l'enregistrement des spectres RMN hétéronucléaires  $^1\text{H}$ ,  $^{15}\text{N}$ ,  $^{13}\text{C}$ . Différents échantillons de protéine K1 préparés en milieu riche ont été utilisés pour les expériences de GST *pull-down*, de *Northwestern* et *Southwestern* et de recherche de partenaires.

### **1.3.2. Purification du domaine K2**

Le protocole de purification du domaine K2 en fusion à la GST (GST-K2) pour obtenir le domaine K2 est le même que celui de la protéine GST-K1 décrit dans le paragraphe précédent. Après la coupure à la Tev-protéase, nous avons voulu tester une étape de purification supplémentaire sur colonne de Ni-NTA agarose (QIAGEN) afin de séparer la Tev-protéase qui a une étiquette histidine de la protéine K2 issue de la coupure. Nous avons observé que la protéine K2 restait accrochée à la colonne nickel. Il est possible que le cluster d'histidine observée sur la structure du domaine K2 soit responsable de cette affinité de la protéine pour les ions nickel. Cette étape de purification supplémentaire n'a donc pas été retenue par la suite. Néanmoins, nous obtenons, là encore, en fin de purification sur colonne *Glutathion Sepharose 4B* et coupure à la Tev-protéase, une protéine suffisamment pure (estimation de la pureté de la protéine sur gel).

Un échantillon de protéine K2 préparé à partir de 3 litres de milieu riche a été concentré à 5,3mg/ml (0.4mM) et utilisé pour effectuer des mesures de dichroïsme circulaire et pour enregistrer des spectres RMN  $^1\text{H}$  dans le but de vérifier que le domaine K2 était replié en solution. Différents échantillons de protéine K2 préparés en milieu riche ont été utilisés pour les expériences de GST *pull-down*, de *Northwestern* et *Southwestern* et de recherche de partenaires.

### **1.3.3. Purification du domaine K3**

L'ensemble de la purification du domaine K3 est réalisé en chambre froide à 4°C selon le protocole classique. Une première étape de purification par affinité sur une colonne d'*IgG Sepharose 6 Fast Flow* (Amersham Biosciences) est réalisée pour séparer la protéine ZZ-K3 de l'ensemble des protéines solubles du surnageant obtenu après lyse cellulaire. Ensuite, une coupure à la Tev-protéase de la protéine de fusion ZZ-K3 immobilisée sur colonne permet de récupérer le domaine K3, en laissant l'étiquette ZZ accrochée à la colonne. L'efficacité de la coupure de la Tev-protéase sur colonne est très bonne et proche de 100%. Pour obtenir des échantillons très purs et bien monodisperses, en particulier pour les expériences SAXS, une étape de filtration sur gel a été ajoutée. Selon le volume à injecter, nous avons utilisé une colonne HiPrep 26/60 Sephacryl S-100 High Resolution (Amersham Pharmacia Biotech) ou une colonne Superdex 200 HR 10/30 (Amersham Pharmacia Biotech) qui ont respectivement un optimum de séparation des protéines entre 1kDa et 100kDa et entre 10kDa et 600kDa.

Un échantillon de protéine K3 préparé à partir de 3 litres de milieu riche a été concentré jusqu'à 12mg/ml et utilisé pour enregistrer des spectres RMN  $^1\text{H}$  dans le but de vérifier que le domaine K3 était replié en solution et de comparer ses spectres avec ceux du domaine K2. Un

échantillon de protéine K3 préparé à partir de 5 litres de milieu minimum marqué  $^{15}\text{N}$  a été concentré à 9.6mg/ml (0.5mM) et utilisé pour les expériences de SAXS à Lausanne et Orsay et pour des expériences de titration par RMN. Un échantillon de protéine K3 préparé à partir de 6 litres de milieu minimum marqué  $^{13}\text{C}$ ,  $^{15}\text{N}$  a été concentré à 5.3mg/ml (0.3mM) dans un tampon 50mM NaCl, 20mM Tris-HCl pH7 et utilisé pour l'enregistrement des spectres RMN hétéronucléaires  $^1\text{H}$ ,  $^{15}\text{N}$ ,  $^{13}\text{C}$ . Différents échantillons de protéine K3 préparés en milieu riche ont été utilisés pour préparer des anticorps monoclonaux dirigés contre ce domaine de KIN17, pour les expériences de *Northwestern* et *Southwestern* et pour les études d'interaction.

### **1.3.4. Purification des autres régions utilisées dans les différentes expériences**

La préparation de la protéine ZZ-K1 pour les expériences de *pull-down* préparatif (paragraphe 3.3) a été effectuée en suivant le protocole décrit pour la purification de la protéine ZZ-K3. Les protéines GST-K1, GST-K1 long, GST-K2 et GST-K2 long sont préparés selon le protocole de purification classique de protéine en fusion à la GST. Les mutants K302E et K391E ont été préparés par mutagenèse dirigée en utilisant le kit QuickChange Site-Directed Mutagenesis Kit (Stratagene) à partir de la protéine K1. Les protéines ont été ensuite purifiées en suivant le même protocole que celui utilisé pour préparer la protéine K1. La protéine entière KIN17 utilisée a toujours été préparée par Laurent Miccoli du LGR/DRR/CEA Fontenay-aux-Roses selon le protocole décrit dans Miccoli et al. (2002).

## **1.4. L'expérience de pull-down entre les domaines de hKIN17**

Des cultures de 0.5 ou 1L en milieu LB pour exprimer le domaine K1 en fusion à la GST (GST-K1) et à l'étiquette ZZ (ZZ-K1), le domaine K2 en fusion à la GST (GST-K2), et le domaine K3 en fusion à la GST (GST-K3) et à l'étiquette ZZ (ZZ-K3) sont préparées et les protéines en fusion avec la GST sont purifiées selon le protocole classique. Les surnageants des cultures de 500ml dans lesquelles les protéines ZZ-K1 et ZZ-K3 ont été surexprimées sont déposés sur 1ml de billes d'*IgG Sepharose 6 Fast Flow* (Amersham Biosciences). Les billes sur lesquelles les protéines ZZ-K1 et ZZ-K3 sont accrochées sont lavées en tampon 150mM NaCl, 50mM Tris pH7,6. En final, les billes avec les protéines de fusion sont resuspendues à 50% dans le tampon. 50 $\mu\text{l}$  des billes sont alors prélevés pour chaque condition d'interaction du pull-down et réparti dans des tubes à filtres (nanosep MF 0.22 $\mu\text{m}$ , Pall). Pour préparer les colonnes contrôles ZZ, la Tev-protéase est ajoutée à 4% pour la nuit sur la colonne afin de couper les protéines de fusion ZZ-K1 et ZZ-K3 et le lendemain les colonnes sont lavées avec le tampon pour éliminer les domaines K1 et K3 coupés.

D'après les capacités des billes d'*IgG Sepharose 6 Fast Flow*, nous pouvons estimer qu'une quantité de l'ordre de 100 $\mu\text{g}$  de protéines est fixée sur 50 $\mu\text{l}$  de billes. Les billes des colonnes d'interaction sont resuspendues dans le tampon 150mM NaCl, 50mM Tris pH7,6 et 100 $\mu\text{g}$  de protéine GST-K1 ou 100 $\mu\text{g}$  de protéine GST-K2 ou 50 $\mu\text{g}$  de protéine GST-K3 sont déposées, dans un volume total de 500 $\mu\text{l}$ . L'incubation dure une heure à 4°C sous faible agitation. Les billes sont ensuite centrifugées une minute à 1000g et lavées 3 fois avec 500 $\mu\text{l}$  de tampon. Le bleu de charge est ajouté directement sur les billes et les échantillons sont analysés sur gel SDS-PAGE 15% coloré au bleu de Coomassie.

---

## 2. LES METHODES UTILISEES POUR L'ETUDE STRUCTURALE DE LA PROTEINE HKIN17

---

### 2.1. Méthodes d'analyse de séquences et de prédictions de structure

Pour déterminer les domaines structuraux de la protéine hKIN17, les programmes suivants ont été utilisés :

DISOPRED2 <http://bioinf.cs.ucl.ac.uk/disopred/>

HCA [http://bioserv.rpbs.jussieu.fr/RPBS/cgi-bin/Ressource.cgi?chzn\\_lg=fr&chzn\\_rsrc=HCA](http://bioserv.rpbs.jussieu.fr/RPBS/cgi-bin/Ressource.cgi?chzn_lg=fr&chzn_rsrc=HCA)

SMART <http://smart.embl-heidelberg.de/>

PFAM <http://www.sanger.ac.uk/Software/Pfam/>

### 2.2. Cristallographie du domaine K1

#### 2.2.1. Présentation de la cristallographie aux rayons X

Pour étudier la structure de macromolécules avec un système optique, il faut disposer d'une source de rayonnement compatible avec la résolution à atteindre. Pour pouvoir observer l'arrangement des atomes constituant une molécule, le rayonnement doit avoir une longueur d'onde du même ordre de grandeur que les distances interatomiques, de l'ordre de l'angström. Différents rayonnements ayant cette gamme de longueur d'onde peuvent être obtenus avec des électrons, des neutrons ou des photons X. Ainsi, les méthodes de microscopie et de diffraction électroniques, de diffraction des neutrons et de diffraction des rayons X peuvent apporter des informations structurales à l'échelle atomique. Les rayons X peuvent être obtenus avec des générateurs à anode tournante bombardant une anode de cuivre. Le cuivre excité émet des rayons X à la longueur d'onde de 1,54 Å correspondant à la transition  $K\alpha$ . Les synchrotrons sont des sources de rayons X bien plus performantes : elles produisent des rayons X de faible divergence et de haute brillance permettant de réduire considérablement les temps d'exposition des cristaux et d'améliorer significativement la qualité des données en augmentant le rapport signal/bruit des mesures d'intensités. De plus, ce rayonnement synchrotron couvre de manière continue une très large gamme spectrale, pouvant aller de l'infrarouge (1 eV) aux rayons X durs (50 keV).

Dans le cas des rayons X, la diffusion par une molécule isolée est trop faible pour avoir suffisamment de signal ce qui oblige à faire diffuser un grand nombre de molécules simultanément afin d'additionner leurs contributions. Pour avoir une diffusion cohérente de toutes les molécules, celles-ci doivent être ordonnées dans l'espace. Cet état est atteint dans le cristal qui est un agencement périodique de molécules.

L'interaction entre le rayonnement X et les électrons de la molécule se traduit par la diffusion de ce rayonnement de façon élastique, c'est-à-dire qu'il n'y a pas de perte d'énergie entre l'onde incidente et l'onde diffusée.

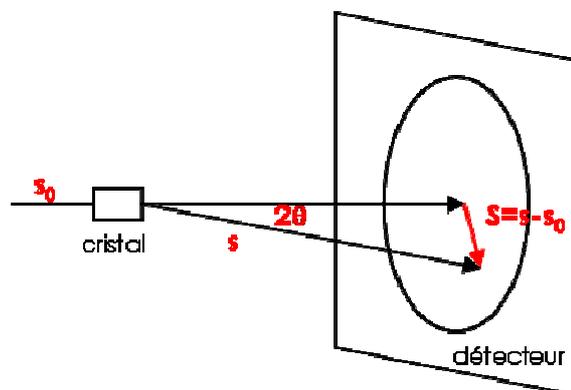


Figure 2-1 : Représentation schématique de la diffusion élastique d'un rayonnement X.  $2\theta$  est l'angle entre la radiation incidente et diffusée.

Dans le cas d'un cristal où la matière est ordonnée de façon périodique, on parle de diffraction puisque les ondes diffusées vont interagir par interférences constructives dans des directions privilégiées de l'espace. On peut définir des familles de plans réticulaires d'indices entiers  $h, k, l$  dont la distance interréticulaire est donnée par  $d_{hkl}$ . La diffraction des rayons X par un cristal résulte de la réflexion des rayons X par les familles de plans réticulaires et n'existe que pour des valeurs de  $\theta$  particulières vérifiant

$$2d_{hkl}\sin\theta = n\lambda$$

où  $\lambda$  est la longueur d'onde incidente et  $n$  un entier. On peut montrer que la norme de  $\mathbf{S}$ , le vecteur de diffusion défini comme  $\mathbf{S} = \mathbf{s} - \mathbf{s}_0$ , est  $2\sin\theta/\lambda = 1/d_{hkl}$ . La résolution de la réflexion  $hkl$  est définie par  $d_{hkl} = 1/S = \lambda/2\sin\theta$ .

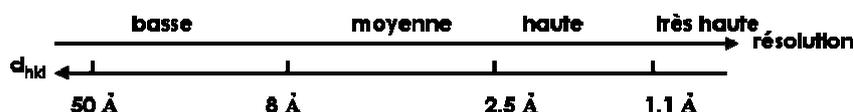


Figure 2-2 : Gammes de résolution des réflexions.

La limite ultime de la résolution des réflexions est fixée par la longueur d'onde ( $d_{\text{limite}} = \lambda/2$ ). Dans la pratique, les désordres statiques et dynamiques dans les cristaux produisent un affaiblissement rapide des intensités diffractées quand  $S$  augmente. Il existe une limite de résolution au-delà de laquelle les réflexions sont trop faibles pour pouvoir être mesurées. La limite de résolution est donc surtout une fonction de la qualité du cristal et des conditions expérimentales de mesure des réflexions.

Le cristal peut être représenté comme le produit de convolution d'un motif (la macromolécule) par un réseau (caractérisé par la maille et le groupe d'espace du cristal). Le motif est répété à chaque noeud du réseau suivant les symétries du groupe d'espace. Le cliché de diffraction est la transformée de Fourier de la densité électronique de l'objet et est donc le produit des transformées de Fourier du motif par celle du réseau. La transformée de Fourier du réseau est un réseau, le réseau réciproque. Il vient donc que le cliché de diffraction représente la transformée de Fourier de la densité électronique échantillonnée aux noeuds du réseau réciproque. Les équations suivantes établissent la relation entre le facteur de structure  $F$  et la densité électronique  $\rho$ .

$$F_{hkl} = \iiint \rho(xyz) e^{2\pi(hx+ky+lz)} dx dy dz$$

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

où  $V$  est le volume de la maille cristalline engendrée par les vecteurs  $\mathbf{a}$ ,  $\mathbf{b}$  et  $\mathbf{c}$  ( $V = \mathbf{a} \cdot \mathbf{b} \wedge \mathbf{c}$ ),  $x$ ,  $y$ ,  $z$  les coordonnées dans l'espace direct et  $h$ ,  $k$ ,  $l$  les indices du réseau réciproque.

Le facteur de structure peut s'exprimer sous la forme d'un nombre complexe mettant en avant les deux composantes de l'onde, l'amplitude et la phase.

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

L'information répartie de manière discontinue au niveau de chaque noeud du réseau réciproque est suffisante pour reconstituer l'objet initial pourvu que l'on connaisse l'amplitude et la phase de l'onde diffusée en chaque point. Or en absence de lentille, l'information de phase est perdue. Lors d'une expérience de diffraction, les clichés obtenus permettent seulement de déterminer l'intensité de l'onde diffractée dans une direction  $hkl$ , qui est proportionnelle au carré du module du facteur de structure. Il est donc nécessaire de retrouver l'information de phase manquante pour pouvoir effectuer la sommation de Fourier qui donne une image de la densité électronique.

## 2.2.2. Résolution de structure

L'obtention des cristaux du domaine K1 a fait l'objet de la publication de l'article 1. La procédure de phasage qui a permis d'obtenir la structure du domaine K1 est décrite dans l'article 1 et l'article 2.

Pour la résolution de structure, les programmes suivants ont été utilisés :

CCP4 <http://www.ccp4.ac.uk/>

SHELXD <http://shelx.uni-ac.gwdg.de/SHELX/>

HySS <http://www.phenix-online.org/download/documentation/phenix-1.24.1b/hyss/overview.html>

SHARP et autoSHARP <http://www.globalphasing.com/sharp/>

ARP/wARP <http://www.embl-hamburg.de/ARP/>

### 2.2.2.1. Cristallogenèse

La méthode utilisée est celle de la diffusion de vapeur en goutte assise. Dans cette technique de cristallisation, une goutte d'un mélange de la solution de protéine et de l'agent précipitant (des sels non chaotropiques, des solvants organiques ou des polymères) est placée sur un support de surface concave, disposé dans un puit dans lequel la solution du précipitant est à une concentration supérieure à celle de la goutte. Le système est fermé avec une lamelle et scellé hermétiquement avec de la graisse de silicone. L'équilibre des concentrations s'établit entre la goutte et le réservoir par diffusion des espèces volatiles entraînant une diminution du volume de la goutte et donc une augmentation de la concentration de la protéine et de l'agent précipitant jusqu'à atteindre l'état de supersaturation favorable à la formation de germes cristallins. Pour atteindre cet état de sursaturation il faut agir simultanément sur deux aspects : l'augmentation de la concentration effective de la protéine et

la modulation des forces répulsives ou attractives entre les molécules de protéine. Pour cela, l'utilisation des agents précipitants aux propriétés variables est nécessaire : les sels vont agir sur la force ionique du milieu, les polymères sur la structuration de la couche de solvation autour de la protéine et les solvants organiques sur la constante diélectrique du milieu. De même, une grande diversité d'additifs avec des propriétés particulières peut être utilisée.

#### 2.2.2.2. Traitement des données

Le traitement des données a pour but d'extraire les réflexions indexées par les entiers  $h$ ,  $k$ ,  $l$  et mesurées par les amplitudes  $F$  et leurs écart-types associés  $\sigma_F$  à partir des clichés de diffraction. Ce traitement est effectué de manière automatique avec les programmes DENZO (Otwinowski et al. 1997), MOSFLM (Leslie 1992) ou XDS (Kabsch 1993). Le traitement des données contient deux étapes distinctes : d'une part la détermination de la maille et du groupe de Laue qui caractérisent le réseau cristallin et d'autre part la mesure des intensités qui caractérisent le motif à l'intérieur du cristal.

#### 2.2.2.3. Résolution du problème de phase

Le problème majeur de la cristallographie des macromolécules tient dans la perte d'information des phases des ondes diffractées. Différentes méthodes existent pour résoudre le problème des phases : les méthodes directes utilisées pour résoudre la structure de petites molécules, les méthodes de phasage *de novo* (remplacement isomorphe et utilisation de la diffusion anormale) qui nécessitent l'introduction d'atomes lourds, la méthode de remplacement moléculaire lorsqu'une structure proche est connue et les méthodes utilisant les dégâts d'irradiation. Ces différentes méthodes de phasage sont souvent combinées afin d'obtenir des phases initiales de bonne qualité donnant une carte de densité électronique interprétable.

Les méthodes de phasage par remplacement isomorphe consistent à comparer les intensités de diffraction d'un cristal de la protéine native aux intensités de cristaux dans lesquels des atomes lourds sont fixés à la protéine. Les atomes introduits sont des atomes lourds car ils doivent posséder un nombre d'électrons supérieurs au nombre d'électrons de la protéine pour que les intensités des cristaux de dérivés soient significativement différentes de celles du cristal natif. La contrainte de cette méthode est l'isomorphisme des structures cristallines de la protéine native et des dérivés lourds. Si l'isomorphisme est conservé, on a la relation vectorielle entre les facteurs de structure:  $\mathbf{F}_{PH} = \mathbf{F}_P + \mathbf{F}_H$  avec  $\mathbf{F}_{PH}$ ,  $\mathbf{F}_P$  et  $\mathbf{F}_H$  les facteurs de structure du dérivé lourd, de la protéine native et des atomes lourds respectivement. Les amplitudes  $|\mathbf{F}_{PH}|$ ,  $|\mathbf{F}_P|$  et  $|\mathbf{F}_H|$  sont accessibles expérimentalement. La résolution de la position des atomes lourds peut se faire à l'aide des méthodes de Patterson ou des méthodes directes et une fois la sous-structure des atomes lourds déterminée,  $\mathbf{F}_H$  peut être calculé en module et en phase.  $\mathbf{F}_P$  se déduit facilement de  $\mathbf{F}_{PH}$  et  $\mathbf{F}_H$  par  $\mathbf{F}_P = \mathbf{F}_{PH} - \mathbf{F}_H$ . Si  $\mathbf{F}_{PH}$  et  $\mathbf{F}_P$  ne sont pas colinéaires, deux valeurs de phases  $\alpha_P$  sont possibles. L'information apportée par un seul dérivé lourd (SIR) laisse donc une ambiguïté sur la valeur des phases mais avec un deuxième dérivé, l'indétermination est levée. Cette méthode MIR (Multiple Isomorphous Replacement) a été la première méthode de phasage utilisée pour résoudre la structure de macromolécules (Green et al. 1954). L'utilisation de la diffusion anormale comme source d'information supplémentaire, dans les méthodes SIRAS/MIRAS (Single/Multiple

Isomorphous Replacement with Anomalous Scattering) permet aussi de lever l'indétermination de phase.

Les méthodes utilisant la diffusion anormale (Hendrickson 1991) sont basées sur le phénomène suivant : lorsque la longueur d'onde du rayonnement incident est proche d'un seuil d'absorption d'un atome dans le cristal (la fréquence de l'onde incidente est proche d'une fréquence de transition électronique de l'atome), le phénomène de résonance lors de l'interaction des rayons X avec les électrons de l'atome ne peut plus être négligé. L'onde diffusée par l'atome n'est plus en phase avec l'onde de référence diffusée par un électron à l'origine et cela se traduit par un facteur de diffusion de la forme :

$$f^{\lambda} = f^0 + f' + i f''$$

Les valeurs des paramètres  $f'$  et  $f''$  sont déterminées en mesurant les variations de la diffusion anormale au voisinage du seuil d'absorption du diffuseur. Comme cela est exprimé dans l'équation ci-dessus, le facteur de structure du diffuseur anormal possède trois composantes dont une composante imaginaire. La conséquence est que la loi de Friedel n'est plus vérifiée, c'est-à-dire que  $F_{hkl}$  ( $F_+$ ) et  $F_{-h-k-l}$  ( $F_-$ ) ne sont plus complexes conjugués. La différence anormale correspond à la quantité Dano =  $|F_+| - |F_-|$ . Les amplitudes  $|F_+|$  et  $|F_-|$  sont accessibles expérimentalement. La résolution de la position des atomes anormaux peut se faire à l'aide des méthodes de Patterson ou des méthodes directes et la position des atomes lourds dans la maille permet de calculer les contributions,  $F_H$  et  $F_H''$  de ces atomes au facteur de structure selon

$$F_H = \sum_j (f_j^0 + f_j') \exp(i2\pi\mathbf{h} \cdot \mathbf{r}_j)$$

$$F_H'' = i \sum_j f_j'' \exp(i2\pi\mathbf{h} \cdot \mathbf{r}_j)$$

Les autres atomes de la protéine contribuent au facteur de structure  $F_P$  avec

$$F_P = \sum_j f_j \exp(i2\pi\mathbf{h} \cdot \mathbf{r}_j) = |F_P| \exp(i\alpha_P)$$

Ainsi,  $F_+ = F_P + F_H + F_H''$  et  $(F_-)^* = F_P + F_H - F_H''$  et  $F_P$  se déduit par  $F_P = F_+ - (F_H + F_H'')$  =  $(F_-)^* - (F_H - F_H'')$ , donnant deux valeurs possibles de phases  $\alpha_P$ . La solution de phase dans le cas de la méthode SAD (Single-wavelength Anomalous Diffraction) est ambiguë mais, contrairement à la méthode SIR, les deux solutions ne sont pas équiprobables. La méthode SAD présente l'intérêt de ne nécessiter qu'un jeu unique de données, réduisant ainsi la durée d'exposition du cristal aux rayons X et donc les risques de dommage dus à l'irradiation. La méthode MAD (Multiple-wavelength Anomalous Diffraction) utilise des données de diffraction enregistrées à plusieurs longueurs d'onde, en général autour d'un seuil d'absorption du diffuseur. Son utilisation est devenue possible grâce au rayonnement synchrotron, de longueur d'onde ajustable. Le problème du manque d'isomorphisme est surmonté dans ces méthodes car tous les enregistrements se font sur le même cristal et c'est la structure du cristal dérivé qui est déterminé. Cependant, ces méthodes nécessitent des mesures des intensités de diffraction très précises en raison de la faiblesse relative des différences anormales.

#### 2.2.2.4. Affinement des structures

Les phases peuvent ensuite être améliorées par des méthodes de modification de densité électronique dans l'espace direct (nivellement de solvant, ajustement des histogrammes de densité, moyenne par les symétries non-cristallographiques). À partir de la densité modifiée, de nouvelles phases sont calculées et combinées avec les phases expérimentales. Quand la densité électronique devient interprétable, un modèle initial de la molécule est construit dans

la densité électronique calculée. La procédure d'affinement permet de corriger le modèle initial afin d'obtenir le meilleur accord entre les facteurs de structure observés et ceux calculés à partir du modèle.

En pratique, on procède en alternant des étapes de reconstruction manuelle du modèle initial avec des étapes d'affinement numérique des paramètres du modèle. La reconstruction manuelle du modèle est effectuée sur station graphique en utilisant des logiciels de visualisation du modèle et des cartes de densité électroniques (TURBO (Roussel et al. 1989), XtalView (McRee 1999), COOT (Emsley et al. 2004)). Les cartes de densité utilisées sont essentiellement des cartes résiduelle ou différence qui montrent le désaccord entre le modèle et les facteurs de structure observés : (i) la carte  $F_o - F_c$  ( $F_{obs} - F_{calc}$ ,  $\Phi_{calc}$ ) représentant  $\rho_{obs}(xyz) - \rho_{calc}(xyz)$ . Dans cette carte, un pic positif s'interprète comme des atomes manquants dans le modèle, un pic négatif sur un atome du modèle indique que celui-ci est probablement mal placé ; (ii) la carte  $2F_o - F_c$  ( $2F_{obs} - F_{calc}$ ,  $\Phi_{calc}$ ) permet de visualiser la densité électronique tout en tenant compte des imperfections du modèle. L'affinement numérique, à l'aide du programme REFMAC5 (CCP4 1994), permet de réduire l'écart entre les facteurs de structure observés et calculés en affinant les paramètres géométriques du modèle. En général, le nombre d'observations issues des facteurs de structure expérimentaux est inférieur au nombre de paramètres à affiner (position des atomes et facteur d'agitation thermique), ce qui ne permet pas à l'affinement de converger convenablement. Le nombre d'observations est augmenté en imposant des contraintes géométriques et stéréochimiques au modèle (longueurs de liaison et angles de liaisons standard).

## 2.3. Expériences RMN sur les différents domaines

### 2.3.1. Présentation de la résonance magnétique nucléaire

La RMN utilise les propriétés magnétiques des noyaux des atomes placés dans un champ magnétique très intense. Le paramètre qui caractérise l'état quantique du noyau et détermine son énergie en présence d'un champ magnétique est le spin  $I$ . Seuls les noyaux ayant un spin non nul donne un signal en RMN. En fait, la plupart des atomes intéressants en biologie (H, P, C, N) ont au moins un isotope dont le noyau a un spin  $1/2$ . Les noyaux de ces isotopes ( $^1\text{H}$ ,  $^{31}\text{P}$ ,  $^{13}\text{C}$ ,  $^{15}\text{N}$ ) se comportent comme des petits aimants. Il existe  $2I+1=2$  niveaux d'énergie différents pour ces noyaux. En présence d'un champ magnétique externe très intense (une bobine supraconductrice est plongée dans l'hélium liquide pour créer un champ statique de 10 à 20 Teslas), une partie des spins se trouve dans l'état énergétiquement le plus favorable et il en résulte une aimantation macroscopique nette dans l'échantillon parallèle au champ statique  $B_0$ . Dès que cette aimantation bascule hors de sa position d'équilibre, elle est soumise à un couple qui tend à la faire tourner autour de l'axe du champ  $B_0$  (mouvement de précession). Dans le cas des spins nucléaires, la fréquence de précession est  $\omega = \gamma B_0$  où  $\gamma$  est le rapport gyromagnétique.

Au moyen d'un second champ magnétique  $B_1$  appliqué pendant une durée courte, l'aimantation des spins peut être basculée d'un angle  $\theta$  que l'on peut contrôler. Le nuage des électrons qui entoure les noyaux crée un champ magnétique local qui se superpose au champ principal  $B_0$ . Le champ effectif vu par le noyau est donc légèrement différent  $B_{eff} = B_0 (1 - \sigma)$

où  $\sigma$  correspond à la constante d'écran. Il dépend de la nature des atomes voisins dans la molécule. La fréquence de résonance réelle du noyau est alors  $\omega_{\text{eff}} = \gamma B_{\text{eff}}$ . Lors du retour à l'équilibre de l'aimantation, chaque atome résonne à une fréquence qui lui est propre et qui dépend de son environnement magnétique. Il émet un signal de précession magnétique à cette fréquence et avec une certaine constante d'amortissement.

Au moyen d'une bobine, le signal électromagnétique (signal de précession libre) émis par les spins en rotation est détecté. La superposition des signaux provenant de tous les atomes est enregistrée puis analysée. Le spectre RMN correspond à la transformée de Fourier du signal de précession libre. En jouant sur les différents types d'atomes et en utilisant des séquences d'excitation particulières, certaines propriétés géométriques comme la proximité de deux atomes à travers l'espace ou leur proximité dans le squelette de la molécule peuvent être étudiées de façon sélective.

L'analyse des spectres de RMN ne donne pas accès immédiatement à la structure tridimensionnelle de la molécule en solution. Il faut d'abord attribuer chaque fréquence de résonance à l'atome correspondant dans la molécule, puis utiliser les données géométriques observées (distances et angles dièdres entre atomes) pour calculer la structure tridimensionnelle de la molécule en solution. Ce calcul se fait par modélisation moléculaire en utilisant aussi les données connues sur la structure chimique (connections entre atomes) et spatiale (longueurs des liaisons, valeurs des angles) des protéines. Il permet d'obtenir une famille de structures toutes compatibles avec la géométrie d'une protéine et les données expérimentales. La qualité de ces structures dépend de la précision et du nombre de distances et d'angles mesurés.

La RMN en phase liquide est la méthode classique amenant à la structure atomique de macromolécules en solution. Elle nécessite néanmoins des concentrations importantes de protéines (de l'ordre de 1mM avec une sonde classique et 0.25mM avec une cryosonde) et comporte des contraintes significatives sur la taille des molécules étudiées (30 - 40 kDa au plus) comme sur leurs propriétés de solubilité. Imprécise sur les molécules allongées et les boucles, elle permet d'étudier des molécules partiellement structurées. Le travail en solution donne la possibilité de faire varier de manière continue les paramètres ioniques et thermodynamiques (température, pH, force ionique, dénaturants, détergents). La RMN du solide ouvre quant à elle des possibilités intéressantes pour l'étude de macromolécules sous forme de poudre (donc non cristallisable) et pour les protéines membranaires.

## **2.3.2. Expériences RMN réalisées**

### **2.3.2.1. Echantillons de protéine préparés pour les études RMN**

Les différents échantillons pour l'enregistrement des expériences RMN ont été généralement préparés de la même manière. L'échantillon de protéine concentrée autour du millimolaire est centrifugé 10 minutes à 10 000g pour culotter la protéine qui serait agrégée. Dans le tube RMN, aux 550 $\mu$ l de protéine sont ajoutés 10% D<sub>2</sub>O, 1mM TSP (3-(trimethylsilyl)[2,2,3,3-<sup>2</sup>H<sub>4</sub>] propionate) utilisé comme référence interne de déplacement chimique <sup>1</sup>H, 0.05% NaN<sub>3</sub>, un cocktail d'inhibiteurs de protéases (SIGMA), 1mM PMSF ou

AEBSF et 1mM TCEP lorsque la protéine a des cystéines libres. Le pH est mesuré et ajusté à la valeur voulue en fin de préparation de l'échantillon RMN.

### 2.3.2.2. Enregistrement des différentes expériences RMN

Les expériences RMN ont été enregistrées sur trois types de spectromètre RMN au laboratoire : un spectromètre Bruker DRX-500, un spectromètre Bruker DRX-600 et un spectromètre Bruker Avance 700MHz, tous équipés d'une sonde trois noyaux ( $^1\text{H}$ ,  $^{15}\text{N}$ ,  $^{13}\text{C}$ ). Les données de RMN enregistrées ont été traitées et transformées à l'aide des logiciels Xwinnmr (Bruker), TopSpin (Bruker) ou NMRPipe (Delaglio et al. 1995). Les spectres RMN ont ensuite été analysés et exploités à l'aide du logiciel Felix (Molecular Simulations) ou Sparky (<http://www.cgl.ucsf.edu/home/sparky/>).

## 2.4. Mesures de dichroïsme circulaire

Les mesures de dichroïsme circulaire sur les domaines K2 et K3 ont été effectuées sur un spectromètre Jasco J-815, à température ambiante. Les échantillons de protéine étaient concentrés à 20 $\mu\text{M}$  dans un tampon 15mM NaCl, 5mM phosphate pH6 pour K2 et dans un tampon 10mM NaCl, 5mM Tris-HCl pH7.6 pour K3, dans une cuve de 1mm d'épaisseur (volume de 250 $\mu\text{l}$ ). Des mesures ont été enregistrées pour des longueurs d'onde allant de 190 à 250nm (UV lointain) par pas de 0,5nm avec un temps d'acquisition d'une seconde sur les échantillons de protéine et sur les tampons correspondants afin d'obtenir un blanc à soustraire.

Les mesures brutes de dichroïsme circulaire donnent les valeurs de theta en mdeg en fonction de la longueur d'onde. A partir de ces données, l'ellipticité molaire par résidu en fonction de la longueur d'onde est calculée en utilisant la formule suivante :

$[\text{Theta}]_{\text{MRES}} (\text{deg.cm}^2.\text{dmol}^{-1}.\text{res}^{-1}) = 0.1 * \text{theta} (\text{mdeg}) / (l (\text{cm}) \times c (\text{M}) \times n)$  avec  $l$  la longueur du trajet optique (cm),  $c$  la concentration de la protéine (M) et  $n$  le nombre de résidus dans la protéine.

## 2.5. Mesures de DLS

La technique de DLS (Dynamic Light Scattering) mesure des fluctuations de l'intensité de la lumière diffusée par la protéine en solution en fonction du temps qui apparaissent lorsque les particules sont soumises au mouvement Brownien. L'analyse de ces fluctuations d'intensité permet de déterminer un coefficient de diffusion des particules qui sont converties en distribution de taille puis en rayon hydrodynamique. Il est également possible d'en déduire le taux de polydispersité de la protéine, qui représente une déviation standard de la taille de la particule reportée à son rayon. Cette valeur permet d'apprécier le taux d'agrégation et d'oligomérisation de la protéine. Une polydispersité de la solution de protéine inférieure à 30% est en général considérée comme acceptable pour son étude par SAXS mais il est préférable de trouver des conditions expérimentales (concentration, composition de tampon) qui limite la polydispersité à 20% maximum.

Les mesures ont été effectuées avec l'appareil DynaPro du Laboratoire de Génétique des Virus (Gif-sur-Yvette) dirigé par S. Bressanelli. De manière générale, ces mesures ont été effectuées sur des échantillons de protéine avant leur étude par SAXS pour vérifier l'homogénéité de la solution de protéine et dans le cas où la protéine existe sous forme monodisperse estimer le rayon hydrodynamique de l'objet en solution. Avant les mesure, l'échantillon de protéine (20 $\mu$ l) était centrifugé 10 minutes à 10 000rpm. L'enregistrement et le traitement des mesures ont été effectués avec le logiciel Dynamics associé à l'appareil. Une mesure dure tant que 30 à 50 acquisitions de faible barre d'erreur n'ont pas été enregistrées.

## 2.6. Mesures de SAXS

### 2.6.1. Présentation de la diffusion aux petits angles

La diffusion aux petits angles (SAS, Small Angle Scattering) de rayons X et de neutrons est un outil utilisé pour étudier la conformation en solution de macromolécules biologiques de taille comprise entre quelques kDaltons et plusieurs MDaltons (Koch et al. 2003; Vachette et al. 2003). Les informations contenues dans les données de SAS sont illustrées de manière qualitative sur la Figure 2-3.

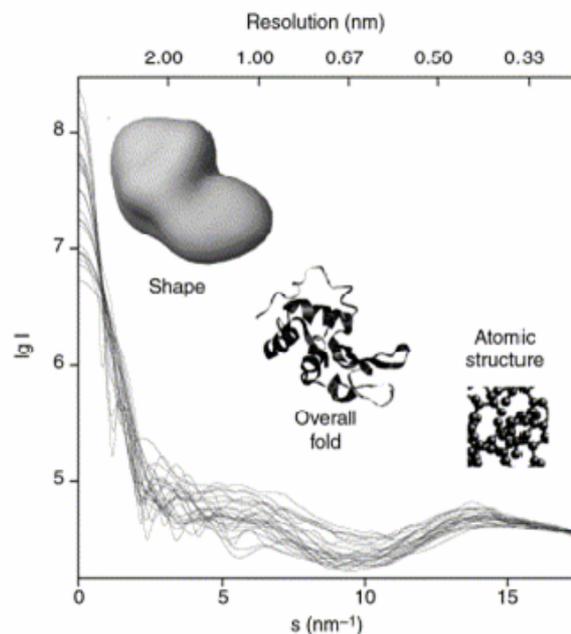


Figure 2-3 : Courbes de diffusion de rayons X calculées pour les modèles atomiques de 25 protéines différentes avec des masses moléculaires allant de 10 à 300 kDa. Les intensités de diffusion sont à une échelle logarithmique, l'axe du haut montre la résolution spatiale ( $\Delta = 2\pi/s$ ) et les différents schémas indiquent les différents niveaux d'organisation structurale (Svergun et al. 2002).

Aux petits angles (équivalent à 2-3 nm de résolution), les courbes décroissent rapidement en fonction de  $s$  et sont essentiellement déterminées par la forme des particules. A résolution moyenne (équivalent à une résolution entre 2 et 0.5 nm), les différences entre les courbes sont moins prononcées et au-dessus d'une résolution de 0.5 nm les courbes sont très similaires. Ainsi, les données SAS contiennent des informations concernant les caractéristiques structurales telles que la forme, la structure tertiaire et quaternaire des particules jusqu'à une résolution de 0.5 nm environ mais ne conviennent pas pour l'analyse de structure atomique.

Les données de SAS peuvent également donner des informations sur les interactions dans la solution.

Dans une expérience de diffusion aux petits angles de rayons X (SAXS, Small Angle X-ray Scattering), la solution de macromolécule est exposée aux rayons X et l'intensité diffusée par cette solution est enregistrée (Figure 2-4).

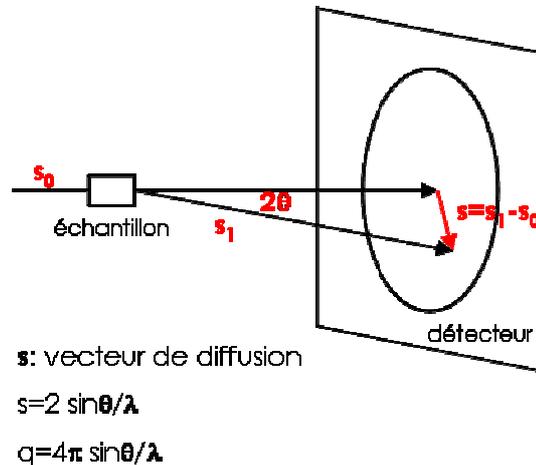


Figure 2-4 : Représentation schématique d'une expérience de SAXS.  $2\theta$  est l'angle entre la radiation incidente et diffusée. Certains auteurs utilisent  $s$  comme étant égal à  $4\pi \sin\theta / \lambda$ , ce qui peut être une source de confusion. Dans ce manuscrit, nous avons tenté de bien distinguer  $q$  et  $s$ .

Les positions et orientations aléatoires des particules en solution dues à l'agitation thermique conduisent à une distribution isotrope des intensités. Pour une solution monodisperse de particules sans interaction (solution idéale), l'intensité diffusée est proportionnelle à la diffusion d'une particule moyennée sur toutes les orientations. Ainsi la courbe de diffusion contient seulement des informations sur la structure tridimensionnelle moyennée de la particule. La diffusion d'une particule étant proportionnelle à la différence au carré du contraste entre la densité électronique de la particule  $\rho(r)$  et celle du solvant  $\rho_0$ , il vient que

$$I(q) = \left\langle \left| \iiint_V (\rho(r) - \rho_0) e^{i\vec{q} \cdot \vec{r}} d^3r \right|^2 \right\rangle.$$

L'intensité diffusée  $I$  est enregistrée comme une fonction du vecteur de diffusion. Le profil de diffusion de la macromolécule est obtenu en soustrayant le signal du tampon d'équilibre.

## 2.6.2. Enregistrement des données de SAXS

### 2.6.2.1. Mesures à l'EPFL, Lausanne

L'appareil SAXS est un système de la compagnie Molecular Metrology et a été installé par la société JJ X-Ray (<http://www.jjxray.dk/sxmm.html>). Il est équipé d'une source à micro-foyer MM002 de la compagnie OSMIC (RIGAKU). L'anode en cuivre fournit un faisceau de longueur d'onde  $\lambda = 1.54 \text{ \AA}$ . L'appareil comprend un détecteur à gaz (chambre multifils, Molecular Metrology). La Figure 2-5 montre une vue d'ensemble de l'appareil.

L'échantillon est placé dans un capillaire de 1,0 mm de diamètre fixé dans un porte-échantillon et dans une enceinte sous vide, ce porte-échantillon est maintenu à une température constante au choix (nous avons travaillé souvent à 4°C) grâce à un système de portoir réfrigéré (Figure 2-5). Un même porte-échantillon est utilisable plusieurs fois, en particulier pour enregistrer les données sur une protéine et sur le tampon correspondant. Pour remplir le capillaire, il est confortable d'avoir 50-60µl de solution.

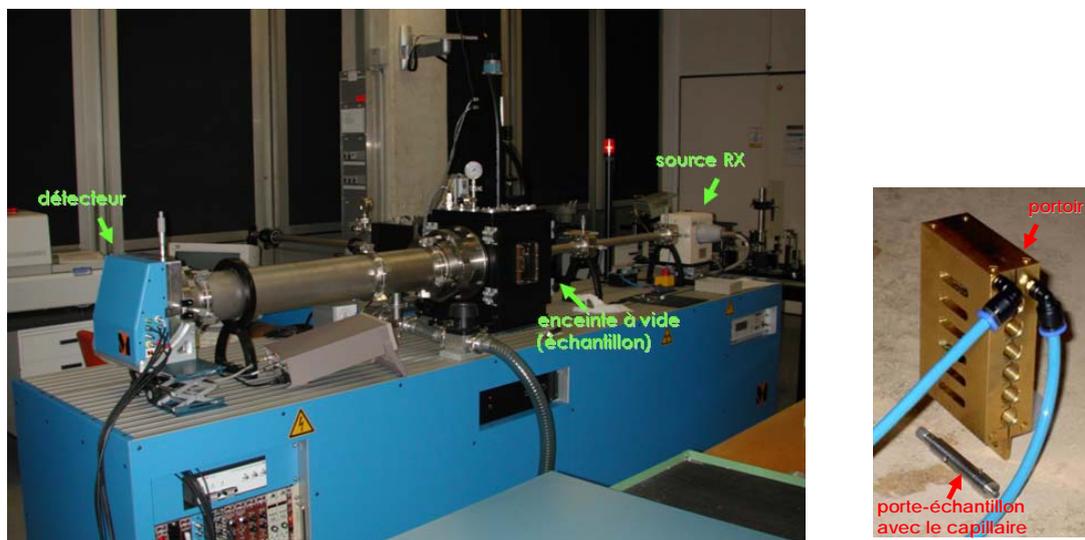


Figure 2-5 : Photo du montage de mesures SAXS de l'EPFL à Lausanne à gauche et photo d'un porte-échantillon et du portoir réfrigéré utilisé sur l'installation SAXS de l'EPFL.

Le traitement des données enregistrées (réduction 1D) se fait grâce à l'utilisation du logiciel SAXSGUI associé à l'installation et programmé en langage Matlab. Une mesure sur un échantillon de béhénate d'argent est utilisée pour la calibration et le centrage des spectres.

Deux séries de mesures ont été enregistrées sur la protéine entière KIN17 et une série de mesures a été enregistrée sur le domaine K3.

protéine <b>KIN17</b> entière dans 500mM NaCl, 10mM DTT, 20mM phosphate pH7.4 mesures à température ambiante	
concentration de la protéine	durée d'exposition
2 mg/ml	10 000s (3 heures)
1 mg/ml	10 000s (3 heures)
tampon	10 800s (3 heures)
protéine <b>KIN17</b> entière dans 150mM NaCl, 10% glycerol, 20mM phosphate pH7.4, anti-protéases mesures à 4°C	
tampon	10 000s (3 heures)
2 mg/ml	10 000s (3 heures)
tampon	100 000s (1 jour)
1 mg/ml	100 000s (1 jour)
protéine <b>K3</b> dans 150mM NaCl, 10mM DTT, 50mM Tris HCl pH7.6 mesures à 4°C	
tampon	10 800s (3 heures)
9,6 mg/ml	10 800s (3 heures)
4,8 mg/ml	10 800s (3 heures)
tampon	14 500s (4 heures)

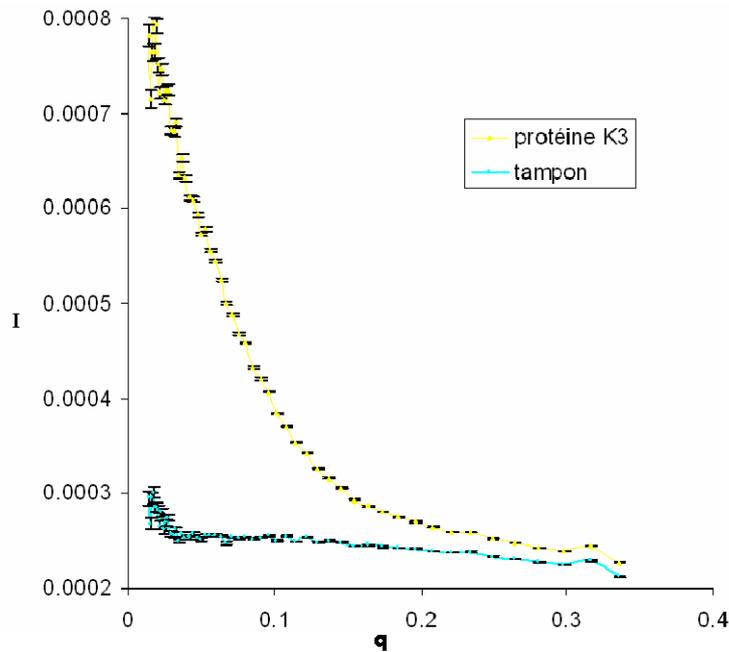


Figure 2-6 : Courbes de l'intensité diffusée en fonction du vecteur de diffusion pour la solution de protéine à 9.6mg/ml et pour le tampon correspondant, enregistrées à Lausanne pendant 3 heures. L'intensité  $I$  est donnée en unité arbitraire et  $q$  en  $\text{\AA}^{-1}$ . Les barres d'erreur ont été ajoutées.

Sur la Figure 2-6, le signal du bruit de fond, c'est-à-dire le signal du tampon, apparaît être très bas indiquant qu'il n'y a que peu de diffusion parasite. Cela est d'autant plus important que nous souhaitons le meilleur rapport signal sur bruit alors que nous travaillons à l'aide d'un appareil de laboratoire et donc d'intensité limitée et sur une protéine de petite taille et de concentration limitée et donc de pouvoir de diffusion limité.

Après traitement des données avec Primus (Konarev et al. 2003), nous obtenons une courbe de diffusion de chaque échantillon de protéine à laquelle le signal du tampon correspondant a été soustrait (Figure 2-7-gauche). Les courbes ont également été normées pour la concentration de la protéine (Figure 2-7-droite).

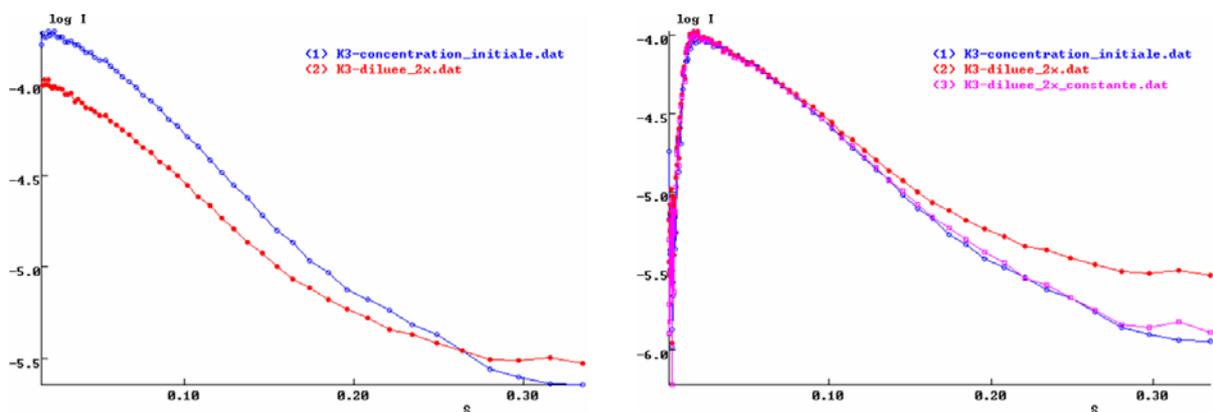


Figure 2-7 : A gauche, superposition des courbes de diffusion de la protéine K3 à la concentration initiale (en bleu) et à la concentration correspondant à une dilution par deux (rouge) après soustraction par le tampon correspondant. A droite, superposition des courbes de diffusion de la protéine K3 normées par la concentration.

Les courbes normées par la concentration se superposent très bien aux petits angles suggérant qu'il n'y a pas d'effet des interactions intermoléculaires (répulsion ou attraction) sur la courbe de diffusion. Cependant, vers les grands angles, la courbe à concentration/2

présente une intensité nettement supérieure à celle de la courbe à concentration initiale. La soustraction du tampon pour la courbe à concentration/2 n'a pas été effectuée correctement. La soustraction d'une constante permet de superposer les 2 courbes (courbe magenta sur la Figure 2-7). Dans la suite, nous n'utiliserons pas ces données enregistrées à la concentration diluée deux fois.

#### 2.6.2.2. Mesures à l'IBBMC, Orsay

L'appareil SAXS est un appareil Nanostar. La source est une anode tournante Microstar avec un foyer de taille apparente  $0.1 \times 0.1 \text{ mm}^2$ . L'anode en cuivre fournit un faisceau de longueur d'onde  $\lambda=1.54\text{\AA}$ . Le faisceau est focalisé par une optique multicouches dite Montel (2 miroirs de Bragg à  $90^\circ$ ) identique à l'optique utilisée sur le montage de Lausanne. Trois collimateurs de diamètres respectifs 0,75, 0,4 et 1,0 mm définissent le faisceau direct. Le détecteur HiStar est un détecteur bi-dimensionnel à gaz (mélange Xe/CO<sub>2</sub>) de 100 mm de diamètre. La Figure 2-8 montre une vue d'ensemble de l'appareil.



Figure 2-8 : Photo du montage de mesures SAXS de l'IBBMC à Orsay. Des détails supplémentaires sont disponibles à l'adresse <http://www.ibbmc.u-psud.fr/ressources/saxs/nanostar.inc.php>.

L'échantillon est placé dans un capillaire de quartz de 1,8 mm de diamètre environ, maintenu à une température constante de  $20^\circ\text{C}$  dans une enceinte sous vide. 25-30 $\mu\text{l}$  de solution sont suffisants pour faire une mesure.

Une série de mesures sur le domaine K3 a été enregistrée sur cet appareil. Des données ont été enregistrées sur des solutions de protéine à deux concentrations, 5.8mg/ml et 2.9mg/ml, dans 150mM NaCl, 10mM DTT, 50mM Tris HCl pH7,6. Des données ont également été enregistrées sur un échantillon de lysozyme à 4.4mg/ml pour la détermination de la masse moléculaire du domaine K3 en solution. Trois spectres de 2 heures ont été enregistrés pour chaque cellule avec le tampon et avec l'échantillon de protéine correspondant. La diffusion d'un échantillon de carbone vitreux enregistrée avant et après chaque spectre est utilisée pour normer chaque spectre à l'intensité transmise. Les données sont réduites à l'aide du programme d'acquisition et de réduction des données associé à l'appareil Nanostar.

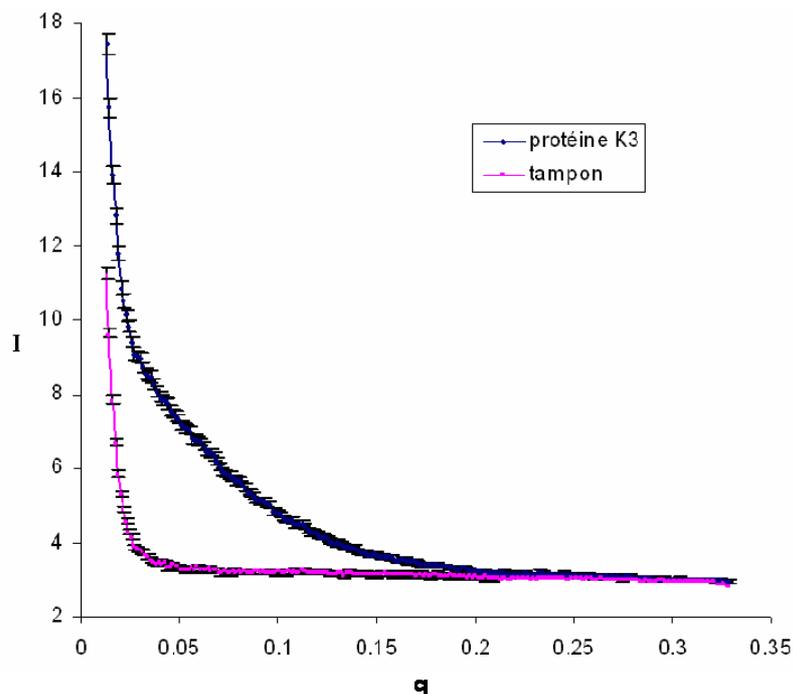


Figure 2-9 : Courbes de l'intensité en fonction du vecteur de diffusion pour la solution de protéine à 2.9mg/ml et pour le tampon correspondant, enregistrées à Orsay pendant 2 heures. L'intensité  $I$  est donnée en unité arbitraire et  $q$  en  $\text{\AA}^{-1}$ . Les barres d'erreur ont été ajoutées.

Les mesures faites à Orsay comptaient parmi les premiers enregistrements effectués sur l'appareil dont le réglage n'était pas encore optimisé, d'où le signal de diffusion parasite important aux très petits angles (Figure 2-9).

Après traitement des données avec Primus (Konarev et al. 2003), nous obtenons une courbe de diffusion de chaque échantillon de protéine à laquelle le signal du tampon correspondant a été soustrait (Figure 2-10-gauche). Les courbes ont également été normées pour la concentration de la protéine (Figure 2-10-droite).

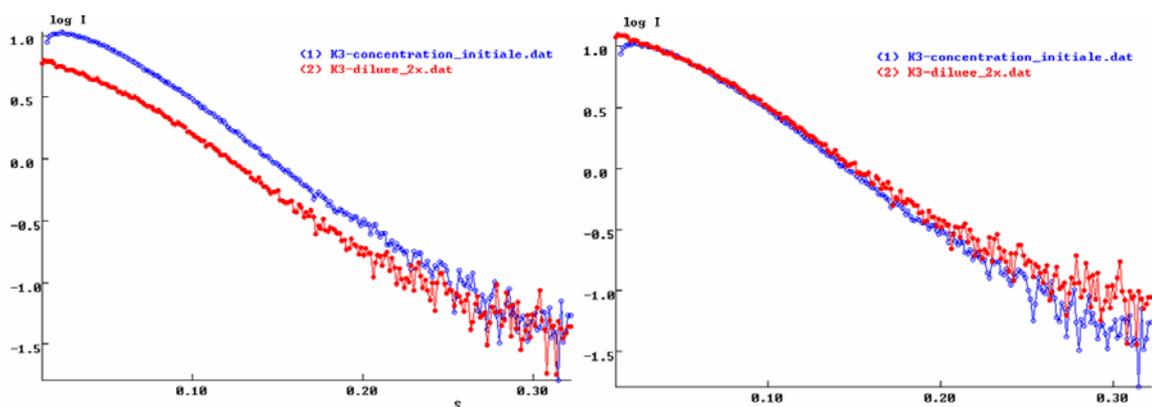


Figure 2-10 : A gauche, superposition des courbes de diffusion de la protéine K3 à la concentration initiale (en bleu) et à la concentration correspondant à une dilution par deux (rouge) après soustraction par le tampon correspondant. A droite, superposition des courbes de diffusion de la protéine K3 normées par la concentration.

Les courbes normées par la concentration se superposent très bien aux petits angles, jusqu'à environ  $q=0.1$ , suggérant qu'il n'y a pas d'effet des interactions intermoléculaires (répulsion ou attraction) sur la courbe de diffusion. Les différences observées aux très petits

angles sont essentiellement dues à une soustraction incorrecte de la forte diffusion parasite consécutive au mauvais réglage de l'appareil. Par ailleurs, les courbes de diffusion enregistrées à Orsay sont très voisines de la courbe de diffusion enregistrée à Lausanne sur la solution de protéine à concentration initiale.

### 2.6.3. Paramètres structuraux globaux

Les données sont analysées dans Primus (Konarev et al. 2003) : superposition des spectres, normation des spectres par la concentration, moyennation des spectres, soustraction des spectres de tampon aux spectres de protéines, détermination du rayon de giration et de  $I(0)$  à partir de la courbe de Guinier.

Tous les programmes utilisés dans ces études sont inclus dans la suite ATSAS 2.1 disponible à l'adresse <http://www.embl-hamburg.de/ExternalInfo/Research/Sax/software.html>

#### 2.6.3.1. Détermination de la masse moléculaire des particules en solution

L'intensité à l'origine (à  $s=0$ ) normalisée par la concentration  $c$  de particules en solution peut s'écrire sous la forme :

$$I(0)/c = N_A M (1-\rho_0\Psi)^2 / \mu^2$$

avec  $\Psi$  le volume électronique spécifique moyen,  $\mu$  le rapport  $M/n$  (pour une protéine, une bonne approximation de ce rapport est 1.87),  $M$  la masse moléculaire de la particule et  $n$  le nombre d'électrons de la particule,  $N_A$  est le nombre d'Avogadro.

Il apparaît que l'intensité à l'origine normalisée par la concentration de particules est proportionnelle à la masse moléculaire  $M$  des particules en solution. Le terme  $N_A(1-\rho_0\Psi)^2 / \mu^2$  peut être considéré comme étant constant pour toutes les protéines. Pour le déterminer, une solution de protéine de masse moléculaire connue (par exemple une solution de lysozyme) et de concentration connue est utilisée. Alors la masse moléculaire des particules en solution peut être déduite de la connaissance de la concentration des particules  $c$  et de la valeur de l'intensité du faisceau incident à l'origine ( $M= I(0)/(c \times \text{constante})$ ).

#### 2.6.3.2. Estimation du rayon de giration par la relation de Guinier

Pour des valeurs de  $s$  telles que  $2\pi s R_g < 1.2$ , Guinier (Guinier 1939) a montré que :

$$I(s) \approx I(0) \exp(-4\pi^2 R_g^2 s^2 / 3)$$

où  $R_g$ , le rayon de giration de la particule, correspond à la distance quadratique moyenne de tous les points de la particule au centre de gravité pondéré par la masse diffusante (le contraste de densité électronique)

$$R_g^2 = \frac{\int_{V_r} (\rho(r) - \rho_0) r^2 dV_r}{\int_{V_r} (\rho(r) - \rho_0) dV_r}$$

Le rayon de giration d'un objet mesure le degré de (non)sphéricité de la particule en solution. Ainsi, pour un volume donné, le plus petit rayon de giration est celui d'une sphère. En pratique, la pente de la droite  $\ln(I(s))$  en fonction de  $s^2$  (courbe de Guinier) permet de déduire

la valeur du rayon de giration de la particule. L'ordonnée à l'origine de la droite donne la valeur de  $I(0)$ .

### 2.6.3.3. Fonction de distribution des distances $P(r)$

La fonction  $P(r)$  représente l'histogramme des distances entre toutes les paires de points de la particule. La fonction  $P(r)$  est nulle lorsque la distance entre deux points est supérieure à  $D_{\max}$ , la dimension maximale de la particule. La fonction  $P(r)$  est calculée à partir de  $I(s)$  en utilisant la méthode de transformée de Fourier indirecte (Glatter 1977).

$$I(s) = 4\pi \int_0^{D_{\max}} p(r) \frac{\sin(2\pi rs)}{2\pi rs} dr$$

Le programme GNOM (Svergun 1992) a été utilisé pour calculer la fonction de distribution des distances interatomiques  $P(r)$  et en déduire le rayon de giration et la distance maximale dans l'objet en solution. La valeur de  $R_g$  est déduite de la fonction  $P(r)$  selon :

$$R_g^2 = \frac{\int_0^{D_{\max}} r^2 p(r) dr}{2 \int_0^{D_{\max}} p(r) dr}$$

La détermination du  $R_g$  à partir de la fonction  $P(r)$  utilise l'ensemble de la courbe de diffusion et non pas seulement les intensités aux petits angles comme dans la méthode de Guinier. Elle est donc moins sensible aux interactions entre les particules ou à la présence d'oligomères en solution.

La fonction  $P(r)$  donne également une indication sur la forme globale de la particule et sur sa structure en domaines. Sur la Figure 2-11 sont représentées des courbes  $P(r)$  pour différents types de protéines. Les molécules globulaires ont une  $P(r)$  en forme de cloche avec un maximum à  $D_{\max}/2$  et les particules allongées ont un maximum de  $P(r)$  pour  $r < D_{\max}/2$ .

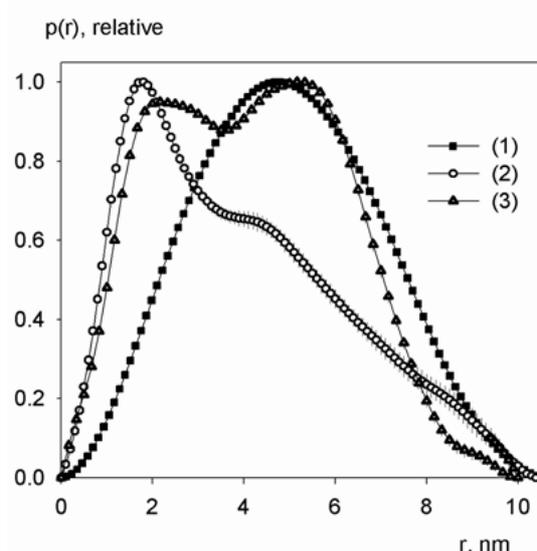


Figure 2-11 : Courbes  $P(r)$  typiques calculées à partir de données expérimentales enregistrées sur (1) une protéine globulaire compacte, (2) une protéine allongée et (3) une protéine à deux domaines (Vachette et al. 2003).

## 2.6.4. Détermination de la forme des particules

L'analyse de données de SAXS consiste à déterminer un modèle de la structure tridimensionnelle d'une macromolécule à partir d'un profil de diffusion à une dimension. Excepté pour le cas trivial d'une particule sphérique, la solution n'est clairement pas unique et plusieurs modèles différents peuvent être en accord avec la même courbe de SAXS. Une approche commune est de superposer un jeu de modèles résultants de reconstructions indépendantes pour voir quels sont les caractères les plus représentés, donc plus ou moins les plus fiables et éliminer les caractéristiques de chaque modèle.

### 2.6.4.1. Modélisation *ab initio*

Nous avons utilisé en parallèle deux méthodes de construction de modèles *ab initio* pour déterminer l'enveloppe à basse résolution de la particule en solution à partir de la courbe de diffusion. L'hypothèse faite en amont de ces méthodes est que la protéine adopte une forme définie en solution. Une protéine structurée occupe un volume compact et les éléments décrivant ce volume doivent être connexes et leur distribution être compacte, deux contraintes qui réduisent très fortement le nombre de configurations possibles.

La méthode implémentée dans le programme DAMMIN (Svergun 1999) décrit un volume à l'aide d'un ensemble de pseudo-atomes. Une sphère de diamètre égal à la taille maximale des particules est remplie par M petites sphères (pseudo-atomes) et chaque pseudo-atome est attribué soit à la particule, soit au solvant. Le programme DAMMIN cherche une configuration compacte de pseudo-atomes qui soit en accord avec les données de diffusion. La minimisation de la fonction d'accord entre les courbes calculée et expérimentale est faite par recuit simulé en partant d'une configuration initiale aléatoire. La fonction d'accord à minimiser est de la forme  $f(x) = \chi^2 + \alpha P(x)$  où  $\chi^2$  mesure l'écart entre les données expérimentales et calculées,  $P(x)$  correspond à un terme de pénalité (compacité et connectivité de la molécule) et  $\alpha$  est le poids de la pénalité.

La méthode implémentée dans le programme GASBOR (Svergun et al. 2001) lui ressemble mais ici la protéine est décrite comme un ensemble de pseudo-résidus centrés sur les C $\alpha$  et séparés de 0.38 nm. Le programme GASBOR cherche un arrangement spatial des pseudo-résidus compatible avec la connectivité d'une chaîne qui soit en accord avec le profil de diffusion expérimental.

Il est clair que ces méthodes donnent plusieurs solutions (plusieurs distributions spatiales de sphères). Faire tourner ces programmes de détermination de la forme de la particule plusieurs fois avec différentes configurations aléatoires initiales génère de nombreux modèles, qui sont tous en accord avec le profil de diffusion. Les modèles obtenus indépendamment les uns des autres peuvent être superposés. Le programme SUPCOMB (Kozin et al. 2000) peut aligner et superposer deux modèles représentés par un ensemble de points par minimisation de la NSD (normalized spatial discrepancy), une pseudo-distance qui donne une mesure de la non similarité entre les deux modèles. Les programmes DAMSEL et DAMSUP de la suite de programme DAMAVER (Volkov et al. 2003) sont utilisés pour aligner les différents modèles deux à deux obtenus avec les programmes DAMMIN et GASBOR et pour déterminer une pseudo-distance (Normalized Spatial Discrepancy NSD) entre les différents modèles et un modèle pris comme référence. Les modèles de faible NSD par rapport à ce modèle de référence sont superposés pour faire apparaître les caractéristiques les plus représentées dans

tous les modèles. Il n'y a pas de modèle moyen calculé car la courbe de diffusion de ce modèle ne correspondrait pas à la courbe de diffusion expérimentale.

#### 2.6.4.2. Modélisation à partir de modèles existants

Le programme BUNCH (Petoukhov et al. 2005) est utilisé pour calculer des modèles d'une protéine lorsque les structures des domaines individuels de la protéine sont disponibles. Le programme BUNCH représente la protéine en utilisant la description atomique à haute résolution pour les domaines connus et par des pseudo-résidus pour les parties inconnues. Les amplitudes partielles de diffusion des domaines connus sont calculées grâce au programme CRY SOL (Svergun et al. 1995). Un calcul Prebunch est lancé avant chaque calcul Bunch pour construire un fichier de coordonnées initial. Les amplitudes des pseudo-résidus de la partie inconnue sont calculées en utilisant le facteur de forme d'un résidu moyen. Un protocole de recuit simulé est employé pour trouver les positions et orientations optimales des domaines dont on a les modèles à haute résolution et les conformations probables des chaînes de pseudo-résidus entre les domaines pour que la courbe de diffusion calculée soit en accord avec la courbe de diffusion expérimentale, en tenant compte de contraintes de connectivité et de compacité.

## **2.7. Combinaison des informations à haute et basse résolution**

La facilité et la vitesse de préparation des échantillons et de collecte de données SAS rendent cette méthode facilement employable en même temps que d'autres méthodes. L'utilisation conjointe du SAS, des autres techniques de détermination de structures à haute et basse résolution et des outils bioinformatiques est un outil puissant pour l'étude de l'organisation structurale de systèmes complexes et hiérarchisés (cf Figure 2-12). Dans la suite de ce paragraphe, nous ne détaillons que certains aspects des utilisations conjointes des méthodes à basse résolution et haute résolution en rapport avec les travaux de thèse.

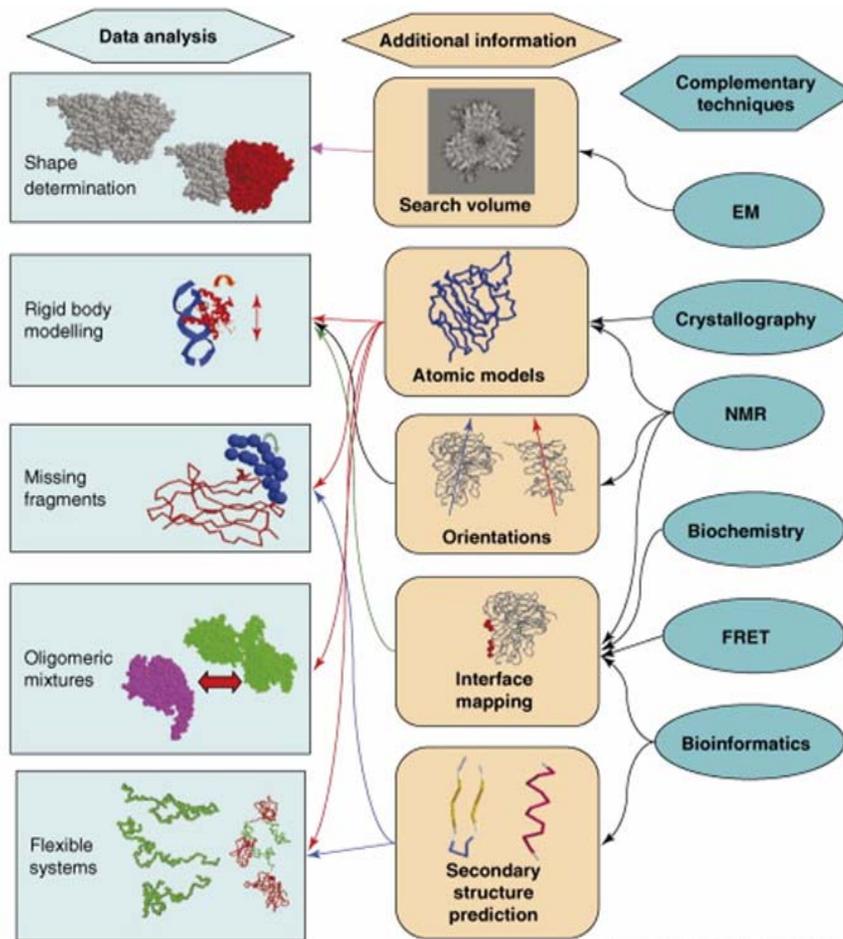


Figure 2-12 : Apports des approches combinant haute et basse résolution dans la détermination des structures de protéines ou complexes de protéines (Petoukhov et al. 2007).

### 2.7.1. Limites des méthodes à haute résolution pour l'étude des régions flexibles

La RMN peut fournir des informations sur la structure des protéines en solution à l'échelle atomique mais elle est encore limitée par la taille des protéines. Par ailleurs, en cristallographie aux rayons X, les régions flexibles sont souvent enlevées des protéines pour faciliter leur cristallisation. C'est pourquoi les domaines de grandes protéines sont souvent étudiés séparément. Les techniques de détermination de structure à haute résolution n'apparaissent donc pas bien adaptées pour étudier les protéines flexibles. En effet, la flexibilité inhérente ou les hétérogénéités conformationnelles dans les protéines résultent souvent en l'absence de boucles ou même de domaines entiers dans les modèles RMN et de cristallographie. Pourtant, la caractérisation des protéines flexibles en solution est cruciale car elle peut fournir les bases structurales pour la compréhension de la fonction de ces protéines. Le principal avantage du SAXS est de donner des informations directes sur la structure globale de la macromolécule en solution, sans les artefacts spécifiques à la cristallographie et sans les limitations de la RMN.

### **2.7.2. Combinaison des informations à haute et basse résolution pour l'étude de protéines multi-domaines et de complexes**

L'analyse structurale de complexes fonctionnels, qui sont souvent de nature transitoire et flexible, nécessite une synergie d'approches complémentaires couvrant une large gamme de taille de macromolécules, de conditions expérimentales et de résolution spatiale et temporelle.

Les analyses par SAS peuvent révéler les structures de protéines multi-domaines (Par exemple, Marquez et al. 2003)) et de complexes de protéine. Toutes les parties d'une protéine ou toutes les protéines d'un complexe contribuent à l'intensité SAS et la configuration de la protéine ou du complexe peut être déterminée en fixant les parties connues de la structure et en ajoutant les parties manquantes pour fitter le profil SAS avec l'hypothèse que la structure tertiaire des sous-unités est conservée. Ainsi, le SAS peut permettre de caractériser l'assemblage, l'architecture et les différents états conformationnels de complexes moléculaires réversibles qui se forment en réponse à un stress de l'environnement ou à un état cellulaire. Ces complexes dynamiques et réversibles sont critiques à certains carrefours dans les cellules, en particulier dans les contrôles et les coordinations de différents mécanismes cellulaires. Dans ce cadre là, l'article de Tsutakawa et al. (2007) présentent différents exemples où le SAXS a été utilisé en combinaison avec la cristallographie aux rayons X pour élucider des structures de protéines et de complexes de protéines impliqués dans des mécanismes de réparation de l'ADN. Dans un autre exemple, la combinaison du SAXS avec des analyses RMN a permis de construire un modèle à basse résolution du complexe dimérique de la protéine L20C avec un ARNr (Raibaud et al. 2003).

Bien que des progrès aient été accomplis dans les dernières années, dans beaucoup de cas encore il reste une ambiguïté sur la façon de placer les structures à haute résolution des domaines isolés avec les orientations appropriées dans les enveloppes à basse résolution. Les données de couplage dipolaire résiduel (RDC) ou de relaxation hétéronucléaire de la RMN peuvent fournir de telles contraintes de rotation afin de définir l'orientation relative des domaines. De plus, des expériences RMN de cartographie de variations des déplacements chimiques et d'augmentation de la relaxation paramagnétique par marquage de spin (*paramagnetic relaxation enhancements from spin-labeling, PRE*) peuvent fournir des informations sur les interfaces moléculaires dans un complexe. Différentes méthodes pour combiner les contraintes RMN et les données de SAXS ont été proposées (Mattinen et al. 2002; Grishaev et al. 2005; Gabel et al. 2006) ; les contraintes RMN (RDC en particulier) permettant de réduire les degrés de liberté angulaires (contraintes d'orientation) et les données de diffusion aux petits angles permettant de limiter les degrés de liberté de translation grâce aux contraintes de translation telles que les distances inter-domaines ou le rayon de giration.

### **2.7.3. Combinaison des informations à haute et basse résolution pour la caractérisation de changements conformationnels**

Le SAS est également puissant comme méthode hybride puisque la combinaison avec d'autres informations structurales peut fournir des informations significatives sur la conformation de protéine en solution.

La cristallographie des protéines peut permettre d'observer des changements conformationnels grâce à des sous-unités non symétriquement reliées de l'unité asymétrique, des régions des valeurs accrues de B et à l'aide de cristaux distincts avec différents cofacteurs ou partenaires. Cependant, le réseau cristallin limite souvent les mouvements des domaines des protéines. Dans ce contexte, la comparaison entre la particule en solution et la structure dans le cristal peut être informative et peut révéler des perturbations causées par des contacts cristallins qui empêchent des mouvements de domaines ou distordent la structure. Par exemple, une étude combinant des expériences RMN, SAXS et des simulations de dynamiques moléculaires a révélé des grandes différences entre la structure en solution et dans le cristal de la protéine Grb2 (Yuzawa et al. 2001). Dans une autre étude, en plus de données de mutagenèse et de calorimétrie, les expériences SAXS ont permis de valider la position et l'orientation du domaine histone dans le modèle de SOS (Sondermann et al. 2005). L'étude par SAXS de la protéine Spa15 a confirmé la structure du dimère allongé de la protéine observée dans des cristaux (van Eerde et al. 2004).

---

### **3. LES EXPERIENCES POUR L'ETUDE FONCTIONNELLE DE LA PROTEINE HKIN17**

---

#### **3.1. Etudes d'interaction de la protéine hKIN17 avec les acides nucléiques**

##### **3.1.1. Expérience de Southwestern**

Des quantités équivalentes (entre 3 et 10 µg selon les essais) des différentes protéines sont déposées sur un gel SDS-PAGE 15%. Les protéines sont ensuite transférées sur une membrane de nitrocellulose (Schleicher & Schuell Bioscience) par transfert semi-sec et elles sont visualisées avec du rouge Ponceau. La membrane est mise à incuber avec du tampon d'hybridation (10mM Tris pH7.4, 50mM NaCl, 0.3% Tween20, 2mM β-mercaptoéthanol, 2mM MgCl<sub>2</sub>, 20µM ZnCl<sub>2</sub>). En parallèle, les ADN sont marqués radioactivement par phosphorylation, c'est-à-dire qu'un phosphate marqué au <sup>32</sup>P est ajouté à leur extrémité 5'. L'ADN radioactif est dilué dans du tampon d'hybridation pour avoir de l'ordre de 10<sup>6</sup> coups par ml de tampon, est chauffé quelques minutes à 90°C et laissé refroidir à température ambiante. L'hybridation est ensuite réalisée en mettant en contact la sonde ADN avec la membrane dans un four à hybrider (Hybridiser) sous faible agitation, pendant au moins une heure, à température donnée. Puis, les membranes sont lavées avec du tampon d'hybridation deux fois pendant 15 minutes. Finalement, les membranes sont mises à exposer au PhosphoImager, ou sur film hyper-sensibles (films MS Kodak) pendant une trentaine de minutes à -80°C ou sur films plus résolutifs (films MR Kodak) pendant la nuit à -80°C.

##### **3.1.2. Expérience de dot-blot**

Différentes quantités de protéines (0, 0.1, 0.2 et 0.5µg) sont déposées sur une membrane de nitrocellulose 0.22µm. Les dépôts sont faits à l'aide du système Bio-dot de Biorad et sont de même volume pour chaque protéine (de 2 à 5µl). La suite de l'expérience est exactement la même que pour le Southwestern décrit ci-dessus.

##### **3.1.3. Etudes d'interaction des domaines de hKIN17 avec l'ARN**

Ces expériences sont décrites dans l'article 2 (page 87).

## 3.2. Etudes d'interaction de la protéine hKIN17 avec les histones

### 3.2.1. Expériences de GST pull-down entre le domaine C-terminal et les histones

Pour chaque condition, 100µl de billes de *Glutathion Sepharose 4B* (GS4B, Amersham Biosciences) sont équilibrés dans le tampon 20mM Tris-HCl pH 7.4, 0.5% TritonX-100 et 0.5 ou 0.75 ou 1M NaCl. Les billes sont mises à incuber avec 100µg de protéine (GST, GST-K1) dans un volume total de 500µl et dans des tubes filtres centrifuges (nanosep MF 0.22µm, Pall) pendant une heure à 4°C sous agitation douce. Les billes sont ensuite centrifugées à 600g, pendant 2 minutes et lavées deux fois avec 500µl de tampon. Puis, 100µg d'un mélange d'histones de thymus de veau (Roche) sont ajoutés aux billes dans un volume total de 500µl et l'incubation dure une heure à 4°C sous agitation douce. Les billes sont ensuite lavées cinq fois avec 500µl de tampon. Le bleu de charge est ajouté directement sur les billes et les échantillons sont analysés sur gel SDS-PAGE 15%, coloré au Bleu de Coomassie.

L'expérience de pull-down a été répétée en condition d'interaction 20mM Tris-HCl pH 7.4, 0.5% TritonX-100 et 500mM NaCl. Les deux bandes d'histones visibles pour chaque protéine appât sur gel coloré au bleu colloïdal ont été découpées et digérées selon le protocole décrit plus loin dans le paragraphe 3.5.2. L'identification par spectrométrie de masse MS a été effectuée au laboratoire DCMR de l'Ecole Polytechnique par JP. le Caer et J. Chamot-Rooke.

L'expérience a été répétée selon le même protocole avec les histones isolées. Les histones utilisées sont des préparations d'histone H3 et d'histone H4 de thymus de veau (Roche Diagnostics). L'interaction a eu lieu dans le tampon 20mM Tris-HCl pH 7.4, 0.5% TritonX-100 et 1M NaCl. L'histone H4 s'agrégeant sur les billes, les lavages ont été effectués avec une centrifugation plus forte (jusqu'à 1200g).

L'expérience a ensuite été répétée en ajoutant une étape d'éluion des histones retenues. Pour chaque condition, 10µl de billes de *GS4B* sont équilibrés dans le tampon 20mM Tris-HCl pH 7.4, 0.5% Triton X-100 et 0.3 ou 0.5 ou 1M NaCl. Les billes sont mises à incuber avec 10µg de protéine (GST, GST-K1) ou sans protéine pour la condition 'billes seules' dans un volume total de 500µl et dans des tubes filtres centrifuges pendant 45 minutes à 4°C sous agitation douce. Les billes sont ensuite lavées deux fois avec 500µl de tampon. Puis, 10µg d'histone H3 de thymus de veau sont ajoutés aux billes dans un volume total de 500µl et l'incubation dure toute la nuit à 4°C sous agitation douce. Les billes sont lavées cinq fois avec 500µl de tampon et sont éluées avec 20µl de tampon GSH 100mM, 50mM Tris HCl. Le bleu de charge est ajouté aux éluions et sur les billes après l'éluion et les échantillons sont analysés sur gel SDS-PAGE 15%, coloré au Bleu de Coomassie.

La comparaison de l'interaction avec l'histone H3 de thymus de veau et l'histone H3 recombinante a ensuite été effectuée. Pour chaque condition, 10µl de billes de *GS4B* sont équilibrés dans le tampon 50mM Tris-HCl pH 7.4, 0.5% Triton X-100 et 500mM NaCl. Les billes sont mises à incuber avec 5µg de protéine (GST, GST-K1 et GST-Asf1 donnée par F. Oschenbein du laboratoire) dans un volume total de 500µl et dans des tubes filtres centrifuges pendant une heure à 4°C sous agitation douce. Les billes sont ensuite lavées deux fois avec 500µl de tampon. Puis, 5µg d'histone H3 de thymus de veau ou 5µg d'histone H3

recombinante (Upstate) selon les conditions sont ajoutés aux billes dans un volume total de 500µl et l'incubation dure une heure trente à 4°C sous agitation douce. Les billes sont lavées cinq fois avec 500µl de tampon et sont éluées avec 20µl de tampon 100mM GSH, 50mM Tris HCl pH8. Le bleu de charge est ajouté aux éluions et sur les billes après l'éluion et les échantillons sont analysés sur gel SDS-PAGE 15%. Après migration, le gel est transféré sur membrane de nitrocellulose pour un immunoblotting avec des anticorps anti-H3 (Abcam) comme premier anticorps et des anticorps anti-lapin (Abcam) comme second anticorps et une révélation à l'ECL (Amersham Biosciences).

L'expérience qui a permis d'obtenir l'histone H3 retenue par la protéine GST-K1 en grande quantité pour l'identification des modifications post-traductionnelles par spectrométrie de masse a été effectuée selon le protocole suivant. Pour chaque condition, 15µl de billes de *GS4B* sont lavés dans du tampon PBS. Les billes sont mises à incuber avec 100µg de protéine (GST, GST-K1) dans un volume total de 500µl et dans des tubes filtres centrifuges pendant une heure à 4°C sous agitation douce. Les billes sont ensuite lavées et équilibrées quatre fois avec 500µl de tampon 50mM Tris-HCl pH 7.4, 0.5% Triton X-100 et 300 ou 500mM NaCl. Puis, 200µg d'histone H3 de thymus de veau sont ajoutés aux billes dans un volume total de 500µl et l'incubation dure deux heures trente à 4°C sous agitation douce. Les billes sont lavées cinq fois avec 500µl de tampon et les billes sont transférées dans des tubes filtres centrifuges propres pour l'éluion. Les billes sont éluées avec 40µl puis 10µl de tampon 20mM GSH, 0.1% triton-X100, 150mM NaCl, 100mM Tris HCl pH8,1. Le bleu de charge est ajouté aux éluions et les échantillons sont analysés sur gel SDS-PAGE 15%. La bande d'histone H3 retenue par GST-K1 et visible sur gel coloré au bleu de Coomassie a été découpée et envoyée à l'IPBS de Toulouse (collaboration avec l'équipe de B. Monsarrat) pour une identification par spectrométrie de masse LC-MS/MS.

### **3.2.2. Mise en évidence de l'interaction du domaine K1 avec l'histone H3 par far western**

2, 10 et 20 pmoles de protéines, BSA (Sigma) ou histone H3 de thymus de veau (Roche), sont déposées sur une membrane de nitrocellulose (Sigma) à l'aide du système Bio-dot de Biorad. La membrane est bloquée pendant une heure à température ambiante dans du TBS-T (150mM NaCl, 20mM Tris-HCl pH8 et 0.1% Tween20) avec du lait (lait en poudre commercial) à 2%. La membrane est ensuite mise à incuber avec 1µM de protéine K1 dans du tampon TBS-T et lait 2%, pendant deux heures, sous agitation douce à 4°C. La membrane est lavée trois fois pendant deux minutes dans du TBS-T. Elle est ensuite transférée pendant 5 minutes dans une solution de formaldéhyde à 5% dans PBS et préalablement chauffée 10 minutes à 65°C. Cette étape permet de consolider les interactions existantes entre les protéines par *cross-linking*. La membrane est ensuite lavée plusieurs fois avec un grand volume de TBS-T pour bien enlever le formaldéhyde. L'immunoblotting de la membrane est réalisé avec les anticorps mAbK36 (Biard et al. 2002) comme premier anticorps et des anticorps GAMPO (Pierce) comme second anticorps.

