

Thèse  
présentée pour obtenir le grade de  
Docteur de l'Université Louis Pasteur  
Strasbourg I

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

par Muriel Uhring

**Contribution à l'étude structure/fonction  
du facteur de transcription TFIID**

Soutenue publiquement le 16 décembre 2004

**Membres du Jury**

Directeur de thèse : Dr Dino MORAS  
Rapporteur interne : Prof Etienne WEISS  
Rapporteur externe : Dr Eric LE CAM  
Rapporteur externe : Prof Yves MECHULAM  
Examineurs : Dr Arnaud POTERSZMAN  
Dr Patrick SCHULTZ

*à mes parents,  
à Eric,*

## *Remerciements*

*Tout d'abord je veux remercier les membres du jury d'avoir bien voulu accepter d'évaluer ce travail de thèse : le Professeur Etienne Weiss, le Docteur Eric Le Cam ainsi que le Docteur Yves Mechulam.*

*Je remercie mon directeur de thèse, le Docteur Dino Moras, pour m'avoir donné l'opportunité d'effectuer ma thèse dans un environnement scientifique et technique de qualité au sein de son laboratoire.*

*Je remercie particulièrement le Docteur Arnaud Poterszman qui, tout au long de ces années, a su me guider dans mes choix, en toute liberté et dans la bonne humeur.*

*Je remercie vivement le Docteur Jean-Marc Egly et toute son équipe pour le partage et l'échange de connaissances.*

*Je remercie sincèrement le Docteur Patrick Schultz pour m'avoir fait découvrir avec plaisir la microscopie électronique. Mille mercis à Corinne de m'avoir initié avec joie aux techniques microscopiques, à Christine pour son aide précieuse à tout moment et à toute épreuve, à Véronique pour ces belles discussions partagées, merci à toutes pour votre aide. Merci à Bruno et Sacha notre bien-aimé suisse-américano.*

*Je tiens à remercier toutes les personnes des équipes communes pour leur disponibilité, leur expertise et leur gentillesse, et particulièrement Jean-Luc et Isabelle (la culture de cellules), Serge (les séquences), Pascal (la synthèse de peptides), Edouard et Pascal (la synthèse d'oligonucléotides), Noëlle et Hélène (la spectrométrie de masse), Jean-Marie (le clonage), Maité et Hélène (la bibliothèque).*

*Je remercie chaudement tous les acteurs de l'équipe TFIHH pour avoir instauré une sincère camaraderie : Florence, Wassim et Marc (et pour tous les cafés partagés !)*

*Merci à toute l'équipe du laboratoire, pour leur aide de tous les jours : Alain, Alexandra, Anne-Catherine, André, Bénédicte, Catherine, Christiane, Christophe, David et*

*Lydia, Denis, Didier, Gilbert, Isabelle, Jean, Fabrice, Luc, Marc, Natacha, Odile, Olivier, Pierre, Sylvia, Sylvie, Raymond, Thierry, Thomas, Vincent C., Vincent D., Yann.*

*Merci à notre équipe voisine, James, Cyril, Liliane (pour son écoute et sa présence), Georges (en retraite joyeuse et active), Renata (aux nombreuses qualités, culinaires, et autres).*

*Merci aux personnes qui ont corrigé ce manuscrit ou qui ont participé, même indirectement, à son avancement.*

*De manière générale, je remercie tout le personnel de l'IGBMC qui regorge de personnalités talentueuses et enthousiasmantes.*

*Inciter les jeunes filles à s'engager dans des carrières scientifiques et techniques, promouvoir l'image de la science chez les femmes et l'image des femmes dans les sciences, c'est à travers l'association « Femmes et Sciences » que je m'y suis engagée.*

*Merci à l'Association de la Recherche contre le Cancer qui m'a en partie financée au cours de mes études doctorales.*

*Merci à tous mes amis, qui m'ont entourée pendant la rédaction de ce manuscrit, bercée par mes airs d'opéra préférés. Merci à mes amis d'enfance, Nicolas, Christian, Jean-David, Vali et Christian. Merci à Brat, Sandrine et Véronique. Merci à Isabelle, Grace et Sylvie. Pour répondre à leurs curiosités scientifiques, ils trouveront en fin de manuscrit un aperçu plus éthéré...*

*Merci de tout cœur à ma famille, à mon frère, Frédéric et à ma belle sœur, Betty, à Adalric et Louise, à mes beaux-parents Yolande et Jean-Marie. Merci à Raymond.*

*Et je veux plus que jamais remercier ma tendre moitié, Eric, pour sa patience, son soutien continu et avec qui je me réjouis de partager ma vie.*

*Comment remercier comme ils le méritent mes chers parents, sans qui rien ne serait...  
« La bonté de mon père est plus haute que la montagne, la bonté de ma mère plus profonde que l'océan... »*

La diversité se trouve à la racine même de la biologie. Les gènes, qui constituent le patrimoine de l'espèce, s'associent et se séparent au fil des générations, formant ces combinaisons toujours différentes et toujours fugitives que sont les individus. C'est cette combinaison infinie des gènes qui rend chacun de nous unique. C'est elle qui donne à l'espèce sa richesse et sa variété.

François Jacob -Extraits-

## *Avant Propos*

Le génome humain, composé d'environ  $3 \times 10^9$  paires de bases (pb), comprendrait quelque 30,000 gènes répartis plus ou moins uniformément le long des 48 chromosomes (Gruetzner et al., 1999), (Venter et al., 2001), (Lander et al., 2001). D'autres estimations soutiennent que le génome humain est plutôt composé de 50,000 à 120,000 gènes (Aparicio, 2000), (Liang et al., 2000)). Au-delà des chiffres, une interrogation fondamentale demeure : quels sont les mécanismes qui orchestrent l'expression de tous ces gènes? La quête de la réponse à cette question a contribué à l'essor et au développement de la biologie cellulaire, moléculaire et structurale au milieu du vingtième siècle et de nombreuses équipes cherchent aujourd'hui à élucider les mystères de l'expression génique (Broder et Venter, 2000), (Bentley, 2000b).

Au cours du développement de l'embryon humain, les cellules, contenant pourtant la même information génétique, se spécialisent et se multiplient pour former divers organes et tissus. Comme certains gènes remplissent une fonction très spécifique, il est impératif qu'ils soient exprimés aux moments opportuns et dans les tissus appropriés. Sans régulation de la transcription, l'expression anarchique des gènes nuirait vraisemblablement au maintien du métabolisme cellulaire normal et conséquemment au développement et à la survie de l'organisme lui-même. La régulation de l'expression des gènes apparaît donc comme un phénomène indispensable à la vie. Les cellules sont ainsi pourvues de divers mécanismes complexes qui assurent l'expression ordonnée de chacun de leurs gènes. Tout dérèglement de l'un de ces mécanismes peut avoir des conséquences dramatiques non seulement pour la cellule mais également pour l'organisme entier. Les médecins et les biologistes du développement fondent beaucoup d'espoir sur la contribution que pourrait apporter la compréhension de l'expression différentielle des gènes à la lutte contre certaines maladies congénitales, au vieillissement et au cancer (Bentley, 2000a), (Warren et al., 2002).

Ainsi, les différentes étapes menant d'un gène à son produit, et plus particulièrement la transcription, sont finement contrôlées. La cellule doit également préserver l'intégrité de son génome. Pour cela, elle a développé différents mécanismes qui permettent de détecter et de réparer les nombreuses lésions occasionnées à l'ADN par des agents exogènes (rayonnements ionisants et ultraviolets, agents cancérigènes et autres produits chimiques), mais aussi par des agents endogènes (radicaux libres, instabilité chimique des bases). Ces deux processus fondamentaux, transcription et réparation de l'ADN, agissent de manière coordonnée. C'est ainsi qu'un complexe multi protéique se trouve à l'interface des deux mécanismes cellulaires : c'est le facteur TFIIH. Ce complexe de 10 sous-unités a été initialement associé à la transcription des gènes de classe II (Lu et al., 1992) et le clonage des différentes sous-unités qui le composent a révélé une forte relation avec la réparation de l'ADN par NER (Nucleotide Excision Repair) (Wang et al., 1994b), (Schaeffer et al., 1994; Hwang et al., 1996b).

L'objet de mes études doctorales a été d'étudier les relations qui existent entre la structure et la fonction du facteur multi protéique TFIIH portant les activités enzymatiques hélicase et kinase. Du fait de sa grande taille et de sa complexité, diverses approches ont été utilisées pour cette étude. Tout d'abord, l'architecture du facteur entier a été abordée en combinant microscopie électronique et analyse biochimique des interactions protéines-protéines. Ensuite, des études qui relèvent de la génomique structurale, ont visé à déterminer la structure à l'échelle atomique d'un homologue de l'hélicase humaine de TFIIH chez une archaebactérie. Finalement, des travaux combinant données structurales et fonctionnelles, ont eu pour but de comprendre comment TFIIH s'intègre dans des complexes physiologiques.

## Abréviations

**ADN** acide désoxyribonucléique

**ADNss** acide désoxyribonucléique simple brin

**ATP** adénosine tri-phosphate

**ARN** acide ribonucléique

**ARNm** acide ribonucléique messenger

**BER** réparation par excision de base

**CAK** kinase activant les Cdk (*Cdk-activating kinase*)

**Cdk** kinase dépendante des cyclines (*cyclin-dependent kinase*)

**CS** syndrome de Cockayne

**CTD** domaine carboxy-terminal de la grande sous-unité de l'ARN polymérase II  
(*carboxy-terminal domain*)

**3D** trois dimensions

**FTC** fonction de transfert de contraste

**GST** glutathione S transférase

**HCV** virus de l'hépatite C

**MAT1** *ménage à trois 1*

**NER** réparation par excision resynthèse des nucléotides

**pb** paire de base

**PEG** polyéthylène glycol

**TAF<sub>II</sub>** facteur de classe II associé à la TBP (*TBP associated factor*)

**TBP** protéine de liaison à la boîte TATA (*TATA binding protein*)

**TFII** facteur de transcription de classe II

**TTD** trichotiodystrophie

**SDS** sodium dodecyl sulfate

**Sf9** *Spodoptera frugiperda*

**ss** *Sulfolobus solfataricus*

**TF** transformée de Fourier

**UV** rayonnement ultraviolet

**XP** Xeroderma Pigmentosum



# TABLE DES MATIERES

## **Chapitre I : INTRODUCTION**

A. <u>Mise en situation</u> .....	1
1. La chromatine.....	1
2. Le promoteur.....	4
B. <u>La machinerie transcriptionnelle de base</u> .....	6
1. Le complexe de préinitiation de la transcription.....	6
2. Assemblage du complexe de préinitiation.....	8
3. Données structurales sur la transcription.....	10
C. <u>La réparation de l'ADN</u> .....	13
1. La réparation par excision resynthèse des nucléotides (NER).....	13
2. Les étapes du NER.....	14
3. Données structurales sur la réparation de l'ADN par NER.....	16
D. <u>Le facteur de transcription TFIIH</u> .....	17
1. Présentation du facteur TFIIH.....	17
2. Les sous-unités de TFIIH.....	19
2.1 Les sous-unités du core.....	19
2.2. Les sous-unités du CAK.....	21
2.3. La dixième sous-unité TTD-A/p8.....	23
3. Données structurales sur TFIIH.....	24
4. TFIIH, un facteur aux interacteurs multiples.....	25
5. TFIIH à l'interface entre réparation et transcription de l'ADN.....	28
E. <u>Synthèse Médecine et Sciences</u> « Les hélicases et les maladies associées : Quand la double hélice ne s'ouvre plus ».....	29

## Chapitre II : ORGANISATION QUATERNAIRE DE TFIIH

### I. La microscopie électronique

A. <u>Généralités sur la microscopie</u> .....	32
1. Coloration négative.....	32
2. Cryomicroscopie.....	34
3. Cryocoloration négative.....	36
4. Coloration négative à froid.....	37
5. Cristaux bidimensionnels.....	37
B. <u>L'analyse d'image de molécules isolées</u> .....	38
1. Pré-traitement des images .....	39
2. Alignement en rotation et en translation.....	39
3. Partitionnement des images.....	40
3.1. L'analyse factorielle des correspondances.....	40
3.2. La classification hiérarchique ascendante.....	41
4. Alignement multi-références.....	41
C. <u>La reconstruction tridimensionnelle d'un objet à partir de ses projections</u> .....	43
1. Reconstruction cylindrique moyenne.....	45
2. Méthode conique aléatoire de collection des données.....	45
3. Méthode des lignes communes.....	46
D. <u>Limitations du microscope et fonction de transfert de contraste</u> .....	47
E. <u>Mesure de la résolution des reconstructions</u> .....	50

### II. Etude du facteur TFIIH par microscopie électronique

A. <u>Objectifs du travail</u> .....	51
B. <u>Modèle de TFIIH en microscopie par coloration négative à froid</u> .....	52

1. La coloration négative à froid.....	52
2. Sélection des particules.....	54
3. L'analyse d'images.....	55
<b>C. <u>Purification et caractérisation du complexe core-TFIIH recombinant</u>.....</b>	<b>61</b>
1. Purification du complexe recombinant core-TFIIH dans les cellules d'insectes.....	61
2. Test d'activité.....	63
3. Caractérisation par spectrométrie de masse.....	66
4. Structure en microscopie électronique.....	69

### **III. Organisation moléculaire du facteur de transcription TFIIH**

A. <u>Objectif de l'étude</u> .....	72
B. <u>Modèle général sur l'architecture de TFIIH</u> .....	72

### **Chapitre III : L'HELICASE XPD : dissection de la protéine humaine, vers la structure de son homologue archaebactérien**

A. <u>Généralités sur les hélicases</u> .....	74
1. Les familles d'hélicases, conservation des motifs.....	75
2. Mesure de l'activité déroulante.....	76
3. Mesure de l'activité ATPase.....	80
4. Mécanisme d'action.....	80
5. Déplacement de protéines... ..	81
B. <u>L'hélicase XPD humaine</u> .....	82
1. L'hélicase XPD du complexe TFIIH.....	82
2. Dissection de l'hélicase humaine XPD dans le système d'expression baculovirus.....	84
3. Optimisation de la production du fragment C-terminal de XPD .....	86
4. Coexpression de XPD avec ses partenaires.....	89
C. <u>L'hélicase XPD d'archaebactérie</u> .....	93

1. Généralités sur les archaebactéries.....	93
2. Les protéines de la chromatine chez les archaebactéries.....	95
3. La transcription chez les archaebactéries.....	97
4. Régulation de la transcription.....	100
5. La réparation par excision resynthèse des nucléotides chez les archaebactéries.....	101
<b>D. Etude structurale de l'homologue de l'hélicase XPD chez <i>Sulfolobus solfataricus</i>.....</b>	<b>105</b>
1. Motifs conservés.....	105
2. Purification de l'hélicase putative XPD de <i>Sulfolobus solfataricus</i> .....	107
3. Caractérisations biophysiques.....	110
4. Cristallisation de l'hélicase putative XPDss.....	111
4.1. Généralités sur la cristallisation.....	111
4.2. La cristallisation par la méthode de diffusion en phase vapeur.....	112
4.3. Enregistrement des données.....	114
4.4. Le problème de la phase en cristallographie.....	115
4.5. Cristallisation de la putative hélicase XPDss.....	116

#### **Chapitre IV : TFIIH en complexes avec ses partenaires physiologiques**

<b>Article 2</b> : Immobilization of biotinylated DNA on 2-D streptavidine crystals.....	120
--	-----

<b>Article 3</b> : Molecular organization and oligomeric state of TFIIH.....	126
--	-----

#### **Chapitre V : Production et caractérisation des complexes recombinants**

<b>Article 4</b> : Expression of FLAG Fusion Proteins in Insect Cells : Application to the Multi-Subunit Transcription/DNA Repair Factor TFIIH.....	127
---	-----

<b>Article 5</b> : Influence of affinity tags for protein production in insect cells .....	128
--	-----

<b>Chapitre VI : Discussion.....</b>	<b>129</b>
--------------------------------------	------------

<b>Chapitre VII : Matériel et Méthodes.....</b>	<b>143</b>
---	------------

A. <u>Biologie moléculaire</u> .....	143
1. Matériel.....	143
1.1. Souches bactériennes.....	143
1.2. Vecteurs de clonage et d'expression.....	144
2. Méthodes.....	144
2.1. Amplification par PCR « Polymerase Chain Reaction ».....	144
2.2. Mutagenèse dirigée.....	146
2.3. Electrophorèse en gel d'agarose.....	148
2.4. Estimation de la quantité d'ADN.....	149
2.5. Précipitation des ADN.....	149
B. <u>Clonage</u> .....	150
1. Stratégies de clonage.....	150
1.1. Digestion de l'ADN par des enzymes des restriction.....	150
1.2. Déphosphorylation des extrémités 5' phosphate du vecteur.....	150
1.3. Extraction au phénol-chloroforme.....	151
1.4. Purification et extraction de fragment PCR.....	151
1.5. Ligation.....	152
1.6. Transformation à partir du mélange de ligation.....	153
1.7. Obtention du vecteur de clonage et vérification de la séquence.....	153
2. Amplification des vecteurs et préparation d'ADN plasmidique.....	154
2.1. maxipreparation d'ADN plasmidique sur colonne Nucleobond .....	154
2.2. maxipreparation d'ADN plasmidique par double gradient de césium.....	155
3. Milieux de culture.....	156
4. Transformation des bactéries.....	157
4.1. Méthode chimique.....	157
4.1.1. Préparation des bactéries chimio-compétentes.....	157
4.1.2. Transformation chimique.....	158

4.2. Méthode par électroporation.....	158
4.2.1. Préparation des bactéries électro-compétentes.....	158
4.2.2. Transformation par électroporation.....	159
5. La technologie Gateway.....	160
5.1. La recombinaison.....	160
5.2. Les enzymes recombinantes.....	160
5.3. Les réactions de recombinaison.....	161
C. <u>Systemes d'expression</u> .....	162
1. Systeme d'expression dans <i>Escherichia coli</i> .....	162
2. Systeme d'expression dans les cellules d'insectes infectées par un baculovirus.....	162
D. <u>Biochimie</u> .....	166
Cultures et production de protéines dans <i>Escherichia coli</i> .....	166
Cultures et production de protéines dans les cellules d'insectes.....	167
1. Purification.....	167
1.1. Sonication.....	167
1.2. Ultracentrifugation.....	168
1.3. Chromatographie d'affinité sur résine de chélation de cations métalliques.....	168
1.4. Protéolyse ménagée par la thrombine ou par la TEV du peptide hexahistidine...	168
1.5. Héparine.....	169
1.6. Autres protéines de fusion .....	169
1.7. Gel filtration.....	170
1.8. Immunopurification par le système FLAG.....	172
1.9. Interaction de protéines recombinantes.....	173
2. Caractérisation des protéines purifiées.....	176

2.1. Electrophorèse sur gel de polyacrylamide en conditions dénaturantes.....	176
2.2. Révélation au bleu de Coomassie.....	177
2.3. Révélation au nitrate d'argent.....	178
2.4. Western Blot .....	179
3. Dosage des protéines.....	180
3.1. Spectrométrie de masse.....	181
3.2. Digestion trypsique.....	181
3.3. Microséquencage.....	181
4. Tests d'activité.....	181
4.1. Marquage radioactif d'un oligonucléotide au [ <sup>-32</sup> P] ATP.....	181
4.2. Hybridation des oligonucléotides et réaction de déroulement de brins.....	183
4.3. Test hélicase.....	184
4.4. Test de transcription <i>in vitro</i> .....	186
<b>Bibliographie</b> .....	189
<b>Digression</b> "Connaître la structure d'une molécule pour comprendre sa fonction".....	230

# *Index des Figures et Tableaux*

## **Chapitre I : Introduction**

---

- Figure 1 : Activation de la chromatine
- Figure 2 : La décondensation de la chromatine rend les différentes séquences d'ADN régulatrices accessibles aux facteurs de transcription
- Figure 3 : Deux modèles pour l'assemblage du complexe de préinitiation de la transcription
- Figure 4 : Représentation schématique des structures de RAP30 et RPB14.4
- Figure 5 : Structure de la RNA polymérase II de levure à 2.6Å
- Figure 6 : La réparation par excision resynthèse des nucléotides (NER) chez les eucaryotes
- Figure 7 : Représentation schématique de la structure de hXPA et yPCNA
- Figure 8 : Représentation des structures de TFIIH résolues à ce jour
- Tableau 1 : Facteurs généraux de transcription humains
- Tableau 2 : Structures de complexes protéiques impliquées dans la transcription
- Tableau 3 : Facteurs de réparation par excision resynthèse de nucléotides
- Tableau 4 : Structures de complexes protéiques impliquées dans la réparation
- Tableau 5 : Composition du facteur TFIIH
- Tableau 6 : Facteurs de transcription interagissant avec TFIIH

## **Chapitre II : Organisation quaternaire de TFIIH**

---

- Figure 9 : Congélation rapide de la grille à l'aide d'une « guillotine »
- Figure 10 : La cryocoloration négative : représentation schématique de la préparation des échantillons
- Figure 11 : L'analyse d'images en deux dimensions
- Figure 12 : Convention des angles d'Euler telle qu'elle est établie dans le logiciel IMAGIC
- Figure 13 : Trois méthodes pour combiner l'information de projection dans une structure 3D
- Figure 14 : Principe de la méthode conique aléatoire de collection de données
- Figure 15 : Les principes des lignes communes
- Figure 16 : Principe de formation de l'image d'un objet
- Figure 17 : Le microscope électronique à canon à effet de champ (FEG) Tecnai F20 200kV
- Figure 18 : Champ de molécules de TFIIH après coloration à froid
- Figure 19 : Extraction des molécules de TFIIH
- Figure 20 : Galeries d'images moyennes de classe de TFIIH observées en coloration à froid
- Figure 21 : Analyse d'image de TFIIH
- Figure 22 : Reconstruction 3D du modèle de TFIIH obtenu en coloration à froid
- Figure 23 : Purification du complexe recombinant core-TFIIH
- Figure 24 : Comparaison de la production de complexes recombinants core-TFIIH



Figure 25 : Comparaison de l'activité des complexes recombinants  
Figure 26 : Principe de la technologie CypherGen  
Figure 27 : Champ de particules de core-TFIID observées en coloration négative  
Figure 28 : Classification des particules de core-TFIID  
Figure 29 : Structure en microscopie électronique du facteur recombinant core-TFIID  
Figure 29A : Modèle général de l'architecture de TFIID (A) Tableau résumant les interactions entre sous-unités (B) Représentation schématique du modèle