### 3.3. Expériences de pull-down pour la recherche des partenaires protéiques associés aux domaines K1 et K2

#### 3.3.1. Présentation de la méthode de pull down

C'est une méthode de purification de protéines associée à une protéine appât fusionnée à une étiquette (GST, 6xHis, ZZ, maltose, streptavidine par exemple). La protéine de fusion est incubée avec un mélange de protéines d'extraits cellulaires avant d'être retenue avec sa ou ses protéines partenaires par chromatographie d'affinité (Figure 3-1). La protéine peut aussi être fixée sur la colonne avant de faire passer l'extrait protéique. Le complexe peut être formé *in vivo* ou *in vitro*. Après différents lavages, l'éluion des protéines partenaires peut se faire de différentes manières : à l'aide d'un compétiteur, d'une protéase spécifique coupant entre l'étiquette et la protéine X ou d'un gradient de force ionique.

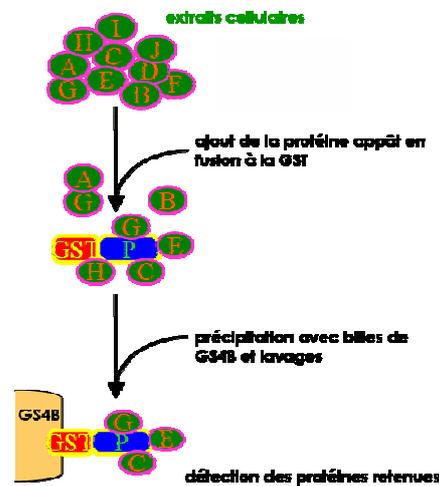


Figure 3-1 : Représentation schématique des différentes étapes d'un GST pull down.

L'avantage de cette technique est d'être relativement simple à mettre en œuvre et rapide (Markillie et al. 2005). Différents contrôles soient indispensables : billes d'affinité + mélange de protéines, billes d'affinité + étiquette + mélange de protéines. Ces contrôles n'excluent pas que des faux positifs puissent être identifiés et donc les résultats obtenus doivent être validés par d'autres techniques. Dans le cas du GST *pull down*, la taille importante de la GST peut entraîner un encombrement stérique qui peut gêner la fixation de certaines protéines. Dans ce cas là, l'étiquette GST peut être mis en C-terminal si elle était en N-terminal et inversement ou un espaceur peut être ajouté entre l'étiquette et la protéine d'intérêt ou enfin l'expérience peut être répétée avec une autre étiquette. Par ailleurs, le GST *pull down* est peu adapté à l'étude des dimérisations car la GST est connue pour dimériser.

### **3.3.2. Préparation d'extraits totaux et nucléaires de cellules HeLa marqués au <sup>35</sup>S**

La procédure suivante est appliquée pour obtenir des extraits cellulaires de 10 boîtes de Pétri de cellules HeLa. Les cellules HeLa sont mises en culture dans du milieu DMEM (Dulbecco's Modified Eagle Medium, Gibco) auquel 10% de sérum de veau fœtal, 2mM glutamine, 100U/ml de pénicilline et 100U/ml de streptomycine sont ajoutés. Les cultures sont arrêtées lorsque les cellules arrivent à confluence. Quatre heures avant l'arrêt des cultures, les cellules sont transférées dans du milieu DMEM sans méthionine auquel 2% de sérum de veau fœtal sans méthionine et 100 µCi/ml de méthionine marquée au <sup>35</sup>S sont ajoutés. Les cellules sont ensuite récupérées en grattant les boîtes de culture avec 3ml de PBS froid et centrifugées à 4500g pendant 10 minutes. Le culot de cellules est gardé et lavé avec 10ml de PBS froid (=le culot est resuspendu puis re-centrifugé). Les cellules sont ensuite lysées en ajoutant du tampon de lyse (4ml de PBS, 1% Triton, une pastille d'inhibiteurs de protéases (Roche)) et en laissant incubé environ 15 minutes sur la glace. La solution est centrifugée à 15000g pendant 15 minutes et le surnageant correspond aux extraits totaux. Pour la préparation des extraits nucléaires, le même protocole est suivi pour la culture des cellules HeLa. Par contre, le protocole de lyse est différent et est expliqué en détails dans (Andrews et al. 1991).

### **3.3.3. Essais préliminaires de GST-pulldown**

40 µl de billes de *Glutathion Sepharose 4B* (Amersham Biosciences) à 50% *slurry* sont incubés pendant 1 heure à 4°C avec 40 µg de protéine (GST, GST-K1 et GST-K2). Les billes sont centrifugées à 500g pendant 5 minutes, le surnageant est jeté. Les billes sont lavées 2 fois avec 10 volumes de billes (200µl) de PBS froid et centrifugées à 500g pendant 5 minutes. Ensuite, les extraits nucléaires ou totaux préparés comme décrit ci-dessus sont ajoutés aux billes. 500µl d'extraits, ce qui correspond aux cellules d'une boîte de Pétri à confluence, sont mis à incubé avec les billes en batch pendant une heure à 4°C avec une agitation douce sur roue. Les mélanges billes-extraits sont transférés dans des petites colonnes pour des lavages différents (en 150mM, 300mM et 1M NaCl). Les billes sont centrifugées à 10 000g pendant une minute à 4°C et lavées avec 500 µl de tampon PBS avec 150mM NaCl ou 300mM NaCl ou 1M NaCl. Les lavages sont répétés cinq fois pour bien éliminer les protéines qui sont retenues de manière non-spécifique. A la fin, les billes sont resuspendues dans 500 µl de PBS froid et transférées dans un tube eppendorf. Après une courte centrifugation à 10 000g, le surnageant est retiré et du bleu de charge 4X est ajouté sur les billes. Les échantillons sont analysés sur gel SDS-PAGE et les résultats sont visualisés par autoradiographie en utilisant un PhosphorImager Storm.

### **3.3.4. Protocole de pulldown préparatif**

En tenant compte de la capacité des billes de *Glutathion Sepharose 4B* (Amersham Biosciences), 200 µl de billes GS4B à 50% *slurry* sont incubées pendant une heure à 4°C avec 200µg de protéine (GST, GST-K1 et GST-K2 ou ZZ et ZZ-K1). Les billes sont ensuite centrifugées à 600g, pendant 5 minutes, à 4°C et le surnageant est éliminé. Les billes sont

lavées deux fois en ajoutant cinq volumes de PBS froid et en centrifugeant à 600g, pendant 5 minutes à 4°C. Indépendamment, les extraits nucléaires HeLa (Cilbiotech) sont dilués pour avoir l'équivalent de 20 millions de cellules HeLa par ml dans du tampon PBS à 300mM NaCl, 1% NP40 et cocktails d'inhibiteurs de protéases et les extraits totaux HeLa (Cilbiotech) et nucléaires sont centrifugés à 18 000 rpm pendant 30 minutes à 4°C pour éliminer les précipités. Pour les extraits nucléaires traités à la benzonase, la benzonase (Novagen) a été ajoutée à 25U/ml final et 10mM MgCl<sub>2</sub> ont également été ajoutés pour permettre l'activité de la nucléase. Ensuite, les billes sont incubées avec des extraits nucléaires ou totaux correspondant à environ 10<sup>8</sup>-10<sup>9</sup> cellules HeLa pendant deux à trois heures à 4°C. Après incubation, les billes sont lavées trois fois avec 5ml de PBS froid à 300mM NaCl et centrifugées à 1 500g pendant 5 minutes à 4°C. Après ces premiers lavages qui ont permis d'enlever les principales protéines se liant de manière non-spécifique, les extraits peuvent être remis à incuber avec les billes pour favoriser les interactions spécifiques. Les billes sont ensuite resuspendues dans 400µl de PBS froid à 300mM NaCl pour être transférées dans des petites colonnes afin d'être encore lavées cinq fois avec 500µl de PBS froid à 300mM NaCl. Les protéines sont ensuite éluées avec cinq volumes de billes de PBS froid à 1M NaCl. Les protéines éluées sont précipitées au TCA avant d'être déposées sur gel SDS-PAGE. Le gel est coloré au bleu de Coomassie ou au bleu colloïdal. Par contre, le protocole de coloration à l'argent doit être modifié pour éviter des problèmes d'interférence entre le formaldéhyde utilisé pendant la coloration et la spectrométrie de masse (Richert et al. 2004).

### **3.4. Expériences d'isolement de complexes associés au domaine C-terminal de hKIN17**

#### **3.4.1. Expériences de filtration sur gel**

##### 3.4.1.1. Principe de la filtration sur gel

C'est une méthode de chromatographie d'exclusion, appelée aussi tamis moléculaire. La protéine cible est mise à incuber avec un extrait protéique et, si une ou plusieurs protéine(s) de l'extrait interagissent de manière stable avec la protéine cible, la protéine cible en complexe sortira plus vite de la colonne que la protéine isolée. Pour déterminer dans quelle fraction d'éluution la protéine d'intérêt se trouve, elle doit être facilement repérable : grâce à une détection avec un anticorps, grâce à un marquage radioactif ou grâce à une étiquette repérable. La taille du complexe peut être estimée et il est ensuite possible de séparer les différents constituants du complexe et de les analyser. Cependant avec cette méthode, il n'est pas sûr que toutes les protéines isolées dans les mêmes fractions que la protéine cible interagissent directement avec elle ou soient impliquées dans le même complexe.

##### 3.4.1.2. Filtration sur gel des extraits nucléaires incubés avec le domaine C-ter

Les extraits nucléaires HeLa (Cilbiotech) sont concentrés pour arriver à une concentration de protéines autour de 30mg/ml et équilibrés dans le tampon d'interaction constitué de 4%

glycérol, 100mM NaCl, 20mM Tris pH8, 1mM Na<sub>2</sub>MoO<sub>4</sub>, 0.28% NAG (n-octyl-D-glucopyranoside), cocktail d'anti-protéases (1mM PMSF, 1mM EDTA, 1µg/ml aprotinine, 10µg/ml leupeptine, 10µg/ml pepstatine, 1mM benzamidine). La protéine K1 est ajoutée à 10µM final et l'incubation avec les extraits nucléaires dure plusieurs heures sur glace. Quelques mg de cytochrome c, un indicateur coloré pour la filtration sur gel, sont ajoutés au mélange qui est injecté sur une colonne C16/100 de Sephacryl S200 (Pharmacia) piloté par le système Gradifrac (Pharmacia Biotech). Le fractionnement dure la nuit (0.3ml/min) à 4°C et les fractions sont de 2,7 ml. Les mêmes conditions de fractionnement ont été utilisées pour faire la calibration de la filtration sur gel avec un mélange standard (Biorad, 151-1901).

Pour l'étude de l'effet de la concentration du domaine K1 dans les extraits nucléaires, nous avons suivi le protocole suivant. 3ml d'extraits nucléaires correspondant approximativement à  $7.5 \cdot 10^8$  cellules HeLa sont séparés en deux après équilibration dans le tampon d'interaction décrit ci-dessus. La première moitié des extraits est incubé avec la protéine K1 à 45µM pendant la nuit, sur glace (condition contrôle). Le lendemain, 10% de ce mélange sont prélevés et ajoutés à l'autre moitié des extraits nucléaires pour une incubation pendant la nuit, sur glace (condition 10%(EN+K1)). Les deux mélanges sont successivement fractionnés selon la même procédure que celle décrite ci-dessus.

L'expérience qui a permis d'étudier l'effet d'un traitement des extraits nucléaires par des nucléases a été réalisée selon le protocole suivant. 3ml d'extraits nucléaires correspondant approximativement à  $7.5 \cdot 10^8$  cellules HeLa sont séparés en trois après équilibration dans le tampon d'interaction décrit ci-dessus. 1ml d'extraits nucléaires est incubé directement avec la protéine à 20µM pendant plusieurs heures sur glace (condition contrôle) avant d'être fractionné selon la même procédure que celle décrite ci-dessus. 1ml d'extraits nucléaires est incubé avec de la RNaseA (SIGMA) à 10 µg/ml sur glace pendant 15 minutes avant d'ajouter la protéine K1 à 20µM et de poursuivre l'incubation plusieurs heures (condition traitée avec RNase). Le mélange est ensuite fractionné de la même manière que le contrôle.

### **3.4.2. Expériences de gel ND-IEF**

#### 3.4.2.1. Principe du gel ND-IEF

La technique de gel ND-IEF (*non-denaturing isoelectrofocusing*), encore appelé gel IEF natif, est une méthode qui permet de fractionner des protéines et des complexes en fonction de leur pI apparent. Le ND-IEF est constitué d'un gel granuleux réparti dans un puits horizontal et peut être réalisé dans différentes gammes de pH en choisissant les ampholytes appropriés. La résolution de séparation des protéines est d'autant meilleure que le gel IEF est long. La quantité de protéines qui peut être chargée sur le gel est établie expérimentalement pour éviter une surcharge du gel qui peut conduire à un blocage de la migration à cause de la précipitation de certaines protéines, spécialement pour des pH entre 5 et 6,5. Le temps nécessaire pour que la focalisation des protéines soit atteinte est estimé en utilisant un marqueur coloré, le cytochrome C qui se concentre du côté de la cathode, bien que les protéines ne focalisent pas toutes avec la même facilité. Plusieurs facteurs peuvent influencer la forme de la bande de focalisation d'une protéine: (i) la distribution de charges dans une protéine influence sa migration ; (ii) certaines protéines ont plusieurs isoformes dont les pI diffèrent et qui vont se focaliser dans différentes fractions ; (iii) la forme et l'hydrophobicité de certaines protéines

influencent leur migration (Galat et al. 2004). Le temps de focalisation souvent long peut faciliter la dégradation des protéines, c'est pourquoi un cocktail d'anti-protéases est ajouté au mélange à fractionner.

#### 3.4.2.2. Gel ND-IEF des extraits nucléaires incubés avec le domaine C-ter

1ml d'extraits nucléaires correspondant approximativement à  $2.5 \cdot 10^8$  cellules HeLa est incubé avec la protéine K1 à  $45 \mu\text{M}$  et dialysé sur la nuit contre le même tampon d'interaction que précédemment. Le lendemain, le mélange pour le gel ND-IEF est préparé : dans 5ml d'eau sont ajoutés 5mg de NAG, 100 $\mu\text{l}$  de chaque ampholine (pharmalyte 3-10, pharmalyte 5-8, ampholine 4-6 et pharmalyte 8-10.5, Amersham), 5% de glycérol, 5% xylitol et 100 $\mu\text{l}$  de cytochrome c à 10mg/ml comme indicateur coloré de la migration et de la focalisation des protéines. Ce mélange est ajouté aux extraits nucléaires incubés avec la protéine et réparti dans le puit de gel ND-IEF avec quelques mg de billes Ultrodex (Amersham Biosciences) pour obtenir un gel de texture convenable. La migration du gel a lieu à  $4^\circ\text{C}$  et jusqu'à ce que le cytochrome c apparaisse très focalisé. Après migration, le gel est découpé en 13 tranches de pH et 100 $\mu\text{l}$  d'eau sont ajoutés à chaque fraction avant la mesure de pH. Une petite partie de ces fractions est déposée sur un gel SDS-PAGE 15% pour l'immunoblotting et l'autre partie est déposée sur un gel SDS-PAGE 12% pour la coloration à l'argent.

L'expérience qui a permis d'étudier l'effet d'un traitement des extraits nucléaires par des nucléases a été réalisée selon le même protocole que celui décrit ci-dessus à la différence que les extraits nucléaires (1ml) après dialyse dans le tampon d'interaction ont été traités avec 100 unités de benzonase (Novagen) et 1mM  $\text{MgCl}_2$  pendant plusieurs heures sur glace avant d'y ajouter la protéine K1.

### 3.4.3. Immuno-précipitation

L'immunoprécipitation est en fait un cas particulier du *pull down* et l'affinité utilisée ici est une affinité antigène/anticorps. L'immunoprécipitation se fait en utilisant la protéine A ou la protéine G, couplée de façon covalente à différents supports (billes d'agarose, de sépharose, billes magnétiques), ces protéines ayant une affinité forte pour les anticorps. Différents types d'éluents peuvent être utilisés pour récupérer les protéines retenues par la protéine appât : directement avec un tampon SDS/DTT pour gel, avec des tampons de force ionique croissante ou avec un peptide compétiteur. L'avantage de cette méthode est que le complexe est formé *in vivo* avec la protéine native. De plus, l'affinité antigène-anticorps étant très forte, des lavages à forte force ionique peuvent être faits pour ne récupérer que les protéines partenaires fortement liées sans risque de dissocier le complexe protéine-anticorps. Cependant, dans le cas de complexes de tailles importantes, le ou les épitopes de la protéine peuvent être masqués et ainsi rendre cette méthode inefficace bien que l'utilisation d'un anticorps polyclonal puisse limiter cet effet.

L'anticorps utilisé pour l'immunoprécipitation du domaine K1 est l'anticorps monoclonal mAbK36 (Biard et al. 2002). Par contre, pour l'immunoprécipitation de la protéine hKIN17 entière les deux anticorps mAbK36 et mAbK58 (Biard et al., 2002) ont été utilisés. Ces anticorps sont préparés par l'équipe du Dr. C. Creminon du SPI/DSV/CEA Saclay.

### 3.4.4. Expérience préparative pour isoler les complexes

Le protocole reprend les différentes procédures décrites précédemment pour la filtration sur gel et le gel ND-IEF. Nous apportons ici quelques précisions expérimentales propres à cet essai préparatif. Nous avons utilisé des extraits nucléaires correspondant à 4x20ml de  $5.10^9$  cellules Hela + 10ml de  $2.5.10^9$  cellules Hela, auxquels le cocktail d'inhibiteurs de protéases suivant a été ajouté : 1mM AEBSF, 1mM EDTA, 5 $\mu$ g/ml aprotinine, 1 $\mu$ g/ml leupeptine, 1 $\mu$ g/ml pepstatine, 1mM benzamidine, 1 $\mu$ M bestatine, 1 $\mu$ M phosphoramidon, 1 $\mu$ g/ml antipaine, 1 $\mu$ M E64. Ils sont concentrés dans une cellule ultrafiltration Amicon de 400ml avec une membrane YM1 jusqu'à un volume de 70ml environ ; nous observons alors l'apparition d'agrégats blancs. La protéine K1 est ajoutée à une concentration de 40  $\mu$ M, 1mM  $Na_2MoO_4$ , 0.25% NAG et 10mg de cytochrome c sont également ajoutés.

De cette préparation, 20ml sont fractionnés sur filtration sur gel avec une colonne C26/100 de Sephacryl S200 (Pharmacia) piloté par le système Gradifrac (Pharmacia Biotech). Le fractionnement dure la nuit (0.5ml/min) et les fractions sont de 4,5 ml. Les fractions sont concentrées en trois parties dans la cellule ultrafiltration Amicon de 400ml avec la membrane YM1 : concentration des fractions 43-80 et lavage avec l'eau jusqu'à 15ml, concentration des fractions 33-42 et lavage avec l'eau jusqu'à 30ml (pic 2) et concentration des fractions 20-32 et lavage avec l'eau jusqu'à 45ml (pic 1).

Le reste de la préparation (50ml) est fractionné sur gel ND-IEF. Pour cela, il est mélangé avec 50mg de NAG, 5ml de glycérol 50%, 2.5g de xylitol, 2ml d'ampholines (1ml pharmalyte 3-10, 400 $\mu$ l pharmalyte 5-8, 400 $\mu$ l ampholine 6-8 et 200 $\mu$ l pharmalyte 8-10.5), 50mg de DTT, 1mM EDTA, 3g d'ultradex et l'ensemble est réparti dans un grand puits de gel ND-IEF. La migration a lieu à 4°C: 15mA, 60V pendant 5 heures (jour J), 300V sur la nuit, 900V sur la journée (J+1). Une bande blanchâtre apparaît du côté de l'anode suggérant qu'il y a trop de protéines. Le gel est donc séparé en deux dans deux puits et la migration est poursuivie à 500V sur la nuit, 1000V sur la journée (J+2), 750V sur la nuit et 1000V sur la matinée (J+3). Lorsque la migration est terminée, les fractions de pH sont récoltées pour chaque gel et les fractions équivalentes en pH des deux gels sont mélangées. Ainsi, 14 fractions sont isolées, leur pH est mesuré et elles sont déposées dans des colonnes d'élution afin de récupérer les protéines de ces fractions en faisant une élution à l'eau. Les protéines issues des fractions 1 à 5 sont mélangées et concentrées dans une cellule ultrafiltration Amicon de 400ml avec la membrane YM1. A la fin de la concentration, le pH est remonté à 7.

L'immunoprécipitation 1 est réalisée avec les fractions 1-5 du gel ND-IEF concentrées. Avant de commencer l'immunoprécipitation, la solution est centrifugée à 4500g, pendant 10 minutes, filtrée à 0.22 $\mu$ m et au final le volume est de 20ml. 500 $\mu$ g d'anticorps mAbK36 sont directement ajoutés et l'incubation sur glace dure quatre heures. Ensuite, 400 $\mu$ l de billes de PGS (Amersham) sont ajoutées et l'incubation est poursuivie sous faible agitation à 4°C pendant deux heures. Puis, le mélange est centrifugé à 6000g, le surnageant est retiré et les billes sont lavées trois fois avec 1ml de tampon d'interaction et deux fois avec 1ml de tampon d'interaction sans NAG. Le tampon urée pour pouvoir faire une migration sur gel 2D est déposé directement sur les billes. L'immunoprécipitation 2 est réalisée avec les fractions concentrées en pic 1 de la filtration sur gel de la même manière que ci-dessus sauf qu'après filtration, le volume du pic 1 est de 50ml environ et que 250 $\mu$ g d'anticorps mAbK36 sont ajoutés et que l'incubation dure 6 heures. Des taches visibles sur les gels 2D des protéines isolées par immunoprécipitation sont découpées et préparées pour des analyses LC-MS/MS effectuées au laboratoire DCMR de l'Ecole Polytechnique.

## 3.5. Identification des protéines isolées par spectrométrie de masse

### 3.5.1. Méthodes pour identifier les protéines isolées

#### 3.5.1.1. Séparation des protéines

Le gel dénaturant, de type SDS-PAGE à une dimension est la façon la plus simple de procéder. Il permet de séparer les protéines selon leur taille. Mais avec cette méthode, la résolution des bandes de protéines est limitée bien qu'elle puisse être améliorée sur gel de gradient, avec la taille du gel ou avec la nature du tampon de migration.

Le gel à deux dimensions (Gorg et al. 2004), une technique développée par O'Farrell en 1975, combine deux séparations séquentielles des protéines : isoélectrofocalisation (IEF) dans une première dimension où les protéines sont séparées en fonction de leurs points isoélectriques, puis séparation selon la taille des protéines sur gel SDS-PAGE dans la deuxième dimension.

Des systèmes de chromatographie très résolutifs en haute pression (type HPLC) peuvent également être utilisés pour la séparation des différents constituants d'un mélange et même être couplés directement à des analyses en spectrométrie de masse.

#### 3.5.1.2. Identification des protéines

Le *Western blot* (*immunoblotting*) peut être utilisé pour révéler la présence d'une protéine si on dispose d'anticorps dirigé contre cette protéine.

Le micro séquençage peut être utilisé pour identifier les protéines isolées. Après séparation sur gel, les protéines sont transférées sur une membrane de PVDF et colorées au bleu de Coomassie. Les bandes de protéines sont alors découpées pour être séquencées par dégradation d'Edman. La séquence des seuls premiers acides aminés (10 à 20) est réalisée et est souvent suffisante pour identifier les protéines. Cependant, si l'extrémité N-terminale de la protéine est bloquée (formylée en particulier), le microséquençage n'est pas possible. De plus, l'utilisation de cette méthode est limitée par sa sensibilité moyenne, nécessitant quelques dizaines de ng de protéines.

La méthode plus largement utilisée pour identifier les protéines est la digestion des protéines à la trypsine (ou par une autre protéase) suivie de l'analyse des peptides par spectrométrie de masse. Les protéines sont séparées sur gel coloré au bleu de Coomassie, au bleu colloïdal ou au nitrate d'argent et les bandes de protéines qui paraissent intéressantes sont soumises à une digestion par la trypsine qui permet le clivage de la chaîne protéique en position C-terminale des acides aminés de type lysine et arginine. Les peptides résultants sont analysés par spectrométrie de masse et l'ensemble des masses des peptides constitue la carte massique de la protéine. Cette liste de masses, caractéristique de la protéine initiale, est comparée avec les masses des digestions théoriques calculées pour les protéines présentes dans les bases de données via des moteurs de recherche d'empreinte peptidique.

### **3.5.2. Identification des protéines par spectrométrie de masse**

#### 3.5.2.1. Digestion à la trypsine des protéines dans le gel

Pendant toute la durée de l'expérience, il faut faire attention à manipuler très proprement pour éviter des contaminations par les kératines de la peau ou des cheveux car ces protéines donnent un signal très fort en spectrométrie de masse. Pour cela, il est conseillé d'utiliser des tubes Treff, de nettoyer la paillasse à l'éthanol avant l'expérience, de toujours porter des gants et une blouse et d'éviter les vêtements en laine.

Le gel est découpé en bandes de protéines d'intérêt. Chaque morceau de gel est coupé en plus petits morceaux d'environ 1mm/1mm et lavé dans 100-150µl d'eau pendant 5 minutes, sous agitation douce. Le surnageant est retiré et les morceaux de gel sont recouverts d'acétonitrile 100% pendant 10 minutes sous agitation avant d'être séchés au speedvac. Cette étape de lavage est nécessaire pour éliminer les colorants et le SDS qui peuvent inhiber les protéases. Les étapes suivantes de réduction et alkylation peuvent être facultatives selon les protéines. Le gel est mis à gonfler avec 10mM DTT, 0.1M  $\text{NH}_4\text{HCO}_3$  pendant 30 minutes à 56°C sous agitation. L'échantillon est ensuite ramené à température ambiante et le surnageant est éliminé. Le gel est alors recouvert avec environ un volume (50µl par exemple) d'acétonitrile qui est rapidement remplacé par 55mM iodoacétamide, 0.1M  $\text{NH}_4\text{HCO}_3$  et le mélange est mis à l'obscurité pendant 20 minutes, à température ambiante. Le surnageant est éliminé et le gel est lavé avec un volume de 0.1M  $\text{NH}_4\text{HCO}_3$  pendant 10 minutes sous agitation. Un volume d'acétonitrile est ajouté et le gel est remis à agiter pendant 15 minutes. Le surnageant est enlevé et un dernier lavage est effectué avec un volume d'acétonitrile. Le surnageant est éliminé assez rapidement et le gel est séché au speedvac.

Pour l'étape de digestion tryptique, le gel est d'abord recouvert de 25mM  $\text{NH}_4\text{HCO}_3$ , 5mM  $\text{NaCl}_2$  avec 12.5ng/µl de trypsine (Roche) et mis à incuber pendant environ 40 minutes dans la glace pour permettre à la trypsine de diffuser dans tout le gel mais de ne pas être encore active. Alors, le surnageant est retiré pour éliminer l'excès de trypsine et remplacé par du tampon 25mM  $\text{NH}_4\text{HCO}_3$  pour maintenir le gel immergé. Le gel est incubé à 37°C pendant plusieurs heures (2 heures ou la nuit par exemple) et la trypsine est maintenant active. La durée de la digestion peut être optimisée pour augmenter ou diminuer le pourcentage de coupure. Les peptides issus de la digestion sont ensuite extraits du gel. Pour cela, un volume d'acide formique est ajouté au gel, mis à agiter pendant 15 minutes et soniqué quelques minutes. Le surnageant contenant les peptides est transféré dans un tube propre. Le gel est lavé une seconde fois avec un volume d'acétonitrile, vortexé et le surnageant qui contient encore des peptides est ajouté au surnageant précédent. Les surnageants sont séchés au speedvac puis repris dans 10µl d'acide formique 5% avant d'être dessalés directement sur Zip-Tip (Millipore). Enfin, les échantillons sont séchés au speedvac et stockés à -20°C en attendant l'analyse par spectrométrie de masse.

Pour les analyses LC-MS/MS des protéines isolées sur gel bidimensionnel, le protocole est le même que celui décrit plus haut pour préparer les peptides pour une analyse MALDI à quelques différences près : les étapes de réduction et alkylation ne sont pas faites car ces réactions ont déjà lieu lors de la préparation des protéines pour le gel 2D, la digestion est réalisée avec 100 à 500ng de trypsine par tache de protéines et dure une nuit à 37°C, enfin l'étape de Zip-Tip est supprimé.

## 3.5.2.2. Enregistrement des spectres de masse

De façon classique, un spectromètre de masse est composé (i) d'une source d'ions, enceinte au sein de laquelle les ions sont formés à partir de l'échantillon vaporisé, (ii) d'un analyseur qui trie les ions en fonction de  $m/z$  où  $m$  est la masse de l'ion et  $z$  la valeur absolue du nombre de charge, (iii) d'un détecteur qui compte les ions (valeur de l'intensité de chaque  $m/z$ ) et (iv) d'un enregistreur qui amplifie et traite le signal et permet de visualiser les spectres (Figure 3-2). Le spectre de masse montre donc la variation du courant ionique détecté en fonction de  $m/z$ .

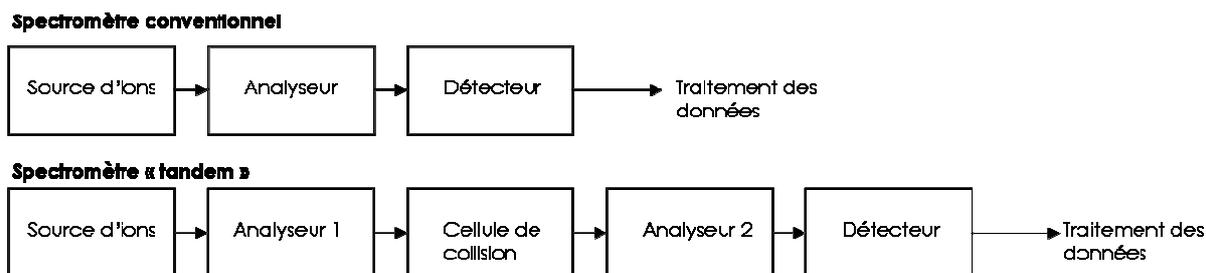


Figure 3-2 : Schéma des spectromètres conventionnel et tandem.

De nombreuses techniques d'ionisation, d'analyse et de détection ont été développées. Plusieurs paramètres sont à prendre en compte pour le choix de la méthode : les propriétés physico-chimiques de l'échantillon à analyser, la quantité d'échantillon disponible et surtout le type d'information recherchée (poids moléculaire, information de structure, formule brute...). Les sources d'ionisation les plus employées pour les composés organiques sont les sources par Ionisation Electronique ou Ionisation Chimique alors que celles utilisées couramment pour les biomolécules sont l'électrospray (ESI) ou le nano-electrospray (nanoESI) ainsi que l'ionisation par désorption laser assistée par matrice (MALDI). Un analyseur de type "temps de vol" (TOF) sépare les ions en fonction du temps qu'ils mettent pour traverser le tube de vol, alors que dans une cellule ICR, les ions sont séparés en fonction de leur fréquence de rotation dans le champ magnétique.

Les appareils construits pour réaliser les expériences MS/MS, les spectromètres tandem, sont constitués (1) d'un premier analyseur qui permet de sélectionner un ion précurseur (ion parent) issu de la source d'ions, (2) d'une chambre de collision où s'effectue une phase d'activation de l'ion précurseur qui conduit à son excitation et à sa fragmentation et (3) d'un second analyseur qui permet de détecter les ions fragments (ions fils) (Figure 3-3). Selon les instruments utilisés, les étapes de sélection et d'analyse des fragments se font dans l'espace (tQ, Q-TOF, TOF-TOF) ou dans le temps (Trappe ionique, FT-ICR, Orbitrap).

Les analyses par spectrométrie de masse MALDI-TOF et LC-MS/MS ont été réalisées au Laboratoire DCMR de l'École Polytechnique en collaboration avec l'équipe de J.Chamot-Rooke, par Jean-Pierre le Caer et Christian Malosse.



Figure 3-3 : A gauche, spectromètre FT-ICR Bruker BioAPEX III (7 Tesla) du laboratoire DCMR de l'Ecole Polytechnique. A droite, spectromètre QToF-Premier de ce même laboratoire.

Pour une analyse en MALDI, 2,5 $\mu$ l de matrice MALDI DHB à 10mg/ml ont été ajoutés aux peptides avant de mettre les tubes à soniquer quelques minutes. Pour chaque bande de protéine, il y a eu deux analyses en spectrométrie de masse, c'est-à-dire que deux dépôts de 1 $\mu$ l sur la cible à irradier. Les analyses des produits issus de la digestion à la trypsine ont été effectuées à l'aide d'un spectromètre Bruker FT ICR-MS avec un champ magnétique de 7T. La sensibilité pour la source d'ionisation MALDI est d'environ 100 fmoles de produit déposé. Les analyses LC-MS/MS des produits issus de la digestion à la trypsine ont été effectuées sur un spectromètre QToF-Premier (Waters Corp., Milford, MA, USA) équipé d'une source d'ionisation nano-ESI (LockSpray). Les spectres MS/MS ont été enregistrés dans une gamme de masse moléculaire de 40 à 2000 Da en utilisant une énergie de collision dépendant de l'état de charge des ions. Les données ont été analysées en utilisant le programme PLGS de Waters. Le couplage a été réalisé avec une chaîne nano-LC Agilent 1100 series. La colonne utilisée est une Zorbax 300SB-C18 3,5 $\mu$ m 150x0.75mm (Agilent).

### 3.5.2.3. Analyse des spectres de masse

Les spectres de masse correspondent aux masses monoisotopiques des peptides. Les masses qui correspondent à des peptides bien connus pour appartenir à des polymères du gel, des tubes ou à des fragments provenant de l'autodigestion de la trypsine sont retirées des spectres avant l'analyse. Cependant, les peptides résultants de l'autodigestion de la trypsine ont pu être utilisés pour faire une calibration interne des spectres. Les spectres ont été analysés en utilisant les logiciels Mascot (<http://www.matrixscience.com>) et Profound (<http://prowl.rockefeller.edu/prowl-cgi/profound.exe>). La banque NCBI<sub>nr</sub> des protéines humaines a été choisie puisque les extraits proviennent de cellules HeLa. Cette banque de données contient les résultats de toutes les digestions tryptiques théoriquement possibles pour l'ensemble des protéines humaines. L'accord entre les pics observés dans le spectre de masse et les protéines proposées est évalué en utilisant différents algorithmes basés sur les données et les conditions expérimentales. Le score associé à chaque résultat évalue le risque qu'un accord entre les données expérimentales et les données théoriques soit aléatoire ou vrai. Un seuil communément admis est qu'un événement est significatif s'il apparaît au hasard avec une fréquence de moins de 5%. Ainsi, une protéine est correctement identifiée si elle correspond au meilleur score. De plus, pour chaque résultat de Profound et Mascot, plusieurs valeurs sont calculées et permettent d'évaluer les identifications proposées : l'intensité de couverture des intensités dans le spectre par les peptides qui correspondent à la protéine proposée, la couverture de la séquence de la protéine par les peptides, le rapport entre le

nombre de peptides en accord avec la protéine proposée sur le nombre total de peptides dans le spectre et l'écart de masse entre les masses des peptides de la banque de données et les valeurs de masses monoisotopiques des peptides du spectre. Cependant, la significativité dépend fortement de la qualité des données (nombre de peptides, précision des mesures) et donc les protéines qui ne sont pas évaluées comme les meilleurs résultats doivent quand même être examinées avec attention. Pour certaines identifications, les protéines proposées par les logiciels d'analyse des données de spectrométrie de masse (Profound et Mascot) peuvent avoir une masse moléculaire différente de celle évaluée à partir du gel SDS-PAGE. Des modifications post-traductionnelles, telles que la phosphorylation, la glycosylation et le clivage de protéines pour donner des produits catalytiquement actifs, peuvent expliquer des écarts de poids moléculaire. Bien que nous travaillions avec des extraits cellulaires à 4°C auxquels sont ajoutés des inhibiteurs de protéases, il ne peut pas être exclu qu'il puisse y avoir également des dégradations de protéines. Enfin, cette approche est limitée car elle est mal adaptée aux mélanges de protéines et elle n'est directement applicable qu'aux protéines présentes dans les banques de données.

C'est pourquoi, une autre approche complémentaire est largement employée. Elle consiste à séparer les différents fragments de digestion par chromatographie liquide et à fragmenter par MS/MS le plus de peptides possible. Les données de poids moléculaire des peptides de digestion ainsi que quelques éléments de séquence obtenus par MS/MS peuvent être insérés dans des banques de données de protéines, qui comparent les masses de peptides expérimentales aux masses théoriques déduites de chacune des séquences de protéines présentes dans ces banques de données après avoir effectué une digestion fictive *in silico* de chacune par la trypsine, et permettent ainsi l'identification de la protéine analysée. En outre, une analyse en mode MS/MS permet, dans le cas de mélanges complexes, de sélectionner spécifiquement une espèce parmi les autres pour obtenir son spectre de fragmentation. Même si les expériences de MS/MS sur tous les peptides donnent accès à de nombreux fragments de séquences, elles ne permettent ni de trouver les séquences totales de tous les peptides, ni de déterminer leurs positions exactes au sein de la séquence protéique. Dans ce cas, une ou plusieurs digestions supplémentaires peuvent être effectuées sur la protéine avec une enzyme différente afin d'améliorer la couverture de séquence obtenue.

### **3.6. Expériences concernant la recherche d'un rôle éventuel de hKIN17 dans l'épissage**

Les extraits nucléaires utilisés dans les expériences suivantes ont été préparés selon le protocole de Dignam et al. (1983) et les snRNPs et l'ARN pré-messager MINX radioactif ont été préparés selon les protocoles employés dans le laboratoire du Pr. R. Lührmann (MPI, Göttingen, Allemagne).

#### **3.6.1. Pull-down de réaction d'épissage**

Les billes de Ni-Nta agarose (QIAGEN) sont lavées avec le tampon d'épissage (65 mM KCl, 3.6 mM MgCl<sub>2</sub>, 20mM Cp, 2mM ATP dans eau) pour faire une suspension à 50% de billes. Les billes à 50% (200µl) sont mises à incuber avec 50µl de protéine His-KIN17 (à 10mg/ml) dans 1ml de tampon d'épissage, pendant une heure environ, sous agitation à 4°C.

Les billes sont lavées deux fois avec du tampon d'épissage et sont transférées dans un tube propre. Les réactifs de la réaction d'épissage sont ajoutés : 2ml d'extrait nucléaire à 10mg/ml, 15nM d'ARN radioactif (ARN pré-messager MINX), 25mM KCl, 3mM MgCl<sub>2</sub>, 20mM Cp, 2mM ATP dans un volume total de 5ml. L'ensemble est mis à incuber à 30°C pendant 45 minutes sous agitation pour que la réaction d'épissage ait lieu. Après cette incubation, le surnageant des billes est enlevé et gardé (flow-through). Les billes sont ensuite lavées avec 150mM NaCl, 10mM imidazole, 20mM Tris pH8 deux fois puis dans 150mM NaCl, 20mM imidazole, 20mM Tris pH8 une fois avant d'être éluées avec un tampon à 200mM imidazole. L'ARN est extrait au PCI (Phenol/Chloroforme/alcool Isoamylique 25:24:1), analysé sur un gel PAGE dénaturant 8.3M urée-10% polyacrylamide et visualisé par coloration à l'argent et /ou autoradiographie. Les protéines sont analysées sur un gel dénaturant SDS-PAGE 15% et visualisées par coloration au Bleu de Coomassie.

### 3.6.2. Pull-down des snRNPs

Les billes de Ni-Nta agarose (QIAGEN) et les billes de GS4B (Amersham Biosciences) sont lavées avec du PBS pour faire une suspension à 50% de billes. Les billes à 50% (200µl) sont mises à incuber avec 100 µg de protéines His-KIN17, GST, GST-K1 ou GST-K3 dans 1ml de PBS, pendant 30 minutes environ, sous agitation à 4°C. Les billes sont lavées trois fois avec du PBS et 100µg de snRNPs sont ajoutés. L'ensemble est mis à incuber à 4°C pendant deux heures environ sous agitation. Après cette incubation, les billes sont lavées six fois avec 500µl de PBS. Pour récupérer les protéines, soit le bleu de charge est déposé directement sur les billes soit les protéines sont extraites. Les protéines sont analysées sur un gel dénaturant SDS-PAGE 15% et visualisées par coloration au Bleu de Coomassie.

### 3.6.3. Recherche des conditions d'immunodéplétion

Les billes de *Protein G Sepharose* (Amersham Biosciences) sont lavées cinq fois avec 500µl de PBS pH8 froid. Deux anticorps anti-KIN17 différents sont utilisés (Biard et al. 2002). Les trois conditions suivantes sont préparées :

Mock	ΔK36	ΔK58
150 µl de PGS	150 µl de PGS	150 µl de PGS
500 µl de PBS pH8	500 µg d'anticorps K36 (500 µl)	500 µg d'anticorps K36 (250 µl)
12,5 µl de BSA (20 mg/ml)	12,5 µl de BSA (20 mg/ml)	250 µl de PBS pH8
2,5 µl d'ARNt (10 mg/ml)	2,5 µl d'ARNt (10 mg/ml)	12,5 µl de BSA (20 mg/ml)
		2,5 µl d'ARNt (10 mg/ml)

Les tubes sont mis à incuber sous agitation tournante, à 4°C sur la nuit (2 heures suffisent). Le lendemain, chaque tube est séparé en deux pour que les billes soient lavées soit en 200mM NaCl, soit en 500mM NaCl, avec du tampon IPP200 (20mM Hepes pH7.9, 200mM NaCl, 1.5mM MgCl<sub>2</sub>, 0.05% NP40) ou IPP500 (20mM Hepes pH7.9, 500mM NaCl, 1.5mM MgCl<sub>2</sub>, 0.05% NP40) respectivement. Chaque tube est encore séparé en deux puisque l'immunodéplétion se fait en deux étapes. Pour chaque étape, les conditions sont les suivantes :

à 200mM NaCl	à 500mM NaCl
37,5 µl de PGS (150/4) +/- anticorps 200 µl d'extrait nucléaire à 10 mg/ml 63,5 µl de tampon IPP200	37,5 µl de PGS (150/4) +/- anticorps 200 µl d'extrait nucléaire à 10 mg/ml 50 µl de tampon IPP500 13,5 µl de 5M NaCl

Les tubes sont laissés à incuber sous agitation tournante, à 4°C pendant une heure pour la première étape d'immunodépletion. Ensuite le surnageant est récupéré et transféré sur le reste de billes pour la seconde étape d'immunodépletion et les billes sont mises à incuber pendant une heure de la même manière que précédemment. Finalement, le surnageant est récupéré et correspond aux extraits nucléaires immuno-déplétés.

### 3.6.4. Test d'épissage avec les extraits nucléaires déplétés

Les extraits nucléaires (par aliquot de 25µl) sont dialysés deux fois une heure contre 500ml de tampon D (20mM Hepes pH8.0, 10% glycérol, 100mM KCl, 0.2mM EDTA, 1.5mM MgCl<sub>2</sub>, 0.5mM DTT) sous agitation douce à 4°C. Les extraits nucléaires après dialyse sont centrifugés jusqu'à 7000rpm et sont prêts pour la réaction d'épissage. Puisque nous avons une trentaine de conditions à tester, un mix est préparé (33X) : 9,9µl de 100mM MgCl<sub>2</sub>, 8,25µl de 100mM ATP, 16,5µl de 500mM CP, 20µl de MINX radioactif (30000cpm/µl, 0,15pmole/µl), 11,25µl de MINX froid (0,1pmole/µl) et 16,6µl d'eau pour arriver à un volume total de 82.5µl à diviser en 2,5µl par réaction d'épissage. Chaque réaction d'épissage est préparée de la manière suivante :

20	25	30	35	40	% extrait nucléaire
2.5	3.1	3.75	4.4	5	µl extrait nucléaire
0.5	0.5	0.4	0.3	0.25	µl KCl 1M (60mM final)
2.5µl du mix volume total de la réaction d'épissage : 12.5µl					(10nM MINX, 15000cpm, 3mM MgCl <sub>2</sub> finaux)

La réaction d'épissage est incubée pendant 45 minutes à 30°C et est stoppée en la mettant sur glace. Ensuite, elle est soumise à une digestion par la protéinase K : 1µl de protéinase K (à 10mg/ml) et 50µl de tampon associé à cette enzyme sont ajoutés dans chaque réaction d'épissage. L'ARN est ensuite extrait et déposé sur gel 8.3M urée-14% polyacrylamide. La radioactivité est détectée par une exposition du gel au Phosphorimager ou sur un film.

### 3.6.5. Test de complémentation

Les extraits nucléaires (par aliquot de 25µl) sont dialysés deux fois une heure contre 500ml de tampon D sous agitation douce à 4°C. La protéine recombinante HsKIN17, concentrée à 10mg/ml dans glycérol 40%, est aussi dialysée dans le tampon D avant d'être diluée pour obtenir une solution à 20pmole/µl. Nous préparons un mix avec les conditions classiques d'épissage, 40% d'extrait nucléaire et deux concentrations de protéines recombinantes (KIN17 1x : 2pmole/µl et KIN17 2x : 4pmole/µl) :

	mock	ΔK58	ΔK58 + KIN17 1x	ΔK58 + KIN17 2x
100mM MgCl <sub>2</sub>	1.65	1.65	1.54	1.43

100mM ATP	1.38	1.38	1.38	1.38
500mM CP	2.75	2.75	2.75	2.75
MINX radioactif	1.1	1.1	1.1	1.1
MINX froid	6.88	6.88	6.88	6.88
extrait nucléaire	27.5	27.5	27.5	27.5
1M KCl	1.38	1.38	0.55	
KIN17 recombinante			8.25	13.75
Eau	26.11	26.11	18.8	13.76

La concentration d'ARN dans la réaction d'épissage est de 14nM. Le mix est réparti en cinq fractions de 7 $\mu$ l incubés à 30°C pour suivre la réaction d'épissage pendant différentes durées (0, 15, 30, 45 et 60 minutes) et en cinq fractions de 5 $\mu$ l incubés à 30°C pendant différentes durées (0, 2, 5, 10 et 20 minutes) pour suivre la formation des complexes de l'épissage. Le gel des réactions d'épissage est réalisé de la même manière que précédemment. Pour le gel des complexes, les échantillons sont préparés de la manière suivante : les réactions d'épissage sont stoppées sur glace aux différents temps et 2.15 $\mu$ l de mix héparine (17 $\mu$ l d'héparine à 2mg/ml pour 34 $\mu$ l de sucrose 20%) sont ajoutés et bien mélangés à la réaction. Les tubes sont chauffés une minute à 30°C avant d'être déposés sur gel 2% LMP agarose. Le gel est migré à 120V pendant 3 heures 15 environ, puis fixé dans 10% acide acétique, 10% méthanol pendant 30 minutes, séché sur papier Whatmann et exposé au Phosphorimager ou à un film sensible à la radioactivité.



# **RESULTATS**



<b>1.</b>	<b>LA PROTEINE HKIN17 EST UNE PROTEINE MODULAIRE .....</b>	<b>66</b>
1.1.	PREDICTION DES REGIONS DESTRUCTUREES .....	66
1.2.	IDENTIFICATION DE DOMAINES CONNUS DANS LES REGIONS GLOBULAIRES .....	67
1.3.	VALIDATION EXPERIMENTALE DES PREDICTIONS DE REGIONS GLOBULAIRES .....	67
1.4.	DELIMITATION DES DOMAINES POUR L'ANALYSE STRUCTURALE .....	68
<b>2.</b>	<b>LA CARACTERISATION STRUCTURALE DU DOMAINE C-TERMINAL.....</b>	<b>69</b>
2.1.	CRISTALLISATION DU DOMAINE C-TERMINAL.....	69
2.1.1.	<i>Criblage des conditions de cristallisation et obtention de premiers cristaux.....</i>	<i>74</i>
2.1.2.	<i>Optimisation des conditions de cristallisation.....</i>	<i>75</i>
2.1.3.	<i>Recherche de dérivés lourds et préparation des dérivés tungstate.....</i>	<i>78</i>
2.2.	DETERMINATION DES PHASES .....	80
2.2.1.	<i>Enregistrement d'un jeu MAD sur un dérivé tungstate et estimation du signal anomal.....</i>	<i>80</i>
2.2.2.	<i>Essai de phasage avec le jeu de données du dérivé tungstate .....</i>	<i>82</i>
2.2.3.	<i>Résolution du problème de phasage avec autoSHARP.....</i>	<i>84</i>
2.3.	STRUCTURE A HAUTE RESOLUTION DU DOMAINE C-TERMINAL PAR RADIO-CRISTALLOGRAPHIE .....	85
2.4.	ETUDE PAR RMN DU DOMAINE C-TERMINAL .....	99
2.4.1.	<i>Attribution des pics de résonance du spectre <sup>1</sup>H, <sup>15</sup>N-HSQC .....</i>	<i>99</i>
2.4.2.	<i>Dynamique du domaine C-terminal .....</i>	<i>100</i>
2.4.3.	<i>Topologie du domaine C-terminal.....</i>	<i>101</i>
<b>3.</b>	<b>LA CARACTERISATION STRUCTURALE DE LA REGION N-TERMINALE .....</b>	<b>103</b>
3.1.	DETERMINATION DE LA STRUCTURE TRIDIMENSIONNELLE DU DOMAINE 51-160 PAR RMN .....	103
3.2.	ESSAI DE DETERMINATION DE LA STRUCTURE DE LA REGION 1-160 PAR RMN .....	117
3.3.	ETUDE STRUCTURALE DE LA REGION 1-160 PAR SAXS .....	117
3.3.1.	<i>Extraction des paramètres structuraux .....</i>	<i>118</i>
3.3.1.1.	<i>Le rayon de giration à partir de la courbe de Guinier.....</i>	<i>118</i>
3.3.1.2.	<i>La région 1-160 est sous forme monomérique en solution.....</i>	<i>118</i>
3.3.1.3.	<i>La fonction de distribution des distances interatomiques P(r).....</i>	<i>119</i>
3.3.2.	<i>Hypothèses sur la région 1-160 compatibles avec les données de SAXS .....</i>	<i>120</i>
3.3.2.1.	<i>Les données SAXS sont compatibles avec la présence d'un domaine globulaire dans la région N-terminale en amont du domaine winged helix .....</i>	<i>120</i>
3.3.2.2.	<i>Les données SAXS sont compatibles avec la présence d'un doigt de zinc dans la région N-terminale en amont du domaine winged helix.....</i>	<i>121</i>
3.3.2.3.	<i>L'analyse de la flexibilité de la région N-terminale en solution.....</i>	<i>122</i>
3.3.3.	<i>Construction ab initio de modèles de la région 1-160.....</i>	<i>123</i>
3.3.3.1.	<i>Les modèles obtenus avec le programme DAMMIN .....</i>	<i>123</i>
3.3.3.2.	<i>Les modèles obtenus avec le programme GASBOR.....</i>	<i>124</i>
3.4.	DONNEES SAXS ET RMN COMPATIBLES POUR PROPOSER UN MODELE DE LA REGION 1-160 .....	126
3.4.1.	<i>Effet de la région N-terminale en amont du winged helix sur ce domaine.....</i>	<i>126</i>
3.4.2.	<i>Calcul de modèles de la région 1-160.....</i>	<i>127</i>
<b>4.</b>	<b>VERS LA STRUCTURE DE LA PROTEINE ENTIERE.....</b>	<b>130</b>
4.1.	ABSENCE D'INTERACTION ENTRE LES REGIONS N- ET C-TERMINALES .....	130
4.2.	PREPARATION DE LA PROTEINE ENTIERE POUR LES ETUDES STRUCTURALES .....	131
4.3.	ENREGISTREMENT DE DONNEES DE SAXS SUR LA PROTEINE ENTIERE.....	133



## 1. LA PROTEINE hKIN17 EST UNE PROTEINE MODULAIRE

La protéine hKIN17 est, comme nous allons le présenter dans ce paragraphe, une protéine modulaire. Dans l'objectif de sa caractérisation structurale, par RMN et/ou radiocristallographie, nous avons adopté une approche d'étude domaine par domaine. Une étape déterminante de ce travail est alors la prédiction exacte des limites des domaines solubles susceptibles de se replier de manière autonome.

La délimitation des domaines a été réalisée par le Dr S.Zinn-Justin avant que j'arrive au laboratoire. J'ai effectué le reste de l'étude.

### 1.1. Prédiction des régions déstructurées

Tout d'abord, les régions globulaires et les régions déstructurées ont été identifiées grâce au programme DISOPRED2 (Ward et al. 2004). Les régions désordonnées sont flexibles, dynamiques et peuvent être partiellement ou complètement étendues en solution. Elles sont souvent caractérisées par une faible complexité de séquence et une composition enrichie en résidus aromatiques et hydrophiles, alors que les régions correspondant aux domaines globulaires sont riches en résidus hydrophobes qui constituent leur cœur structural (33% de résidus hydrophobes en moyenne dans un domaine globulaire).

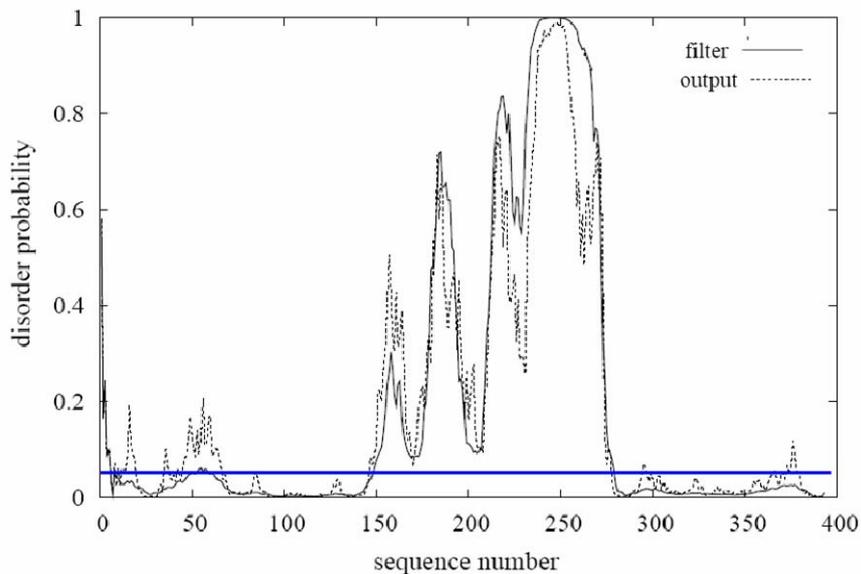


Figure 1-1: Prédiction des régions déstructurées au sein de la protéine humaine KIN17 par le programme DISOPRED2. Le graphe montre la probabilité d'être désordonné en fonction de la position dans la séquence. La ligne bleue horizontale correspond au seuil structuré/déstructuré choisi par défaut par le programme. La courbe en trait plein (*filter*) correspond au résultat de la prédiction de DISOPRED2 et la courbe en pointillés (*output*) correspond au résultat de DISOPREDsvm.

D'après les résultats de DISOPRED2 (Figure 1-1), il apparaît que la protéine humaine KIN17 est constituée de deux régions globulaires, l'une en N-terminal (résidus 6 à 150) et l'autre en C-terminal (résidus 278 à 393). Par contre, la région centrale de la protéine (résidus 151 à 277) est prédite comme une région déstructurée. Une analyse complémentaire basée sur

l'approche HCA (Hydrophobic Cluster Analysis, (Gaboriaud et al. 1987)) de la protéine hKIN17, qui met en valeur les paquets de résidus hydrophobes correspondant aux régions structurées, donne les mêmes limites de domaines.

## 1.2. Identification de domaines connus dans les régions globulaires

Nous avons cherché à savoir si les régions globulaires de la protéine hKIN17 correspondaient à des domaines fonctionnels. La recherche de domaines a été réalisée en utilisant le moteur de recherche de la SMART sur les bases de données SMART (Schultz et al. 1998) et PFAM (Sonnhammer et al. 1997; Bateman et al. 2004). Ces bases de données de domaines répertorient sous forme d'alignements multiples les domaines structuraux connus et de nombreux motifs appartenant probablement à des domaines structuraux.

Ainsi, en N-terminal un doigt de zinc de type Zn-C<sub>2</sub>H<sub>2</sub> a été identifié (résidus 26 à 50). Les doigts de zinc sont des domaines structuraux en général impliqués dans la liaison aux acides nucléiques. De plus, un motif KOW (Kyrpides et al. 1996) a été identifié dans la région C-terminale (résidus 338 à 364). Ce motif est décrit comme appartenant à des protéines se liant à l'ARN.



Figure 1-2: Architecture de la protéine KIN17 humaine, avec les régions potentiellement globulaires sous forme de rectangles bleus et les motifs identifiés par SMART sous forme de rectangles jaunes (doigt de zinc, séquence homologue à RecA, signal de localisation nucléaire, motif KOW).

Sur la

Figure 1-2, la région nommée RecA correspond aux résidus 163 à 201 de la protéine hKIN17 identique à 47% à une région de la protéine RecA d'*E.coli* et elle est le site de la réactivité croisée avec les anticorps anti-RecA (Kannouche et al. 2000). La région nommée NLS correspond à un signal de localisation nucléaire de la protéine détectée par analyse de séquence entre les résidus 240 et 257 (Kannouche et al. 2000).

## 1.3. Validation expérimentale des prédictions de régions globulaires

La protéine hKIN17 entière est purifiée selon le protocole décrit par (Miccoli et al. 2002). Nous avons observé que la protéine est très sensible aux dégradations et que certains échantillons de protéine apparemment pure ont subi de la protéolyse en fin de préparation. Des fragments de protéine hKIN17 issus de la protéolyse ont été séparés sur gel SDS-PAGE et identifiés par séquençage N-terminal.



Figure 1-3: Echantillon de protéine hKIN17 entière protéolysée. A gauche, les fragments issus de la protéolyse sont séparés sur gel SDS-PAGE et les protéines pointées par des flèches rouges ont été séquencées. A droite, la position du N-terminal des zones de protéolyse est notée sur la séquence de la protéine KIN17 humaine par des flèches rouges. La séquence des zones prédites comme globulaires est colorée en bleu.

La plupart des fragments proviennent d'une protéolyse dans la zone prédite comme étant déstructurée. Cela est cohérent avec la grande flexibilité de cette région qui la rend plus sensible à la protéolyse. D'autres fragments proviennent d'une protéolyse dans la région N-terminale, soit au niveau d'une région probablement peu structurée en amont du doigt de zinc, soit dans une zone charnière entre le doigt de zinc et un autre domaine structural.

### 1.4. Délimitation des domaines pour l'analyse structurale

A partir de l'étude précédente, nous avons proposé de résoudre la structure tridimensionnelle des trois régions susceptibles de se replier en domaines solubles de manière autonome. Ces régions sont présentées sur la Figure 1-4.

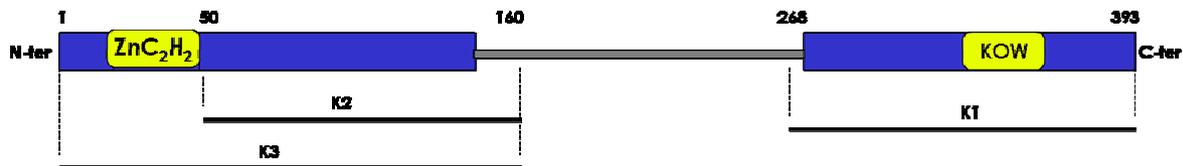


Figure 1-4 : Architecture de la protéine KIN17 humaine, avec les régions potentiellement globulaires sous forme de rectangles bleus et les motifs identifiés par SMART sous forme de rectangles jaunes. Les fragments K1, K2 et K3 choisis pour l'étude structurale de la protéine sont représentés schématiquement par des traits noirs.

Deux de ces régions se trouvent en N-terminal, l'une avec et l'autre sans le doigt de zinc (K3 et K2) puisqu'il peut être difficile de travailler avec les doigts de zinc dans des conditions classiques de préparation des protéines. En particulier, les sites de liaison au zinc, impliquant des résidus cystéines, sont cruciaux pour maintenir le repliement du domaine et peuvent être sujets à des phénomènes d'oxydation *in vitro*. La troisième région d'étude (K1) se trouve en C-terminal de la protéine.

*En conclusion*, pendant la thèse, les domaines K1 (résidus 268 à 393), K2 (résidus 50 à 160) et K3 (résidus 1 à 160) ont été étudiés séparément et en parallèle afin de déterminer leur structure tridimensionnelle et leur rôle fonctionnel dans la protéine, en particulier en recherchant les partenaires biologiques (acides nucléiques ou protéines) avec lesquels ils interagissent.

---

## 2. LA CARACTERISATION STRUCTURALE DU DOMAINE C-TERMINAL

---

Le domaine C-terminal (domaine K1) de la protéine hKIN17 n'existe que chez les eucaryotes supérieurs chez lesquels il apparaît très conservé. D'un point de vue structural, il a été étudié en parallèle par RMN et cristallographie aux rayons X. L'obtention de sa structure à haute résolution par radio-cristallographie a fait l'objet de deux publications, l'une décrivant la cristallogénèse et le phasage et l'autre présentant la structure.

J'ai effectué les études de cristallogénèse du domaine K1, sous la direction du Dr. E. Stura au laboratoire.

J'ai résolu la structure cristallographique du domaine K1 en collaboration avec le Pr. M. Schiltz du LCR/EPFL (Lausanne, Suisse).

J'ai effectué la caractérisation par RMN du domaine K1, encadrée par le Dr. J. Couprie au laboratoire.

### 2.1. Cristallisation du domaine C-terminal

L'article 1 (*Crystallization and halide phasing of the C-terminal domain of human KIN17*, le Maire A. et al., 2006) décrit succinctement les étapes qui ont permis d'obtenir des cristaux du domaine C-terminal diffractant à haute résolution et des phases de haute qualité. A la suite, de cet article, nous décrivons plus en détails la cristallogénèse et le phasage.

Résumé article 1 : *Crystallization and halide phasing of the C-terminal domain of human KIN17* (2006), A le Maire, M Schiltz, S Braud, M Gondry, JB Charbonnier, S Zinn-Justin and E Stura, Acta Crystallographica Section F, F62.

Dans l'article, sont décrits la cristallisation et le phasage initial du domaine C-terminal de la protéine humaine KIN17. Des cristaux diffractant à 1.4 Å de résolution ont été obtenus en condition 10% éthylène glycol, 27% PEG 6000, 500mM LiCl dans 100mM acétate de sodium pH6.3. Le groupe d'espace est P2<sub>1</sub>2<sub>1</sub>2<sub>1</sub> avec des paramètres de maille égaux à a=45.75 Å, b=46.31 Å, c=60.80 Å et il y a une molécule par unité asymétrique. Ce domaine ayant un pI basique, les dérivés d'atomes lourds ont été obtenus par trempage des cristaux avec des ions chargés négativement tels que le tungstate et l'iodure. Le remplacement de LiCl par KI dans la solution cryoprotectante a permis la détermination des phases à partir d'un dérivé iodé conduisant ainsi à l'obtention d'une carte de densité électronique interprétable.

Albane le Maire,<sup>a\*</sup> Marc Schiltz,<sup>b</sup>  
Sandrine Braud,<sup>c</sup> Muriel  
Gondry,<sup>c</sup> Jean-Baptiste  
Charbonnier,<sup>a</sup> Sophie Zinn-  
Justin<sup>a</sup> and Enrico Stura<sup>a</sup>

<sup>a</sup>Laboratoire de Structure des Protéines,  
Département d'Ingénierie et d'Etude des  
Protéines, Commissariat à l'Energie Atomique,  
91191 Gif-sur-Yvette, France, <sup>b</sup>Laboratoire de  
Cristallographie, Ecole Polytechnique Fédérale  
de Lausanne, CH-1015 Lausanne, Switzerland,  
and <sup>c</sup>Laboratoire de Marquage des Protéines,  
Département d'Ingénierie et d'Etude des  
Protéines, Commissariat à l'Energie Atomique,  
91191 Gif-sur-Yvette, France

Correspondence e-mail: albane.lemaire@cea.fr

Received 6 December 2005

Accepted 2 February 2006

## Crystallization and halide phasing of the C-terminal domain of human KIN17

Here, the crystallization and initial phasing of the C-terminal domain of human KIN17, a 45 kDa protein mainly expressed in response to ionizing radiation and overexpressed in certain tumour cell lines, are reported. Crystals diffracting to 1.4 Å resolution were obtained from 10% ethylene glycol, 27% PEG 6000, 500 mM LiCl and 100 mM sodium acetate pH 6.3 in space group  $P2_12_12_1$ , with unit-cell parameters  $a = 45.75$ ,  $b = 46.31$ ,  $c = 60.80$  Å and one molecule in the asymmetric unit. Since this domain has a basic pI, heavy-atom derivatives were obtained by soaking the crystals with negatively charged ions such as tungstate and iodine. The replacement of LiCl by KI in the cryosolution allowed the determination of phases from iodide ions to give an interpretable electron-density map.

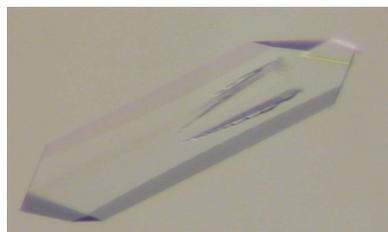
### 1. Introduction

KIN17 is a 45 kDa protein found principally in the nucleus, conserved from yeast to humans and ubiquitously expressed in mammals (Kannouche *et al.*, 2000). It forms intranuclear foci in proliferating cells and is up-regulated in response to UV and  $\gamma$  irradiation (Biard *et al.*, 2002; Blattner *et al.*, 2000; Kannouche *et al.*, 1998; Masson *et al.*, 2003). Human KIN17 is a modular protein comprising four motifs: a zinc finger (amino acids 27–50), a nuclear localization signal (amino acids 239–256), a core domain homologous to bacterial RecA protein (amino acids 161–201) and a KOW motif (amino acids 335–373) (Kyrpides *et al.*, 1996; Ponting, 2002) that is also found in ribosomal protein L24 and in several bacterial transcription factors. KIN17 binds DNA in human cells (Biard *et al.*, 2002) and displays enhanced expression levels in proliferating cultured cells, suggesting a role in nuclear metabolism. KIN17 also directly binds RNA *in vitro* and *in vivo* (Pinon-Lataillade *et al.*, 2004). Its C-terminal domain (residues 268–394) is present from insects to humans and in plants. This domain was crystallized and its three-dimensional structure solved in order to characterize its evolutionarily conserved surfaces and to identify its nucleic acid-binding sites. In particular, we aim to understand the functional role of the KOW motif and test whether this sequence is involved in RNA binding as suggested by Kyrpides *et al.* (1996). Here, we describe and discuss the expression, purification and crystallization and the phasing procedure used to determine the structure of the C-terminal domain of human KIN17 from crystals diffracting to high resolution (1.4 Å).

### 2. Methods, results and discussion

#### 2.1. Crystallization and data collection

**2.1.1. Preparation of the protein.** The gene encoding the region 268–393 of human KIN17 protein was amplified and cloned using the Gateway system (Invitrogen). A soluble expression screening using different strains, seven fusion partners and various expression conditions was performed (Braud *et al.*, 2005). The best results were obtained when transforming *Escherichia coli* BL21(DE3) strain with plasmid pEXP (Invitrogen) encoding the kin17 fragment in fusion with glutathione-S-transferase (GST) and a TEV (tobacco etch virus) protease cleavage site between KIN17 and GST. Bacterial growth was performed at 293 K to an OD<sub>600nm</sub> of 0.8. Expression was induced with 1 mM isopropyl  $\beta$ -D-thiogalactopyranoside (IPTG). The cell



© 2006 International Union of Crystallography  
All rights reserved

pellet was resuspended in 150 mM NaCl, 100 mM Tris pH 8 containing 5% glycerol. Subsequent steps were performed in a cold room. The protein was first purified using glutathione-agarose beads (Sigma) to separate the fusion protein from the other bacterial proteins and again after cleavage with TEV protease to separate the KIN17 fragment from GST, leaving a single non-wild-type glycine at the N-terminus. The purified protein of 14.2 kDa was then concentrated (using Amicon Centricon 3 kDa filters) in PBS (2.7 mM KCl, 1.5 mM  $\text{KH}_2\text{PO}_4$ , 137 mM NaCl and 8.1 mM  $\text{Na}_2\text{HPO}_4$ ) pH 7.0 to a final concentration of 20 mg ml<sup>-1</sup> as estimated from an absorption measurement at 280 nm. Samples were flash-frozen in liquid nitrogen and stored at 193 K. The purity of the protein, as estimated by SDS-PAGE, was greater than 95%.

**2.1.2. Preparation of native crystals.** The sitting-drop vapour-diffusion technique was used in robotic screening to obtain suitable crystallization conditions. 288 conditions were tested using a Tecan robot at 290 K including the 48 unique reagents from Natrix (Hampton Research) and the 240 conditions from the JBScreen (Jena Bioscience). Drops were observed after two weeks; crystals of various shapes appeared in about 30 conditions. After analysis of the conditions yielding crystals, an improvement protocol was designed using the principles of reverse screening (Stura *et al.*, 1994). The first step was to adjust the protein concentration to avoid excessive protein precipitation before the appearance of crystals. Subsequently, the precipitant concentration was adjusted to favour defect-free growth. Lastly, additives were introduced both to limit nucleation and promote slow growth. The optimized crystallization protocol consisted of mixing 1.5 µl of protein diluted to 5 mg ml<sup>-1</sup> in PBS with 1.5 µl reservoir condition consisting of 10% ethylene glycol, 27% PEG 6000, 500 mM LiCl and 100 mM sodium acetate pH 6.3. The choice of ethylene glycol as the additive was based on its ability to slow down crystal growth and on its properties as a cryoprotective solvent. Large crystals (about 400 × 100 × 50 µm) suitable for X-ray diffraction analysis (Fig. 1) were obtained after 2–3 days and were stable for several months.

**2.1.3. Data collection.** After transferring to a cryoprotectant solution [10% ethylene glycol, 15% (v/v) glycerol, 27% PEG 6000, 500 mM LiCl and 100 mM sodium acetate pH 6.3], crystals were flash-frozen in either nitrogen gas or in liquid ethane. All data were collected under cryogenic conditions on beamline BM-30A at the European Synchrotron ESRF (Roth *et al.*, 2002) using a MAR CCD detector.

The native data set was collected to 1.8 Å resolution over a total angular range of 200° using angular increments of 1° per frame and a wavelength of 1.130 Å (Table 1). The programs *DENZO* and *SCALEPACK* from the *HKL* suite (Otwinowski & Minor, 1997) were used for data processing. The native crystals belong to space group  $P2_12_12_1$ , with average unit-cell parameters  $a = 45.75$ ,  $b = 46.31$ ,  $c = 60.80$  Å, and contain one molecule per asymmetric unit, giving an approximate solvent content of 45% (Matthews coefficient  $V_M = 2.3$  Å<sup>3</sup> Da<sup>-1</sup>; Matthews, 1968).

## 2.2. Heavy-atom derivatives and initial phasing

To determine structure-factor phases, various heavy-atom derivatives were tested: tungstate ions were introduced to replace the eventually bound phosphate ions and iodide was tested as a replacement for chloride counterions. Dauter *et al.* (2000) proposed a method of phasing with halide anions which can substitute bound water molecules without any preference for specific coordination geometry and environment. Since the chosen protein region contains

**Table 1**  
Reflection data statistics.

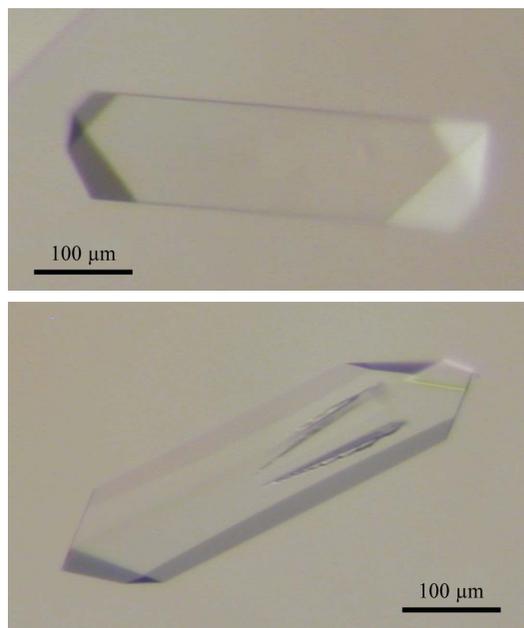
Values in parentheses are for the highest resolution shell.

Data set	Native	Derivate (iodine)
Wavelength (Å)	1.1302	0.920210
Space group	$P2_12_12_1$	$P2_12_12_1$
Unit-cell parameters (Å)	$a = 45.89$ , $b = 46.36$ , $c = 60.97$	$a = 45.98$ , $b = 46.25$ , $c = 60.95$
Resolution range (Å)	99.00–1.81 (1.85–1.81)	46.21–1.45 (1.54–1.45)
No. of observations	154906	156216 (15401)
No. of unique reflections	21818	43097 (6226)
Completeness (%)	94.8 (84.8)	97.0 (86.4)
$\langle I \rangle / \langle \sigma(I) \rangle$	23.57 (5.98)	15.68 (4.72)
$R_{\text{merge}}$ (%)	4.5 (12.2)	6.7 (29.5)

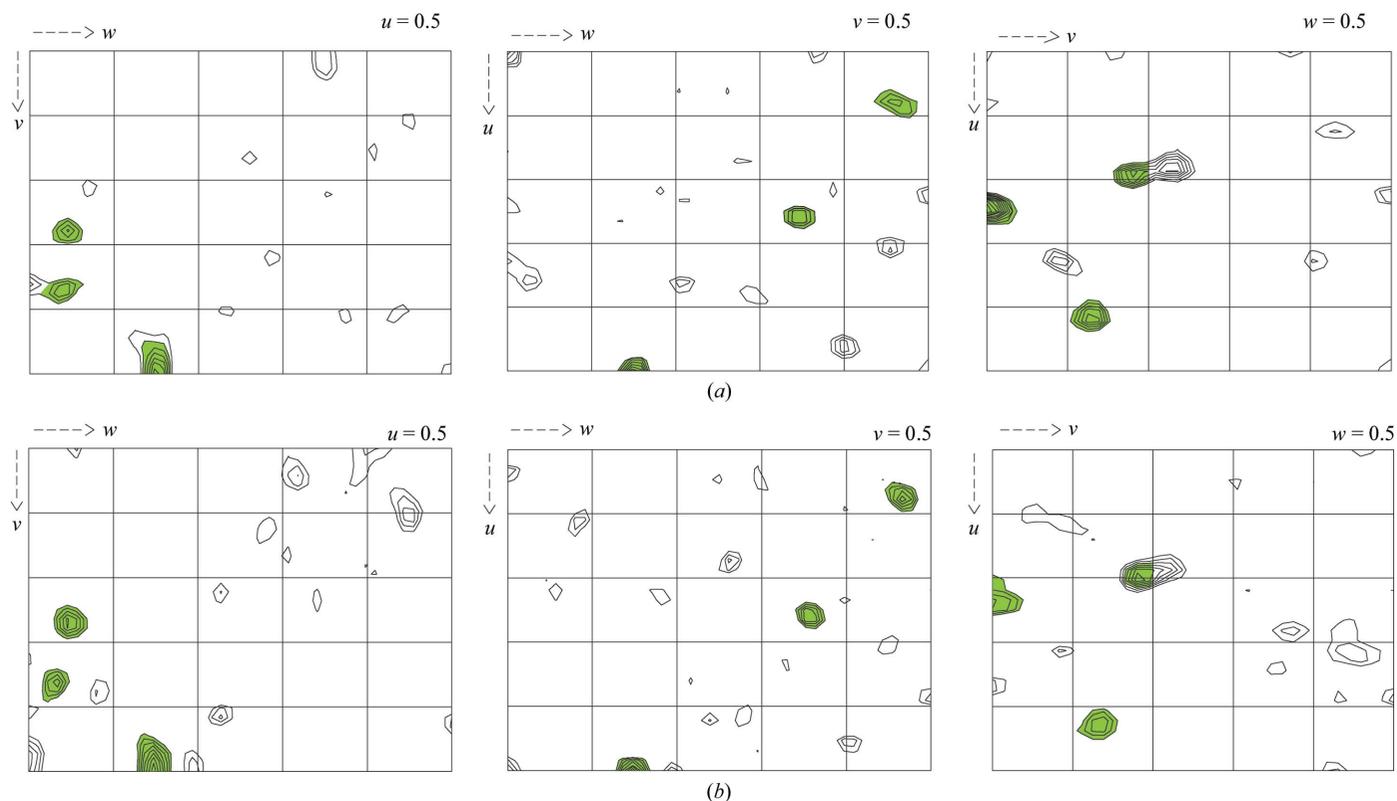
six arginines and 18 lysines, its potential ability to bind halide anions suggested testing this easy way to introduce anomalous scatterers.

Tungstate derivatives were prepared either by soaking {0.5–1 mM tungstate compound [ $\text{WO}_4^{2-}$ ,  $\text{WS}_4^{2-}$  or  $\text{W}(\text{CO}_3)_3^{3-}$ ] in the cryoprotectant solution} or cocrystallization [0.5–2 mM tungstate compound ( $\text{WO}_4^{2-}$  or  $\text{WS}_4^{2-}$ ) in 20% ethylene glycol, 20% PEG 6000 and 100 mM sodium acetate pH 6.3]. Iodine derivatives were prepared by exchanging chloride for iodide during the brief transfer (average handling time 45 s) to a cryosolution composed of 10% ethylene glycol, 15% glycerol, 27% PEG 6000, 100 mM sodium acetate pH 6.3 and 1 M KI. During the exchange phase, no visible cracking of the crystals was observed.

Diffraction data to 1.45 Å were collected using X-rays with a wavelength of 0.92 Å and with an exposure time of 8 s per image (Table 1). The diffraction images were processed with *XDS* (Kabsch, 1993). This data set was isomorphous with the native and scaling of the two data sets was performed using *SCALEIT* (Collaborative Computational Project, Number 4, 1994). The presence of halide sites was assessed by calculating isomorphous difference Patterson maps



**Figure 1**  
Examples of native crystals of the C-terminal domain of human KIN17 obtained in 27% PEG 6000, 10% ethylene glycol, 500 mM LiCl and 100 mM sodium acetate pH 6.3 after 2 days at 290 K. Initial crystals of this form were obtained from the conditions JBS1 D4, JBS2 D1, JBS3 D2, JBS4 B4, JBS4 B6 and JBS5 C2 (Jena Bioscience). Other crystal forms obtained under other conditions were judged to be less suitable for improvement.



**Figure 2** Patterson maps show the presence of at least three iodine sites, marked in green. (a) The three Harker sections of the isomorphous difference Patterson map ( $F_{PHI} - F_P$ ). (b) The three Harker sections of the anomalous difference Patterson map ( $F^+ - F^-$ ). These maps were calculated with data between 15 and 2 Å resolution. Contour levels are at intervals of 0.5 r.m.s. density, starting at 1.5 r.m.s. above the mean density.

and anomalous difference Patterson maps using data between 15 and 2 Å. The Harker sections (Fig. 2) revealed at least three strong Harker peaks indicating anomalous sites. Heavy-atom detection using data to 2.0 Å resolution with *RSPS* (Collaborative Computational Project, Number 4, 1994) initially determined three stronger heavy-atom sites. Heavy-atom positions were then used as a seeding solution in *SHARP* (Bricogne *et al.*, 2003). After refinement in *SHARP* and interpretation of residual maps, one new site was identified and added. Using these four sites, *SHARP* was then run to calculate SIRAS phases with all data to 1.45 Å. Phase refinement and density modification were carried out with the program *SOLOMON* (Abrahams & Leslie, 1996) within the *SHARP* interface and yielded a high-quality electron-density map. The refined structure of the protein will be presented elsewhere.

### 3. Conclusions

Like hen egg-white lysozyme (HEWL), a protein with a basic pI that is frequently used as a model in crystallization studies, the C-terminal domain of KIN17 seems to have an affinity for halide ions which can occupy ordered solvent sites around the molecule in the crystal (Dauter & Dauter, 1999; Lim *et al.*, 1998; Steinrauf, 1998). Iodide anions preferentially bind to lysine and arginine (Dauter *et al.*, 2000). It appeared that these anions bind tightly to our fragment of KIN17, as indicated by their high anomalous difference peaks, good occupancies and low *B* factors. We confirm that this method of introducing anomalously scattering halide ions into protein crystals is a fast and simple way to solve structures. The method is particularly well suited to the present case since the exposed surface of the protein is relatively small, giving rise to only a small number of halide sites. The

detection of iodine sites was thus straightforward, even with conventional Patterson maps (Dauter *et al.*, 2000; Usòn *et al.*, 2003). Of all the heavy-atom compounds tested, KI was the only one to produce derivatives that yielded good phasing information. Surprisingly, the tungstate derivatives did not lead to interpretable phases even when tungstate sites could be clearly detected and refined. The refinement of the structure is currently in progress, as are crystallization trials on the other domains and on the entire protein. Analysis of the crystal structure of the C-terminal domain of KIN17 will allow an investigation at an atomic resolution of how KIN17 binds nucleic acids.

We thank Dr J. Angulo-Mora for providing human kin17 cDNA and Dr H. van Tilbeurgh for the use of the Tecan robot. We thank Dr J.-L. Ferrer (beamline FIP-BM30A) for help with synchrotron data collection at the ESRF (Grenoble).

### References

- Abrahams, J. P. & Leslie, A. G. W. (1996). *Acta Cryst.* **D52**, 30–42.
- Biard, D. S., Miccoli, L., Despras, E., Frobert, Y., Creminon, C. & Angulo, J. F. (2002). *J. Biol. Chem.* **277**, 19156–19165.
- Blattner, C., Kannouche, P., Litfin, M., Bender, K., Rahmsdorf, H. J., Angulo, J. F. & Herrlich, P. (2000). *Mol. Cell. Biol.* **20**, 3616–3625.
- Braud, S., Moutiez, S., Belin, P., Abello, N., Drevet, P., Zinn-Justin, S., Courçon, M., Masson, C., Dassa, J., Charbonnier, J., Boulain, J., Ménez, A., Genet, R. & Gondry, M. (2005). *J. Proteome Res.* **4**, 2137–2147.
- Bricogne, G., Vornrhein, C., Flensburg, C., Schiltz, M. & Paciorek, W. (2003). *Acta Cryst.* **D59**, 2023–2030.
- Collaborative Computational Project, Number 4 (1994). **D50**, 760–763.
- Dauter, Z. & Dauter, M. (1999). *J. Mol. Biol.* **289**, 93–101.
- Dauter, Z., Dauter, M. & Rajashankar, K. R. (2000). *Acta Cryst.* **D56**, 232–237.
- Kabsch, W. (1993). *J. Appl. Cryst.* **26**, 795–800.

- Kannouche, P., Mauffrey, P., Pinon-Lataillade, G., Mattei, M. G., Sarasin, A., Daya-Grosjean, L. & Angulo, J. F. (2000). *Carcinogenesis*, **21**, 1701–1710.
- Kannouche, P., Pinon-Lataillade, G., Tissier, A., Chevalier-Lagente, O., Sarasin, A., Mezzina, M. & Angulo, J. F. (1998). *Carcinogenesis*, **19**, 781–789.
- Kyrpides, N. C., Woese, C. R. & Ouzounis, C. A. (1996). *Trends Biochem. Sci.* **21**, 425–6.
- Lim, K., Nadarajah, A., Forsythe, E. L. & Pusey, M. L. (1998). *Acta Cryst.* **D54**, 899–904.
- Masson, C., Menea, F., Pinon-Lataillade, G., Frobert, Y., Chevillard, S., Radicella, J. P., Sarasin, A. & Angulo, J. F. (2003). *Proc. Natl Acad. Sci. USA*, **100**, 616–621.
- Matthews, B. W. (1968). *J. Mol. Biol.* **33**, 491–497.
- Otwinowski, Z. & Minor, W. (1997). *Methods Enzymol.* **276**, 307–326.
- Pinon-Lataillade, G., Masson, C., Bernardino-Sgherri, J., Henriot, V., Mauffrey, P., Frobert, Y., Araneda, S. & Angulo, J. F. (2004). *J. Cell Sci.* **117**, 3691–3702.
- Ponting, C. P. (2002). *Nucleic Acids Res.* **30**, 3643–3652.
- Roth, M., Carpentier, P., Kaikati, O., Joly, J., Charrault, P., Pirocchi, M., Kahn, R., Fanchon, E., Jacquamet, L., Borel, F., Bertoni, A., Israel-Gouy, P. & Ferrer, J.-L. (2002). *Acta Cryst.* **D58**, 805–814.
- Steinrauf, L. K. (1998). *Acta Cryst.* **D54**, 767–780.
- Stura, E. A., Satterthwait, A. C., Calvo, J. C., Kaslow, D. C. & Wilson, I. A. (1994). *Acta Cryst.* **D50**, 448–455.
- Usón, I., Schmidt, B., von Bulow, R., Grimme, S., von Figura, K., Dauter, M., Rajashankar, K. R., Dauter, Z. & Sheldrick, G. M. (2003). *Acta Cryst.* **D59**, 57–66.