### Chapitre III : L'hélicase XPD

---

Figure 30 : Mécanisme d'action d'une hélicase  
Figure 31 : Mesure de l'activité déroulante d'une hélicase  
Figure 32 : Les mutations de l'hélicase XPD  
Figure 33 : Dissection de l'hélicase humaine XPD  
Figure 34 : Comparaison de la production du fragment C-terminal de la protéine XPD  
Figure 35 : Coexpression entre XPD et p44 dans les cellules d'insectes  
Figure 36 : Coexpression entre XPD et MAT1 dans les cellules d'insectes  
Figure 36A : Coexpression entre XPD et p62 dans les cellules d'insectes  
Figure 37 : Arbre phylogénique représentant les trois règnes  
Figure 38 : Le complexe de préinitiation chez les archaebactéries  
Figure 39 : Structure cristallographique du complexe ternaire TBP/TFB/DNA chez les archaebactéries et structure du complexe eucaryote TBP/TFIIB/DNA  
Figure 40 : Les archaebactéries possèdent un système de réparation par excision resynthèse des nucléotides  
Figure 41 : Tableau récapitulatif des motifs hélicases retrouvés chez XPDhs et chez XPD de *Sulfolobus solfataricus*  
Figure 42 : Alignement des hélicases d'archaebactéries  
Figure 43 : Purification de l'hélicase XPDss de *Sulfolobus solfataricus*  
Figure 44 : Diagramme schématique de solubilité  
Figure 45 : Principe de la méthode par « goutte suspendue » et de « la goutte assise »  
Figure 46 : Cristallisation de la putative hélicase XPDss sous sa forme native  
Figure 47 : Observation de molécules de TFIID liées à l'ADN  
Tableau 7 : Tableau récapitulatif des motifs conservés chez les hélicases  
Tableau 8 : Distribution des protéines liant l'ADN et des protéines de la réparation dans les génomes d'archaebactéries séquencés à ce jour

## Chapitre VI : Discussion

---

Figure 48 : Comparaison des structures du core-TFIID de levure et humain

Figure 49 : Modèle d'assemblage du complexe TFIID

Figure 50 : Localisation du facteur TFIIE sur la polymérase

## Chapitre VII : Matériel et Méthodes

---

Figure 51 : Méthode d'«Overlap extension »

Figure 52 : Système d'expression dans *Escherichia coli*

Figure 53 : Système d'expression dans les cellules d'insectes

Figure 54 : Le système d'expression Bac-to-Bac®

Figure 55 : Immunopurification par le système Flag

Figure 56 : Purification de l'oligonucléotide marqué au [  $^{-32}\text{P}$ ]ATP

Figure 57 : Construction et purification de la sonde hélicase

Figure 58 : Test de transcription *in vitro*

# Chapitre I : Introduction

## A. Mise en situation

### 1. La chromatine

Dans le noyau d'une cellule humaine, l'ADN est fractionné en 46 chromosomes et est condensé sous forme de chromatine (Johnson et al., 1998). Ainsi, une molécule d'ADN linéaire longue de 10 centimètres sera compactée en un chromosome mesurant à peine 10 micromètres de long, ce qui implique une condensation d'un facteur 10,000. La condensation de l'ADN en chromatine s'organise de manière séquentielle et ordonnée. En premier lieu, 146 pb d'ADN s'enroulent autour d'un octamère d'histones pour former un nucléosome condensant l'ADN par un facteur de 7 (Roth et al., 1992). Jusqu'à 95 % de l'ADN génomique serait ainsi associé à ces nucléosomes qui constituent le premier niveau d'organisation de la chromatine. Dans un second niveau d'organisation, les nucléosomes se compactent et forment une hélice au rythme de 6 nucléosomes par tour produisant une fibre de 30 nanomètres d'épaisseur. Cette fibre est finalement condensée en euchromatine (condensation légère) ou en hétérochromatine (condensation prononcée) constituant le troisième et dernier niveau d'organisation.

Il semble que seuls les gènes localisés dans l'euchromatine peuvent être potentiellement transcrits (Smale, 1997). Cette structure organisationnelle de l'ADN dans le noyau constituerait donc en elle-même un mécanisme de répression de la transcription des gènes. Pour activer la transcription d'un gène donné dans une cellule vivante, la chromatine comprise dans la région de contrôle du gène se doit d'être modifiée ou altérée de façon à être permissive à la transcription (Cobb et al., 2000). De ce fait, il est essentiel que les structures complexes de la chromatine formées par le regroupement de réseaux de nucléosomes soient supprimées. De plus, les nucléosomes spécifiques qui recouvrent les séquences régulatrices doivent être accessibles aux protéines régulatrices. Finalement, les nucléosomes situés dans la séquence transcrite doivent permettre le passage de l'ARN polymérase.

Bien que les mécanismes demeurent en partie incompris, cette réorganisation locale de la chromatine et la modification des nucléosomes entraînent l'accroissement de l'accessibilité des facteurs de transcription et de la machinerie générale de transcription aux séquences régulatrices du gène (Davie et Chadee, 1998).

L'acétylation des histones est l'un des mécanismes impliqué dans la régulation de la chromatine (Eberharter et Becker, 2002). Naturellement porteuses d'une charge nette positive, les histones possèdent une forte affinité pour l'ADN qui est chargé négativement. Le positionnement précis d'un nucléosome aura comme conséquence de bloquer l'accès des facteurs de transcription aux séquences régulatrices. Par contre, l'acétylation des histones viendra réduire l'affinité de ces dernières pour l'ADN en enlevant la charge nette positive des histones, facilitant l'assemblage des facteurs de transcription à l'ADN. Donc de façon générale, l'hyperacétylation est associée à la chromatine qui est permissive à la transcription alors que l'hypoacétylation est associée à des régions du génome où il y a peu de transcription (Felsenfeld et al., 2000). L'ajout ou la perte d'un groupement acétyle sur les résidus lysine des histones H3 et H4 s'effectue par l'entremise des acétylases et déacétylases, respectivement.

Récemment, un lien entre la méthylation de l'ADN et la déacétylation des histones a été fait. En effet, la présence d'un groupement méthyle sur les résidus cytosines des CpG (cytosine-guanine) présents dans la séquence du gène est généralement associée à une absence de transcription et l'absence de méthylation est associée avec la possibilité d'être transcrit (Razin, 1998). Certaines études ont permis d'observer une répression stable de l'activité de transcription de différents gènes suite à leur méthylation (Schubeler et al., 2000). Inversement, d'autres études ont montré que des gènes inactifs pouvaient produire des transcrits suite à la déméthylation des CpG par la 5-azacytidine (Chiurazzi et al., 1998). Ainsi, la méthylation de l'ADN semblerait jouer un rôle très important dans la régulation des gènes en inhibant la transcription. Cette inhibition pourrait découler soit de la réduction de l'affinité entre les facteurs de transcription activateurs et leurs séquences d'ADN cibles, soit par la liaison de répresseurs ayant une grande affinité pour l'ADN méthylé. Un de ces répresseurs identifiés, MeCP2 (*methyl-CpG-binding*

*protein 2* ; protéine 2 qui se lie aux CpG méthylés), interagit également avec des histones déacétylases, établissant ainsi un lien moléculaire direct entre la méthylation et la déacétylation (Nan et al., 1998). L'acétylation des histones et la méthylation des CpG sont parmi les processus participant à la restructuration de la chromatine qui sont les plus étudiés (Lee et al., 2004; Zhang et al., 2004; Trievel, 2004; He et Lehming, 2003). Les facteurs de transcription joueraient un rôle déterminant dans ces processus (Kang et al., 2004; Tsai et Fondell, 2004; Follows et al., 2003; Johnson et al., 2003).

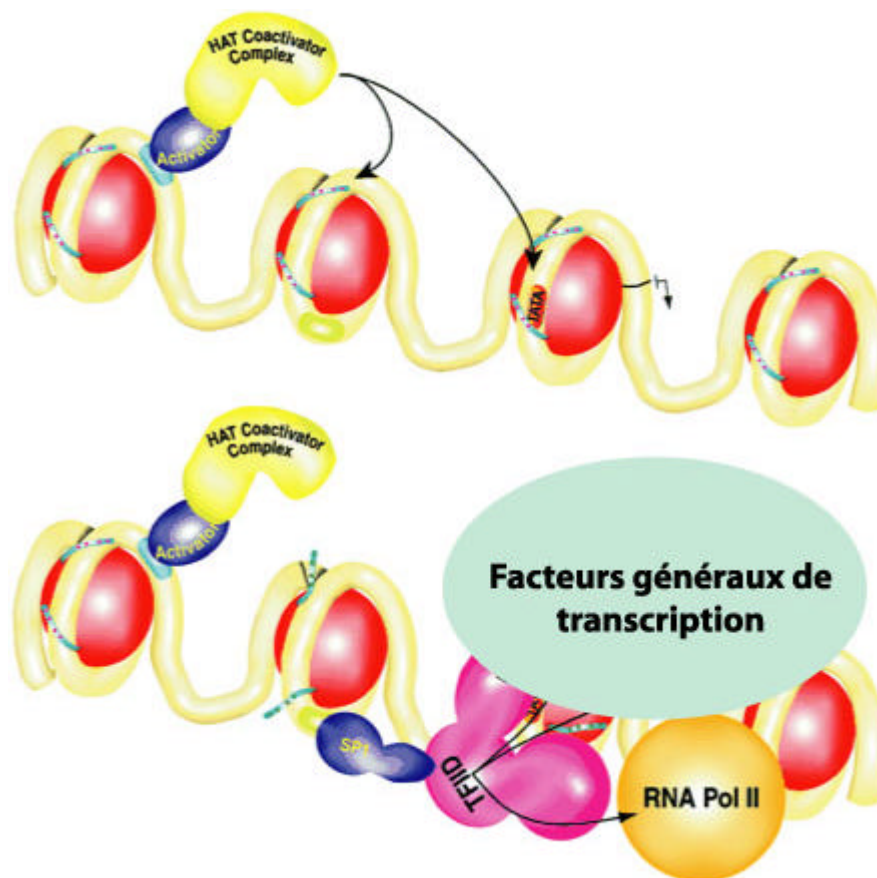


Figure 1 : Activation de la chromatine : l'acétylation des histones rend la chromatine permissive à la transcription, facilitant l'accès des facteurs généraux de transcription à l'ADN. L'acétylation des histones se fait sur la partie NH<sub>2</sub>-terminale. Elle neutralise la charge positive de la partie N-terminale des histones. Des modifications électrostatiques expliquent alors le changement de conformation du nucléosome, les charges négatives de l'ADN étant moins attirées par la partie N-terminale des histones maintenant neutre.

## 2. Le promoteur

La plupart des gènes de classe II, c'est-à-dire ceux qui codent pour les acides ribonucléiques messagers (ARNm), sont constitués de deux séquences d'ADN distinctes : la séquence régulatrice et la séquence transcrite qui est composée d'exons et d'introns. La séquence régulatrice, également appelée séquence promotrice, est située en amont de la séquence transcrite contrôle la production du messenger. Cette région peut parfois s'étendre sur des milliers de pb d'ADN.