*Compléments à l'article 1 concernant la cristallo-genèse.*

L'étape limitante d'une étude cristallographique est souvent l'obtention de cristaux de la macromolécule. Un cristal est un assemblage de molécules ordonnées régulièrement dans l'espace grâce à des interactions non covalentes. La cristallo-genèse implique un rapprochement des macromolécules en augmentant les interactions entre les molécules de protéines et en excluant une partie du solvant. La connaissance de la physico-chimie des macromolécules n'est pas suffisante pour établir une méthode rationnelle de cristallo-genèse qui reste pour une grande part très empirique. Une fois les cristaux obtenus, ceux-ci doivent être améliorés afin d'atteindre la plus haute résolution possible pour résoudre la structure avec la plus grande précision.

Nous allons décrire ici les différentes étapes de cristallo-genèse qui ont permis d'obtenir des cristaux du domaine C-terminal de hKIN17 diffractant à haute résolution. Nous allons tout d'abord présenter la recherche des premières conditions de cristallisation qui ont permis d'obtenir des premiers cristaux avant de décrire les différentes étapes d'optimisation des conditions en vue d'améliorer la taille et la qualité des cristaux. L'article décrit les conditions de purification de la protéine K1. L'échantillon de protéine K1 utilisé n'était pur qu'à 95% environ. Cependant, cette pureté limitée n'a pas gêné la croissance des cristaux et n'a pas empêché d'obtenir de beaux monocristaux. Nous n'avons pas vérifié la reproductibilité puisque tous les cristaux ont été obtenus avec le même lot de protéine. La cristallisation de K1 était favorisée par le fait que la protéine est très stable dans le temps.

### **2.1.1. Criblage des conditions de cristallisation et obtention de premiers cristaux**

La protéine purifiée (260  $\mu$ l) a été utilisée à une concentration de 20 mg/ml pour rechercher des conditions de cristallisation en robot, comme cela est indiqué dans l'article. Les boîtes de cristallisation ont été observées après 15 jours à 17°C. Dans 26 conditions, des cristaux, des microcristaux, des précipités cristallins, des sphérulites, des aiguilles ou des oursins ont été observés (Figure 2-1). Un récapitulatif de ces conditions de cristallisation est donné en annexe 2.

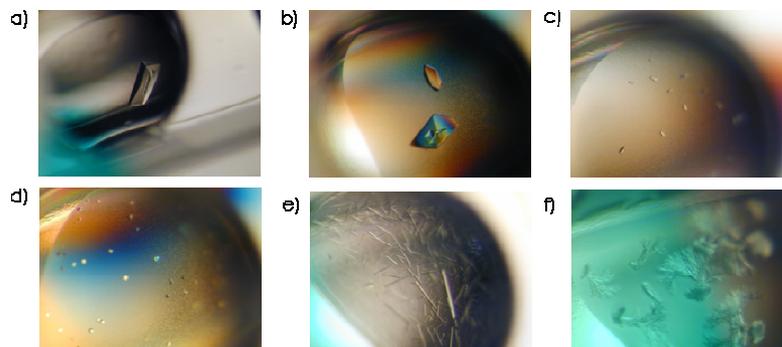


Figure 2-1: Images de quelques gouttes de cristallisation dans lesquelles un cristal (a), des cristaux en présence de précipité (b), des microcristaux (c), des sphérulites (d) des aiguilles (e) ou des oursins (f) ont été observés. Les conditions associées à chacune sont référencées dans le tableau en annexe 2 par la même lettre que celle indiquée à côté de chaque image.

Les cristaux observés dans plusieurs conditions ont des formes variées suggérant différentes formes cristallines mais il faudrait le vérifier en déterminant les groupes d'espace de chacun de ces cristaux. En effet, changer les conditions de solvant peut affecter la nature et le nombre des contacts cristallins et conduire à différentes formes cristallines. Les contacts cristallins impliquent des reconnaissances moléculaires de régions complémentaires de la protéine par liaisons hydrogène, contacts de van der Waals, ponts salins. La stabilité des cristaux de protéine et aussi leur pouvoir de diffraction reposent sur la nature et la force de ces contacts cristallins dans le réseau de cristal.

Ces conditions ont donné des pistes de cristallisation que nous avons ensuite cherché à reproduire et à optimiser afin de déterminer des conditions permettant d'obtenir des cristaux de bonne qualité et diffractant à haute résolution.

### 2.1.2. Optimisation des conditions de cristallisation

Dans ce qui est présenté par la suite, nous avons utilisé la méthode de *reverse screening* développée par E. Stura (Stura et al. 1994). Cela revient à explorer le diagramme de solubilité de la protéine, en étudiant l'effet de la concentration en agent précipitant, de la concentration en protéine, d'additifs (ions), du pH et de la température sur la solubilité de la protéine. Les conditions physico-chimiques nécessaires pour atteindre la nucléation ne sont pas les mêmes que celles nécessaires pour la croissance des cristaux bien que les deux événements dépendent du degré de supersaturation de la protéine. Alors, en connaissant les paramètres qui jouent sur la supersaturation, il devient possible de découpler la nucléation de la croissance.

a. En même temps que nous avons cherché à reproduire certaines des conditions précédentes qui ont donné des cristaux, nous avons fait varier la concentration en agent précipitant. Nous avons gardé la même concentration de protéine que celle utilisée pour le criblage en robot de cristallisation (20mg/ml).

Tableau 2-1: Effet de la concentration en agent précipitant. Pour chaque condition, le bilan de l'observation après un jour (J+1) ou deux jours (J+2) des gouttes de cristallisation est donné.

Condition de cristallisation		Observation
0.1M NaMes pH6.5	20% à 35% de PEG 4000	à J+1, plein et trop de cristaux
0.1M Na acetate, 0.1M NaMes pH6.5	20% à 35% de PEG 2000	à J+1, beaucoup de cristaux et des précipités pour les plus fortes concentrations de PEG
0.2M LiSO4, 0.1M Tris HCl pH8.5	25 à 35% de PEG 3000	rien
0.5M LiCl, 0.1M Tris HCl pH8.5	20 à 30% de PEG 8000	à J+2, un petit cristal pour 25% PEG qui devient un beau cristal à J+3, mais il colle au fond du puits



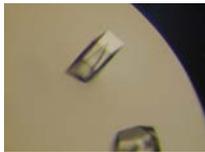
0.5M LiCl, 0.1M NaMes pH6.5	20% à 35% de PEG 6000	rien
1M LiCl, 0.1M Na acetate pH6.3	20% à 35% de PEG 6000	à J+2, beaux cristaux + peau pour 30% et 35% de PEG

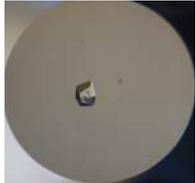


Nous observons que, dans certaines conditions, une partie de la protéine passe dans la phase insoluble formant une sorte de un gel suggérant que la protéine est trop concentrée. De plus, les cristaux collent au fond du puits dans ce gel et sont donc difficiles à prélever. Pour la suite, nous avons décidé de diminuer la concentration de protéine, d'un facteur deux au moins. Dans d'autres conditions, des cristaux apparaissent très rapidement. Or si la croissance initiale est trop rapide, elle peut conduire à l'incorporation de défauts dans les cristaux provoquant un arrêt prématuré de la croissance cristalline et/ou des cristaux mal formés. Une bonne croissance cristalline est favorisée si la supersaturation est atteinte très doucement. Pour la suite, nous avons décidé de jouer sur la concentration en PEG pour ralentir la croissance cristalline et favoriser l'obtention de cristaux de bonne qualité et de bonne taille.

**b.** Ainsi, une nouvelle série d'essais de cristallisation a été réalisée avec une concentration en protéine divisée par deux (10 mg/ml). Nous avons également testé différents additifs (LiCl, Li<sub>2</sub>SO<sub>4</sub> et NaCl) dans l'objectif d'augmenter la dureté des cristaux pour une meilleure résolution.

Tableau 2-2 : Effet de la concentration en agent précipitant et du type et de la concentration d'additifs. Pour chaque condition, le bilan de l'observation après un jour (J+1), deux jours (J+2) ou plus des gouttes de cristallisation est donné.

Condition de cristallisation	Observation	Image
0.1M NaMes pH6.5 16% à 26% de PEG 4000	à J+4, un beau cristal pour 26% PEG	
0.1M Na acetate pH6.3 22% à 30% de PEG 6000	à J+1, beaucoup et trop de cristaux pour 27% et 30% PEG	
27% PEG 6000, 0.1, 0.5 et 1M Na acetate pH6.3 de LiCl	à J+1, des beaux cristaux pour 100mM LiCl et à J+3, des cristaux pour 500mM LiCl	

27% PEG 6000, 0.1, 0.5 et 1M 0.1M Na de NaCl acetate pH6.3	à J+1, 1 cristal pour 100mM NaCl	
27% PEG 6000, 0.05, 0.1, 0.25, 0.1M Na 0.5 et 1M de acetate pH6.3 Li <sub>2</sub> SO <sub>4</sub>	à J+1, des cristaux pour 50mM Li <sub>2</sub> SO <sub>4</sub> et à J+2, des cristaux pour 100mM Li <sub>2</sub> SO <sub>4</sub>	

Nous observons que l'ajout des additifs LiCl et NaCl permet de diminuer la nucléation, ce qui est en faveur de l'obtention de monocristaux. Les cristaux observés sont relativement beaux mais présentent quelques défauts, suggérant que la croissance doit encore être ralentie.

Un des cristaux obtenus dans la condition 26% PEG 4000, 0.1M Na Mes pH6.5 a été monté en capillaire et exposé aux rayons X (longueur d'onde=1.5418 Å, distance au détecteur=120mm, exposition 10 min) sur le diffractomètre de laboratoire (générateur de RX Rigaku HR2, détecteur image-plate Mar Research) et diffracte jusqu'à 2.4 Å.

Nous avons testé différentes tailles de goutte puisque la variation du volume des gouttes influence les cinétiques de cristallisation et donc la taille des cristaux de protéine. En gardant la condition de cristallisation 27% de PEG 6000, 500mM LiCl, 0.1M Na acetate pH 6.3, nous testons 6 tailles de goutte (3µl protéine+2µl réservoir, 3µl p+1.5µl r, 1.5µl p+2µl r, 2µl p+2µl r, 2.5µl p+2.5µl r, 3µl p+3µl r). De beaux cristaux sont apparus dans les gouttes de 2µl+2µl et 2.5µl+2.5µl, c'est pourquoi nous continuons avec des gouttes de cristallisation de 2µl protéine+2µl réservoir.

c. Dans certaines conditions précédentes, il n'y a pas eu de cristaux. Par ensemencement, nous cherchons à voir si la nucléation peut être déclenchée dans ces conditions. Du *streak seeding* (ensemencement en ligne) est effectué dans les gouttes qui n'ont pas donné de cristaux dans les boîtes de cristallisation précédentes. Dans cette méthode, un poil de moustache de chat est utilisé pour toucher un cristal existant puis transférer des petits fragments du cristal en faisant une ligne dans une goutte sans cristal. Les fragments de cristaux introduits servent de base à la nucléation. Par conséquent, la croissance cristalline est découplée de la nucléation.



Figure 2-2 : Photos de quatres gouttes après ensemencement à partir d'un cristal de la condition 27% PEG 6000, 0.1M Na acétate pH6.3.

Dès 1 jour après l'ensemencement, nous avons observé des petits cristaux sur les lignes d'ensemencement de toutes les gouttes (sauf pour 0.1M Na acétate pH6.3, 27% PEG 6000, 1M NaCl où des cristaux apparaissent 2 jours après). Dans ces conditions, le degré de supersaturation est tel que l'étape limitante est la nucléation, la nucléation nécessitant une supersaturation supérieure à la croissance. La nucléation est donc, dans ces conditions, plus inhibée que la croissance, ce qui est favorable à l'obtention de monocristaux. De plus, la

plupart des conditions de cristallisation a donné rapidement des cristaux après ensemencement suggérant que la concentration de protéine pouvait encore être diminuée pour ralentir la croissance.

**d.** Ainsi, une nouvelle série d'essais de cristallisation a été réalisée avec une concentration de protéine divisée par quatre (5 mg/ml) et différents additifs ont été testés.

Tableau 2-3 : Effet d'ajout d'additifs. Pour chaque condition, le bilan de l'observation après un jour des gouttes de cristallisation est donné.

Condition de cristallisation	Observation
27% PEG 6000, 500mM LiCl, 10% éthylène glycol 0.1M Na acetate pH6.3	de beaux cristaux
27% PEG 6000, 500mM LiCl, 5% 2-propanol 0.1M Na acetate pH6.3	de beaux cristaux
27% PEG 6000, 500mM LiCl, 10% glycérol 0.1M Na acetate pH6.3	cristaux après ensemencement
27% PEG 6000, 500mM LiCl, 10% MPD 0.1M Na acetate pH6.3	rien

Nous avons ensuite vérifié qu'avec la condition 10% éthylène glycol, 27% PEG 6000, 500mM LiCl 0.1M Na acétate pH6.3 qui parait la plus intéressante, les cristaux diffractent à haute résolution. Un cristal de cette condition a été testé sur le diffractomètre de laboratoire et des données ont été collectées jusqu'à 2.1 Å.

*En conclusion*, la solution de cristallisation suivante a été retenue : 10% éthylène glycol, 27% PEG 6000, 500mM LiCl, 0.1M Na acétate pH 6.3, pour une concentration de protéine de 5 mg/ml et pour des gouttes de 2µl protéine +2µl réservoir. Dans ces conditions, des cristaux apparaissent dès un jour et grossissent jusqu'à environ 400 µm (figure 1 de l'article 1).

### **2.1.3. Recherche de dérivés lourds et préparation des dérivés tungstate**

La méthode de trempage est utilisée pour l'introduction d'atomes lourds (et aussi d'inhibiteurs, activateurs, substrats, cryo-protectants) dans la solution de cristallisation. Dans un cristal de protéine, il y a un fort pourcentage de solvant (de 27 à 95%). De manière générale, les canaux de solvants sont suffisamment larges pour permettre la diffusion de la plupart des petites molécules en tout point de la surface de la protéine accessible en solution excepté pour les régions impliquées dans les contacts cristallins. Cependant, dans certains cas, les contraintes du cristal peuvent empêcher les changements conformationnels ou des réarrangements de la protéine dans le cristal causée par l'introduction de ces molécules. Dans d'autres cas, les forces qui conduisent aux changements conformationnels peuvent être suffisamment fortes pour dépasser les contraintes du réseau perturbant ainsi les contacts cristallins et intermoléculaires conduisant à des cristaux qui se craquent ou se dissolvent.

D'autres réseaux cristallins peuvent être plus flexibles et capables de s'accommoder des changements conformationnels et tandis que les cristaux peuvent craquer initialement, ils peuvent ensuite se ré-organiser en un autre arrangement cristallin et parfois améliorer la cristallinité. Globalement, par trempage, la liaison des atomes lourds n'est pas forcément représentative d'une liaison réelle car les réarrangements de la protéine peuvent être inhibés par le réseau cristallin. Par contre, dans le cas d'une co-cristallisation, la formation du complexe n'est pas contrainte par le cristal mais la solubilité et la conformation du complexe peuvent être suffisamment différentes de celles de la protéine native pour que de nouvelles conditions de cristallisation soient à déterminer.

Pour commencer, nous avons testé par trempage différents atomes lourds à des concentrations de l'ordre du millimolaire et nous avons regardé si les cristaux se craquent en présence de l'atome lourd. Si c'est le cas, la concentration de l'atome lourd peut être diminuée et sinon il faut vérifier que l'atome lourd est bien entré dans le cristal. Ainsi, la recherche d'atomes lourds est un processus d'essais-erreurs. Pratiquement, les cristaux obtenus précédemment étaient transférés dans des gouttes de 20µl de solution avec atome lourd.

Tableau 2-4 : Observations faites suite au trempage de différents atomes lourds du *Heavy Atom Kit* de *Hampton Research* après 20 minutes et après 1 heure.

<b>1mM atomes lourds</b>	TlCl <sub>3</sub>	TlCl	PbCl <sub>2</sub>	(CH <sub>3</sub> ) <sub>3</sub> Pb(CH <sub>3</sub> COO) <sub>2</sub>	K <sub>2</sub> IrCl <sub>6</sub>
à t=20min	craqué	rien	non	craqué	non
à t=1heure	craqué	craqué	bords abimés	craqué	non
<b>1mM atomes lourds</b>	IrCl <sub>3</sub>	K <sub>2</sub> OsO <sub>4</sub>	(NH <sub>4</sub> ) <sub>2</sub> OsBr <sub>6</sub>	Na <sub>2</sub> WO <sub>4</sub>	(NH <sub>4</sub> ) <sub>2</sub> WS <sub>4</sub>
à t=20min	non	?	craqué un peu	craqué un peu	non
à t=1heure	non	?	craqué un peu	craqué	non
<b>1mM atomes lourds</b>	Eu(NO <sub>3</sub> ) <sub>3</sub>	GdCl <sub>3</sub>	LuCl <sub>3</sub>	YbCl <sub>3</sub>	DyCl <sub>3</sub>
à t=20min	non	non	non	non	non
à t=1heure	une craqueure	non	non	non	non
<b>1mM atomes lourds</b>	K <sub>2</sub> ReCl <sub>6</sub>	K <sub>2</sub> PtCl <sub>4</sub>	K <sub>2</sub> PtCl <sub>6</sub>	K <sub>2</sub> Pt(SCN) <sub>6</sub>	HgCl <sub>2</sub>
à t=20min	non	non	non	non	non
à t=1heure	bords abimés	non	non	non	non
<b>1mM atomes lourds</b>	C <sub>7</sub> H <sub>5</sub> ClHgO <sub>2</sub>	HgO	Mersalyl	N <sub>3</sub> O <sub>9</sub> Sm	
à t=20min	non	non	craque un peu	rien	
à t=1heure	un peu ?	non	craque un peu	rien	
<b>atomes lourds</b>	5mM Hg(CH <sub>3</sub> COO) <sub>2</sub>	1 à 0.05mM UO <sub>2</sub> (NO <sub>3</sub> ) <sub>2</sub>		1mM, 0.5mM K <sub>2</sub> HgI <sub>4</sub>	
à t=20min	craqué	craqué		craqué	

Lors de ces essais de trempage dans des solutions d'atomes lourds, la plupart des cristaux de protéine craquent, en particulier avec le tungstate, mais les cristaux ne se dissolvent pas. Il semble que, dans ces conditions là, les atomes lourds jouent plutôt sur les contacts cristallins que sur la solubilité, ce qui suggère d'essayer des co-cristallisations avec ces atomes lourds.

Des essais de co-cristallisation avec 2 types d'atomes lourds, uranyle (UO<sub>2</sub>(NO<sub>3</sub>)<sub>2</sub>) et tungstène (Na<sub>2</sub>WO<sub>4</sub>), et avec des concentrations allant de 0.5 à 1mM dans une solution 10% éthylène glycol, 27% PEG6000, 0.1M Na acétate pH6.3 ont été effectués. Beaucoup de précipités sont observés indiquant que les atomes lourds semblent quand même jouer sur la

solubilité. De nouveaux essais ont donc été effectués en diminuant encore la concentration de protéine (2.5mg/ml) et en diminuant la quantité de précipitant (20% éthylène glycol, 20% PEG6000). De beaux cristaux apparaissaient dans les conditions suivantes : 0.5mM  $\text{UO}_2(\text{NO}_3)_2$  ou 1.5mM  $\text{UO}_2(\text{NO}_3)_2$  ou 0.5 mM  $\text{Na}_2\text{WO}_4$  dans 20% éthylène glycol, 20% PEG6000, 0.1M Na acétate pH6.3. Nous avons également réussi à obtenir des cristaux trempés dans une solution de  $\text{Na}_2\text{WO}_4$  analysables par diffraction aux rayons X en jouant sur la concentration en dérivés tungstate (5mM, 10mM, 15mM, 20mM), sur la durée du trempage (1nuit, plusieurs jours) et sur l'ajout de tungstate dans la solution cryoprotectante.

*En conclusion*, nous avons ainsi préparé divers dérivés tungstate par trempage et co-cristallisation. La préparation des co-cristaux nous a permis d'observer des changements conformationnels causés par le tungstate que nous n'aurions pas pu mettre en évidence avec les cristaux trempés (cf article 2).

## 2.2. Détermination des phases

*Complément à l'article 1 concernant le phasage.*

Dans la conclusion de l'article 1, nous avons souligné que, de manière surprenante, les différents dérivés tungstène préparés ne permettaient pas d'obtenir des phases interprétables bien que les sites d'atomes lourds pouvaient être clairement identifiés. Par la suite, nous avons résolu ce problème et c'est ce que nous allons présenter dans ce paragraphe. En cristallographie des macromolécules biologiques, il existe différentes méthodes de détermination expérimentale des phases utilisant la diffusion anormale, comme cela est décrit dans la partie Matériel et Méthodes (page 30). Pour obtenir des phases à partir des dérivés tungstène, nous avons utilisé la méthode de diffusion anormale à une longueur d'onde (SAD) et à plusieurs longueurs d'onde (MAD).

### 2.2.1. Enregistrement d'un jeu MAD sur un dérivé tungstate et estimation du signal anomal

Le dérivé tungstène consiste en des cristaux du domaine C-terminal de hKIN17 trempés une nuit dans 1mM  $\text{Na}_2\text{WO}_4$  et transférés dans une solution cryo-protectante sans tungstate pendant environ cinq minutes avant congélation dans l'azote. Un jeu MAD a été enregistré sur la ligne FIP-BM30A de l'ESRF (Roth et al. 2002) jusqu'à une résolution de 1.9 Å.

Un spectre d'absorption (Figure 2-3) a été enregistré en mode fluorescence et l'utilisation du programme CHOOCH (Evans et al. 2001) a permis de choisir les longueurs d'onde optimales pour l'enregistrement des données MAD. La longueur d'onde  $w1$  (1.215287Å) correspond au pic où le signal anomal est maximal ( $f''$  maximal), la longueur d'onde  $w2$  (1.215958Å) correspond au pic d'inflexion où le signal dispersif est maximal ( $|f'|$  maximal) et la longueur d'onde  $w3$  (1.203745Å) correspond à une longueur d'onde lointaine de ces deux maxima. Les facteurs de diffusion déduits de ce spectre sont  $f''_{\text{max}} = 31.66$  électrons et  $|f'_{\text{min}}| = 34.54$  électrons.

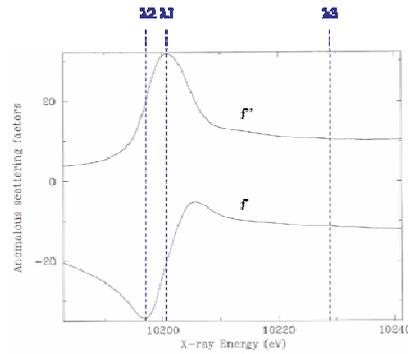


Figure 2-3 : Spectre d'absorption enregistré en mode fluorescence obtenu avec les cristaux du domaine C-terminal de hKIN17 trempés dans 1mM Na<sub>2</sub>WO<sub>4</sub> sur la ligne de lumière BM-30 à l'ESRF. Les trois longueurs d'onde choisies pour l'enregistrement d'un jeu MAD sont notées en bleu.

Des cartes de Patterson de différence anormale ont été calculées pour les trois longueurs d'onde en prenant toutes les données et en appliquant les critères de réjection indiqués par Scaleit (CCP4 1994). Les sections de Harker  $x = \frac{1}{2}$ ,  $y = \frac{1}{2}$ ,  $z = \frac{1}{2}$  sont données sur la Figure 2-4 et révèlent des pics de Harker forts et identiques pour les trois longueurs d'onde.

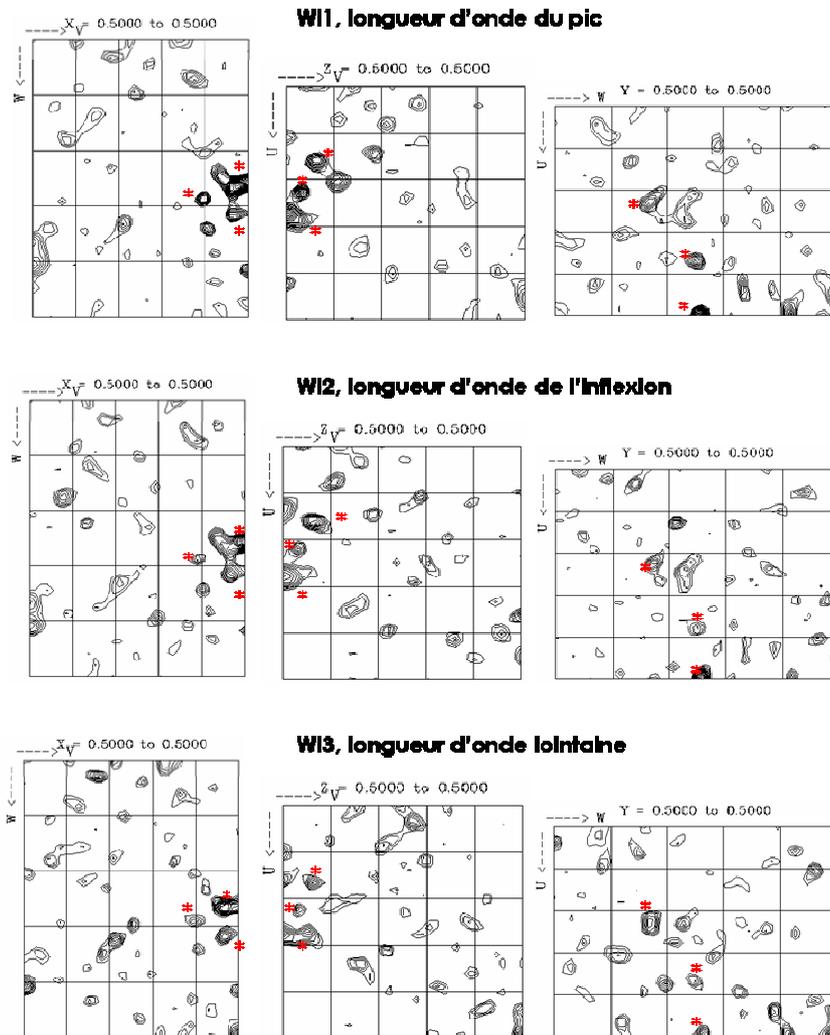


Figure 2-4 : Sections de Harker des cartes de Patterson de différence anormale calculées pour chaque longueur d'onde du jeu MAD enregistré sur les cristaux du domaine C-terminal de hKIN17 trempés dans 1mM Na<sub>2</sub>WO<sub>4</sub>. Le niveau de contour va de 1.5 à 20.0 sigma au-dessus de la densité moyenne par intervalles de 0.5 sigma. Les pics de Harker les plus intenses et retrouvés pour les trois longueurs d'onde sont marqués d'un astérisque rouge.

Par ailleurs, les coefficients de corrélation entre les différences anomales (DANO) des différentes longueurs d'onde ont été calculés avec SFTOOLS (Tableau 2-5).

Tableau 2-5 : Valeur des corrélations entre les différences anomales des trois longueurs d'onde du jeu MAD enregistré sur les cristaux du domaine C-terminal de hKIN17 trempés dans 1mM Na<sub>2</sub>WO<sub>4</sub> selon la coquille de résolution (entre Dmin et Dmax).

Dmin	Dmax	Correl wl1 / wl2	Correl wl1 / wl3	Correl wl2 / wl3
36.81	5.18	0.9671	0.9316	0.9314
5.18	4.12	0.9078	0.8635	0.8407
4.12	3.60	0.8916	0.8431	0.8375
3.60	3.27	0.8932	0.8383	0.8054
3.27	3.03	0.9019	0.8696	0.8720
3.03	2.85	0.9312	0.9021	0.8951
2.85	2.71	0.9317	0.8953	0.8812
2.71	2.59	0.8995	0.8711	0.8521
2.59	2.49	0.9200	0.8916	0.8733
2.49	2.41	0.8926	0.8489	0.8119
2.41	2.33	0.8427	0.8503	0.7718
2.33	2.27	0.8007	0.8014	0.7568
2.27	2.21	0.7298	0.6168	0.5814
2.21	2.15	0.8186	0.7931	0.7660
2.15	2.10	0.6791	0.7254	0.6608
2.10	2.06	0.7040	0.6267	0.5840
2.06	2.02	0.7435	0.7109	0.7084
2.02	1.98	0.6934	0.5379	0.4682
1.98	1.94	0.7799	0.5200	0.4879
1.94	1.91	0.7911	0.6414	0.6515

Les valeurs des corrélations sont très élevées, indiquant que ces différences anomales ne sont pas dues à du bruit aléatoire mais bien dues à une source commune de signal anormal, c'est-à-dire à des sites d'atomes lourds. En général, on considère qu'une corrélation supérieure à 0.3 est significative.

*En conclusion*, nous pouvons identifier clairement le signal anormal provenant des atomes de tungstate.

### **2.2.2. Essai de phasage avec le jeu de données du dérivé tungstate**

Pour les différentes longueurs d'onde, une recherche d'atomes lourds sur les différences anomales a été faite en utilisant en parallèle les programmes SHELXD (Schneider et al. 2002) et HySS (Grosse-Kunstleve et al. 2003). Plusieurs recherches sont lancées en demandant de trouver entre 2 et 10 sites d'atomes lourds. A l'issue de ces recherches, cinq sites sont identiques aux solutions proposées par les programmes SHELXD et HySS.

Cette sous-structure de cinq atomes de tungstate est utilisée pour un affinement SAD avec la longueur d'onde du pic à l'aide du programme SHARP (Bricogne et al. 2003). L'affinement se déroule bien puisque l'affinement des sites (leur position, leur occupation et

les facteurs B) convergent mais les sites affinés ont des faibles taux d'occupation. L'observation des cartes résiduelles montre des pics bien définis et permet d'inclure de nouveaux sites et de poursuivre l'affinement. En final (après trois runs), 16 sites s'affinent correctement, c'est-à-dire que les facteurs de température de tous les sites sont de l'ordre de ceux des atomes de la protéine et les coordonnées de sites convergent bien quoique les sites aient des faibles taux d'occupation.

Cette solution affinée est ensuite utilisée dans un affinement MAD, en incluant les deux autres longueurs d'onde puis en ajoutant un jeu de données natif isomorphe en référence. L'affinement se déroule de la même manière que précédemment. En final, les 16 sites affinés ont toujours des occupations faibles. Il est curieux d'observer que ces sites d'atomes lourds sont tous relativement proches les uns des autres (Tableau 2-6).

Tableau 2-6 : 16 sites d'atomes lourds (tungstate) en fin d'affinement MAD dans Sharp.

CRYST 45.812 46.294 60.689 90.00 90.00 90.00 P212121						
		Coordonnées			Occ	Facteur B
ATOM	1 W	18.03	22.45	20.53	0.20	40.43
ATOM	2 W	15.77	25.04	20.06	0.13	31.37
ATOM	3 W	18.08	25.03	24.36	0.05	11.32
ATOM	4 W	14.80	19.09	21.19	0.05	19.83
ATOM	5 W	13.15	23.15	19.32	0.03	34.84
ATOM	6 W	15.28	22.73	22.49	0.06	13.29
ATOM	7 W	12.22	21.38	20.82	0.09	30.12
ATOM	8 W	17.42	21.86	23.86	0.06	25.28
ATOM	9 W	17.74	24.59	21.53	0.07	20.14
ATOM	10 W	15.80	24.89	16.66	0.04	23.25
ATOM	11 W	14.63	21.14	18.78	0.03	18.79
ATOM	12 W	14.77	19.55	24.60	0.05	26.91
ATOM	13 W	17.88	19.53	22.72	0.05	22.72
ATOM	14 W	17.65	19.98	19.41	0.04	20.08
ATOM	15 W	14.54	18.78	17.71	0.05	34.64
ATOM	16 W	15.44	21.78	15.71	0.05	33.99

Bien que les statistiques de phasage (figure de mérite acentrique supérieure à 0.5 jusqu'à 2 Å, et FOM<sub>acen</sub>/FOM<sub>cen</sub>=0.52845/0.31654 ; pouvoir de phasage anomal égal à 2.557, 1.815 et 1.612 pour w11, w12 et w13 respectivement) soient indicatrices de bonne qualité de phasage, les cartes après aplatissage de solvant restent ininterprétables.

L'explication à ce problème de phasage ne nous est pas apparue clairement. L'analyse des différences anormales avec le module `iotbx.reflection_statistics` du programme HySS montre que les différences anormales paraissent faibles (souvent inférieures à 5%) pour le dérivé tungstate. Il semble donc que le dérivé soit faiblement substitué. C'est effectivement ce qui est observé lors de l'affinement dans SHARP à la fin duquel les atomes ont des taux d'occupation très faibles. Cependant, il est à remarquer que le taux d'occupation affiné dans SHARP n'est qu'un taux relatif à l'échelle des données. Tout de même, le  $f''$  mesuré est très élevé (31.66 au pic) et même pour un site faiblement occupé, le signal peut être à peu près équivalent au signal d'un soufre à grande longueur d'onde qui permet de phaser des structures (Dauter et al. 1999). Et encore plus, 16 sites, même faiblement substitués, devraient être suffisants pour phaser une structure de 127 résidus. De plus, les paramètres peuvent être affinés sans problème, indiquant clairement que même si le signal des atomes lourds est faible, il se distingue nettement du bruit.

A l'issue de ces calculs, nous avons décidé de préparer de nouveaux dérivés au tungstate en augmentant la concentration et/ou le temps de trempage afin d'obtenir des sites plus fortement substitués. Nous avons aussi décidé de ne plus faire de backsoaking ou trempage dans la solution-cryo sans tungstate car les atomes de tungstate qui ne sont peut-être pas liés de manière très forte à la protéine peuvent se dissocier au moment du backsoaking. Nous avons toujours observé le même comportement avec tous les dérivés tungstate préparés : coefficients de corrélation sur les différences anormales élevés, belles cartes de Patterson, solutions d'atomes lourds de HySS et SHELXD nettes et concordantes, affinement des sites d'atomes lourds dans SHARP avec des taux d'occupation assez faibles, statistiques de phasage correctes mais ensuite des cartes de densité électronique après aplatissage de solvant ininterprétables.

A l'issue de ces différents essais de phasage avec le tungstate, nous avons aussi choisi de tester d'autres atomes lourds, en particulier des halides, comme l'iode qui nous a permis de résoudre la structure. Néanmoins, nous étions intéressés de comprendre pourquoi le phasage avec le tungstate ne permettait pas d'obtenir des cartes interprétables. Pour cela, la structure obtenue à partir du phasage SIRAS avec l'iode (cf article 2) a été utilisée pour faire du remplacement moléculaire dans les données de dérivé tungstate à l'aide du programme Refmac5 (CCP4 1994). Nous voulions savoir si la structure de dérivé tungstène permettrait de confirmer la présence des atomes de tungstate, d'expliquer l'assemblage curieux des atomes de tungstate et de comprendre pourquoi ce type de dérivé ne permettait pas d'obtenir des cartes interprétables. Après remplacement moléculaire dans les données du dérivé tungstène, le calcul de carte de Fourier différence anormale a permis de mettre en valeur les sites qui avaient une diffusion anormale et donc de localiser les atomes de tungstate. Les positions trouvées correspondent aux sites utilisés dans l'affinement SHARP pour le phasage. Ces sites ont donc été ajoutés à la structure et affinés (occupations et facteurs B) dans Refmac5. Là encore, l'occupation de ces sites a été affinée à des valeurs très faibles. Les atomes de tungstate ont été regroupés en cluster à des distances de l'ordre de 3 à 4 Å les uns des autres. Nous avons ainsi confirmé la présence des sites de tungstate, faiblement substitués.

### **2.2.3. Résolution du problème de phasage avec autoSHARP**

Un jeu de données a été enregistré sur un co-cristal de la protéine avec 0.5mM Na<sub>2</sub>WO<sub>4</sub>. Ce jeu de données, à la longueur d'onde du pic (1.2144Å), est très redondant car des images ont été enregistrées sur 360°, pour un cristal de groupe d'espace P2<sub>1</sub>2<sub>1</sub>2<sub>1</sub>. Par ailleurs, la complétude des données à haute et basse résolution est très bonne (100% et 99.4% respectivement) et la résolution est de 2.3Å. Un spectre de fluorescence a permis de déduire que  $f'$  et  $f''$  sont égaux à -3.3 et 52.2 au pic. D'après l'analyse avec le module `iotbx.reflection_statistics` de HySS, le signal anormal était élevé à plus de 16%. Mais là encore, le phasage classique dans SHARP a été sans succès bien que l'affinement des sites de tungstate trouvés se déroule très bien.

Finalement, nous avons réussi avec autoSHARP (version 2.2.0) à trouver les sites, les affiner, obtenir une très belle carte de densité et construire presque la totalité du modèle automatiquement avec ARP/wARP. Comme attendu, la structure obtenue se superpose parfaitement à la structure publiée dans l'article 2. La clé du problème a été :

- d'autoriser les sites à être plus proches les uns des autres que la valeur par défaut (3.5Å) à la fois dans SHELXD pour trouver les sites et dans l'analyse des cartes résiduelles pour ajouter

des sites. Cela est nécessaire dans le cas de cluster. Ainsi, nous obtenons un modèle d'atomes lourds plus complet. Pour cela, la ligne *autoSHARP\_SHELX\_Mind\_W="-2"* a été ajoutée dans le fichier *sharpfiles/cardfiles/WhateverJobIdentifieur.shelx.def* et la ligne *autoSHARP\_RESID\_NEWSITE\_DIST\_NOT\_W=2.00* dans le fichier *sharpfiles/cardfiles/WhateverJobIdentifieur.resid.def*

- d'inclure dans le run ARP/wARP tous les sites d'atomes lourds, même ceux qui s'affinent avec une faible occupation.

### 2.3. Structure à haute résolution du domaine C-terminal par radio-cristallographie

L'article 2 présente la structure obtenue à 1.45 Å du domaine C-terminal de hKIN17.

Résumé article 2 : *A tandem of SH3-like domains participates to RNA binding in KIN17, a human protein activated in response to genotoxics* (2006), A le Maire, M Schiltz, E Stura, G Pinon-Lataillade, J Couprie, M Moutiez, M Gondry, J Angulo and S Zinn-Justin, J. Mol. Biol 364, 764-776.

La structure à haute résolution de cette région, présentée dans l'article, révèle un tandem de sous-domaines de type SH3. Ce domaine C-terminal de KIN17 lie les homopolymères de ribonucléotides avec les mêmes préférences que la protéine entière. L'analyse de sa structure complexée au tungstate montre qu'il existe des possibilités d'adaptabilités structurales dans ce domaine. L'interaction avec les atomes de tungstate est médiée par plusieurs lysines localisées dans un sillon chargé positivement et conservé à l'interface entre les deux sous-domaines. Ce sillon pourrait être le site d'interaction avec l'ARN puisque deux mutants de ces lysines ont une affinité réduite pour l'ARN.

# A Tandem of SH3-like Domains Participates in RNA Binding in KIN17, a Human Protein Activated in Response to Genotoxics

Albane le Maire<sup>1</sup>, Marc Schiltz<sup>2</sup>, Enrico A. Stura<sup>1</sup>  
Ghislaine Pinon-Lataillade<sup>3</sup>, Joël Couprie<sup>1</sup>, Mireille Moutiez<sup>1</sup>  
Muriel Gondry<sup>1</sup>, Jaime F. Angulo<sup>3</sup> and Sophie Zinn-Justin<sup>1\*</sup>

<sup>1</sup>CEA, DSV, DIEP  
Gif-sur-Yvette, F-91191  
France

<sup>2</sup>EPFL, Laboratoire de  
Cristallographie, Lausanne  
CH-1015, Switzerland

<sup>3</sup>CEA, DSV, DRP  
Fontenay-aux-Roses  
F-92265, France

The human KIN17 protein is an essential nuclear protein conserved from yeast to human and expressed ubiquitously in mammals. Suppression of *Rts2*, the yeast equivalent of gene *KIN17*, renders the cells unviable, and silencing the human *KIN17* gene slows cell growth dramatically. Moreover, the human gene *KIN17* is up-regulated following exposure to ionizing radiations and UV light, depending on the integrity of the human global genome repair machinery. Its ectopic over-expression blocks S-phase progression by inhibiting DNA synthesis. The C-terminal region of human KIN17 is crucial for this anti-proliferation effect. Its high-resolution structure, presented here, reveals a tandem of SH3-like subdomains. This domain binds to ribonucleotide homopolymers with the same preferences as the whole protein. Analysis of its structure complexed with tungstate shows structural variability within the domain. The interaction with tungstate is mediated by several lysine residues located within a positively charged groove at the interface between the two subdomains. This groove could be the site of interaction with RNA, since mutagenesis of two of these highly conserved lysine residue weakens RNA binding.

© 2006 Elsevier Ltd. All rights reserved.

**Keywords:** KIN17 protein; X-ray structure; SH3-like fold; RNA binding; tungstate

\*Corresponding author

## Introduction

Cells are continuously submitted to attack, either from reactive oxygen species resulting from the normal cellular metabolism or by exogenous agents such as radiation and chemicals. Consequent DNA damage triggers specific signalling to activate the repair machinery. The human gene *KIN17*, located in chromosome 10, participates in the general response to genotoxics. It is up-regulated following DNA damage produced by UV-C or by ionizing radiation,<sup>1,2</sup> and this response to UV light is strictly dependent on the presence of the functional nucleotide excision repair proteins XPA and XPC.<sup>2,3</sup> The ectopic over-

expression of gene *KIN17* modifies the nuclear morphology and inhibits S-phase progression, thus blocking cell growth.<sup>4,5</sup>

*KIN17* encodes for a nuclear protein conserved from yeast to human. Its ubiquitous expression suggests its participation in a regulatory mechanism common to all cell types. Suppression of *Rts2*, the yeast *KIN17* homolog, renders the cells unviable,<sup>6</sup> and silencing the human gene *KIN17* impedes cell growth dramatically.<sup>7</sup> *KIN17* is part of high molecular mass complexes that mediate different types of nucleic acid transactions.<sup>8,9</sup> The mammalian *KIN17* protein binds to curved DNA found at hot-spots of illegitimate recombination in eukaryotic chromosomes,<sup>10–12</sup> and partially complements the bacterial transcriptional regulator in *Escherichia coli* called H-NS.<sup>13</sup> Moreover, proteomic analysis detected protein *KIN17* in several complexes of the human spliceosome, a large RNA–protein assembly consisting of small ribonucleoprotein particles (snRNP) and associated non-small ribo-

Abbreviation used: GST, glutathione-S-transferase.

E-mail address of the corresponding author:  
[sophie.zinn@cea.fr](mailto:sophie.zinn@cea.fr)

nucleoproteins.<sup>14,15</sup> Interaction between KIN17 and RNA was further characterized by Pinon-Lataillade *et al.*,<sup>16</sup> showing that the mouse KIN17 protein interacts with RNA *in vivo* and that human and mouse KIN17 proteins are able to directly bind different types of RNA homopolymers *in vitro*.

Human KIN17 (hKIN17) is a 45 kDa modular protein composed of a 30 kDa N-terminal region conserved in all species, and a 15 kDa tail found only in higher eukaryotes. The N-terminal region exhibits a C<sub>2</sub>H<sub>2</sub> zinc finger motif, which is generally associated with nucleic acid binding function. This domain modulates the interaction between KIN17 and double-stranded DNA as its deletion leads to a reduced DNA binding.<sup>11</sup> In contrast, the C-terminal tail does not affect the DNA binding of KIN17.<sup>11</sup> It is involved in the anti-proliferative effect of KIN17. Indeed, cell growth is not affected by over-expression of KIN17 protein lacking the C-terminal tail, whereas it is inhibited by over-expression of the whole protein.<sup>5</sup>

The C-terminal domain of hKIN17 comprises a 27 residue KOW motif,<sup>17</sup> shared among three families of ribosomal proteins (microbial L24 and eukaryotic eL26 and eL27), the essential bacterial transcriptional elongation factor NusG, the eukaryotic chromatin elongation factor Spt5p, the T54 proteins and the higher eukaryotic KIN17 proteins.<sup>18</sup> In the three-dimensional structures of the ribosomal proteins<sup>19</sup> and the transcription factor NusG,<sup>20</sup> the KOW motif occurs in the first two strands and adjacent loops of a larger src homology 3 (SH<sub>3</sub>)-like domain. L24 protein interacts with several RNAs in the ribosome, which is in agreement with the original proposal that KOW is an RNA-binding motif.<sup>17</sup>

Here, we report the high-resolution X-ray structure of the C-terminal domain of hKIN17 and show that this domain contributes to the RNA-binding properties of the whole protein. We also present a structure of this domain complexed to tungstate that allowed us to identify two highly conserved lysine residues involved in binding to this ion analogous to phosphate, as well as in RNA binding.

## Results

### The C-terminal region of hKIN17 exhibits an all- $\beta$ structure consisting of two tightly packed SH<sub>3</sub>-like subdomains

The boundaries of the C-terminal domain of hKIN17 protein (amino acid residues 268–393) were delineated on the basis of residue conservation from sequence alignment of all known KIN17 proteins (Figure 1(a)) and on the basis of hydrophobic amino acid cluster analysis<sup>21</sup> of the hKIN17 sequence. This domain forms an independently folded unit, as it can be expressed as a soluble and folded protein in solution.

The structure of the isolated C-terminal domain of hKIN17 was determined by means of single

isomorphous replacement with anomalous scattering (SIRAS) phasing using data collected on crystals soaked with iodine as described.<sup>22</sup> The structural model, including 211 water molecules, was refined to a resolution of 1.45 Å, leading to final *R* and *R*<sub>free</sub> values of 13.7% and 17.5%, respectively (Table 1). The geometry can be considered as very good, with 94.2% of the residues in the most preferred  $\phi/\psi$  conformations and only one residue in a disallowed region of the Ramachandran plot. All residues in the X-ray crystal structure are well ordered, except for the first seven residues, (the non-wild-type glycine introduced by the cloning vector and the first six residues) for which there is no electron density, and for which a signal corresponding to random coil is seen in NMR experiments (data not shown). This domain is roughly ellipsoidal, with dimensions of 40 Å × 30 Å × 25 Å and organized into two major subdomains (Figure 1(b)). Subdomain A (spanning residues 284–334 of the hKIN17 sequence) and subdomain B (spanning residues 340–391 of the hKIN17 sequence) both adopt a similar five-stranded  $\beta$ -barrel-like fold and are connected to each other by a short linker of five residues. In each subdomain, the first and smaller  $\beta$ -sheet is formed from strands  $\beta$ -1<sub>A(or B)</sub>, part of  $\beta$ -2<sub>A(or B)</sub> and  $\beta$ -5<sub>A(or B)</sub>, and the second and larger  $\beta$ -sheet is formed from the remainder of  $\beta$ -2<sub>A(or B)</sub>,  $\beta$ -3<sub>A(or B)</sub> and  $\beta$ -4<sub>A(or B)</sub>. The two  $\beta$ -sheets pack against each other at approximately right-angles because of the helical turn between  $\beta$ -4<sub>A(or B)</sub> and  $\beta$ -5<sub>A(or B)</sub>. Interestingly, the N-terminal segment 276–283 and the C-terminal fragment 391–393 form the interface between the two subdomains. They interact with both subdomains and the C-terminal fragment interacts with the linker region (Figure 1(b)). The surface area of each subdomain buried in the interface of the tandem structure is approximately 1150 Å<sup>2</sup>. This is a relatively large protein–protein interface with a classical percentage of non-polar residues (58 %).<sup>23</sup>

The overall backbone structure of the two subdomains superimposes well, with an rmsd of 1.16 Å on 29 C $\alpha$  atoms, excluding the three variable loops, despite a low level of sequence identity (9% identity; 54% sequence similarity) (Figure 2). A DALI search using the individual A and B subdomain structures shows that these subdomains display structural analogy with known protein domains adopting an SH<sub>3</sub>-like fold. Domain III of NusG (PDB code 1M1G) is the best DALI match for both subdomains, and shows 17% and 20% identity with subdomains A and B, respectively. A high DALI score (>6) is obtained also for the dihydrofolate reductase (PDB code 1vie), the Tudor domain of SMN protein (PDB code 1gv5), the SH<sub>3</sub> domain of the protein kinase p56lck (PDB code 1lck-A) and the ribosomal protein L21e (PDB code 1ffk-N). Thus, the C-terminal domain of hKIN17 has structural similarity with various SH<sub>3</sub>-like domains, even if it lacks similarities in either primary sequence or charge distribution.

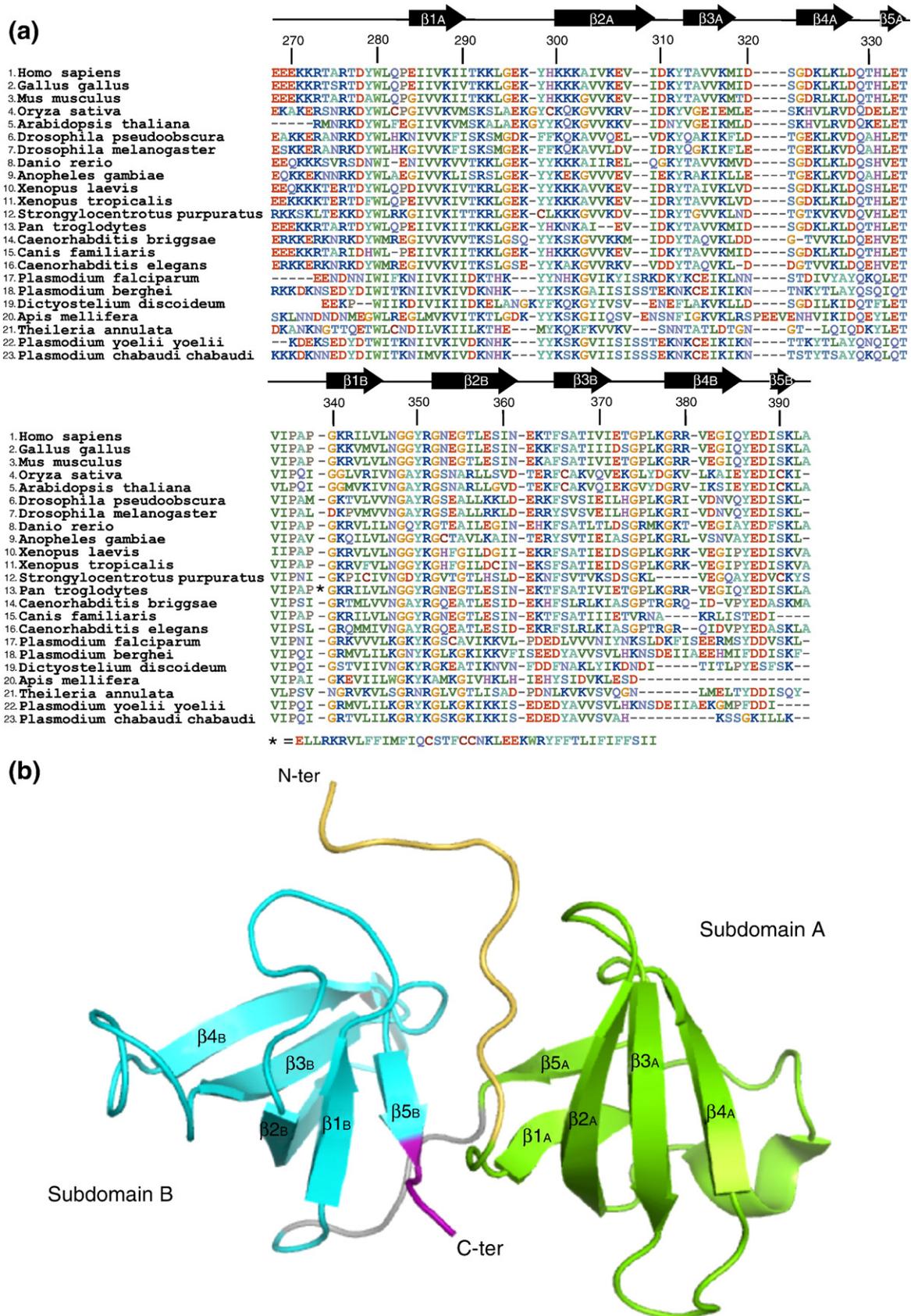


Figure 1 (legend on opposite page)

**Table 1.** Phasing and refinement statistics

<b>A. Phasing</b>	
FoM (acentric/centric)	0.4961/0.3944
Phasing power iso (acentric/centric)	1.980/1.573
Phasing power anomalous	1.250
<b>B. Refinement</b>	
Space group	P2 <sub>1</sub> 2 <sub>1</sub> 2 <sub>1</sub>
Resolution range (Å)	32.62–1.45
Reflections used in refinement	22,249
R (%)	13.7
R <sub>free</sub> (%)	17.5
Total number of atoms	1181
Number of	
Protein atoms	954
Iodine atoms	12
Water molecules	211
Average B-factor (Å <sup>2</sup> )	15.3
r.m.s.d from ideality	
Bond lengths (Å)	0.013
Bond angles (deg.)	1.519
Chirality (deg.)	0.091
Ramachandran plot	
Most-favoured regions (%)	94.2
Additionally allowed regions (%)	4.8
Generously allowed regions (%)	0
Disallowed regions (%)	1 <sup>a</sup>

<sup>a</sup> K311 in loop β<sub>2A</sub>β<sub>3A</sub> is located in a disallowed region of the Ramachandran plot; however, its backbone conformation was well supported by the electron density.

### The C-terminal domain of hKIN17 binds RNA *in vitro*

Since hKIN17 protein is able to recognize RNA homopolymers and can bind RNA *in vivo*,<sup>16</sup> we investigated which region(s) of the protein interact(s) with RNA. Three constructs containing either the N-terminal globular region 1–160 (shown to be folded by NMR analysis; data not shown), the same region without the zinc finger 68–160 (shown to be folded by NMR analysis),<sup>24</sup> or the C-terminal domain 268–393 described here were tested for their ability to bind RNA using the same <sup>32</sup>P-labelled RNA probe as that used to characterize the RNA interaction properties of the full-length hKIN17.<sup>16</sup> The North-western analysis showed that the N-terminal domain comprising the zinc finger interacts efficiently with RNA *in vitro*, and that the deletion of the zinc finger from this domain drastically reduces RNA binding (Figure 3). The C-terminal domain showed an intermediate RNA-binding activity. In comparison, a South-western analysis showed that only the

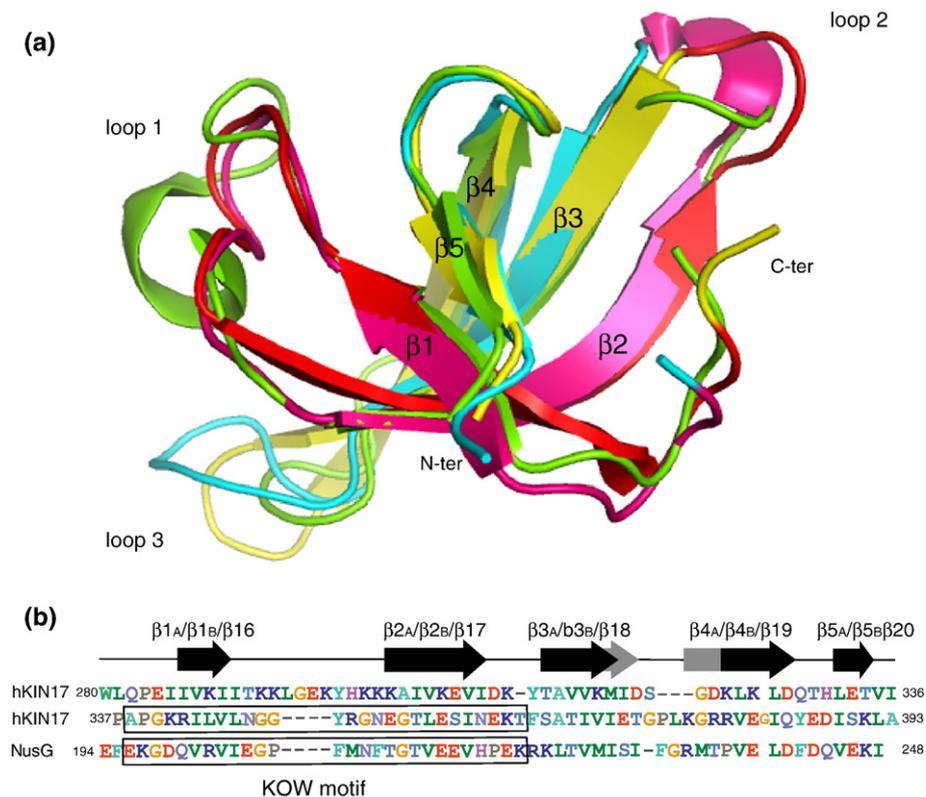
N-terminal domain binds significantly to a double-stranded DNA known to interact with the whole hKIN17 protein (data not shown).

To characterize further the RNA-binding properties of the C-terminal domain, we carried out binding assays with ribonucleotide homopolymers immobilized on agarose beads. This kind of assay has been useful in assessing RNA binding for many other RNA-binding proteins and domains.<sup>25–27</sup> The C-terminal domain of hKIN17 fused to glutathione-S-transferase (GST) binds to all four RNA homopolymers at 100 mM NaCl, with a clear preference for poly(rG) and poly(rU) (Figure 4(a)). The possibility of GST interference was eliminated by repeating the test with GST alone, since no signal was detected using a specific anti-GST antibody. The cleaved protein was also assayed for RNA binding with results comparable to those obtained with the GST-hKIN17 C-terminal domain fusion (Figure 4(a)). To rule out the possibility of non-specific binding due to electrostatic interactions, the assay was repeated at increasing ionic strength varying from 0.1 M to 1 M NaCl, showing that binding of the C-terminal domain of hKIN17 to poly(rG) and poly(rU) is still significant in concentrations of NaCl up to 0.25 M (Figure 4(b)). Such resistance of the C-terminal domain-poly(rG) or poly(rU) interaction to significant ionic strength suggested that, in addition to ionic contacts, hydrophobic interactions are involved in the RNA–protein interaction. The protein also retains its RNA-binding capacity in the presence of either 1 mg/ml of *E. coli* tRNA or 1 mg/ml of heparin (Figure 4(b)). Finally, the heparin-resistant binding,<sup>28,29</sup> and the tRNA-resistant binding, demonstrate the stringency and specificity of the binding of the C-terminal domain of hKIN17 to RNA. A North-western assay confirms the interaction of the C-terminal domain of hKIN17 with poly(rG) and poly(rU) (Supplementary Data). On the basis of these experiments, we conclude that the C-terminal domain of hKIN17 is an RNA-binding domain *in vitro*.

### Tungstate-binding sites reveal important determinants for the binding of RNA

The binding of a protein to RNA is often mediated by interaction with the phosphate backbone of the nucleic acids, and positively charged amino acids are generally found in contact with the phosphate groups.<sup>30</sup> In order to

**Figure 1.** Sequence and structure of the hKIN17 C-terminal domain. (a) Multiple sequence alignment of the C-terminal region of 23 eukaryotic KIN17 proteins. The numbering scheme refers to the sequence of human KIN17. Positively charged residues are in blue, negatively charged residues are in red and hydrophobic residues are in green. The secondary structure elements corresponding to the hKIN17 C-terminal crystal structure are indicated above the alignment. The length of the hKIN17 C-terminal domain is well conserved, except in the Chimpanzee (*Pan troglodytes*) sequence, which harbours a 42 residue insertion between the two subdomains. (b) A ribbon representation of the 3D structure of the hKIN17 C-terminal domain. It consists of two SH3-like subdomains; the first one (A) is in green and the second one (B) is in blue. The short linker between the two subdomains is in grey (segment 335–339) and the two segments involved in the interface between the subdomains are in orange (N-terminal segment 274–283) and in violet (C-terminal segment 391–393).



**Figure 2.** Structural analogues of the two SH<sub>3</sub>-like subdomains. (a) Superposition of the two subdomains of hKIN17 C-terminal domain (colours as in Figure 1) and the domain III of *aaeNusG* (PDB code 1M1G; yellow). The rmsd calculated between the backbones of the two subdomains except for the three variable loops reaches 1.16 Å (on 29 C°). The rmsd calculated between the backbones of subdomain B of hKIN17 C-terminal domain and its closest DALI match, the domain III of *aaeNusG* except for the three variable loops reaches 1.20 Å (on 29 C°). The sequences presenting the KOW motif form the first two β-strands and their adjacent loops (in red for NusG and in pink for KIN17). (b) Sequence alignment of the two subdomains of hKIN17 C-terminal domain with the domain III of *aaeNusG*. A β-bulge occurs in the second subdomain causing the insertion of residue G383. Longer β strands (β18 and β19) for the domain III of *aaeNusG* are drawn in grey. The KOW motifs of the second subdomain of hKIN17 C-terminal domain and the domain III of *aaeNusG* are boxed.

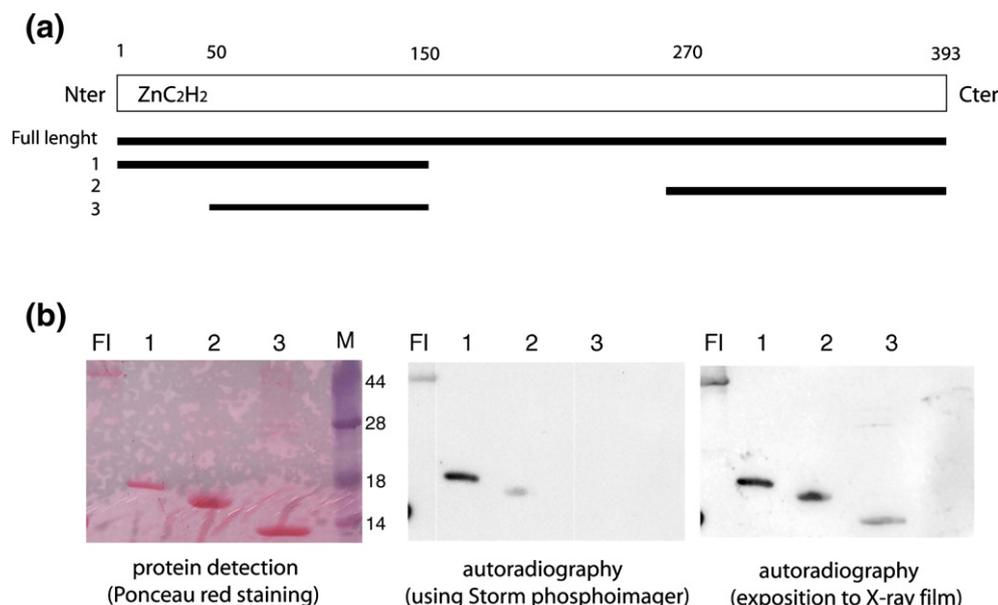
identify the positively charged residues of hKIN17 involved in RNA interaction, we screened the surface of the C-terminal domain using tungstate, which can mimic phosphate groups,<sup>31</sup> and has been useful for identifying phosphate binding sites in other crystal structures.<sup>32–36</sup> As tungsten is substantially heavier than phosphorus, has more electrons and can give rise to a strong anomalous signal, tungstate ions can be identified easily by X-ray diffraction methods. We collected data at the W *L*<sub>III</sub> absorption edge from a crystal soaked in Na<sub>2</sub>WO<sub>4</sub> solution and a crystal obtained by co-crystallizing with Na<sub>2</sub>WO<sub>4</sub> to a resolution of 1.91 Å and 2.30 Å, respectively. Both derivative structures revealed important features of the binding of tungstate anions to hKIN17 C-terminal domain.

The initial 2*F*<sub>o</sub>–*F*<sub>c</sub> and *F*<sub>o</sub>–*F*<sub>c</sub> electron density maps obtained from soaked crystals showed a large patch of electron density, defining a cluster of tungsten atoms, confirmed by the anomalous signal. Twelve strong peaks (>10 σ for the anomalous difference) characterise the entire cluster but only six of these have a refined

occupancy greater than 0.1. The tungstate anions are all located around the cationic NH<sub>3</sub><sup>+</sup> side-chain groups of K302 in β2 of subdomain A, of K341 in β1 of subdomain B and of K391 at the C terminus of β5 in subdomain B of hKIN17 C-terminal domain (Figure 5(a)). A larger cluster of 12 tungstate anions, at almost the same location as in the soaked crystals, is found for the co-crystal (Figure 5(b)). The interaction is characterised by two additional lysine residues (K377' and K297') from molecules related by crystal symmetry and another minor site where one tungsten ion is nestled between the side-chains of R273 and of K317' from a symmetry-related molecule.

#### Derivative structures reveal structural variability within the C-terminal domain

The structure of the soaked crystal is isomorphous with that of the native domain. On the other hand, the co-crystal of the C-terminal domain of hKIN17 with Na<sub>2</sub>WO<sub>4</sub> changes its cell parameters from *a*=45.75 Å and *b*=46.31 Å in the native crystal to *a*=47.67 Å and *b*=48.72 Å, with concomitant



**Figure 3.** RNA-binding activity of deletions of the hKIN17 protein. (a) A representation of the full-length and deleted hKIN17 proteins. The deleted protein (1) corresponds to the N-terminal tail of hKIN17 and the deleted protein (3) corresponds to the same region without the zinc finger. The deleted protein (2) corresponds to the C-terminal domain of hKIN17. Residues from 150 to 270 of the hKIN17 sequence correspond to a flexible region, without any predicted folded domain. (b) The full-length KIN17 protein (FI) and the deleted proteins were detected by staining with Ponceau red on a nitrocellulose membrane to estimate the amount of input protein. North-western blotting was carried out with the same membrane incubated with a  $^{32}\text{P}$ -labelled RNA probe in 0.05 M NaCl. M corresponds to a pre-stained molecular mass marker. The zinc finger and the C-terminal domains are the hKIN17 RNA-binding domains.

differences in crystal packing. To characterise these changes, we compare the models obtained from the crystal soaked in  $\text{Na}_2\text{WO}_4$  and the  $\text{Na}_2\text{WO}_4$  co-crystal (Table 2). As shown in Figure 5(a), the hKIN17 C-terminal domain exhibits conformational changes in loops, in particular in the second subdomain, as well as a slight shift in the orientation of the two  $\text{SH}_3$ -like subdomains relative to each other. However, the A–B subdomains interface is mostly unchanged.

#### The C-terminal domain of hKIN17 binds RNA through a surface comprising the highly conserved lysine residues 302 and 391

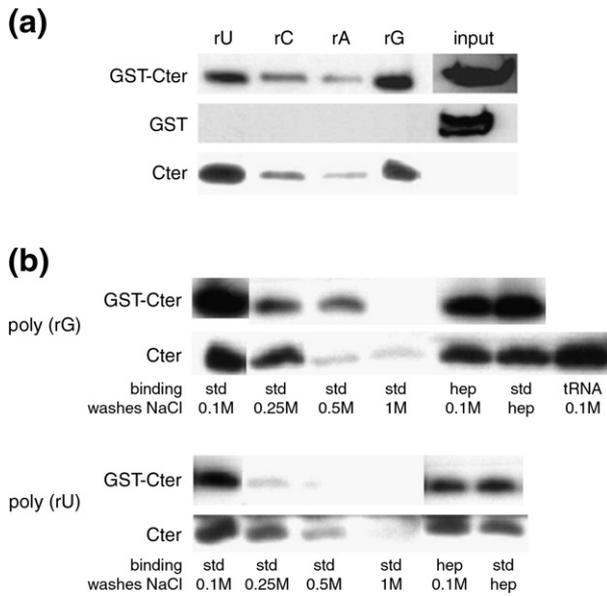
The tungstate ions interact with several lysine residues from the hKIN17 C-terminal domain, two of which, K302 and K391, are strictly conserved among KIN17 sequences (Figure 1(a)). To confirm that K302 and K391 are also involved in hKIN17 binding to RNA, two mutants were produced where these lysine residues were changed into negatively charged residues: K302E and K391E. Binding of the mutants to RNA was assayed by two independent methods. A binding assay using affinity chromatography to ribonucleotide homopolymers immobilized on agarose beads showed that both mutants K302E and K391E exhibited a reduced RNA-binding ability compared to the wild-type (Figure 6). K302E mutation abolishes the poly(rC) and poly(rA)-

binding activity and clearly reduces the poly(rU) and poly(rG)-binding activity of the C-terminal domain of hKIN17. The K391E mutant has a reduced RNA-binding activity, in particular concerning its poly(rU) binding activity. These results were further confirmed by performing a North-western analysis using  $^{32}\text{P}$  end-labelled poly(G) and poly(U) RNAs. This experiment showed that the K302E and K391E mutations abolish the poly(rU)-binding activity and reduce the poly(rG)-binding activity of the C-terminal domain of protein hKIN17 (Supplementary Data). NMR spectroscopy of the mutant proteins shows that the charge reversal mutants have a three-dimensional structure similar to that of the native protein (data not shown), indicating that the surface mutations do not affect their global folding or structural integrity. Thus, we conclude that the two highly conserved lysine residues K302 and K391 are implicated in RNA binding of the C-terminal domain of hKIN17.

## Discussion

#### Contribution of the hKIN17 C-terminal domain to RNA binding

A KOW motif was identified from sequence analysis in the C-terminal domain of hKIN17.



**Figure 4.** RNA-binding properties of the hKIN17 C-terminal domain. (a) Purified recombinant hKIN17 C-terminal domain (Cter) or in fusion with a GST-tag and purified GST protein as a negative control were incubated with each ribonucleotide homopolymers and washed with 0.1M NaCl. The proteins retained by the immobilized homopolymer RNAs were analysed by immunoblotting. hKIN17 C-terminal domain shows a clear preference for poly(rG) and poly(rU) homopolymers. (b) Characterization of the poly(rG) and poly(rU) RNA interaction with the hKIN17 C-terminal domain alone or in fusion with a GST-tag. Purified recombinant proteins were mixed with immobilized poly(rG) and poly(rU) under various binding conditions, as indicated. Std corresponds to the following binding buffer: 20 mM Tris-HCl (pH 7.4), 2.5 mM MgCl<sub>2</sub>, 0.5% (v/v) Triton X-100, 100 mM NaCl. The retained proteins were washed under the conditions described for each lane. Hep corresponds to incubation with 1 mg/ml of heparin. The hKIN17 C-terminal domain binds poly(rG) and poly(rU) in salt-resistant and heparin-resistant conditions.

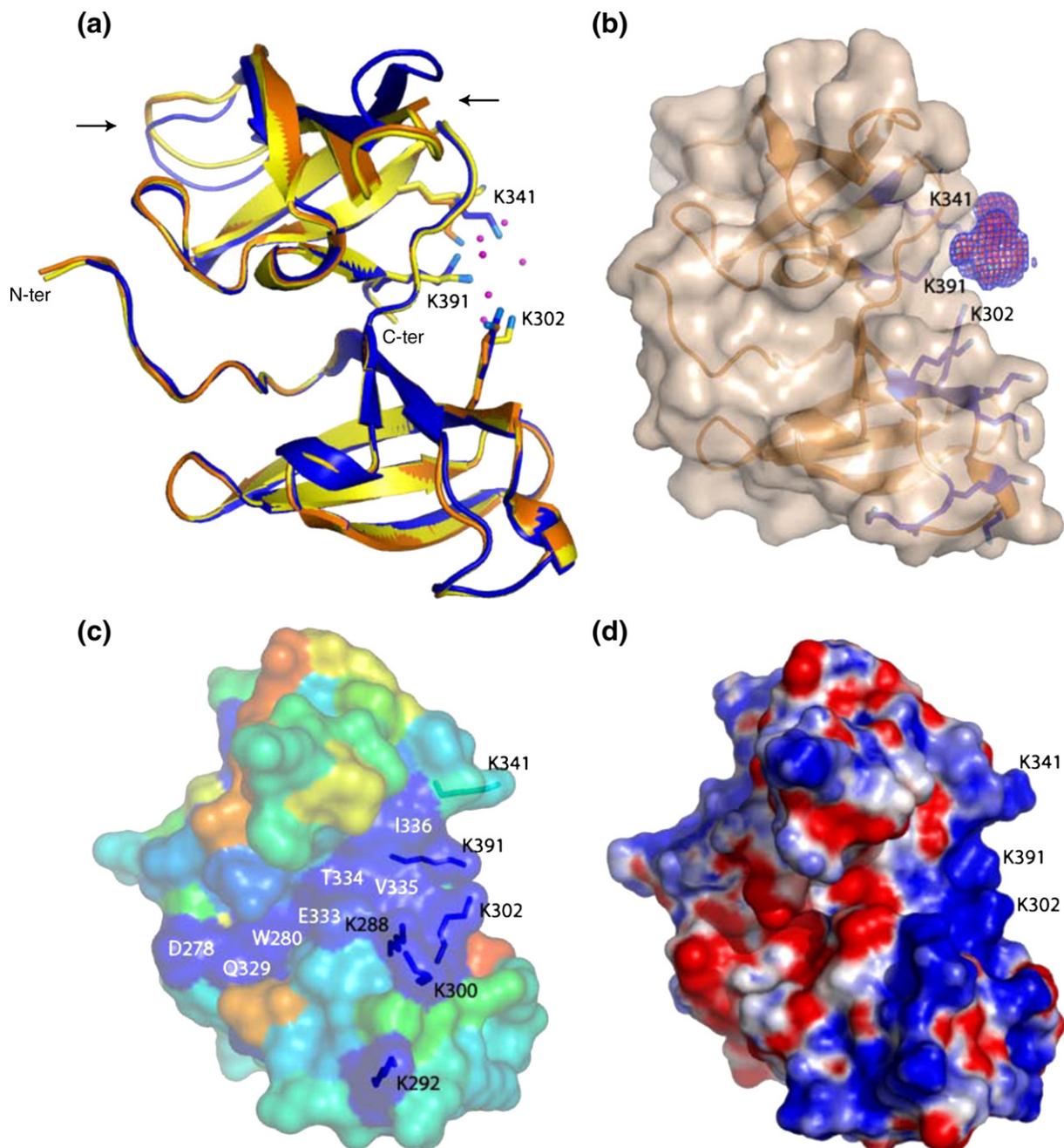
From the known structure of the large ribosomal subunit,<sup>19</sup> the KOW sequence motif occurs within an SH<sub>3</sub>-like fold. It was thus predicted that the hKIN17 C-terminal domain should exhibit an SH<sub>3</sub>-like fold. Surprisingly, it contains two structurally similar SH<sub>3</sub>-like folds. The predicted KOW sequence element begins in the loop before  $\beta_{1B}$  strand and finishes in the loop  $\beta_{2B}\beta_{3B}$ . Loop  $\beta_{1\beta 2}$  is four residues larger in subdomain A compared to subdomain B, and this insertion prevents the detection of a KOW motif in subdomain A, even if the canonical conserved glycine exists at position 348. On the contrary, loops  $\beta_{2\beta 3}$  and  $\beta_{3\beta 4}$  are larger in subdomain B. Our RNA interaction studies strengthen the notion that KOW sequence containing domains are involved in RNA binding.

Tungstate derivatives were prepared either by soaking preformed crystals in a solution of Na<sub>2</sub>WO<sub>4</sub> or by co-crystallising in the presence of Na<sub>2</sub>WO<sub>4</sub>.

Methodologically, the two approaches are complementary. In soaked crystals, structural rearrangements can be inhibited by the rigidity of the crystal lattice, while during co-crystallization, artefactual sites can be created at ligand-mediated crystal contacts. In our soaking experiments, we observed no cracking or dissolution of the crystals and a protein conformation identical with that of the free form of KIN17, while the co-crystals showed that conformational changes at the interface between the two subdomains and in loops  $\beta_{2\beta 3}$  and  $\beta_{3\beta 4}$  occurred as the complex was formed in solution. All these conformational changes that occurred in derivatives were not of sufficient amplitude to overcome the crystal lattice constraints. In the co-crystals, a larger domain rearrangement was possible and unrestrained by crystal interactions, and the loops could adopt a more appropriate conformation. These rearrangements are likely to originate at the tungsten-interacting lysine residues and are amplified at longer distances, culminating in loops  $\beta_{2\beta 3}$  and  $\beta_{3\beta 4}$  of subdomain B.

Conservatively, we considered meaningful only the tungstate-binding sites common to soaked crystals and co-crystals; namely, K302, K341 and K391. Identification of these lysine residues involved in the interaction with tungstate, a phosphate analog, allowed us to locate putative binding sites of the hKIN17 C-terminal domain with the RNA phosphoribose backbone. Mutagenesis studies confirmed the role of K302 and K391 in RNA binding. This is quite typical, as lysine and arginine residues are often involved in binding to the phosphate backbone of nucleic acids.<sup>30</sup> We suggest that the flexibility observed after tungstate binding might allow the protein to adapt to its RNA target.

The RNA-binding activity of hKIN17 seems to be distributed over several domains in the whole protein. At the N terminus, the C<sub>2</sub>H<sub>2</sub> zinc-finger domain shows a dual RNA/DNA-binding function *in vitro*, as reported recently for other zinc-finger motifs.<sup>37</sup> At the C terminus, the tandem SH<sub>3</sub>-like domain binds only RNA *in vitro*. Our methodology that uses tungstate, an electron-dense phosphate mimic, is effective in delimiting in the crystal structure the volume occupied by the RNA phosphate backbone in those stretches where protein-phosphate interactions are dominant, but is unable to define any sequence specificity. The nucleotide specificity suggested by the differential binding of the various RNA homopolymers could arise from base-protein interactions or a preference for a particular phosphate backbone conformation found in its biological RNA target. Given that the hKIN17 C-terminal domain has the same preferences for poly(rG) and poly(rU) as the whole hKIN17 protein, it is likely that this region of the molecule is a major determinant of the *in vivo* RNA binding of the whole hKIN17. From the structural study of several protein-RNA complexes, it was shown that the use of multiple domains allows extended



**Figure 5.** Localisation of the tungstate binding sites in the hKIN17 C-terminal domain. (a) Superimposition on subdomains A of the C-terminal domain of hKIN17 from native (orange),  $\text{Na}_2\text{WO}_4$ -soaked crystal (yellow) and from the co-crystal (blue). Lysine residues involved in tungstate binding are shown in the corresponding backbone colour. Tungstate anions from the soaked crystal are shown in magenta. Quantification of the conformation changes within the subdomains (in particular, in loops  $\beta 2\beta 3$  and  $\beta 3\beta 4$  of subdomain B, as indicated by arrows) and of the different positioning of subdomain B is reported in Table 2. (b) Ribbon and surface representation of the hKIN17 C-terminal domain structure. The omit map was calculated using CNS<sup>58</sup> around the tungstate anions cluster with data collected for the co-crystal. Key lysine residues surrounding the tungstate anions are labelled. (c) A projection of the CONSURF results on the surface of C-terminal domain of hKIN17 indicates the degree of conservation from dark blue (highly conserved) to red (variable). The orientation of the structure has been rotated slightly relative to (b), in order to distinguish the five lysine residues in the positively charged groove. The side-chains of these lysine residues putatively involved in RNA binding are displayed. (d) Surface representation of the C-terminal domain of hKIN17 coloured by electrostatic potential and shown in the same orientation as (c). Blue and red colours indicate positive and negative electrostatic surface potential, respectively.

interactions between proteins and target RNA sequences, thereby contributing to increase binding affinity and specificity. RNA-binding domains

with limited specificity at the individual level can achieve high functional selectivity *via* cooperativity between their modular domains or *via* the

**Table 2.** Superimposition of the two subdomains of the hKIN17 C-terminal domain

	Native versus soaked crystal	Native versus co-crystal
Superimposition on subdomains A	rmsd <sub>subdomains A</sub> 0.083	rmsd <sub>subdomains A</sub> 0.424
	rmsd <sub>subdomains B</sub> 0.206	rmsd <sub>subdomains B</sub> 1.801
Superimposition on subdomains B	rmsd <sub>subdomains A</sub> 0.212	rmsd <sub>subdomains A</sub> 3.174
	rmsd <sub>subdomains B</sub> 0.115	rmsd <sub>subdomains B</sub> 0.781

The rmsd values are calculated on 51 C $\alpha$  atoms for subdomains A (spanning residues 284–334 of the hKIN17 sequence) and 52 C $\alpha$  atoms for subdomains B (spanning residues 340–391 of the hKIN17 sequence).

formation of complexes.<sup>38</sup> In the case of hKIN17, the tandem SH<sub>3</sub>-like domain could cooperate with the N-terminal zinc-finger domain to associate optimally with specific RNAs.

#### A highly conserved groove, comprising K302 and K391 and other positively charged residues, is characteristic of the KIN17 family

Analysis of phylogenetic conservation of the C-terminal domain of hKIN17 using PSI-blast revealed that this domain is conserved in KIN17 proteins from metazoans, plants and parasites. Apart from the conserved buried amino acid residues that form part of the protein core, the CONSURF results projected onto the three-dimensional structure of hKIN17 C-terminal domain reveals conserved residues on the surface of the protein that could be structural determinants for interactions with other biomolecules.

Five lysine residues (K288, K292, K300, K302 and K391) are part of a highly conserved and positively charged surface (Figure 5(c)), surrounded by two interfacial exposed and conserved hydrophobic residues (V335 and I336). This surface is essentially located on the  $\beta_1\beta_2$  hairpin of the first SH<sub>3</sub>-like domain; it also comprises two residues of the linker and one residue at the C terminus of the second SH<sub>3</sub>-like domain. On the basis of the surface charge distribution of RNA, we proposed that this positively charged groove may act as an RNA-binding region. This is in agreement with the conclusion that K302 and K391 are involved in RNA binding. A similar situation, where lysine residues from different subdomains located on the groove borders are involved in RNA binding, is found for the *Bacillus stearothermophilus* ribosomal protein L2. In the double- $\beta$ -barrel protein, one of the subdomains adopts an SH<sub>3</sub>-like fold and the other adopts an OB-fold.<sup>39</sup> The two subdomains are connected by a four-residue  $_310$  helix and are packed tightly approximately perpendicular one to the other. On the basis of mutagenesis data, the authors proposed that protein L2 has a large contact surface for the target RNA in the cleft formed by the interface region of the two subdomains. This RNA-binding

surface of the ribosomal protein L2 comprises positively charged and hydrophobic residues of both subdomains.

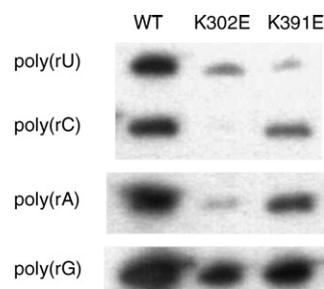
At the interface between the two SH<sub>3</sub>-like domains of hKIN17 and next to the positively charged surface, another set of highly conserved residues (D278, E333 and E387) forms a mainly negatively charged cleft, and three other conserved residues (W280, Q329 and T334) form the bottom of the cleft. These residues may constitute a surface for other interactions.

The search for analogs of the C-terminal domain of hKIN17 using PSI-blast highlighted significant sequence similarity with the C-terminal region of the metazoan and plant transcription elongation factors SPT5, belonging to the DSIF complex, which exerts both positive and negative effects on RNA polymerase II.<sup>40,41</sup> In addition, the search identified the poorly studied T54 proteins from metazoa and plants. These sequences are not as well conserved as the KIN17 sequence, in particular concerning the positively charged groove.

#### Functional significance of the tandem SH<sub>3</sub>-like domain

In general, the interaction of the ribosomal SH<sub>3</sub>-like domains with RNA is mediated by the variable loops that connect the  $\beta$ -strands.<sup>42</sup> Variability in the length of the three first loops ( $\beta_1\beta_2$ ,  $\beta_2\beta_3$ ,  $\beta_3\beta_4$ ) is a characteristic of SH<sub>3</sub>-like domains and is mainly responsible for the high level of functional versatility of this fold. Consistently, a DALI search using the structures of the hKIN17 SH<sub>3</sub>-like subdomains revealed that these are structurally highly similar to SH<sub>3</sub>-like domains, exhibiting various functions, from signal transduction (tyrosine kinase) to nucleic acid binding (ribosomal protein L21e).

Only a few three-dimensional structures of tandem SH<sub>3</sub>-like domains have been determined. The tandem Tudor domain of human 53BP1,<sup>43,44</sup> the tandem Tudor domain of FMRP,<sup>45</sup> the double chromodomains of human CHD1<sup>46</sup> and the double



**Figure 6.** K302E and K391E mutants have a reduced RNA binding activity. The purified wild-type hKIN17 C-terminal domain and the K302E and K391E mutant domains, in fusion with GST, were incubated with each ribonucleotide homopolymers and washed with 0.1 M NaCl. The proteins retained by the immobilized homopolymer RNAs were analysed by immunoblotting. Both mutants show a reduced binding activity for the four homopolymers.

Tudor domain of JMJD2A<sup>47</sup> were all reported to bind to methylated peptides through a conserved aromatic cage. However, hKIN17 SH<sub>3</sub>-like subdomains do not exhibit any solvent-exposed aromatic clusters, suggesting that they do not target similar methylated peptides. Furthermore, the association of a region of 53BP1, containing the tandem Tudor domains, with IR-induced foci is sensitive to treatment with ribonuclease, suggesting that another function of double Tudor of 53BP1 is to bind RNA.<sup>48</sup> Thus, RNA binding could represent a more general function of tandem SH<sub>3</sub>-like domains. In this study, a novel RNA-binding site in a double SH<sub>3</sub>-like domain is characterized even if it does not exclude the possibility that this domain is involved also in protein-protein interaction.