L'élément promoteur de base le mieux caractérisé est la boîte TATA. Cet élément est localisé entre 25 et 30 paires de bases en amont du site d'initiation de la transcription et présente une séquence consensus TATAA (Lescure et al., 1994). Ce motif lie la TBP (*Tata box Binding Protein*) et cette dernière recrute par la suite l'ARN polymérase II et les facteurs généraux au site d'initiation de la transcription (SIT). Pour les autres gènes dont l'expression est limitée à certains tissus et qui n'ont pas de boîte TATA, le positionnement de l'ARN polymérase II au SIT se fait via la séquence initiateur (*initiator*) (Emami et al., 1998) (Kaufmann et Smale, 1994). L'élément initiateur est reconnu par la protéine CIF (*cofactor of Inr function*), qui recruterait le facteur TFIID (Kaufmann et al., 1998; Kaufmann et al., 1996) ou par les sous-unités TAF<sub>II</sub>S (*TBP-associated factor*) du facteur TFIID (Martinez et al., 1994), (Chalkley et Verrijzer, 1999), (Verrijzer et al., 1995). L'interaction des différents facteurs avec ces éléments a pour but d'initier l'assemblage précis du complexe de pré-initiation de la transcription.

Pour activer un gène précis, différents stimuli internes et/ou externes entraînent la liaison d'un certain nombre de facteurs de transcription particuliers à des séquences d'ADN régulatrices situées sur le promoteur (Felsenfeld et al., 1996). Ces facteurs provoquent l'ouverture locale de la chromatine, ce qui permet à d'autres facteurs de transcription de se lier à leurs séquences d'ADN régulatrices qui étaient inaccessibles auparavant (Kurdistani et al., 2004). Ces dernières protéines vont recruter et surtout stabiliser la liaison de l'holoenzyme au SIT du promoteur permettant une initiation efficace de la transcription (Yildirim et Doruker, 2004), (Borukhov et Nudler, 2003). Les facteurs de transcription qui se lient à la région

stimulatrice stabiliseraient encore davantage la liaison de l'holoenzyme au SIT (Gill, 2001a), (Zhang et al., 2002).

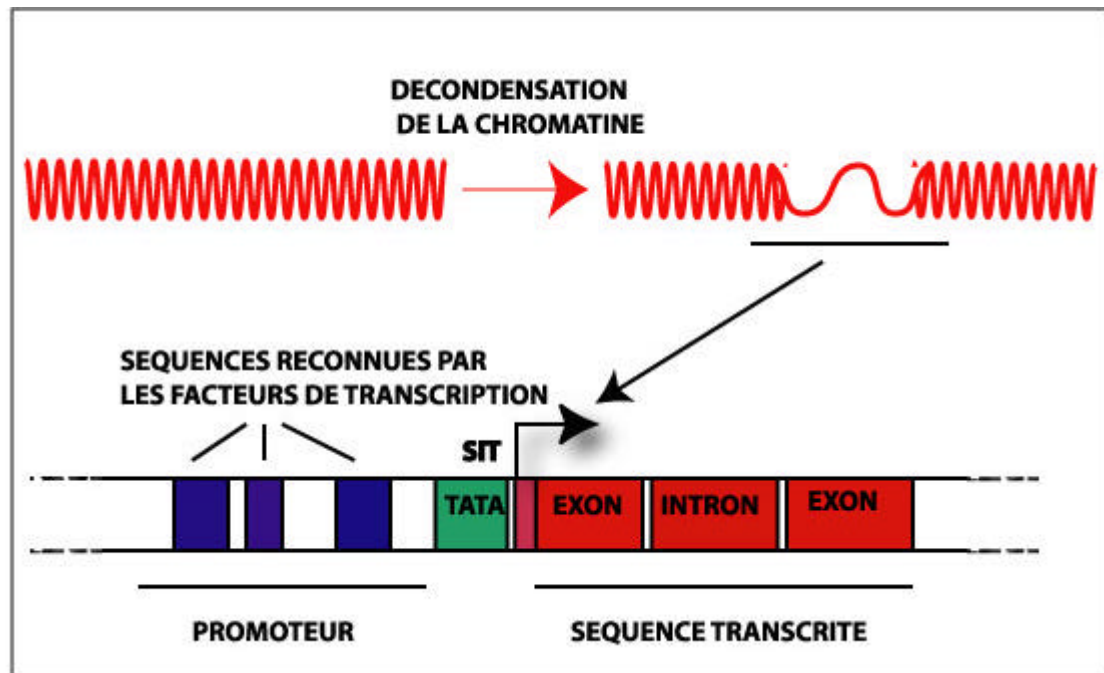


Figure 2 : La décondensation de la chromatine rend les différentes séquences d'ADN régulatrices accessibles aux facteurs de transcription.

Alors que certains facteurs, les « activateurs », favorisent la transcription, d'autres, les « répresseurs », en contrepartie l'inhibent. L'action de ces répresseurs, tout comme la perte d'activateurs, peut mener à la condensation de la chromatine ou à la déstabilisation de la liaison de l'holoenzyme au SIT (Gaston et Jayaraman, 2003; Rojo, 2001). Ces événements provoquent l'arrêt de toute nouvelle initiation de la transcription.

## B. La machinerie transcriptionnelle de base

### 1. Le complexe de préinitiation de la transcription

La régulation des gènes implique l'interaction entre la chromatine, les protéines activatrices et modulatrices ainsi que la machinerie générale de transcription. La machinerie basale de transcription (tableau 1), également nommée holoenzyme, comporte pour les gènes de classe II : l'ARN polymérase II (ARN Pol II) et les facteurs généraux de transcription (TFIIA, TFIIIB, TFIID, TFIIIE, TFIIF et TFIIH). Un complexe de co-activateurs fait également partie de l'holoenzyme, mais il diffère d'un gène à l'autre (Lewis et Reinberg, 2003). L'ARN polymérase II est composée de 12 sous-unités ; elle possède une extrémité carboxy-terminale unique comportant plusieurs répétitions de l'heptapeptide de séquence consensus Tyrosine-Sérine-Proline-Thréonine-Sérine-Proline-Sérine (Cabrejos et al., 2004) (Meininghaus et al., 2000). Chez les mammifères les répétitions de l'heptapeptide sont au nombre de 52. Localisé à l'extérieur du complexe de l'ARN Polymérase et donc accessible, ce domaine carboxy-terminal est important pour l'initiation et la terminaison de la transcription, l'épissage de l'ARN et la formation des extrémités poly-A de l'ARNm mature (Bird et al., 2004).

Parmi tous les facteurs généraux de la transcription, TFIID est le seul qui puisse se lier au promoteur proximal de façon spécifique et indépendante. La liaison de TBP (*TATA box binding protein* ; protéine qui se lie à la boîte TATA), un constituant de TFIID, sur le motif conservé thymine-adénine-thymine-adénine-adénine-adénine (TATAAA) par un contact direct avec l'ADN constitue l'évènement déclencheur de l'assemblage du complexe de pré-initiation de la transcription (Kim et Iyer, 2004). Outre la TBP, le complexe TFIID est composé d'au moins 10 autres facteurs qui sont associés à la TBP : les facteurs associés à la TBP (*TBP associated factors, TAF*) (Pereira et al., 2004; Furukawa et Tanese, 2000). Les fonctions des TAF sont multiples. Les TAF sont impliqués dans la reconnaissance de promoteur proximal, dans l'activation de la transcription par interaction avec les protéines activatrices, dans le remodelage de la chromatine et ils interagissent avec les autres facteurs généraux de la transcription.



Facteur général de transcription	Sous-unités	Propriétés
TFIIA	37kDa ( )	
	19kDa ( )	
	13kDa ( )	
TFIIB	35kDa	Fixation asymétrique au promoteur
TFIID (1000kDa)	TBP 38kDa	Lie la TATA box
	TAF <sub>II</sub> 250 250kDa	Kinase
	TAF <sub>II</sub> 150 150kDa	Lie le promoteur en amont
	TAF <sub>II</sub> 135 135kDa	
	TAF <sub>II</sub> 100 100kDa	
	TAF <sub>II</sub> 80 80kDa	Similarité à l'histone H4
	TAF <sub>II</sub> 65 65kDa	
	TAF <sub>II</sub> 55 55kDa	
	TAF <sub>II</sub> 31 31kDa	Similarité à l'histone H3
	TAF <sub>II</sub> 30 30kDa	
	TAF <sub>II</sub> 28 28kDa	Histone-fold
	TAF <sub>II</sub> 20 20kDa	Similarité à l'histone H2B
	TAF <sub>II</sub> 18 18kDa	Histone-fold
TFIIE	56kDa ( )	Recrutement de TFIIH avec TFIIE
	34kDa ( )	Lie l'ADN
TFIIF	58kDa (RAP74)	Stimule l'élongation
	26kDa (RAP30)	homologie
TFIIH (468kDa)	Voir §D	
ARN polymérase II (500kDa)	hRPB1 217kDa	
	hRPB2 130kDa	
	hRPB3 30kDa	
	hRPB4 14.5kDa	
	hRPB5 22kDa	
	hRPB6 14kDa	
	hRPB7 17.8kDa	
	hRPB8 15.4kDa	
	hRPB9 13.8kDa	
	hRPB10 10kDa	
	hRPB10 9kDa	
hRPB11 13.3kDa		

Tableau 1 : Facteurs généraux de transcription humains

## 2. Assemblage du complexe de préinitiation

Il existe deux modes d'assemblage du complexe de préinitiation de la transcription (figure 2) : le modèle séquentiel et le modèle de l'holoenzyme (Murakami et al., 2002), (Gill, 2001b), (Parvin et Young, 1998).

L'assemblage du complexe de pré-initiation de la transcription se fait par étape dans le modèle séquentiel. La première étape est la liaison de TBP/TFIID à la séquence TATA située à 30 paires de bases environ du site d'initiation de la transcription. Les facteurs TFIIA et TFIIIB arriveraient ensuite pour stabiliser le complexe TBP/ADN (Fan et al., 2004). Constitué de 3 sous-unités, TFIIA serait également un acteur important pour la transcription activée mais non pour la transcription basale. Dans un troisième temps, TFIIIB, qui avait été recruté par le facteur TFIID via la TBP, ira recruter à son tour un complexe formé par l'association entre TFIIIF et l'ARN polymérase II et positionnera cette dernière au site d'initiation de la transcription. TFIIIF, qui est composé de 2 sous-unités, facilite le recrutement de l'ARN polymérase II au promoteur, l'initiation de la transcription ainsi que l'élongation, et est le seul facteur qui reste associé à la polymérase pendant l'élongation. Les facteurs TFIIIE et TFIIH s'intègrent ensuite à ce complexe via une interaction TFIIIE/ARN polymérase II et TFIIIE/TFIIH (Bushnell et al., 1996) (Maxon et al., 1994).

Les fonctions principales de TFIIIE consistent à réguler les activités de TFIIH et d'initiation la transcription (Bushnell et al., 1996). En plus de jouer un rôle dans l'initiation de la transcription en collaboration avec TFIIIE et TFIIIF par son activité hélicase, TFIIH participe également à l'ouverture des brins d'ADN dans le système de réparation de l'ADN par excision-resynthèse de nucléotides, avec cette même activité hélicase (Hwang et al., 1996a). Finalement, TFIIH est impliqué dans la phosphorylation du domaine carboxy-terminal de l'ARN pol II (Watanabe et al., 2000). Le dernier élément de la machinerie basale de transcription, le complexe des co-activateurs, servirait de pont entre les facteurs généraux de la transcription (TFIID, TFIIIB, etc.) et les autres facteurs de transcription (activateurs, répresseurs, etc.). Ce complexe serait également étroitement lié au domaine carboxy-terminal de la polymérase (Riedl et Egly, 2000).