## Conclusion

RNA-binding proteins are involved in a wide range of cellular processes in the nucleus and in the cytoplasm, including regulation of pre-mRNA splicing, mRNA stability, translation efficiency and transport of RNAs between the nucleus and the cytoplasm. An RNA-binding function for hKIN17 is in accordance with the identification of hKIN17 as a component of the human spliceosome.<sup>14,15</sup> Furthermore, a Tudor-like domain structurally similar to domain III of NusG and to the subdomains of the hKIN17 C-terminal domain was found recently in the bacterial transcription-repair coupling factor (TRCF), where it mediates interaction with the  $\beta$  subunit of RNA polymerase.<sup>49</sup> Consistently, the C-terminal domain of hKIN17 as the whole protein co-immunoprecipitate with the RNA polymerase II in HeLa nuclear extracts (G.P.-L., unpublished results). Further investigation will be required to identify which cellular RNA interacts with hKIN17 and to understand how hKIN17 functions in RNA metabolism.

## Materials and Methods

### Cloning, expression and purification of the domains of hKIN17 protein

The C-terminal domain comprising residues 268–393 of hKIN17 protein was prepared as described.<sup>22</sup> Mutant proteins (K302E and K391E) were engineered by site-directed mutagenesis using QuikChange Site-Directed Mutagenesis Kit (Stratagene) and expressed and purified as the wild-type protein. The region comprising residues 1–160 of hKIN17 protein and including the zinc finger (residues 23–50) was prepared as described.<sup>24</sup> The region comprising residues 51–160 of hKIN17 protein was prepared following the protocol used for the C-terminal domain.<sup>22</sup>

### Crystallization, X-ray data collection and refinement

Crystallization and initial phasing of the native structure have been described.<sup>22</sup> Briefly, a halide heavy-atom

derivative was prepared by brief-soaking of a native crystal in a solution of KI. SHARP was run to calculate single isomorphous replacement with anomalous scattering phases at 1.45 Å resolution.<sup>50</sup> After solvent flattening with SOLOMON,<sup>51</sup> the polypeptide chain was traced using a combination of ARP/wARP<sup>52</sup> and manual rebuilding. Model refinement was carried out with REFMAC5.<sup>53</sup>

Tungstate derivatives were prepared by both soaking and co-crystallization techniques as described.<sup>22</sup> Native crystals were soaked overnight in the crystallisation solution (100 mM sodium acetate (pH 6.3), 10% (v/v) ethylene glycol, 27% (w/v) PEG 6000, 0.5 M LiCl) with 1 mM Na<sub>2</sub>WO<sub>4</sub> added. The hKIN17 C-terminal domain was co-crystallized with tungstate in a different crystallisation solution (100 mM sodium acetate (pH 6.3), 25% ethylene glycol, 20% PEG 6000) with 0.5 mM Na<sub>2</sub>WO<sub>4</sub> added and co-crystals were then transferred to the cryoprotectant solution (100 mM sodium acetate (pH 6.3), 25% ethylene glycol, 20% PEG 6000) with 100 mM Na<sub>2</sub>WO<sub>4</sub> added.

Data were collected using synchrotron radiation at a wavelength of 1.2144 Å, an energy where tungsten scatters X-rays with an anomalous contribution ( $f''$ ) greater than ten electrons. Molecular replacement was carried out using Molrep and refinement using REFMAC5<sup>53</sup> with data collected on a W-soaked crystal. Strong electron density peaks were identified in an anomalous difference Fourier map calculated to locate the tungsten atoms. With the data collected on a co-crystal, AutoSHARP<sup>54</sup> was used to calculate single wavelength anomalous diffraction (SAD) phases. It was found to be crucial to reduce the default value for the minimum allowed distance between sites for heavy-atom detection in SHELXD as well as for analysis of residual maps in AutoSHARP, since the tungstate anions are best represented as a cluster. Model refinement was then carried out with REFMAC5.<sup>53</sup>

### North-western blot analysis

The 1200 bp RNA probe corresponding to *Kin17* ORF was prepared as described.<sup>16</sup> To allow renaturation of the recombinant proteins after they were separated by SDS-PAGE and transferred onto nitrocellulose membranes, the protein blots were treated for 1–2 h at room temperature in a hybridization buffer (10 mM Tris-HCl (pH 7.4), 50 mM NaCl, 1X Denhardt's solution, 1 mM EDTA) in the presence of 1  $\mu$ g/ml of *E. coli* tRNA (Sigma). The blots were then probed at room temperature for 1 h with radiolabelled RNA probe (100,000 cpm for a 2 ml reaction volume) in a hybridization buffer containing 2  $\mu$ g/ml of *E. coli* tRNA. Blots were washed twice for 15 min in the hybridization buffer and exposed to X-ray film or to a PhosphorImager screen.

### RNA homopolymer binding assay

RNA homopolymer-conjugated agarose beads (poly(rC), poly(rG), poly(rU)) and conjugated Sepharose beads (poly(rA)) were obtained from Sigma. Binding of the purified recombinant protein to RNA homopolymers was performed as described.<sup>16</sup> A portion (20  $\mu$ l) of each RNA homopolymer bead was incubated with an equal amount of purified protein (3  $\mu$ g), for 30 min at 4 °C with gentle agitation in a total volume of 500  $\mu$ l of binding buffer (20 mM Tris-HCl (pH 7.4), 2.5 mM MgCl<sub>2</sub>, 0.5% (v/v) Triton X-100 and 100 mM, 250 mM, 500 mM or 1 M NaCl). After incubation, the beads were washed five times with 500  $\mu$ l of binding buffer. Some binding reactions or washes

were performed in the presence of 1 mg/ml of heparin or 1 mg/ml of *E. coli* tRNA. Proteins bound to the RNA homopolymer beads were eluted by boiling in SDS loading buffer and loaded onto an SDS/15% (w/v) polyacrylamide electrophoresis gel. After electrophoretic separation, proteins were transferred onto a nitrocellulose filter (Schleicher and Schuell) and detected with anti-GST HRP conjugate antibody (Amersham Biosciences) or anti-KIN17 mAb Igk36 antibody,<sup>55</sup> depending on the presence or the absence of a GST tag fused to the proteins.

### NMR spectroscopy studies

NMR samples (about 1 mM) were prepared in PBS (pH 7.0) containing 10% <sup>2</sup>H<sub>2</sub>O, 1 mM EDTA, a protease inhibitor cocktail (Sigma), 1 mM NaN<sub>3</sub> and 3-(trimethylsilyl) [2,2,3,3-<sup>2</sup>H<sub>4</sub>] propionate (TSP) was added as a chemical shift reference. The 1D <sup>1</sup>H, 2D <sup>1</sup>H-<sup>1</sup>H nuclear Overhauser effect spectroscopy (NOESY) and 2D <sup>1</sup>H-<sup>1</sup>H total correlated spectroscopy (TOCSY) experiments were carried out at 297 K on a Bruker DRX-500 spectrometer. All spectra were processed and compared with the program Xwinnmr (Bruker).

### Sequence analysis

The amino acid sequences comparisons were optimized using the program MUSCLE,<sup>56</sup> and were submitted to CONSURF,<sup>57</sup> in order to calculate for each residue a conservation score based on residue similarity and phylogenetic distance.

### Protein Data Bank accession number

The coordinates and structure factors of the X-ray structure have been deposited with the RCSB Protein Data Bank under accession number 2CKK.

## Acknowledgements

We thank C. Vonrhein for assistance with auto-SHARP and for helpful discussions. We thank JB. Charbonnier for critical review of the manuscript. We thank B. Gilquin for helpful discussions and MH. Le Du for assistance with crystallography. The laboratory of JF. Angulo work was supported by contract N° 8702 of Electricité de France.

## Supplementary Data

Supplementary data associated with this article can be found, in the online version, at [doi:10.1016/j.jmb.2006.09.033](https://doi.org/10.1016/j.jmb.2006.09.033)

## References

1. Biard, D. S., Saintigny, Y., Maratrat, M., Paris, F., Martin, M. & Angulo, J. F. (1997). Enhanced expres-

- sion of the Kin17 protein immediately after low doses of ionizing radiation. *Radiat. Res.* **147**, 442–450.
2. Masson, C., Menaa, F., Pinon-Lataillade, G., Frobert, Y., Chevillard, S., Radicella, J. P. *et al.* (2003). Global genome repair is required to activate KIN17, a UVC-responsive gene involved in DNA replication. *Proc. Natl Acad. Sci. USA*, **100**, 616–621.
3. Kannouche, P., Mauffrey, P., Pinon-Lataillade, G., Mattei, M. G., Sarasin, A., Daya-Grosjean, L. & Angulo, J. F. (2000). Molecular cloning and characterization of the human KIN17 cDNA encoding a component of the UVC response that is conserved among metazoans. *Carcinogenesis*, **21**, 1701–1710.
4. Biard, D. S., Kannouche, P., Lannuzel-Drogou, C., Mauffrey, P., Apiou, F. & Angulo, J. F. (1999). Ectopic expression of (Mm)Kin17 protein inhibits cell proliferation of human tumor-derived cells. *Expt. Cell Res.* **250**, 499–509.
5. Kannouche, P. & Angulo, J. F. (1999). Overexpression of kin17 protein disrupts nuclear morphology and inhibits the growth of mammalian cells. *J. Cell Sci.* **112**, 3215–3224.
6. Winzeler, E. A., Shoemaker, D. D., Astromoff, A., Liang, H., Anderson, K., Andre, B. *et al.* (1999). Functional characterization of the *S. cerevisiae* genome by gene deletion and parallel analysis. *Science*, **285**, 901–906.
7. Biard, D. S., Despras, E., Sarasin, A. & Angulo, J. F. (2005). Development of new EBV-based vectors for stable expression of small interfering RNA to mimic human syndromes: application to NER gene silencing. *Mol. Cancer Res.* **3**, 519–529.
8. Biard, D. S., Miccoli, L., Despras, E., Harper, F., Pichard, E., Creminon, C. & Angulo, J. F. (2003). Participation of kin17 protein in replication factories and in other DNA transactions mediated by high molecular weight nuclear complexes. *Mol. Cancer Res.* **1**, 519–531.
9. Miccoli, L., Frouin, I., Novac, O., Di Paola, D., Harper, F., Zannis-Hadjopoulos, M. *et al.* (2005). The human stress-activated protein kin17 belongs to the multi-protein DNA replication complex and associates in vivo with mammalian replication origins. *Mol. Cell Biol.* **25**, 3814–3830.
10. Mazin, A., Milot, E., Devoret, R. & Chartrand, P. (1994). KIN17, a mouse nuclear protein, binds to bent DNA fragments that are found at illegitimate recombination junctions in mammalian cells. *Mol. Gen. Genet.* **244**, 435–438.
11. Mazin, A., Timchenko, T., Menissier-de Murcia, J., Schreiber, V., Angulo, J. F., de Murcia, G. & Devoret, R. (1994). Kin17, a mouse nuclear zinc finger protein that binds preferentially to curved DNA. *Nucl. Acids Res.* **22**, 4335–4341.
12. Milot, E., Belmaaza, A., Wallenburg, J. C., Gusew, N., Bradley, W. E. & Chartrand, P. (1992). Chromosomal illegitimate recombination in mammalian cells is associated with intrinsically bent DNA elements. *EMBO J.* **11**, 5063–5070.
13. Timchenko, T., Bailone, A. & Devoret, R. (1996). Btcd, a mouse protein that binds to curved DNA, can substitute in *Escherichia coli* for H-NS, a bacterial nucleoid protein. *EMBO J.* **15**, 3986–3992.
14. Makarov, E. M., Makarova, O. V., Urlaub, H., Gentzel, M., Will, C. L., Wilm, M. & Luhrmann, R. (2002). Small nuclear ribonucleoprotein remodeling during catalytic activation of the spliceosome. *Science*, **298**, 2205–2208.
15. Rappsilber, J., Ryder, U., Lamond, A. I. & Mann, M.

- (2002). Large-scale proteomic analysis of the human spliceosome. *Genome Res.* **12**, 1231–1245.
16. Pinon-Lataillade, G., Masson, C., Bernardino-Sgherri, J., Henriot, V., Mauffrey, P., Frobert, Y. *et al.* (2004). KIN17 encodes an RNA-binding protein and is expressed during mouse spermatogenesis. *J. Cell Sci.* **117**, 3691–3702.
17. Kyrpides, N. C., Woese, C. R. & Ouzounis, C. A. (1996). KOW: a novel motif linking a bacterial transcription factor with ribosomal proteins. *Trends Biochem. Sci.* **21**, 425–426.
18. Ponting, C. P. (2002). Novel domains and orthologues of eukaryotic transcription elongation factors. *Nucl. Acids Res.* **30**, 3643–3652.
19. Ban, N., Nissen, P., Hansen, J., Moore, P. B. & Steitz, T. A. (2000). The complete atomic structure of the large ribosomal subunit at 2.4 Å resolution. *Science*, **289**, 905–920.
20. Steiner, T., Kaiser, J. T., Marinkovic, S., Huber, R. & Wahl, M. C. (2002). Crystal structures of transcription factor NusG in light of its nucleic acid- and protein-binding activities. *EMBO J.* **21**, 4641–4653.
21. Gaboriaud, C., Bissery, V., Benchetrit, T. & Mornon, J. P. (1987). Hydrophobic cluster analysis: an efficient new way to compare and analyse amino acid sequences. *FEBS Letters*, **224**, 149–155.
22. le Maire, A., Schiltz, M., Braud, S., Gondry, M., Charbonnier, J. B., Zinn-Justin, S. & Stura, E. (2006). Crystallization and halide phasing of the C-terminal domain of human KIN17. *Acta Crystallog. sect. F*, **62**, 245–248.
23. Lo Conte, L., Chothia, C. & Janin, J. (1999). The atomic structure of protein-protein recognition sites. *J. Mol. Biol.* **285**, 2177–2198.
24. Carlier, L., le Maire, A., Braud, S., Masson, C., Gondry, M., Zinn-Justin, S. *et al.* (2006). NMR assignment of region 51–160 of human KIN17, a DNA and RNA-binding protein. *J. Biomol. NMR*. (Electronic publication).
25. Adinolfi, S., Bagni, C., Musco, G., Gibson, T., Mazzarella, L. & Pastore, A. (1999). Dissecting FMR1, the protein responsible for fragile X syndrome, in its structural and functional domains. *Rna*, **5**, 1248–1258.
26. Bertrand, S., Burlet, P., Clermont, O., Huber, C., Fondrat, C., Thierry-Mieg, D. *et al.* (1999). The RNA-binding properties of SMN: deletion analysis of the zebrafish orthologue defines domains conserved in evolution. *Hum. Mol. Genet.* **8**, 775–782.
27. Siomi, H., Siomi, M. C., Nussbaum, R. L. & Dreyfuss, G. (1993). The protein product of the fragile X gene, FMR1, has characteristics of an RNA-binding protein. *Cell*, **74**, 291–298.
28. Pinol-Roma, S., Choi, Y. D., Matunis, M. J. & Dreyfuss, G. (1988). Immunopurification of heterogeneous nuclear ribonucleoprotein particles reveals an assortment of RNA-binding proteins. *Genes Dev.* **2**, 215–227.
29. Swanson, M. S. & Dreyfuss, G. (1988). Classification and purification of proteins of heterogeneous nuclear ribonucleoprotein particles by RNA-binding specificities. *Mol. Cell Biol.* **8**, 2237–2241.
30. Jones, S., Daley, D. T., Luscombe, N. M., Berman, H. M. & Thornton, J. M. (2001). Protein-RNA interactions: a structural analysis. *Nucl. Acids Res.* **29**, 943–954.
31. Egloff, M. P., Cohen, P. T., Reinemer, P. & Barford, D. (1995). Crystal structure of the catalytic subunit of human protein phosphatase 1 and its complex with tungstate. *J. Mol. Biol.* **254**, 942–959.
32. Barford, D., Flint, A. J. & Tonks, N. K. (1994). Crystal structure of human protein tyrosine phosphatase 1B. *Science*, **263**, 1397–1404.
33. Davies, D. R., Interthal, H., Champoux, J. J. & Hol, W. G. (2002). Insights into substrate binding and catalytic mechanism of human tyrosyl-DNA phosphodiesterase (Tdp1) from vanadate and tungstate-inhibited structures. *J. Mol. Biol.* **324**, 917–932.
34. Gray, C. H., Good, V. M., Tonks, N. K. & Barford, D. (2003). The structure of the cell cycle protein Cdc14 reveals a proline-directed protein phosphatase. *EMBO J.* **22**, 3524–3535.
35. Longenecker, K. L., Roach, P. J. & Hurley, T. D. (1996). Three-dimensional structure of mammalian casein kinase I: molecular basis for phosphate recognition. *J. Mol. Biol.* **257**, 618–631.
36. Stuckey, J. A., Schubert, H. L., Fauman, E. B., Zhang, Z. Y., Dixon, J. E. & Saper, M. A. (1994). Crystal structure of Yersinia protein tyrosine phosphatase at 2.5 Å and the complex with tungstate. *Nature*, **370**, 571–575.
37. Brown, R. S. (2005). Zinc finger proteins: getting a grip on RNA. *Curr. Opin. Struct. Biol.* **15**, 94–98.
38. Singh, R. & Valcarcel, J. (2005). Building specificity with nonspecific RNA-binding proteins. *Nature Struct. Mol. Biol.* **12**, 645–653.
39. Nakagawa, A., Nakashima, T., Taniguchi, M., Hosaka, H., Kimura, M. & Tanaka, I. (1999). The three-dimensional structure of the RNA-binding domain of ribosomal protein L2; a protein at the peptidyl transferase center of the ribosome. *EMBO J.* **18**, 1459–1467.
40. Yamaguchi, Y., Narita, T., Inukai, N., Wada, T. & Handa, H. (2001). SPT genes: key players in the regulation of transcription, chromatin structure and other cellular processes. *J. Biochem. (Tokyo)*, **129**, 185–191.
41. Zorio, D. A. & Bentley, D. L. (2001). Transcription elongation: the ‘Foggy’ is lifting ellipsis. *Curr. Biol.* **11**, R144–R146.
42. Klein, D. J., Moore, P. B. & Steitz, T. A. (2004). The roles of ribosomal proteins in the structure assembly, and evolution of the large ribosomal subunit. *J. Mol. Biol.* **340**, 141–177.
43. Charier, G., Couprie, J., Alpha-Bazin, B., Meyer, V., Quemeneur, E., Guerois, R. *et al.* (2004). The Tudor tandem of 53BP1: a new structural motif involved in DNA and RG-rich peptide binding. *Structure*, **12**, 1551–1562.
44. Huyen, Y., Zgheib, O., Ditullio, R. A., Jr, Gorgoulis, V. G., Zacharatos, P., Petty, T. J. *et al.* (2004). Methylated lysine 79 of histone H3 targets 53BP1 to DNA double-strand breaks. *Nature*, **432**, 406–411.
45. Ramos, A., Hollingworth, D., Adinolfi, S., Castets, M., Kelly, G., Frenkiel, T. A. *et al.* (2006). The structure of the N-terminal domain of the fragile X mental retardation protein: a platform for protein-protein interaction. *Structure*, **14**, 21–31.
46. Flanagan, J. F., Mi, L. Z., Chruszcz, M., Cymborowski, M., Clines, K. L., Kim, Y. *et al.* (2005). Double chromodomains cooperate to recognize the methylated histone H3 tail. *Nature*, **438**, 1181–1185.
47. Huang, Y., Fang, J., Bedford, M. T., Zhang, Y. & Xu, R. M. (2006). Recognition of histone H3 lysine-4 methylation by the double Tudor domain of JMJD2A. *Science*, **312**, 748–751.
48. Pryde, F., Khalili, S., Robertson, K., Selfridge, J., Ritchie, A. M., Melton, D. W. *et al.* (2005). 53BP1 exchanges slowly at the sites of DNA damage and appears to require RNA for its association with chromatin. *J. Cell Sci.* **118**, 2043–2055.

49. Deaconescu, A. M., Chambers, A. L., Smith, A. J., Nickels, B. E., Hochschild, A., Savery, N. J. & Darst, S. A. (2006). Structural basis for bacterial transcription-coupled DNA repair. *Cell*, **124**, 507–520.
50. Bricogne, G., Vornrhein, C., Flensburg, C., Schiltz, M. & Paciorek, W. (2003). Generation, representation and flow of phase information in structure determination: recent developments in and around SHARP 2.0. *Acta Crystallog. sect. D*, **59**, 2023–2030.
51. Abrahams, J. P. & Leslie, A. G. (1996). Methods used in the structure determination of bovine mitochondrial F1 ATPase. *Acta Crystallog. sect. D*, **52**, 30–42.
52. Perrakis, A., Morris, R. & Lamzin, V. S. (1999). Automated protein model building combined with iterative structure refinement. *Nature Struct. Biol.* **6**, 458–463.
53. Murshudov, G. N., Vagin, A. A. & Dodson, E. J. (1997). Refinement of macromolecular structures by the maximum-likelihood method. *Acta Crystallog. sect. D*, **53**, 240–255.
54. Vornrhein, C., Blanc, E., Roversi, P. & Bricogne, G. (2005). Automated structure solution with auto-SHARP. In *Crystallographic Methods* (Doublé, S., ed), Humana Press, Totowa, NJ, USA.
55. Miccoli, L., Biard, D. S., Creminon, C. & Angulo, J. F. (2002). Human kin17 protein directly interacts with the simian virus 40 large T antigen and inhibits DNA replication. *Cancer Res.* **62**, 5425–5435.
56. Edgar, R. C. (2004). MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucl. Acids Res.* **32**, 1792–1797.
57. Glaser, F., Pupko, T., Paz, I., Bell, R. E., Bechor-Shental, D., Martz, E. & Ben-Tal, N. (2003). ConSurf: identification of functional regions in proteins by surface-mapping of phylogenetic information. *Bioinformatics*, **19**, 163–164.
58. Brunger, A. T., Adams, P. D., Clore, G. M., DeLano, W. L., Gros, P., Grosse-Kunstleve, R. W. *et al.* (1998). Crystallography and NMR system: a new software suite for macromolecular structure determination. *Acta Crystallog. sect. D*, **54**, 905–921.

*Edited by J. Doudna*

(Received 5 July 2006; received in revised form 24 August 2006; accepted 11 September 2006)  
Available online 16 September 2006

## 2.4. Etude par RMN du domaine C-terminal

L'étude structurale du domaine C-terminal de hKIN17 (domaine K1) par RMN a été menée en parallèle à l'étude par radiocristallographie. Cette étude nous a permis d'attribuer le spectre  $^1\text{H}$ ,  $^{15}\text{N}$ -HSQC de la protéine en vue de son utilisation dans des expériences d'interaction et d'obtenir des informations sur la flexibilité de la protéine.

### 2.4.1. Attribution des pics de résonance du spectre $^1\text{H}$ , $^{15}\text{N}$ -HSQC

Le spectre 1D  $^1\text{H}$  du domaine C-terminal de KIN17 (Figure 2-5) montre que ce domaine est replié en solution dans les conditions de concentration, de tampon, de pH et de température de cette expérience RMN.

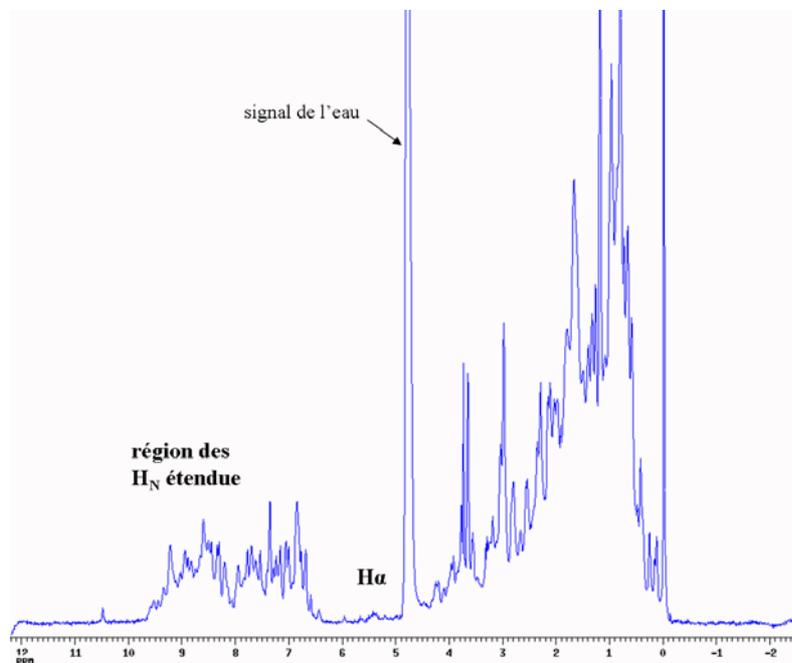


Figure 2-5 : Spectre 1D proton à 500 MHz d'un échantillon du domaine C-terminal de hKIN17 à 1mM, à 25°C en tampon phosphate pH7.

Nous avons donc poursuivi l'étude par des enregistrements de spectres 2D et 3D hétéronucléaires sur un échantillon unique de protéine doublement marquée  $^{13}\text{C}$ - $^{15}\text{N}$ . Les protéines doublement marquées possèdent des noyaux actifs tout le long de la chaîne principale et sur les chaînes latérales ; les nombreuses possibilités de transferts d'aimantation et leur efficacité rendent les expériences RMN hétéronucléaires ( $^1\text{H}$ ,  $^{15}\text{N}$  et  $^{13}\text{C}$ ) à plusieurs dimensions très riches en information. Ces expériences ont été utilisées pour une attribution rapide et sans ambiguïté des fréquences de résonance  $^1\text{H}$ ,  $^{15}\text{N}$  et  $^{13}\text{C}$  du domaine C-terminal de hKIN17.

L'attribution des systèmes de spins a été réalisée en combinant les informations des spectres HNCA (Hn, Ca et Ca-1), HNCACB (Hn, Cb et Cb-1), CBCA(CO)NH (Hn, Ca-1 et Cb-1), HNCO (Hn, Co-1) et HCACOCANH (Hn, Co et Co-1). Ces expériences ont permis d'identifier les valeurs des déplacements chimiques des carbones  $^{13}\text{CO}$ ,  $^{13}\text{C}\alpha$  et  $^{13}\text{C}\beta$  des

résidus  $i$  et  $i-1$  relatifs à chaque groupement amide. Les expériences  $^1\text{H}$ ,  $^{15}\text{N}$ -TOCSY HSQC (Hn, Ha et Hb...) et  $^1\text{H}$ ,  $^{15}\text{N}$ -NOESY HSQC (Hn, Ha-1 et Hb-1...) ont aussi été utilisées. L'interprétation et la comparaison de toutes ces expériences ont permis d'abord de connecter les résidus deux à deux puis de reconstituer l'ensemble de la séquence peptidique à partir des acides aminés facilement identifiables tels que l'alanine, la glycine, l'isoleucine, la proline, la sérine, la thréonine et la valine. Ainsi, nous avons pu attribuer un acide aminé à chaque système de spin, et donc associer un résidu du domaine K1 à chacun des pics de corrélation du spectre  $^1\text{H}$ ,  $^{15}\text{N}$ -HSQC (Figure 2-6).

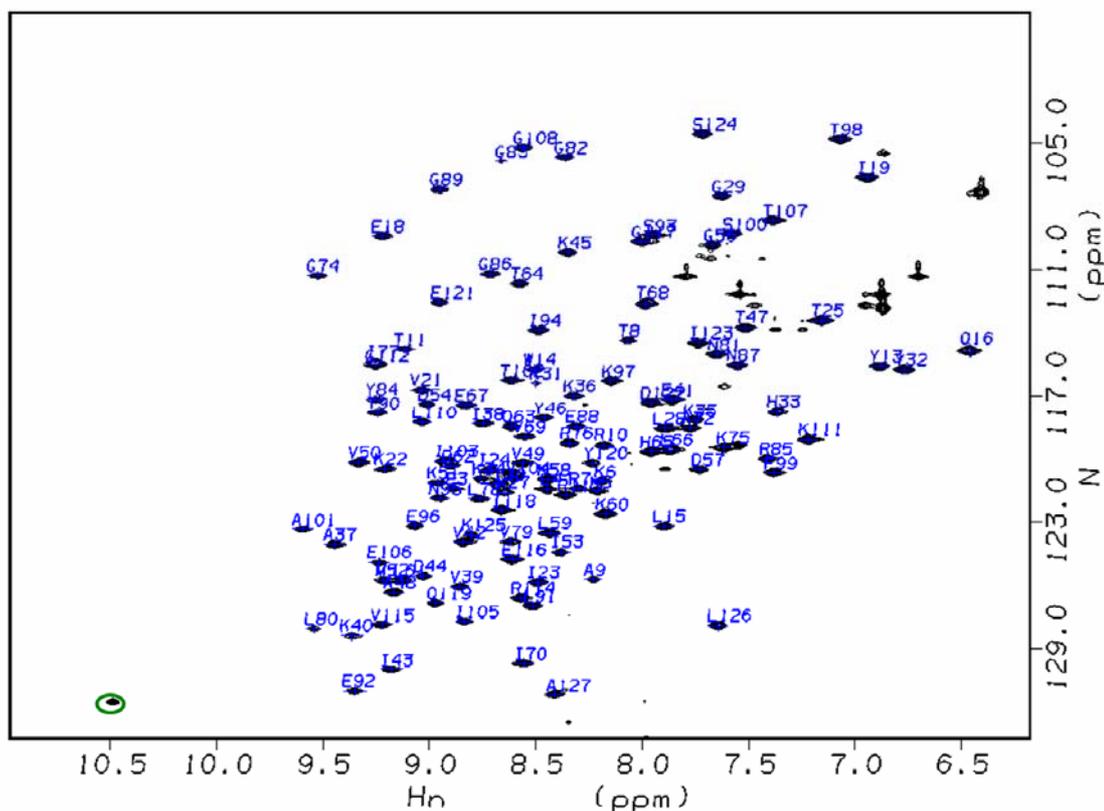


Figure 2-6 : Spectre  $^1\text{H}$ ,  $^{15}\text{N}$ -HSQC à 600 MHz d'un échantillon du domaine C-terminal de hKIN17 marquée  $^{15}\text{N}$  à 1mM, à 25°C en tampon phosphate pH7. L'attribution complète des résonances des protons amide du squelette est indiquée. La corrélation entourée en vert correspond au NH indole du tryptophane et les autres corrélations non attribuées correspondent aux  $\text{NH}_2$  de la chaîne latérale d'asparagines et de glutamines.

### 2.4.2. Dynamique du domaine C-terminal

Une expérience de mesure des nOes hétéronucléaires  $^1\text{H}$ - $^{15}\text{N}$  a été réalisée. Celle-ci permet de distinguer les parties flexibles ou déstructurées de la protéine et les régions repliées. Les nOes dépendent uniquement des mouvements rapides de la liaison  $\text{H}_\text{N}$ - $\text{N}$  dans le référentiel de l'aimant et de la distance  $\text{H}_\text{N}$ - $\text{N}$  mais sont indépendants des protons environnants (Renner et al. 2002). La mesure de l'aimantation  $^{15}\text{N}$  dans deux conditions d'irradiation radiofréquence, en saturant ou non le spin du proton associé au sein de la liaison NH, permet d'avoir une estimation des mouvements des NH de la protéine, sur une échelle de temps entre la picoseconde et la nanoseconde. Le rapport de l'intensité des pics  $^1\text{H}$ - $^{15}\text{N}$  dans ces deux conditions est calculé pour chaque résidu et si la différence est supérieure à 0,5 le résidu est dans une région rigide, si elle est comprise entre 0 et 0,5 le résidu est dans une

région flexible (telle que des boucles) et si elle est négative, le résidu appartient à une région déstructurée. La Figure 2-7 montre la valeur des nOe hétéronucléaires pour les résidus du domaine C-terminal de hKIN17.

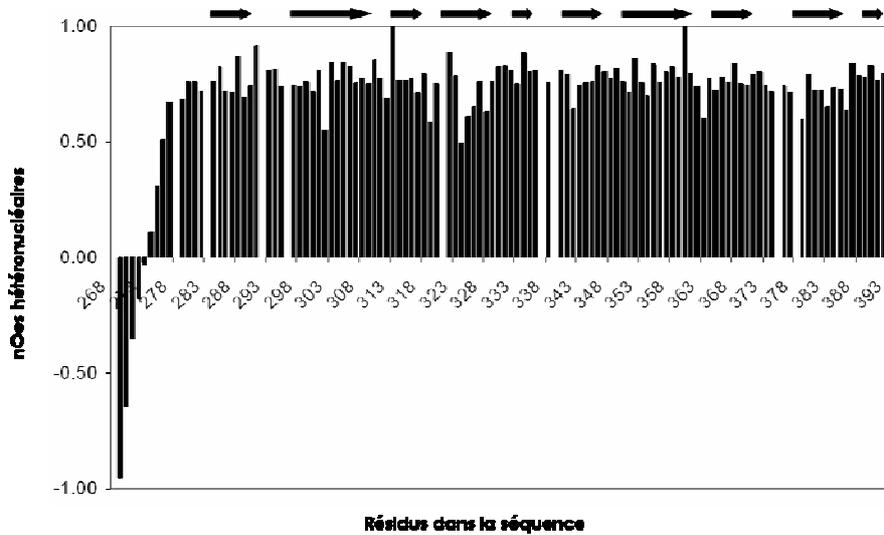


Figure 2-7 : Mesure de l'effet Overhauser hétéronucléaire  $^1\text{H}$ - $^{15}\text{N}$  pour la chaîne principale du domaine C-terminal de hKIN17. Les brins  $\beta$  déduits de la structure cristallographique du domaine C-terminal sont notés au-dessus de l'histogramme.

L'erreur sur les mesures est probablement assez grande car celles-ci n'ont été faites qu'une fois. En général, l'expérience est faite plusieurs fois afin de minimiser et d'estimer l'erreur commise. De plus, il manque certaines mesures, en particulier pour des résidus dans des boucles car pour ces résidus les taches de corrélation ne sont pas visibles sur le spectre  $^1\text{H}$ ,  $^{15}\text{N}$ -HSQC. Néanmoins, les nOes hétéronucléaires mesurés sont en accord avec le modèle obtenu par radiocristallographie du domaine C-terminal de hKIN17. En effet, le signal des premiers résidus correspond à un signal de résidus dans une région désordonnée et pour ces mêmes résidus nous n'avons pas observé de densité électronique. Pour le reste de la protéine, le signal correspond à des résidus dans des régions repliées. La différence entre les régions correspondant aux feuillets  $\beta$  et aux boucles n'apparaît pas clairement probablement parce que les boucles sont limitées en taille et ne sont donc pas très flexibles. De même, le linker entre les deux domaines, de taille limitée, n'apparaît pas flexible.

### 2.4.3. Topologie du domaine C-terminal

L'attribution des déplacements chimiques HA, CA, CB, CO, N a été utilisée pour déterminer les angles de torsion du squelette  $\Phi$  et  $\Psi$  des résidus attribués avec le logiciel TALOS (Cornilescu et al. 1999). Ce logiciel permet ainsi de prédire dans quelle région du diagramme de Ramachandran se trouve chaque résidu. Il donne pour chacune de ses prédictions un indice de confiance. Les valeurs des angles  $\Phi$  et  $\Psi$  pour lesquels des résultats avec un indice de confiance supérieur à 8 ont été obtenus sont comparés aux valeurs des angles  $\Phi$  et  $\Psi$  déduits de la structure obtenue par cristallographie (Figure 2-8).

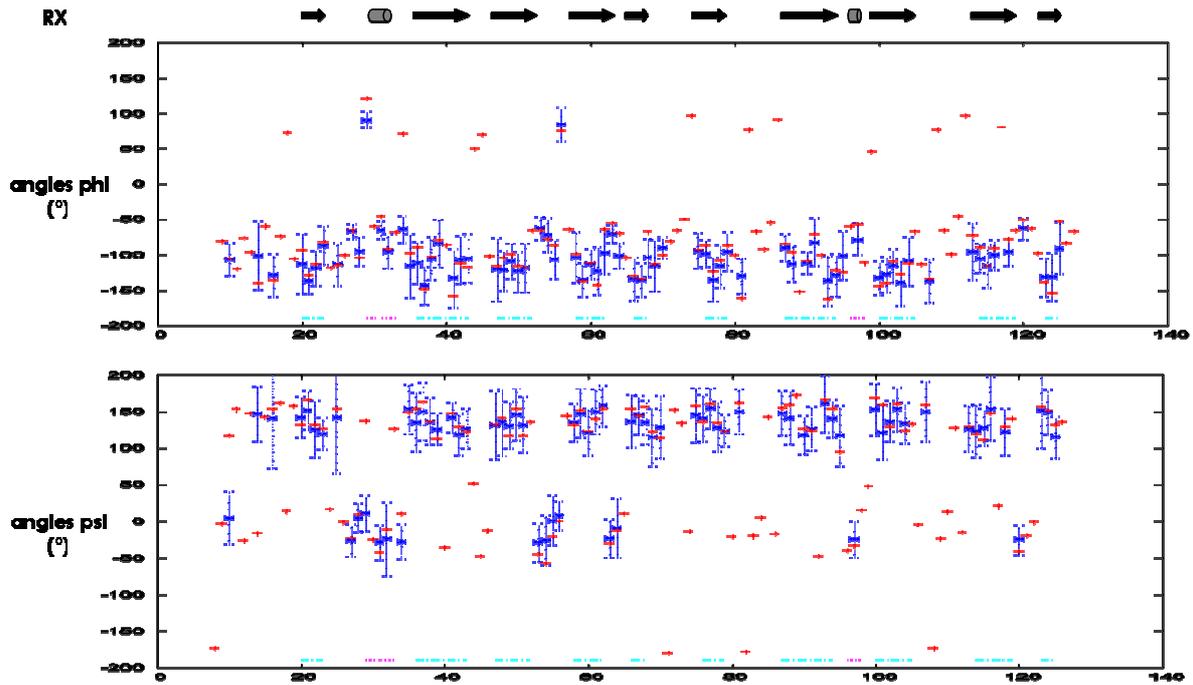


Figure 2-8 : Comparaison des angles  $\Phi$  et  $\Psi$  déduits de la structure cristallographique (croix rouges) et prédits par TALOS (carrés bleus) pour le domaine C-terminal de hKIN17. Les barres d'erreurs pour les prédictions de Talos correspondent aux barres d'erreurs généralement utilisées lors de modélisation sous contrainte RMN. Les structures secondaires (flèches pour les feuillets  $\beta$  et cylindres pour les hélices  $\alpha$ ) déduites de la structure obtenue par cristallographie aux rayons X sont également représentées.

En tenant compte des barres d'erreurs, les valeurs des angles  $\Phi$  et  $\Psi$  prédits par TALOS et des angles déduits de la structure cristallographique sont les mêmes, sauf pour un résidu dans la boucle entre  $\beta_1$  et  $\beta_2$ . Cela suggère que le domaine C-terminal adopte la même structure en solution que dans les cristaux.

*En conclusion*, cette étude en RMN, en parallèle de l'obtention de la structure à haute résolution du domaine C-terminal de hKIN17 par radiocristallographie, a permis de montrer que le domaine a la même structure en solution que dans les cristaux et l'attribution du spectre  $^1\text{H}$ ,  $^{15}\text{N}$ -HSQC est utilisable dans le cadre d'étude d'interaction avec des partenaires biologiques de ce domaine grâce à la méthode de cartographie des variations de déplacements chimiques. La position d'un pic de corrélation est très sensible à l'environnement chimique de l'amide correspondant et donc des variations de pH, de température ou la liaison à une autre molécule peuvent induire des changements dans le spectre  $^1\text{H}$ - $^{15}\text{N}$ , HSQC de la protéine. Dans le cas d'une interaction protéine-protéine, les changements dans le spectre sont associés à quelques résidus qui sont soit directement impliqués dans ou proche du site d'interaction, soit impliqués dans des réarrangements conformationnels induits par la liaison.

### 3. LA CARACTERISATION STRUCTURALE DE LA REGION N-TERMINALE

---

La région N-terminale de hKIN17 est très conservée de la levure à l'homme, en incluant les parasites et les plantes. Elle est constituée de deux domaines structuraux : un domaine prédit en doigt de zinc des résidus 26 à 50 et un autre domaine globulaire des résidus 51 à 160. La structure de ce dernier domaine (domaine K2) a été déterminée par RMN et a permis de montrer qu'il appartient à la famille des *winged helix*. L'étude de l'ensemble de la région N-terminale (domaine K3) a été réalisée en combinant des informations de mesures RMN et SAXS.

J'ai initié l'étude structurale du domaine K2 et la structure par RMN a été déterminée par L. Carlier de l'Université de Rouen.

J'ai encadré S.Gras, étudiante de master, pour l'étude par RMN du domaine K3.

J'ai réalisé l'étude du domaine K3 par SAXS en allant travailler avec le Pr. M. Schiltz du LCR/EPFL (Lausanne, Suisse) et avec le Dr. P.Vachette de l'IBBMC (Orsay). Le Dr. B. Gilquin a été impliqué dans l'étape finale de modélisation du domaine K3.

#### 3.1. Détermination de la structure tridimensionnelle du domaine 51-160 par RMN

J'ai initié l'étude structurale du domaine K2 mais la structure tridimensionnelle a été déterminée par L.Carlier de l'Université de Rouen.

L'article 3 décrit les conditions de préparation de la protéine K2 ainsi que les expériences RMN enregistrées pour l'attribution du spectre  $^1\text{H}$ ,  $^{15}\text{N}$ -HSQC de ce domaine de la protéine hKIN17.

Résumé article 3 : *NMR assignment of region 51–160 of human KIN17, a DNA and RNA-binding protein* (2006) L Carlier, A le Maire, S Braud, C Masson, M Gondry, S Zinn-Justin, L Guilhaudis, I Milazzo, D Davoust, B Gilquin & J Couprie, Journal of Biomolecular NMR, Vol 36, Suppl 5

Dans l'article, sont reportées les attributions  $^1\text{H}$ ,  $^{15}\text{N}$ ,  $^{13}\text{C}$  de la région 51–160 de la protéine hKIN17 qui constituent une première étape vers l'obtention de la structure atomique de cette région et l'élucidation de sa fonction biologique. Les résidus de ce domaine de KIN17 ont tous été identifiés : 99% des déplacements chimiques du squelette et pour les chaînes latérales, tous les aliphatiques et 88% des noyaux aromatiques ont été attribués. L'attribution a été déposée à la BMRB sous le numéro 6938.

## Letter to the Editor

### NMR assignment of region 51–160 of human KIN17, a DNA and RNA-binding protein

DOI 10.1007/s10858-006-0013-y

KIN17 is a 45 kDa nuclear protein which is remarkably conserved in eukaryotes, ubiquitously expressed in mammals, and forms intra-nuclear foci in proliferating cells. Major features of KIN17 are its ability to bind DNA and RNA and to be up-regulated in response to UV and irradiation, suggesting a role in DNA replication, the DNA damage response or RNA processing (Pinon-Lataillade et al., 2004; Miccoli et al., 2005). Human KIN17 comprises an N-terminal zinc finger (27–50) and a C-terminal KOW motif (335–373). Here we report the  $^1\text{H}$ ,  $^{15}\text{N}$ ,  $^{13}\text{C}$  assignments of the region 51–160 of human KIN17 as a preliminary step toward obtaining the atomic structure of this region and elucidating its biological function.

Residues corresponding to the KIN17 domain were all identified: 99% of the backbone chemical shifts, and for the side chains, all the aliphatic and 88% of the aromatic  $^1\text{H}$ – $^{13}\text{C}$  nuclei were assigned. BMRB deposit with accession number 6938.

References: Miccoli et al. (2005) *Mol. Cell. Biol.*, **25**, 3814–3830; Pinon-Lataillade et al. (2004) *J. Cell. Sci.*, **117**, 3691–3702.

Ludovic Carlier<sup>a</sup>, Albane le Maire<sup>b</sup>, Sandrine Braud<sup>b</sup>, Cédric Masson<sup>b</sup>, Muriel Gondry<sup>b</sup>, Sophie Zinn-Justin<sup>a</sup>, Laure Guilhaudis<sup>a</sup>, Isabelle Milazzo<sup>a</sup>, Daniel Davoust<sup>a</sup>, Bernard Gilquin<sup>b</sup> & Joël Couprie<sup>b,\*</sup>  
<sup>a</sup>*Equipe de Chimie Organique et Biologie Structurale, CNRS UMR 6014, IFRMP 23, Université de Rouen, France;* <sup>b</sup>*Département d'Ingénierie et d'Etude des Protéines, CEA Saclay, 91191, Gif-sur-Yvette, France*

\*To whom correspondence should be addressed. E-mail: joel.couprie@cea.fr

**Supplementary material** is available in electronic format at <http://dx.dio.org/10.1007/10858-006-0013-y>.

**Letter to the Editor (Supplemental Material)**

**NMR assignment of region 51-160 of human KIN17,  
a DNA and RNA-binding protein**

Ludovic Carlier<sup>a</sup>, Albane le Maire<sup>b</sup>, Sandrine Braud<sup>b</sup>, Cédric Masson<sup>b</sup>, Muriel Gondry<sup>b</sup>,  
Sophie Zinn-Justin<sup>b</sup>, Laure Guilhaudis<sup>a</sup>, Isabelle Milazzo<sup>a</sup>, Daniel Davoust<sup>a</sup>,  
Bernard Gilquin<sup>b</sup> and Joël Couprie<sup>b,\*</sup>

<sup>a</sup> *Equipe de Chimie Organique et Biologie Structurale, IFRMP 23, CNRS UMR 6014, Université de Rouen*

<sup>b</sup> *Département d'Ingénierie et d'Etude des Protéines, CEA Saclay, 91191 Gif-sur-Yvette, France.*

\* To whom correspondence should be addressed: joel.couprie@cea.fr

Keywords: NMR assignments, KIN17, DNA-binding, RNA-binding, nuclear metabolism

## Biological context

KIN17 is a 45 kDa nuclear protein conserved in eukaryotes and initially identified on the basis of its cross-reactivity with antibodies raised against bacterial RecA, a recombination-repair enzyme (Kannouche et al., 2000). In mammals, the KIN17 protein is ubiquitously expressed. It forms intranuclear foci in proliferating cells and is up-regulated after UV or  $\gamma$  irradiation (Biard et al., 2002; Kannouche et al., 1998). The other features of KIN17 are its ability to bind DNA and RNA *in vitro* and *in vivo* (Biard et al., 2002; Pinon-Lataillade et al., 2004), to associate with multiprotein DNA replication complexes (Miccoli et al., 2005), and to complement the functions of the bacterial transcriptional factor H-NS (Timchenko et al., 1996), suggesting a role in nuclear metabolism.

Human KIN17 is a modular protein comprising four motifs: a zinc finger (28-50) at the N-terminus, a core domain homologous to RecA protein (163-201), a nuclear localization signal (240-257), and a KOW motif (335-373) found in several bacterial transcription factors (Kannouche et al., 2000; Ponting et al., 2002). On the basis of sequence alignment and secondary structure prediction, we identified a globular region located between the zinc finger and the core domain. We here report the  $^1\text{H}$ ,  $^{15}\text{N}$ ,  $^{13}\text{C}$  assignments of the region 51-160 as a preliminary step toward obtaining the atomic structure of this KIN17 domain and elucidating its biological function.

## Methods and experiments

The gene encoding the region 51-160 of human KIN17 was amplified by polymerase chain reaction (PCR) and cloned into several expression vectors using the Gateway system (Invitrogen). A protein expression screening using different strains and fusion partners was performed (Braud et al., *in press*) and highest levels of expression and solubility were obtained when transforming *E. coli* Rosetta(DE3) strain with plasmid pEXP-TH5 (Invitrogen). This vector encodes a 6xHis tag, the Z-domain from Staphylococcal Protein A (ZZ), a TEV protease (Tobacco Etch Virus) cleavage site, and the KIN17 51-160 domain. Protein expression was induced by the addition of 1 mM of isopropyl-1-thio- $\beta$ -D-galactopyranoside (IPTG) to the bacterial culture at a cell density of 1.2 absorbance units measured at 595 nm. The bacterial culture was further grown for 14 hours at 20°C.

The fusion protein was purified using IgG Sepharose Fast Flow (Amersham) in order to separate the fusion protein from other bacterial proteins and to remove the (His)<sub>6</sub>-ZZ tag after an on-column cleavage using recombinant (His)<sub>6</sub>-TEV protease. A second step of purification was performed using Ni-NTA column (Qiagen) to remove (His)<sub>6</sub>-TEV protease and residual (His)<sub>6</sub>-ZZ tag.

Uniformly labelled <sup>15</sup>N protein was produced in minimal medium M9 containing 1 g.L<sup>-1</sup> of (<sup>15</sup>NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub> (Boehringer) as the sole nitrogen source. Uniformly labelled <sup>13</sup>C/ <sup>15</sup>N protein was produced in a rich medium prepared from uniformly labelled <sup>13</sup>C/<sup>15</sup>N *Spirulina maxima* cyanobacteria. NMR samples containing about 0.7 mM protein were prepared in 50 mM phosphate buffer pH 6.0 containing 150 mM NaCl, 1 mM EDTA, a protease inhibitor cocktail (SIGMA), 1 mM TCEP, and 1 mM NaN<sub>3</sub> in either 90% H<sub>2</sub>O/10% D<sub>2</sub>O or in 100% D<sub>2</sub>O. 3-(trimethylsilyl)[2,2,3,3-<sup>2</sup>H<sub>4</sub>] propionate (TSP) was added as an internal <sup>1</sup>H chemical shift reference. <sup>13</sup>C and <sup>15</sup>N chemical shifts were referenced indirectly to TSP, using the absolute frequency ratios (Wishart et al., 1995).

NMR experiments were performed at 303 K on Bruker AVANCE DMX-600 or Varian INOVA-800 spectrometers equipped with triple-resonance (<sup>1</sup>H, <sup>15</sup>N, <sup>13</sup>C) cryoprobes. Spectra used for sequential backbone assignment were as follows: 2D <sup>15</sup>N-<sup>1</sup>H HSQC, 3D HNCO, 3D HNCA, 3D CBCA(CO)NH, 3D CBCANH, 3D HNHA, 3D (HCA)CO(CAN)NH experiments. Aliphatic and aromatic <sup>1</sup>H and <sup>13</sup>C side chain assignments were obtained using 3D HBHA(CO)NH, 3D HCCH-TOCSY, 3D HCCH-COSY, and 3D <sup>13</sup>C-HSQC NOESY experiments. All NMR data were processed with NMRPipe software (Delaglio et al., 1995) and analysed with Felix software (Molecular Simulations).

### **Extent of assignments and data deposition**

For cloning strategy reasons, this KIN17 domain contains the region 51-160 of human KIN17 plus an additional glycine residue at the N terminus. Figure 1 shows the <sup>15</sup>N-<sup>1</sup>H HSQC spectrum of the 111 residues protein numbered from 1 to 111. All the amide <sup>1</sup>H and <sup>15</sup>N backbone resonances were assigned except for N-terminal amino acids G1 and Q2, and R53 whose NH is unobservable presumably due to loop mobility or chemical exchange since this residue is located between secondary structure elements.

Compared to the expected number of resonances, nine additional weakened cross-peaks were observed in the  $^{15}\text{N}$ - $^1\text{H}$  HSQC spectrum. However, backbone and side chain assignments, as well as 3D  $^{15}\text{N}$  HSQC-NOESY patterns revealed that 6 spin systems split in two (R3, Q4, L5, L6, A8, S9). Those residues are located before the proline P12 and we suppose that the minor form is due to a partial isomerisation of P12. The observation of a characteristic strong NOE between  $\text{H}_\beta$  of P12 and  $\text{H}_\alpha$  of N11 indicated that the predominant form of this proline is in a *trans* conformation.

Finally, all residues corresponding to the region 51-160 of KIN17 were identified: the backbone assignment is complete for  $^1\text{H}_\alpha$ ,  $^{13}\text{C}_\alpha$  and  $^{13}\text{CO}$ , and it reaches 98% for  $^1\text{H}_\text{N}$  and  $^{15}\text{N}$  nuclei. All the aliphatic side chain resonances were assigned and 88% of the aromatic  $^1\text{H}$  and  $^{13}\text{C}$  resonances were identified, including all Tyr and His aromatic rings, 3 Phe out of 6, and 2 Trp out of 3. Assignments of  $\text{NH}_2$  side chain resonances of asparagines and glutamines were also completed except for N42. The  $^1\text{H}$ ,  $^{13}\text{C}$  and  $^{15}\text{N}$  chemical shifts have been deposited in the BioMagResBank (<http://www.bmrb.wisc.edu>) under accession number 6938.

## Acknowledgement

The Bruker DMX-600 spectrometer was supported by grants of the Conseil Regional de Haute-Normandie (France). We are grateful to Dr Adrien Favier who kindly recorded 3D HSQC NOESY experiments on Varian INOVA-800 spectrometer at I.B.S Jean-Pierre Ebel (Grenoble, France), and to Marie Courçon and Mireille Moutiez for the protein expression screening. We also thank the Centre de Ressources Informatiques de Haute-Normandie (France) for NMR software facilities.

## References

- Braud, S., Moutiez, M., Belin, P., Abello, N., Drevet, P., Zinn-Justin, S., Courçon, M., Masson, C., Dassa, J., Charbonnier, J.B., Boulain, J.C., Ménez, A., Genet, R., Gondry, M. (2005). *J. Proteome Res.*, In press.
- Biard, D.S., Miccoli, L., Despras, E., Frobert, Y., Creminon, C. and Angulo, J.F. (2002) *J. Biol.Chem.*, **277**(21), 19156-19165.
- Delaglio, F., Grzesiek, S., Vuister, G. W., Zhu, G., Pfeifer J. and Bax, A. (1995) *J. Biomol. NMR.*, **6**, 277-293.

Kannouche, P., Mauffrey, P., Pinon-Lataillade, G., Mattei, MG., Sarasin, A., Daya-Grosjean, L. and Angulo, JF. (2000) *Carcinogenesis*, **21**(9), 1701-1710.

Kannouche, P., Pinon-Lataillade, G., Tissier, A., Chevalier-Lagente, O., Sarasin, A., Mezzina, M. and Angulo, JF. (1998) *Carcinogenesis*, **19**(5), 781-789.

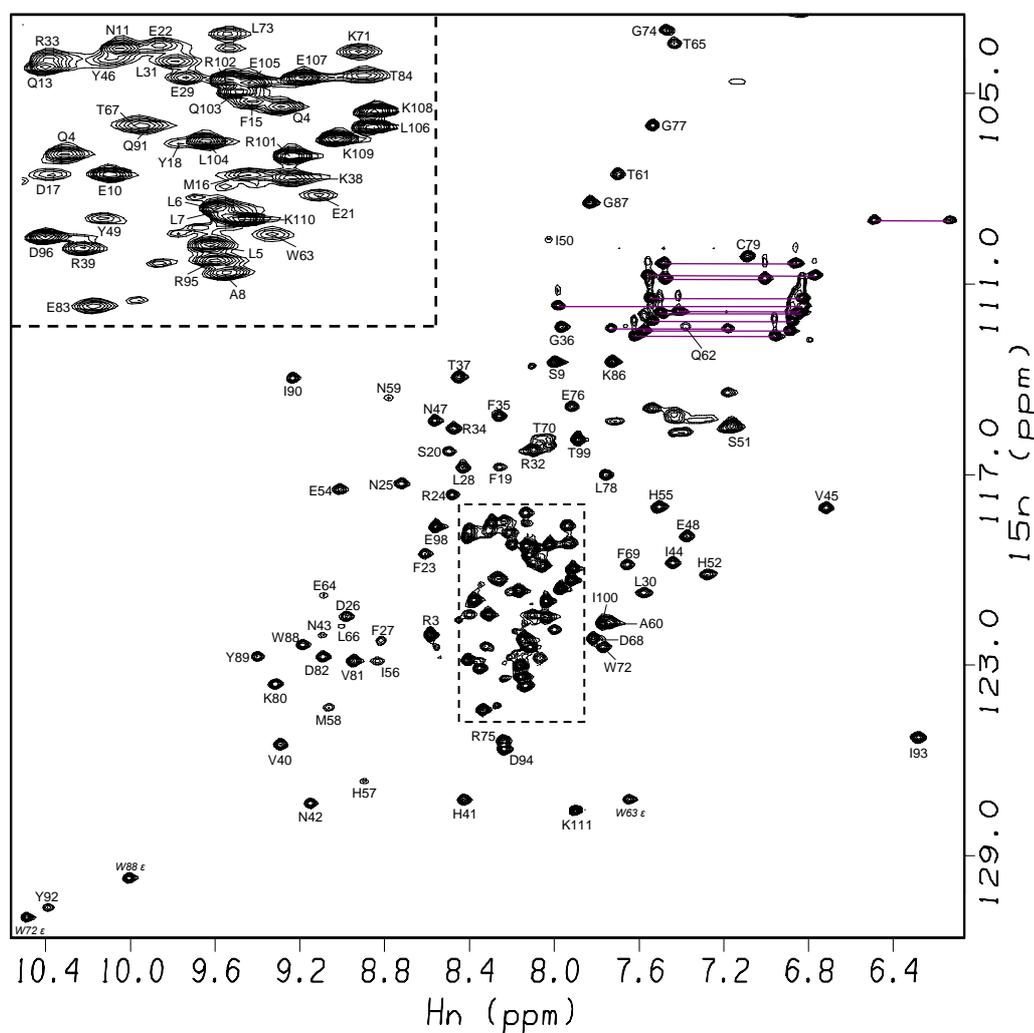
Miccoli, L., Frouin, I., Novac, O., Di Paola, D., Harper, F., Zannis-Hadjopoulos, M., Maga, G., Biard, D.S.F., Angulo, J.F. (2005). *Mol. Cell. Biol.*, **25**(9), 3814-3830.

Pinon-Lataillade, G., Masson, C., Bernardino-Sgherri, J., Henriot, V., Mauffrey, P., Frobert, Y., Araneda, S. and Angulo, JF. (2004) *J. Cell. Sci.*, **117**(16), 3691-3702

Ponting, CP. (2002) *Nucleic Acids Res.*, **30**(17), 3643-3652.

Timchenko, T., Bailone, A. and Devoret, R. (1996) *EMBO J.*, **15**, 3986-3992.

Wishart, DS., Bigam, CG., Yao, J., Abildgaard, F., Dyson, HJ., Oldfield, E., Markley, JL. and Sykes, BD. (1995) *J. Biol. NMR*, **6**, 135-140



**Figure 1:** 2D  $^{15}\text{N}$ - $^1\text{H}$  HSQC spectrum of the region 51-160 of human KIN17. The assignments are indicated with the one-letter amino acid code.  $\text{NH}_2$  side chain resonances of asparagines and glutamine are connected by purple horizontal lines.

L'article 4 présente la structure tridimensionnelle du domaine K2 obtenue par RMN.

Résumé article 4 : *Solution structure of the region 51-160 of human KIN17 reveals an atypical Winged Helix domain* (2007) L Carlier, J Couprie, A le Maire, L Guilhaudis, I Milazzo, M Courçon, M Moutiez, M Gondry, D Davoust, B Gilquin and S Zinn-Justin, Protein Science (sous presse).

Dans l'article, est présentée la structure en solution de la région 51-160 de la protéine KIN17 humaine. Nous montrons que ce domaine se replie en un « tonnelet orthogonal » de trois hélices  $\alpha$  contre un feuillet  $\beta$  anti-parallèle constitué de trois brins. Ce repliement appartient à la famille des *winged helix*. La comparaison structurale de ce domaine avec d'autres domaines WH révèle que le domaine WH de KIN17 présente une hélice  $3_{10}$  additionnelle et très conservée. De plus, l'hélice H3, impliquée classiquement dans l'interaction des WH avec l'ADN, n'est pas positivement chargée dans le domaine WH de KIN17. Ainsi, nous proposons que la région 51-160 de la protéine humaine KIN17 soit plutôt impliquée dans une interaction protéine-protéine par sa surface conservée centrée sur l'hélice  $3_{10}$ .

## PROTEIN STRUCTURE REPORT

## Solution structure of the region 51–160 of human KIN17 reveals an atypical winged helix domain

LUDOVIC CARLIER,<sup>1</sup> JOËL COUPRIE,<sup>2</sup> ALBANE LE MAIRE,<sup>2</sup> LAURE GUILHAUDIS,<sup>1</sup> ISABELLE MILAZZO-SEGALAS,<sup>1</sup> MARIE COURÇON,<sup>2</sup> MIREILLE MOUTIEZ,<sup>2</sup> MURIEL GONDRIY,<sup>2</sup> DANIEL DAVOUST,<sup>1</sup> BERNARD GILQUIN,<sup>2</sup> AND SOPHIE ZINN-JUSTIN<sup>2</sup><sup>1</sup>Equipe de Chimie Organique et Biologie Structurale, IFRMP 23, CNRS UMR 6014, Université de Rouen, 76821 Mont-Saint-Aignan, France<sup>2</sup>CEA de Saclay, Institut de Biologie et Technologies iBiTec-S, 91190 Gif-sur-Yvette, France

(RECEIVED June 21, 2007; FINAL REVISION August 22, 2007; ACCEPTED September 20, 2007)

## Abstract

Human KIN17 is a 45-kDa eukaryotic DNA- and RNA-binding protein that plays an important role in nuclear metabolism and in particular in the general response to genotoxics. Its amino acids sequence contains a zinc finger motif (residues 28–50) within a 30-kDa N-terminal region conserved from yeast to human, and a 15-kDa C-terminal tandem of SH3-like subdomains (residues 268–393) only found in higher eukaryotes. Here we report the solution structure of the region 51–160 of human KIN17. We show that this fragment folds into a three- $\alpha$ -helix bundle packed against a three-stranded  $\beta$ -sheet. It belongs to the winged helix (WH) family. Structural comparison with analogous WH domains reveals that KIN17 WH module presents an additional and highly conserved  $3_{10}$ -helix. Moreover, KIN17 WH helix H3 is not positively charged as in classical DNA-binding WH domains. Thus, human KIN17 region 51–160 might rather be involved in protein–protein interaction through its conserved surface centered on the  $3_{10}$ -helix.

**Keywords:** KIN17 protein; winged helix fold; NMR structure; nuclear metabolism**Supplemental material:** see [www.proteinscience.org](http://www.proteinscience.org)

Maintenance of the genomic integrity is crucial for all living organisms. Since cells are continuously submitted to genotoxic stresses that induce DNA damages, they have evolved sophisticated mechanisms to detect, signal the presence of, and repair these lesions. In humans, the nucleotide excision repair pathway is an efficient way to remove DNA lesions induced by UV or ionizing radiation. The human gene *KIN17*, located in chromosome 10, participates in the general response to genotoxics. It is up-regulated following DNA damages produced by UV

or  $\gamma$  irradiation, and this response is strictly dependent on the presence of nucleotide excision repair proteins XPA and XPC (Masson et al. 2003).

The *KIN17* gene encodes a ubiquitously expressed nuclear protein that forms intranuclear foci in proliferating cells (Biard et al. 2002). Major features of the mammalian KIN17 protein are its ability to bind DNA and RNA in vitro and in vivo, to associate with multiprotein DNA replication complexes, and, when overexpressed, to inhibit the S-phase progression of the replication process (for review, see Angulo et al. 2005). Furthermore, the mouse KIN17 protein binds preferentially to curved DNA found at hot spots of illegitimate recombination in eukaryotic chromosomes (Mazin et al. 1994a). Taken together, these observations clearly indicate that KIN17 plays an important role in nuclear metabolism.

Reprint request to: Sophie Zinn-Justin, LBSR (bât 144), iBiTec-S, CEA Saclay, 91190 Gif-sur-Yvette, France; e-mail: [sophie.zinn@cea.fr](mailto:sophie.zinn@cea.fr); fax: 33-1-69084712.

Article and publication are at <http://www.proteinscience.org/cgi/doi/10.1110/ps.073079107>.

Human KIN17 is a 45-kDa modular protein composed of a 30-kDa N-terminal region conserved from yeast to human, and a 15-kDa tail found only in higher eukaryotes. The recently solved crystal structure of its C-terminal domain (residues 268–393) reveals a tandem of SH3-like subdomain that contributes to RNA-binding of the whole protein (le Maire et al. 2006). Deletion of this C-terminal domain does not affect the DNA binding of KIN17 (Mazin et al. 1994b). The N-terminal region comprises a C<sub>2</sub>H<sub>2</sub> zinc finger motif (residues 28–50), which exhibits DNA- and RNA-binding properties (Mazin et al. 1994b; le Maire et al. 2006), and a globular domain (residues 51–160) of unknown function. Interestingly, if deletion of the zinc finger motif reduces KIN17 DNA binding, the core of the protein (residues 71–281) which comprises the globular region 51–160, still binds DNA (Mazin et al. 1994b). Here, we report the high-resolution NMR structure of human KIN17 region 51–160 as an important step toward identifying the structural and biological functions of this domain.

## Results and Discussion

### Structure determination

The three-dimensional structure of human KIN17 region 51–160 was characterized using heteronuclear NMR and molecular modeling. As reported previously (Carlier et al. 2006), <sup>1</sup>H, <sup>15</sup>N, and <sup>13</sup>C chemical shift assignments were completed for 95% of the backbone atoms and 90% of the side-chain resonances (BMRB accession number 6938). The solution structure was calculated using 2743 NOESY-derived proton–proton distances, 18 hydrogen bonds, and 140  $\phi$  and  $\psi$  torsion angle values. Table 1 summarizes the structural statistics for the ensemble of the 12 best conformers that represent the solution structure of human KIN17 region 51–160. Analysis of this final ensemble shows that no distance violations larger than 0.5 Å are present and that the covalent geometry is respected. Atomic coordinates have been deposited in the RSCB Protein Data Bank under accession number 2V1N.

### Structure description

A backbone superposition of the 12 best conformers is shown in Figure 1A as a stereo diagram. The first 14 N-terminal residues and the five last C-terminal amino acids (residues 51–64 and 156–160, respectively) are disordered, whereas fragment 65–155 adopts a well-defined  $\alpha/\beta$  fold. The root mean square deviation (RMSD) calculated from the mean structure between residues 65 and 155 yields 0.43 Å for the backbone and 1.03 Å for the heavy atoms. The three-dimensional structure is constituted of four  $\alpha$ -helices, H1 (residues 65–83), H2 (residues 91–99), H3 (residues 115–124), and H4 (residues 147–155); three  $\beta$ -strands, S1 (residues 88–90),

**Table 1.** Structural statistics for the ensemble of 12 final conformers

Number of nOe assigned	3413
Number of restraints	
Distance restraints	2743
Dihedral restraints	140
from TALOS prediction	136
from analysis of HNHA or HMQC-J	4
Hydrogen bonds	18
Number of violations	
nOe distance > 0.5 Å	0
Dihedral restraints > 10°	0
RMSD from idealized covalent geometry	
Bond (Å)	0.016 ± 0.0002
Angle (°)	3.3 ± 0.03
Improper (°)	2.7 ± 0.1
RMSD on restraints	
Distance (Å)	0.032 ± 0.0012
Dihedral (°)	1.1 ± 0.07
Coordinate precision (residues 65–155)	
Backbone (Å)	0.43 ± 0.08
Heavy atoms (Å)	1.03 ± 0.12
Ramachandran plot (residues 65–155)	
Most favored regions	88.6%
Additional allowed regions	10.1%
Generously allowed regions	0.8%
Disallowed regions	0.5%

S2 (residues 128–133), and S3 (residues 136–141); and one <sub>3</sub><sub>10</sub>-helix, H2.5 (residues 107–109), organized into an H1-S1-H2-H2.5-H3-S2-S3-H4 topology. Thus, the amino-terminal half of the domain is mainly  $\alpha$ -helical, whereas the carboxyl-terminal half is dominated by two  $\beta$ -strands arranged in a twisted antiparallel  $\beta$ -sheet (Fig. 1B). The  $\alpha$ -helices H1, H2, and H3 form a central orthogonal bundle. They are amphipathic, and their hydrophobic core is constituted by Met-65, Phe-72, Phe-76, Leu-77, Leu-79, Leu-80 (H1), Val-94, Tyr-95, Tyr-98 (H2), Leu-115, Phe-118, Trp-121, and Leu-122 (H3). The three-stranded  $\beta$ -sheet is packed onto the three-helix bundle via a hydrophobic interface in which Phe-76, Leu-77, Leu-80 (H1), Leu-115, Thr-119, Leu-122, and Gly-123 (H3) contact the hydrophobic  $\beta$ -sheet face composed of Cys-128, Val-130 (S2), Trp-137, Ile-139, and Tyr-141 (S3).

The three-amino-acid <sub>3</sub><sub>10</sub>-helix H2.5 is located in a highly conserved long loop (residues 100–114) that connects  $\alpha$ -helices H2 and H3 (Figs. 1B, 2A). The presence of such a short <sub>3</sub><sub>10</sub>-helix in the unstructured region H2–H3 is supported by unambiguous assignment of several interresidue <sup>1</sup>H–<sup>1</sup>H nOes, in particular two consecutive  $d_{\alpha N}(i, i + 2)$  and one  $d_{\alpha N}(i, i + 3)$  connectivities that involve residues 107–110 (Supplemental Fig. S1). In addition,  $\phi$  and  $\psi$  dihedral restraints obtained for Met-107 and Asn-108 from TALOS prediction are consistent with a <sub>3</sub><sub>10</sub>-helix conformation (Supplemental Table S1). In the NMR structures, the <sub>3</sub><sub>10</sub>-helix H2.5 is stabilized by



sensitive to rapid motions (i.e., picosecond to nanosecond timescale), as a function of residue number. Negative and small values (<0.25) are observed for the extremities 52–60 and 157–160, confirming the mobility of the N- and C-terminal parts of the domain. Intermediate values (between 0.35 and 0.6) are found for the C-terminal region 144–156, indicating partially restricted motions in the helical region H4. Consistently, only a small number of medium- and long-range  $^1\text{H}$ - $^1\text{H}$  nOes were unambiguously assigned to residues 52–60 and 144–160 (Supplemental Fig. S1). Finally, almost all residues belonging to the three-helix bundle, the antiparallel  $\beta$ -sheet, and the long loop H2–H3 (i.e., from Met-65 to Asp-143) display a  $^1\text{H} \rightarrow ^{15}\text{N}$  nOe value close to 0.8 characteristic of highly restricted rapid motions. This last region adopts a stable conformation on the picosecond to nanosecond timescale.

#### *Region 51–160 of human KIN17 adopts a winged helix fold*

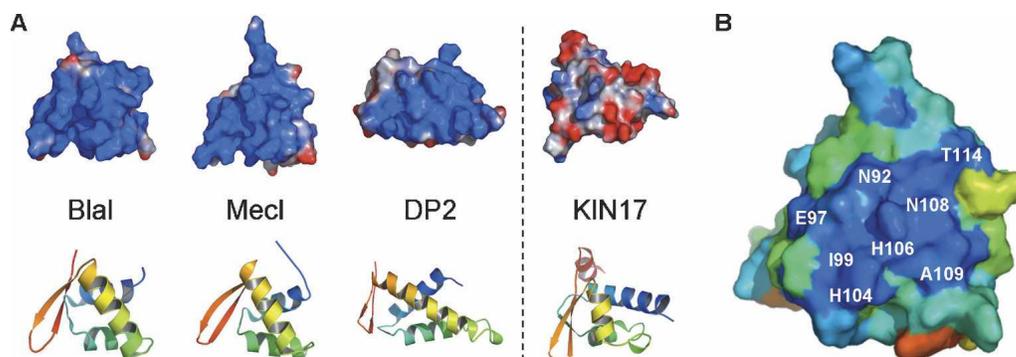
The three-dimensional structure of human KIN17 fragment 65–155 was submitted to the DALI server ([www.ebi.ac.uk/dali/](http://www.ebi.ac.uk/dali/)). It is close to the three-dimensional structure of numerous proteins belonging to the winged helix family (WH), which is a member of the DNA-binding helix-turn-helix superfamily (Gajiwala and Burley 2000). The WH motif is a compact  $\alpha/\beta$  fold consisting of three  $\alpha$ -helices (H1, H2, and H3), three  $\beta$ -strands (S1, S2, and S3), and two loops called wings W1 and W2, organized in an H1-S1-H2-H3-S2-W1-S3-W2 topology. The WH domain of human KIN17 is very close to the canonical topology. The only difference occurs at the level of the H2–H3 loop, where an atypical well-defined  $3_{10}$ -helix is observed. Analysis of the phylogenetic conservation of human KIN17 region 51–160 was achieved using the PSI-blast server ([www.ncbi.nlm.nih.gov/BLAST/](http://www.ncbi.nlm.nih.gov/BLAST/)). Figure 2A shows that this region is particularly well conserved from yeast to human. Most residues from the hydrophobic core of the domain are conserved (31% sequence identity and 77% sequence similarity over 14 residues), as well as 8 of the 15 amino acids belonging to the long loop H2–H3. Such analysis suggests that all eukaryotic KIN17 proteins share a common WH domain at their amino-terminal region that is structurally similar to the human KIN17 region 51–160.

#### *Possible functions of the human KIN17 winged helix domain*

The WH module is mainly known for its ability to interact with nucleic acids and in particular with DNA. Interestingly, the DALI tool shows that human KIN17 region 51–160 displays structural analogy with the DNA-binding WH domain of BlaI (PDB code 1XSD, Z-score 5.1, sequence identity 13%), MecI (1OKR, Z-score 5.0, sequence identity

9%), and DP2 (1CF7, Z-score 4.8, sequence identity 16%), whose structures in complex with a DNA fragment have been solved by crystallography (Zheng et al. 1999; Garcia-Castellanos et al. 2004; Safo et al. 2005). In these structures, the recognition helix H3 penetrates DNA major groove to make base-specific contacts while the wing W1 interacts with the phosphodiester backbone of the minor groove to increase the binding affinity. This DNA recognition mode is typical of WH proteins (Gajiwala and Burley 2000). Superposition of the KIN17 WH domain onto the WH–DNA complex of BlaI, MecI, and DP2 reveals a remarkable structural similarity despite a low level of sequence identity (Supplemental Fig. S2). However, significant differences are observed in the region comprising the recognition helix H3 and the loop H2–H3. In KIN17, loop H2–H3 is longer by at least seven residues whereas helix H3 is about five amino acids shorter. Moreover, the presence of the additional  $3_{10}$ -helix H2.5 in KIN17 loop H2–H3 results in a substantially different positioning of the recognition helix H3 that does not superimpose on that of BlaI, MecI, and DP2. Comparison of the electrostatic potential surfaces also shows structural differences between the WH domain of KIN17 and the DNA-binding WH proteins (Fig. 3A). In all the structures of WH–DNA complexes, basic residues belonging to helix H3 and/or wing W1 form a concave positively charged surface that is generally well conserved. In the case of KIN17, the corresponding surface is rather slightly negative. Moreover, analysis of a sequence alignment of 20 KIN17 proteins using the program ConSurf reveals that the conservation of this surface is not significant ([consurf.tau.ac.il/](http://consurf.tau.ac.il/)). Considering these observations, we were convinced that the WH domain of KIN17 does not interact with DNA through the typical recognition mode of WH proteins. Southwestern blotting studies were further performed using different KIN17 fragments but no interaction was observed between the recombinant domain 51–160 and a double-stranded DNA fragment recognized *in vivo* and *in vitro* by the whole protein (data not shown).

Interestingly, three residues of helix H2 (Asn-92, Glu-97, and Ile-99) and five residues of loop H2–H3 (His-104, His-106, Asn-108, Ala-109, and Thr-114) constitute a highly conserved and neutral surface located on a face opposite to the H3–W1 face (Fig. 3B). This surface could be involved in protein–protein interactions as observed in the recently published structure of the RAP74–FCP1 complex, in which the WH domain of RAP74 uses its neutral H2–H3 surface to bind the FCP1 peptide (Nguyen et al. 2003). Since the KIN17 WH module (residues 65–143) is close to a zinc finger motif (residues 28–50), its conserved surface could also participate with intramolecular interactions with this zinc finger domain. Structural investigation of intramolecular domain interactions will now be carried out in order to provide additional clues on the function of KIN17 WH module.



**Figure 3.** (A) Electrostatic properties at the surface of WH proteins. (*Top panel*) Representation of the electrostatic potential at the surface of the WH domains of BlaI (PDB code 1XSD), MecI (1OKR), DP2 (1CF7), and KIN17. The domains are oriented so as to show the H3–W1 face. Blue, white, and red colors indicate positive, neutral, and negative electrostatic surface potential, respectively. (*Lower panel*) Ribbon representation of these structures in the same orientation as in the *top panel*. The canonical secondary structure elements are highlighted according to the following color code: dark blue, H1; light blue, S1; green, H2; yellow, H3; red, S2, S3, and W1. (B) A projection of the ConSurf results on the H2–H2.5 surface of the human KIN17 WH indicates the degree of conservation from dark blue (highly conserved) to red (variable).

## Materials and Methods

### Sample preparation

Region 51–160 of human KIN17 was overexpressed in *Escherichia coli* strain Rosetta DE3 transformed with a construct generated in pEXP-TH5 plasmid (Invitrogen). This vector encodes a 6xHis-tag, the Z-domain from Staphylococcal Protein A (ZZ), a TEV protease (Tobacco Etch Virus) cleavage site, and the KIN17 region. The recombinant protein was prepared from cells grown to an  $OD_{595\text{ nm}}$  of 1.2 at 37°C in a bacterial culture supplemented by 100  $\mu\text{g}/\text{mL}$  ampicillin and 35  $\mu\text{g}/\text{mL}$  chloramphenicol. Protein expression was then induced by addition of 1 mM isopropyl-1-thio- $\beta$ -D-galactopyranoside (IPTG), and culture was further grown for 14 h at 20°C. The cells were harvested by centrifugation, resuspended in a Tris-HCl buffer (100 mM Tris-HCl pH 8.0, 150 mM NaCl, 5% glycerol, 1 mM EDTA, 1 mM PMSF), and lysed with an Eaton press (Rassant). The fusion protein was purified using IgG Sepharose Fast Flow (Amersham) and digested on-column with recombinant (His)<sub>6</sub>-TEV protease. The cleaved KIN17 domain was further purified on nickel-nitriloacetic acid resin with a Ni-NTA column (Qiagen) to remove the (His)<sub>6</sub>-TEV protease as well as the residual (His)<sub>6</sub>-ZZ fragment. Because of the cloning strategy, the protein resulting from the cleavage contains the region 51–160 of human KIN17 plus an additional glycine residue at the N terminus. Uniformly labeled <sup>15</sup>N protein was produced in minimum medium M9 containing 1 g.L<sup>-1</sup> of (<sup>15</sup>NH<sub>4</sub>)<sub>2</sub>SO<sub>4</sub> (Boehringer) as the sole nitrogen source. Uniformly labeled <sup>13</sup>C/<sup>15</sup>N protein was produced in a rich medium prepared from uniformly labeled <sup>13</sup>C/<sup>15</sup>N *Spirulina maxima* cyanobacteria. The resulting protein was characterized by electrospray ionization mass spectroscopy. NMR samples (~0.7 mM) were prepared in 50 mM phosphate buffer (pH 6.0) containing 150 mM NaCl, 1 mM EDTA, a protease inhibitor cocktail (Sigma), 1 mM TCEP, and 1 mM NaN<sub>3</sub> in either 90% H<sub>2</sub>O/10% D<sub>2</sub>O or in 100% D<sub>2</sub>O. 3-(Trimethylsilyl)[2,2,3,3,3-<sup>2</sup>H<sub>4</sub>] propionate (TSP) was added as an internal <sup>1</sup>H chemical shift reference. <sup>13</sup>C and <sup>15</sup>N chemical shifts were referenced indirectly to TSP, using the absolute frequency ratios.

### NMR spectroscopy and structure calculation

All NMR experiments were performed at 30°C on Bruker AVANCE DMX-600 or VARIAN INOVA-800 spectrometers equipped with a triple resonance cryoprobe. <sup>1</sup>H, <sup>15</sup>N, and <sup>13</sup>C resonance assignment of human KIN17 region 51–160 has been described previously (Carlier et al. 2006). A <sup>15</sup>N-<sup>1</sup>H heteronuclear nOe experiment was carried out at 600 MHz using the classical pulse program (Farrow et al. 1994). Two spectra were acquired in an interleaved fashion with and without saturation (5s), respectively. Uncertainty in the <sup>1</sup>H → <sup>15</sup>N nOe values were calculated from the peak intensity errors estimated from the root mean square baseline noise of both spectra. The <sup>1</sup>H-<sup>1</sup>H nuclear Overhauser effect cross-peak volumes used for structure calculation were measured on three NOESY experiments (a <sup>15</sup>N-HSQC-NOESY in H<sub>2</sub>O with a 80-ms mixing time and a <sup>13</sup>C-HSQC-NOESY in D<sub>2</sub>O with a 100-ms mixing time recorded at 800 MHz, and a <sup>13</sup>C-HSQC-NOESY in the <sup>13</sup>C aromatic region recorded at 600 MHz with a 120-ms mixing time).  $\phi$  and  $\psi$  torsion angle values were deduced either from the backbone chemical shifts using the program TALOS (Cornilescu et al. 1999) or from the analysis of the HNHA and HMQC-J experiments. Hydrogen-bond restraints were derived from slowly exchanging amide protons, identified by measuring the amide proton exchange rates from <sup>1</sup>H-<sup>15</sup>N HSQC spectra recorded at different times on a protein sample dissolved in D<sub>2</sub>O (Supplemental Table S2). Oxygen atoms involved in hydrogen-bond restraints were deduced from the secondary structure determined by the TALOS program and from the  $\beta$ -sheet topology. All spectra were processed with the programs Xwinnmr (Bruker) or NMRPipe (Delaglio et al. 1995) and analyzed using Felix (Accelrys).

The solution structure of human KIN17 region 51–160 was solved on the basis of the analysis of 3413 nOe cross-peaks (1466 were picked on the <sup>15</sup>N-HSQC-NOESY, 1919 on the aliphatic <sup>13</sup>C-HSQC-NOESY, and 28 on the aromatic <sup>13</sup>C-HSQC-NOESY, respectively). The semiautomated iterative assignment procedure INCA was applied for nOe assignment and three-dimensional structure construction (Savarin et al. 2001). Structures were calculated using a simulated annealing protocol in CNS 1.0 (Brunger et al. 1998). At the last step, 400 structures were generated using a total of 2743 distance

restraints, 18 hydrogen bonds, 136  $\phi$  and  $\psi$  torsion angle values obtained from TALOS prediction, and 4  $\phi$  angle values deduced from HNHA and HMQC-J experiments. The 12 best structures were selected and refined with a standard energy function (CHARMM22), including an electrostatic energy term. This term was calculated with no net charge on the side-chain atoms and with a distance-gated dielectric constant

### Electronic supplemental material

Supplemental material includes an overview of the NMR restraints used for structure calculation of human KIN17 region 51–160 and a superimposition of this domain onto the X-ray winged helix–DNA structure of three analogous proteins.

### Acknowledgments

The authors are grateful to Dr. Adrien Favier, who kindly recorded NOESY experiments in the framework of the I.B.S Jean-Pierre Ebel national facility (Grenoble, France). We also thank the Conseil Régional de Haute-Normandie and the Centre de Ressources Informatiques de Haute-Normandie (France) for NMR and molecular modeling facilities.