Le second modèle, dit de l'holoenzyme, permet l'association au site promoteur de toutes les composants du complexe de pré-initiation en une fois par la pré-existence de ce même complexe non-lié au promoteur. Ce modèle a été proposé suite à la purification du complexe *in vivo*.

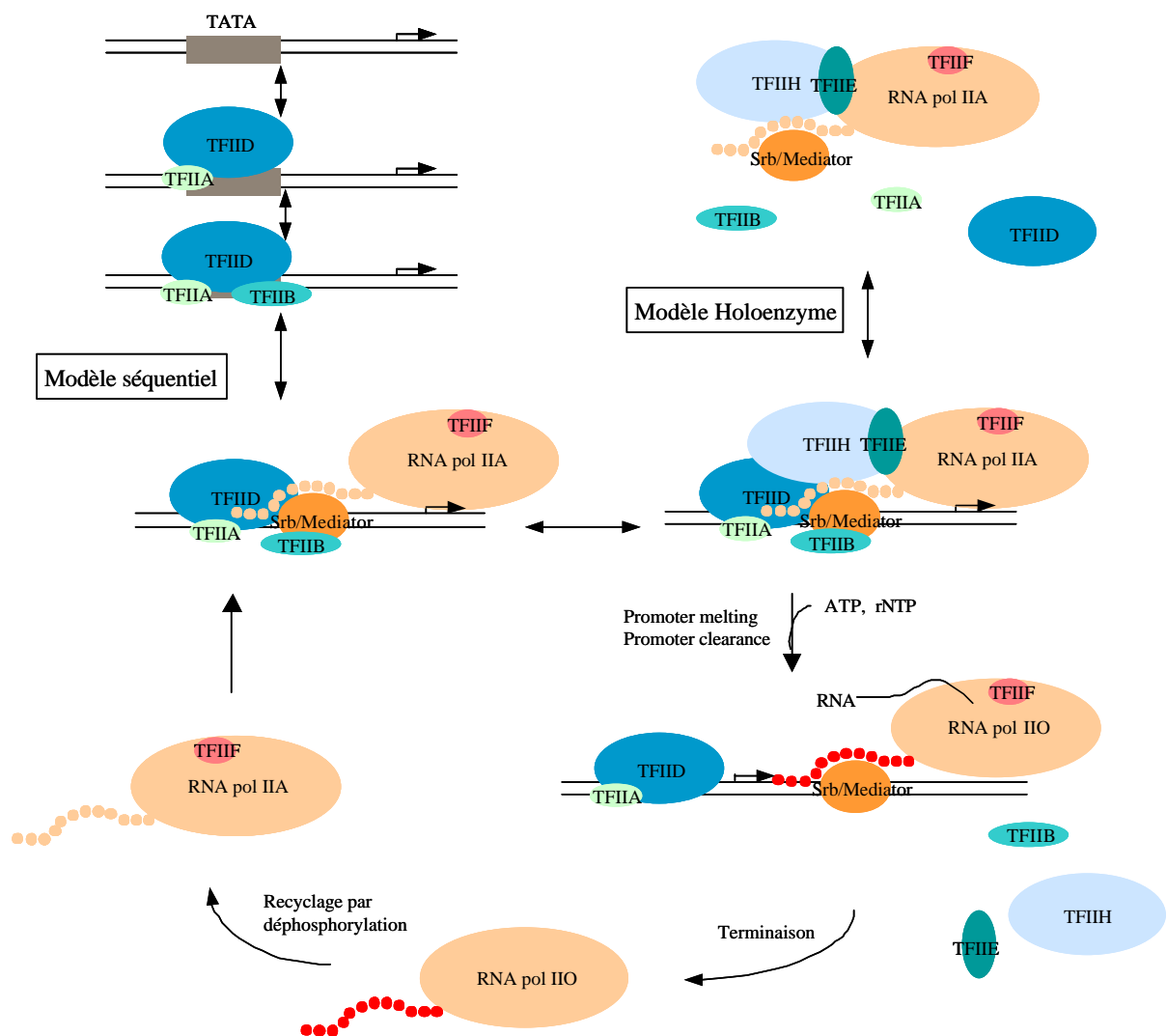


Figure 3 : Deux modèles pour l'assemblage du complexe de préinitiation de la transcription

### 3. Données structurales sur la transcription

La compréhension fine d'un mécanisme enzymatique implique l'accès à la structure tridimensionnelle des protéines et des acides nucléiques. De nombreuses structures de composants de la machinerie transcriptionnelle sont à présent connues. Nous allons faire le point des structures disponibles à ce jour (tableau 2).

En 1992, soit moins de trois ans après le clonage de son gène, la structure atomique de la TBP (TATA box Binding Protein) a été résolue en complexe avec un ADN double brin (Kim et al., 1993b; Kim et al., 1993a). Ceci constitue la base de l'initiation de la transcription. La structure de la TBP/ADN en complexe avec TFIIA ou TFIIB a ensuite été obtenue (Geiger et al., 1996), (Tan et al., 1996), (Nikolov et al., 1995), (Tsai et Sigler, 2000).

La résolution de composants isolés du complexe de préinitiation de la transcription a été engagée et les structures de protéines impliquées dans la transcription sont disponibles : les histone-like dTAF42/dTAF62 et hTAF18/hTAF28, le complexe TBP/TAF230 et un fragment de TAF250 (Hoffmann et al., 1996), (Birck et al., 1998), (Liu et al., 1998), (Jacobson et al., 2000), les sous-unités RPB8, RPB5 et RPB14.4 de la RNA polymérase II (Rio-Portilla et al., 1999), (Yee et al., 2000), (Krapp et al., 1998), la cycline H de TFIIH (Kim et al., 1996b), (Andersen et al., 1996). La sous-unité RAP30 de TFIIF (Groft et al., 1998) et les sous-unités et de TFIIE (Okuda et al., 2000) (Meinhart et al., 2003). La structure de la RNA polymérase II de levure a été publiée à une résolution de 2.6Å (Cramer et al., 2001) (Figure 5).

Nom	Taille	Organisme	Code PDB	Résolution	Technique
RPB5	Entier	<i>S. cerevisiae</i>	1DZF	1.9 Å	Rayons X
	Entier	<i>M. thermoauto.</i>	1EIK	-	RMN
RPB8	Entier	<i>S. cerevisiae</i>	1A1D	-	RMN
RPB14.4	Entier	Humain	1QKL	-	RMN
	Entier	<i>M. Janashii</i>	1HMJ	-	RMN
ARN pol II	Entier	<i>S. cerevisiae</i>	1NT9	2.8Å	Rayons X
ARN Pol II	Entier	<i>E.coli</i>	1Q8I	2 Å	Rayons X
TAF18/28	14-75	Humain	1BH8	2.6Å	Rayons X
	113-201	Humain	1BH9	2.6Å	Rayons X
TAF40/60	11-95 (TAF42)	Drosophile	1TAF	2.4Å	Rayons X
	1-82 (TAF62)	Drosophile			Rayons X
TBP/TAF230	60-249 (TBP)	<i>S. cerevisiae</i>	1TBA	-	RMN
	11-77 (TAF230)	Drosophile		-	RMN
TAF250	1359-1638	Humain	1EQF	2.1Å	Rayons X
TFIIB	2-59	Humain	1DL6	-	RMN
	113-316	Humain	1VOL	2.7Å	Rayons X
	112-316	Humain	1TFB	-	RMN
TBP	entier	<i>S. cerevisiae</i>	1TGH	1.9 Å	Rayons X
TFIIA	1-54 et 208-286 (TOA1), 2-122 (TOA2)	<i>S. cerevisiae</i>	1YTF	2.5 Å	Rayons X
Cycline H	Entier	Humain	1JKW	2.6Å	Rayons X
RAP30	164-249	Humain	1BBY	-	RMN
TFIIEb	66-146	Humain	1D8J	-	RMN
TFIIEa	1-110	<i>Sulfolobus</i>	1QH1	2.9 Å	Rayons X
RAP74	363-517	Human	1J2X	1.02 Å	Rayons X

Tableau 2 : Structures de complexes protéiques impliquées dans la transcription.

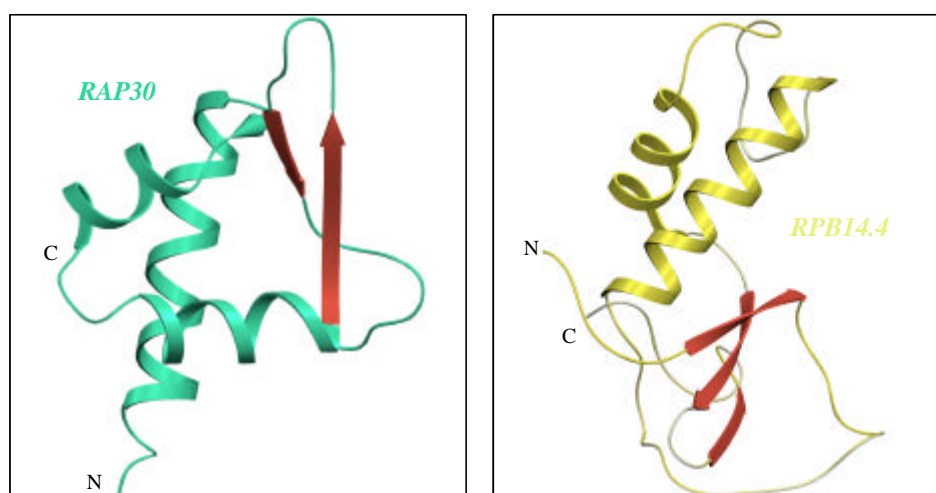


Figure 4 : Représentation schématique des structures de RAP30 et RPB14.4.

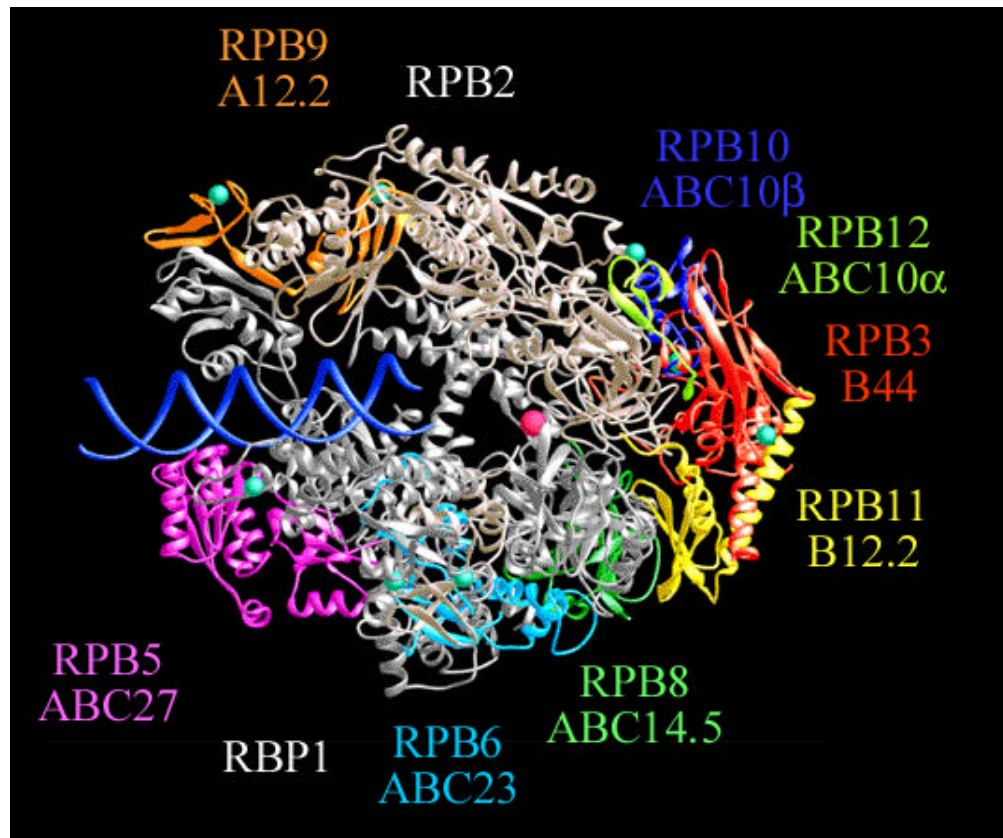


Figure 5 : Structure de la ARN polymérase II de levure à 2.8Å de résolution (Cramer et al., 2001).

## C. La réparation de l'ADN

### 1. La réparation par excision resynthèse des nucléotides (NER)

Il existe plusieurs mécanismes dans les cellules eucaryotes pour maintenir intègre l'information génétique. Outre l'activité d'édition des ADN polymérase, chaque cellule dispose de différents processus : (i) la réparation par excision de base (BER), (ii) la réparation par excision de nucléotides (NER), (iii) la réparation double brin, (iv) la réparation de mésappariement, (v) la réparation directe, (vi) et la réparation post-répllicative (Bootsma et al., 1995; Ng et al., 2003; Winkler et al., 2001; Slupphaug et al., 2003; Plosky et al., 2002; Kaina et al., 2001; van Hoffen et al., 2003).

Le processus de réparation dans lequel TFIIH est impliqué est le NER. Les composants de ce mécanisme sont à présent tous connus et le système a été reconstitué avec des protéines recombinantes chez la levure et chez l'homme (tableau 3) (Wang et al., 1994a; Sancar, 1996; Evans et al., 1997; de Laat et al., 1999; Volker et al., 2001c; Riedl et al., 2003a).

Protéines	PM	Homologues <i>S. cerevisiae</i>	Propriétés
XPA	31kDa	RAD14	Reconnaissance des dommages
RPA	70kDa	RFA1	Liaison de l'ADN simple brin
	34kDa	RFA2	
	11kDa	RFA3	
XPC/HHR23B	125kDa (XPC)	RAD4	Reconnaissance des dommages
	58kDa (HHR23B)	RAD23	
XPF	112kDa (ERCC4)	RAD1	Nucléase, incision 5'
	33kDa (ERCC1)	RAD10	
XPG	135kDa (ERCC5)	RAD2	Nucléase, incision 3'
RFC	140kDa		
	40kDa		
PCNA	32kDa	POL30	Interagit avec l'ADN polymérase
Pol /			resynthèse
DNA Ligase I	102kDa	CDC9	Ligation

Tableau 3 : Facteurs de réparation par excision-resynthèse de nucléotides

## 2. Les étapes du NER

La lésion sur l'ADN est reconnue par les facteurs XPC/HHR23B (Fitch et al., 2003), qui recrutent XPA, RPA et TFIIH (Tapias et al., 2004; You et al., 2003). Les complexes ERCC1/XPF et XPG rejoignent la région endommagée. Les incisions en 3' par XPG (Clarkson, 2003; Araujo et al., 2001), et en 5' par ERCC1/XPF sont effectuées après ouverture de l'ADN par TFIIH (Gaillard et Wood, 2001) (Feaver et al., 2000). L'oligonucléotide comprenant la zone endommagée est libéré et la région simple brin est protégée par RPA. L'ADN polymérase  $\delta$  en présence des facteurs PCNA et RFC permet la re-synthèse de la zone simple brin (Shivji et al., 1995; Ellison et Stillman, 2003). La dernière liaison phosphodiester est réalisée par la DNA ligase I (Wu et al., 1999).

TFIIH intervient dans la reconnaissance de la lésion sur l'ADN. Il interagit directement avec le complexe XPC/HHR23B (Volker et al., 2001b). De plus, il interagit avec d'autres composants de la machinerie de réparation, XPA (Park et al., 1995) et XPG, et il stimule l'activité de la nucléase XPF/ERCC1. Les sous-unités hélicase XPD (Coin et al., 1999c; Winkler et al., 2000a; Sung et al., 1996b) et XPB (Guzder et al., 1996) jouent un rôle direct dans la réaction par NER en permettant l'ouverture de l'ADN autour de la lésion, nécessaire aux coupures effectuées en 3' et en 5' par XPG et XPF/ERCC1, respectivement. De ce fait, les mutations des gènes codant pour XPB et XPD, observées dans certaines pathologies, affectent l'activité et/ou la conformation de ces protéines, réduisant l'action de TFIIH dans le système de réparation par NER (Lehmann, 2003; Van Brabant et al., 2000; Berneburg et Lehmann, 2001). Le complexe CAK quant à lui n'est pas nécessaire à la réparation, en effet, des travaux effectués par reconstitution *in vitro* montrent un effet négatif du complexe CAK sur la réparation (Sandrock et Egly, 2001b; Mu et al., 1996).



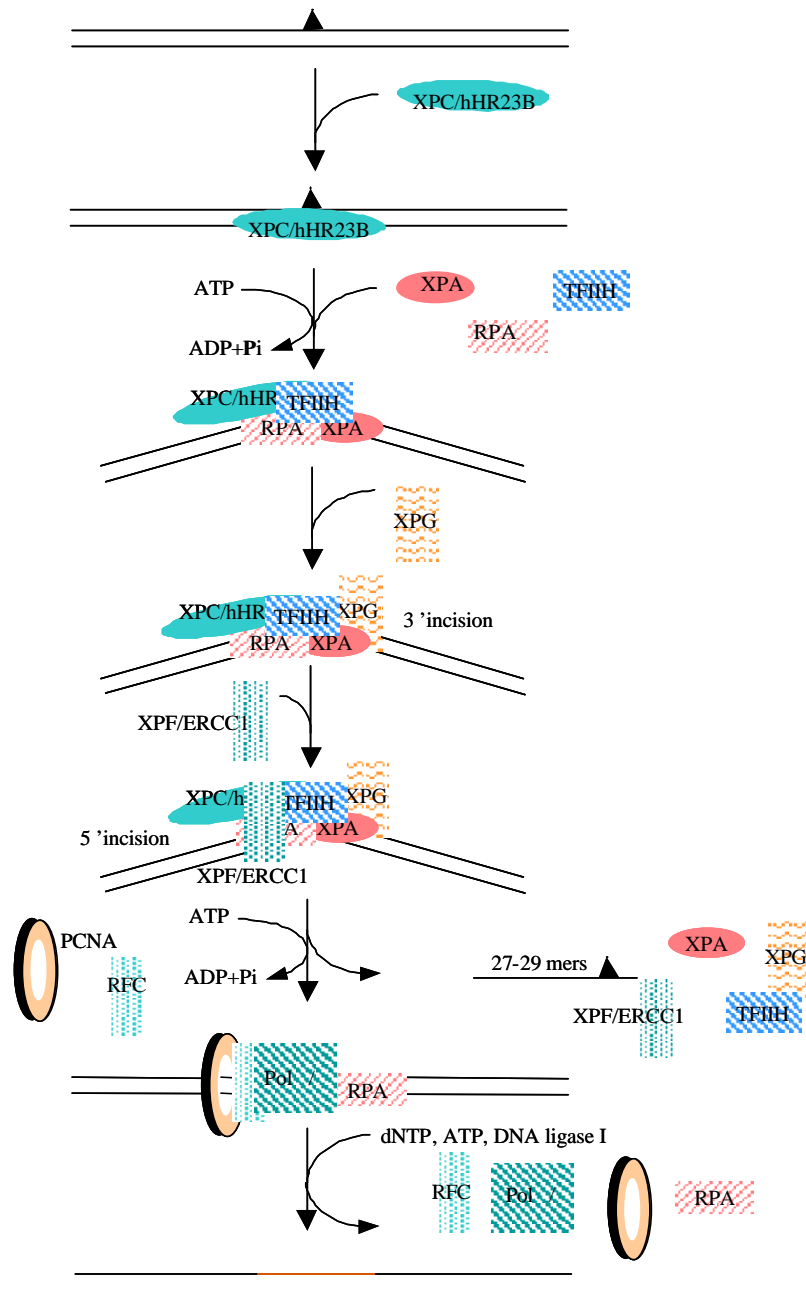


Figure 6 : La réparation par excision-resynthèse des nucléotides (NER) chez les eucaryotes.

### 3. Données structurales sur la réparation de l'ADN par NER

Comparativement à la transcription, peu de données structurales sont connues sur le système NER. Ceci reflète la difficulté d'obtenir ces complexes transitoires en quantité suffisante pour des études structurales. Les données disponibles sont rassemblées dans le tableau 4.

Nom	Taille	Organisme	Code PDB	Résolution	Technique
PCNA	Entier	<i>S. cerevisiae</i>	1PLQ	2.3Å	Rayons X
RPA-DNA	183-420	Humain	1JMC	2.4Å	Rayons X
RPA14/32	43-171 (RPA32)	Humain	1QUQ	2.5 Å	Rayons X
	Entier (RPA14)	Humain			
XPA	98-219	Humain	1XPA	-	RMN
XPC	Domaine de liaison	humain	1PVE	-	RMN
XPF	1-143	<i>Pyrococcus</i>	1J25	1.78 Å	Rayons X

Tableau 4 : Structures de complexes protéiques impliquées dans la réparation

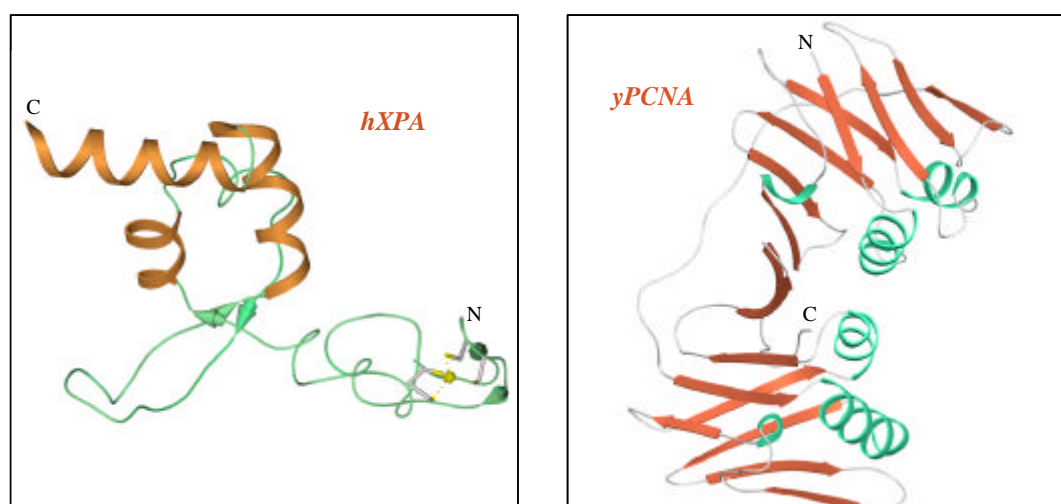


Figure 7 : Représentation schématique de la structure de hXPA et yPCNA (Ikegami et al., 1998), (Krishna et al., 1994).