### References

- Angulo, J.F., Mauffrey, P., Pinon-Lataillade, G., Miccoli, L., and Biard, D.S. 2005. Putative roles of kin17, a mammalian protein binding curved DNA, in transcription and replication. In *DNA conformation and transcription*, pp. 75–89. Landes Biosciences, Austin, TX.
- Biard, D.S., Miccoli, L., Despras, E., Frobort, Y., Creminon, C., and Angulo, J.F. 2002. Ionizing radiation triggers chromatin-bound kin17 complex formation in human cells. *J. Biol. Chem.* **277**: 19156–19165.
- Brunger, A.T., Adams, P.D., Clore, G.M., DeLano, W.L., Gros, P., Grosse-Kunstleve, R.W., Jiang, J.S., Kuszewski, J., Nilges, M., Pannu, N.S., et al. 1998. Crystallography & NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr. D Biol. Crystallogr.* **54**: 905–921.
- Carlier, L., le Maire, A., Braud, S., Masson, C., Gondry, M., Zinn-Justin, S., Guilhaudis, L., Milazzo, I., Davoust, D., Gilquin, B., et al. 2006. NMR assignment of region 51–160 of human KIN17, a DNA and RNA-binding protein. *J. Biomol. NMR* **36**(Suppl 1): 29. doi: 10.1007/s10858-006-0013-y.
- Cornilescu, G., Delaglio, F., and Bax, A. 1999. Protein backbone angle restraints from searching a database for chemical shift and sequence homology. *J. Biomol. NMR* **13**: 289–302.
- Delaglio, F., Grzesiek, S., Vuister, G.W., Zhu, G., Pfeifer, J., and Bax, A. 1995. NMRPipe: A multidimensional spectral processing system based on UNIX pipes. *J. Biomol. NMR* **6**: 277–293.
- Farrow, N.A., Muhandiram, R., Singer, A.U., Pascal, S.M., Kay, C.M., Gish, G., Shoelson, S.E., Pawson, T., Forman-Kay, J.D., and Kay, L.E. 1994. Backbone dynamics of a free and phosphopeptide-complexed Src homology 2 domain studied by  $^{15}\text{N}$  NMR relaxation. *Biochemistry* **33**: 5984–6003.
- Gajiwala, K.S. and Burley, S.K. 2000. Winged helix proteins. *Curr. Opin. Struct. Biol.* **10**: 110–116.
- García-Castellanos, R., Mallorqui-Fernandez, G., Marrero, A., Potempa, J., Coll, M., and Gomis-Ruth, F.X. 2004. On the transcriptional regulation of methicillin resistance: MecI repressor in complex with its operator. *J. Biol. Chem.* **279**: 17888–17896.
- le Maire, A., Schiltz, M., Stura, E.A., Pinon-Lataillade, G., Couprie, J., Moutiez, M., Gondry, M., Angulo, J.F., and Zinn-Justin, S. 2006. A tandem of SH3-like domains participates in RNA binding in KIN17, a human protein activated in response to genotoxics. *J. Mol. Biol.* **364**: 764–776.
- Masson, C., Mena, F., Pinon-Lataillade, G., Frobort, Y., Chevillard, S., and Radicella, J.P. 2003. Global genome repair is required to activate KIN17, a UVC-responsive gene involved in DNA replication. *Proc. Natl. Acad. Sci.* **100**: 616–621.
- Mazin, A., Milot, E., Devoret, R., and Chartrand, P. 1994a. KIN17, a mouse nuclear protein, binds to bent DNA fragments that are found at illegitimate recombination junctions in mammalian cells. *Mol. Gen. Genet.* **244**: 435–438.
- Mazin, A., Timchenko, T., Menissier-de Murcia, J., Schreiber, V., Angulo, J.F., de Murcia, G., and Devoret, R. 1994b. Kin17, a mouse nuclear zinc finger protein that binds preferentially to curved DNA. *Nucleic Acids Res.* **22**: 4335–4341.
- Nguyen, B.D., Abbott, K.L., Potempa, K., Kobor, M.S., Archambault, J., Greenblatt, J., Legault, P., and Omichinski, J.G. 2003. NMR structure of a complex containing the TFIIF subunit RAP74 and the RNA polymerase II carboxyl-terminal domain phosphatase FCPI. *Proc. Natl. Acad. Sci.* **100**: 5688–5693.
- Safo, M.K., Zhao, Q., Ko, T.P., Musayev, F.N., Robinson, H., Scarsdale, N., Wang, A.H., and Archer, G.L. 2005. Crystal structures of the Blal repressor from *Staphylococcus aureus* and its complex with DNA: Insights into transcriptional regulation of the bla and mec operons. *J. Bacteriol.* **187**: 1833–1844.
- Savarin, P., Zinn-Justin, S., and Gilquin, B. 2001. Variability in automated assignment of NOESY spectra and three-dimensional structure determination: A test case on three small disulfide-bonded proteins. *J. Biomol. NMR* **19**: 49–62.
- Zheng, N., Fraenkel, E., Pabo, C.O., and Pavletich, N.P. 1999. Structural basis of DNA recognition by the heterodimeric cell cycle transcription factor E2F-DP. *Genes Dev.* **13**: 666–674.

### 3.2. Essai de détermination de la structure de la région 1-160 par RMN

Ayant déterminé la structure du domaine K2 par RMN, nous avons cherché à déterminer la structure de la région N-terminale entière (résidus 1 à 160) en utilisant la même méthode. Sur le spectre  $^1\text{H}$ ,  $^{15}\text{N}$ -HSQC du domaine K3 (Figure 3-9), environ 120 systèmes de spins sont identifiés contre les 156 attendus (161 résidus dont 4 prolines et une glycine N-terminale). Parmi ces systèmes de spins, environ 80% correspondent à des résidus du domaine K2. Mais surtout, il apparaît que près de la moitié des résonances  $^1\text{H}$ - $^{15}\text{N}$  correspondant aux groupements amides de la région N-terminale 1-50 de K3 sont absentes sur le spectre. Les taches de corrélation du spectre  $^1\text{H}$ ,  $^{15}\text{N}$ -HSQC correspondant à des résidus propres à la protéine K3 se répartissent en deux groupes : (i) des taches de corrélation au centre du spectre dans une zone correspondant à des résidus dans une région déstructurée et (ii) des taches de corrélations réparties dans le spectre  $^1\text{H}$ ,  $^{15}\text{N}$ -HSQC mais de faible intensité suggérant la présence d'échange conformationnel à l'échelle de la milliseconde.

Afin d'obtenir les informations nécessaires à la détermination de la structure tridimensionnelle de la protéine K3 par RMN, nous avons enregistré les expériences trois noyaux suivantes : HNC0, HNCA, CBCACONH, HNCACB et HBHACONH sur un échantillon de protéine K3 doublement marquée  $^{15}\text{N}$ - $^{13}\text{C}$ . L'attribution du domaine K2 a été retrouvée à partir des données extraites de ces expériences RMN 3D et en s'aidant de l'attribution du domaine K2 seul réalisée par L. Carlier (cf article 3). Par contre, l'attribution des systèmes de spins correspondant à des résidus de la partie en amont du domaine K2 n'a pas été possible à cause du manque de données. Nous n'avons donc pas assez de données pour déterminer la structure tridimensionnelle par RMN de la région N-terminale de hKIN17 (K3). Il est possible que (i) l'extrémité 1-25 en amont du domaine prédit en doigt de zinc soit dégradée pendant les expériences ou en échange conformationnel entre une forme structurée et une forme déstructurée ou que (ii) le domaine prédit en doigt zinc soit lui-même en échange conformationnel ou que sa position par rapport au domaine *winged helix* fluctue. Une partie des réponses à ces questions pourraient être apportées par l'étude du domaine prédit en doigt de zinc seul (résidus 26-50). Les premiers essais pour préparer ce domaine ont été pour le moment infructueux. L'enregistrement d'un spectre  $^1\text{H}$ ,  $^{15}\text{N}$ -HSQC de ce domaine permettrait de savoir combien de pics de corrélation sont observables pour cette région isolée et s'il est possible de déterminer la structure de ce domaine.

### 3.3. Etude structurale de la région 1-160 par SAXS

Le but de cette étude par SAXS est de savoir si des informations sur la présence possible de région(s) repliées ou dépliée(s) et sur la flexibilité de la protéine peuvent être apportées et si un modèle pour cette région N-terminale de la protéine hKIN17 peut être proposé.

### 3.3.1. Extraction des paramètres structuraux

Des données de SAXS ont été enregistrées au LCR/EPFL (Lausanne, Suisse) en collaboration avec le Pr. M.Schiltz et à l'IBBMC (Orsay, France) en collaboration avec le Dr.P.Vachette. Les intensités de diffusion de la solution de protéine K3 ont été mesurées dans une gamme du vecteur de diffusion égale à  $0,01 \text{ \AA}^{-1} < q < 0,3 \text{ \AA}^{-1}$ , ce qui correspond dans l'espace réel à  $628 \text{ \AA} > r > 21 \text{ \AA}$ .

#### 3.3.1.1. Le rayon de giration à partir de la courbe de Guinier

Nous avons commencé par déterminer la valeur du rayon de giration de la protéine K3 à partir de la courbe de Guinier, grâce au programme PRIMUS (Konarev et al. 2003).

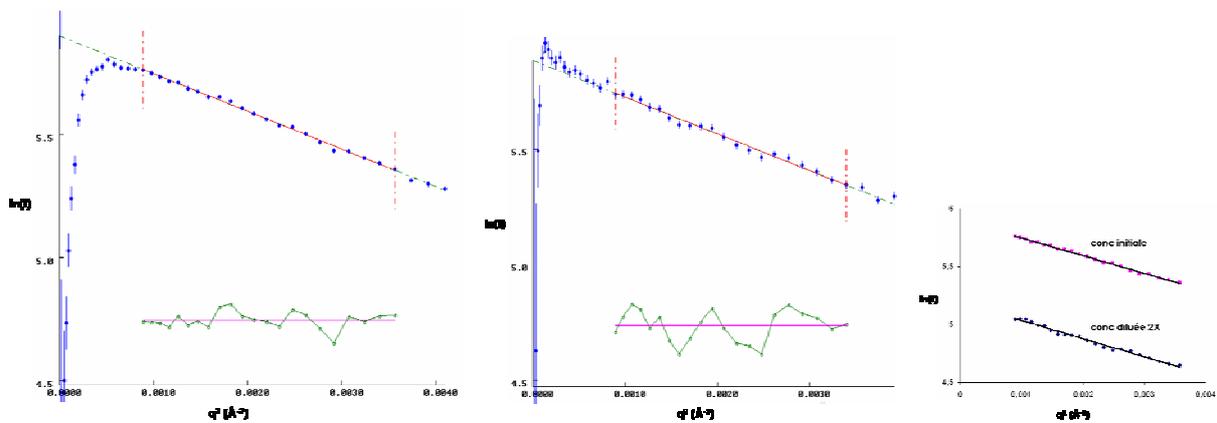


Figure 3-1 : A gauche et au centre) Courbes de Guinier pour la protéine K3 à deux concentrations différentes (concentration initiale et concentration diluée) pour les données enregistrées à Orsay et à droite) Superposition de ces deux zones de Guinier.

D'après les graphes de la Figure 3-1, la zone de Guinier ( $0.65 < qR_g < 1.3$ ) des courbes de diffusion enregistrées à Orsay sur deux concentrations de protéine K3 est linéaire (Figure 3-1-a) suggérant que dans les deux conditions, la solution de protéine est bien monodisperse et qu'il n'y a pas d'effet majeur d'interaction entre particules. De plus, les courbes aux deux concentrations ont la même pente (Figure 3-1-b), confirmant qu'il n'y a pas d'interaction intermoléculaire (pas d'attraction ou de répulsion des particules). Le rayon de giration déduit de ces courbes est de  $21.8 \pm 0.3 \text{ \AA}$  et  $21.5 \pm 0.2 \text{ \AA}$  pour les solutions de protéine à concentration initiale et diluée deux fois, respectivement. La courbe de Guinier à partir des données enregistrées à Lausanne a beaucoup moins de points. Néanmoins, une valeur du rayon de giration peut en être déduite,  $R_g = 20.8 \pm 0.5 \text{ \AA}$ . Nous pouvons donc conclure que le domaine K3 dans les conditions des expériences de SAXS a un rayon de giration autour de  $21 \pm 0.5 \text{ \AA}$ . Des protéines globulaires (myoglobine, calmoduline, élastase) de poids moléculaire proche de celui du domaine K3 ont des rayons de giration autour de  $15-17 \text{ \AA}$ , ce qui suggère que le domaine K3 n'est pas complètement globulaire.

#### 3.3.1.2. La région 1-160 est sous forme monomérique en solution

En parallèle des mesures sur la protéine K3 faites à Orsay, une mesure a été faite sur une solution de concentration connue ( $c=4.4\text{mg/ml}$ ) de lysozyme. Ainsi à partir des valeurs de  $R_g$

et  $I(0)$  (=intensité à l'origine), nous pouvons estimer la masse moléculaire de la protéine en solution. A partir de la courbe de Guinier du lysozyme, le  $R_g$  déduit ( $R_g=15.1 \pm 0.3 \text{ \AA}$ ) correspond bien à la valeur connue pour le  $R_g$  du lysozyme et la valeur de  $I(0)$  est égale à 181. Ainsi  $I(0)/c$  vaut 41. La masse moléculaire (MM) du lysozyme étant de 14.3kDa, la valeur de la constante de proportionnalité entre  $I(0)/c$  et MM est de 2.9 ( $I(0)/c=2.9$  pour  $MM=1 \text{ kDa}$ ). Pour la protéine K3, nous pouvons donc déduire :

Concentration	5.8 mg/ml	2.9 mg/ml
$I(0)$ déduite du Guinier	366	180
$I(0)/c$ calculé	64	63
masse moléculaire déduite	21.9 kDa	21.6 kDa

Les masses moléculaires déduites des mesures sur les solutions de protéine K3 à deux concentrations sont proches de la masse moléculaire théorique de la protéine en monomère (19.5 kDa). Nous pouvons donc conclure que la protéine K3 est essentiellement sous forme de monomère en solution, dans la gamme de concentration utilisée.

### 3.3.1.3. La fonction de distribution des distances interatomiques $P(r)$

Les courbes  $P(r)$  présentées sur la Figure 3-2 ont été calculées par le programme GNOM (Svergun 1992) à partir des données d'Orsay. Ce sont des courbes en forme de cloche avec un maximum pour  $R$  autour de 21-22  $\text{\AA}$ . Les meilleures solutions ont été obtenues avec un  $D_{max}$  compris entre 75 et 80  $\text{\AA}$ . C'est également ce qui a été observé avec les données de Lausanne. Les courbes  $P(r)$  calculées à partir des données de Lausanne et à partir des données d'Orsay se superposent bien (Figure 3-2).

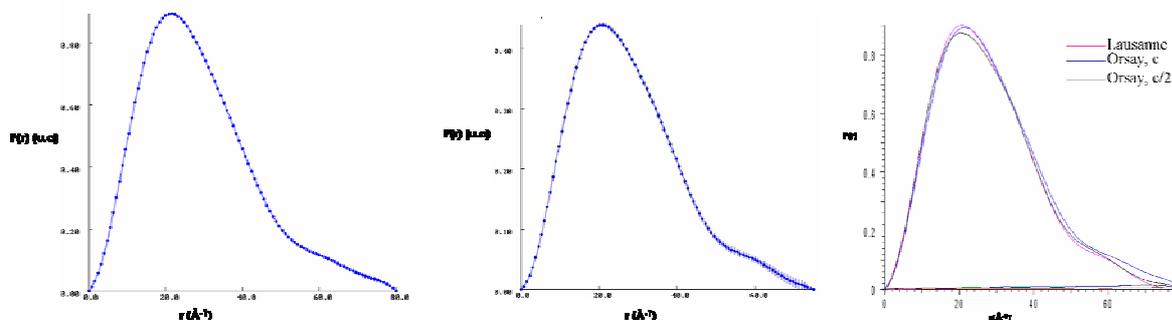


Figure 3-2 : Courbe  $P(r)$  pour la protéine K3 à deux concentrations différentes (concentration initiale à gauche et concentration diluée deux fois au centre), en utilisant les données enregistrées à Orsay. A droite, superposition des courbes  $P(r)$  calculées à partir des données de Lausanne (données à concentration initiale) et à partir des données d'Orsay.

La valeur des rayons de giration déduits de ces courbes  $P(r)$  est de  $22.7 \pm 0.2 \text{ \AA}$  et  $21.6 \pm 0.2 \text{ \AA}$  pour la protéine K3 à la concentration initiale et à la concentration diluée deux fois, respectivement. De la même manière, le rayon de giration a été déduit des courbes  $P(r)$  calculées à partir des données de Lausanne. Sa valeur est de  $21.8 \pm 0.3 \text{ \AA}$  et  $22.2 \pm 0.3 \text{ \AA}$  pour la protéine K3 à la concentration initiale et à la concentration diluée deux fois, respectivement. Ces valeurs de rayon de giration sont légèrement plus élevées que celles déduites des courbes de Guinier. Cependant, le rayon de giration déduit de la  $P(r)$  est souvent plus fiable que celui déduit de la courbe de Guinier car le calcul de  $P(r)$  tient compte de toute la courbe  $I(s)$  alors que la courbe de Guinier dépend seulement des petits angles.

En conclusion, l'analyse des courbes enregistrées à Orsay et à Lausanne montre que les données ont un bon accord entre elles et que les paramètres structuraux déduits sont très voisins. Le rayon de giration et le Dmax de la protéine K3 sont  $21.5 \pm 0.5$  Å et  $80 \pm 5$  Å respectivement et ces valeurs sont tout à fait compatibles avec une protéine de la taille de K3. Dans la suite de l'analyse des données de SAXS, afin de ne pas surcharger le manuscrit, nous ne présenterons que les résultats obtenus avec les données de Lausanne. Nous obtenons des résultats similaires avec les données d'Orsay.

### 3.3.2. Hypothèses sur la région 1-160 compatibles avec les données de SAXS

#### 3.3.2.1. Les données SAXS sont compatibles avec la présence d'un domaine globulaire dans la région N-terminale en amont du domaine *winged helix*

Par défaut, dans le programme Bunch (Petoukhov et al. 2005), les régions à prédire sont considérées comme globulaires et donc il y a une contrainte très forte sur le rayon de giration de ces régions. L'utilisateur peut relâcher cette contrainte pour voir les effets sur la qualité de l'accord entre les données expérimentales et calculées. Nous avons effectué des calculs avec et sans cette contrainte sur le rayon de giration de la région en amont du domaine *winged helix*. Nous avons observé que les courbes de diffusion des modèles contraints (poids sur  $R_g=1$ ) présentent un accord avec les données expérimentales aussi bon que celles des modèles sans contrainte (poids sur  $R_g=0$ ). L'existence d'un domaine structuré est donc compatible avec les données expérimentales.

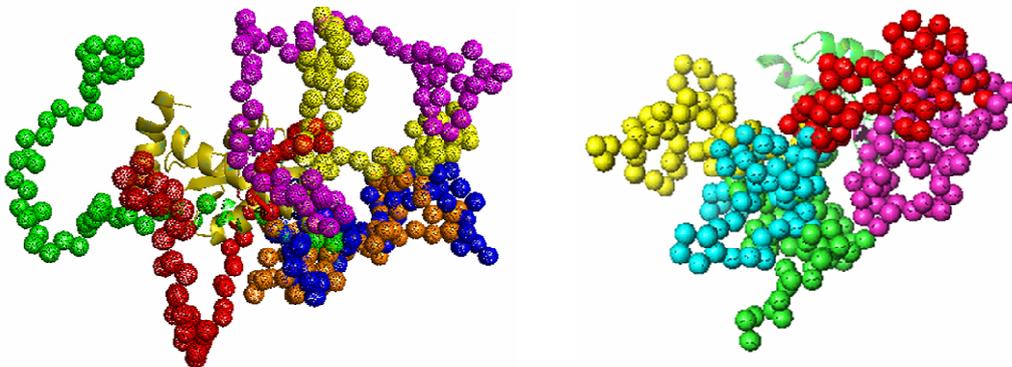


Figure 3-3 : Superposition des modèles de la protéine K3 calculés avec le programme Bunch où le seul domaine connu est le domaine *winged helix*, avec une pénalité sur le Rg de la zone à prédire égale à 0 et à 1 à gauche et à droite, respectivement. Les domaines *winged helix* de chaque modèle sont superposés et chaque modèle est représenté par une couleur.

L'effet de la contrainte sur le rayon de giration de la zone à prédire est bien visible sur la Figure 3-3. Pour les modèles où cette contrainte est relâchée, la région N-terminale de K3 est très étendue et l'espace exploré par cette région dans les différents modèles est très large. En revanche, pour les modèles où cette contrainte est forte, les zones prédites sont effectivement globulaires et plutôt distribuées sur un même côté du domaine *winged helix*.

Nous avons également montré par des mesures de dichroïsme circulaire effectuées sur les protéines K3 et K2, dans les mêmes conditions expérimentales la présence de structure secondaire dans la région N-terminale en amont du *winged helix*. En comparant les spectres de dichroïsme circulaire, le pourcentage d'hélice de la région 1-50 a pu être estimé.

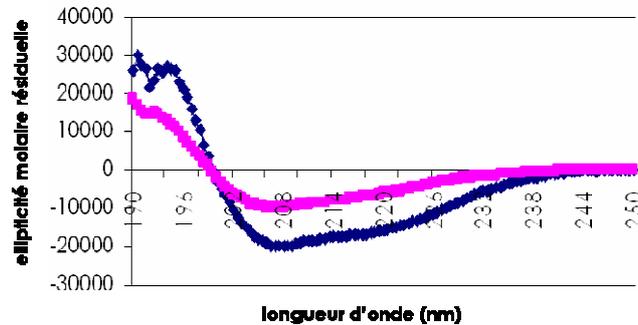


Figure 3-4 : Spectres de dichroïsme circulaire superposés des protéines K2 (rose) et K3 (bleu) enregistrés dans les mêmes conditions expérimentales.

La Figure 3-4 montre que l'amplitude du signal de la protéine K3 est bien supérieure à celle du domaine K2, la différence étant attribuée à la présence des 50 résidus supplémentaires dans K3 par rapport à K2. Sachant que l'ellipticité molaire résiduelle de 100% d'hélice théorique pour 50 résidus à 220nm est de -36320 et que la différence d'ellipticité molaire résiduelle mesurée entre K3 et K2 à 220nm est de -9636, le pourcentage d'hélice dans la structure pour les 50 résidus supplémentaires peut être déduit et est égal à 26,5%. Cette valeur n'est qu'une estimation théorique à partir de mesure de dichroïsme circulaire mais elle indique la présence de structure secondaire dans la région 1-50.

### 3.3.2.2. Les données SAXS sont compatibles avec la présence d'un doigt de zinc dans la région N-terminale en amont du domaine *winged helix*

Pour représenter la région de hKIN17 prédite en doigt de zinc, nous avons construit un modèle à partir de la structures tridimensionnelle d'un doigt de zinc avec le même motif CXXC ..... HXXXXXH. Nous avons retenu la structure obtenue par RMN du doigt de zinc de la protéine humaines '*enhancer binding protein*' (pdb : 3znf, (Omichinski et al. 1990)) qui a 20% d'identité avec la séquence du doigt de zinc de hKIN17. Ce doigt de zinc consiste en deux brins  $\beta$  antiparallèles connectés par un coude et en une hélice  $\alpha$  classique (Figure 3-5).

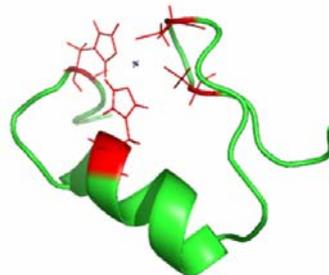


Figure 3-5 : Structure tridimensionnelle du doigt de zinc de la protéine humaines '*enhancer binding protein*' (3znf). Les résidus du motif C2H2 sont colorés en rouge et l'emplacement de l'atome de zinc est représenté par une croix noire au centre du motif.

Les résidus du fichier 3znf sont remplacés par les résidus correspondants de la séquence de hKIN17 et le modèle est affiné dans Modeller (Sali et al. 1995), essentiellement pour s'affranchir de clashes entre résidus.

Prebunch a été utilisé pour construire les modèles initiaux à partir des structures des domaines doigt de zinc et *winged helix* et en générant de manière aléatoire l'extrémité N-terminale. Ensuite, Bunch a été utilisé pour calculer les modèles de la protéine K3. Nous avons observé que, dans le cas où la pénalité sur le rayon de giration est nulle, les modélisations obtenues en donnant le domaine *winged helix* et le domaine doigt de zinc rendent compte des données expérimentales aussi bien voire mieux que celles obtenues en ne donnant que le domaine *winged helix*. En effet, les valeurs de  $\chi$ , mesurant l'accord entre les données expérimentales et la courbe de diffusion calculée à partir du modèle se situent autour de 0.92 lorsque les deux modèles globulaires sont donnés contre 1.10 lorsque le domaine *winged helix* seulement est donné. En revanche, les modèles sont équivalents en terme d'accord avec les données expérimentales lorsque la contrainte sur le rayon de giration est à 1. La courbe expérimentale s'accommode donc parfaitement de la présence d'un domaine doigt de zinc structuré dans la protéine K3.

### 3.3.2.3. L'analyse de la flexibilité de la région N-terminale en solution

Le programme EOM (Bernado et al. 2007) a été utilisé pour analyser la conformation de la protéine K3. Un calcul a été effectué en donnant la structure du domaine doigt de zinc homologue à celui prédit dans KIN17 (cf paragraphe 3.3.2.2) et du domaine *winged helix* et les régions non connues comme étant non structurées (mode *random*). Les résultats sont présentés sur la Figure 3-6. La distribution de conformations correspond à 10 000 modèles générés initialement et de façon aléatoire pour les fragments de structure non connue. L'ensemble optimisé correspond à 1 000 modèles répartis en 50 meilleurs ensembles de 20 conformations et de  $\chi$  global égal à 0.91.

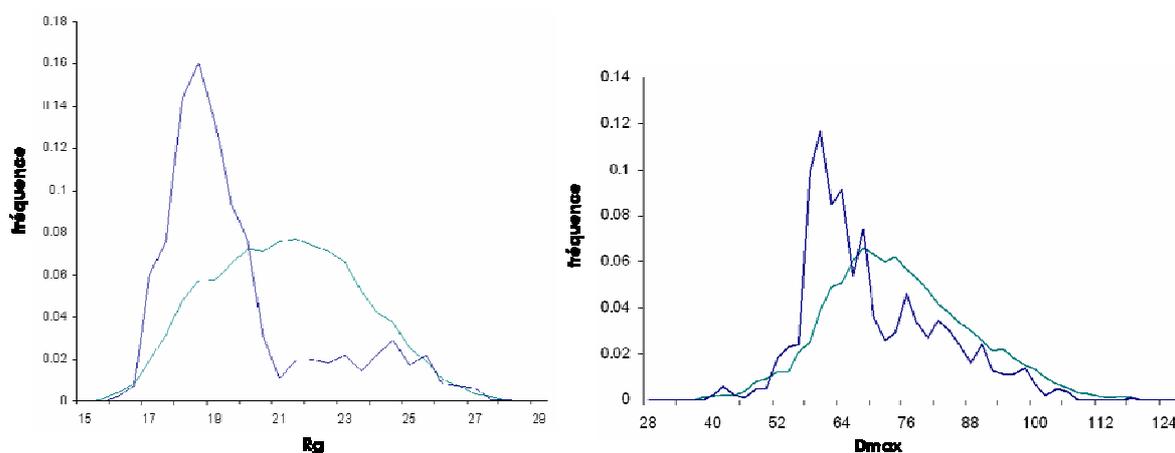


Figure 3-6 : Distribution de taille (Rg et Dmax) pour la distribution des 10 000 conformations (en vert) et pour l'ensemble optimisé de 1 000 conformations (en bleu).

Le rayon de giration de l'ensemble optimisé de conformations est déplacé vers des plus petites valeurs par rapport à celui estimé à partir de la distribution des conformations aléatoires (17-18 Å contre 21-22 Å). Il en est de même pour la distance maximale (60-65 Å

contre 70-75 Å). Nous pouvons donc en déduire que dans la protéine K3 en solution, le linker entre les deux domaines et peut-être également l'extrémité N-terminale ne sont pas complètement dépliés.

*En conclusion*, l'hypothèse de la présence d'un doigt de zinc replié dans la protéine K3 est compatible avec les données de SAXS enregistrées sur cette protéine. Cela va également dans le sens de la comparaison du signal de dichroïsme circulaire de la protéine K2 et de la protéine K3, présentée au paragraphe 3.3.2.1. Si cette hypothèse est acceptée, nous pouvons proposer un modèle assez compact des deux domaines, doigt de zinc et *winged helix*. En effet, (i) l'analyse EOM montre que le linker entre les deux domaines est plutôt compact, (ii) la fonction de distribution  $P(r)$  ne présente qu'un seul maximum. En revanche, l'extrémité N-terminale de la protéine K3 semble être plutôt mobile et désordonnée car (i) elle est sensible à la protéolyse (cf paragraphe 1.3) et (ii) des pics correspondant à des résidus déstructurés ont été observés dans le spectre  $^1\text{H}$ - $^{15}\text{N}$ , HSQC de la protéine et pourrait être attribués à cette extrémité.

### 3.3.3. Construction *ab initio* de modèles de la région 1-160

Nous avons utilisé deux méthodes de construction de modèles *ab initio* implémentées dans les programmes DAMMIN (Svergun 1999) et GASBOR (Svergun et al. 2001) pour déterminer l'enveloppe de la protéine K3 en solution. Puisque ces méthodes *ab initio* ne donnent pas un modèle tridimensionnel unique, plusieurs reconstructions indépendantes ont été effectuées. Les programmes DAMSEL et DAMSUP de la suite de programme DAMAVER (Volkov et al. 2003) ont été utilisés pour aligner les différentes reconstructions et déterminer le modèle le plus proche en moyenne (sur toutes les distances deux à deux) de tous les modèles et donc le plus représentatif.

#### 3.3.3.1. Les modèles obtenus avec le programme DAMMIN

Sept modèles de la protéine K3 ont été calculés avec le programme DAMMIN et les reconstructions ont été analysées pour proposer un modèle possible de l'enveloppe de la protéine K3.

Tableau 3-1: Résultats des analyses par DAMSEL et DAMSUP des reconstructions obtenues avec le programme DAMMIN pour la protéine K3. Les fichiers de la forme D-K3-X.pdb sont les résultats de DAMMIN et les fichiers de la forme D-K3-X-r.pdb sont les résultats de la superposition par DAMSUP.

Résultats de damsel.exe	Résultats de damsups.exe	
Mean value of NSD = 0.681	D-K3-7.pdb	Reference file
Variation of NSD = 0.030	D-K3-5-r.pdb	NSD = 0.585
Recommend to discard files with NSD > Mean + 2*Variation	D-K3-3-r.pdb	NSD = 0.651
	D-K3-1-r.pdb	NSD = 0.600
	D-K3-6-r.pdb	NSD = 0.657
	D-K3-4-r.pdb	NSD = 0.710
D-K3-7.pdb Ref // Aver NSD = 0.654		
D-K3-5.pdb Include // Aver NSD = 0.658		
D-K3-3.pdb Include // Aver NSD = 0.667		
D-K3-1.pdb Include // Aver NSD = 0.673		
D-K3-6.pdb Include // Aver NSD = 0.681		
D-K3-4.pdb Include // Aver NSD = 0.694		
D-K3-2.pdb Discard // Aver NSD = 0.742		

D'après les résultats répertoriés dans le Tableau 3-1, le modèle le plus proche des autres en moyenne (distances deux à deux) est le fichier D-K3-7.pdb ; il a la plus faible valeur de NSD (pseudo-distance moyenne). Lorsque la valeur de la NSD est inférieure à 1, les modèles sont considérés comme étant voisins. En revanche, le modèle K3-2, le plus éloigné des autres modèles, est éliminé de la comparaison.

Le programme DAMSUP effectue des translations et rotations des modèles fournis par DAMMIN pour optimiser la NSD par rapport au modèle de référence et la superposition des volumes.

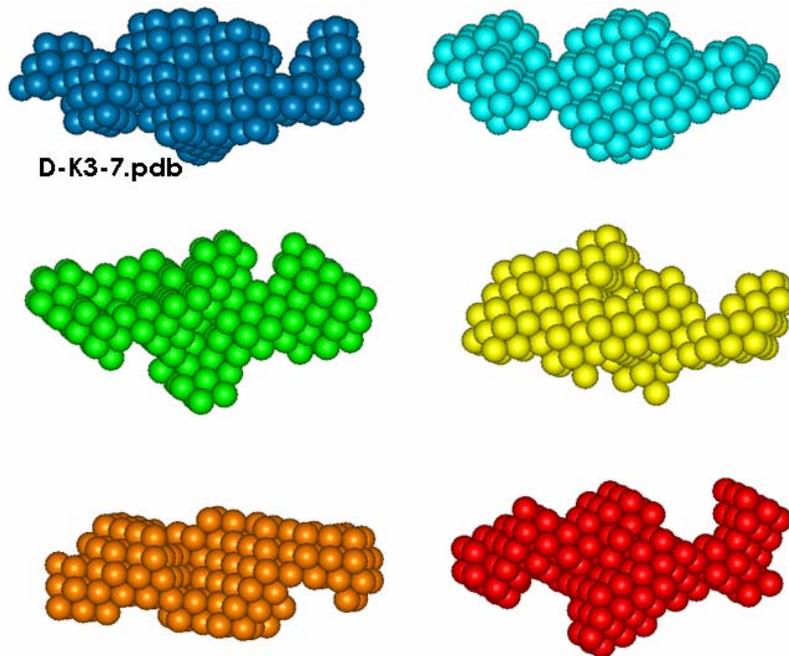


Figure 3-7 : Les 6 modèles calculés avec DAMMIN et juxtaposés après superposition par DAMSUP. D-K3-7.pdb est le modèle le plus proche en moyenne des autres modèles.

Les modèles de la protéine K3 calculés avec le programme DAMMIN ont en commun d'avoir un volume moyen allongé, un peu large dans la deuxième dimension et mince dans la dernière direction (Figure 3-7). Une structure en deux ou trois domaines pourrait être suggérée mais l'interprétation des données ne peut pas être plus précise.

### 3.3.3.2. Les modèles obtenus avec le programme GASBOR

De la même manière, sept modèles calculés avec le programme GASBOR ont été comparés et superposés pour proposer un modèle de la conformation globale de la protéine K3.

Tableau 3-2 : Résultats des analyses par DAMSEL et DAMSUP des reconstructions obtenues avec le programme GASBOR pour la protéine K3. Les fichiers de la forme G-K3-X.pdb sont les résultats de GASBOR et les fichiers de la forme G-K3-X-r.pdb sont les résultats de la superposition par DAMSUP.

Résultats de damsels.exe	Résultats de damsups.exe	
Mean value of NSD = 1.124	G-K3-5.pdb	Reference file
Variation of NSD = 0.030	G-K3-2-r.pdb	NSD = 1.063
Recommend to discard files with NSD > Mean +	G-K3-6-r.pdb	NSD = 0.993

2*Variation		G-K3-4-r.pdb	NSD = 1.133
		G-K3-0-r.pdb	NSD = 1.154
G-K3-5.pdb	Ref // Aver NSD = 1.085	G-K3-1-r.pdb	NSD = 1.084
G-K3-2.pdb	Include // Aver NSD = 1.091	G-K3-3-r.pdb	NSD = 1.080
G-K3-6.pdb	Include // Aver NSD = 1.108		
G-K3-4.pdb	Include // Aver NSD = 1.129		
G-K3-0.pdb	Include // Aver NSD = 1.142		
G-K3-1.pdb	Include // Aver NSD = 1.150		
G-K3-3.pdb	Include // Aver NSD = 1.160		

D'après les résultats répertoriés dans le Tableau 3-2, le modèle le plus proche des autres en moyenne est le fichier G-K3-5. Ici, les valeurs de NSD sont un peu plus élevées comparativement aux NSD calculés avec les modèles de DAMMIN. Ceci est en partie dû au fait que les modèles de DAMMIN utilisent des pseudo-atomes placés sur un réseau hexagonal tandis que les pseudo-résidus de GASBOR explorent l'espace sans autre restriction que celle de se trouver à une distance de 3.8 Å d'au moins un autre pseudo-résidu.

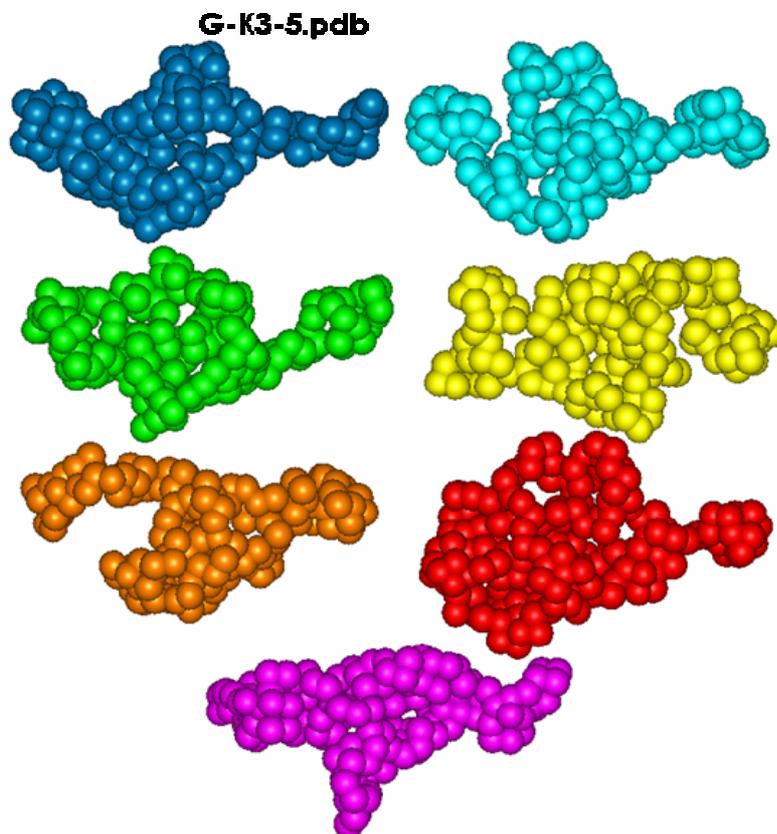


Figure 3-8 : Les 7 modèles calculés avec GASBOR et juxtaposés après superposition par DAMSUP. G-K3-5.pdb est le modèle le plus proche en moyenne des autres modèles.

Les modèles de la protéine K3 calculés avec le programme GASBOR ont encore en commun d'avoir une forme allongée et un grand domaine peut être distingué (Figure 3-8).

### 3.4. Données SAXS et RMN compatibles pour proposer un modèle de la région 1-160

Nous avons montré que la présence d'un doigt de zinc replié et qu'un modèle assez compact des deux domaines, doigt de zinc et *winged helix*, dans la protéine K3 sont compatibles avec les données de SAXS enregistrées sur cette protéine. Nous avons ensuite cherché à proposer un modèle plus précis que les modèles obtenus à partir des constructions *ab initio* de la région N-terminale en combinant les données RMN et SAXS dont nous disposons.

#### 3.4.1. Effet de la région N-terminale en amont du *winged helix* sur ce domaine

Nous avons commencé par chercher à préciser l'arrangement relatif des deux domaines (doigt de zinc et *winged helix*) dans la protéine K3. En comparant les pics de corrélation  $^1\text{H}$ ,  $^{15}\text{N}$  des résidus du domaine K2 sur les spectres  $^1\text{H}$ ,  $^{15}\text{N}$ -HSQC de K2 et K3 enregistrés dans des conditions équivalentes, nous avons proposé une zone éventuelle d'interaction entre le domaine *winged helix* et la région N-terminale en amont de ce domaine.

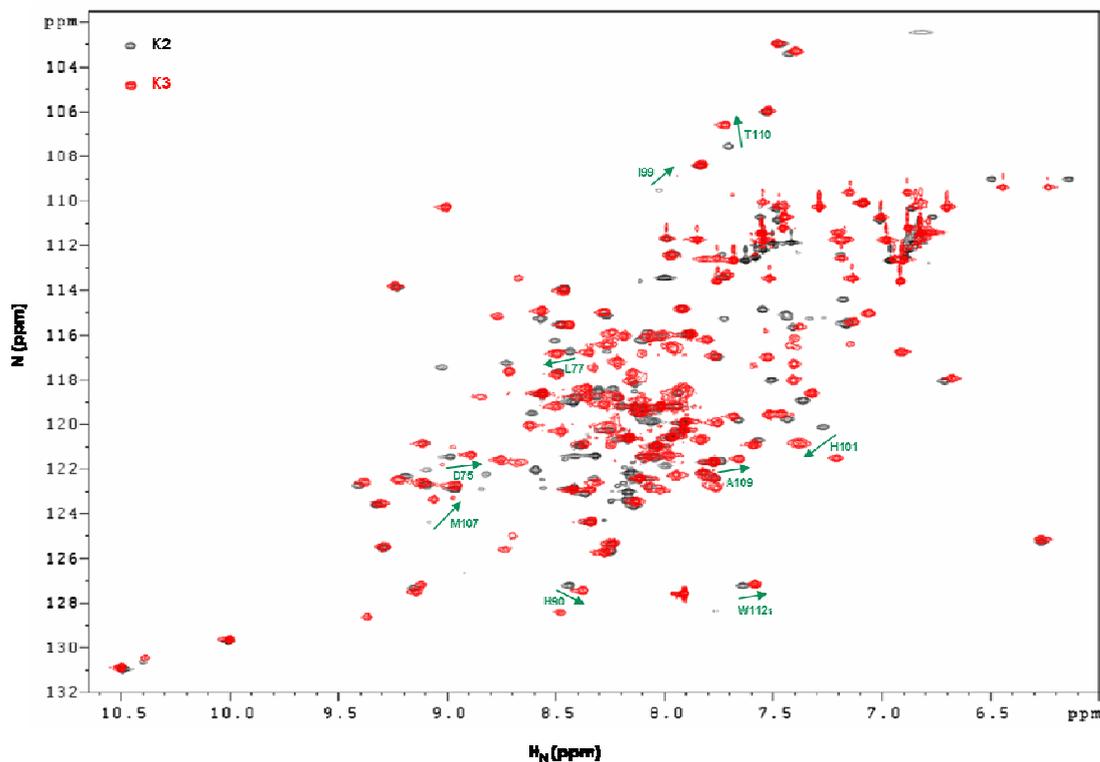


Figure 3-9 : Superposition des  $^1\text{H}$ ,  $^{15}\text{N}$ -HSQC des protéines K2 et K3, enregistrés dans des conditions équivalentes à 600 MHz. Les résidus qui présentent une variation significative de déplacements chimiques  $^1\text{H}$  et  $^{15}\text{N}$  sont mis en évidence ( $\Delta\delta_m > 0.7$  ppm où  $\Delta\delta_m = |\Delta\delta_{\text{H}_\text{N}}| + |\Delta\delta_{\text{H}}/10|$ ).

Comme le montre la superposition des spectres  $^1\text{H}$ ,  $^{15}\text{N}$ -HSQC (Figure 3-9), la plupart des pics de corrélation relatifs aux résidus de K2 n'ont subi aucune ou une très faible modification des déplacements chimiques  $^1\text{H}$  et  $^{15}\text{N}$ . Nous pouvons donc en déduire que le domaine *winged*

*helix* a la même structure qu'il soit seul ou dans le domaine K3. Cependant, les pics de corrélation correspondant aux résidus de l'extrémité N-terminale (51-67) du domaine K2 ont disparu ou ont été fortement déplacés dans le spectre  $^1\text{H}$ ,  $^{15}\text{N}$ -HSQC de la protéine K3. Cela est attendu car ces résidus correspondent au linker entre le domaine K2 et la région en amont et donc leur environnement chimique est fortement perturbé dans K3 par rapport à K2. D'autres pics de corrélation sont déplacés significativement ( $\Delta\delta_m > 0.7$  ppm où  $= |\Delta\delta\text{H}_\text{N}| + |\Delta\delta\text{N}_\text{H}/10|$ ) et correspondent à des résidus de la boucle H2-H3 et de l'hélice  $3_{10}$ . Il s'agit des résidus I99, H101, M107, A109 et T110. Des modifications d'intensité de pic sont également observées au niveau de la boucle H2-H3. Ainsi, les pics de corrélation  $^1\text{H}$ - $^{15}\text{N}$  correspondant aux résidus E103, H104, I105, H106, et N108, et qui appartiennent à des zones dégagées du spectre HSQC de K2, n'ont pas été retrouvées sur le spectre HSQC de K3. Ainsi la région du *winged helix* qui comprend ces résidus (Figure 3-10) soit change de conformation lorsque le winged helix est accompagné du reste de la région N-terminale (résidus 1-51), soit est en interaction avec la région N-terminale en amont. Il est intéressant de noter que cette région correspond à la région très conservée (Figure 3-10) et hydrophobe du *winged helix*.

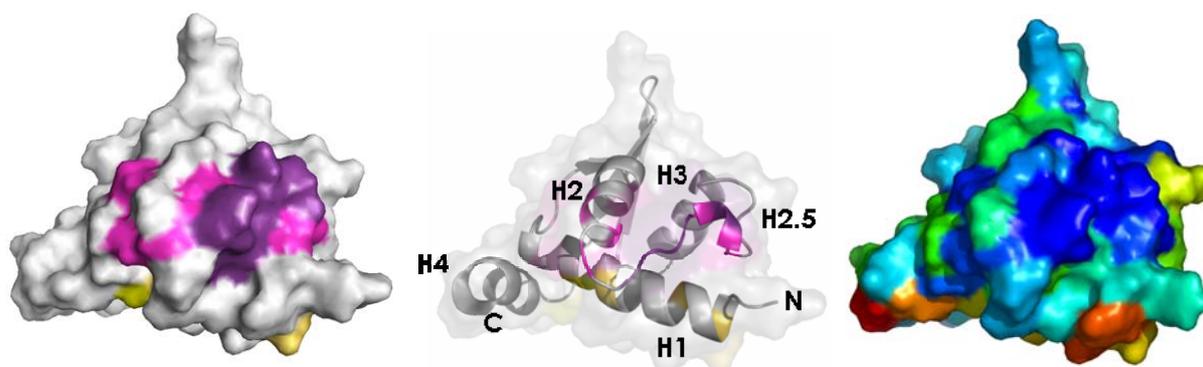


Figure 3-10 : Mise en évidence de l'effet de l'ajout de la région N-terminale en amont du domaine *winged helix* sur celui-ci. Les résidus dont la modification moyenne de déplacements chimiques  $\Delta\delta_m$  est significative sont colorés en violet sur la surface du *winged helix*. Les résidus dont le pic de corrélation n'a pas été retrouvé sont colorés en rose sur la surface du *winged helix*. Les résidus de l'hélice H1 dont la valeur de  $\Delta\delta_m$  est significative sont colorés en jaune.

Des modifications significatives de déplacement chimique apparaissent également au niveau de l'hélice H1 (résidus D75, L77 et E78), sur une surface opposée à la surface décrite précédemment (Figure 3-10). Nous proposons que la variation de déplacement chimique de ces résidus soit plutôt due à une modification structurale de l'hélice H1 dans K3 ou des résidus du linker en amont de cette hélice.

*En conclusion*, le domaine *winged helix* subit des changements conformationnels après l'addition de la région N-terminale de KIN17 : au niveau du linker entre les deux régions et dans la région très conservée du *winged helix*. Cela donne des indications sur un positionnement possible de région N-terminale, incluant le domaine prédit en doigt de zinc, par rapport au domaine *winged helix*.

### 3.4.2. Calcul de modèles de la région 1-160

Nous avons commencé par introduire les contraintes RMN pour optimiser la position relative du domaine doigt de zinc et du domaine *winged helix*. Dans cette première étape, nous utilisons la structure du doigt de zinc homologue à celui prédit dans hKIN17 (paragraphe

3.3.2.2) et la structure RMN du domaine *winged helix*. Les domaines sont gardés indépendants afin de permettre un grand échantillonnage des orientations relatives. En revanche, une contrainte de distances entre le C-terminal du doigt de zinc et le N-terminal du *winged helix* est introduite dans le programme CHARMM (Brooks et al. 1983). Les contraintes RMN ont été déduites des variations de déplacement chimique des résidus I99, H101, M107, A109 et T110 du domaine *winged helix* lorsque la région N-terminale en amont du domaine *winged helix* lui est ajoutée (paragraphe 3.4.1). Ces contraintes ont été introduites sous forme de contraintes ambiguës dans le programme CHARMM en deux temps : (1) affinement avec la contrainte que chaque C $\alpha$  des résidus I99, H101, M107, A109 et T110 du *winged helix* soit à une distance maximale de 5 Å du C $\alpha$  d'un résidu quelconque du doigt de zinc et (2) raffinement avec la contrainte qu'un atome quelconque des résidus I99, H101, M107, A109 et T110 du *winged helix* soit à une distance maximale de 3 Å d'un atome quelconque d'un résidu quelconque du doigt de zinc. De cette manière, 9 000 structures initiales ont été générées puis optimisées par dynamique moléculaire sous contraintes.

Parmi ces structures, les 18 structures d'énergie de contrainte minimum ont été sélectionnées. A partir de ces structures initiales, le linker et l'extrémité N-terminale de la protéine K3 ont été générées pour construire des modèles de la protéine K3 compatibles avec les données de SAXS. Pour cela, le programme BUNCH a été utilisé en gardant les positions des domaines doigt de zinc et *winged helix* fixes.

Nous avons trié les modèles obtenus en partitionnant l'ensemble selon les valeurs de la distance quadratique moyenne (rmsd) entre les doigts de zinc de chaque modèle. Deux modèles appartiennent à un même groupe si les deux doigts de zinc diffèrent de moins de 2 Å. Ainsi six groupes de modèles ont été obtenus (Figure 3-11).

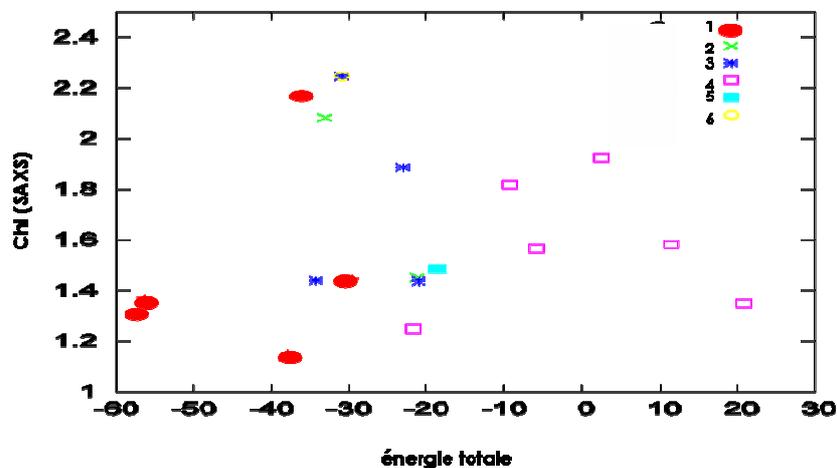


Figure 3-11 : Répartition en fonction de la valeur de l'énergie totale et de l'accord avec les données de SAXS (chi) des 18 structures calculées après introduction des contraintes RMN et SAXS. Les structures sont également regroupées en groupes (1 à 6) et une couleur est associée à chaque groupe.

Ainsi, nous observons que le groupe de modèles 1 constitue le meilleur cluster puisque quatre modèles parmi les cinq qui le constituent ont à la fois un bon accord avec les données de SAXS (chi faible) et une énergie totale faible. Ces quatre structures sont superposées sur la Figure 3-12.

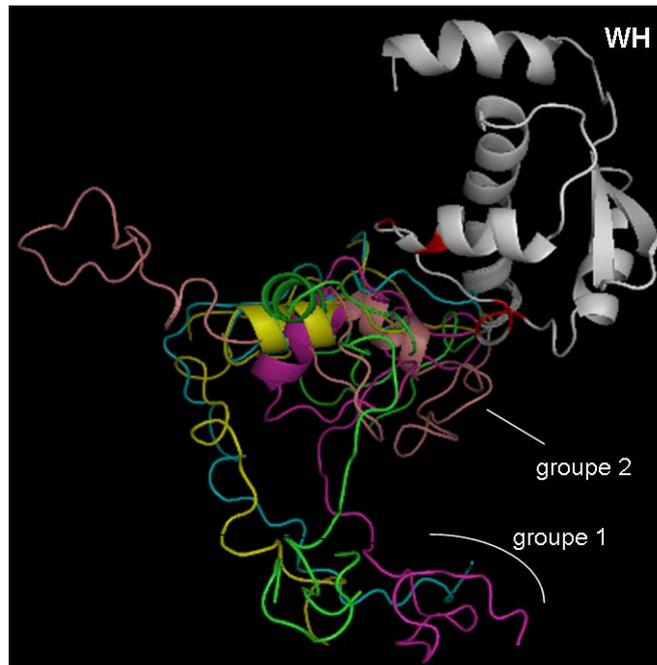


Figure 3-12 : Superposition des quatre meilleures structures du groupe 1 et de la meilleure structure du groupe 2. Les domaines *winged helix* (WH) sont superposés et représentés en blanc, exceptés les résidus I99, H101, M107, A109 et T110 colorés en rouge.

Nous observons que les modèles sont différents bien qu'ils vérifient tous les contraintes de la RMN et du SAXS. L'extrémité N-terminale, qui apparaît toujours tournée vers l'extérieur probablement pour satisfaire la contrainte de rayon de giration, explore un grand espace. La mobilité et flexibilité observées dans cette région sont confirmées par la tendance de cette région à être protéolysée. Les doigts de zinc des différents modèles adoptent différentes orientations par rapport au *winged helix* mais pour les structures du groupe 1, les doigts de zinc semblent tous orientés perpendiculairement à région de K2 qui comprend les résidus I99, H101, M107, A109 et T110 (boucle H2-H3 et hélice 3<sub>10</sub>) alors que dans le groupe 2, les doigts de zinc ont plutôt une orientation parallèle.

*En conclusion*, par cette analyse, nous avons montré que les contraintes déduites des expériences RMN et que les contraintes issues des données SAXS sont compatibles. L'analyse de ces données nous a permis de proposer une distribution de modèles voisins pour la protéine K3. Pour privilégier un modèle plutôt qu'un autre, en particulier une orientation du doigt de zinc par rapport au *winged helix*, des contraintes supplémentaires sont nécessaires. La taille de cette protéine K3 ainsi que la présence de régions flexibles et sensibles à la protéolyse limitent également la quantité d'information fournie par les données de SAXS.

## 4. VERS LA STRUCTURE DE LA PROTEINE ENTIERE

En parallèle de l'étude structurale des différents domaines de la protéine hKIN17, nous étions intéressés par la détermination de la structure de la protéine entière.

Le Dr L. Miccoli du LGR/DRR/CEA Fontenay-aux-Roses a purifié la protéine hKIN17 utilisée dans ces études. J'ai réalisé les mesures de DLS au Laboratoire de Virologie du CNRS/Gif-sur-Yvette dirigé par S. Bressanelli. J'ai réalisé l'étude de la protéine entière hKIN17 par SAXS en allant travailler avec le Pr. M. Schiltz du LCR/EPFL (Lausanne, Suisse) et avec le Dr. P. Vachette de l'IBBMC d'Orsay.

### 4.1. Absence d'interaction entre les régions N- et C-terminales

Des interactions éventuelles entre les différents domaines de la protéine hKIN17 (K1, K2 et K3) ont été étudiées par des expériences biochimiques de *pull-down* : un des domaines de la protéine est immobilisé sur une colonne et un autre domaine est ajouté et incubé avec le premier. Après lavage, s'il y a interaction, le second domaine reste retenu par le premier et sinon il est éliminé.

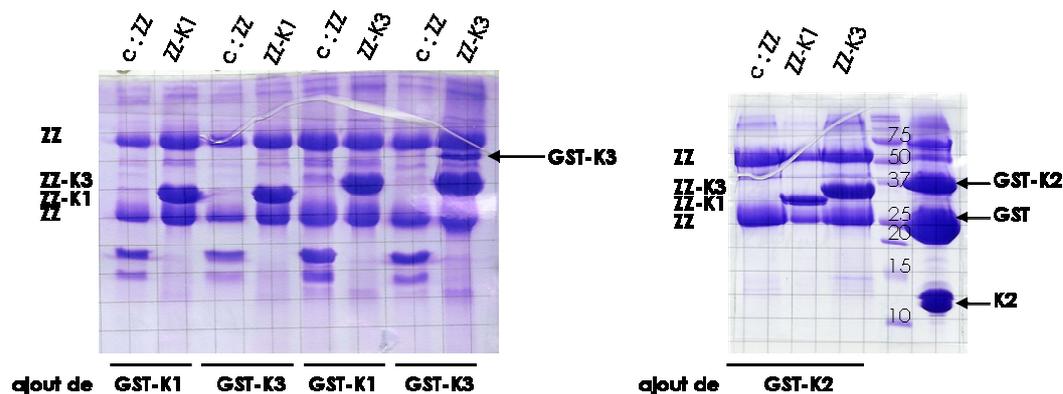


Figure 4-1 : Etude d'interaction entre les différents domaines de la protéine hKIN17, sur gel 15% SDS-PAGE coloré au bleu de Coomassie. Les domaines K1 et K3 fusionnés à l'IgG sont accrochés sur des billes d'IgG-sepharose. Les contrôles consistent en la protéine de fusion IgG accrochée sur des billes d'IgG-sepharose. Les protéines GST-K1, GST-K2 et GST-K3 sont ajoutés à chacune des colonnes. Les résultats présentés ici sur gel 15% SDS-PAGE coloré au bleu de Coomassie correspondent aux protéines retenues après lavage.

D'après les résultats présentés sur la Figure 4-1, nous pouvons conclure que les différents domaines de la protéine KIN17 n'interagissent pas entre eux. En revanche, le domaine K3 en fusion à la GST est retenu sur la colonne chargée d'IgG-K3 et pas sur la colonne contrôle avec les IgG seules. Ce domaine K3 ne retient pas non plus le domaine K2 en fusion à la GST. Nous avons également confirmé ce résultat en observant une interaction entre le domaine K3 seul et le domaine K3 fusionné à une étiquette ZZ et immobilisé sur colonne IgG. Cela suggère que la région N-terminale de KIN17, en particulier les 50 premiers résidus comprenant le doigt de zinc, est une région potentielle de dimérisation *in vitro*.

*En conclusion*, les études biochimiques d'interaction entre les domaines suggèrent que la protéine KIN17 adopte une conformation étendue avec une absence d'interaction entre les régions C- et N-terminales.

## 4.2. Préparation de la protéine entière pour les études structurales

La protéine hKIN17 entière a été purifiée selon le protocole décrit dans (Miccoli et al. 2002). La protéine avec six histidines ajoutées en N-terminal est purifiée à partir de cellules d'insectes sf9 infectées par un baculovirus codant pour cette protéine. En fin de purification, la protéine (MW=46kDa, pI=9.01) est pure à environ 90%, dans un tampon 20mM phosphate pH 7.4, 500mM NaCl, 10% glycérol. Avant de le passer en filtration sur gel, un mélange d'inhibiteurs de protéases (1mM AEBSF, 5µg/ml aprotinine, 1µM bestatine, 1µg/ml leupeptine, 1µg/ml pepstatine, 1µM phosphoramidon, 1µg/ml antipain, 1mM benzamidine, 1µM E64) a été ajouté car la protéine KIN17 est sensible à la protéolyse.

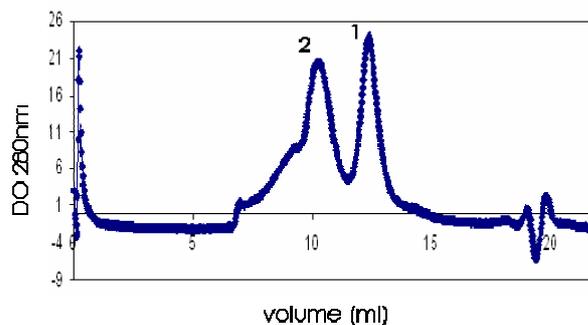


Figure 4-2 : Gel filtration de la protéine hKIN17 après purification et après centrifugation de 10 minutes à 10 000g, sur une colonne Superdex 200 de 24 ml (Amersham) dans un tampon 20 mM phosphate pH 7.4, 500 mM NaCl.

Une filtration sur gel de la préparation de protéine hKIN17 (Figure 4-2) montre deux pics (PM apparent du pic 2 = 307 000 kDa ( $V_e = 10.21$ ) et PM apparent du pic 1 = 105 000 kDa ( $V_e = 12.36$ )) suggérant la coexistence de deux formes en solution. Le volume d'éluion attendu sur cette colonne pour hKIN17 d'après sa masse moléculaire et en faisant l'hypothèse d'une protéine globulaire est de 14ml. Nous proposons que la protéine hKIN17, probablement non globulaire en solution, existe sous la forme de monomère et de dimère en solution.

Les études structurales de la protéine KIN17, par radiocristallographie ou par SAXS, nécessitent une préparation homogène et monodisperse de la protéine c'est-à-dire dans laquelle KIN17 n'existe que sous une forme. Faisant l'hypothèse que le dimère de KIN17 se forme grâce à des ponts disulfures entre deux monomères, au niveau de la région doigt de zinc d'après les résultats du paragraphe précédent, nous étudions l'effet d'un ajout d'une grande quantité de réducteur (10mM DTT) dans la solution de protéine.

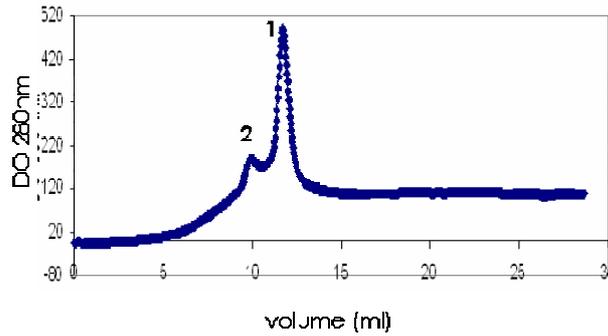


Figure 4-3 : Gel filtration de la protéine hKIN17 après purification et ajout de 10mM DTT dans l'échantillon et après centrifugation de 10 minutes à 10 000g, sur une colonne Superdex 200 de 24 ml (Amersham) dans un tampon 20 mM phosphate pH 7.4, 500 mM NaCl, 10mM DTT.

Une filtration sur gel de la préparation de protéine hKIN17 à laquelle 10mM DTT ont été ajoutés (Figure 4-3) montre que le pic correspondant à la protéine en dimère (pic 2,  $V_e = 9.94$  ml) disparaît quasiment. Les fractions correspondant au pic 1, autrement dit à la protéine KIN17 monomérique, sont concentrées jusqu'à 2mg/ml.

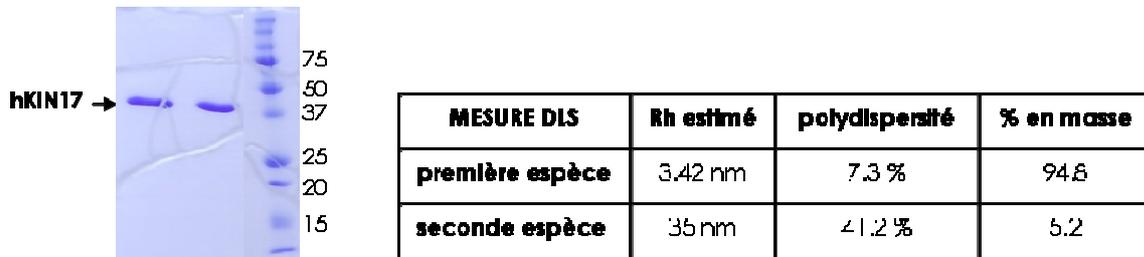


Figure 4-4 : A gauche, protéine concentré à 2mg/ml déposée sur gel SDS-PAGE 12%, coloré au bleu de Coomassie. A droite, résultats des mesures de DLS sur le même échantillon.

Après cette étape de filtration sur gel, la protéine concentrée paraît très pure puisqu'il n'apparaît qu'une bande à la taille attendue (autour de 46kDa) sur un gel SDS-PAGE (Figure 4-4). De plus, la monodispersité de l'échantillon a été vérifié par DLS. Il y a bien une espèce très majoritaire, de rayon hydrodynamique estimé à 3.42nm correspondant à une protéine globulaire de masse moléculaire estimée autour de 60 kDa, et la polydispersité de l'échantillon est très faible. Par contre, l'état d'oligomérisation de la protéine ne peut pas être déterminé avec certitude avec des mesures de DLS puisque différents facteurs, tels qu'une tendance à l'agrégation ou une protéine de forme allongée, conduisent à un Rh plus élevé que celui d'une protéine globulaire de la même masse. Les mesures de DLS ne donnent qu'une limite supérieure de l'état d'oligomérisation de la protéine. Ainsi, ces résultats confirment les résultats de la filtration sur gel, en faveur d'un monomère de KIN17 dans la solution additionnée de DTT. Nous avons également vérifié que l'échantillon reste sous forme monomérique en le repassant sur une filtration sur gel.

### 4.3. Enregistrement de données de SAXS sur la protéine entière

L'obtention de cristaux de la protéine entière hKIN17 est apparue difficile du fait de la présence d'une grande région déstructurée et sensible à la protéolyse au centre de la protéine (paragraphe Prédiction des régions déstructurées 1.1 et 1.3).

Par contre, puisque nous disposons de la structure des différents domaines de la protéine, nous avons plutôt privilégié une étude par SAXS de la protéine hKIN17 entière. Selon la même approche que l'étude de la région K3 de la protéine, le profil de diffusion en solution de la protéine multi-domaine peut être analysé en utilisant les modèles des structures à haute résolution des domaines avec l'hypothèse que leur structure tertiaire est conservée. Ce type d'approche a été utilisé avec succès pour l'étude de différentes protéines modulaires : par exemple il a été montré à partir de données de SAXS que la protéine Btk, composée de quatre domaines globulaires (SH3, SH2, kinase et PH) adopte une conformation étendue en solution (Marquez et al. 2003), de même pour la protéine PCPE constituée de deux domaines CUB et d'un domaine BTR (Bernocco et al. 2003).

Une première série de mesures a été effectuée sur un échantillon de protéine KIN17 entière concentrée à 2 mg/ml dans un tampon à 500mM NaCl. Des courbes de diffusion ont été enregistrées à concentration initiale et à concentration diluée deux fois à température ambiante et pendant 3 heures sur l'appareil de laboratoire de l'EPFL.

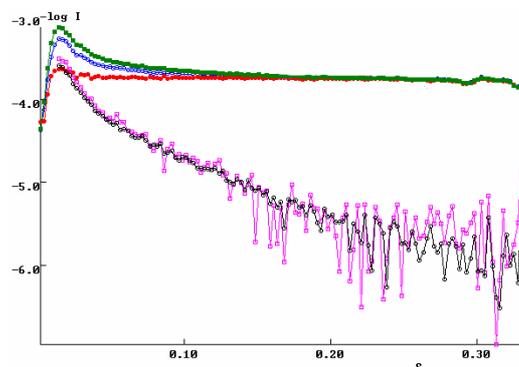


Figure 4-5 : Superposition des courbes de diffusion avant soustraction de la protéine KIN17 entière à la concentration initiale en vert et diluée deux fois en bleu à droite et du tampon correspondant en rouge et superposition des courbes de diffusion après soustraction et correction par la concentration de la protéine KIN17 entière à la concentration initiale en noire et diluée deux fois en rose. Ces données ont été enregistrées à l'EPFL, pendant une durée de 3 heures.

Les courbes de diffusion après soustraction et correction par la concentration de la protéine KIN17 entière à la concentration initiale et diluée deux fois se superposent bien (Figure 4-5). A partir de la zone du Guinier, nous pouvons estimer le rayon de giration de la protéine à 63Å pour la courbe à 1mg/ml et à 61Å pour 2 mg/ml. Cette valeur de  $R_g$  est trop élevée pour une protéine de 45kDa, même si l'on fait l'hypothèse que la protéine adopte une conformation étendue en solution. Cette valeur élevée de  $R_g$  peut plutôt être attribuée à la présence d'oligomères de la protéine en solution, suggérant que dans ces conditions expérimentales, la protéine est fortement associée.

Une deuxième série de mesures a été effectuée sur un échantillon de protéine KIN17 entière concentrée à 2 mg/ml dans un tampon à 150mM NaCl. Des courbes de diffusion ont

été enregistrées à concentration initiale et à concentration diluée deux fois à 4°C grâce au système de refroidissement du capillaire installé sur l'appareil de laboratoire de l'EPFL.

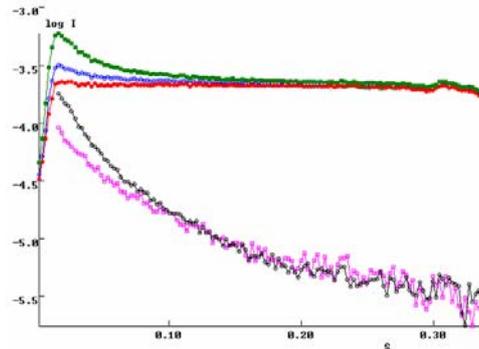


Figure 4-6 : Superposition des courbes de diffusion avant soustraction de la protéine KIN17 entière à la concentration initiale en vert et diluée deux fois en bleu à droite et du tampon correspondant en rouge et superposition des courbes de diffusion après soustraction et correction par la concentration de la protéine KIN17 entière à la concentration initiale en noire et diluée deux fois en rose. Ces données ont été enregistrées à l'EPFL.

Cette fois-ci, les deux courbes après soustraction et correction par la concentration ne se superposent plus aux petits angles (Figure 4-6). Par conséquent, les rayons de giration estimés à partir de la zone de Guinier de ces courbes sont différents :  $R_g$  estimé à 34Å pour la courbe à 1mg/ml et  $R_g$  estimé à 50Å pour la courbe à 2mg/ml. Ici, nous voyons que le rayon de giration change avec la concentration de la protéine. La dépendance du rayon de giration en concentration suggère qu'il y a, en solution, un équilibre d'association de la protéine.

*En conclusion*, nous avons observé que la protéine a une tendance à s'agréger, d'autant plus que la force ionique du tampon est élevée. Avec les données actuelles, nous n'avons pas pu calculer de fonction  $P(r)$  et de modèle de la protéine. Avant de poursuivre avec de nouveaux enregistrements de données de SAXS, nous avons à mener un travail de physico-chimie pour voir si en particulier il est possible d'éliminer l'auto association de la protéine entière en diminuant la concentration en sel ou en jouant sur le pH du tampon de la protéine. Des mesures de DLS peuvent être utilisées pour cette étude.

---

Dans cette première partie des résultats, nous avons montré que la protéine hKIN17 est constituée de deux régions globulaires, en N- et C-terminal, séparées par une grande région flexible. Cette organisation modulaire de la protéine nous a incités à aborder son étude structurale en cherchant à déterminer la structure tridimensionnelle de chacune des régions globulaires. Après avoir mis au point les conditions de production et de purification des différents domaines de la protéine hKIN17, nous avons montré qu'elle est constituée de trois domaines dont nous avons déterminé la structure : un domaine prédit en doigt de zinc et un domaine *winged helix* en N-terminal et un double domaine de type SH3 en C-terminal.

L'enjeu de la seconde partie du travail de thèse est d'établir une relation entre la structure atomique des domaines de la protéine KIN17, les interactions qu'ils peuvent avoir avec des partenaires protéiques ou acides nucléiques et le rôle de ces interactions dans la fonction globale de la protéine KIN17 dans la cellule. A partir de la structure tridimensionnelle des domaines, il s'agit de rechercher quels sont les fonctions, les partenaires et les mécanismes d'action associés à ce type de repliement dans d'autres protéines et de déterminer si les domaines possèdent les propriétés structurales requises pour ces fonctions. Ainsi,

(1) les doigts de zinc sont des motifs très communément trouvés en interaction avec les acides nucléiques et en particulier l'ADN. Dans le cadre de la thèse, nous montrons que la région N-terminale comprenant le doigt de zinc est impliquée dans l'interaction de la protéine hKIN17 avec l'ADN. Par contre, la présence d'un motif KOW (Kyrpides et al. 1996) dans la région C-terminale a orienté notre travail vers la caractérisation de l'interaction de ce domaine avec l'ARN.

(2) dans certaines protéines, des doubles domaines en tonneau  $\beta$  de type SH3 sont impliqués dans la reconnaissance de lysines modifiées d'histones. Ainsi, le double domaine Tudor de 53BP1 reconnaît l'histone H4 lorsque la lysine 20 est diméthylée (Charier et al. 2004; Botuyan et al. 2006) et cette interaction est nécessaire pour la relocalisation de la protéine aux sites de cassures double brin de l'ADN endommagé. Le double Tudor de JMJD2A est responsable de l'adressage de la protéine vers les régions de la chromatine enrichies en méthylation des lysines 4 de l'histone H3 et 20 de l'histone H4 impliqué dans la régulation de la transcription (Huang et al. 2006). Les deux domaines Chromo de la protéine CHD1 coopèrent pour lier l'histone H3 méthylée sur la lysine 4 (Flanagan et al. 2005). Ces exemples de double domaine de type SH3 capables de reconnaître des modifications épigénétiques d'histones nous ont encouragés à tester si le domaine C-terminal de hKIN17 pouvait également être impliqué dans ce type d'interaction.

(3) nous avons également cherché à identifier des partenaires associés aux domaines K1 et K2 de la protéine hKIN17. En effet, le domaine *winged helix* de la protéine KIN17 ne semble pas avoir les caractéristiques structurales observées pour les domaines *winged helix* qui interagissent avec l'ADN mais il possède une interface hydrophobe très conservée qui pourrait être impliquée dans des interactions protéine-protéine intramoléculaires ou intermoléculaires.

(4) par ailleurs, la protéine hKIN17 ayant été observée associée à un complexe de 600 kDa dans des extraits nucléaires de cellules HeLa (Miccoli et al. 2005), nous avons cherché à isoler des complexes associées à la protéine hKIN17.

(5) ayant identifié des protéines du spliceosome dans les travaux (3) et (4), nous avons cherché à préciser le rôle de la protéine hKIN17 dans les mécanismes d'épissage des ARN pré-messagers.

<b>1. LES DOMAINES IMPLIQUES DANS L'INTERACTION DE LA PROTEINE AVEC LES ACIDES NUCLEIQUES .....</b>	<b>137</b>
1.1. INTERACTION DE LA PROTEINE AVEC L'ADN .....	137
1.1.1. Domaines impliqués dans l'interaction de la protéine avec l'ADN .....	137
1.1.2. Reconnaissance d'ADN courbe et non courbe .....	139
1.1.3. Peut-on caractériser au niveau moléculaire l'interaction de la région N-terminale avec l'ADN ?	140
1.2. INTERACTION DE LA PROTEINE AVEC L'ARN.....	142
<b>2. L'INTERACTION DU DOMAINE C-TERMINAL AVEC LES HISTONES .....</b>	<b>144</b>
2.1. LIAISON DU DOMAINE C-TERMINAL AUX HISTONES H3 ET H4 .....	144
2.2. LIAISON DU DOMAINE C-TERMINAL A L'HISTONE H3 MODIFIEE.....	147
<b>3. LA RECHERCHE DE PARTENAIRES PROTEIQUES ASSOCIES AUX DOMAINES N- ET C-TERMINAUX .....</b>	<b>150</b>
3.1. MISE AU POINT DU GST-PULLDOWN AVEC DES EXTRAITS MARQUES AU <sup>35</sup> S .....	150
3.2. PULL-DOWN A GRANDE ECHELLE ET COUPLE A L'IDENTIFICATION DES PROTEINES PAR SPECTROMETRIE DE MASSE .....	151
<b>4. L'ISOLEMENT DE COMPLEXES ASSOCIES AU DOMAINE C-TERMINAL .....</b>	<b>155</b>
4.1. ASSOCIATION DU DOMAINE C-TERMINAL A UN COMPLEXE DE HAUT POIDS MOLECULAIRE.....	155
4.2. ASSOCIATION DU DOMAINE C-TERMINAL A UN COMPLEXE DE P <sub>I</sub> GLOBAL ACIDE .....	157
4.3. COMBINAISON DE LA FILTRATION MOLECULAIRE ET DU GEL ND-IEF.....	159
4.4. IDENTIFICATION DES PROTEINES PRESENTES DANS LE COMPLEXE ASSOCIE AU DOMAINE C-TER.....	159
4.4.1. Isolement du ou des complexes.....	159
4.4.2. Immunoprécipitation et identification .....	161
<b>5. EST-CE QUE LA PROTEINE JOUE UN ROLE DANS L'EPISSAGE DES ARN PRE-MESSAGERS ? .....</b>	<b>164</b>
5.1. PULL-DOWN DE REACTION D'EPISSAGE.....	164
5.2. PULL-DOWN DES SNRNPS.....	165
5.3. TEST D'EPISSAGE PAR IMMUNODEPLETION-COMPLEMENTATION .....	166
5.3.1. Obtention d'extraits nucléaires immuno-déplétés .....	167
5.3.2. Efficacité d'épissage des extraits nucléaires immuno-déplétés et contrôles .....	168
5.3.3. Complémentation de l'épissage par ajout de protéine recombinante .....	169
5.3.4. Comment optimiser les conditions d'épissage et de complémentation ?.....	170
5.3.5. Effet de la déplétion de KIN17 sur la formation des complexes d'épissage.....	170



---

## 1. LES DOMAINES IMPLIQUES DANS L'INTERACTION DE LA PROTEINE AVEC LES ACIDES NUCLEIQUES

---

La protéine KIN17 humaine est capable de lier l'ADN simple et double brin, avec une préférence pour l'ADN courbe ; en particulier la protéine KIN17 se lie à l'ADN courbe trouvé au niveau des spots de recombinaison illégitime dans les cellules humaines (Angulo et al. 1991b; Mazin et al. 1994a; Mazin et al. 1994b; Timchenko et al. 1996; Biard et al. 2002). De plus, à l'aide d'un mutant de la protéine pour le doigt de zinc, il a été montré que la protéine KIN17 de souris lie efficacement l'ADN double brin si son doigt de zinc est fonctionnel (Mazin et al. 1994b). Par ailleurs, l'interaction de la protéine entière KIN17 de mammifères (souris et homme) avec l'ARN a été caractérisée par G. Pinon-Lataillade du LGR-DRR-CEA Fontenay-aux-Roses (Pinon-Lataillade et al. 2004). Cette étude a montré que la protéine KIN17 de souris lie l'ARN *in vivo* et que les protéines KIN17 de souris et d'homme sont capables d'interagir directement *in vitro* avec différents types d'homopolymères d'ARN. Ainsi, ces différentes études ont montré que la protéine KIN17 de mammifères se lie aux acides nucléiques, ADN et ARN. Dans ce cadre là, nous avons été intéressés par déterminer les régions de la protéine impliquées dans ces interactions avec les acides nucléiques.

J'ai réalisé certaines études de *Southwestern* et les études de *Northwestern* avec les Dr. G. Pinon-Lataillade et J. Angulo-Mora du LGR/DRR/CEA Fontenay-aux-Roses.

J'ai réalisé les autres études de *Southwestern* avec le Dr. J.Acker du SBGM/DBJC/DSV/CEA Saclay.

### 1.1. Interaction de la protéine avec l'ADN

Dans le travail présenté ci-dessous, nous avons cherché (i) à préciser les régions de la protéine hKIN17 impliquées dans l'interaction avec l'ADN, (ii) à comprendre les bases moléculaires associées à ces interactions et en particulier (iii) à caractériser la discrimination entre ADN courbe et non courbe.

#### 1.1.1. Domaines impliqués dans l'interaction de la protéine avec l'ADN

La recherche du ou des domaines de la protéine hKIN17 impliqué dans l'interaction de la protéine entière avec l'ADN a été menée en utilisant la méthode du *Southwestern*. Cette méthode consiste à incuber une membrane sur laquelle les protéines d'intérêt ont été transférées avec un ADN radioactif et à révéler sur support radiosensible les protéines capables de retenir la cible radioactive dans les conditions d'incubation de l'expérience. Un premier essai a été réalisé en hybridant un ADN double brin long de 500 nucléotides, provenant d'une origine de répllication reconnue par la protéine hKIN17 entière (J. Angulo-Mora, résultats non publiés), avec les différents domaines structuraux de KIN17 (K1, K2 et K3) (Figure 1-1).

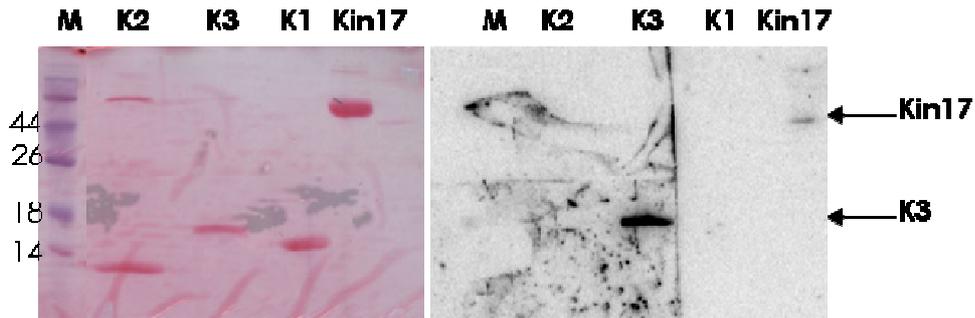


Figure 1-1 : Etude de l'interaction des domaines structuraux de KIN17 (K1, K2 et K3) avec un long ADN double brin de 500nt par *Southwestern*. A gauche, les protéines sont colorées au rouge Ponceau et apparaissent en quantité équivalente. A droite, la radioactivité est révélée au Phosphoimager.

Cette expérience confirme l'interaction de la protéine hKIN17 avec l'ADN et montre que le domaine N-terminal comprenant le doigt de zinc (K3) interagit fortement avec cet ADN dans les conditions de l'expérience. En revanche, le domaine C-terminal (K1) et le domaine N-terminal sans le doigt de zinc (K2) ne sont pas capables de retenir l'ADN suggérant que le doigt de zinc est responsable de l'interaction de la protéine avec l'ADN.

Nous avons confirmé l'interaction du domaine K3 avec l'ADN par la méthode de *dot-blot*. La procédure expérimentale est la même que pour le *Southwestern* à la seule différence que les protéines sont ici déposées directement sur la membrane de nitrocellulose, sans séparation électrophorétique préalable sur gel acrylamide, limitant ainsi le risque de dénaturation ou de mauvaise renaturation des protéines. Pour cette expérience, les différents domaines structuraux de KIN17 ainsi que la protéine entière ont été déposés directement sur une membrane de nitrocellulose. La BSA a été utilisée en contrôle négatif car il n'est pas décrit qu'elle interagit avec l'ADN. La membrane de nitrocellulose a été mise à hybrider avec le même ADN double brin de 500 nucléotides que précédemment.

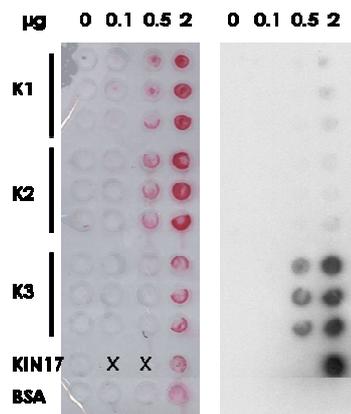


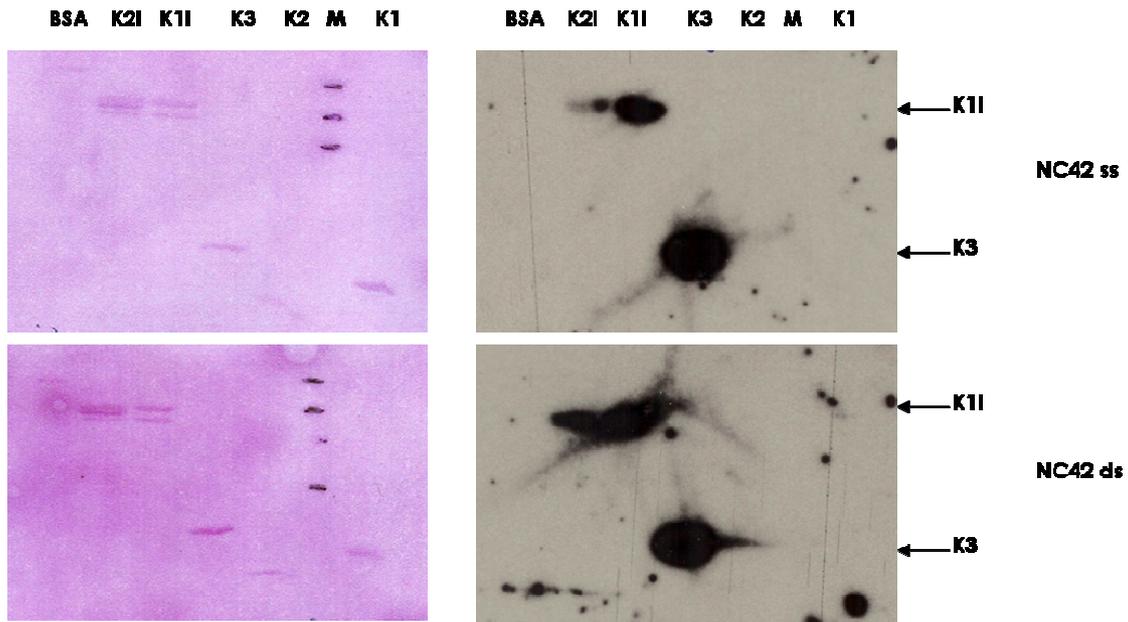
Figure 1-2 : Etude de l'interaction des domaines structuraux de KIN17 (K1, K2 et K3) avec un long ADN double brin de 500nt par *dot-blot*. Les protéines KIN17 et BSA servent respectivement de contrôle positif et négatif. A gauche, les protéines sont colorées au rouge Ponceau et les quantités de protéine déposée sont indiquées au dessus de la membrane. A droite, la radioactivité est révélée au Phosphoimager. Les deux croix signifient que les dépôts sont absents à ces positions.

Cette expérience confirme les résultats obtenus par *Southwestern* puisque l'ADN radioactif est fortement retenu par la protéine KIN17 entière et par le domaine N-terminal comprenant le doigt de zinc. Cependant, à ce stade là, nous ne savons pas si le doigt de zinc

seul est nécessaire et suffisant pour l'interaction ou si la combinaison des deux domaines, doigt de zinc et *winged helix*, est indispensable pour l'interaction de la protéine hKIN17 avec l'ADN. Il serait donc intéressant de répéter cette étude avec le doigt de zinc seul et purifié. De plus, nous ne pouvons pas exclure que la région centrale déstructurée puisse également être impliquée dans l'interaction de la protéine KIN17 avec l'ADN.

### 1.1.2. Reconnaissance d'ADN courbe et non courbe

Dans les essais précédents, l'ADN utilisé est un ADN long de 500 nucléotides rendant difficile toute étude d'interaction d'un point de vue structural (par RMN par exemple) ; c'est pourquoi nous avons cherché à réduire la taille de cet ADN. Puisqu'il avait été montré que la protéine KIN17 peut discriminer deux oligonucléotides de 42 nucléotides, un ADN courbe simple brin et un ADN non courbe simple brin (Mazin et al. 1994b), nous avons poursuivi les expériences de *Southwestern* avec ces oligonucléotides. De plus, l'interaction avec l'ADN de constructions plus longues des domaines K1 et K2 (K1long=K11 et K2long=K21, Figure 1-1 dans la partie Matériel et Méthodes) comprenant la région flexible en amont et en aval respectivement de chacun des domaines a été testée afin de voir si la région déstructurée au centre de hKIN17 joue un rôle dans l'interaction de la protéine avec les différents ADN.



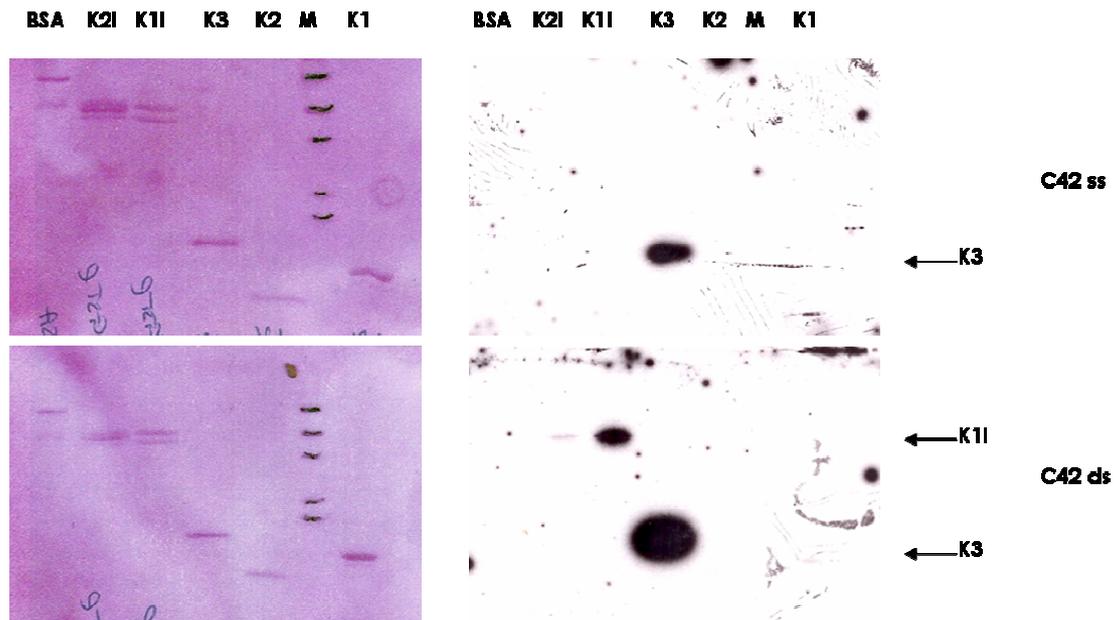


Figure 1-3 : Etude de l'interaction des domaines structuraux de KIN17 (K1, K2 et K3) et des domaines comprenant la région flexible de KIN17 en fusion à la GST (K1I, K2I) avec les ADN courbe et non courbe de 42 nucléotides (C42 et NC42) sous forme simple et double brin par Southwestern. La BSA (Bovine Serum Albumine) est utilisée comme contrôle négatif. M représente le marqueur de poids moléculaire. A gauche, les protéines sont colorées au rouge Ponceau et apparaissent en quantité équivalente. A droite, la radioactivité est révélée sur film MS après 30 minutes d'exposition pour les membranes incubées avec le NC42 et la radioactivité est révélée sur film MR après une nuit d'exposition pour les membranes incubées avec le C42.

Ces résultats montrent que le domaine N-terminal (K3) se lie fortement aux ADN de 42 nucléotides, courbe et non courbe, simple et double brin. Cependant, avec ce type d'expérience, nous ne pouvons pas comparer l'affinité du domaine pour les différents oligonucléotides. Cette liaison du domaine K3 aux deux séquences d'ADN (ADN courbe et non-courbe) suggère que la région N-terminale de hKIN17 se lie plutôt de manière non spécifique à l'ADN.

De plus, le domaine K1long en fusion à la GST est capable de retenir l'ADN (NC42ss, NC42ds, C42ds et peut être même C42ss) alors que le domaine K1 ne l'est pas, suggérant que la région flexible en amont de la région C-terminale pourrait aussi participer à l'interaction de la protéine hKIN17 avec l'ADN. Ainsi, il pourrait y avoir une coopérativité dans l'interaction de plusieurs régions de hKIN17 avec l'ADN pour augmenter l'affinité de la protéine entière pour l'ADN.

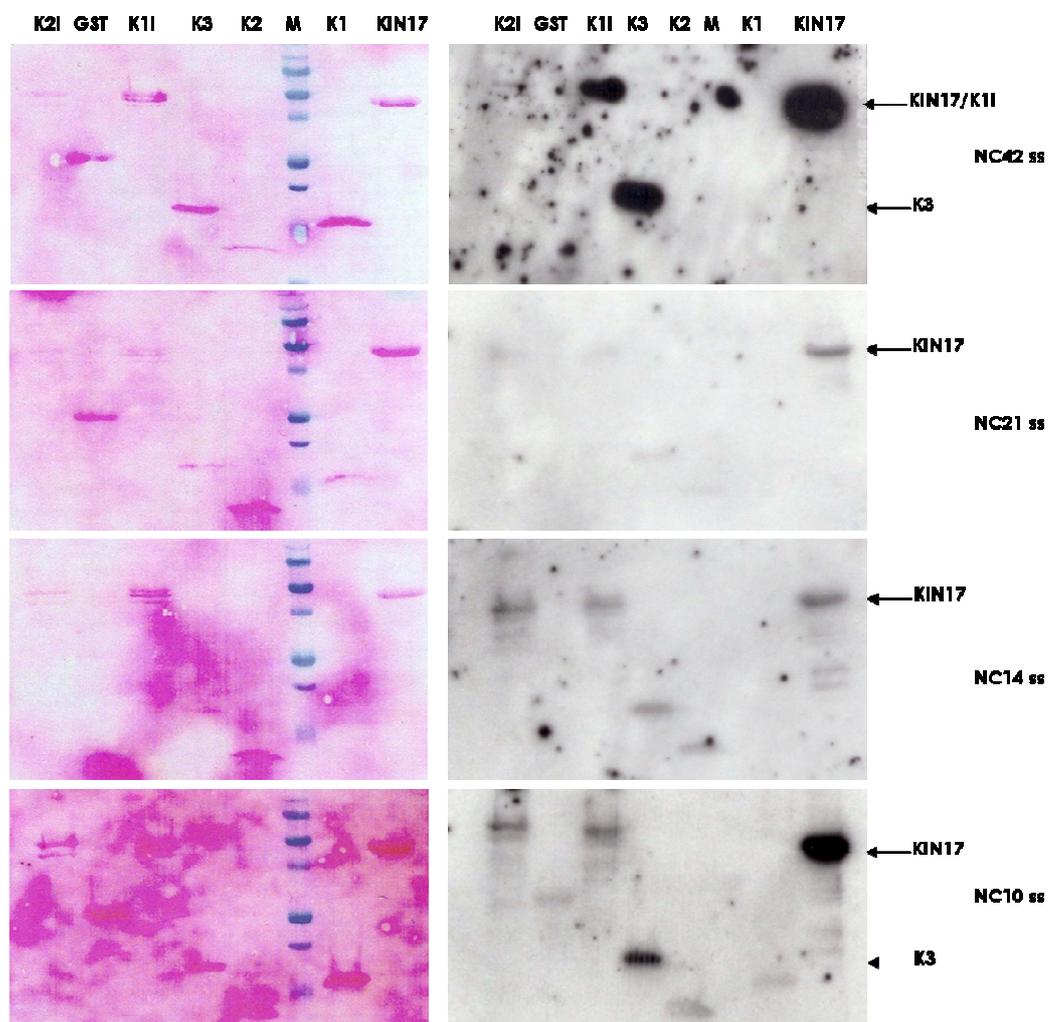
### 1.1.3. Peut-on caractériser au niveau moléculaire l'interaction de la région N-terminale avec l'ADN ?

Dans l'objectif d'évaluer l'affinité de l'interaction entre le domaine K3 et l'ADN et de cartographier la zone d'interaction par RMN, nous avons cherché à réduire la taille de l'ADN nécessaire pour observer la liaison avec le domaine K3. En effet, une interaction forte entre une protéine et un ADN de grande taille ne peut pas être mise en évidence par RMN car le spectre de la protéine dans le complexe de grande taille aurait des raies élargies avec le risque que le signal ne devienne plus détectable. Nous avons donc réalisé différents essais de

Southwestern avec des sondes réduites (Figure 1-4) à partir des deux sondes, courbe et non courbe, de 42 nucléotides testées précédemment. Les ADN 42mères courbe et non-courbe sont constitués de motifs répétés. Les ADN plus courts ont donc été choisis en tenant compte de ces répétitions.

**ADN NC42** : 5'-TCTAATCTCTCTCTAATCTCTTCTAATCTCTCTCTAATCTCT-3'  
**ADN NC21** : 5'-TCTAATCTCTCTCTAATCTCT-3'  
**ADN NC14** : 5'-TCTAATCTCTCTCT-3'  
**ADN NC10** : 5'-TCTAATCTCT-3'  
**ADN C42** : 5'-AAAAATCTCTAAAAAATCTCTAAAAAATCTCTAAAAAATCTCT-3'  
**ADN C21** : 5'-AAAAATCTCTAAAAAATCTCT-3'  
**ADN C10** : 5'-AAAAATCTCT-3'

Figure 1-4 : Séquences nucléotidiques des ADN utilisés pour les essais de Southwestern avec la protéine KIN17 et ses différents domaines.



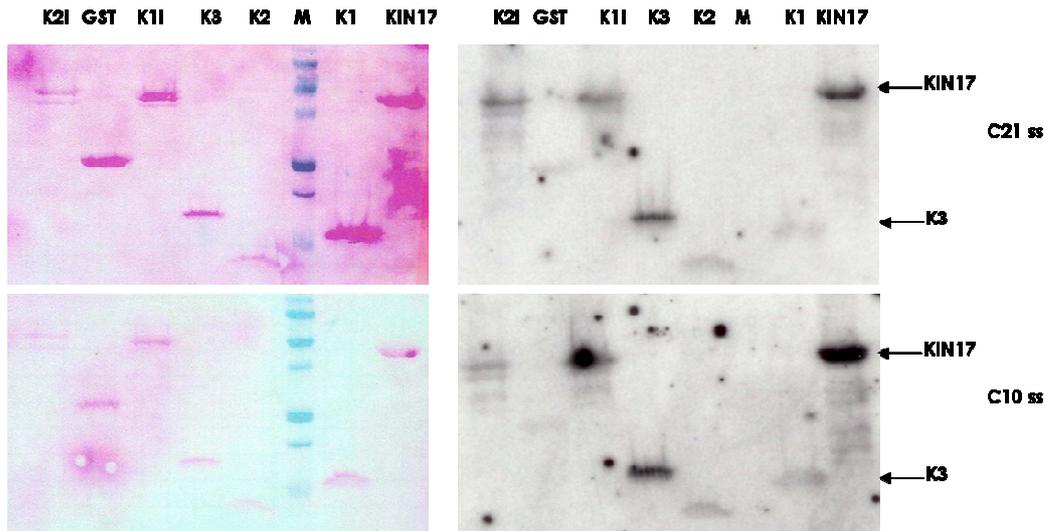


Figure 1-5 : Etude de l'interaction des domaines structuraux de KIN17 (K1, K2 et K3) et des domaines comprenant la région flexible de KIN17 en fusion à la GST (K11, K21) avec les ADN de différentes tailles par *Southwestern*. La GST est utilisée comme contrôle négatif. A gauche, les protéines sont colorées au rouge Ponceau et apparaissent en quantité équivalente. A droite, la radioactivité est révélée sur film MR après une nuit d'exposition.

Ces résultats montrent que les ADN courts (de 21, 14 et 10 nucléotides) sont moins bien retenus que les ADN de 42 nucléotides par le domaine K3. Cela peut paraître surprenant du fait de la taille de la région N-ter (diamètre de l'ordre de 20Å) et donc de la taille de la séquence d'ADN qu'elle peut *a priori* lier. Un ADN de moins d'une dizaine de nucléotides doit être suffisant pour interagir avec un doigt de zinc et l'acide nucléique (Brown 2005). Il est ainsi possible qu'un ADN de 42 nucléotides puisse retenir plusieurs molécules K3 de la membrane et donc apparaître se lier plus fortement qu'un ADN de plus petite taille. En revanche, la protéine entière hKIN17 semble très bien retenir les deux ADN, non courbe et courbe simple brin de 10 nucléotides.

Nous prévoyons d'étudier par la méthode du retard sur gel l'interaction de la région N-terminale de hKIN17 avec l'ADN non courbe de 42 nucléotides pour lequel une forte interaction a été observée en *Southwestern*. Nous aurons ainsi accès à l'affinité de l'interaction et nous mettrons en évidence la formation d'un ou de plusieurs complexes successif(s) protéine-ADN. Puisque le domaine K3 est également capable de retenir les ADN NC10ss et C10ss, nous pouvons envisager d'étudier l'interaction du domaine avec ces ADN courts par RMN.

## 1.2. Interaction de la protéine avec l'ARN

De la même manière que pour l'interaction de hKIN17 avec l'ADN, nous avons d'abord identifié les domaines de la protéine hKIN17 impliqués dans l'interaction avec l'ARN, dans le but de préciser ensuite les bases moléculaires de cette interaction.

La plupart des résultats concernant ce travail ont été publiés dans l'article 2 (page 86). Ainsi, il a été montré par *Northwestern* que la région N-terminale de hKIN17 comprenant le doigt de zinc et le domaine C-terminal retiennent l'ARN *in vitro* (figure 3 de l'article 2). Le

domaine N-terminal, en particulier le doigt de zinc, semble donc avoir la capacité de lier à la fois l'ARN et l'ADN.

L'interaction du domaine C-terminal avec l'ARN a été ensuite précisée en étudiant l'affinité de ce domaine pour les différents types d'homopolymères. Nous avons montré que les préférences de ce domaine pour les homopolymères poly(U) et poly(G) étaient les mêmes que celles de la protéine KIN17 entière (figure 4 de l'article 2). A partir d'informations cristallographiques et d'études de mutants, nous avons proposé qu'une région conservée et chargée positivement du domaine K1 puisse être impliquée dans l'interaction avec l'ARN (figures 5 et 6 de l'article 2). Nous avons vérifié par RMN que la structure des domaines K1 mutés était conservée par rapport à celle du domaine K1 natif. Pour cela, nous avons enregistré les spectres 1D et les spectres  $^1\text{H}$ - $^1\text{H}$ , TOCSY et  $^1\text{H}$ - $^1\text{H}$ , NOESY de chacun des domaines (K1 natif, K36 et K125) pour les comparer. Tous les spectres de mutants se superposent très bien à ceux du natif, en particulier nous n'avons pas observé d'élargissement des raies sur les spectres 1D. A l'heure actuelle, nous n'avons pas obtenu de structure du domaine C-terminal de KIN17 en complexe avec l'ARN. En effet, nous considérons que, pour que la structure de ce complexe ait un intérêt biologique, il est nécessaire de savoir s'il y a une séquence ou une forme particulière d'ARN qui interagit avec KIN17 avant de déterminer la structure du complexe correspondant.

## 2. L'INTERACTION DU DOMAINE C-TERMINAL AVEC LES HISTONES

Différents travaux ont montré que des doubles domaines de la « Royal Family » dans des protéines de la signalisation ou de la réparation des dommages de l'ADN interagissent avec des histones modifiées. Ainsi, le double domaine Tudor de la protéine 53BP1 reconnaît la lysine 20 diméthylée de l'histone H4 (Botuyan et al. 2006), le double domaine Chromo de la protéine CHD1 reconnaît la lysine 4 méthylée de l'histone H3 (Flanagan et al. 2005). Nous avons utilisé la méthode de GST pull-down pour mettre en évidence et étudier l'interaction directe du domaine C-terminal de la protéine hKIN17, composé de deux sous-domaines de type SH3, avec les histones.

J'ai effectué l'ensemble des expériences de GST pull down.

J'ai collaboré avec l'équipe du Dr. J.Chamot-Rooke du laboratoire DCMR de l'Ecole Polytechnique et avec l'équipe du Dr. B. Monsarrat de l'IPBS de Toulouse pour les analyses par spectrométrie de masse.

### 2.1. Liaison du domaine C-terminal aux histones H3 et H4

Un premier essai de GST pull-down a été réalisé pour voir s'il y avait une interaction entre le domaine K1 de KIN17 en fusion à la GST et les histones de thymus de veau. D'après l'expérience du laboratoire et d'après les résultats de pull-down publiés dans Huyen et al., 2004, il est apparu qu'il est nécessaire de travailler dans des conditions stringentes (forte concentration en sel, ajout de Triton X-100) lorsque des histones entières sont utilisées pour des études d'interaction *in vitro* afin de limiter les phénomènes d'agrégation. En particulier, des études ont montré que l'histone H3 a tendance à s'associer en solution de force ionique modérée (Morris et al. 1977) et à pH neutre (Edwards et al. 1970). Ainsi plusieurs conditions en sel allant de 0.5 à 1M NaCl ont été testées. Les billes chargées en GST ou en protéines de fusion sont mises à incuber avec un mélange d'histones de thymus de veau et après un certain nombre de lavages, elles sont directement déposées sur gel SDS-PAGE (Figure 2-1).

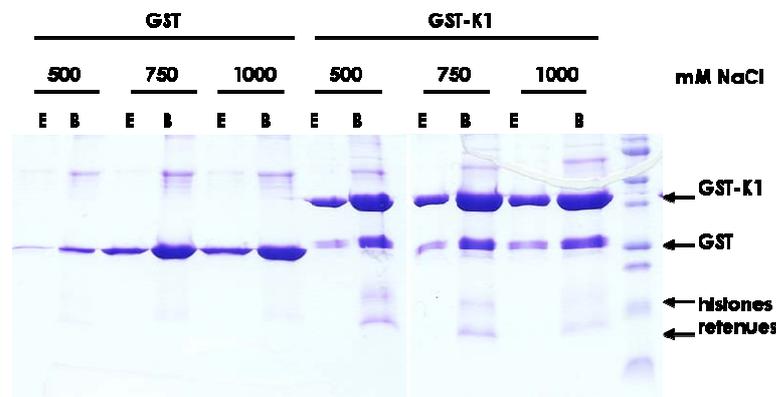


Figure 2-1 : Etude de l'interaction d'un mélange d'histones de thymus de veau avec les protéines GST et le domaine K1 de KIN17 en fusion à la GST (GST-K1) par la méthode du GST pull-down. Sur un gel SDS-PAGE 15% coloré au bleu de Coomassie sont déposées les billes chargées en protéine avant (Av) et après (Ap) l'incubation avec les histones. Pour chaque protéine, trois conditions différentes de tampon d'interaction sont testés : 20mM Tris-HCl pH 7.4, 0.5% Triton X-100 et 0.5 ou 0.75 ou 1M NaCl.

Les résultats de la Figure 2-1 montrent que le domaine K1 est capable de retenir assez fortement les histones puisqu'en 1M NaCl des histones sont encore visibles sur le gel en bleu de Coomassie. Le contrôle (GST seule) ne lie pas les histones pour toutes les conditions d'interaction testées. Ensuite, les histones retenues ont été identifiées sans ambiguïté par spectrométrie de masse LC-MS au laboratoire DCMR de l'Ecole Polytechnique. La bande la plus haute, au-dessus de 15kDa, correspond à l'histone H3 (34% de couverture) et la bande la plus basse correspond à l'histone H4 (58% de couverture). Ainsi, le domaine K1 de hKIN17 est capable de lier les histones H3/H4 mais pas les histones H2A/H2B. Les histones H3 et H4 sont associées dans le nucléosome (Luger et al. 1997) et forment un complexe stable *in vitro* (Eickbush et al. 1978). Par conséquent, la question se pose de savoir si l'interaction du domaine K1 de KIN17 avec les histones a lieu par l'intermédiaire de l'histone H3 ou de l'histone H4.

C'est pourquoi, nous avons ensuite testés en GST pull-down l'interaction du domaine K1 avec chacune des histones H3 et H4 séparément. Là encore, les billes chargées en GST ou en protéines de fusion sont mises à incuber avec chacune des histones H3 et H4 de thymus de veau et après un certain nombre de lavages, elles sont directement déposées sur gel SDS-PAGE (Figure 2-2).

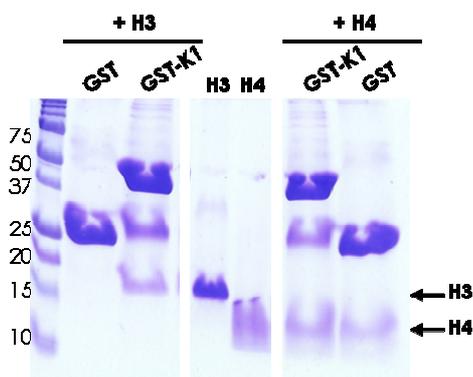


Figure 2-2 : Etude de l'interaction de l'histone H3 et de l'histone H4 de thymus de veau avec les protéines GST et le domaine K1 de KIN17 en fusion à la GST (GST-K1) par la méthode du GST pull-down. Sur un gel SDS-PAGE 15% coloré au bleu de Coomassie sont déposées les billes chargées en protéine après l'incubation avec les deux histones. La condition d'interaction de cet essai est : 20mM Tris-HCl pH 7.4, 0.5% Triton X-100 et 0.5M NaCl. Une fraction des histones H3 et H4 est déposée sur gel (puits notés H3 et H4).

D'après les résultats présentés sur la Figure 2-2, le domaine K1 de hKIN17 est capable de lier l'histone H3 alors que le contrôle (GST seule) ne l'est pas. En revanche, il n'est pas possible de conclure sur une interaction éventuelle avec l'histone H4 puisque l'histone H4 est retenue de la même manière par toutes les protéines, y compris le contrôle GST. Nous ne sommes pas parvenus à trouver de meilleures conditions expérimentales pour l'interaction avec l'histone H4.

Nous avons répété l'expérience de GST pull-down avec l'histone H3 en rajoutant une étape d'éluion des billes après les lavages pour s'assurer que les histones n'étaient pas simplement agrégées sur les billes chargées en protéines mais qu'elles étaient plutôt associées à la protéine GST-K1.

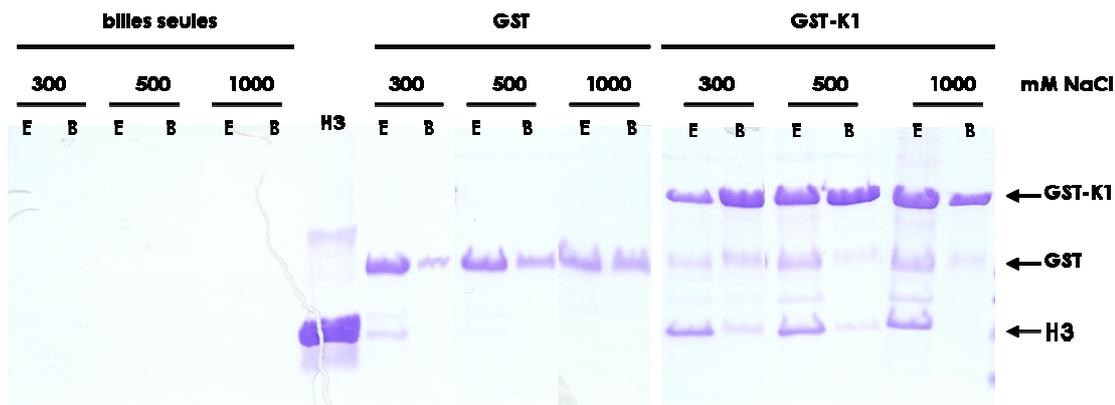


Figure 2-3 : Etude de l'interaction de l'histone H3 de thymus de veau avec les billes de GS4B non chargées en protéines, avec la protéine GST et avec le domaine K1 de KIN17 en fusion à la GST (GST-K1) par la méthode du GST pull-down. Sur un gel SDS-PAGE 15% coloré au bleu de Comassie sont déposées les éluions (E) et les billes après élution (B). Pour chaque protéine, trois conditions différentes de tampon d'interaction sont testées : 50mM Tris-HCl pH 7.4, 0.5% Triton X-100 et 0.3 ou 0.5 ou 1M NaCl.

D'après les gels de la Figure 2-3, nous montrons que l'histone H3 est éluee en même temps que la protéine d'intérêt (GST-K1) des billes de GS4B. Cette fois-ci, il apparaît aussi que l'interaction de GST-K1 avec l'histone H3 reste forte en 1M NaCl. En accord avec les essais précédents, la protéine GST ne retient pas l'histone H3 dans les conditions d'interactions fortes en sel (0.5 et 1M NaCl). De même, nous nous sommes assurés que l'histone H3 ne s'agrége pas sur les billes de GS4B nues.

Nous avons confirmé l'interaction du domaine C-terminal avec l'histone H3 par une autre méthode, le *far western*. La protéine H3 de thymus de veau et la protéine BSA comme contrôle négatif sont déposées en différentes quantités sur une membrane qui est mise à incuber avec le domaine C-terminal de la protéine hKIN17.

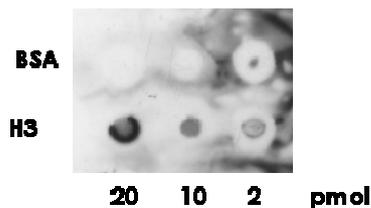


Figure 2-4 : Interaction du domaine C-terminal de KIN17 avec l'histone H3 par *far western*. Trois quantités différentes de protéines (2, 10 et 20 pmol) sont déposées comme indiqué. Les mêmes quantités de BSA sont déposées. La présence de la protéine K1 est détectée avec des anticorps anti-KIN17.

La Figure 2-4 montre que le domaine K1 est détecté aux endroits où l'histone H3 a été déposée et d'autant plus fortement que la quantité d'histone déposée est grande. En revanche, aucun signal de K1 n'est détecté aux endroits où la BSA a été déposée. Ainsi, le domaine C-terminal de KIN17 a été spécifiquement retenu par l'histone H3 de thymus de veau.

*En conclusion*, dans nos conditions expérimentales, le domaine C-terminal de KIN17 est capable de lier l'histone H3. Par contre, nous ne pouvons pas exclure une interaction éventuelle de la protéine avec l'histone H4.

## 2.2. Liaison du domaine C-terminal à l'histone H3 modifiée

Les histones de thymus de veau, en tant qu'histones eucaryotes, ont subi des modifications post-traductionnelles *in vivo* (acétylation, phosphorylation, méthylation, ubiquitylation, ADP-ribosylation, sumoylation, glycosylation ou biotinylation). Nous étions donc intéressés par savoir si l'interaction du domaine K1 de hKIN17 était dépendante de la présence de ces modifications.

Pour cela, nous avons comparé l'interaction du domaine K1 avec l'histone H3 de thymus de veau et avec l'histone H3 recombinante, c'est-à-dire sans modifications post-traductionnelles. La protéine Asf1 est un chaperon d'histones (Munakata et al. 2000) qui interagit avec l'histone H3 indépendamment de son état de modifications post-traductionnelles. Elle est donc utilisée comme contrôle positif dans ces expériences de GST pull-down.

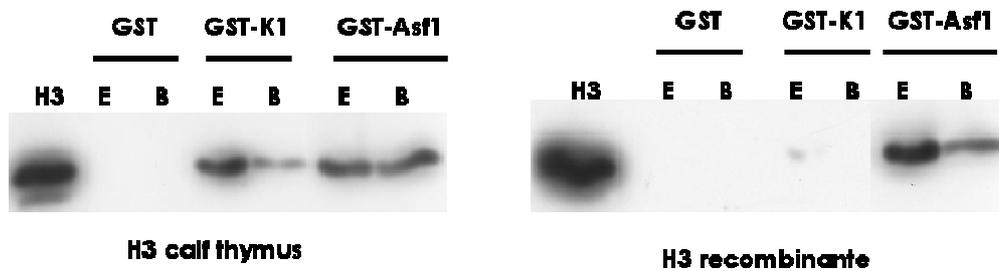


Figure 2-5 : Comparaison de l'interaction *in vitro* du domaine de hKIN17 avec l'histone H3 de thymus de veau et l'histone H3 recombinante. Des quantités égales de protéines fusionnées à la GST et de GST sont déposées sur des billes de GS4B. Les histones H3 retenues sont visualisées par immunodétection avec des anticorps anti-H3. La protéine Asf1 connue pour être un chaperon d'histones sert de contrôle positif et la protéine GST de contrôle négatif. La condition d'interaction de cet essai est : 50mM Tris-HCl pH 7.4, 0.5% Triton X-100 et 0.5M NaCl.

Alors que la protéine Asf1 interagit fortement et indifféremment avec l'histone H3 de thymus de veau et l'histone H3 recombinante, le domaine K1 de hKIN17 ne retient que l'histone H3 de thymus de veau (Figure 2-5). Il semble donc que la présence de modifications post-traductionnelles sur l'histone H3 soit cruciale pour la reconnaissance de cette histone par le domaine C-terminal de hKIN17.

Au sein du mélange d'histones de thymus de veau, de multiples combinaisons de modifications post-traductionnelles sont présentes et nous avons donc été intéressés par l'identification de la ou des modification(s) post-traductionnelle(s) particulière(s) reconnue(s) par le domaine C-terminal de hKIN17. Pour cela, nous avons répété l'expérience de GST pull-down avec des grandes quantités de protéines de fusion et d'histones (Figure 2-6) afin d'identifier par spectrométrie de masse les modifications des histones retenues par la protéine GST-K1. Afin de les comparer à l'ensemble des modifications présentes dans le mélange initial, les histones utilisées pour faire l'expérience de GST pull-down ont également été analysées (Input).

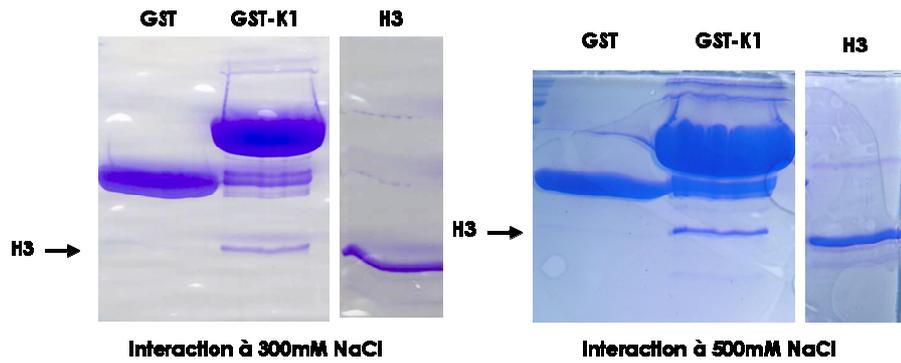


Figure 2-6 : Résultat de l'interaction de l'histone H3 de thymus de veau avec la protéine GST-K1 en grande quantité, dans deux conditions d'interaction différentes (300 et 500 mM NaCl). Les bandes d'histone H3 visibles au Bleu de Comassie ont été découpées, digérées à la trypsine et analysées par spectrométrie de masse LC MS/MS.

Les analyses par spectrométrie de masse LC MS/MS (LC Packing/Qq-ToF QStarXL, Applied Biosciences) ont été effectuées par le laboratoire de B.Monsarrat à l'IPBS de Toulouse. La trypsine est une endoprotéase qui hydrolyse les liaisons peptidiques dans lesquelles un acide aminé basique (Lysine, Arginine) engage sa fonction acide. La séquence de l'histone H3 est assez enrichie en résidus lysine et arginine, en particulier dans sa queue N-terminale (Figure 2-7). Certains peptides générés par la digestion à la trypsine vont donc être très courts et par conséquent pas visibles en spectrométrie de masse, pour laquelle 5 ou 6 résidus sont nécessaires pour constituer un peptide visible. En revanche, lorsque ces résidus sont modifiés, la trypsine n'est plus capable de digérer les protéines au niveau de ces résidus et donc de nouveaux peptides sont générés.

ARTKQTARKS TGGKAPRKQL ATKAARKSAP ATGGVKKPHR YRPGTVALRE IRRYOKSTEL LIRKLPFOR  
L VREIAQDFKT DLRFQSSAVM ALQEACEAYL VGLFEDTNLC AIHAKRVTIM PKDIQLARRI RGERA

peptide et modifications identifiés	Input	GST-K1
peptide 9-17 et 10-17		
- méthylation K9, acétylation K14	oui	non
- diméthylation K9, acétylation K14	oui	oui
- triméthylation K9, acétylation K14	oui	non
- acétylation K14	oui	non
peptide 18-26		
- acétylation K23	oui	oui
- acétylation K18, acétylation K23	oui	oui
peptide 27-36		
- méthylation K27	oui	oui
- diméthylation K27	oui	oui
- diméthylation K27 et méthylation K36	oui	non
peptide 73-83		
- méthylation K79	oui	oui
- diméthylation K79	oui	oui

Figure 2-7 : Séquence de l'histone H3 bovine en haut. Les peptides et modifications post-traductionnelles identifiés par spectrométrie de masse LC MS/MS au sein de l'histone H3 utilisée pour le GST pull-down (input) et au sein de l'histone H3 retenue par la protéine GST-K1 sont donnés dans le tableau.

Les modifications post-traductionnelles identifiées dans l'histone H3 retenue par le domaine K1 de hKIN17 sont comparées aux modifications post-traductionnelles des histones utilisées pour faire l'expérience de GST pull-down (Figure 2-7). Cependant, il faut être très prudent puisque les analyses par spectrométrie de masse ne sont pas quantitatives et donc l'absence d'un peptide donné ne signifie pas forcément qu'il n'existe pas dans la préparation mais seulement qu'il n'a pas été identifié. Nous pouvons tout de même noter que dans l'input il existe les différents peptides où la lysine 9 de l'histone H3 est non ou mono-, di- ou triméthylée alors que dans l'histone H3 retenue par GST-K1, il n'existe que le peptide correspondant à la lysine 9 diméthylée. Nous avons donc retenu cette modification (diméthylation de la lysine 9) comme pouvant jouer un rôle dans l'interaction du domaine C-terminal de KIN17 avec l'histone H3. Nous avons utilisé la méthode de cartographie des variations de déplacements chimiques par RMN pour étudier l'interaction entre le domaine K1 de hKIN17 et un peptide de l'histone H3 contenant la lysine 9 diméthylée. Nous n'avons observé qu'une très faible interaction, de l'ordre du millimolaire, de ce peptide avec le domaine K1. De la même manière, nous avons étudié l'interaction de ce domaine K1 avec un peptide d'histone comprenant une histone acétylée (lysine 18). Avec ce peptide, nous n'observons pas d'interaction dans les conditions de la RMN.

*En conclusion*, ces travaux ont montré que le domaine C-terminal de KIN17 était capable d'interagir avec l'histone H3 et éventuellement avec l'histone H4. De plus, nous avons précisé que cette interaction était dépendante de la présence de modifications post-traductionnelles sur l'histone H3. Nous avons ensuite cherché à identifier par spectrométrie de masse la ou les modifications post-traductionnelles responsable(s) de l'interaction de l'histone H3 avec le domaine K1 mais cette méthode ne nous a pas permis de les identifier avec succès.

### 3. LA RECHERCHE DE PARTENAIRES PROTEIQUES ASSOCIES AUX DOMAINES N- ET C-TERMINAUX

Pour la recherche de partenaires associées à la protéine hKIN17, nous avons d'abord choisi la méthode *GST-pulldown* pour isoler les partenaires protéiques associés à chacun des domaines de la protéine (domaines K1 et K2).

J'ai effectué les expériences de mises au point avec le Dr. L. Miccoli au LGR/DRR/CEA Fontenay-aux-Roses.

J'ai effectué les expériences de pull-down à grande échelle au laboratoire.

J'ai collaboré avec l'équipe du Dr. J.Chamot-Rooke du laboratoire DCMR de l'Ecole Polytechnique pour les analyses de spectrométrie de masse.

#### 3.1. Mise au point du GST-pulldown avec des extraits marqués au <sup>35</sup>S

Avant de mener des expériences de *GST-pulldown* à grande échelle, nous avons effectué des essais avec des quantités restreintes d'extraits cellulaires HeLa afin de voir si cette méthode pouvait nous permettre de retenir des partenaires protéiques avec l'un ou l'autre des domaines de hKIN17.

Ces expériences de mise au point ont été réalisées avec des extraits totaux et nucléaires que nous avons nous-mêmes préparés à partir de cellules HeLa en culture comme cela est décrit dans le paragraphe 3.3.2 de la partie Matériel et Méthodes. Les protéines des cellules HeLa ont été marquées radioactivement par incorporation de méthionine marquée au <sup>35</sup>S. Le marquage radioactif des protéines permet de travailler avec des petites quantités d'extraits HeLa et de protéines de fusion en étant en même temps très sensible dans la détection des protéines retenues. Ces essais ont été réalisés avec les domaines K1 et K2 en fusion à la GST et avec la protéine GST comme contrôle négatif.

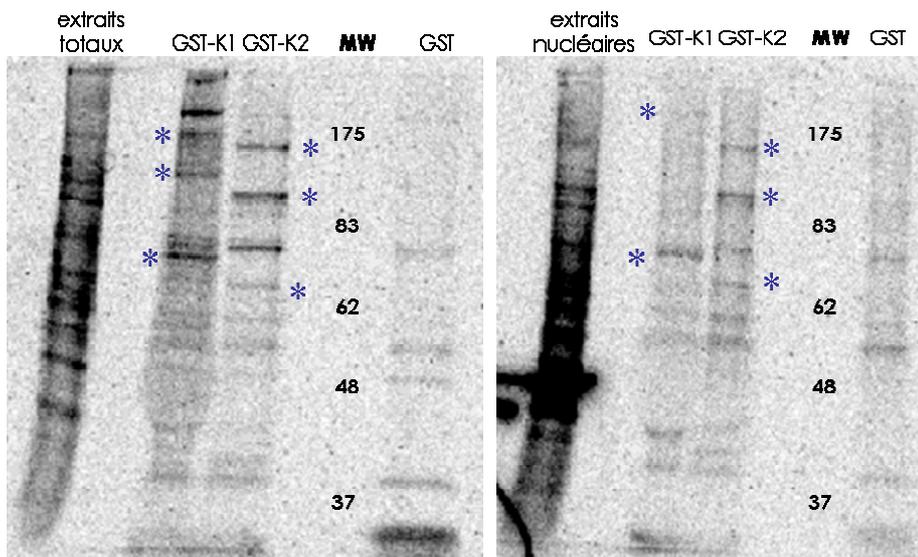


Figure 3-1 : Les billes avec les protéines GST, GST-K1 et GST-K2 après interaction avec les extraits totaux HeLa ou les extraits nucléaires HeLa et après lavage en 300mM NaCl sont déposées sur gel SDS-PAGE et les

protéines marquées au  $^{35}\text{S}$  sont révélées au PhosphoImager. Les bandes de protéines marquées par un astérisque bleu correspondent à des bandes de protéines fortes qui ne sont observées que pour l'une des fusions (GST-K1 ou GST-K2) et pas pour le contrôle GST.

Les profils de protéines des extraits totaux ou des extraits nucléaires de cellules HeLa retenues par les protéines GST, GST-K1 et GST-K2 sont comparés :

- Certaines bandes de protéines sont communes aux puits du contrôle GST et des fusions GST-K1 et GST-K2 (Figure 3-1). Ces protéines doivent correspondre à des protéines liées aux colonnes ou aux protéines de fusion de manière non spécifique.

- Comme cela est mis en évidence sur la Figure 3-1, les protéines de fusion GST-K1 et GST-K2 retiennent des protéines différentes.

- La protéine GST-K2 ne semble pas retenir de protéine supplémentaire lorsque des extraits totaux sont utilisés par rapport à des extraits nucléaires. En revanche, il apparaît que des protéines supplémentaires sont retenues par GST-K1 lorsque des extraits totaux sont utilisés comparativement à des extraits nucléaires.

*En conclusion*, il semble que des protéines partenaires interagissent spécifiquement avec chacun des domaines K1 et K2 de hKIN17 et sont donc potentiellement intéressantes à identifier. De plus, le domaine K2 semble avoir essentiellement des partenaires nucléaires alors que le domaine K1 lui aurait plutôt des partenaires nucléaires et cytoplasmiques.

### **3.2. Pull-down à grande échelle et couplé à l'identification des protéines par spectrométrie de masse**

Afin d'identifier les protéines retenues par les différents domaines de KIN17 par spectrométrie de masse, nous avons reproduit les expériences précédentes avec de plus grandes quantités de protéines de fusion et d'extraits cellulaires HeLa à incubation. On peut s'attendre à ce que, pour une protéine partenaire donnée, le nombre de protéines retenues par les protéines de fusion soit faible par rapport au nombre de protéines dans les extraits car les interactions des protéines de fusion avec les protéines partenaires sont probablement de faible affinité. Par ailleurs, l'identification de protéines séparées sur gel SDS-PAGE par spectrométrie de masse est possible si un minimum de protéines peut être extrait du gel d'autant plus que l'extraction des peptides par digestion à la trypsine dans le gel n'est pas totale.

Sur la Figure 3-2 ci-dessous sont présentés les résultats de trois expériences successives de pull-down avec les domaines K1 et K2 de la protéine hKIN17 en fusion avec une étiquette GST ou ZZ. La plupart de ces essais ont été réalisés avec des extraits nucléaires de cellules HeLa.

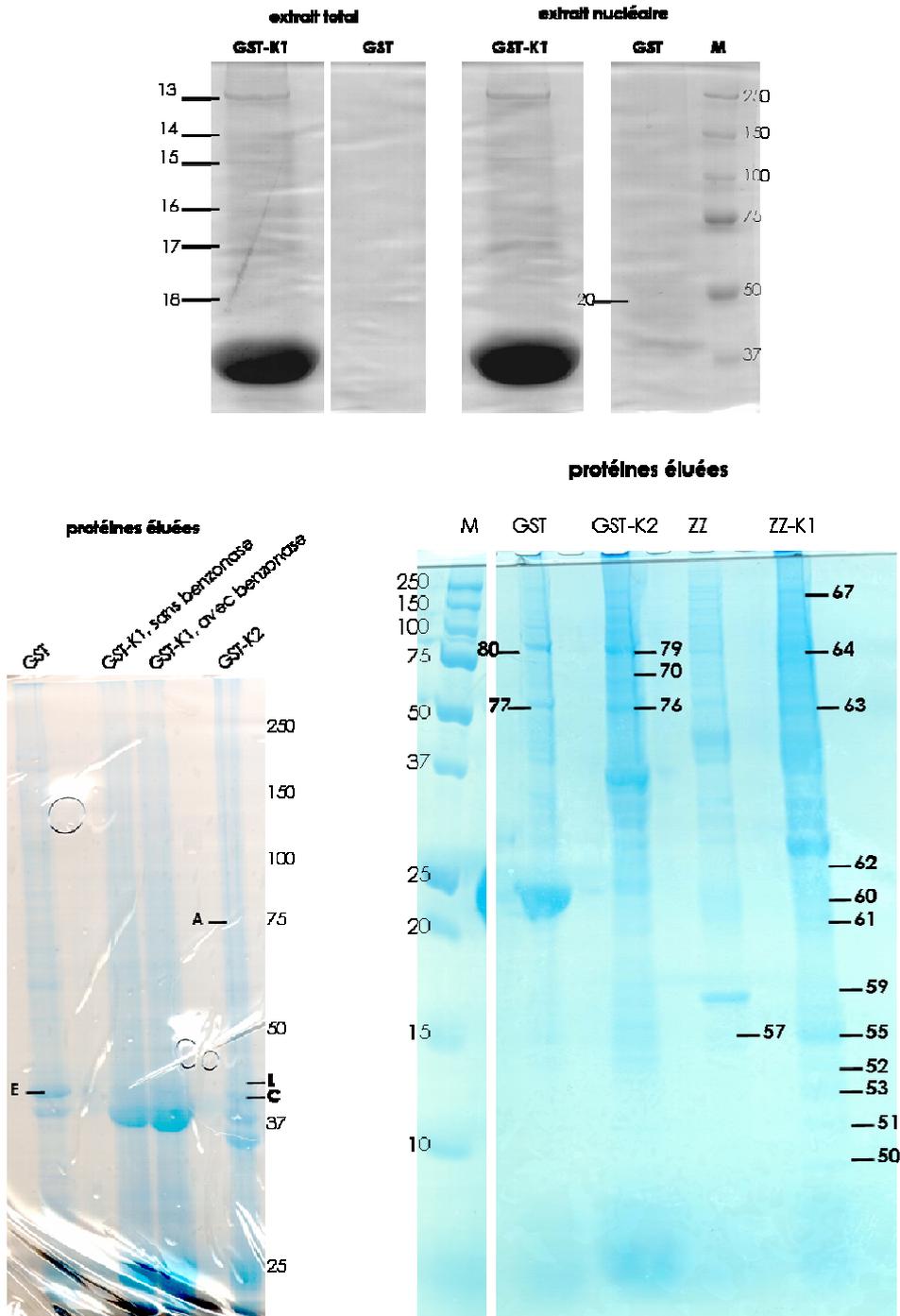


Figure 3-2 : En haut, gels des protéines éluées après lavage en 300mM NaCl des billes incubées avec GST ou GST-K1 + extraits total (à gauche) ou nucléaire (à droite) HeLa. Ce sont des gels SDS-PAGE 8% colorés au Bleu de Coomassie pour voir les protéines de grande taille. Les bandes de protéines numérotées ont été analysées par spectrométrie de masse pour identifier les protéines.

En bas, à gauche, les protéines retenues par les protéines GST, GST-K1 et GST-K2 et éluées avec 1M NaCl sont analysées sur un gel SDS-PAGE 8% coloré au bleu colloïdal. Les bandes de protéines notées par une lettre ont été analysées par spectrométrie de masse pour les identifier.

En bas, à droite, les protéines retenues par les protéines GST, GST-K2, ZZ et ZZ-K1 et éluées avec 1M NaCl et les protéines qui restent sur les billes après élution sont analysées sur un gel gradient (Protean II Ready Gel, 10.20%Tris-HCl, Biorad) coloré au bleu colloïdal. Les bandes de protéines numérotées ont été analysées par spectrométrie de masse pour les identifier.

Le premier essai a été réalisé pour rechercher des partenaires du domaine C-terminal de hKIN17 (K1) en fusion à la GST avec des extraits totaux et nucléaires de cellules HeLa commerciaux (Cilbiotech, Belgique) dans les conditions expérimentales déterminées lors de la mise au point du *GST pull-down* avec les extraits marqués au  $^{35}\text{S}$ . En comparant les deux gels de la Figure 3-2-haut, nous voyons les mêmes bandes de protéines assez fortes lorsque les extraits totaux ou nucléaires sont utilisés. Contrairement à ce qui avait été observé précédemment (paragraphe 3.1), la protéine GST-K1 ne semble pas retenir plus de protéines des extraits totaux par rapport aux extraits nucléaires. Dans l'essai suivant, des extraits nucléaires (Cilbiotech) ont été traités à la benzonase avant d'être incubés avec les protéines appât. Ce traitement des extraits par une nucléase pourrait limiter les interactions indirectes, via l'ADN, entre les protéines. Par ailleurs, les protéines retenues par les protéines appât ont été éluées des billes avec 1M NaCl dans le but de s'affranchir des protéines qui se fixeraient sur les billes de façon non spécifique. Globalement, d'après la Figure 3-2-b, la coloration au bleu colloïdal ne permet pas de détecter un grand nombre de protéines retenues par les fusions. Les profils des protéines retenues par la fusion GST-K1 lorsque cette dernière est incubée avec des extraits traités ou non avec de la benzonase diffèrent très peu. Nous avons également effectué le pull-down avec le domaine K1 en fusion avec une autre étiquette (ZZ) comme protéine appât (Figure 3-2-c). Cette autre construction du domaine K1 peut rendre le domaine plus accessible à ses partenaires et les protéines contaminantes retenues par des interactions non spécifiques avec le contrôle (étiquette ZZ sur la colonne IgG) ne sont pas les mêmes que celles retenues par le contrôle GST.

Certaines bandes de protéines des gels de cette Figure 3-2 sont annotées et correspondent à des protéines qui ont été identifiées par spectrométrie de masse au laboratoire DCMR de Polytechnique en collaboration avec l'équipe de Julia Chamot-Rooke. Les protéines ont été analysées par spectrométrie de masse MALDI-TOF et identifiées grâce à l'utilisation des logiciels *Profound* et/ou *Mascot*. Les résultats de ces identifications sont répertoriés dans l'annexe 3.

D'après les résultats de l'identification des protéines :

- parmi les protéines identifiées, certaines correspondent à des protéines très abondantes dans la cellule, comme par exemple la cyclophiline, l'ubiquitine, la cofiline, la protéine Ran et doivent plutôt être retenues via des interactions non-spécifiques. Il apparaît également important de séquencer les protéines retenues par les contrôles (étiquette GST ou ZZ) afin d'identifier les protéines contaminantes et non spécifiques.

- les expériences de pull down avec le domaine K1 ont été réalisées avec deux types d'étiquette de fusion (GST et ZZ). Pour l'ensemble des essais effectués, nous observons que le domaine K1 retient peu de protéines des extraits. Les quelques protéines identifiées (la protéine CAD, une protéine de « fusion » qui réalise quatre activités enzymatiques (GATase, CPSase, ATCase and DHOase), la protéine tight junction protein 2 qui joue un rôle dans les jonctions serrées et les jonctions adhérentes et la protéine pyruvate kinase qui intervient dans la régulation de la glycolyse) ne semblent pas avoir de lien fonctionnel avec la protéine KIN17. De manière plus intéressante, parmi les protéines retenues par le domaine K1, nous avons identifié la protéine SmD2, une des sept protéines Sm impliquées dans la formation des complexes de l'épissage, bien que cela puisse paraître surprenant de la trouver isolée des autres protéines Sm.

- nous avons identifié de manière assez sûre, un certain nombre de partenaires potentiels du domaine K2 : le facteur eIF-4A, les protéines p68 et DDX21 appartenant à la famille des protéines DEAD-box, des hélicases de l'ARN.

*En conclusion*, d'après ces essais de pulldown avec des extraits de cellules HeLa, il apparaît que la mise en œuvre et l'interprétation de ce type d'expériences restent difficiles et souvent incertaines. Néanmoins, nous pouvons proposer quelques protéines potentiellement associées au domaine N-terminal (K2) de hKIN17. Nous discuterons plus loin (Partie Discussion) des protéines identifiées et du sens que pourrait avoir l'interaction de la protéine KIN17 avec elles. En revanche, cette méthode de pulldown n'a pas permis d'identifier des partenaires du domaine K1, bien que deux types de fusion aient été utilisés. Puisque le domaine K1 en fusion à la GST ou à ZZ a toujours une étiquette de 6 histidines en N-terminal, il serait intéressant d'effectuer deux affinités successives (colonne nickel puis GS4B ou colonne nickel puis IgG) pour augmenter la spécificité d'interaction, comme dans l'approche Tap-Tag (Rigaut et al. 1999; Puig et al. 2001) par exemple qui a fait largement ses preuves.

## 4. L'ISOLEMENT DE COMPLEXES ASSOCIES AU DOMAINE C-TERMINAL

L'isolement de protéines associées au domaine C-terminal de la protéine KIN17 par des méthodes d'affinité (*pull-down*, cf paragraphe 3) ayant été infructueuse dans notre cas, nous avons adopté une stratégie alternative à la recherche de partenaires qui consiste en la purification de complexes. Pour cela, nous avons mis au point un protocole de purification des complexes associés à cette protéine par des méthodes biochimiques telles que la filtration sur gel et le gel ND-IEF. L'intérêt ici est que nous n'avons pas d'étiquette associée à la protéine qui pourrait interférer avec l'assemblage des complexes ou la fonction de la protéine.

Nous avons adopté la stratégie la plus couramment employée pour la caractérisation de complexes de protéines qui consiste en la purification de ces complexes intacts, en quantité suffisante et de pureté appropriée suivie de l'identification des différents composants par spectrométrie de masse. Plusieurs étapes de purification ont été mises au point pour isoler les complexes bien que nous ayons essayé de réduire la durée totale de l'expérience pour préserver des complexes stables.

J'ai été très aidée et conseillée par le Dr. A. Galat du SIMOPRO/CEA pour les expériences d'isolement de complexes

J'ai collaboré avec l'équipe du Dr. J.Chamot-Rooke du laboratoire DCMR de l'Ecole Polytechnique pour les analyses de spectrométrie de masse, réalisées par le Dr. C. Malosse.

### 4.1. Association du domaine C-terminal à un complexe de haut poids moléculaire

Nous avons d'abord cherché à isoler des complexes potentiellement formés avec le domaine C-terminal (K1) de la protéine hKIN17 dans des extraits de cellules HeLa par filtration moléculaire. La technique de filtration sur gel est une méthode chromatographique qui permet de fractionner des protéines et des complexes en fonction de leur taille apparente. Le domaine C-terminal de la protéine hKIN17 est ajouté en excès (10 à 40  $\mu\text{M}$ ) à des extraits nucléaires de cellules Hela avec lesquels il est mis à incuber. Après équilibration, le mélange de protéines est fractionné sur une colonne de gel filtration.

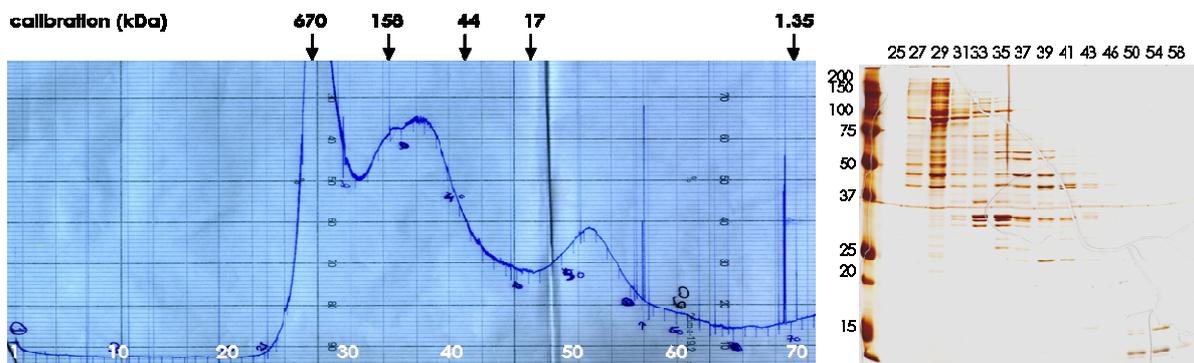


Figure 4-1 : A gauche, profil de la filtration sur gel sur colonne Sephacryl S200 d'un mélange d'extraits nucléaires de cellules HeLa et du domaine K1 à 10 $\mu\text{M}$ . Une calibration de la colonne a été faite dans les mêmes conditions avec un mélange commercial (Biorad, 151-1901) contenant les protéines thyroglobuline,  $\gamma$ -globuline,

ovalbumine, myoglobine et la vitamine B-12 et les résultats de cette calibration sont reportés au-dessus du profil de gel filtration. Le numéro des fractions est indiqué en-dessous du profil en blanc. A droite, gel SDS-PAGE 12% coloré au nitrate d'argent de certaines fractions entre les fractions 25 et 58 de la filtration sur gel.

La Figure 4-1 montre que la séparation des protéines selon leur taille est efficace sur la colonne, c'est-à-dire que les premières fractions sont enrichies en protéines de haut poids moléculaire et les dernières fractions en protéines de petit poids moléculaire. Ensuite, la présence du domaine K1 dans les différentes fractions est recherchée par immuno-blotting.



Figure 4-2 : Immunoblot, avec des anticorps anti-KIN17 (K36) reconnaissant le domaine K1, des fractions 24 à 58 de la filtration sur gel présentée sur la Figure 4-1. Le puits K1 correspond au domaine K1 directement déposée sur gel comme contrôle de l'immunoblotting.

D'après les résultats de la Figure 4-2, la protéine est éluée essentiellement dans les fractions correspondant à son poids moléculaire (autour de 15 kDa, à partir des fractions 50). Il apparaît aussi que la protéine traîne dans les fractions 40 à 50 précédant celles correspondant à son volume d'élué observé lorsqu'elle est déposée seule sur filtration sur gel. De plus, une fraction de K1 apparaît dans des fractions de haut poids moléculaire, suggérant que, dans ces fractions, le domaine K1 est retenu par des protéines et/ou des acides nucléiques dans un ou des complexes de haut poids moléculaire. La colonne de filtration sur gel Sephacryl S200 utilisée permet de séparer de façon optimale des protéines entre 10 et 250 kDa. Le ou les complexe(s) que nous observons dans les fractions au début de la séparation est (sont) donc un (des) complexe(s) de taille supérieure à 300 kDa mais nous ne pouvons pas en déterminer la taille précise avec cette expérience.

Nous avons ensuite regardé si la protéine était toujours associée à des complexes de haut poids moléculaire lorsque sa concentration était diminuée d'un facteur 10 pour s'assurer que ce qui était observé n'était pas un effet dû à la concentration élevée de protéine. Pour cela, deux filtrations sur gel ont été comparées : la première réalisée dans les conditions classiques de concentration du domaine K1 et d'extraits nucléaires (1.5 ml d'extraits nucléaires avec le domaine K1 à 45  $\mu$ M) et la seconde réalisée en prenant 10% du mélange précédent ajouté à 1.5 ml d'extraits nucléaires.

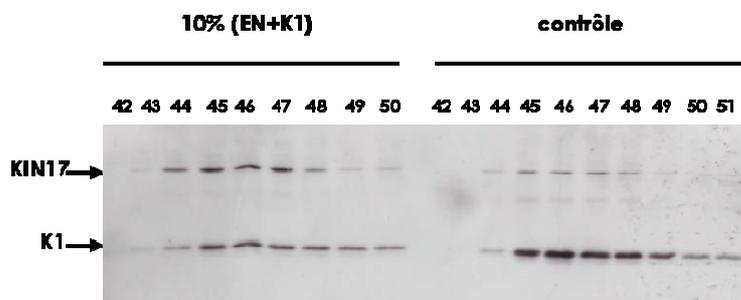


Figure 4-3 : Immunoblot avec les anticorps K36 des fractions 42 à 50, fractions entre lesquelles les retards d'élué des protéines K1 et KIN17 entière sont observés pour les deux filtrations sur gel. La filtration sur gel contrôle est réalisée dans les conditions classiques et la filtration sur gel 10% (EN+K1) est réalisée comme cela est décrit dans le texte. Les puits 42 à 50 correspondent aux puits 25 à 36 de la figure précédente.

D'après les résultats présentés en Figure 4-3, le retard d'éluion du domaine K1 est observé dans les mêmes fractions de haut poids moléculaire pour les deux filtrations sur gel. Par contre, l'intensité du domaine K1 détectée par immunoblot est un peu différente pour les deux filtrations sur gel. Alors qu'on aurait pu s'attendre à observer la même quantité de protéine retardée dans les deux expériences puisque K1 est ajoutée en large excès, il semble qu'il y ait plus de domaine K1 retardé pour la première filtration (contrôle). En même temps, la protéine KIN17 entière est aussi détectée dans les fractions de haut poids moléculaire et il semble qu'il y ait plus de protéine KIN17 entière détectée pour la seconde filtration sur gel (10% EN+K1). Ainsi, la présence du domaine K1 dans les complexes de haut poids moléculaire est toujours observée même lorsque la protéine est ajoutée en plus faible quantité et la présence de la protéine entière KIN17 dans ces mêmes fractions est d'autant moins forte qu'il y a plus de domaine K1. Nous avons confirmé la présence de la protéine KIN17 entière dans les fractions de haut poids moléculaire en fractionnant les extraits nucléaires sans ajouter de protéine. Cela est en accord avec l'observation montrant que la protéine KIN17 est associée à un complexe de 600 kDa dans des extraits nucléaires de cellules HeLa (Miccoli et al. 2005). Par ailleurs, une filtration moléculaire d'extraits nucléaires auxquels le domaine N-terminal de KIN17 (K3) est ajouté en excès ne montre pas de retard d'éluion de la protéine K3. Ces expériences suggèrent que la présence de la protéine entière KIN17 dans des complexes de haut poids moléculaire est attribuable essentiellement à sa région C-terminale.

Nous avons ensuite étudié l'effet de différents composés sur la présence du domaine K1 dans des complexes de haut poids moléculaire. En particulier, l'effet d'un traitement des extraits nucléaires avec des nucléases (DNase et RNase) avant équilibrage avec le domaine K1 a été testé. Le retard du domaine K1 est observé indifféremment avec les extraits nucléaires contrôle et traité, suggérant que l'ajout de nucléase dans les extraits nucléaires n'empêche pas le domaine K1 d'être associée aux complexes de haut poids moléculaire. Deux hypothèses sont alors possibles : il n'y a pas d'acides nucléiques dans les complexes isolés ou les acides nucléiques ne sont pas accessibles aux nucléases.

## **4.2. Association du domaine C-terminal à un complexe de pI global acide**

Nous avons ensuite cherché à isoler des complexes potentiellement formés avec le domaine C-terminal (K1) de la protéine hKIN17 dans des extraits de cellules HeLa par gel ND-IEF. La technique de gel ND-IEF permet de fractionner des protéines et des complexes en fonction de leur pI apparent, en concentrant les protéines. Le domaine C-terminal de hKIN17 est ajouté en excès (45  $\mu$ M) à des extraits nucléaires de cellules HeLa avec lesquels il est mis à incuber. Après équilibrage, le mélange de protéines est fractionné sur gel ND-IEF. Après migration et focalisation des protéines, le gel est découpé en 13 fractions, du côté acide vers le côté basique.

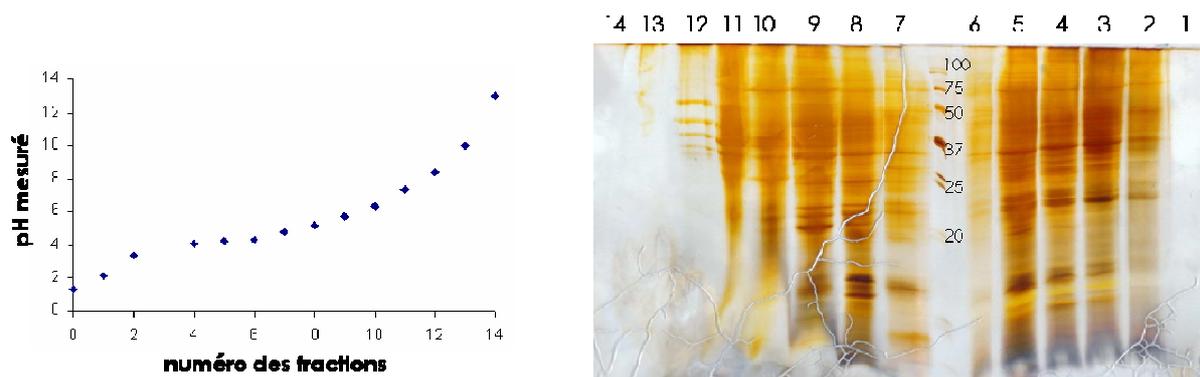


Figure 4-4 : A gauche, dépendance du pH en fonction du numéro des fractions issues de la séparation d'un mélange d'extraits nucléaires de cellules HeLa et du domaine K1 à 45 $\mu$ M sur gel ND-IEF. Les fractions 0 et 14 correspondent respectivement aux électrodes positives et négatives du gel. A droite, gel SDS-PAGE 12% coloré au nitrate d'argent des fractions 1 à 14 du gel ND-IEF.

Le pH varie de manière régulière entre 1,30 et 13,00 indiquant que le gel ND-IEF s'est bien déroulé. De plus, d'après le gel des protéines des différentes fractions, il apparaît assez clairement que des protéines se sont focalisées dans des gammes de pH restreintes. La présence du domaine K1 de la protéine hKIN17 dans les différentes fractions est recherchée par immuno-blotting.

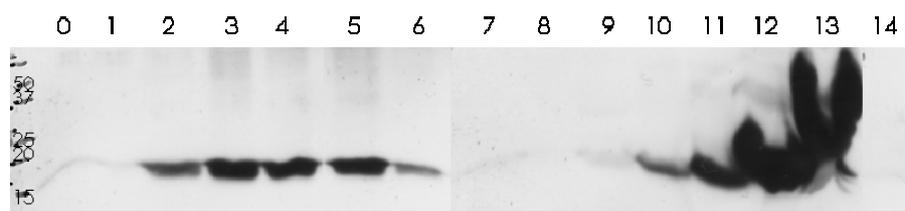


Figure 4-5 : Immunoblot avec les anticorps K36 des fractions 0 à 14 du gel ND-IEF présenté sur la Figure 4-4.

D'après les résultats de la Figure 4-5, le domaine K1 est essentiellement focalisé dans les fractions de pH basique correspondant à son pI (pI=9.1). Il apparaît aussi qu'une fraction du domaine K1 est focalisée dans des fractions à pH plus acide (pH autour de 4), suggérant que le domaine K1 a été entraîné dans cette zone de pH par d'autres protéines et/ou des acides nucléiques et/ou des particules cellulaires ou des biomembranes. Sur l'immunoblot, on ne détecte pas la présence de la protéine entière KIN17. En revanche, la même expérience de fractionnement sur gel ND-IEF a été faite avec des extraits nucléaires seuls, c'est-à-dire non enrichis en domaine K1 ou KIN17. L'immunoblotting des fractions de ce gel a permis de détecter une faible bande de KIN17 entière endogène dans les fractions basiques correspondant à son pI (pI=9.6) et dans les fractions autour de pH 4. Ces résultats montrent que le domaine C-terminal de la protéine KIN17 est associé à un complexe de pI global acide. Il est ensuite intéressant (i) d'étudier quelles sont les autres protéines présentes dans ces fractions à pH acides, en particulier pour savoir si ce sont des protéines à pI acides ou d'autres types de protéines également entraînées ici et (ii) d'étudier si le pH acide est dû à la présence d'acides nucléiques dans ces fractions.

Nous avons ensuite réalisé la même expérience avec des extraits nucléaires traités à la benzonase, une endonucléase capable de dégrader toutes les formes d'ADN et d'ARN, pour voir si cela changeait la répartition du domaine K1 dans les fractions du gel ND-IEF. La

protéine apparaît toujours dans les fractions à pI basique et dans les fractions autour de pH 4. Deux hypothèses peuvent expliquer ces résultats : il n'y a pas d'acides nucléiques dans les fractions à pH 4 ou les acides nucléiques présents dans ces fractions sont saturés en protéines et donc ne sont pas accessibles pour la benzonase ; des acides nucléiques nus se focalisant plutôt au niveau de l'électrode positive.

### **4.3. Combinaison de la filtration moléculaire et du gel ND-IEF**

Lorsque les extraits nucléaires équilibrés avec le domaine K1 sont d'abord fractionnés sur colonne de filtration sur gel et que les fractions correspondant aux complexes de haut poids moléculaire sont concentrées et séparées sur gel ND-IEF, le domaine K1 est alors détectable par immuno-blotting dans les gammes de pH correspondant à son pI mais aussi dans les fractions de pH acide autour de 4,5. Inversement, lorsque les extraits nucléaires équilibrés avec le domaine K1 sont d'abord fractionnés sur gel ND-IEF et que les fractions de pH acide (entre 3,2 et 4,8 environ) sont éluées, concentrées et séparées sur une colonne de filtration sur gel, le domaine K1 est détectable par immuno-blotting dans les fractions correspondant à son poids moléculaire mais aussi dans les fractions de haut poids moléculaire, correspondant aux complexes de haut poids moléculaire.

Les fractions du gel ND-IEF ou de la filtration sur gel dans lesquelles le domaine K1 est détecté à un pH différent de son pI ou à une taille différente de celle attendue pour sa masse moléculaire sont riches en protéines. Les deux expériences précédentes suggèrent que parmi les nombreuses protéines, des protéines sont communes aux complexes isolés par ND-IEF et par filtration sur gel. Ainsi, pour isoler les protéines en complexe avec le domaine C-terminal de la protéine KIN17, nous pouvons choisir (i) soit de faire successivement les deux purifications de complexes, l'ordre étant indifférent ; des protéines contaminantes différentes sont alors éliminées par gel ND-IEF et filtration sur gel, (ii) soit de faire en parallèle les deux purifications de complexes et comparer les protéines isolées pour ne retenir que les protéines communes.

### **4.4. Identification des protéines présentes dans le complexe associé au domaine C-ter**

#### ***4.4.1. Isolement du ou des complexes***

Nous avons ensuite cherché à isoler les protéines en complexe avec le domaine C-terminal de la protéine hKIN17. La Figure 4-6 présente schématiquement les différentes étapes de l'expérience préparative.

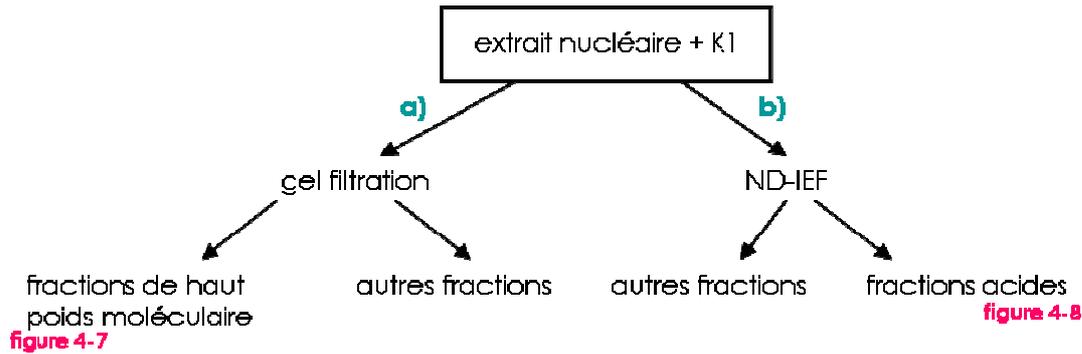


Figure 4-6 : Schéma présentant les différentes étapes de l'expérience préparative des complexes associés au domaine C-terminal de hKIN17 dont les résultats sont présentés ci-dessous.

a) Une fraction des extraits nucléaires équilibrés avec le domaine K1 a été fractionnée par filtration sur gel. La filtration sur gel a bien permis de fractionner les protéines selon leur taille et par immunoblotting, le domaine K1 est encore observé dans les fractions correspondant à son poids moléculaire et dans des fractions de haut poids moléculaire. Les fractions sont donc concentrées en trois groupes : les fractions de haut poids moléculaire dans lesquelles le domaine K1 est détecté (pic 1, Figure 4-7), les fractions de poids moléculaire intermédiaire (pic 2) et les fractions de petits poids moléculaire (pic 3).

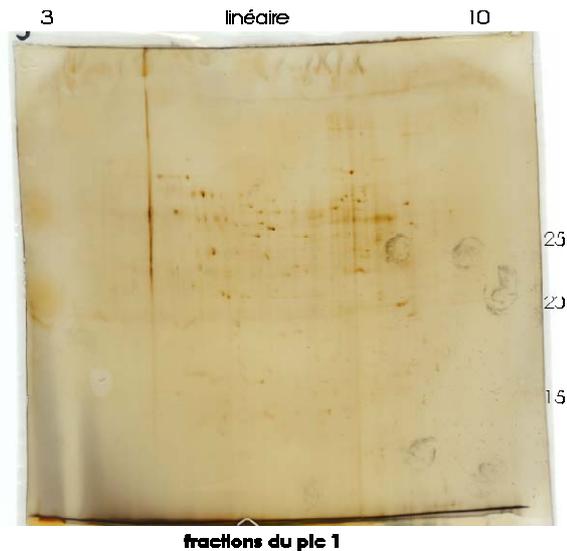


Figure 4-7 : Gel bi-dimensionnel (3-10 linéaire, 15%) coloré au nitrate d'argent d'un aliquote des fractions du pic 1 de la filtration sur gel sur colonne Sephacryl S200 d'extraits nucléaires équilibrés avec le domaine K1 (à 40µM).

b) En parallèle, le reste des extraits nucléaires équilibrés avec le domaine K1 a été fractionné sur un gel ND-IEF. La répartition en pH est satisfaisante et varie de 1,28 à 12,48. Alors que des focalisations spécifiques sont observées pour les différentes protéines, le domaine K1, en plus d'être détecté dans les fractions basiques proches de son pI, est détecté par immunoblotting dans les fractions à pH acide. Sur la Figure 4-8 est donné le gel bidimensionnel coloré au nitrate d'argent d'une fraction à pH acide du gel ND-IEF dans laquelle la présence du domaine K1 a été détectée.

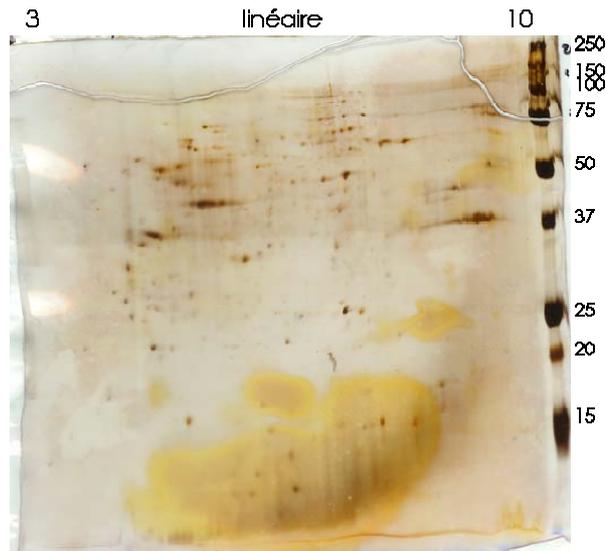


Figure 4-8 : Gel bi-dimensionnel (3-10 linéaire, 15%) coloré au nitrate d'argent d'un aliquote d'une fraction à pH acide (pH=3-4) du gel ND-IEF d'extraits nucléaires équilibrés avec le domaine K1 (à 40 $\mu$ M).

#### 4.4.2. Immunoprécipitation et identification

Il apparaît des taches en commun entre les gels 2D des protéines isolées dans les fractions de haut poids moléculaire par filtration moléculaire et les gels 2D des protéines isolées dans les fractions de pH acide par gel ND-IEF, suggérant qu'il existe des protéines communes entre ces fractions. Pour une recherche plus spécifique des protéines associées au domaine C-terminal de KIN17 dans les fractions de haut poids moléculaire ou de pH acide, nous avons utilisé la méthode d'immunoprécipitation avec des anticorps anti-KIN17. Nous avons eu plusieurs approches en parallèle représentées schématiquement sur la Figure 4-9.

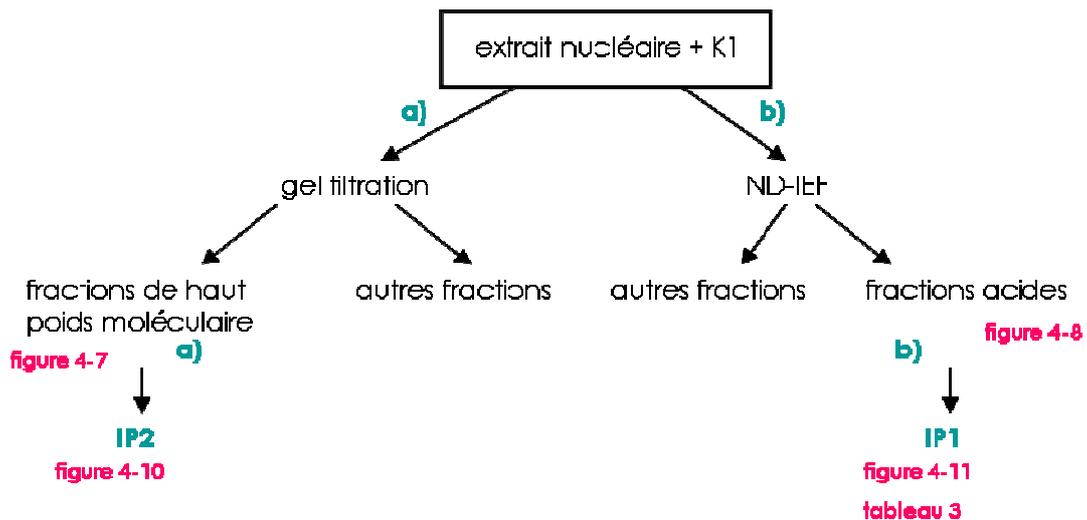


Figure 4-9 : Schéma présentant les différentes étapes de l'expérience dont les résultats sont présentés ci-dessous.

a) Une immunoprécipitation (IP2) avec des anticorps anti-KIN17 (mAbK36) a été effectuée avec les fractions de haut poids moléculaire issues de la filtration sur gel d'un

mélange d'extraits nucléaires équilibrés avec le domaine K1 et dans lesquelles K1 a été détectée par immuno-blotting (pic 1).

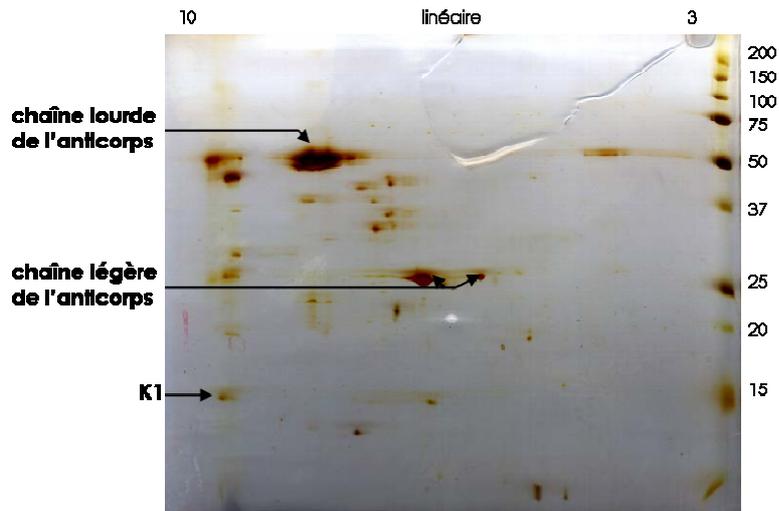


Figure 4-10 : Gel bi-dimensionnel (3-10 linéaire, 15%) coloré au nitrate d'argent des protéines isolées par immunoprécipitation avec des anticorps anti-KIN17 à partir des fractions de haut poids moléculaire d'une filtration sur gel d'un mélange d'extraits nucléaires équilibrés avec le domaine K1.

D'après le gel de la Figure 4-10, l'immunoprécipitation a été efficace puisque de la protéine K1 a été immunoprécipitée par les anticorps. De plus, d'autres taches de protéines sont visibles sur gel et peuvent potentiellement correspondre à des protéines en interaction avec le domaine C-terminal de KIN17. Le séquençage par spectrométrie de masse MS/MS de ces protéines a été infructueux du fait d'un signal trop faible, provenant d'une quantité trop faible de protéine.

b) Une immunoprécipitation (IP1) avec des anticorps anti-KIN17 (mAb K36) a été effectuée avec les fractions acides concentrées du gel ND-IEF dans lesquelles le domaine K1 a été détecté par immuno-blotting.

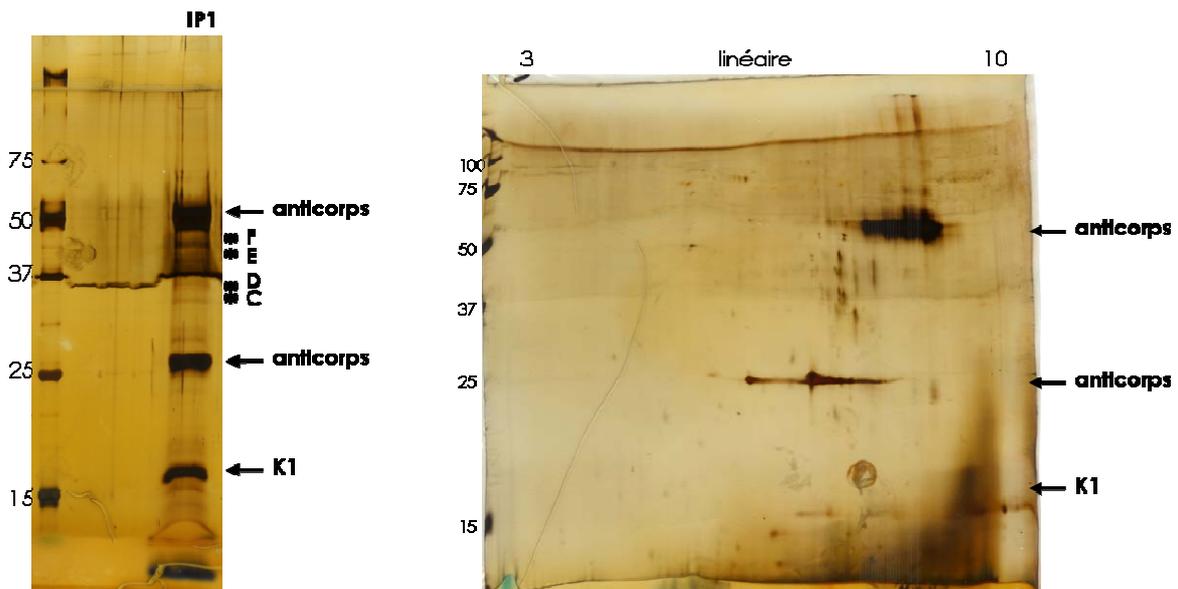


Figure 4-11 : Résultat de l'immunoprécipitation du domaine K1 à partir des fractions acides du gel ND-IEF sur gel 1D 12% et sur un gel bidimensionnel (3-10 linéaire, 15%) colorés au nitrate d'argent. Les bandes de protéines marquées d'un astérisque correspondent aux bandes de protéines séquencées par spectrométrie de masse MS/MS.

Le gel analytique de l'immunoprécipitation (Figure 4-11) montre que l'immunoprécipitation a été efficace puisque la bande correspondant à la protéine K1 est forte comparativement aux bandes correspondant aux anticorps. Un gel préparatif avec tout le reste des protéines immunoprécipitées a ensuite été réalisé et coloré au bleu de Coomassie. Quatre bandes de protéines, bien visibles à la coloration et correspondant aux bandes de protéines marquées d'un astérisque sur la Figure 1-1, ont été analysées par spectrométrie de masse MS/MS.

Tableau 3 : Résultat de l'identification des protéines C, D, E et F visibles sur le gel présenté en Figure 4-11 par spectrométrie de masse MS/MS.

bande sur gel	PM (kDa) estimé sur gel	identification
C	32	hnRNP A2B1
D	35	hnRNP A2B1
E	45	pas assez de matériel
F	45	pas assez de matériel

Nous avons identifié la protéine hnRNPA2/B1 comme un composant du complexe associé au domaine C-terminal de hKIN17. Cette protéine a un pI basique (pI théorique= 8.97), suggérant que sa présence dans les fractions acides des gels ND-IEF est aussi due à son association à un complexe de pI global acide. Par ailleurs, une immunoprécipitation de la protéine hKIN17 endogène avec un mélange d'anticorps mAbK36 et mAbK58 dans des extraits totaux de cellules HeLa a permis de retenir également cette protéine hnRNPA2/B1. La protéine hnRNPA2/B1 est une protéine nucléaire appartenant à la famille des ribonucléoprotéines, des protéines abondantes en interaction avec les hnRNA dans les complexes du spliceosome.

*En conclusion*, à partir de ces résultats, il paraît difficile de dire à quel complexe KIN17 appartient puisque nous n'avons pour le moment pas isolé beaucoup de protéines en complexe avec hKIN17. C'est pourquoi, nous prévoyons d'effectuer des identifications supplémentaires.

---

## 5. EST-CE QUE LA PROTEINE JOUE UN ROLE DANS L'ÉPISSAGE DES ARN PRE-MESSAGERS ?

---

Différentes observations présentées en introduction suggèrent une intervention de la protéine hKIN17 dans les mécanismes d'épissage. En particulier, hKIN17 se lie directement à l'ARN *in vivo* (Pinon-Lataillade et al. 2004) et des analyses par protéomique du spliceosome humain (Makarov et al. 2002; Rappsilber et al. 2002; Makarova et al. 2004) ont identifié la protéine KIN17 comme étant associée à certains complexes de l'épissage, en particulier le complexe B. De plus, parmi les protéines identifiées comme partenaires potentiels des domaines K1 ou K2, nous avons retenu plusieurs protéines intervenant dans le métabolisme de l'ARN. Les protéines p68, DDX21 et eIF-4A, toutes des hélicases de l'ARN, appartiennent à la famille des DEAD box et une des sept protéines Sm, la protéine SmD2 même s'il peut paraître intrigant de la trouver seule. Dans les complexes associés au domaine C-terminal et à la protéine hKIN17, nous avons également identifié une ribonucléoprotéine, hnRNPA2/B1. Ces identifications nous ont encouragés à déterminer si la protéine KIN17 humaine a effectivement un rôle biologique dans les mécanismes d'épissage des ARN pré-messagers.

Je suis allée à plusieurs reprises dans le laboratoire du Pr. R. Lührmann du MPI, Göttingen, pour effectuer les expériences suivantes.

### 5.1. Pull-down de réaction d'épissage

Cette expérience a été réalisée pour tester si la protéine KIN17 s'associe à des protéines impliquées dans l'épissage des ARN pré-messagers, en particulier aux protéines Sm puisque la protéine SmD2 a été identifiée par spectrométrie de masse parmi les protéines retenues par le domaine K1 en fusion à la GST (cf Paragraphe 3.2).