## D. Le facteur de transcription TFIIH

### 1. Présentation du facteur TFIIH

Le facteur général de transcription des gènes de classe II, TFIIH, a d'abord été purifié chez le rat (facteur *a*) et chez la levure (facteur *b*) puis chez l'homme (Conaway et al., 1996), (Svejstrup et al., 1994), (Marinoni et al., 1997a). Le facteur TFIIH est un complexe multi-protéique de plus de 460 kDa, composé de dix sous-unités conservées de la levure à l'homme (voir Tableau 5). Il a longtemps été décrit comme le seul facteur de transcription possédant plusieurs activités enzymatiques comme des activités ATPase dépendantes de l'ADN, liées à des activités hélicases ainsi qu'une activité kinase. L'implication des deux hélicases de TFIIH (XPB et XPD) dans le système de réparation de l'ADN par excision-resynthèse de nucléotides, et la caractérisation du sous-complexe kinasique de TFIIH correspondant au complexe CAK nécessaire à la progression du cycle cellulaire, font de TFIIH un complexe majeur dans au moins trois processus cellulaires clés : la transcription, la réparation de l'ADN, et le cycle cellulaire (pour des revues, (Drapkin et Reinberg, 1994), (Seroz et al., 1995), (Bhatia et al., 1996), (Coin et Egly, 1998), (Zurita et Merino, 2003), (Kwek et al., 2004).

Le facteur TFIIH peut être isolé en deux sous-complexes majeurs, le sous-complexe *core* constitué de XPB, p62, p52, p44 et p34, et le sous-complexe CAK constitué de Cdk7, de la cycline H et de MAT1. La sous-unité XPD est généralement associée au sous-complexe *core*, mais peut également être associée au sous-complexe CAK. XPD pourrait ainsi permettre l'ancrage du sous-complexe CAK au *core* pour former le facteur TFIIH (Drapkin et al., 1996), (Reardon et al., 1996), (Rossignol et al., 1997).

Chez la levure *S. cerevisiae*, le sous-complexe *core* est constitué de Rad25, Rad3, Tfb1, Tfb2, Ssl1 et Tfb4 (les homologues respectifs de XPB, XPD, p62, p52, p44 et p34), alors que Kin28, Ccl1 et Tfb3/Rig2 (les homologues respectifs de Cdk7, cycline H et MAT1) forment un sous-complexe trimérique équivalent au sous-complexe CAK humain (Feaver et al., 1994b).

Sous-unité	Masse moléculaire (kDa)	Motif(s)	Activité(s)	
XPB (Rad25)	89	site de liaison à l'ATP, motifs hélicase, NLS, HTH	hélicase 3' -> 5'	<b>core</b>
XPD (Rad3)	80	site de liaison à l'ATP, motifs hélicase	hélicase 5' -> 3'	
p62 (Tfb1)	62	Domaine PH		
p52 (Tfb2)	52	NLS		
p44 (Ssl1)	44	motifs de liaison au zinc	liaison à l'ADN, régulateur de l'activité hélicase de XPD	
p34 (Tfb4)	34	"doigt" de zinc		
p8	8		stabilisation du complexe	
Cdk7 (Kin28)	38	domaine catalytique des protéines kinase, boucle T, NLS	kinase	
cycline H (Ccl1)	35	motif de type "cyclin fold"	régulateur de Cdk7	
MAT1 (Tfb3/Rig2)	33	motif de liaison au zinc de type RING- finger	régulateur de Cdk7	

Tableau 5 : Composition du facteur TFIIH.

Les sous-unités du facteur TFIIH humain et de levure (entre parenthèses) sont indiquées. Les deux principaux sous-complexes (*core* et *CAK*) sont indiqués.

NLS : séquence de localisation nucléaire (*nuclear localization sequence*).

HTH : motif hélice-coude-hélice (*helix-turn-helix*).

---

## 2. Les sous-unités de TFIIH

### 2.1 Les sous-unités du *core*

#### L'hélicase XPD

Le gène codant pour XPD a été caractérisé par transfection d'ADN génomique de cellules Hela dans des cellules de hamster sensibles aux UV et déficientes dans le système de réparation par incision. La protéine XPD est composée de 760 résidus et présente les motifs des hélicases. Elle possède une activité hélicase de polarité 5' -> 3', liée à une activité ATPase dépendante de l'ADN (Sung et al., 1996a). L'activité hélicase de XPD n'est pas indispensable à l'activité transcriptionnelle de TFIIH (Coin et al., 1999a), (Tirode et al., 1999), mais est nécessaire en réparation (Winkler et al., 2000b). XPD peut également être la cible d'activateurs transcriptionnels et ainsi moduler l'activité de la transcription (Tong et al., 1995).

Les gènes codant pour des homologues de XPD ont été caractérisés dans de nombreuses organismes comme *S. pombe* (Murray et al., 1992), le hamster (Kirchner et al., 1994), le poisson *Xiphophorus maculatus* (Della et al., 1995) et chez de nombreuses archaebactéries (voir Chapitre III).

#### L'hélicase XPB

La protéine XPB est capable de compléter le groupe B d'une maladie génétique rare, le Xeroderma pigmentosum (XP), à l'origine d'une déficience dans la réparation de l'ADN par le système NER. Le gène codant pour XPB a été caractérisé chez la levure comme *Ssl2 (suppressor of stem-loop)* (Gulyas et Donahue, 1992). Chez l'homme, XPB a été identifiée comme étant une sous-unité du facteur TFIIH.

La protéine XPB est composée de 782 résidus et possède les motifs hélicase. Elle possède une activité hélicase de polarité 3' -> 5', liée à une activité ATPase dépendante de l'ADN (Ma et al., 1994). L'activité hélicase de XPB est nécessaire en transcription et en réparation (Coin et al., 1999b) (Moreland et al., 1999). Les gènes codant pour des homologues de XPB ont été caractérisés chez de nombreux

organismes comme la drosophile (Mounkes et al., 1992), les plantes telles qu'*Arabidopsis thaliana* (Ribeiro et al., 1998), et dans un organisme unicellulaire procaryote, agent de la lèpre, *Mycobacterium leprae* (Poterszman et al., 1997).

#### La sous-unité p62

La protéine p62, composée de 548 résidus, fut la première sous-unité de TFIIH dont le gène a été cloné (Fischer et al., 1992). La protéine p62 joue un rôle en transcription et en réparation (Matsui et al., 1995) (Wang et al., 1995b) (Sweder et al., 1996). La protéine p62 possède un domaine PH (pleckstrin homology) (Gervais et al., 2004b). Ce domaine interagit physiquement avec la protéine XPG du complexe de réparation et apparaît nécessaire à la réparation de l'ADN. Il n'est pas nécessaire à l'assemblage de TFIIH ni à la transcription basale. De nombreux régulateurs interagissent avec TFIIH via la sous-unité p62. Par exemple, la protéine p62 peut interagir avec le facteur de transcription E2F, permettant le recrutement de TFIIH. La kinase de TFIIH phosphorylerait alors E2F, déclenchant sa dégradation par la voie de l'ubiquitine (Pearson et Greenblatt, 1997) (Vandel et Kouzarides, 1999).

#### La sous-unité p52

La protéine p52, composée de 436 résidus, est la dernière des sous-unités de TFIIH dont le gène a été cloné (Feaver et al., 1997; Marinoni et al., 1997b). Cette protéine ne présente pas de motifs particuliers permettant de lui conférer une fonction. Néanmoins, elle semble nécessaire à la viabilité de la cellule et jouerait un rôle en transcription et en réparation. La sous-unité p52 régule la fonction de l'hélicase XPB par des interactions protéine-protéine (Jawhari et al., 2002a).

#### La sous-unité p44

La protéine p44 est composée de 395 résidus (Humbert et al., 1994) et a d'abord été identifiée, chez *S. cerevisiae*, comme une protéine impliquée dans l'initiation de la traduction et appelée Ssl1 (*suppressor of stem-loop*) (Yoon et al., 1992). L'homologue de la protéine p44 a été caractérisé chez la levure *S. pombe* (Adachi et

al., 1999). La protéine p44 pourrait interagir avec l'ADN et présente dans sa partie C-terminale des motifs de liaison au zinc. Trois atomes de zinc peuvent être fixés par un motif "doigt de zinc" (*zinc finger*) de type C4 et par un nouveau motif de type C6H2 apparenté au *RING-finger* (Fribourg et al., 2000). p44 peut jouer un rôle en réparation en interagissant avec des protéines impliquées dans ce processus comme CSA et XPG (Henning et al., 1995) (Iyer et al., 1996) mais également avec p62, et surtout en modulant l'activité de l'hélicase XPD (Coin et al., 1998a). Il faut signaler que le gène codant pour p44 est dupliqué chez l'homme avec une copie centromérique (codant pour p44c) et une copie télomérique (codant pour p44t). Les protéines correspondantes diffèrent l'une de l'autre, dans leur séquence primaire, par trois acides aminés. Les gènes codant pour p44 se trouvent dans une région chromosomique à proximité du gène SMN (*survival motor neurone*). Cette région subit des réarrangements chromosomiques souvent délétères chez les patients atteints de la maladie de Werdnig-Hoffmann se traduisant par une atrophie musculaire spinale (Biros et Forrest, 1999; Biroş, I et al., 1997; Erdem et al., 1999).

#### La sous-unité p34

La protéine p34 se compose de 303 résidus dont l'homologue chez la levure est la protéine Tfb4. Cette protéine présente un motif de liaison au zinc correspondant au motif de type C4 de p44. Il a été montré que, comme les autres sous-unités du *core* de TFIIH, la protéine Tfb4 serait impliquée dans la transcription et dans la réparation (Feaver et al., 1999).

## 2.2. Les sous-unités du CAK

#### La sous-unité catalytique Cdk7

La protéine Cdk7 a été caractérisée chez de nombreuses espèces allant du xénope à l'homme. Cdk7 se compose de 346 résidus et présente les éléments caractéristiques des Cdks dont le site de liaison à l'ATP, l'équivalent de l'hélice PSTAIRE, et la boucle T où se situent les résidus phosphorylables Ser164 et Thr170, chez l'homme (Shiekhattar et al., 1995; Adamczewski et al., 1996). À la différence

des autres Cdk, Cdk7 ne présente pas les sites de phosphorylation inhibiteurs de l'activité correspondant aux résidus Thr14 et Tyr15 chez Cdk1, situés au niveau du site de liaison de l'ATP. De plus, Cdk7 possède, à son extrémité C-terminale, une séquence de localisation nucléaire bi-partite KRKR (Boulikas, 1996) située en position 328 à 331 chez l'homme. La séquence primaire de Cdk7 présente environ 40% d'identité avec celle des autres Cdk chez les vertébrés (Shuttleworth, 1995). Cdk7 est conservée parmi les eucaryotes : sa séquence primaire présente respectivement 47%, 84%, 86% et 93% d'identité avec les protéines de *S. cerevisiae*, de l'étoile de mer (*starfish*), de la drosophile et la souris.