L'expérience consiste à fixer la protéine His-KIN17 (hsKIN17) sur une colonne nickel et à incuber ces billes avec une réaction d'épissage d'un ARN pré-messenger MINX (MINX est une construction artificielle d'un ARN pré-messenger avec un seul exon dérivé d'une unité de transcription de l'adénovirus 2). Après différents lavages, la présence des snRNAs et des protéines retenues par la protéine HsKIN17 est révélée sur gel (Figure 5-1).

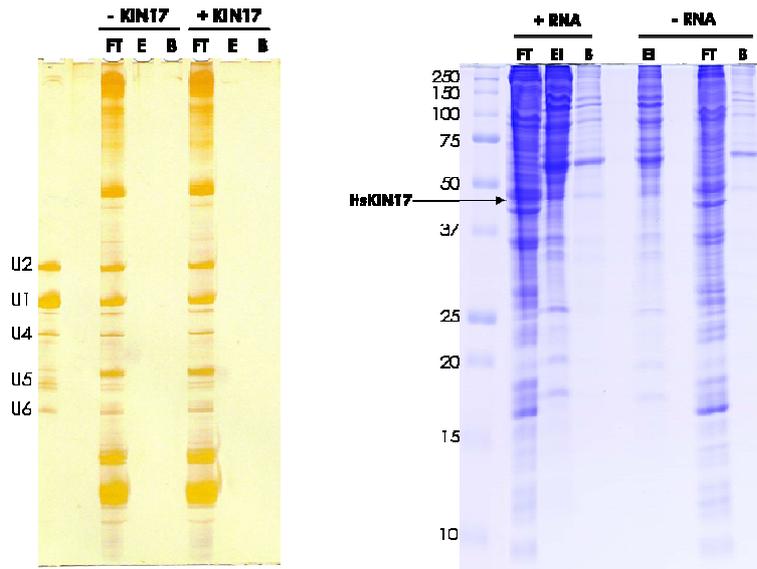


Figure 5-1 : Résultats du pull-down avec la protéine recombinante His-hKIN17 d'une réaction d'épissage d'un ARN pré-messager MINX de 45 minutes. A gauche, visualisation sur un gel 8.3M urée-10% polyacrylamide coloré à l'argent de l'ARN extrait du flow-through (FT), de l'élution (E) et des billes après élution (B). A droite, visualisation sur un gel SDS-PAGE 15% coloré au bleu de Coomassie des protéines du flow-through (FT), de l'élution (E) et des billes après élution (B).

D'après les résultats de la Figure 5-1, la protéine HsKIN17 ne retient aucun des snRNAs. En effet, le profil de l'élution et des billes de nickel après incubation de la réaction d'épissage avec la protéine HsKIN17 ne présente pas d'enrichissement en l'un des snRNA par rapport à celui du contrôle (sans KIN17). De même, la protéine HsKIN17 semble ne retenir aucune protéine particulière puisque le profil protéique d'élution et des billes après incubation de la réaction d'épissage avec la protéine HsKIN17 ressemble au profil du contrôle (sans ARN). De la même manière, une immunoprécipitation de la réaction d'épissage avec des anticorps anti-KIN17 (K36 et K58) a montré que les snRNAs (de U1 à U6) ne semblent pas spécifiquement retenus par la protéine KIN17.

## 5.2. Pull-down des snRNPs

Dans cette expérience, les protéines appât (KIN17 entière ou les domaines K1 et K3 en fusion à la GST) sont fixées sur des billes. Les billes sont ensuite mises à incuber avec un mélange de snRNPs, c'est-à-dire avec les différents complexes de l'épissage. Après différents lavages, la présence des snRNAs et des protéines retenues par les protéines appât est révélée sur gel (Figure 5-2).

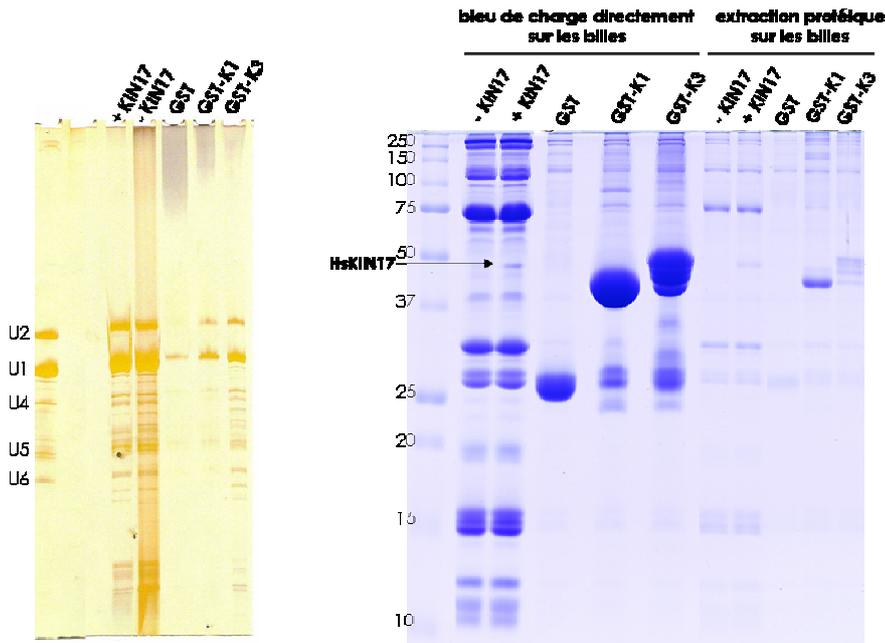


Figure 5-2 : Résultats du pull-down avec la protéine recombinante His-hKIN17 et les domaines K1 et K3 en fusion à la GST d'un mélange de snRNPs. La GST sert de contrôle au pull-down avec les domaines. A gauche, visualisation sur un gel 8.3M urée-10% polyacrylamide coloré à l'argent de l'ARN extrait des billes. A droite, visualisation sur un gel SDS-PAGE 15% coloré au bleu de Coomassie des protéines des billes sur lesquelles le bleu de charge est déposé directement et des protéines extraites des billes.

D'après les résultats de la Figure 5-2, la protéine HsKIN17 ne retient pas plus de protéines que le contrôle (sans HsKIN17). Puisque le contrôle de la colonne nickel sans protéine HsKIN17 retient déjà beaucoup de protéines, nous avons refait l'expérience en ajoutant une étape de *pre-clearing* sur la colonne nickel. L'utilisation d'extraits après *pre-clearing* ne nous a pas non plus permis d'isoler des protéines associées à la protéine HsKIN17. Les domaines de la protéine en fusion à la GST ne retiennent pas significativement plus de protéines que la GST seule.

*En conclusion*, nous n'avons pas réussi à isoler de protéines des réaction d'épissage dans des expériences de *pull down* avec la protéine KIN17, entière ou sous forme de domaines, comme protéine appât. Il est possible que les interactions de la protéine KIN17 avec d'autres composants du *spliceosome* soient transitoires et de faible affinité et donc difficilement isolables dans les conditions expérimentales utilisées.

### 5.3. Test d'épissage par immunodéplétion-complémentation

Tout d'abord, la protéine est retirée physiquement des extraits nucléaires et l'effet de la déplétion de la protéine sur l'efficacité de l'épissage est étudié. Dans un second temps, la protéine recombinante est rajoutée aux extraits déplétés pour voir si un épissage normal est retrouvé. Ces deux étapes sont nécessaires pour conclure à un rôle de la protéine étudiée dans l'épissage. Le principe de cette expérience peut être schématisé de la manière suivante (Figure 5-3) :

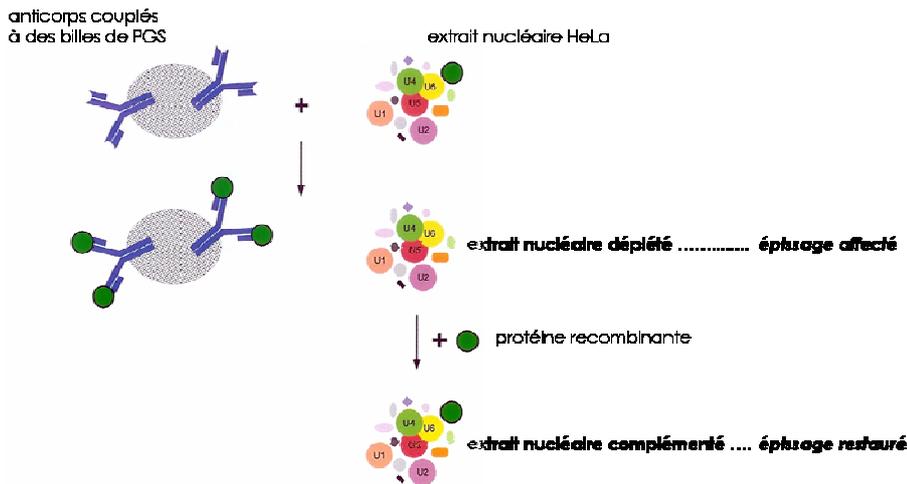


Figure 5-3 : Représentation schématique des étapes principales d'un test d'épissage par immunodéplétion-complémentation. Cette figure a été inspirée par Berkta Akyildiz du MPI, Göttingen.

### 5.3.1. Obtention d'extraits nucléaires immuno-déplétés

Pour commencer, nous avons cherché les conditions optimales pour obtenir des extraits nucléaires déplétés en KIN17. Pour l'immuno-déplétion, nous disposons de deux anticorps monoclonaux (mAbK36 et mAbK58) dirigés contre la protéine KIN17. Nous avons également testé la déplétion en différentes conditions de sel (200mM et 500mM NaCl).

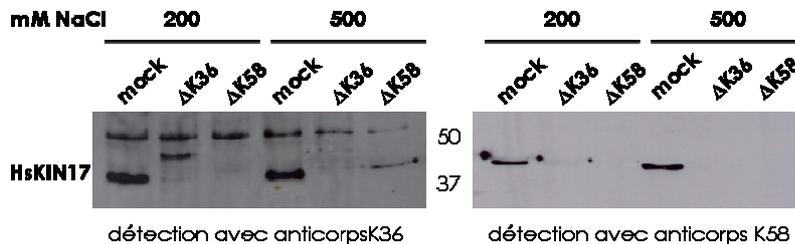


Figure 5-4 : Western-blot de l'immuno-déplétion d'extraits nucléaires avec K36 et K58, en 200mM NaCl ou 500mM NaCl. La condition *mock* correspond au contrôle, sans anticorps.

L'immuno-déplétion de KIN17 avec l'un ou l'autre des anticorps monoclonaux K36 et K58, en 200 et 500 mM NaCl est très efficace puisque le signal de KIN17 disparaît totalement sur le western blot alors qu'il est de forte intensité pour les contrôles (Figure 5-4). De plus, l'immuno-déplétion en forte concentration de sel (500mM NaCl) permet de limiter la co-déplétion de protéines qui pourraient être associées à KIN17 puisque on peut supposer qu'à une telle concentration en sels les interactions avec KIN17 sont dissociées.

### 5.3.2. Efficacité d'épissage des extraits nucléaires immuno-déplétés et contrôles

Les extraits nucléaires déplétés sont ensuite testés pour évaluer l'efficacité d'épissage *in vitro* d'un fragment d'ARN pré-messager (MINX) radioactif et voir quelle(s) condition(s) affecte(nt) le plus l'épissage par rapport au contrôle. Pour cela, des essais avec différentes quantités d'extraits nucléaires (20, 25, 30, 35 et 40%) dans la réaction d'épissage ont été réalisés (Figure 5-5).

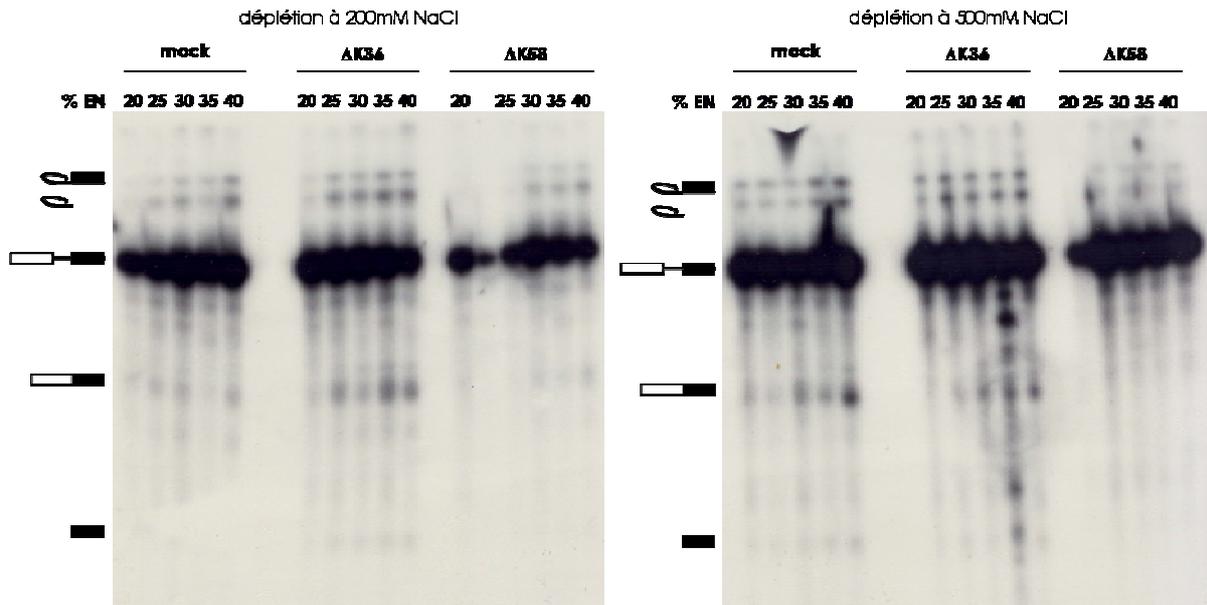


Figure 5-5 : Epissage *in vitro* d'un ARN pré-messager MINX radioactif avec des extraits nucléaires contrôle (mock) et immuno-déplétés en KIN17 avec K58 et K36 en 200mM et 500mM NaCl, pour différentes proportions d'extraits nucléaires dans la réaction d'épissage. L'ARN est analysé sur un gel 8.3M urée-14% acrylamide et visualisé par exposition sur un film radio-sensible. La position de l'ARN pré-messager et des intermédiaires et produits de l'épissage est indiquée sur la gauche.

La plus grande différence d'efficacité d'épissage entre les extraits contrôles et immuno-déplétés est observée pour les extraits nucléaires immuno-déplétés avec mAb K58 en 500mM NaCl comparativement aux extraits nucléaires contrôle en 500mM NaCl. Nous avons ensuite confirmé cette différence d'efficacité d'épissage (Figure 5-6).

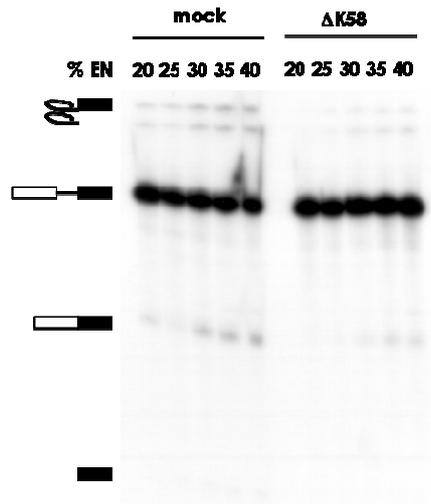


Figure 5-6 : Epissage *in vitro* d'un ARN pré-messager MINX radioactif avec des extraits nucléaires contrôle (mock) et immuno-déplété en KIN17 avec K58 en 500mM NaCl, pour différentes proportions d'extraits nucléaires dans la réaction d'épissage. L'ARN est analysé sur un gel 8.3M urée-14% acrylamide et visualisé par autoradiographie au Phosphorimager. La position de l'ARN pré-messager et des intermédiaires et produits de l'épissage est indiquée sur la gauche.

### 5.3.3. Complémentation de l'épissage par ajout de protéine recombinante

Dans un second temps, nous testons si l'épissage peut être restauré au niveau du contrôle par ajout de protéine KIN17 recombinante dans les extraits nucléaires déplétés. Pour cela, les conditions d'épissage utilisées sont classiques et la durée de la réaction varie (0, 15, 30, 45 et 60 minutes). La quantité d'extrait nucléaire utilisé est de 40% puisque d'après les résultats précédents (Figure 5-7), une différence significative de l'efficacité d'épissage est observée pour cette proportion d'extrait nucléaire dans la réaction d'épissage. Pour la complémentation, deux quantités de protéine KIN17, recombinante et purifiée selon le protocole décrit dans Miccoli et al., 2002, sont testées (2 pmole/ $\mu$ l et 4 pmole/ $\mu$ l) puisque nous n'avons *a priori* pas d'idées de la quantité de protéine nécessaire pour restaurer l'épissage.

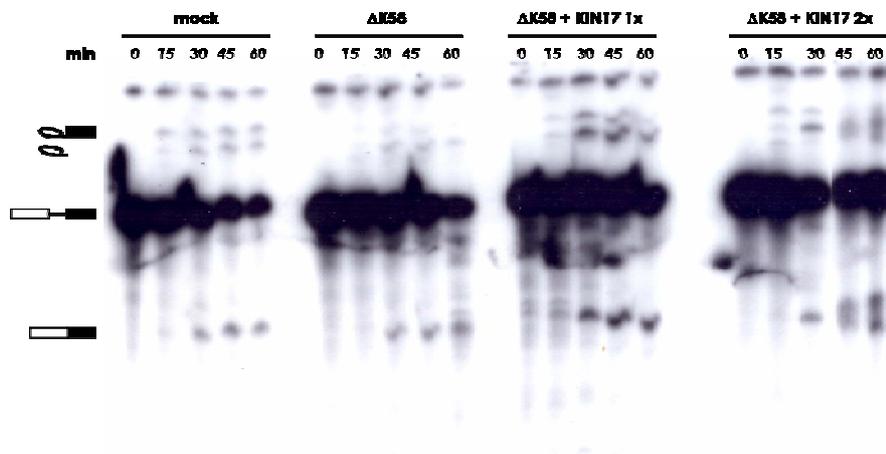


Figure 5-7 : Epissage *in vitro* d'un ARN pré-messager MINX radioactif avec des extraits nucléaires contrôle (mock), déplété en KIN17 ( $\Delta$ K58), complétement avec KIN17 recombinante à 2 pmole/ $\mu$ l (KIN17 1x) et

complémenté avec KIN17 recombinante à 4 pmole/ $\mu$ l (KIN17 2x), pour différentes durées de réaction. L'ARN est analysé sur un gel 8.3M urée-15% acrylamide et visualisé par exposition sur un film radio-sensible. La position de l'ARN pré-messager et des intermédiaires et produits de l'épissage est indiquée sur la gauche.

Dans cet essai, la différence entre l'efficacité d'épissage avec les extraits déplété et contrôle est moins nette que précédemment. Par contre, il semble que l'ajout de protéine recombinante augmente l'efficacité d'épissage. Il reste néanmoins difficile de conclure puisque le signal avec l'extrait déplété est assez fort comparativement à celui du contrôle.

#### **5.3.4. Comment optimiser les conditions d'épissage et de complémentation ?**

La protéine KIN17 n'est pas une enzyme ou un composant essentiel de l'épissage. Il se peut donc qu'elle intervienne faiblement dans l'épissage. Ainsi, pour observer un effet de la déplétion de cette protéine et de la complémentation avec cette protéine, il est nécessaire d'optimiser les conditions d'épissage.

Afin d'améliorer cette expérience, nous avons travaillé également avec une autre concentration d'extrait nucléaire (20%) dans la réaction d'épissage. En effet, à cette concentration d'extrait nucléaire, d'après les résultats de la Figure 5-6, des produits de l'épissage sont observés pour l'extrait contrôle mais pas pour l'extrait déplété. De plus, en ajoutant encore moins d'extrait nucléaire, la quantité de protéine hKIN17 qui pourrait rester dans l'extrait immuno-déplété est diminuée. Ce dernier point peut être critique si la protéine n'est pas un facteur essentiel de l'épissage ; pour voir un effet il faut alors enlever le maximum de protéine. Malheureusement, nous n'avons pas observé d'effets plus clairs dans ces conditions. Nous avons également fait varier la quantité d'ARN à épisser de 10 à 20mM. Là encore, nous n'avons pas obtenu d'effets plus clairs dans ces conditions. Par ailleurs, l'ARN pré-messager MINX est un ARN simple à épisser, qui ne nécessite peut-être pas l'intervention de beaucoup de co-facteurs, nécessaires dans le cas de l'épissage d'un ARN pré-messager plus complexe. C'est pourquoi, nous avons aussi testé un autre ARN à épisser, l'ARN pré-messager de la globine (composé de deux exons et d'un intron). Là encore, l'effet n'apparaît pas plus nettement.

Il se peut aussi que la protéine recombinante produite dans Baculovirus ne permette pas d'avoir une complémentation efficace. Elle peut aussi n'être pas bien repliée en fin de purification et donc ne pas être fonctionnelle.

#### **5.3.5. Effet de la déplétion de KIN17 sur la formation des complexes d'épissage**

Pour déterminer à quelles étapes la déplétion de KIN17 perturbe la formation des complexes de l'épissage, nous avons suivi la formation des différents complexes d'épissage sur gel natif. Le complexe H apparaît initialement, le complexe A apparaît dès 2 minutes, puis le complexe B apparaît après 5 minutes et le complexe C apparaît à 10 minutes. Nous avons donc suivi des réactions d'épissage dans des extraits nucléaires contrôle ou déplétés aux différents temps correspondants à la formation des complexes de l'épissage.

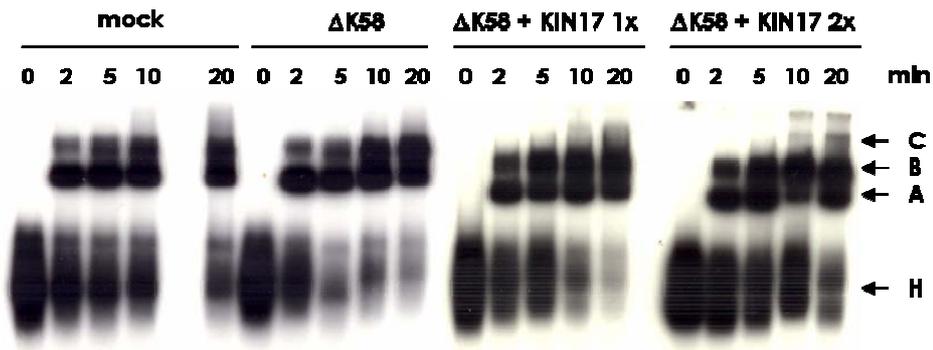


Figure 5-8 : L'assemblage du spliceosome dans des extraits nucléaires contrôle (mock), déplété en KIN17, complémenté avec KIN17 à 2 pmole/ $\mu$ l (KIN17 1x) et complémenté avec KIN17 à 4 pmole/ $\mu$ l (KIN17 2x) est analysé sur gel natif aux temps indiqués. La position des complexes H, A, B et C est indiqué.

L'ajout de protéine recombinante à 4 pmole/ $\mu$ l dans les extraits nucléaires immuno-déplétés en KIN17 permet de voir la formation du complexe C alors que dans la réaction d'épissage avec les extraits nucléaires immuno-déplétés il semble y avoir une accumulation de complexe B. Cependant, l'interprétation de ces résultats est limitée car nous n'obtenons pas le profil classique pour les extraits nucléaires mock. Nous observons plutôt que le complexe A reste assez fort jusqu'à 10-20 minutes. Néanmoins, dans plusieurs essais, nous avons remarqué que nous n'observons jamais la bande correspondant au complexe C pour les réactions d'épissage avec les extraits nucléaires déplétés.

*En conclusion*, ces expériences ont montré que les extraits HeLa immuno-déplétés en protéine KIN17 avaient une efficacité d'épissage diminuée. En revanche, nous ne sommes pas parvenus à obtenir un test de complémentation concluant avec la protéine KIN17 recombinante. Il est toutefois possible que la protéine KIN17 recombinante purifiée ne soit pas active et donc ne permette pas de compléter l'épissage lorsqu'elle est ajoutée aux extraits déplétés.



# **DISCUSSION ET PERSPECTIVES**



<b>1. DISCUSSION ET PERSPECTIVES .....</b>	<b>173</b>
1.1. L'ETUDE STRUCTURALE DE LA PROTEINE hKIN17 .....	173
1.1.1. Etude des différents domaines structuraux de la protéine hKIN17 .....	173
1.2. CE QU'APPORTE LA DETERMINATION DE LA STRUCTURE TRIDIMENSIONNELLE DES DOMAINES DE LA PROTEINE hKIN17 SUR LEUR FONCTION DANS LA PROTEINE.....	174
1.2.1. Le domaine doigt de zinc de type $C_2H_2$ lie les acides nucléiques.....	175
1.2.2. Le domaine winged helix seul ne lie pas les acides nucléiques.....	176
1.2.3. Le domaine winged-helix interagit avec les 60 premiers résidus de hKIN17.....	176
1.2.4. Le domaine winged helix a des partenaires protéiques impliqués dans le métabolisme de l'ARN	178
1.2.5. Le double domaine de type SH3 lie aussi l'ARN .....	179
1.2.6. Le double domaine SH3 reconnaît les histones modifiées.....	180
1.2.7. Le double domaine SH3 est responsable de la présence de hKIN17 au sein d'un complexe acide de haut poids moléculaire.....	181
<b>2. CONCLUSION.....</b>	<b>183</b>

Dans le cadre de la thèse, nous avons avancé dans la description structurale et fonctionnelle de la protéine hKIN17. Qu'apportent ces études sur la protéine hKIN17 et sur son ou ses rôles dans la cellule ? Quelles perspectives pouvons-nous proposer à ces travaux ?



---

## 1. DISCUSSION ET PERSPECTIVES

---

### 1.1. L'étude structurale de la protéine hKIN17

#### 1.1.1. Etude des différents domaines structuraux de la protéine hKIN17

Nous avons tout d'abord montré, par des analyses bioinformatiques, que la protéine hKIN17 est constituée de deux régions globulaires situées en N- et C-terminal et séparées par une grande région flexible. Cette organisation modulaire nous a incités à aborder l'étude structurale de hKIN17 en cherchant à déterminer la structure tridimensionnelle de chacune de ses régions globulaires. Par cristallographie aux rayons X, RMN et/ou SAXS, nous avons montré que la protéine hKIN17 est constituée de trois domaines structuraux : un domaine prédit en doigt de zinc et un domaine *winged helix* dans sa région N-terminale et un double domaine de type SH3 dans sa région C-terminale. Seul le motif en doigt de zinc avait été prédit par bioinformatique. Nous disposons ainsi d'informations à l'échelle atomique sur l'organisation de la protéine.

De plus, les expériences de *pulldown* montrent que la région N-terminale n'interagit pas avec le domaine C-terminal, suggérant qu'en l'absence de partenaires, la région centrale de la protéine hKIN17 est plutôt désordonnée et étendue et que la protéine adopte une conformation allongée (Figure 1-1). Nous avons tenté de confirmer cette proposition en utilisant la diffusion des rayons X aux petits angles (SAXS). Cependant, ces études sont limitées car (i) la production de la protéine entière ne permet pas d'obtenir de grande quantité de protéine pure ; (ii) la protéine entière est très sensible aux phénomènes de protéolyse et de dégradation ; et (iii) les échantillons issus de différentes purifications ne sont pas toujours équivalents en terme de stabilité, rendant difficile la reproductibilité des expériences. Il serait intéressant de poursuivre cette étude en recherchant d'abord à optimiser les conditions physico-chimiques de l'échantillon (tampon, température) et les conditions d'enregistrement des spectres pour limiter les phénomènes d'agrégation et de dégradation de la protéine. La connaissance d'un partenaire (acide nucléique ou protéine) suffisamment affin pour cette protéine pourrait faciliter la détermination de sa structure tridimensionnelle en stabilisant une conformation liée de la protéine au sein d'un complexe.

Nous avons également envisagé d'utiliser la méthode de co-cristallisation avec des anticorps développée par E. Stura au laboratoire (Graille et al. 2001). En effet, nous disposons d'anticorps monoclonaux dirigés contre la protéine (Biard et al. 2002) et des fragments Fab de l'anticorps mAb K36, dirigé contre la région structurée C-terminale de la protéine, ont été préparés par Claude Vidaud du CEA/DSV/SBTN. Cette méthode permet, par l'interaction avec les anticorps, de fixer une configuration de la protéine, facilitant ainsi sa cristallisation. Cependant, nous pensons qu'il serait plus intéressant d'utiliser des anticorps dirigés contre la région centrale flexible qui apparaît la plus gênante pour la cristallisation de la protéine hKIN17.

Dans notre modèle de protéine étendue, les domaines N- et C-terminaux de hKIN17 peuvent lier des partenaires différents de façon concomitante (Figure 1-1, gauche).

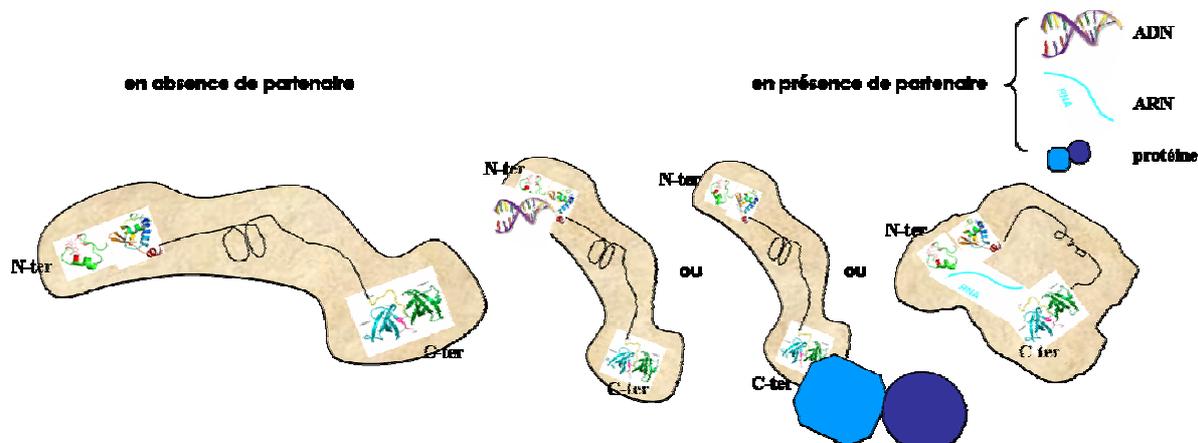


Figure 1-1 : A gauche, modèle de la protéine hKIN17 en l'absence de partenaires. A droite, différentes possibilités de conformations de la protéine hKIN17 en présence de partenaires (ADN, ARN ou protéine).

En revanche, il est possible qu'en présence de certains partenaires, les régions N- et C-terminales se rapprochent (Figure 1-1, droite). En particulier, cela peut être le cas lors de l'interaction de la protéine hKIN17 avec l'ARN puisque le domaine doigt de zinc et le double domaine de type SH3 lient tous les deux l'ARN. L'identification d'une séquence d'ARN interagissant simultanément avec les deux régions globulaires de hKIN17 pourrait permettre une étude par SAXS ou par radio-cristallographie de la protéine ainsi stabilisée. Au contraire avec d'autres partenaires, la région centrale allongée peut permettre aux domaines N- et C-terminaux de lier des partenaires différents de façon concomitante.

## 1.2. Ce qu'apporte la détermination de la structure tridimensionnelle des domaines de la protéine hKIN17 sur leur fonction dans la protéine

L'enjeu est d'établir une relation entre la structure atomique des domaines de la protéine KIN17, les interactions qu'ils peuvent avoir avec des partenaires protéiques ou acides nucléiques et le rôle de ces interactions dans la fonction globale de la protéine KIN17 dans la cellule. A partir de la structure tridimensionnelle des domaines, il s'agit de rechercher quels sont les fonctions, les partenaires et les mécanismes d'action associés à ce type de repliement dans d'autres protéines et de déterminer si les domaines possèdent les propriétés structurales requises pour ces fonctions. Cependant, la recherche de la fonction des domaines à partir de leur structure reste limitée car (i) des repliements différents peuvent converger vers une même fonction, en permettant par exemple aux résidus critiques pour une réaction catalytique de se trouver correctement positionnés quel que soit l'échafaudage qui les porte, (ii) réciproquement, de nombreux exemples d'évolution divergente d'une même architecture élémentaire vers des fonctions très distinctes ont été mis en évidence et (iii) l'étude de la protéine domaine par domaine exclut les fonctions biologiques qui nécessitent une combinaison de plusieurs domaines pour pouvoir avoir lieu.

### 1.2.1. Le domaine doigt de zinc de type $C_2H_2$ lie les acides nucléiques

Les doigts de zinc, des petits domaines très abondants chez les eucaryotes supérieurs, apparaissent être des domaines à fonctions multiples. D'après les résultats obtenus, le domaine doigt de zinc de type  $C_2H_2$  de la protéine hKIN17 est capable de lier l'ADN et l'ARN. Il pourrait donc être impliqué dans l'interaction de la protéine hKIN17 avec l'ADN et l'ARN. C'est aussi le cas de doigts de zinc trouvés dans la protéine TFIIIA, une protéine capable de lier l'ADN et l'ARN, où chaque doigt de zinc de type  $C_2H_2$  établit différents contacts avec chaque type d'acides nucléiques (Bogenhagen 1993; Lu et al. 2003). En général, les doigts de zinc en interaction avec l'ADN reconnaissent des bases spécifiques dans le grand sillon de l'ADN : des contacts ont lieu entre l'hélice  $\alpha$  et le grand sillon de l'ADN par l'intermédiaire de liaisons hydrogène avec les positions -1, 3 et 6 de l'hélice du doigt de zinc. La structure secondaire et tertiaire de l'ARN est plus complexe que celle de l'ADN et comprend des boucles et des éléments en double hélice. Les doigts de zinc en interaction avec l'ARN peuvent reconnaître ces différents éléments de structure en plus d'éléments de séquence.

Il apparaît assez clairement que la protéine hKIN17 a une affinité pour l'ADN. Il a même été précisé que la protéine KIN17 s'associe fortement *in vivo* à des fragments d'ADN contenant des origines de réplication dans des cellules humaines HeLa et dans des cellules de singe CV-1 (Miccoli et al. 2005). C'est pourquoi, nous avons commencé l'étude de l'interaction des domaines de hKIN17 avec l'ADN en utilisant un ADN provenant d'une origine de réplication. Puis, nous avons cherché à préciser quel ADN était reconnu par le domaine doigt de zinc. Théoriquement, un doigt de zinc de type  $C_2H_2$ , composé de deux brins  $\beta$  antiparallèles et d'une hélice  $\alpha$ , ne se lie que sur une séquence d'ADN de trois ou quatre paires de bases consécutives (Brown 2005). Par conséquent, la reconnaissance d'un ADN plus long passe par l'interaction de plusieurs doigts de zinc avec l'ADN. Pour illustration, les quatre structures suivantes de complexes de doigts de zinc avec un ADN ont été résolues : les trois doigts de zinc successifs de la protéine Zif268 en complexe avec un ADN double brin de 11 paires de bases (Pavletich et al. 1991), les deux doigts de zinc successifs de la protéine Tramtrack en complexe avec un ADN double brin de 18 paires de bases (Fairall et al. 1993), les cinq doigts de zinc de la protéine GLI en complexe avec un ADN double brin de 21 paires de bases (Pavletich et al. 1993) et les six doigts de zinc N-terminaux de la protéine TFIIIA en complexe avec un ADN double brin de 31 paires de bases (Nolte et al. 1998). Dans notre cas, les ADN d'une quarantaine de nucléotides étant composés de quelques répétitions d'ADN d'une dizaine de nucléotides, la juxtaposition de plusieurs sites de liaison à hKIN17 au sein du 40-mère est une explication possible à la meilleure affinité apparente du domaine K3 pour cet oligonucléotide par rapport au 10-mère. Par ailleurs, il se peut que le doigt de zinc de la protéine hKIN17 ait une affinité plus élevée pour une séquence particulière d'ADN et que parmi les ADN utilisés cette séquence n'apparaisse pas. Ceci est soutenu par le fait que K3 reconnaît les oligonucléotides de 10 paires de bases 5'-TCTAATCTCT-3' et 5'-AAAATCTTCT-3' mais ne reconnaît pas la séquence 5'-AACTCGAGTT-3' (l'expérience de *southwestern* correspondant à ce dernier oligonucléotide n'est pas montrée dans la thèse). Néanmoins, la séquence cible de K3 est difficile à prédire : il a été observé que les arginines avaient des préférences pour les guanosines, les acides aspartiques pour les adénosines et cytidines et les leucines pour les thymidines mais ces observations sont insuffisantes pour définir un code de reconnaissance des doigts de zinc pour l'ADN (Brown 2005).

### **1.2.2. Le domaine *winged helix* seul ne lie pas les acides nucléiques**

Ce domaine ne présente pas les caractéristiques structurales observées pour les domaines *winged helix* classiques appartenant à des facteurs de transcription et en interaction avec l'ADN. En effet, l'hélice H3 impliquée généralement dans l'interaction des domaines *winged helix* avec l'ADN n'est pas constituée de résidus chargés positivement dans la protéine hKIN17. Des détails supplémentaires sur l'analyse structurale du domaine *winged helix* de la protéine hKIN17 sont donnés dans la thèse de Ludovic Carlier (Carlier 2006). Nous avons confirmé ces analyses en montrant que le domaine *winged helix* de hKIN17 n'est pas capable, seul, de lier les acides nucléiques (paragraphe XXX).

### **1.2.3. Le domaine *winged-helix* interagit avec les 60 premiers résidus de hKIN17**

Nous avons également étudié la région N-terminale entière (région K3 comprenant les domaines doigt de zinc et *winged helix*) puisqu'il est possible que cette combinaison de domaines forme une unité fonctionnelle avec des propriétés qui lui soient propres. Une recherche dans la banque de données SMART montre qu'il existe d'autres protéines (140 protéines détectées) possédant à la fois un doigt de zinc et un domaine *winged-helix* mais dans toutes les séquences de ces protéines, les deux domaines sont éloignés d'environ une centaine de résidus alors que le linker s'étend sur une quinzaine de résidus dans hKIN17. Le domaine *winged helix* de hKIN17 possède une surface hydrophobe très conservée au niveau de la boucle H2-H3 et de l'hélice 3<sub>10</sub> additionnelle. Au sein de la région K3 cette surface du *winged helix* interagit avec les 60 premiers résidus. Elle est donc impliquée dans une interaction intramoléculaire et les résidus interagissant avec cette surface sont probablement très conservés. Il est difficile cependant de distinguer, parmi les 60 premiers résidus incluant le doigt de zinc, les résidus conservés impliqués dans cette interaction intramoléculaire des résidus conservés impliqués dans la liaison à l'ADN et dans le motif C2H2. A l'issue de l'étude structurale par RMN et SAXS du domaine K3, nous proposons un ensemble de modèles structuraux de cette région N-terminale.

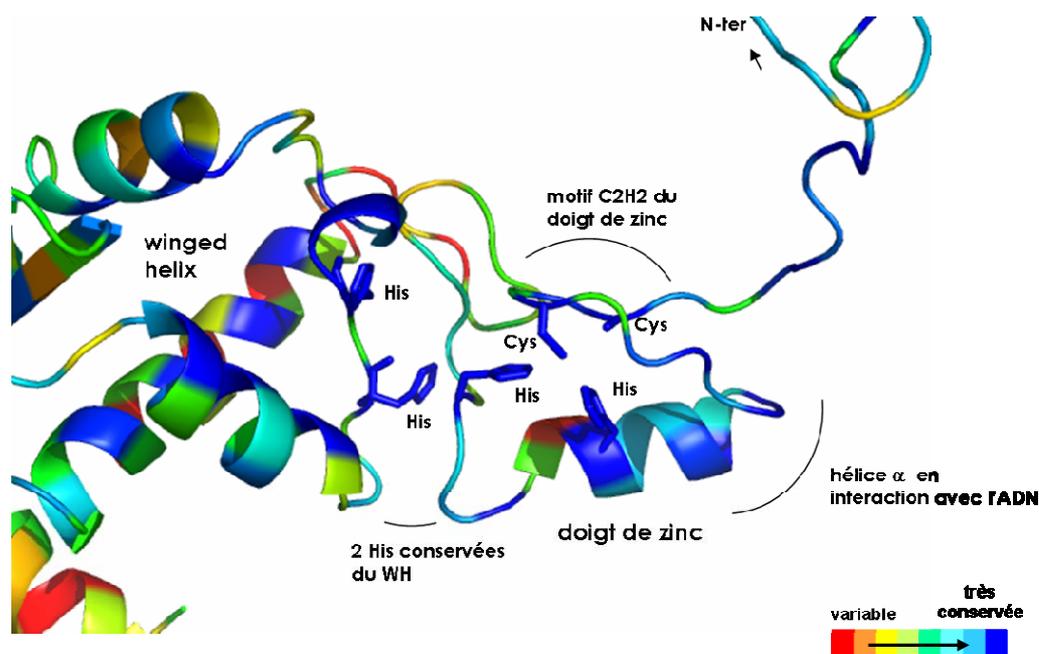


Figure 1-2 : Projection des résultats de CONSURF sur le meilleur modèle de la région N-terminale de hKIN17 obtenu par combinaison des données RMN et SAXS (paragraphe 3.4.2 de Résultats sur l'étude structurale). Les histidines et cystéines du motif doigt de zinc sont mises en évidence, ainsi que les deux histidines du *winged helix* strictement conservées dans toutes les séquences de KIN17.

Dans les modèles proposés (Figure 1-2), l'extrémité N-terminale de l'hélice  $\alpha$  du doigt de zinc est disponible pour l'interaction du domaine avec l'ADN. Il est même possible que la région N-terminale en amont qui est très conservée soit également impliquée dans l'interaction de la protéine avec l'ADN. Par ailleurs, dans ces modèles, le motif des deux histidines et deux cystéines du doigt de zinc est en regard des deux histidines conservées du *winged helix* suggérant que ces résidus forment ensemble un cluster de liaison à un ou deux ions. Par EXAFS, nous pourrions vérifier la présence d'un ou deux ions dans la protéine.

Pour privilégier un modèle de la région N-terminale plutôt qu'un autre, des informations supplémentaires sont nécessaires :

(i) puisque nous faisons l'hypothèse qu'il y a un doigt de zinc dans la région N-terminale, nous devons valider la présence et le repliement correct de ce domaine doigt de zinc dans la protéine ;

(iii) puisque nous faisons l'hypothèse que des résidus du doigt de zinc sont en contact avec le *winged helix*, nous devons donc valider les résidus impliqués dans l'interaction entre les domaines doigt de zinc et le *winged helix*. Pour cela, il faudrait faire une titration des variations de déplacement chimique par RMN du domaine *winged helix* par le domaine doigt de zinc. Nous travaillons aujourd'hui sur une nouvelle construction permettant d'exprimer dans la bactérie le doigt de zinc avec une étiquette Histidine non clivable afin de pouvoir réaliser cette titration. Nous pourrions aussi imaginer faire la titration avec la protéine de fusion GST-doigt de zinc que l'on sait produire (sans parvenir à cliver la protéine de fusion).

### 1.2.4. Le domaine *winged helix* a des partenaires protéiques impliqués dans le métabolisme de l'ARN

Nous avons recherché les partenaires protéiques associés au domaine *winged helix* par la méthode de GST-pulldown avec des extraits cellulaires HeLa. Les résultats indiquent que les protéines partenaires de ce domaine sont essentiellement nucléaires. Parmi les protéines purifiées comme partenaires potentiels du domaine K2, plusieurs protéines intervenant dans le métabolisme de l'ARN ont été identifiées : les protéines p68 (DDX5), DDX21 et eIF-4A, qui sont des hélicases à ARN appartenant à la famille DEAD box. De manière cohérente, G. Pinon-Lataillade du LGR/DRR/CEA Fontenay-aux-Roses a montré que la protéine p68 coimmunoprécipite avec KIN17 dans des cellules HeLa et dans des testicules de souris. La protéine p68 régule l'expression des ARN messagers, en particulier lors de la transcription (Bates et al. 2005) et de l'épissage (Guil et al. 2003). La protéine p68 est aussi essentielle à la maturation des ARN pre-ribosomiaux et l'absence d'expression de p68 conduit à la mort cellulaire (Jalal et al. 2007). DDX21 régule la transcription du gène *c-Jun* et son activité hélicase est impliquée dans cette fonction. De plus, DDX21 est nécessaire à la maturation de l'ARN ribosomal 20S en 18S (Fuller-Pace 2006). Enfin, la protéine eIF4A-I est impliquée dans la liaison de l'ARN messenger au ribosome (Svitkin et al. 2001). De manière cohérente, il a été suggéré un rôle de régulateur de la transcription pour la protéine KIN17. De plus, des interactions entre *rts2*, l'homologue de KIN17 dans la levure comprenant uniquement la région N-terminale, et la protéine ribosomale P0 de la sous-unité 60S ont été observées par Tap-Tag (Gavin et al. 2002).

Par ailleurs, la protéine hKIN17 ainsi que les protéines p68, DDX21 et eIF-4A ont été identifiées dans la première analyse par protéomique du spliceosome humain à grande échelle (Rappsilber et al. 2002). Les protéines hKIN17 et p68 ont également été identifiées dans le complexe B activé, B\* (Makarov et al. 2002; Makarova et al. 2004). Dans une analyse en parallèle de quatre purifications du complexe B (Deckert et al. 2006), la protéine p68 a été détectée et les protéines hKIN17 et eIF-4A n'ont été détectées qu'une seule fois dans ce complexe B. Par ailleurs, des protéines de filaments de type actine et des cyclophilines, qui ont été retenues dans nos expériences de pull down, ont été détectées lors de la purification des complexes de l'épissage assemblés *in vivo* dans des cellules (Chen et al. 2007a).

Avant les résultats de ces différentes analyses par protéomique des complexes de l'épissage, la protéine KIN17 n'avait pas de connexion fonctionnelle décrite avec les mécanismes d'épissage des ARN pré-messagers. Dans ce type d'analyses, il peut être difficile de distinguer parmi les protéines identifiées celles qui sont strictement un composant du *spliceosome* et celles qui sont plutôt des contaminants. En particulier, des protéines peuvent se purifier avec le *spliceosome* simplement parce que ce sont des protéines qui lient l'ARN. Jurica et Moore (Jurica et al. 2003) proposent que des protéines identifiées, comme KIN17, dans au moins deux complexes préparés dans des conditions expérimentales différentes peuvent être impliquées dans des mécanismes d'épissage des ARN pré-messagers. Mais, sans validation fonctionnelle, il est prématuré de classer ces protéines, telle que KIN17, comme des facteurs de l'épissage. Cela nous a encouragé à étudier l'intervention de la protéine KIN17 dans les événements d'épissage des ARN pré-messagers.

Pour cela, dans le laboratoire du Pr. R. Lührmann, nous avons regardé si la protéine hKIN17 s'associe à des complexes de l'épissage et nous avons réalisé un suivi de l'efficacité d'épissage dans des extraits nucléaires déplétés en hKIN17 puis complémentés avec de la protéine KIN17 recombinante. A partir des résultats obtenus, nous pouvons proposer que la

protéine KIN17 n'est probablement pas un composant essentiel du *spliceosome* et qu'elle peut intervenir faiblement et de manière transitoire dans les mécanismes d'épissage des ARN pré-messagers.

Devant les doutes sur le rôle éventuel de la protéine KIN17 dans l'épissage, le Pr. R. Lührmann nous a proposé d'effectuer un autre type d'expérience. Pour cela, des essais ont été effectués dans le laboratoire du Pr. M. Carmo-Fonseca de l'Institut de Médecine Moléculaire de l'Université de Lisbonne (Portugal) pour voir si la protéine KIN17 est recrutée au niveau des sites de transcription du gène beta-globine dans des cellules de souris. La méthode développée par cette équipe (Custodio et al. 2004) a déjà permis de montrer que les protéines du complexe de jonction d'exon (EJC) co-localisent avec des composants du *spliceosome* aux sites de transcription et d'épissage dans le nucléoplasme. L'épissage des ARN pré-messagers du gène beta-globine est nécessaire pour un recrutement efficace de l'EJC et le ciblage ultérieur des mRNPs résultants aux voies d'export nucléaire. Les premiers essais où la présence de la protéine KIN17 était recherchée par immunofluorescence ont bien marché techniquement mais n'ont pas permis de détecter une accumulation de la protéine au site de transcription du gène beta-globine. Cela peut indiquer que la protéine se lie de manière transitoire pour un temps très court aux transcrits naissants. Cependant, nous ne pouvons pas exclure la possibilité que l'épitope de la protéine ne soit pas accessible à l'anticorps lorsque la protéine s'assemble au transcrits naissants. Des essais supplémentaires ont été effectués avec des cellules de souris transfectées avec un vecteur codant pour la protéine KIN17 avec une étiquette GFP, mais les résultats sont les mêmes que précédemment. En conclusion, à l'heure actuelle, nous n'avons pas réussi à montrer de manière non ambiguë que la protéine hKIN17 a un rôle dans les mécanismes d'épissage des ARN pré-messagers.

### **1.2.5. Le double domaine de type SH3 lie aussi l'ARN**

La structure à haute résolution de la région C-terminale de la protéine hKIN17 montre qu'elle est constituée de deux domaines de type SH3. De manière générale, les domaines de type SH3 présentent une grande diversité de fonctions impliquant des interactions avec des acides nucléiques ou avec des partenaires protéiques. Les deux domaines sont reliés par un court linker suggérant qu'ils forment une seule unité fonctionnelle.

La présence d'un motif KOW (Kyrpides et al. 1996) dans cette région a orienté notre travail vers la caractérisation de l'interaction de ce domaine avec l'ARN, comme cela est présenté dans l'article 2. Des motifs KOW ont été observés en particulier dans des protéines du ribosome qui adoptent un repliement de type SH3 et qui sont en interaction avec l'ARN. Il semble que le domaine C-terminal de hKIN17 a réellement une affinité pour l'ARN comparativement aux autres acides nucléiques puisqu'il n'est pas capable de lier l'ADN simple et double brin d'après les résultats des expériences de Southernwestern présentées dans le paragraphe XXX. Nous ne sommes pas parvenus à isoler une séquence d'ARN suffisamment affine pour KIN17 qui nous permette d'envisager des études de complexe, par RMN ou radiocristallographie. Des études de *footprinting* pourraient permettre de réduire la taille de l'ARN nécessaire pour observer une interaction avec le domaine K1 de KIN17. Néanmoins, l'étude de complexes du domaine K1 avec le tungstate nous permet de proposer que l'interaction du domaine C-terminal de la protéine avec l'ARN a lieu par la surface très conservée et chargée positivement à l'interface entre les deux sous-domaines. Comme ce qui

a été observé pour les protéines ribosomales (Ban et al. 2000), l'interaction de ce domaine de hKIN17 avec l'ARN n'implique pas directement le motif KOW.

Il apparaît que les domaines N- et C-terminaux sont tous les deux capables de lier l'ARN. L'étude de différents complexes protéine-ARN (Singh et al. 2005) montre que l'utilisation de plusieurs domaines d'interaction avec l'ARN permet d'augmenter l'affinité et la spécificité entre la protéine et l'ARN cible. Ainsi, des domaines de liaison à l'ARN qui ont une spécificité limitée individuellement peuvent atteindre une grande sélectivité fonctionnelle par coopérativité. Dans la protéine hKIN17, nous pouvons proposer que le double domaine de type SH3 et le domaine doigt de zinc coopèrent pour interagir de façon optimale avec des ARN spécifiques.

### **1.2.6. Le double domaine SH3 reconnaît les histones modifiées**

Dans certaines protéines, des doubles domaines en tonneau  $\beta$  de type SH3 sont impliqués dans la reconnaissance de lysines modifiées d'histones. Ces modifications épigénétiques des histones jouent un rôle dans la régulation de la transcription et dans la réparation des dommages de l'ADN (Corsini et al. 2007). Les structures tridimensionnelles des domaines 53BP1, JMJD2A et CHD1 en complexe avec des peptides d'histones contenant les lysines modifiées ont été déterminées par cristallographie aux rayons X (Flanagan et al. 2005; Botuyan et al. 2006; Chen et al. 2007b). La reconnaissance de peptides méthylés par tous ces domaines est similaire et implique une cage de deux à quatre résidus aromatiques ainsi qu'un résidu acide en face de l'ion méthylammonium de la lysine modifiée. De manière surprenante, il n'y a pas de cage aromatique de ce type dans le domaine C-terminal de hKIN17, suggérant un autre type d'interaction avec les histones. Nous avons ensuite cherché à préciser quel(s) résidu(s) modifié(s) de l'histone H3 est (sont) impliqué(s) dans l'interaction avec le double domaine de type SH3 de hKIN17. Pour cela, nous avons procédé à une identification par spectrométrie de masse MS/MS des résidus modifiés de l'histone H3 retenue par le domaine C-terminal de KIN17 (paragraphe XXX). Cela nous a ensuite amené à étudier plus précisément par RMN l'interaction du domaine K1 avec deux peptides de l'histone H3 comprenant deux types de modifications, pour l'un une diméthylation de la lysine 9 et pour l'autre une acétylation de la lysine 18.

Par RMN, nous n'observons pas d'interaction du domaine K1 avec le peptide d'histone acétylé et une très faible interaction, de l'ordre du millimolaire, avec le peptide d'histone méthylé. Il y a plusieurs possibilités : (i) une interaction très faible *in vitro* a quand même une signification biologique et peut-être renforcée par une interaction concomitante de la protéine hKIN17 avec l'ADN dans le contexte du nucléosome ; ou (ii) cette interaction de l'ordre du millimolaire est non spécifique et la méthode d'identification des résidus modifiés par spectrométrie de masse n'a pas permis d'identifier les modifications d'histones significatives dans cette interaction du domaine C-terminal de la protéine hKIN17 avec l'histone H3. Une méthode plus efficace de recherche des modifications d'histones (essentiellement des méthylations) à l'aide de puce de protéines a été développée par l'équipe de MT Bedford (Kim et al. 2006) et nous avons proposé à cette équipe de rechercher si le domaine C-terminal est capable de retenir un des peptides d'histones testés. Par ailleurs, nous avons conscience qu'il serait intéressant de valider ces résultats avec la protéine entière et par la suite par des études *in vivo*.

### **1.2.7. Le double domaine SH3 est responsable de la présence de hKIN17 au sein d'un complexe acide de haut poids moléculaire**

Par la méthode de GST pull-down d'extraits nucléaires HeLa, nous avons identifié avec succès des partenaires associés au domaine *winged helix* de la protéine hKIN17. En revanche, cette méthode ne nous a pas permis d'identifier des partenaires protéiques associés au double domaine de type SH3. Plusieurs raisons peuvent être invoquées : (i) l'addition d'une étiquette (GST ou ZZ) à la protéine peut changer ses propriétés conduisant à des changements de composition et de stabilité des complexes dans lesquels elle est impliquée et (ii) les interactions transitoires et de faible affinité peuvent être dissociées lors des nombreux lavages. Par contre, les expériences de filtration sur gel et gel ND-IEF d'extraits nucléaires nous ont permis d'isoler un complexe de haut poids moléculaire de pI global acide associé au domaine C-terminal de la protéine hKIN17. Le succès de cette méthode réside sûrement dans le fait qu'elle permet d'isoler des interactions de la protéine hKIN17 ou d'un domaine de la protéine avec un partenaire protéique ou acide nucléique qui peuvent n'avoir lieu que dans le cadre d'un complexe fonctionnel et pas avec la protéine isolée ou en fusion à une étiquette.

Nous avons montré que la protéine hKIN17 endogène était également présente, grâce à son domaine C-terminal, dans ce ou ces complexe(s) de haut poids moléculaire et de pI basique. De plus, sur gel ND-IEF, une fraction de la protéine hKIN17 apparaît dégradée à pH basique alors que la protéine à pH acide est détectée à son poids moléculaire (A.Galat, communication personnelle), suggérant que la présence de la protéine dans un complexe la protège de la protéolyse.

Après avoir isolé le ou les complexe(s) associés au domaine C-terminal et à la protéine hKIN17, nous n'étions pas tant intéressés par l'identification des partenaires en interaction directe avec la protéine que par l'identification de composants connus et présents dans ce ou ces complexe(s). Pour cela, nous avons adopté la stratégie la plus couramment employée pour la caractérisation de complexes de protéines qui consiste en la purification de ces complexes intacts, en quantité suffisante et de pureté appropriée suivie de l'identification des différents composants par spectrométrie de masse. La protéine hKIN17 pouvant appartenir à différents complexes, des protéines de différents complexes peuvent être isolées.

Dans les différentes méthodes, nous avons mesuré l'importance d'avoir différentes étapes de purifications ou d'isolement des complexes afin de réduire les interactions non spécifiques, avec en particulier des protéines très abondantes dans les cellules comme les protéines du cytosquelette, les facteurs de traduction, les protéines ribosomales et les chaperones ((Dziembowski et al. 2004). Ainsi, nous avons conjugué les expériences de filtration sur gel et gel ND-IEF avec une étape d'immunoprécipitation avec des anticorps anti-KIN17. Après immunoprécipitation, nous n'observons qu'un nombre limité de protéines sur gel comparativement au nombre de protéines observées dans les fractions de haut poids moléculaire et de pI acide.

Jusqu'à maintenant, nous n'avons pas obtenu une identification exhaustive des protéines retenues après l'immunoprécipitation. Lors de la préparation des échantillons pour la spectrométrie de masse, un certain nombre de bandes de protéine visibles ont été sélectionnées. Le choix des protéines à identifier est donc une étape où de l'information est perdue. Une autre approche, utilisée dans différents laboratoires, consiste à identifier toutes les protéines d'un puits en découpant la bande du gel correspondant au puits en petits morceaux, indépendamment des protéines visibles ou non. De plus, certaines identifications

par spectrométrie de masse n'ont pas donné de résultats fiables du fait d'un signal trop faible dû à des quantités trop faibles de peptides issus de la digestion à la trypsine. Ainsi, pour le moment, nous n'avons identifié que la protéine hnRNP A2/B1 comme composant des complexes. Les hnRNP (pour *heterogeneous nuclear ribonucleoprotein*) s'associent avec les ARN pré-messagers naissants et sont impliqués dans le métabolisme des ARN ; en particulier dans la régulation de l'épissage. Il est évident que la protéine hnRNP A2/B1 ne représente pas l'ensemble des protéines présentes dans le ou les complexes avec hKIN17. De nouvelles identifications de protéines isolées par immunoprécipitation sont en cours afin d'obtenir une description plus complète des complexes associés à la protéine hKIN17.

---

## 2. CONCLUSION

---

La protéine KIN17 apparaît comme une protéine multifonctionnelle. Elle est présente dans une grande diversité d'organismes eucaryotes (animaux, plantes, champignons et protistes) et donc des utilisations supplémentaires de la protéine ont pu apparaître au cours de l'évolution. On peut imaginer que (i) la protéine avait déjà initialement plusieurs fonctions cellulaires ou (ii) la protéine avait une fonction initiale puis des interactions supplémentaires et d'affinités plus faibles sont apparues et ont été le support de nouvelles fonctions. Nous privilégions cette seconde hypothèse d'après la conservation de séquence de la protéine KIN17 : la région N-terminale existe dans tous les organismes alors que la région C-terminale de la protéine n'est présente que chez les Eucaryotes supérieurs.

La fonction exercée par la protéine KIN17 peut aussi varier selon

(i) sa localisation dans la cellule. La protéine KIN17 est décrite comme une protéine nucléaire mais elle a été également détectée dans le cytoplasme et dans la fraction de membranes légères constituée des particules du type mitochondries, ribosomes, Golgi (A.Galat, communication personnelle). C'est pourquoi, nous prévoyons de rechercher des partenaires associés à la protéine dans les différents compartiments cellulaires (nucléaire, cytosolique, membranes légères) ;

(ii) des modifications post-traductionnelles de la protéine, en lien avec différentes activations (suite à un stress UV par exemple). Le point isoélectrique global de la protéine KIN17 varie selon sa localisation dans la cellule (A.Galat, communication personnelle). Des essais sont en cours pour voir si la protéine existe sous forme phosphorylée ;  
et enfin (iii) son état d'oligomérisation.

Ces quelques mécanismes de régulation de la fonction d'une protéine ne sont pas mutuellement exclusifs et la protéine peut les utiliser en combinaison pour effectuer l'une ou l'autre de ses fonctions.

La protéine KIN17 possède des partenaires différents selon le domaine étudié : des acides nucléiques, des composants de la chromatine (histones), des protéines associées au métabolisme de l'ARN. Il apparaît aujourd'hui que la transcription et la traduction, les deux mécanismes majeurs conduisant à l'expression des gènes, communiquent directement entre elles en partageant des protéines. De même, les mécanismes de contrôle de l'expression des gènes sont en interaction avec les mécanismes de signalisation et de réparation des dommages de l'ADN. Tous ces mécanismes sont régulés via des interactions protéines-protéines et/ou des interactions protéines-ADN/ARN caractérisées par une grande labilité (interactions de faibles affinités), impliquant souvent des partenaires multiples, agissant en synergie ou en compétition.

A partir des résultats obtenus au cours de ce travail de thèse, il paraît intéressant de développer l'étude du rôle de la protéine hKIN17 dans le métabolisme de l'ARN. Néanmoins, nous n'avons pas réussi jusqu'à maintenant à montrer qu'elle était un facteur essentiel de l'épissage des ARN pré-messagers. L'analyse en cours des partenaires associés à la protéine dans les différents compartiments cellulaires devraient suggérer un rôle pour hKIN17 dans d'autres mécanismes associés à l'expression des gènes (transcription des gènes, maturation et export nucléaire des ARN messagers, biogénèse du ribosome, régulation de la traduction, surveillance des ARN messagers ou silencing médié par des petits ARN), que nous pourrions alors valider *in vivo*.



# **BIBLIOGRAPHIE**



- Andrews, N.C., and Faller, D.V. (1991). A rapid micropreparation technique for extraction of DNA-binding proteins from limiting numbers of mammalian cells. *Nucleic Acids Res* 19(9): 2499.
- Angulo, J.F., Rouer, E., Benarous, R., and Devoret, R. (1991a). Identification of a mouse cDNA fragment whose expressed polypeptide reacts with anti-recA antibodies. *Biochimie* 73(2-3): 251-256.
- Angulo, J.F., Rouer, E., Mazin, A., Mattei, M.G., Tissier, A., Horellou, P., Benarous, R., and Devoret, R. (1991b). Identification and expression of the cDNA of KIN17, a zinc-finger gene located on mouse chromosome 2, encoding a new DNA-binding protein. *Nucleic Acids Res* 19(19): 5117-5123.
- Bakkenist, C.J., and Kastan, M.B. (2004). Initiating cellular stress responses. *Cell* 118(1): 9-17.
- Ban, N., Nissen, P., Hansen, J., Moore, P.B., and Steitz, T.A. (2000). The complete atomic structure of the large ribosomal subunit at 2.4 Å resolution. *Science* 289(5481): 905-920.
- Bateman, A., Coin, L., Durbin, R., Finn, R.D., Hollich, V., Griffiths-Jones, S., Khanna, A., Marshall, M., Moxon, S., Sonnhammer, E.L., et al. (2004). The Pfam protein families database. *Nucleic Acids Res* 32(Database issue): D138-141.
- Bates, G.J., Nicol, S.M., Wilson, B.J., Jacobs, A.M., Bourdon, J.C., Wardrop, J., Gregory, D.J., Lane, D.P., Perkins, N.D., and Fuller-Pace, F.V. (2005). The DEAD box protein p68: a novel transcriptional coactivator of the p53 tumour suppressor. *Embo J* 24(3): 543-553.
- Bernado, P., Mylonas, E., Petoukhov, M.V., Blackledge, M., and Svergun, D.I. (2007). Structural characterization of flexible proteins using small-angle X-ray scattering. *J Am Chem Soc* 129(17): 5656-5664.
- Bernocco, S., Steiglitz, B.M., Svergun, D.I., Petoukhov, M.V., Ruggiero, F., Ricard-Blum, S., Ebel, C., Geourjon, C., Deleage, G., Font, B., et al. (2003). Low resolution structure determination shows procollagen C-proteinase enhancer to be an elongated multidomain glycoprotein. *J Biol Chem* 278(9): 7199-7205.
- Biard, D.S., Kannouche, P., Lannuzel-Drogou, C., Mauffrey, P., Apiou, F., and Angulo, J.F. (1999). Ectopic expression of (Mm)Kin17 protein inhibits cell proliferation of human tumor-derived cells. *Exp Cell Res* 250(2): 499-509.
- Biard, D.S., Miccoli, L., Despras, E., Frobert, Y., Creminon, C., and Angulo, J.F. (2002). Ionizing radiation triggers chromatin-bound kin17 complex formation in human cells. *J Biol Chem* 277(21): 19156-19165.
- Biard, D.S., Miccoli, L., Despras, E., Harper, F., Pichard, E., Creminon, C., and Angulo, J.F. (2003). Participation of kin17 protein in replication factories and in other DNA transactions mediated by high molecular weight nuclear complexes. *Mol Cancer Res* 1(7): 519-531.
- Biard, D.S., Saintigny, Y., Maratrat, M., Paris, F., Martin, M., and Angulo, J.F. (1997). Enhanced expression of the Kin17 protein immediately after low doses of ionizing radiation. *Radiat Res* 147(4): 442-450.
- Blattner, C., Kannouche, P., Litfin, M., Bender, K., Rahmsdorf, H.J., Angulo, J.F., and Herrlich, P. (2000). UV-Induced stabilization of c-fos and other short-lived mRNAs. *Mol Cell Biol* 20(10): 3616-3625.
- Bogenhagen, D.F. (1993). Proteolytic footprinting of transcription factor TFIIIA reveals different tightly binding sites for 5S RNA and 5S DNA. *Mol Cell Biol* 13(9): 5149-5158.

- Botuyan, M.V., Lee, J., Ward, I.M., Kim, J.E., Thompson, J.R., Chen, J., and Mer, G. (2006). Structural basis for the methylation state-specific recognition of histone H4-K20 by 53BP1 and Crb2 in DNA repair. *Cell* 127(7): 1361-1373.
- Braud, S., Moutiez, M., Belin, P., Abello, N., Drevet, P., Zinn-Justin, S., Courcon, M., Masson, C., Dassa, J., Charbonnier, J.B., et al. (2005). Dual expression system suitable for high-throughput fluorescence-based screening and production of soluble proteins. *J Proteome Res* 4(6): 2137-2147.
- Bricogne, G., Vornrhein, C., Flensburg, C., Schiltz, M., and Paciorek, W. (2003). Generation, representation and flow of phase information in structure determination: recent developments in and around SHARP 2.0. *Acta Crystallogr D Biol Crystallogr* 59(Pt 11): 2023-2030.
- Brooks, B.R., Brucoleri, R.E., Olafson, B.D., States, D.J., Swaminathan, S., and Karplus, M. (1983). CHARMM: A Program for Macromolecular Energy, Minimization, and Dynamics Calculations. *J. Comp. Chem.* 4: 187-217.
- Brown, R.S. (2005). Zinc finger proteins: getting a grip on RNA. *Curr Opin Struct Biol* 15(1): 94-98.
- Carlier, L. (2006). Caractérisation de la région 51-160 de la protéine KIN17 humaine par RMN et Modélisation Moléculaire. In *Biologie Structurale et Fonctionnelle*. Université de Rouen, Rouen.
- CCP4. (1994). Collaborative Computational Project Number 4. *Acta Crystallogr* 50: 760-763.
- Charier, G., Couprie, J., Alpha-Bazin, B., Meyer, V., Quemeneur, E., Guerois, R., Callebaut, I., Gilquin, B., and Zinn-Justin, S. (2004). The Tudor tandem of 53BP1: a new structural motif involved in DNA and RG-rich peptide binding. *Structure* 12(9): 1551-1562.
- Chen, Y.I., Moore, R.E., Ge, H.Y., Young, M.K., Lee, T.D., and Stevens, S.W. (2007a). Proteomic analysis of in vivo-assembled pre-mRNA splicing complexes expands the catalog of participating factors. *Nucleic Acids Res* 35(12): 3928-3944.
- Chen, Z., Zang, J., Kappler, J., Hong, X., Crawford, F., Wang, Q., Lan, F., Jiang, C., Whetstine, J., Dai, S., et al. (2007b). Structural basis of the recognition of a methylated histone tail by JMJD2A. *Proc Natl Acad Sci U S A* 104(26): 10818-10823.
- Cornilescu, G., Delaglio, F., and Bax, A. (1999). Protein backbone angle restraints from searching a database for chemical shift and sequence homology. *J Biomol NMR* 13(3): 289-302.
- Corsini, L., and Sattler, M. (2007). Tudor hooks up with DNA repair. *Nat Struct Mol Biol* 14(2): 98-99.
- Cox, M.M. (1999). Recombinational DNA repair in bacteria and the RecA protein. *Prog Nucleic Acid Res Mol Biol* 63: 311-366.
- Custodio, N., Carvalho, C., Condado, I., Antoniou, M., Blencowe, B.J., and Carmo-Fonseca, M. (2004). In vivo recruitment of exon junction complex proteins to transcription sites in mammalian cell nuclei. *Rna* 10(4): 622-633.
- Dauter, Z., Dauter, M., de La Fortelle, E., Bricogne, G., and Sheldrick, G.M. (1999). Can anomalous signal of sulfur become a tool for solving protein crystal structures? *J Mol Biol* 289(1): 83-92.
- Davierwala, A.P., Haynes, J., Li, Z., Brost, R.L., Robinson, M.D., Yu, L., Mnaimneh, S., Ding, H., Zhu, H., Chen, Y., et al. (2005). The synthetic genetic interaction spectrum of essential genes. *Nat Genet* 37(10): 1147-1152.
- de la Cruz, X., Lois, S., Sanchez-Molina, S., and Martinez-Balbas, M.A. (2005). Do protein motifs read the histone code? *Bioessays* 27(2): 164-175.
- Deckert, J., Hartmuth, K., Boehringer, D., Behzadnia, N., Will, C.L., Kastner, B., Stark, H., Urlaub, H., and Luhrmann, R. (2006). Protein composition and electron microscopy

- structure of affinity-purified human spliceosomal B complexes isolated under physiological conditions. *Mol Cell Biol* 26(14): 5528-5543.
- Delaglio, F., Grzesiek, S., Vuister, G.W., Zhu, G., Pfeifer, J., and Bax, A. (1995). NMRPipe: a multidimensional spectral processing system based on UNIX pipes. *J Biomol NMR* 6(3): 277-293.
- Despras, E., Pfeiffer, P., Salles, B., Calsou, P., Kuhfittig-Kulle, S., Angulo, J.F., and Biard, D.S. (2007). Long-term XPC silencing reduces DNA double-strand break repair. *Cancer Res* 67(6): 2526-2534.
- Dignam, J.D., Lebovitz, R.M., and Roeder, R.G. (1983). Accurate transcription initiation by RNA polymerase II in a soluble extract from isolated mammalian nuclei. *Nucleic Acids Res* 11(5): 1475-1489.
- Dziembowski, A., and Seraphin, B. (2004). Recent developments in the analysis of protein complexes. *FEBS Lett* 556(1-3): 1-6.
- Edwards, P.A., and Shooter, K.V. (1970). A study of calf thymus histone fraction F2(b) by gel filtration. *Biochem J* 120(1): 61-66.
- Eickbush, T.H., and Moudrianakis, E.N. (1978). The histone core complex: an octamer assembled by two sets of protein-protein interactions. *Biochemistry* 17(23): 4955-4964.
- Emsley, P., and Cowtan, K. (2004). Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* 60(Pt 12 Pt 1): 2126-2132.
- Evans, G., and Pettifer, R.F. (2001). CHOOCH: a program for deriving anomalous-scattering factors from X-ray fluorescence spectra. *J. Appl. Cryst.* 34: 82-86.
- Fairall, L., Schwabe, J.W., Chapman, L., Finch, J.T., and Rhodes, D. (1993). The crystal structure of a two zinc-finger peptide reveals an extension to the rules for zinc-finger/DNA recognition. *Nature* 366(6454): 483-487.
- Flanagan, J.F., Mi, L.Z., Chruszcz, M., Cymborowski, M., Clines, K.L., Kim, Y., Minor, W., Rastinejad, F., and Khorasanizadeh, S. (2005). Double chromodomains cooperate to recognize the methylated histone H3 tail. *Nature* 438(7071): 1181-1185.
- Fuller-Pace, F.V. (2006). DExD/H box RNA helicases: multifunctional proteins with important roles in transcriptional regulation. *Nucleic Acids Res* 34(15): 4206-4215.
- Gabel, F., Simon, B., and Sattler, M. (2006). A target function for quaternary structural refinement from small angle scattering and NMR orientational restraints. *Eur Biophys J* 35(4): 313-327.
- Gaboriaud, C., Bissery, V., Benchetrit, T., and Mornon, J.P. (1987). Hydrophobic cluster analysis: an efficient new way to compare and analyse amino acid sequences. *FEBS Lett* 224(1): 149-155.
- Galat, A., and Thai, R. (2004). Long-size isoelectric focusing (IEF) in flexible silicone tubes: an application to semi-preparative fractionation of proteins. *Int J Biol Macromol* 34(3): 163-171.
- Gavin, A.C., Bosche, M., Krause, R., Grandi, P., Marzioch, M., Bauer, A., Schultz, J., Rick, J.M., Michon, A.M., Cruciat, C.M., et al. (2002). Functional organization of the yeast proteome by systematic analysis of protein complexes. *Nature* 415(6868): 141-147.
- Giaever, G., Chu, A.M., Ni, L., Connelly, C., Riles, L., Veronneau, S., Dow, S., Lucau-Danila, A., Anderson, K., Andre, B., et al. (2002). Functional profiling of the *Saccharomyces cerevisiae* genome. *Nature* 418(6896): 387-391.
- Glatter, O. (1977). A new method for the evaluation of small-angle scattering data. *J. Appl. Cryst.* 10(5): 415-421.
- Gorg, A., Weiss, W., and Dunn, M.J. (2004). Current two-dimensional electrophoresis technology for proteomics. *Proteomics* 4(12): 3665-3685.

- Graille, M., Stura, E.A., Housden, N.G., Beckingham, J.A., Bottomley, S.P., Beale, D., Taussig, M.J., Sutton, B.J., Gore, M.G., and Charbonnier, J.B. (2001). Complex between *Peptostreptococcus magnus* protein L and a human antibody reveals structural convergence in the interaction modes of Fab binding proteins. *Structure* 9(8): 679-687.
- Green, D., Ingram, V., and Perutz, M. (1954). The structure of haemoglobin IV. Sign of determination by isomorphous replacement method. *Proc. R. Soc. London SER. A* 225: 287-307.
- Grishaev, A., Wu, J., Trewthella, J., and Bax, A. (2005). Refinement of multidomain protein structures by combination of solution small-angle X-ray scattering and NMR data. *J Am Chem Soc* 127(47): 16621-16628.
- Grosse-Kunstleve, R.W., and Adams, P.D. (2003). Substructure search procedures for macromolecular structures. *Acta Crystallogr D Biol Crystallogr* 59(Pt 11): 1966-1973.
- Guil, S., Gattoni, R., Carrascal, M., Abian, J., Stevenin, J., and Bach-Elias, M. (2003). Roles of hnRNP A1, SR proteins, and p68 helicase in c-H-ras alternative splicing regulation. *Mol Cell Biol* 23(8): 2927-2941.
- Guinier, A. (1939). *Ann. Phys.* 12: 161-237.
- Hartmuth, K., Urlaub, H., Vornlocher, H.P., Will, C.L., Gentzel, M., Wilm, M., and Lührmann, R. (2002). Protein composition of human prespliceosomes isolated by a tobramycin affinity-selection method. *Proc Natl Acad Sci U S A* 99(26): 16719-16724.
- Hazbun, T.R., Malmstrom, L., Anderson, S., Graczyk, B.J., Fox, B., Riffle, M., Sundin, B.A., Aranda, J.D., McDonald, W.H., Chiu, C.H., et al. (2003). Assigning function to yeast proteins by integration of technologies. *Mol Cell* 12(6): 1353-1365.
- Hendrickson, W.A. (1991). Determination of macromolecular structures from anomalous diffraction of synchrotron radiation. *Science* 254(5028): 51-58.
- Hittelman, W.N., and Pollard, M. (1984). Visualization of chromatin events associated with repair of ultraviolet light-induced damage by premature chromosome condensation. *Carcinogenesis* 5(10): 1277-1285.
- Hoeijmakers, J.H. (2001). Genome maintenance mechanisms for preventing cancer. *Nature* 411(6835): 366-374.
- Huang, J., Sengupta, R., Espejo, A.B., Lee, M.G., Dorsey, J.A., Richter, M., Opravil, S., Shiekhhattar, R., Bedford, M.T., Jenuwein, T., et al. (2007). p53 is regulated by the lysine demethylase LSD1. *Nature* 449(7158): 105-108.
- Huang, Y., Fang, J., Bedford, M.T., Zhang, Y., and Xu, R.M. (2006). Recognition of histone H3 lysine-4 methylation by the double tudor domain of JMJD2A. *Science* 312(5774): 748-751.
- Jaberaboansari, A., Nelson, G.B., Roti Roti, J.L., and Wheeler, K.T. (1988). Postirradiation alterations of neuronal chromatin structure. *Radiat Res* 114(1): 94-104.
- Jalal, C., Uhlmann-Schiffler, H., and Stahl, H. (2007). Redundant role of DEAD box proteins p68 (Ddx5) and p72/p82 (Ddx17) in ribosome biogenesis and cell proliferation. *Nucleic Acids Res* 35(11): 3590-3601.
- Jurica, M.S., Licklider, L.J., Gygi, S.R., Grigorieff, N., and Moore, M.J. (2002). Purification and characterization of native spliceosomes suitable for three-dimensional structural analysis. *Rna* 8(4): 426-439.
- Jurica, M.S., and Moore, M.J. (2003). Pre-mRNA splicing: awash in a sea of proteins. *Mol Cell* 12(1): 5-14.
- Kabsch, W.J. (1993). Automatic processing of rotation diffraction data from crystals of initially unknown symmetry and cell constants. *J. Appl. Cryst.* 26: 795-800.

- Kannouche, P., and Angulo, J.F. (1999). Overexpression of kin17 protein disrupts nuclear morphology and inhibits the growth of mammalian cells. *J Cell Sci* 112 ( Pt 19): 3215-3224.
- Kannouche, P., Mauffrey, P., Pinon-Lataillade, G., Mattei, M.G., Sarasin, A., Daya-Grosjean, L., and Angulo, J.F. (2000). Molecular cloning and characterization of the human KIN17 cDNA encoding a component of the UVC response that is conserved among metazoans. *Carcinogenesis* 21(9): 1701-1710.
- Kannouche, P., Pinon-Lataillade, G., Tissier, A., Chevalier-Lagente, O., Sarasin, A., Mezzina, M., and Angulo, J.F. (1998). The nuclear concentration of kin17, a mouse protein that binds to curved DNA, increases during cell proliferation and after UV irradiation. *Carcinogenesis* 19(5): 781-789.
- Kaplan, Y., and Kupiec, M. (2007). A role for the yeast cell cycle/splicing factor Cdc40 in the G1/S transition. *Curr Genet* 51(2): 123-140.
- Kim, J., Daniel, J., Espejo, A., Lake, A., Krishna, M., Xia, L., Zhang, Y., and Bedford, M.T. (2006). Tudor, MBT and chromo domains gauge the degree of lysine methylation. *EMBO Rep* 7(4): 397-403.
- Koch, M.H., Vachette, P., and Svergun, D.I. (2003). Small-angle scattering: a view on the properties, structures and structural changes of biological macromolecules in solution. *Q Rev Biophys* 36(2): 147-227.
- Konarev, P.V., Volkov, V.V., Sokolova, A.V., Koch, M.H.J., and Svergun, D.I. (2003). PRIMUS - a Windows-PC based system for small-angle scattering data analysis. *J. Appl. Cryst.* 36: 1277-1282.
- Kozin, M.B., and Svergun, D.I. (2000). Automated matching of high- and low-resolution structural models. *J. Appl. Cryst.* 34: 33-41.
- Kurumizaka, H., Aihara, H., Ikawa, S., Kashima, T., Bazemore, L.R., Kawasaki, K., Sarai, A., Radding, C.M., and Shibata, T. (1996). A possible role of the C-terminal domain of the RecA protein. A gateway model for double-stranded DNA binding. *J Biol Chem* 271(52): 33515-33524.
- Kyrpides, N.C., Woese, C.R., and Ouzounis, C.A. (1996). KOW: a novel motif linking a bacterial transcription factor with ribosomal proteins. *Trends Biochem Sci* 21(11): 425-426.
- Leslie, A.G.W. (1992). Recent changes to the MOSFLM package for processing film and image plate data. *Joint CCP4 + ESF-EAMCB Newsletter on Protein Crystallography* 26.
- Ljungman, M. (1989). Pretreatment with UV light renders the chromatin in human fibroblasts more susceptible to the DNA-damaging agents bleomycin, gamma radiation and 8-methoxypsoralen. *Carcinogenesis* 10(3): 447-451.
- Lu, D., Searles, M.A., and Klug, A. (2003). Crystal structure of a zinc-finger-RNA complex reveals two modes of molecular recognition. *Nature* 426(6962): 96-100.
- Luger, K., Mader, A.W., Richmond, R.K., Sargent, D.F., and Richmond, T.J. (1997). Crystal structure of the nucleosome core particle at 2.8 Å resolution. *Nature* 389(6648): 251-260.
- Makarov, E.M., Makarova, O.V., Urlaub, H., Gentzel, M., Will, C.L., Wilm, M., and Luhrmann, R. (2002). Small nuclear ribonucleoprotein remodeling during catalytic activation of the spliceosome. *Science* 298(5601): 2205-2208.
- Makarova, O.V., Makarov, E.M., Urlaub, H., Will, C.L., Gentzel, M., Wilm, M., and Luhrmann, R. (2004). A subset of human 35S U5 proteins, including Prp19, function prior to catalytic step 1 of splicing. *Embo J* 23(12): 2381-2391.
- Markillie, L.M., Lin, C.T., Adkins, J.N., Auberry, D.L., Hill, E.A., Hooker, B.S., Moore, P.A., Moore, R.J., Shi, L., Wiley, H.S., et al. (2005). Simple protein complex

- purification and identification method for high-throughput mapping of protein interaction networks. *J Proteome Res* 4(2): 268-274.
- Marquez, J.A., Smith, C.I., Petoukhov, M.V., Lo Surdo, P., Mattsson, P.T., Knekt, M., Westlund, A., Scheffzek, K., Saraste, M., and Svergun, D.I. (2003). Conformation of full-length Bruton tyrosine kinase (Btk) from synchrotron X-ray solution scattering. *Embo J* 22(18): 4616-4624.
- Masson, C., Mena, F., Pinon-Lataillade, G., Frobert, Y., Chevillard, S., Radicella, J.P., Sarasin, A., and Angulo, J.F. (2003). Global genome repair is required to activate KIN17, a UVC-responsive gene involved in DNA replication. *Proc Natl Acad Sci U S A* 100(2): 616-621.
- Masson, C., Mena, F., Pinon-Lataillade, G., Frobert, Y., Radicella, J.P., and Angulo, J.F. (2001). Identification of KIN (KIN17), a human gene encoding a nuclear DNA-binding protein, as a novel component of the TP53-independent response to ionizing radiation. *Radiat Res* 156(5 Pt 1): 535-544.
- Mattinen, M.L., Paakkonen, K., Ikonen, T., Craven, J., Drakenberg, T., Serimaa, R., Waltho, J., and Annala, A. (2002). Quaternary structure built from subunits combining NMR and small-angle x-ray scattering data. *Biophys J* 83(2): 1177-1183.
- Maurer-Stroh, S., Dickens, N.J., Hughes-Davies, L., Kouzarides, T., Eisenhaber, F., and Ponting, C.P. (2003). The Tudor domain 'Royal Family': Tudor, plant Agenet, Chromo, PWWP and MBT domains. *Trends Biochem Sci* 28(2): 69-74.
- Mazin, A., Milot, E., Devoret, R., and Chartrand, P. (1994a). KIN17, a mouse nuclear protein, binds to bent DNA fragments that are found at illegitimate recombination junctions in mammalian cells. *Mol Gen Genet* 244(4): 435-438.
- Mazin, A., Timchenko, T., Menissier-de Murcia, J., Schreiber, V., Angulo, J.F., de Murcia, G., and Devoret, R. (1994b). Kin17, a mouse nuclear zinc finger protein that binds preferentially to curved DNA. *Nucleic Acids Res* 22(20): 4335-4341.
- McRee, D.E. (1999). XtalView/Xfit--A versatile program for manipulating atomic coordinates and electron density. *J Struct Biol* 125(2-3): 156-165.
- Melo, J.A., Cohen, J., and Toczyski, D.P. (2001). Two checkpoint complexes are independently recruited to sites of DNA damage in vivo. *Genes Dev* 15(21): 2809-2821.
- Miccoli, L., Biard, D.S., Creminon, C., and Angulo, J.F. (2002). Human kin17 protein directly interacts with the simian virus 40 large T antigen and inhibits DNA replication. *Cancer Res* 62(19): 5425-5435.
- Miccoli, L., Biard, D.S., Frouin, I., Harper, F., Maga, G., and Angulo, J.F. (2003). Selective interactions of human kin17 and RPA proteins with chromatin and the nuclear matrix in a DNA damage- and cell cycle-regulated manner. *Nucleic Acids Res* 31(14): 4162-4175.
- Miccoli, L., Frouin, I., Novac, O., Di Paola, D., Harper, F., Zannis-Hadjopoulos, M., Maga, G., Biard, D.S., and Angulo, J.F. (2005). The human stress-activated protein kin17 belongs to the multiprotein DNA replication complex and associates in vivo with mammalian replication origins. *Mol Cell Biol* 25(9): 3814-3830.
- Morris, G., and Lewis, P.N. (1977). Conformational studies on histone H3 and its CNBr peptides. *Eur J Biochem* 77(3): 471-477.
- Motoyama, N., and Naka, K. (2004). DNA damage tumor suppressor genes and genomic instability. *Curr Opin Genet Dev* 14(1): 11-16.
- Munakata, T., Adachi, N., Yokoyama, N., Kuzuhara, T., and Horikoshi, M. (2000). A human homologue of yeast anti-silencing factor has histone chaperone activity. *Genes Cells* 5(3): 221-233.

- Neubauer, G., King, A., Rappsilber, J., Calvio, C., Watson, M., Ajuh, P., Sleeman, J., Lamond, A., and Mann, M. (1998). Mass spectrometry and EST-database searching allows characterization of the multi-protein spliceosome complex. *Nat Genet* 20(1): 46-50.
- Newport, J., and Yan, H. (1996). Organization of DNA into foci during replication. *Curr Opin Cell Biol* 8(3): 365-368.
- Nolte, R.T., Conlin, R.M., Harrison, S.C., and Brown, R.S. (1998). Differing roles for zinc fingers in DNA recognition: structure of a six-finger transcription factor IIIA complex. *Proc Natl Acad Sci U S A* 95(6): 2938-2943.
- Nyberg, K.A., Michelson, R.J., Putnam, C.W., and Weinert, T.A. (2002). Toward maintaining the genome: DNA damage and replication checkpoints. *Annu Rev Genet* 36: 617-656.
- Omichinski, J.G., Clore, G.M., Appella, E., Sakaguchi, K., and Gronenborn, A.M. (1990). High-resolution three-dimensional structure of a single zinc finger from a human enhancer binding protein in solution. *Biochemistry* 29(40): 9324-9334.
- Otwinowski, Z., and Mjnor, W. (1997). Processing of X-ray Diffraction Data Collected in Oscillation Mode. In *Methods in Enzymology*. (ed. A.P. Eds.), pp. 307-326. C.W. Carter, Jr. & R. M. Sweet, (New York).
- Pavletich, N.P., and Pabo, C.O. (1991). Zinc finger-DNA recognition: crystal structure of a Zif268-DNA complex at 2.1 Å. *Science* 252(5007): 809-817.
- Pavletich, N.P., and Pabo, C.O. (1993). Crystal structure of a five-finger GLI-DNA complex: new perspectives on zinc fingers. *Science* 261(5129): 1701-1707.
- Peterson, C.L., and Cote, J. (2004). Cellular machineries for chromosomal DNA repair. *Genes Dev* 18(6): 602-616.
- Petoukhov, M.V., and Svergun, D.I. (2005). Global rigid body modeling of macromolecular complexes against small-angle scattering data. *Biophys J* 89(2): 1237-1250.
- Petoukhov, M.V., and Svergun, D.I. (2007). Analysis of X-ray and neutron scattering from biomacromolecular solutions. *Curr Opin Struct Biol* 17(5): 562-571.
- Pinon-Lataillade, G., Masson, C., Bernardino-Sgherri, J., Henriot, V., Mauffrey, P., Frobert, Y., Araneda, S., and Angulo, J.F. (2004). KIN17 encodes an RNA-binding protein and is expressed during mouse spermatogenesis. *J Cell Sci* 117(Pt 16): 3691-3702.
- Puig, O., Caspary, F., Rigaut, G., Rutz, B., Bouveret, E., Bragado-Nilsson, E., Wilm, M., and Seraphin, B. (2001). The tandem affinity purification (TAP) method: a general procedure of protein complex purification. *Methods* 24(3): 218-229.
- Raibaud, S., Vachette, P., Guillier, M., Allemant, F., Chiaruttini, C., and Dardel, F. (2003). How bacterial ribosomal protein L20 assembles with 23 S ribosomal RNA and its own messenger RNA. *J Biol Chem* 278(38): 36522-36530.
- Rappsilber, J., Ryder, U., Lamond, A.I., and Mann, M. (2002). Large-scale proteomic analysis of the human spliceosome. *Genome Res* 12(8): 1231-1245.
- Renner, C., Schleicher, M., Moroder, L., and Holak, T.A. (2002). Practical aspects of the 2D 15N-[1h]-NOE experiment. *J Biomol NMR* 23(1): 23-33.
- Richert, S., Luche, S., Chevallet, M., Van Dorsselaer, A., Leize-Wagner, E., and Rabilloud, T. (2004). About the mechanism of interference of silver staining with peptide mass spectrometry. *Proteomics* 4(4): 909-916.
- Rigaut, G., Shevchenko, A., Rutz, B., Wilm, M., Mann, M., and Seraphin, B. (1999). A generic protein purification method for protein complex characterization and proteome exploration. *Nat Biotechnol* 17(10): 1030-1032.
- Roth, M., Carpentier, P., Kaikati, O., Joly, J., Charrault, P., Pirocchi, M., Kahn, R., Fanchon, E., Jacquamet, L., Borel, F., et al. (2002). FIP: a highly automated beamline for multiwavelength anomalous diffraction experiments. *Acta Crystallogr D Biol Crystallogr* 58(Pt 5): 805-814.

- Roussel, A., and Cambillau, C. (1989). Silicon Graphics Geometry Partner Directory. *Silicon Graphics, Mountain View, CA*: 77-78.
- Rubbi, C.P., and Milner, J. (2003). Disruption of the nucleolus mediates stabilization of p53 in response to DNA damage and other stresses. *Embo J* 22(22): 6068-6077.
- Sali, A., Potterton, L., Yuan, F., van Vlijmen, H., and Karplus, M. (1995). Evaluation of comparative protein modeling by MODELLER. *Proteins* 23(3): 318-326.
- Schneider, T.R., and Sheldrick, G.M. (2002). Substructure solution with SHELXD. *Acta Crystallogr D Biol Crystallogr* 58(Pt 10 Pt 2): 1772-1779.
- Schultz, J., Milpetz, F., Bork, P., and Ponting, C.P. (1998). SMART, a simple modular architecture research tool: identification of signaling domains. *Proc Natl Acad Sci U S A* 95(11): 5857-5864.
- Sidik, K., and Smerdon, M.J. (1990). Nucleosome rearrangement in human cells following short patch repair of DNA damaged by bleomycin. *Biochemistry* 29(32): 7501-7511.
- Singh, R., and Valcarcel, J. (2005). Building specificity with nonspecific RNA-binding proteins. *Nat Struct Mol Biol* 12(8): 645-653.
- Smerdon, M.J., Tlsty, T.D., and Lieberman, M.W. (1978). Distribution of ultraviolet-induced DNA repair synthesis in nuclease sensitive and resistant regions of human chromatin. *Biochemistry* 17(12): 2377-2386.
- Sondermann, H., Nagar, B., Bar-Sagi, D., and Kuriyan, J. (2005). Computational docking and solution x-ray scattering predict a membrane-interacting role for the histone domain of the Ras activator son of sevenless. *Proc Natl Acad Sci U S A* 102(46): 16632-16637.
- Sonnhammer, E.L., Eddy, S.R., and Durbin, R. (1997). Pfam: a comprehensive database of protein domain families based on seed alignments. *Proteins* 28(3): 405-420.
- Stura, E.A., Satterthwait, A.C., Calvo, J.C., Kaslow, D.C., and Wilson, I.A. (1994). Reverse screening. *Acta Crystallogr D Biol Crystallogr* 50(Pt 4): 448-455.
- Svergun, D., Barberato, C., and Koch, M.H. (1995). CRY SOL— a Program to Evaluate X-ray Solution Scattering of Biological Macromolecules from Atomic Coordinates. *J. Appl. Cryst.* 28: 768–773.
- Svergun, D.I. (1992). Determination of the regularization parameter in indirect-transform methods using perceptual criteria. *J. Appl. Crystallogr* 25: 495-450.
- Svergun, D.I. (1999). Restoring low resolution structure of biological macromolecules from solution scattering using simulated annealing. *Biophys J* 76(6): 2879-2886.
- Svergun, D.I., and Koch, M.H. (2002). Advances in structure analysis using small-angle scattering in solution. *Curr Opin Struct Biol* 12(5): 654-660.
- Svergun, D.I., Petoukhov, M.V., and Koch, M.H. (2001). Determination of domain structure of proteins from X-ray solution scattering. *Biophys J* 80(6): 2946-2953.
- Svitkin, Y.V., Pause, A., Haghighat, A., Pyronnet, S., Witherell, G., Belsham, G.J., and Sonenberg, N. (2001). The requirement for eukaryotic initiation factor 4A (eIF4A) in translation is in direct proportion to the degree of mRNA 5' secondary structure. *Rna* 7(3): 382-394.
- Timchenko, T., Bailone, A., and Devoret, R. (1996). Btcd, a mouse protein that binds to curved DNA, can substitute in *Escherichia coli* for H-NS, a bacterial nucleoid protein. *Embo J* 15(15): 3986-3992.
- Tsutakawa, S.E., Hura, G.L., Frankel, K.A., Cooper, P.K., and Tainer, J.A. (2007). Structural analysis of flexible proteins in solution by small angle X-ray scattering combined with crystallography. *J Struct Biol* 158(2): 214-223.
- Vachette, P., Koch, M.H., and Svergun, D.I. (2003). Looking behind the beamstop: X-ray solution scattering studies of structure and conformational changes of biological macromolecules. *Methods Enzymol* 374: 584-615.

- van Eerde, A., Hamiaux, C., Perez, J., Parsot, C., and Dijkstra, B.W. (2004). Structure of Spa15, a type III secretion chaperone from *Shigella flexneri* with broad specificity. *EMBO Rep* 5(5): 477-483.
- Volkov, V.V., and Svergun, D.I. (2003). Uniqueness of *ab initio* shape determination in small-angle scattering. *J. Appl. Cryst.* 36: 860-864.
- Ward, J.J., Sodhi, J.S., McGuffin, L.J., Buxton, B.F., and Jones, D.T. (2004). Prediction and functional analysis of native disorder in proteins from the three kingdoms of life. *J Mol Biol* 337(3): 635-645.
- Will, C.L., and Lührmann, R. (2005). Spliceosome structure and function. *RNA World III Gesteland RF, Cech TR, Atkins JF Eds, CSH Laboratory Press*: 369-400.
- Yong, J., Wan, L., and Dreyfuss, G. (2004). Why do cells need an assembly machine for RNA-protein complexes? *Trends Cell Biol* 14(5): 226-232.
- Yuzawa, S., Yokochi, M., Hatanaka, H., Ogura, K., Kataoka, M., Miura, K., Mandiyan, V., Schlessinger, J., and Inagaki, F. (2001). Solution structure of Grb2 reveals extensive flexibility necessary for target recognition. *J Mol Biol* 306(3): 527-537.
- Zhou, Z., Licklider, L.J., Gygi, S.P., and Reed, R. (2002). Comprehensive proteomic analysis of the human spliceosome. *Nature* 419(6903): 182-185.



# **ANNEXES**





g11131248833	<i>Homo sapiens</i>	KRVHNN I VYNEYI - SHREH IHMNATQWETLTF	KVLELLEK
g1119505120	<i>Bos taurus</i>	KRVHNN I VYNEYI - SHREH IHMNATQWETLTF	KVLELLEK
g1173949092	<i>Canis familiaris</i>	KRVHNN I VYNEYI - SHREH IHMNATQWETLTF	KVLELLEK
g11109088173	<i>Musca domestica</i>	KRVHNN I VYNEYI - SHREH IHMNATQWETLTF	KVLELLEK
g1133859884	<i>Mus musculus</i>	KRVHNN I VYNEYI - SHREH IHMNATQWETLTF	KVLELLEK
g11109505121	<i>Rattus norvegicus</i>	KRVHNN I VYNEYI - SHREH IHMNATQWETLTF	KVLELLEK
g11149436978	<i>Omithorhynchus nasutus</i>	KRVHNN I VYNEYI - SHREH IHMNATQWETLTF	KVLELLEK
g11126340352	<i>Monodelphis domestica</i>	KRVHNN I VYNEYI - SHREH IHMNATQWETLTF	KVLELLEK
g11116989569	<i>Strongylocentrotus purpuratus</i>	KRVHNN I VYNEYI - SHREH IHMNATQWETLTF	KVLELLEK
g1182554382	<i>Camoreiro</i>	KRVHNN I VYNEYI - SHREH IHMNATQWETLTF	KVLELLEK
g1117059756	<i>Apis mellifera</i>	KRVHNN I VYNEYI - SHREH IHMNATQWETLTF	KVLELLEK
g1112597869	<i>Drosophila melanogaster</i>	KRVHNN I VYNEYI - SHREH IHMNATQWETLTF	KVLELLEK
g1124667403	<i>Drosophila melanogaster</i>	KRVHNN I VYNEYI - SHREH IHMNATQWETLTF	KVLELLEK
g1191078479	<i>Tribolium castaneum</i>	KRVHNN I VYNEYI - SHREH IHMNATQWETLTF	KVLELLEK
g1117510213	<i>Caenorhabditis elegans</i>	KRVHNN I VYNEYI - SHREH IHMNATQWETLTF	KVLELLEK
g11175493837	<i>Oryza sativa</i>	KRVHNN I VYNEYI - SHREH IHMNATQWETLTF	KVLELLEK
g11148604587	<i>Plasmodium falciparum</i>	KRVHNN I VYNEYI - SHREH IHMNATQWETLTF	KVLELLEK
g11148604587	<i>Plasmodium vivax</i>	KRVHNN I VYNEYI - SHREH IHMNATQWETLTF	KVLELLEK
g1168070251	<i>Plasmodium berghei</i>	KRVHNN I VYNEYI - SHREH IHMNATQWETLTF	KVLELLEK
g1182595924	<i>Plasmodium yoelii</i>	KRVHNN I VYNEYI - SHREH IHMNATQWETLTF	KVLELLEK
g1170945060	<i>Plasmodium chabaudi</i>	KRVHNN I VYNEYI - SHREH IHMNATQWETLTF	KVLELLEK
g11179182424	<i>Coccidioides immitis</i>	KRVHNN I VYNEYI - SHREH IHMNATQWETLTF	KVLELLEK
g11145343619	<i>Coccidioides immitis</i>	KRVHNN I VYNEYI - SHREH IHMNATQWETLTF	KVLELLEK
g1171031989	<i>Theileria parva</i>	KRVHNN I VYNEYI - SHREH IHMNATQWETLTF	KVLELLEK
g1184999764	<i>Theileria annulata</i>	KRVHNN I VYNEYI - SHREH IHMNATQWETLTF	KVLELLEK
g1118374164	<i>Tetrahymena thermophila</i>	KRVHNN I VYNEYI - SHREH IHMNATQWETLTF	KVLELLEK
g1145126029	<i>Gibberellia fujikuroi</i>	KRVHNN I VYNEYI - SHREH IHMNATQWETLTF	KVLELLEK
g1167362968	<i>Cryptosporidium parvum</i>	KRVHNN I VYNEYI - SHREH IHMNATQWETLTF	KVLELLEK
g1167483160	<i>Entamoeba histolytica</i>	KRVHNN I VYNEYI - SHREH IHMNATQWETLTF	KVLELLEK
g11123975648	<i>Trichomonas vaginalis</i>	KRVHNN I VYNEYI - SHREH IHMNATQWETLTF	KVLELLEK
g1171004968	<i>Ustilago maydis</i>	KRVHNN I VYNEYI - SHREH IHMNATQWETLTF	KVLELLEK
g113997819	<i>Magnaporthe oryzae</i>	KRVHNN I VYNEYI - SHREH IHMNATQWETLTF	KVLELLEK
g1111064711	<i>Phaeocephala namodorum</i>	KRVHNN I VYNEYI - SHREH IHMNATQWETLTF	KVLELLEK
g1185098338	<i>Neurospora crassa</i>	KRVHNN I VYNEYI - SHREH IHMNATQWETLTF	KVLELLEK
g1145188098	<i>Ashbya gossypii</i>	KRVHNN I VYNEYI - SHREH IHMNATQWETLTF	KVLELLEK
g11116198231	<i>Chaetomium globosum</i>	KRVHNN I VYNEYI - SHREH IHMNATQWETLTF	KVLELLEK
g1183772548	<i>Aspergillus oryzae</i>	KRVHNN I VYNEYI - SHREH IHMNATQWETLTF	KVLELLEK
g1114522865	<i>Aspergillus niger</i>	KRVHNN I VYNEYI - SHREH IHMNATQWETLTF	KVLELLEK
g1167904746	<i>Aspergillus nidulans</i>	KRVHNN I VYNEYI - SHREH IHMNATQWETLTF	KVLELLEK
g11121712092	<i>Aspergillus clavatus</i>	KRVHNN I VYNEYI - SHREH IHMNATQWETLTF	KVLELLEK
g1170986114	<i>Aspergillus fumigatus</i>	KRVHNN I VYNEYI - SHREH IHMNATQWETLTF	KVLELLEK
g11119474029	<i>Neosartorya fischeri</i>	KRVHNN I VYNEYI - SHREH IHMNATQWETLTF	KVLELLEK
g11116503387	<i>Coprinopsis cinerea</i>	KRVHNN I VYNEYI - SHREH IHMNATQWETLTF	KVLELLEK
g1163246511	<i>Rhizoglyphus mucedoni</i>	KRVHNN I VYNEYI - SHREH IHMNATQWETLTF	KVLELLEK
g1119112830	<i>Schizosaccharomyces pombe</i>	KRVHNN I VYNEYI - SHREH IHMNATQWETLTF	KVLELLEK
g11146421635	<i>Pichia guilliermondii</i>	KRVHNN I VYNEYI - SHREH IHMNATQWETLTF	KVLELLEK
g11126139029	<i>Pichia stipitis</i>	KRVHNN I VYNEYI - SHREH IHMNATQWETLTF	KVLELLEK
g1150292067	<i>Candida glabrata</i>	KRVHNN I VYNEYI - SHREH IHMNATQWETLTF	KVLELLEK
g1168477343	<i>Candida albicans</i>	KRVHNN I VYNEYI - SHREH IHMNATQWETLTF	KVLELLEK
g1160302811	<i>Kyuzenomyces laevis</i>	KRVHNN I VYNEYI - SHREH IHMNATQWETLTF	KVLELLEK
g1150421825	<i>Debaryomyces hansenii</i>	KRVHNN I VYNEYI - SHREH IHMNATQWETLTF	KVLELLEK
g11149239618	<i>Lodderomyces elongisporus</i>	KRVHNN I VYNEYI - SHREH IHMNATQWETLTF	KVLELLEK
g11154294129	<i>Aglonoma scopulatus</i>	KRVHNN I VYNEYI - SHREH IHMNATQWETLTF	KVLELLEK
g1158251638	<i>Cryptococcus neoformans</i>	KRVHNN I VYNEYI - SHREH IHMNATQWETLTF	KVLELLEK
g1150556508	<i>Yarrowia lipolytica</i>	KRVHNN I VYNEYI - SHREH IHMNATQWETLTF	KVLELLEK
g11154691424	<i>Sclerotinia sclerotiorum</i>	KRVHNN I VYNEYI - SHREH IHMNATQWETLTF	KVLELLEK
g11154311507	<i>Botryotinia fuckeliana</i>	KRVHNN I VYNEYI - SHREH IHMNATQWETLTF	KVLELLEK

120

90

g11131248833	<i>Homo sapiens</i>	KRVHNN I VYNEYI - SHREH IHMNATQWETLTF	KVLELLEK
g1119505120	<i>Bos taurus</i>	KRVHNN I VYNEYI - SHREH IHMNATQWETLTF	KVLELLEK
g1173949092	<i>Canis familiaris</i>	KRVHNN I VYNEYI - SHREH IHMNATQWETLTF	KVLELLEK
g11109088173	<i>Musca domestica</i>	KRVHNN I VYNEYI - SHREH IHMNATQWETLTF	KVLELLEK
g1133859884	<i>Mus musculus</i>	KRVHNN I VYNEYI - SHREH IHMNATQWETLTF	KVLELLEK
g11109505121	<i>Rattus norvegicus</i>	KRVHNN I VYNEYI - SHREH IHMNATQWETLTF	KVLELLEK
g11149436978	<i>Omithorhynchus nasutus</i>	KRVHNN I VYNEYI - SHREH IHMNATQWETLTF	KVLELLEK
g11126340352	<i>Monodelphis domestica</i>	KRVHNN I VYNEYI - SHREH IHMNATQWETLTF	KVLELLEK
g11116989569	<i>Strongylocentrotus purpuratus</i>	KRVHNN I VYNEYI - SHREH IHMNATQWETLTF	KVLELLEK
g1182554382	<i>Camoreiro</i>	KRVHNN I VYNEYI - SHREH IHMNATQWETLTF	KVLELLEK
g1117059756	<i>Apis mellifera</i>	KRVHNN I VYNEYI - SHREH IHMNATQWETLTF	KVLELLEK
g1112597869	<i>Drosophila melanogaster</i>	KRVHNN I VYNEYI - SHREH IHMNATQWETLTF	KVLELLEK
g1124667403	<i>Drosophila melanogaster</i>	KRVHNN I VYNEYI - SHREH IHMNATQWETLTF	KVLELLEK
g1191078479	<i>Tribolium castaneum</i>	KRVHNN I VYNEYI - SHREH IHMNATQWETLTF	KVLELLEK
g1117510213	<i>Caenorhabditis elegans</i>	KRVHNN I VYNEYI - SHREH IHMNATQWETLTF	KVLELLEK
g11175493837	<i>Oryza sativa</i>	KRVHNN I VYNEYI - SHREH IHMNATQWETLTF	KVLELLEK
g11148604587	<i>Plasmodium falciparum</i>	KRVHNN I VYNEYI - SHREH IHMNATQWETLTF	KVLELLEK
g11148604587	<i>Plasmodium vivax</i>	KRVHNN I VYNEYI - SHREH IHMNATQWETLTF	KVLELLEK
g1168070251	<i>Plasmodium berghei</i>	KRVHNN I VYNEYI - SHREH IHMNATQWETLTF	KVLELLEK
g1182595924	<i>Plasmodium yoelii</i>	KRVHNN I VYNEYI - SHREH IHMNATQWETLTF	KVLELLEK
g1170945060	<i>Plasmodium chabaudi</i>	KRVHNN I VYNEYI - SHREH IHMNATQWETLTF	KVLELLEK
g11179182424	<i>Coccidioides immitis</i>	KRVHNN I VYNEYI - SHREH IHMNATQWETLTF	KVLELLEK
g11145343619	<i>Coccidioides immitis</i>	KRVHNN I VYNEYI - SHREH IHMNATQWETLTF	KVLELLEK
g1171031989	<i>Theileria parva</i>	KRVHNN I VYNEYI - SHREH IHMNATQWETLTF	KVLELLEK
g1184999764	<i>Theileria annulata</i>	KRVHNN I VYNEYI - SHREH IHMNATQWETLTF	KVLELLEK
g1118374164	<i>Tetrahymena thermophila</i>	KRVHNN I VYNEYI - SHREH IHMNATQWETLTF	KVLELLEK
g1145126029	<i>Gibberellia fujikuroi</i>	KRVHNN I VYNEYI - SHREH IHMNATQWETLTF	KVLELLEK
g1167362968	<i>Cryptosporidium parvum</i>	KRVHNN I VYNEYI - SHREH IHMNATQWETLTF	KVLELLEK
g1167483160	<i>Entamoeba histolytica</i>	KRVHNN I VYNEYI - SHREH IHMNATQWETLTF	KVLELLEK
g11123975648	<i>Trichomonas vaginalis</i>	KRVHNN I VYNEYI - SHREH IHMNATQWETLTF	KVLELLEK
g1171004968	<i>Ustilago maydis</i>	KRVHNN I VYNEYI - SHREH IHMNATQWETLTF	KVLELLEK
g113997819	<i>Magnaporthe oryzae</i>	KRVHNN I VYNEYI - SHREH IHMNATQWETLTF	KVLELLEK
g1111064711	<i>Phaeocephala namodorum</i>	KRVHNN I VYNEYI - SHREH IHMNATQWETLTF	KVLELLEK
g1185098338	<i>Neurospora crassa</i>	KRVHNN I VYNEYI - SHREH IHMNATQWETLTF	KVLELLEK
g1145188098	<i>Ashbya gossypii</i>	KRVHNN I VYNEYI - SHREH IHMNATQWETLTF	KVLELLEK
g11116198231	<i>Chaetomium globosum</i>	KRVHNN I VYNEYI - SHREH IHMNATQWETLTF	KVLELLEK
g1183772548	<i>Aspergillus oryzae</i>	KRVHNN I VYNEYI - SHREH IHMNATQWETLTF	KVLELLEK
g1114522865	<i>Aspergillus niger</i>	KRVHNN I VYNEYI - SHREH IHMNATQWETLTF	KVLELLEK
g1167904746	<i>Aspergillus nidulans</i>	KRVHNN I VYNEYI - SHREH IHMNATQWETLTF	KVLELLEK
g11121712092	<i>Aspergillus clavatus</i>	KRVHNN I VYNEYI - SHREH IHMNATQWETLTF	KVLELLEK
g1170986114	<i>Aspergillus fumigatus</i>	KRVHNN I VYNEYI - SHREH IHMNATQWETLTF	KVLELLEK
g11119474029	<i>Neosartorya fischeri</i>	KRVHNN I VYNEYI - SHREH IHMNATQWETLTF	KVLELLEK
g11116503387	<i>Coprinopsis cinerea</i>	KRVHNN I VYNEYI - SHREH IHMNATQWETLTF	KVLELLEK
g1163246511	<i>Rhizoglyphus mucedoni</i>	KRVHNN I VYNEYI - SHREH IHMNATQWETLTF	KVLELLEK
g1119112830	<i>Schizosaccharomyces pombe</i>	KRVHNN I VYNEYI - SHREH IHMNATQWETLTF	KVLELLEK
g11146421635	<i>Pichia guilliermondii</i>	KRVHNN I VYNEYI - SHREH IHMNATQWETLTF	KVLELLEK
g11126139029	<i>Pichia stipitis</i>	KRVHNN I VYNEYI - SHREH IHMNATQWETLTF	KVLELLEK
g1150292067	<i>Candida glabrata</i>	KRVHNN I VYNEYI - SHREH IHMNATQWETLTF	KVLELLEK
g1168477343	<i>Candida albicans</i>	KRVHNN I VYNEYI - SHREH IHMNATQWETLTF	KVLELLEK
g1160302811	<i>Kyuzenomyces laevis</i>	KRVHNN I VYNEYI - SHREH IHMNATQWETLTF	KVLELLEK
g1150421825	<i>Debaryomyces hansenii</i>	KRVHNN I VYNEYI - SHREH IHMNATQWETLTF	KVLELLEK
g11149239618	<i>Lodderomyces elongisporus</i>	KRVHNN I VYNEYI - SHREH IHMNATQWETLTF	KVLELLEK
g11154294129	<i>Aglonoma scopulatus</i>	KRVHNN I VYNEYI - SHREH IHMNATQWETLTF	KVLELLEK
g1158251638	<i>Cryptococcus neoformans</i>	KRVHNN I VYNEYI - SHREH IHMNATQWETLTF	KVLELLEK
g1150556508	<i>Yarrowia lipolytica</i>	KRVHNN I VYNEYI - SHREH IHMNATQWETLTF	KVLELLEK
g11154691424	<i>Sclerotinia sclerotiorum</i>	KRVHNN I VYNEYI - SHREH IHMNATQWETLTF	KVLELLEK
g11154311507	<i>Botryotinia fuckeliana</i>	KRVHNN I VYNEYI - SHREH IHMNATQWETLTF	KVLELLEK

7113124883 Homo sapiens	LDDDEKTAKFIEEQVRRGGLLE.....GKEQEVPTFT...E-L-SRENDEE.....KVTFF- NL-SRG.....
71119905120 Bos taurus	LDDDEKTAKFIEEQVRRGGLLE.....GKEQEVPTFT...E-L-SRENDEE.....KVTFF- NL-NKG.....
71173949092 Canis familiaris	LDDDEKTAKFIEEQVRRGGLLE.....GKEQEAFTT...E-L-SRENDEE.....KVTFF- NL-NKG.....
71109088173 Musca domestica	LDDDEKTAKFIEEQVRRGGLLE.....GKEQEAFTT...E-L-SRENDEE.....KVTFF- NLS-KG.....
71133850684 Mus musculus	LDDDEKTAKFIEEQVRRGGLLE.....GKEQEPVFT...E-L-SRENDEE.....KVTFF- NL-NKG.....
71110950512 Pantodon nigricolus	LDDDEKTAKFIEEQVRRGGLLE.....GKEQEPVFT...E-L-SRENDEE.....KVTFF- NL-NKG.....
71149438978 Omithorhynchus anatinus	LDDDEKTAKFIEEQVRRGGLLE.....GKEQEPVFT...E-L-SRENDEE.....KVTFF- NL-NKG.....
711126340352 Monodelphis domestica	LDDDEKTAKFIEEQVRRGGLLE.....GKEQCPVFT...E-L-SRENDEE.....KVAFF- NL-NKG.....
71115898559 Strongylocentrotus purpuratus	LDDDEKTAKFIEEQVRRGGLLE.....VTEQNTPTT...E-L-SRONNEE.....KVAFF- NL-NKG.....
71182524382 Danio rerio	LDDDEKTAKFIEEQVRRGREG.....REPTTEPFT...E-L-QRKSDE.....KVAFF- NLEIA.....
71110759766 Apis mellifera	KDDEERMMEFIEKQVKKGGQ.....ESNEQSEVTK...EPLKRLDND...PLVL- NMMK.....
71125978581 Drosophila obscura	KDDEERTDFIEKQIKKAKKDD.....GDSGQEKY...E-LVRE-EAE.....PLKL- DIRLE.....
7124667403 Drosophila melanogaster	KDDEERMADIEEQIKKAKKGG.....EEDGQEKFT...E-L-KRENE.....PLKL- DIRLE.....
71191078478 Tribolium castaneum	KDDEERTLEFVQKVELAQR.....GLTKNEPVT...E-L-IRENEE.....PLKL- DLKLE.....
71117510213 Caenorhabditis elegans	KDDEERHMOIMDGMVQRGELA.....GDDEHEVEAT...E-L-IRDPO.....KIQD- DLNLOILDR.....
71115495383 Oryza sativa	LAEDEGERMIARQIERAAQSMGKTNELGDDASPDGSEGESSEDEYSDS.....ENDHEQDEEAKKAAAGKIAI...ALQRA.....
7118405389 Arabidopsis thaliana	LAEDEGERIERIQIEAAEKKLGGGGE.....GETSGNDEVDDGDDERKEDDL.....RLKS- GVKVG.....
71166825921 Drosophila melanogaster	LNEERQRLQIEKQIKELNKN.....KIEQDEIKPT...E-L-KRENEE.....LLEL- NIKST.....
71118781660 Anopheles gambiae	KDPAERLAEFIEEQVRRGKT.....EEEPCTSGYS...E-L-KRENEE.....TIKI- ELKLG.....
711108094475 Aedes aegypti	KDDEERLAEFIEEQVRRGKN.....EEEPSTSYT...E-L-KRONNEE.....TIKL- DLKLG.....
71124610266 Plasmodium falciparum	YSYEELEKKNIEAIEQAKKGTGFTIESEYTALE.....EKNDEKII...SSIKTNNDTNSIKDKPNNIFLDKLLKG.....
71148804587 Plasmodium vivax	YSYEELEKKNIEAIEQAKKGTGFTIESEYTALE.....NKKDEKIVISSIKATRESSASKA.....GVNV- LIAKA.....
71188070251 Plasmodium berghei	YSYEEEMKKNQVIEAKKGGQFIESEYTG1.....EKKNDEKIISSVK-ATSTNTSINHNDVPKSNIFL- DILKG.....
71182595924 Plasmodium yoelii	YSYEEEMKKNQVIEAKK.....TGQFIOSEYTG1...G-IEKKNDD.....KIII- SSVKP.....
71170945060 Plasmodium chabaudi	YSYEEEMKKNIEVIEAKKGTGQFIESEYTG1.....EKKNDEKIVISSVK-APSTNTTAAHNDGIPKSNIFL- DILKG.....
71119182424 Coccidioides immitis	RQDEEREKLEDDQIKRAQDKAQTG.....TDGDAPEAP...KKLQRYEVE.....KITL- NFASK.....
711145343619 Cryptosporidium hominis	DGSAKQEKLEIQVQRKGVAV.....PAKPEQHEPT.....GKLL.....
71171031985 Theileria parva	ITLEQMHQRILQKLVVEAKES.....GAFHEEVEY...P-LMKDPDQ.....KIVI- NTNLT.....
71184999784 Theileria annulata	ITLEQMHQRILQKLVVEAKES.....GAFDREEY...P-LMKDPDQ.....KIVI- NTNLT.....
71118374164 Tetrahymena thermophila	KQNEELQRMILREQIKRAQAAGN.....TMNKNLEAKNEQMDLEDLEGGQF.....EFGISFLKKG.....
71146126029 Giardia lamblia	FTEDNRR.....DEEKKEEPEA...RELKREGE.....KIKL- SFGAK.....
71167592361 Cryptosporidium hominis	DSKNK.....EAE.....
71166362966 Cryptosporidium parvum	NKOLENTRSSIEQMK.....TAEESKPAVD.....
71187483160 Entamoeba histolytica	MDDQORRLLQEQIDRA.....AEQLADVPAS.....
71123975648 Trichomonas vaginalis	RQDEEREQRLLEQIKRAQK.....GADGAEVEG...KELKREGEK.....
71139978119 Megascops asio	ARSGAEDKMLKMKKARAE.....DKEEEDPEA...RNLQRKEGE.....KVKL- NIGFG.....
7111106471 Phaeocephala irroratorum	EQAKVEASKGEETAKT.....KEEEVDPEA...RNLQRKEGE.....KVKL- NIGFG.....
71185098338 Neurospora crassa	KHEEDIRONMLLQIEFGK.....AEQLADVPAS.....
71145188098 Ashbya gossypii	KQDEEREQMLREQIRRAQDAEARGVL.....
71116158231 Chaetomium globosum	RQDEEREQRLLEQVERARQNA.....
71183772548 Aspergillus oryzae	RQDEEREQRLREQIKRAQSA.....KEEEVDPEA...RNLQRKEGE.....KVKL- NIGFG.....
71146228565 Aspergillus niger	RQDEEREQRLREQIKRAQSA.....KEEEVDPEA...RNLQRKEGE.....KVKL- NIGFG.....
71167904746 Aspergillus nidulans	RQDEEREQRLREQIKRAQNA.....EKDEIDPEA...RVLQRKEGE.....KVKL- NIGFG.....
71121712092 Aspergillus clavatus	RQDEEREQRLREQIKRAQNA.....EKDEIDPEA...RVLQRKEGE.....KVKL- NIGFG.....
71170986114 Aspergillus fumigatus	RQDEEREQRLREQIKRAQNA.....EKDEIDPEA...RVLQRKEGE.....KVKL- NIGFG.....
71119474029 Neosartorya fischeri	RQDEEREQRLREQIKRAQNA.....EKDEIDPEA...RVLQRKEGE.....KVKL- NIGFG.....
71116503367 Coprinopsis cinerea	MSDEERERQLAEQIERAKESODA.....GASRDSPEE...RELKREGE.....KVKL- SLS.....
7116324651 P23 Saccharomyces cerevisiae	KSDEEQRLLEQIKRAVESAGNNEBNDK.....GSREOPVLH...E-IDLSKGN.....PIQL- NLSSS.....
7119112830 Schizosaccharomyces pombe	KNDEEQMRRFLSQIEAG.....LNEESOPEDV.....QVPV- TVQSQ.....
71146421635 Pichia guilliermondii	RQDEQLTNKFLLEQIRRGESAK.....
71126139029 Pichia stipitis	IEAELTESKLEEYLR.....
71150292067 Candida glabrata	DNDEEMDDKLLQRLKRGQE.....
71188477343 Candida albicans	RSQDEIKLRLLEKQIEKA.....
71160302811 Kluyveromyces fragilis	KTIEQMSMKLINDQIKKSGE.....YMNHRKOEFK...E.....
71150421825 Debaryomyces hansenii	HDDEEFGNLEINLIDGKNFL.....EKRRDVTDA...NDVDSIEAER...LLRK- QIERO.....
71149235618 Lodderomyces longisporus	HDDEEERMMIEEQVRRAR.....DRE.....
711154284129 Heliomyces caposulatus	MSDEERKMLLEEQIARAR.....AOADDPDLT...VGLQKKEGE.....KIKL- NL.....
71168261638 Cryptococcus neoformans	-DTETMERKRIEAVKKAEN.....GVVEVKEAV...LTKTDAP1...QLSL- GLKKA.....
711546991424 Sclerotinia sclerotiorum	KGDEEREQMLIREQVRRAKOD.....AARGESEVDEEEROLKREEGEKI...KLSF- GAKPA.....
711154311507 Botryotinia fuckeliana	KGDEEREQMLIREQVRRAKOD.....AEA...RGESEMDQE...ROLKREEGEKI...KLSF- GAKSA.....

210

180

gll13124883J Homoeopians	..ADSSSGAATSSKSSSTLGP.....SALKTIIGSSASVKKRKESSQSSTQSSKKEKKK..KKSAL..	..DEIM..EI
gll119905120J Boutaurus	..ACSS..GAASSKSSSLGP.....SALKTIGTVASVKKRKESSQSAPSKPKK..KKSAL..	..DEIM..EI
gll173949092J Canisfamiliatis	..ACSS..AAASSKSSSLGP.....SALNAIGNAASVKKRKESSQAQSKKKEKK..KKSAL..	..DEIM..EI
gll109088173J Mezocamelus	..ACSSGATSSKSSSTLGP.....SALKTIIGSSASVKKRKESSQAQSKKKEKK..KKSAL..	..DEIM..EI
gll133859684J Mumunuculus	..AGGSAAGATSSKSSSLGP.....SALKLLGSAAGKRRKESQAQSSQAQ..AKK..KKSAL..	..DEIM..EL
gll109505121J Panthuonorychius	..AGSSAAGATSSKSSSLGP.....SALKMLGSAVKKRKESSQAQ..AKK..KKSAL..	..DEIM..EI
gll149436978J Ornithonychanianinus	..ASTSSSTTPKSSSLGP.....SALKMVKGAAAKRKSQSSQAQSKK..KKSAL..	..DEIM..EF
gll126340352J Monodelphisdomestica	..ASTS..GTSKPKSSSLGP.....SAL.....TAASVKKRKESSQLSQSKKKEKK..KKSAL..	..DEIM..EL
gll116898659J Strongylocentrotusgauripuratus	..ASTS..GATTKSSSLVS.....NALKSAAMGsvkrkdsqovpke.....KK..KKSAL..	..DEIM..EM
gll182324382J Canionerio	..A.....AGPSKPSSTLGP.....SALKTTASVKKRKESSLSG..EPKDKK..KKSAL..	..DEII..EM
gll1110759756J Apimelilliera	..KKFPESILGKSALLPK.....RPKLENEEYDKETSSAKIEKESK..LNNDN..	..DNM.....
gll125978681J Dicocephalaceudoobscura	..KKFPQDVLGKSALKR.....PAPEAEKVFVKPKSVAQDSQ.....TRSVL..	..DEII..QEE
gll194076478J Triboliumcaezareum	..POQKDDV.....KPVVFT.....DLKRSLDGPKSKTIKKK..KTENR..KKTAL..	..DEIK..EIE
gll17510213J Caenorhabditiselegans	KLDVKSVASAKISIFDMPVKVKEDEPQSPRSKSGKRSRSRFAAKFKSKSAL..	..DEIKEMEE
gll115453837J Oryziasathva	..VFGPKVNPVLDATKPKVFGFEEDEVSARDKEELAKKKGKDAIAAEA..RRSAL..	..DELM..KEE
gll148804587J Plasmodiumfalciparum	..FALGGVQVATGKERGESKLLFQDEENDKVERGEKRRKSDSGRSEKE..RRSAL..	..DELM..KEE
gll68070251J Plasmodiumberghei	..TMTTTTTNTTTTNTK.....NIFDKLKTNNSSNNYNDQTNPKPYAK..KMSAI..	..EIMFKEK
gll82595924J Plasmodiumyoelii	..SKQQSTPSAVISK.....RPFDALDDGKKEKKIKAAATSGETK..KLSAL..	..DELI..QEE
gll118781660J Anophelesgambiae	..TNIKMESKPKLEK.....RPLEVADEFKKEKFKPGASSSSQFK..KPSAL..	..DELI..REE
gll109884475J Aedesegypti	..DSQNKKKSITDDKNIHRNNTKNRPLSELDLILENEEKHKHNDTSK..LKNNP..	..NYAYISELHNEHK
gll124810266J Plasmodiumfalciparum	..LSGEEGKHVADSLPKRKFVENDVYVKGAAKRRPLSILELLENAEK..RVNT..	..AANLKNQNSHSEMEKQEE
gll166362966J Cryptosporidiumparvum	..NKIKTNDLKAESIKK.....EELSKDKPONKRRPMSLELLIMENEKK..KLNKLLPNLKN	..IANNKSKTEHAHSITKHMENH
gll67483160J Entamoebahistolytica	..TNTNPSTIGHNDVHPKSNIFDLILKKNKIKTNDOLK.....	..SENIKKEE
gll123975648J Trichomonasvaginalis	..NKIKSANDOLKAESTKK.....EEALSCKDSKKNRPMSSLELLIMENEKK..KIKNLLPNLKN	..MANNANPKAEHTHSITNHNENH
gll171004866J Ustilagomaydis	..LTPQENSATSSASEKPD.....	..
gll119182424J Coccidioidesimmitis	..QOTGTNMDLTKRKLQ.....	..
gll145343619J Coccidioideslucimarinus	..TTMSESNKGLNVFKOYRKNKSTIGKOEPEPTKTEGSK.....RKSAL..	..ESIM.....
gll171031985J Theileriaparna	..EENKGVNVFKOYRKNKSSNEKQEPITSKTEGSKRKSLELIMESNKR..KESHY..	..DKAL..
gll8499764J Theileriaannulata	..EQEDEQEENK.....EELISFVQQQNNQOQNEQAAPVTENKIENNTQI..	..NQDDGQI
gll18374164J Tetranychumathemophila	..PAAPDTKASESPAPDTAETADLTKEETIATEKRRE.....	..
gll46126029J Gibberellazase	..AAAGNGASTEQDDE.....	..
gll167592361J Cryptosporidiumtomitris	..IKLSFGAKTAKSTT.....	..
gll185098338J Neurosporaoraesa	..SKANGDSKTESPKQSP.....	..DEV..
gll145188098J Ashbyagocyptii	..SKANGDSKTESPKQSP.....	..
gll1116198231J Chaetomiumglobosum	..SKTNGDSKTESPKQSP.....	..
gll183772548J Aspergillusoryzae	..AKATPPASTEQSRQSP.....	..
gll145228665J Aspergillusniger	..SKAAGEPKAAESRTQSP.....	..
gll167904746J Aspergillumdulans	..SKTNGESKSGAASRTQSP.....	..
gll121712092J Aspergillusclavatus	..SKTNGESKSGAASRTQSP.....	..
gll709886114J Aspergillusfumigatus	..SKTNGESKSGAASRTQSP.....	..
gll119474029J Neosartoryafischeri	..SKTNGESKSVTSRTQSP.....	..
gll116503387J Coprinopsiscinerea	..EEDKKEPSSAT.....	..
gll63246511R7S2 Saccharomyceserevisiae	..SDSHSAQNEFFQRTNTP...TFSSSSSSTLKHKPKNVFAELNKSRRK..	..
gll19112830J Schizosaccharomycespombe	..PIKVLKSTQITKQLQR.....PSKSIQDAASDDEADKPEPKET..KQKIK..	..
gll146421635J Pichiahiemondii	..	..
gll126139029J Pichiachipits	..	..
gll150292067J Candidaglabrata	..	..
gll168477343J Candidaalbicans	..	..
gll150302811J Kluyveromycesaltis	..	..
gll150421825J Debaryomyceshanseii	..KR..	..EKKIRKEL
gll149235618J Lodderomyceselongisporus	..	..
gll154264129J Heliomycesosulatus	..	..
gll168261638J Cryptococcusneoformans	..	..
gll15055405J Yarrowia lipolytica	..VSNPEKTVVLPDSKMPAGDALPLETAKGVKDSSEVLAKDGSKTA..	..
gll11546891424J Sclerotinia sclerotiorum	..VSNSEKTIPEESKTP.....	..
gll154311507J Botryotinia fuckelliana	..	..

g1113124883 Homosapiens	300	.....TARTDY..WLPEI VKI ITKLLGK..YHKKKAI VKEVID..KYTAVVKM DSGD...KLLDQDTHLETVI
g11199005120 Bostaurus	330	.....TARTDY..WLPEI VKI ITKLLGK..YHKKKAI VKEVID..KYTAVVKM DSGD...KLLDQDTHLETVI
g1173949092 Carnifamiliaris	300	.....TARIDH..WLPEI VKI ITKLLGK..YHKKKG VKEVID..KYTAVVKM DSGD...KLLDQDTHLETVI
g11109088173 Musmusculata	300	.....TARTDY..WLPEI VKI ITKLLGK..YHKKKG VKEVID..KYTAVVKM DSGD...KLLDQDTHLETVI
g1133859684 Musmusculus	300	.....TARTDY..WLPEI VKI ITKLLGK..YHKKKG VKEVID..KYTAVVKM DSGD...KLLDQDTHLETVI
g11109505121 Rattusnorvegicus	300	.....TARTDY..WLPEI VKI ITKLLGK..YHKKKG VKEVID..KYTAVVKM DSGD...KLLDQDTHLETVI
g11149438978 Omithorhynchusnatus	300	.....SRTDYS..WLPEI VKI ITKLLGK..YHKKKAVI KEVID..KYTAVVKM DSGD...KLLDQDTHLETVI
g11126340352 Monodelphisdomestica	300	.....SRTDYS..WLPEI VKI ITKLLGK..YHKKKG VKEVID..KYTAVVKM DSGD...KLLDQDTHLETVI
g1119898559 Strongylocentrotuspurpuratus	300	.....SRTDYS..WLPEI VKI ITKLLGK..YHKKKAVK EVID..KYTAVVKM DSGD...KLLDQDTHLETVI
g1182524382 Daniorerio	300	.....SRTDYS..WLPEI VKI ITKLLGK..YHKKKAVK EVID..KYTAVVKM DSGD...KLLDQDTHLETVI
g1110759756 Apis mellifera	300	.....SRTDYS..WLPEI VKI ITKLLGK..YHKKKAVK EVID..KYTAVVKM DSGD...KLLDQDTHLETVI
g1112597858 Drosophilamelanogaster	300	.....SRTDYS..WLPEI VKI ITKLLGK..YHKKKAVK EVID..KYTAVVKM DSGD...KLLDQDTHLETVI
g11910784781 Thoboliumcastaneum	300	.....SRTDYS..WLPEI VKI ITKLLGK..YHKKKAVK EVID..KYTAVVKM DSGD...KLLDQDTHLETVI
g1117510213 Caenorhabditis elegans	300	.....SRTDYS..WLPEI VKI ITKLLGK..YHKKKAVK EVID..KYTAVVKM DSGD...KLLDQDTHLETVI
g11115453837 Oryziasativae	300	.....SRTDYS..WLPEI VKI ITKLLGK..YHKKKAVK EVID..KYTAVVKM DSGD...KLLDQDTHLETVI
g1118405389 Arabidopsis thaliana	300	.....SRTDYS..WLPEI VKI ITKLLGK..YHKKKAVK EVID..KYTAVVKM DSGD...KLLDQDTHLETVI
g1166825921 Drosophilamelanogaster	300	.....SRTDYS..WLPEI VKI ITKLLGK..YHKKKAVK EVID..KYTAVVKM DSGD...KLLDQDTHLETVI
g11178781660 Anophelesgambiae	300	.....SRTDYS..WLPEI VKI ITKLLGK..YHKKKAVK EVID..KYTAVVKM DSGD...KLLDQDTHLETVI
g111088847475 Aedes aegypti	300	.....SRTDYS..WLPEI VKI ITKLLGK..YHKKKAVK EVID..KYTAVVKM DSGD...KLLDQDTHLETVI
g11124810266 Plasmodium falciparum	300	.....SRTDYS..WLPEI VKI ITKLLGK..YHKKKAVK EVID..KYTAVVKM DSGD...KLLDQDTHLETVI
g11148804587 Plasmodium vivax	300	.....SRTDYS..WLPEI VKI ITKLLGK..YHKKKAVK EVID..KYTAVVKM DSGD...KLLDQDTHLETVI
g1168070251 Plasmodium berghei	300	.....SRTDYS..WLPEI VKI ITKLLGK..YHKKKAVK EVID..KYTAVVKM DSGD...KLLDQDTHLETVI
g1182595924 Plasmodium yoelii	300	.....SRTDYS..WLPEI VKI ITKLLGK..YHKKKAVK EVID..KYTAVVKM DSGD...KLLDQDTHLETVI
g1170945606 Plasmodium chabaudi	300	.....SRTDYS..WLPEI VKI ITKLLGK..YHKKKAVK EVID..KYTAVVKM DSGD...KLLDQDTHLETVI
g1119182424 Coccidioides immitis	300	.....SRTDYS..WLPEI VKI ITKLLGK..YHKKKAVK EVID..KYTAVVKM DSGD...KLLDQDTHLETVI
g11145343619 Cryptosporidium parvum	300	.....SRTDYS..WLPEI VKI ITKLLGK..YHKKKAVK EVID..KYTAVVKM DSGD...KLLDQDTHLETVI
g1171031980 Theileria parva	300	.....SRTDYS..WLPEI VKI ITKLLGK..YHKKKAVK EVID..KYTAVVKM DSGD...KLLDQDTHLETVI
g1184989764 Theileria annulata	300	.....SRTDYS..WLPEI VKI ITKLLGK..YHKKKAVK EVID..KYTAVVKM DSGD...KLLDQDTHLETVI
g1118374764 Theileria theileri	300	.....SRTDYS..WLPEI VKI ITKLLGK..YHKKKAVK EVID..KYTAVVKM DSGD...KLLDQDTHLETVI
g1146126029 Gibberella zeae	300	.....SRTDYS..WLPEI VKI ITKLLGK..YHKKKAVK EVID..KYTAVVKM DSGD...KLLDQDTHLETVI
g1167592361 Cryptosporidium parvum	300	.....SRTDYS..WLPEI VKI ITKLLGK..YHKKKAVK EVID..KYTAVVKM DSGD...KLLDQDTHLETVI
g1166362966 Cryptosporidium parvum	300	.....SRTDYS..WLPEI VKI ITKLLGK..YHKKKAVK EVID..KYTAVVKM DSGD...KLLDQDTHLETVI
g1167483160 Eritriaceae hirtellia	300	.....SRTDYS..WLPEI VKI ITKLLGK..YHKKKAVK EVID..KYTAVVKM DSGD...KLLDQDTHLETVI
g11123975648 Trichomonas vaginalis	300	.....SRTDYS..WLPEI VKI ITKLLGK..YHKKKAVK EVID..KYTAVVKM DSGD...KLLDQDTHLETVI
g1171004866 Ustilago maydis	300	.....SRTDYS..WLPEI VKI ITKLLGK..YHKKKAVK EVID..KYTAVVKM DSGD...KLLDQDTHLETVI
g1139878119 Magnaporthe oryzae	300	.....SRTDYS..WLPEI VKI ITKLLGK..YHKKKAVK EVID..KYTAVVKM DSGD...KLLDQDTHLETVI
g1111064711 Phaeosphaeria nodorum	300	.....SRTDYS..WLPEI VKI ITKLLGK..YHKKKAVK EVID..KYTAVVKM DSGD...KLLDQDTHLETVI
g1185098338 Neurospora crassa	300	.....SRTDYS..WLPEI VKI ITKLLGK..YHKKKAVK EVID..KYTAVVKM DSGD...KLLDQDTHLETVI
g1145188098 Aschygosporium	300	.....SRTDYS..WLPEI VKI ITKLLGK..YHKKKAVK EVID..KYTAVVKM DSGD...KLLDQDTHLETVI
g1116198231 Chaetomium globosum	300	.....SRTDYS..WLPEI VKI ITKLLGK..YHKKKAVK EVID..KYTAVVKM DSGD...KLLDQDTHLETVI
g1183772548 Aspergillus oryzae	300	.....SRTDYS..WLPEI VKI ITKLLGK..YHKKKAVK EVID..KYTAVVKM DSGD...KLLDQDTHLETVI
g11146228665 Aspergillus niger	300	.....SRTDYS..WLPEI VKI ITKLLGK..YHKKKAVK EVID..KYTAVVKM DSGD...KLLDQDTHLETVI
g1167904746 Aspergillus nidulans	300	.....SRTDYS..WLPEI VKI ITKLLGK..YHKKKAVK EVID..KYTAVVKM DSGD...KLLDQDTHLETVI
g11121712092 Aspergillus clavatus	300	.....SRTDYS..WLPEI VKI ITKLLGK..YHKKKAVK EVID..KYTAVVKM DSGD...KLLDQDTHLETVI
g1170986114 Aspergillus fumigatus	300	.....SRTDYS..WLPEI VKI ITKLLGK..YHKKKAVK EVID..KYTAVVKM DSGD...KLLDQDTHLETVI
g1119474029 Neosartorya fischeri	300	.....SRTDYS..WLPEI VKI ITKLLGK..YHKKKAVK EVID..KYTAVVKM DSGD...KLLDQDTHLETVI
g1118503387 Coprinopsis cinerea	300	.....SRTDYS..WLPEI VKI ITKLLGK..YHKKKAVK EVID..KYTAVVKM DSGD...KLLDQDTHLETVI
g118324651 Pezizomyces arvensis	300	.....SRTDYS..WLPEI VKI ITKLLGK..YHKKKAVK EVID..KYTAVVKM DSGD...KLLDQDTHLETVI
g1119112830 Schizosaccharomyces pombe	300	.....SRTDYS..WLPEI VKI ITKLLGK..YHKKKAVK EVID..KYTAVVKM DSGD...KLLDQDTHLETVI
g11146421635 Pichia guilliermondii	300	.....SRTDYS..WLPEI VKI ITKLLGK..YHKKKAVK EVID..KYTAVVKM DSGD...KLLDQDTHLETVI
g11126139029 Pichia stipitis	300	.....SRTDYS..WLPEI VKI ITKLLGK..YHKKKAVK EVID..KYTAVVKM DSGD...KLLDQDTHLETVI
g1150292067 Candida glabrata	300	.....SRTDYS..WLPEI VKI ITKLLGK..YHKKKAVK EVID..KYTAVVKM DSGD...KLLDQDTHLETVI
g1188477343 Candida albicans	300	.....SRTDYS..WLPEI VKI ITKLLGK..YHKKKAVK EVID..KYTAVVKM DSGD...KLLDQDTHLETVI
g1150302811 Kuyveromyces haenkei	300	.....SRTDYS..WLPEI VKI ITKLLGK..YHKKKAVK EVID..KYTAVVKM DSGD...KLLDQDTHLETVI
g1150421825 Debaryomyces hansenii	300	.....SRTDYS..WLPEI VKI ITKLLGK..YHKKKAVK EVID..KYTAVVKM DSGD...KLLDQDTHLETVI
g11149235616 Lodderomyces elongisporus	300	.....SRTDYS..WLPEI VKI ITKLLGK..YHKKKAVK EVID..KYTAVVKM DSGD...KLLDQDTHLETVI
g11154284129 Yellow yeasts	300	.....SRTDYS..WLPEI VKI ITKLLGK..YHKKKAVK EVID..KYTAVVKM DSGD...KLLDQDTHLETVI
g1188261636 Cryptococcus neoformans	300	.....SRTDYS..WLPEI VKI ITKLLGK..YHKKKAVK EVID..KYTAVVKM DSGD...KLLDQDTHLETVI
g1150555405 Yarrowia lipolytica	300	.....SRTDYS..WLPEI VKI ITKLLGK..YHKKKAVK EVID..KYTAVVKM DSGD...KLLDQDTHLETVI
g11154691424 Sclerotinia sclerotiorum	300	.....SRTDYS..WLPEI VKI ITKLLGK..YHKKKAVK EVID..KYTAVVKM DSGD...KLLDQDTHLETVI
g11154311507 Botrytis cinerea	300	.....SRTDYS..WLPEI VKI ITKLLGK..YHKKKAVK EVID..KYTAVVKM DSGD...KLLDQDTHLETVI

g 113124883 Homosapiens	360	PAPGKRILVLN-66YRGNEGTLES..INEKFTSATIVIETGPKKGRRVE6 IQYEDISKLA	390	PAPGKRILVLN-66YRGNEGTLES..INEKFTSATIVIETGPKKGRRVE6 IQYEDISKLA
g 119905120 Bosataurus		PAPGKRILVLN-66YRGNEGTLES..INEKFTSATIVIETGPKKGRRVE6 IQYEDISKLA		PAPGKRILVLN-66YRGNEGTLES..INEKFTSATIVIETGPKKGRRVE6 IQYEDISKLA
g 173849092 Canisfamiliars		PAPGKRILVLN-66YRGNEGTLES..INEKFTSATIVIETGPKKGRRVE6 IQYEDISKLA		PAPGKRILVLN-66YRGNEGTLES..INEKFTSATIVIETGPKKGRRVE6 IQYEDISKLA
g 1109088173 Macacamilatta		PAPGKRILVLN-66YRGNEGTLES..INEKFTSATIVIETGPKKGRRVE6 IQYEDISKLA		PAPGKRILVLN-66YRGNEGTLES..INEKFTSATIVIETGPKKGRRVE6 IQYEDISKLA
g 1338595664 Musmusculus		PAPGKRILVLN-66YRGNEGTLES..INEKFTSATIVIETGPKKGRRVE6 IQYEDISKLA		PAPGKRILVLN-66YRGNEGTLES..INEKFTSATIVIETGPKKGRRVE6 IQYEDISKLA
g 10950512 Pattumnorvegicus		PAPGKRILVLN-66YRGNEGTLES..INEKFTSATIVIETGPKKGRRVE6 IQYEDISKLA		PAPGKRILVLN-66YRGNEGTLES..INEKFTSATIVIETGPKKGRRVE6 IQYEDISKLA
g 149436978 Ornithorynchusatinus		PAPGKRILVLN-66YRGNEGTLES..INEKFTSATIVIETGPKKGRRVE6 IQYEDISKLA		PAPGKRILVLN-66YRGNEGTLES..INEKFTSATIVIETGPKKGRRVE6 IQYEDISKLA
g 1126340352 Monodelphisdomestica		PAPGKRILVLN-66YRGNEGTLES..INEKFTSATIVIETGPKKGRRVE6 IQYEDISKLA		PAPGKRILVLN-66YRGNEGTLES..INEKFTSATIVIETGPKKGRRVE6 IQYEDISKLA
g 116896559 Strongyloidespurpuratus		PAPGKRILVLN-66YRGNEGTLES..INEKFTSATIVIETGPKKGRRVE6 IQYEDISKLA		PAPGKRILVLN-66YRGNEGTLES..INEKFTSATIVIETGPKKGRRVE6 IQYEDISKLA
g 82524382 Dantiorio		PAIGKEVILN-66YRGNEGTLES..INEKFTSATIVIETGPKKGRRVE6 IQYEDISKLA		PAIGKEVILN-66YRGNEGTLES..INEKFTSATIVIETGPKKGRRVE6 IQYEDISKLA
g 110759756 Apis mellifera		PAMGKTVLVN-66YRGNEGTLES..INEKFTSATIVIETGPKKGRRVE6 IQYEDISKLA		PAMGKTVLVN-66YRGNEGTLES..INEKFTSATIVIETGPKKGRRVE6 IQYEDISKLA
g 12597858 Drosophilapseudoobscura		PALDKPMVYN-66YRGNEGTLES..INEKFTSATIVIETGPKKGRRVE6 IQYEDISKLA		PALDKPMVYN-66YRGNEGTLES..INEKFTSATIVIETGPKKGRRVE6 IQYEDISKLA
g 24667403 Drosophilamelanogaster		PAIGKLVKVN-66YRGNEGTLES..INEKFTSATIVIETGPKKGRRVE6 IQYEDISKLA		PAIGKLVKVN-66YRGNEGTLES..INEKFTSATIVIETGPKKGRRVE6 IQYEDISKLA
g 9191078478 Triboliumcastaneum		PSLGRQMMIYN-66YRGNEGTLES..INEKFTSATIVIETGPKKGRRVE6 IQYEDISKLA		PSLGRQMMIYN-66YRGNEGTLES..INEKFTSATIVIETGPKKGRRVE6 IQYEDISKLA
g 115453837 Oryza sativa		PQIGGLVRIYN-66YRGNEGTLES..INEKFTSATIVIETGPKKGRRVE6 IQYEDISKLA		PQIGGLVRIYN-66YRGNEGTLES..INEKFTSATIVIETGPKKGRRVE6 IQYEDISKLA
g 18405389 Arabidopsis thaliana		PQIGGMVKIYN-66YRGNEGTLES..INEKFTSATIVIETGPKKGRRVE6 IQYEDISKLA		PQIGGMVKIYN-66YRGNEGTLES..INEKFTSATIVIETGPKKGRRVE6 IQYEDISKLA
g 16682592 Drosophilamelanogaster		PAIGKLVKVN-66YRGNEGTLES..INEKFTSATIVIETGPKKGRRVE6 IQYEDISKLA		PAIGKLVKVN-66YRGNEGTLES..INEKFTSATIVIETGPKKGRRVE6 IQYEDISKLA
g 118781650 Anopheles gambiae		PAIGKLVKVN-66YRGNEGTLES..INEKFTSATIVIETGPKKGRRVE6 IQYEDISKLA		PAIGKLVKVN-66YRGNEGTLES..INEKFTSATIVIETGPKKGRRVE6 IQYEDISKLA
g 110886475 Aedes aegypti		PAIGKLVKVN-66YRGNEGTLES..INEKFTSATIVIETGPKKGRRVE6 IQYEDISKLA		PAIGKLVKVN-66YRGNEGTLES..INEKFTSATIVIETGPKKGRRVE6 IQYEDISKLA
g 1124810266 Plasmodium falciparum		PKIGRHLVILN-66YRGNEGTLES..INEKFTSATIVIETGPKKGRRVE6 IQYEDISKLA		PKIGRHLVILN-66YRGNEGTLES..INEKFTSATIVIETGPKKGRRVE6 IQYEDISKLA
g 14804587 Plasmodium vivax		PQIGGMVKIYN-66YRGNEGTLES..INEKFTSATIVIETGPKKGRRVE6 IQYEDISKLA		PQIGGMVKIYN-66YRGNEGTLES..INEKFTSATIVIETGPKKGRRVE6 IQYEDISKLA
g 168070251 Plasmodium berghei		PISSLELLIMENEKKEK..INEKFTSATIVIETGPKKGRRVE6 IQYEDISKLA		PISSLELLIMENEKKEK..INEKFTSATIVIETGPKKGRRVE6 IQYEDISKLA
g 182595924 Plasmodium yoelii		PQIGRIVLILK-6RY..INEKFTSATIVIETGPKKGRRVE6 IQYEDISKLA		PQIGRIVLILK-6RY..INEKFTSATIVIETGPKKGRRVE6 IQYEDISKLA
g 170945060 Plasmodium chabaudi		PAIGKLVKVN-66YRGNEGTLES..INEKFTSATIVIETGPKKGRRVE6 IQYEDISKLA		PAIGKLVKVN-66YRGNEGTLES..INEKFTSATIVIETGPKKGRRVE6 IQYEDISKLA
g 119192424 Coccidioides immitis		PAIGKLVKVN-66YRGNEGTLES..INEKFTSATIVIETGPKKGRRVE6 IQYEDISKLA		PAIGKLVKVN-66YRGNEGTLES..INEKFTSATIVIETGPKKGRRVE6 IQYEDISKLA
g 145343619 Ostreococcus lucimarinus		PAIGKLVKVN-66YRGNEGTLES..INEKFTSATIVIETGPKKGRRVE6 IQYEDISKLA		PAIGKLVKVN-66YRGNEGTLES..INEKFTSATIVIETGPKKGRRVE6 IQYEDISKLA
g 171031985 Therapsid carnivorans		PAIGKLVKVN-66YRGNEGTLES..INEKFTSATIVIETGPKKGRRVE6 IQYEDISKLA		PAIGKLVKVN-66YRGNEGTLES..INEKFTSATIVIETGPKKGRRVE6 IQYEDISKLA
g 184995764 Therapsid carnivorans		PAIGKLVKVN-66YRGNEGTLES..INEKFTSATIVIETGPKKGRRVE6 IQYEDISKLA		PAIGKLVKVN-66YRGNEGTLES..INEKFTSATIVIETGPKKGRRVE6 IQYEDISKLA
g 118374164 Tetrahymena thermophila		PAIGKLVKVN-66YRGNEGTLES..INEKFTSATIVIETGPKKGRRVE6 IQYEDISKLA		PAIGKLVKVN-66YRGNEGTLES..INEKFTSATIVIETGPKKGRRVE6 IQYEDISKLA
g 146126029 Glomerella zeae		PAIGKLVKVN-66YRGNEGTLES..INEKFTSATIVIETGPKKGRRVE6 IQYEDISKLA		PAIGKLVKVN-66YRGNEGTLES..INEKFTSATIVIETGPKKGRRVE6 IQYEDISKLA
g 167592361 Cryptosporidium hominis		PAIGKLVKVN-66YRGNEGTLES..INEKFTSATIVIETGPKKGRRVE6 IQYEDISKLA		PAIGKLVKVN-66YRGNEGTLES..INEKFTSATIVIETGPKKGRRVE6 IQYEDISKLA
g 166362966 Cryptosporidium parvum		PAIGKLVKVN-66YRGNEGTLES..INEKFTSATIVIETGPKKGRRVE6 IQYEDISKLA		PAIGKLVKVN-66YRGNEGTLES..INEKFTSATIVIETGPKKGRRVE6 IQYEDISKLA
g 167483160 Entamoeba histolytica		PAIGKLVKVN-66YRGNEGTLES..INEKFTSATIVIETGPKKGRRVE6 IQYEDISKLA		PAIGKLVKVN-66YRGNEGTLES..INEKFTSATIVIETGPKKGRRVE6 IQYEDISKLA
g 123975648 Trichomonas vaginalis		PAIGKLVKVN-66YRGNEGTLES..INEKFTSATIVIETGPKKGRRVE6 IQYEDISKLA		PAIGKLVKVN-66YRGNEGTLES..INEKFTSATIVIETGPKKGRRVE6 IQYEDISKLA
g 17004666 Lactaria garyi		PAIGKLVKVN-66YRGNEGTLES..INEKFTSATIVIETGPKKGRRVE6 IQYEDISKLA		PAIGKLVKVN-66YRGNEGTLES..INEKFTSATIVIETGPKKGRRVE6 IQYEDISKLA
g 139578119 Magnaporthe oryzae		PAIGKLVKVN-66YRGNEGTLES..INEKFTSATIVIETGPKKGRRVE6 IQYEDISKLA		PAIGKLVKVN-66YRGNEGTLES..INEKFTSATIVIETGPKKGRRVE6 IQYEDISKLA
g 111064711 Phaeocephala anodonta		PAIGKLVKVN-66YRGNEGTLES..INEKFTSATIVIETGPKKGRRVE6 IQYEDISKLA		PAIGKLVKVN-66YRGNEGTLES..INEKFTSATIVIETGPKKGRRVE6 IQYEDISKLA
g 185096338 Neurospora crassa		PAIGKLVKVN-66YRGNEGTLES..INEKFTSATIVIETGPKKGRRVE6 IQYEDISKLA		PAIGKLVKVN-66YRGNEGTLES..INEKFTSATIVIETGPKKGRRVE6 IQYEDISKLA
g 145186098 Ashbya gossypii		PAIGKLVKVN-66YRGNEGTLES..INEKFTSATIVIETGPKKGRRVE6 IQYEDISKLA		PAIGKLVKVN-66YRGNEGTLES..INEKFTSATIVIETGPKKGRRVE6 IQYEDISKLA
g 116198231 Chaetomium globosum		PAIGKLVKVN-66YRGNEGTLES..INEKFTSATIVIETGPKKGRRVE6 IQYEDISKLA		PAIGKLVKVN-66YRGNEGTLES..INEKFTSATIVIETGPKKGRRVE6 IQYEDISKLA
g 183772548 Aspergillus oryzae		PAIGKLVKVN-66YRGNEGTLES..INEKFTSATIVIETGPKKGRRVE6 IQYEDISKLA		PAIGKLVKVN-66YRGNEGTLES..INEKFTSATIVIETGPKKGRRVE6 IQYEDISKLA
g 145228665 Aspergillus niger		PAIGKLVKVN-66YRGNEGTLES..INEKFTSATIVIETGPKKGRRVE6 IQYEDISKLA		PAIGKLVKVN-66YRGNEGTLES..INEKFTSATIVIETGPKKGRRVE6 IQYEDISKLA
g 167904746 Aspergillus nidulans		PAIGKLVKVN-66YRGNEGTLES..INEKFTSATIVIETGPKKGRRVE6 IQYEDISKLA		PAIGKLVKVN-66YRGNEGTLES..INEKFTSATIVIETGPKKGRRVE6 IQYEDISKLA
g 121712092 Aspergillus solanatus		PAIGKLVKVN-66YRGNEGTLES..INEKFTSATIVIETGPKKGRRVE6 IQYEDISKLA		PAIGKLVKVN-66YRGNEGTLES..INEKFTSATIVIETGPKKGRRVE6 IQYEDISKLA
g 170986114 Aspergillus fumigatus		PAIGKLVKVN-66YRGNEGTLES..INEKFTSATIVIETGPKKGRRVE6 IQYEDISKLA		PAIGKLVKVN-66YRGNEGTLES..INEKFTSATIVIETGPKKGRRVE6 IQYEDISKLA
g 119474029 Microascus furficaria		PAIGKLVKVN-66YRGNEGTLES..INEKFTSATIVIETGPKKGRRVE6 IQYEDISKLA		PAIGKLVKVN-66YRGNEGTLES..INEKFTSATIVIETGPKKGRRVE6 IQYEDISKLA
g 116503387 Coprinopsis striata		PAIGKLVKVN-66YRGNEGTLES..INEKFTSATIVIETGPKKGRRVE6 IQYEDISKLA		PAIGKLVKVN-66YRGNEGTLES..INEKFTSATIVIETGPKKGRRVE6 IQYEDISKLA
g 16324651 RTS2 Saccharomyces cerevisiae		PAIGKLVKVN-66YRGNEGTLES..INEKFTSATIVIETGPKKGRRVE6 IQYEDISKLA		PAIGKLVKVN-66YRGNEGTLES..INEKFTSATIVIETGPKKGRRVE6 IQYEDISKLA
g 19112830 Schizosaccharomyces pombe		PAIGKLVKVN-66YRGNEGTLES..INEKFTSATIVIETGPKKGRRVE6 IQYEDISKLA		PAIGKLVKVN-66YRGNEGTLES..INEKFTSATIVIETGPKKGRRVE6 IQYEDISKLA
g 146421635 Pichiia guilliermondii		PAIGKLVKVN-66YRGNEGTLES..INEKFTSATIVIETGPKKGRRVE6 IQYEDISKLA		PAIGKLVKVN-66YRGNEGTLES..INEKFTSATIVIETGPKKGRRVE6 IQYEDISKLA
g 126139029 Pichia stipitis		PAIGKLVKVN-66YRGNEGTLES..INEKFTSATIVIETGPKKGRRVE6 IQYEDISKLA		PAIGKLVKVN-66YRGNEGTLES..INEKFTSATIVIETGPKKGRRVE6 IQYEDISKLA
g 160292067 Candida glabrata		PAIGKLVKVN-66YRGNEGTLES..INEKFTSATIVIETGPKKGRRVE6 IQYEDISKLA		PAIGKLVKVN-66YRGNEGTLES..INEKFTSATIVIETGPKKGRRVE6 IQYEDISKLA
g 168477343 Candida albicans		PAIGKLVKVN-66YRGNEGTLES..INEKFTSATIVIETGPKKGRRVE6 IQYEDISKLA		PAIGKLVKVN-66YRGNEGTLES..INEKFTSATIVIETGPKKGRRVE6 IQYEDISKLA
g 150302811 Kuyveromyces lactis		PAIGKLVKVN-66YRGNEGTLES..INEKFTSATIVIETGPKKGRRVE6 IQYEDISKLA		PAIGKLVKVN-66YRGNEGTLES..INEKFTSATIVIETGPKKGRRVE6 IQYEDISKLA
g 150421825 Debaryomyces hansenii		PAIGKLVKVN-66YRGNEGTLES..INEKFTSATIVIETGPKKGRRVE6 IQYEDISKLA		PAIGKLVKVN-66YRGNEGTLES..INEKFTSATIVIETGPKKGRRVE6 IQYEDISKLA
g 149235618 Lodderomyces longisporus		PAIGKLVKVN-66YRGNEGTLES..INEKFTSATIVIETGPKKGRRVE6 IQYEDISKLA		PAIGKLVKVN-66YRGNEGTLES..INEKFTSATIVIETGPKKGRRVE6 IQYEDISKLA
g 154284129 Yellowomyces apiculatus		PAIGKLVKVN-66YRGNEGTLES..INEKFTSATIVIETGPKKGRRVE6 IQYEDISKLA		PAIGKLVKVN-66YRGNEGTLES..INEKFTSATIVIETGPKKGRRVE6 IQYEDISKLA
g 158261638 Cryptococcus neoformans		PAIGKLVKVN-66YRGNEGTLES..INEKFTSATIVIETGPKKGRRVE6 IQYEDISKLA		PAIGKLVKVN-66YRGNEGTLES..INEKFTSATIVIETGPKKGRRVE6 IQYEDISKLA
g 150555405 Yarrowia lipolytica		PAIGKLVKVN-66YRGNEGTLES..INEKFTSATIVIETGPKKGRRVE6 IQYEDISKLA		PAIGKLVKVN-66YRGNEGTLES..INEKFTSATIVIETGPKKGRRVE6 IQYEDISKLA
g 1154691124 Soleniasia leuconora		PAIGKLVKVN-66YRGNEGTLES..INEKFTSATIVIETGPKKGRRVE6 IQYEDISKLA		PAIGKLVKVN-66YRGNEGTLES..INEKFTSATIVIETGPKKGRRVE6 IQYEDISKLA
g 1154311507 Botryotinia fuckeliana		PAIGKLVKVN-66YRGNEGTLES..INEKFTSATIVIETGPKKGRRVE6 IQYEDISKLA		PAIGKLVKVN-66YRGNEGTLES..INEKFTSATIVIETGPKKGRRVE6 IQYEDISKLA

## Annexe 2

Récapitulatif des 26 conditions de cristallisation du domaine C-terminal de la protéine hKIN17 dans lesquelles des cristaux, des microcristaux, des aiguilles ou des précipités cristallins ont été observés. Les gouttes de cristallisation ont été préparées en robot de cristallisation (Tecan) à partir des cribles Natrix (Hampton Research) et JBScreen (Jena Biosciences) à 17°C.

Condition de cristallisation	Observation
30% PEG 2000 MME, 0.1M Na Acetate, 0.1M Na MES pH 6.5	un cristal (a)
30% PEG 4000, 0.1M Na MES pH 6.5	un cristal
15% MPD, 0.2M Mg Acetate, 0.1M Na MES pH 6.5	sphérulites (d)
30% MPD, 0.2M Mg Acetate, 0.1M Na MES pH 6.5	très petits cristaux et précipité
30% PEG 8000, 0.2M Ammonium Acetate, 0.01M Mg Acetate, 0.05M Na Cacodylate pH 6.5	petits cristaux
5% PEG 4000, 0.2M Ammonium Acetate, 0.15M Mg Acetate, 0.05M Na HEPES pH 7.0	petits cristaux
5% PEG 8000, 0.1M Ammonium Acetate, 0.02M Mg Chloride, 0.05M Na HEPES pH 7.0	un cristal
10% MPD, 0.05M Ammonium Acetate 0.01M, Mg Chloride, 0.05M Tris HCl pH 7.5	petits cristaux (c)
25% PEG 6000, 0.1M Li Chloride, 0.1M Na HEPES pH 7.5	un cristal
30% MPD, 10% PEG 4000, 0.1M Imidazole HCl pH 8.0	des cristaux et un précipité cristallin
30% PEG 3000, 0.2M Li Sulfate, 0.1M TRIS HCl pH 8.5	aiguilles (e)
30% PEG 4000, 0.2 M Li Sulfate, 0.1 M TRIS HCl pH 8.5	aiguilles
25% PEG 5000 MME, 0.2M Li Sulfate, 0.1M TRIS HCl pH 8.5	aiguilles
30% PEG 4000, 0.2M Na Acetate, 0.1M TRIS HCl pH 8.5	aiguilles
32% PEG 4000, 0.8M Li Chloride 0.1M TRIS HCl pH 8.5	aiguilles, oursins (f)
28% PEG 6000, 0.5 M Li Chloride, 0.1M TRIS HCl pH 8.5	aiguilles
30% PEG 4000, 0.2M Mg Chloride, 0.1M TRIS HCl pH 8.5	aiguilles
17% PEG 20000, 0.1M Mg Chloride, 0.1M TRIS HCl pH 8.5	aiguilles, oursins et peut-être cristaux
12% MPD, 0.05M Mg Chloride, 0.1M TRIS HCl pH 8.5	sphérulites
30% PEG 400, 0.1M K Chloride, 0.01M Mg Chloride, 0.05M Tris HCl pH 8.5	petits cristaux et précipité
1.5M Li Sulfate, 0.1M TRIS HCl pH 8.5	sphérulites
2.0M Mg Chloride, 0.1M TRIS HCl pH 8.5	sphérulites, plaques
1.8M Ammonium Sulfate, 0.025M Mg Sulfate, 0.05M Tris HCl pH8.5	un cristal allongé
25% PEG 4000, 0.1M Na Acetate, 8% iso-Propanol	un cristal
30% PEG 6000, 0.1M Na Acetate, 1M Li Chloride	des cristaux et un précipité (b)
25% PEG 8000, 0.2M Li Chloride	un cristal et un précipité

### Annexe 3

Résultats des identifications par spectrométrie de masse (MALDI-TOF) des bandes de protéines découpées sur les gels présentés en Figure XXX. L'analyse des spectres de masse a été réalisée à l'aide des logiciels Mascott et Profound.

bande de protéine	masse sur gel (kDa)	identification (GeneBank)	masse (kDa)	% recouvrement, Z score
13 GST-K1	250	CAD gi:33989959	245	25,6
14 GST-K1	140	tight junction protein 2 gi:2135255	125	8, 1.38
15 GST-K1	130	pas d'identification claire		
16 GST-K1	75	GST	32.3	31, 1.63
		HsKIN17	45.8	20, 0.88
17 GST-K1	60	GST gi:4929901	32.3	40
18 GST-K1	75	eukaryotic translation elongation factor 1 (EF1a) gi:15530265	50.5	18, 1.16
		GST	32.3	35, 0.72
20 GST <i>contrôle 18</i>	75	eukaryotic translation elongation factor 1 (EF1a)	50.5	11, 1.10

bande de protéine	masse sur gel (kDa)	identification (GeneBank)	masse (kDa)	% recouvrement, Z score
A GST-K2	80	keratin		
C GST-K2	42	gamma-actin gi:7441428	41.99	53, 2.41
E GST <i>contrôle C</i>	42	ACTB protein (actin) gi:15277503	40.54	31, 2.37
L GST-K2	45	eIF-4A I gi:4503529	46.36	20, 2.13

bande de protéine	masse sur gel (kDa)	résultats Profound ou Mascot	masse (kDa)	% recouvrement, Z score
50 ZZ-K1	10	ubiquitin gi:13786827	8.42	80, 2.36
51 ZZ-K1	12	macrophage migration inhibition factor gi:5542151	12.50	21, 1.81
52 ZZ-K1	14	small nuclear ribonucleoprotein D2 gi:4759158	13.62	36, 0.53
53 ZZ-K1	13	profilin gi:3891601	15.01	54, 2.33

55 ZZ-K1	16	cyclophilin gi:1633054	18.09	55, 2.31
57 ZZ <i>contrôle 55</i>	16	cyclophilin	18.09	
59 ZZ-K1	18	cofilin gi :5031635	18.71	42, 2.14
60 ZZ-K1	24	Ran protein (GTP binding protein) gi:5107682	24.58	39, 1.25
61 ZZ-K1	23	Ran protein (GTP binding protein)	23.3	42, 2.41
62 ZZ-K1	26	immunoglobulin (ZZ) gi :21669481	29.30	25, 0.05
63 ZZ-K1	50	pyruvate kinase gi :20178296	58.49	36, 2.13
64 ZZ-K1	80	HSP90 gi :72222	83.62	38, 1.48
67 ZZ-K1	250	actin-binding protein gi:3282771	280.34	5, 1.58

bande de protéine	masse sur gel (kDa)	résultats ProFound ou Mascot	masse (kDa)	% recouvrement, Z score
70 GST-K2	70	p68 protein gi:226021	67.42	24, 1.22
76 GST-K2	55	NR54 protein gi :13124797	54.20	40, 2.34
77 GST <i>contrôle 76</i>	55	NR54 protein	54.20	31, 2.39
79 GST-K2	100	PTB-associated splicing factor gi :4826998	76.25	31, 1.34
		coatomer protein gi :11559929	99.01	32, 1.20
		DDX21 protein gi :31455230	87.84	24
80 GST <i>contrôle 79</i>	100	PTB-associated splicing factor	76.25	18, 2.33



**RESUME :** Très conservée, plus abondante dans les cellules en prolifération que dans les cellules normales, stimulée par les rayonnements UV et ionisants, autant de caractéristiques qui ont suscité notre intérêt pour l'analyse structurale et fonctionnelle de la protéine KIN17 humaine (hKIN17) dans le but de caractériser sa ou ses fonction(s) dans la cellule. Plusieurs propriétés ont été décrites pour la protéine hKIN17 dans différents mécanismes cellulaires tels que la réplication de l'ADN, la réparation de l'ADN et le métabolisme de l'ARN. Pendant ma thèse, j'ai montré en combinant des données de DC, RMN et SAXS que la protéine hKIN17 possède un domaine structuré adoptant probablement le repliement d'un doigt de zinc et un domaine *winged helix* dans sa région N-terminale. Seul le premier domaine lie les acides nucléiques, le deuxième domaine ayant pour partenaires plusieurs hélicases à ARN impliquées dans la transcription et la traduction. J'ai résolu par cristallographie aux rayons X la structure du domaine C-terminal qui est retrouvé uniquement chez les eucaryotes supérieurs et se replie en un double domaine de type SH3. J'ai montré par différentes méthodes biochimiques (filtration sur gel, gel IEF natif) que ce domaine ancre la protéine hKIN17 à un complexe nucléaire acide de haut poids moléculaire dont les composants sont en cours d'identification par spectrométrie de masse. De plus, il lie l'ARN et les histones modifiées. L'ensemble de ces résultats confirment l'implication de la protéine hKIN17 dans le métabolisme de l'ARN et précisent le rôle de chacun des domaines au sein de la protéine.

**MOTS-CLES :** KIN17, biologie structurale, recherche de partenaires protéiques, cristallographie aux rayons X, résonance magnétique nucléaire, diffusion des rayons X aux petits angles, ARN, complexes protéiques.

**TITLE:** Structural and functional characterization of the human KIN17 protein.

**ABSTRACT:** Human KIN17 (hKIN17) is highly conserved, more abundant in proliferating cells than in quiescent cells and overexpressed after UV irradiation and ionizing radiation. All these features raised our interest for its structural and functional analysis at the molecular level in order to describe its functions at the cellular level. Several properties have been described for the hKIN17 protein in different cellular mechanisms such as DNA replication, DNA repair and RNA metabolism. During my project, I showed by combining data from CD, NMR and SAXS that hKIN17 has a structured domain probably exhibiting a zinc finger fold and a winged helix domain in its N-terminus region. Only the first domain binds to nucleic acids. The second domain binds to several RNA helicases involved in transcription and translation. Using X-ray crystallography, I solved the structure of the C-terminal domain which is found only in higher eukaryotes and folds into a double SH3-like domain. I showed using different biochemical methods (molecular filtration, non denaturing IEF) that this domain anchors the hKIN17 protein in a nuclear and acidic complex of high molecular weight, whose components are currently identified by mass spectrometry. Moreover, this domain binds to RNA and modified histones. Taken together, these results confirm the involvement of hKIN17 in RNA metabolism and clarify the role of each domain within the protein.

**KEY WORDS:** KIN17, structural biology, isolation of interacting proteins, X-ray crystallography, nuclear magnetic resonance, small angle X-ray scattering, RNA, protein complexes.