### La sous-unité régulatrice cycline H

Le partenaire de Cdk7 a été caractérisé chez le xénope (Fisher et Morgan, 1994) et chez l'homme (Makela et al., 1994). La cycline H est un polypeptide de 323 résidus. La prédiction de structure secondaire montre que la partie centrale contient deux répétitions directes d'une centaine de résidus, constituées chacune de cinq hélices . La résolution de la structure cristallographique de la cycline H humaine confirme la prédiction de structures secondaires, deux répétitions s'agencent dans un motif de type *cyclin fold* déjà caractérisé chez la cycline A (Brown et al., 1995; Russo et al., 1996). Ce motif constitue la "signature" des cyclines et adopte la même orientation chez les cyclines A et H, phylogénétiquement éloignées. Les résidus impliqués dans l'interaction avec les Cdk sont conservés et adoptent la même orientation chez les deux cyclines. Les deux hélices N- et C-terminales se disposent différemment chez la cycline H par rapport à la cycline A. La région *cyclin box* conservée se situe au niveau de la première répétition.

Chez la levure *S. cerevisiae*, l'homologue de la cycline H, Ccl1 (Valay et al., 1996), est présent sous deux formes provenant de la traduction d'un même ARNm initiée à deux sites distincts. Le nombre d'acides aminés diffère de 19 entre les deux formes de Ccl1 qui sont toutes les deux capables de s'associer à Kin28 et d'être présentes dans le facteur TFIIH.



### La sous-unité MAT1

MAT1 (ménage à trois 1) a été initialement caractérisée chez l'étoile de mer (*starfish*) et chez le xénope (Devault et al., 1995), puis chez l'homme (Tassan et al., 1995; Yee et al., 1995) comme un facteur d'assemblage du complexe CAK. Chez la levure *S. cerevisiae*, l'homologue de MAT1, Tfb3 a été caractérisé. La protéine humaine possède dans sa partie N-terminale un motif de liaison au zinc C3HC4 de type *RING-finger*. De nombreuses protéines possédant ce type de motif interviennent dans des processus cellulaires variés comme l'oncogenèse, l'infection virale, ou la transduction de signaux (pour des revues, (Horn et al., 2004; Subramaniam et al., 2003)). Ces protéines possèdent un motif *RING-finger*, présentent souvent un motif de type *Coiled-coil* (Gervais et al., 2001). MAT1, en s'associant au complexe Cdk7-cycline H, peut stabiliser le complexe CAK et l'activer. MAT1 peut également orienter le complexe CAK vers des substrats à phosphoryler. C'est le cas de p53 (Schneider et al., 1998) et des facteurs Oct (Inamoto et al., 1997) qui sont phosphorylés par le complexe Cdk7-cycline H seulement en présence de MAT1. Une fois intégré dans le complexe TFIIH, MAT1 est capable d'interagir avec les facteurs Oct et pourrait alors intervenir dans le recrutement de TFIIH au niveau des promoteurs des gènes dont la transcription dépend de ces facteurs.

### **2.3. La dixième sous-unité TTD-A/p8**

Récemment une dixième sous-unité du complexe TFIIH a été identifiée chez la levure (Ranish et al., 2004; Giglia-Mari et al., 2004b). On savait que les formes B et D de la trichothiodystrophie étaient causées par des anomalies dans les hélicases XPD et XPD de TFIIH (Bergmann et Egly, 2001; Itin et al., 2001), en revanche aucune des neuf sous-unités de TFIIH n'était mutée chez les patients atteints de trichothiodystrophie de type A (TTD-A), bien qu'un taux réduit du complexe TFIIH soit mesuré chez les malades (Botta et al., 2002). Comme les cellules humaines, les cellules de levure déficientes en TFB5 sont hypersensibles à la destruction par les ultra-violets. Les équipes de Giglia (Giglia-Mari et al., 2004c) ont caractérisé le gène orthologue chez l'homme. Le gène TTD-A codant pour TFB5 a été cloné et l'injection de son ADNc permet de corriger l'anomalie de réparation de l'ADN des cellules TTD-

A. En outre, trois mutations inactivatrices de ce gène dans trois familles indépendantes affectées par la TTD-A ont été mises en évidence (Giglia-Mari et al., 2004a).

### 3. Données structurales sur TFIIH

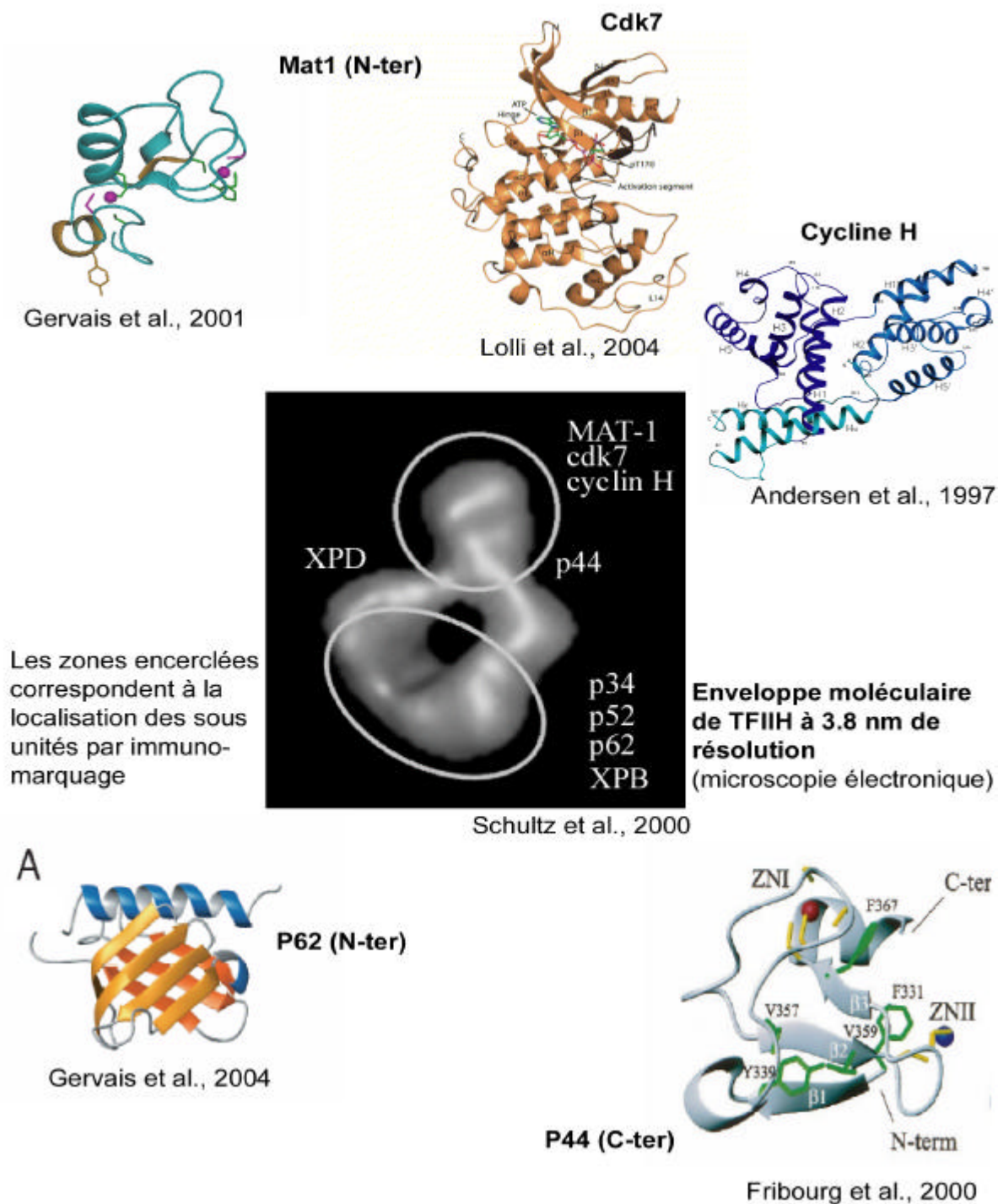


Figure 8 : Représentation des structures de TFIIH résolues à ce jour.

#### 4. TFIIH, un facteur aux interacteurs multiples

Le facteur TFIIH, outre la transcription des gènes de classe II et la réparation, est impliqué dans la transcription des gènes de classe I codant pour les ARN ribosomiaux (Iben et al., 2002). TFIIH serait à la fois associé à une sous-population de l'ARN polymérase I et au facteur TIF-IB-SL1 qui contient une TATA box binding protein et des facteurs associés spécifiques à l'ARN polymérase I (Hoogstraten et al., 2002).

Le facteur de transcription TFIIH, par ces nombreuses interactions, joue un rôle clé dans la transcription des gènes et dans d'autres mécanismes vitaux de la vie cellulaire. Le tableau 6 présente de manière exhaustive les protéines interagissant avec TFIIH, leur effet, et les références qui y renvoient.

TFIIH interagit, entre autres, avec p53. TFIIH interagit avec le suppresseur de tumeur p53 via cinq sous-unités : XPB, XPD, p62, cycline H et MAT1 (Xiao et al., 1994a). Cette interaction entre TFIIH et p53 a à la fois été décrite chez l'homme ainsi que chez la drosophile (Wang et al., 1995a; Wang et al., 2003a). D'une part, ces interactions inhibent les activités enzymatiques de TFIIH : l'interaction entre p53 et les deux hélicases conduit à une inhibition de leurs activités hélicases et l'interaction entre p53 et la cycline H résulte en une diminution significative de l'activité kinase de cdk7 (Xiao et al., 1994a). D'autre part, cdk7 phosphoryle p53, ce qui a pour effet d'augmenter son activité de liaison à des séquences spécifiques d'ADN (Lu et al., 1997). Après une irradiation aux UV, on observe une accumulation de p53 qui coïncide avec un arrêt du cycle cellulaire en phase G1 permettant la réparation de l'ADN. Par ailleurs, il n'est pas possible d'induire l'apoptose médiée par p53 dans des cellules dont les gènes XPB ou XPD sont mutés et la transfection de ces cellules par des vecteurs contenant ces deux gènes leur permet de s'engager dans le processus d'apoptose (Wang et al., 1996). On peut donc imaginer que TFIIH puisse établir un lien entre le blocage de la transcription par une lésion sur l'ADN et l'arrêt du cycle cellulaire par le biais de la phosphorylation de p53 (Lu et al., 1997). Cette activation de p53 pourrait par conséquent mener à l'arrêt du cycle cellulaire.

TFIIH a également un rôle dans la phosphorylation des récepteurs nucléaires. La phosphorylation du résidu Ser77 de RAR (récepteur à l'acide rétinoïque) par cdk7 stimule l'activité transactivatrice de RAR en présence de son ligand (Rochette-Egly et al., 1997). Le(s) mécanisme(s) précis par lequel la phosphorylation régule l'activité des récepteurs nucléaires et l'importance physiologique de cette phosphorylation ne sont pas encore clairs. Certains sites sont constitutivement phosphorylés alors que la phosphorylation d'autres sites est induite par la présence du ligand. On peut imaginer que certaines phosphorylations régulent la liaison du récepteur à son élément de réponse. En effet, la phosphorylation de la Ser389 de RAR influence la capacité des complexes RAR à fixer l'ADN (Rochette-Egly et al., 1997). A l'opposé, il est improbable que la phosphorylation par cdk7 influence la liaison du ligand ou la dimérisation des récepteurs car aucune de ces étapes ne requiert la présence de cette région. Il a enfin été montré que la phosphorylation des récepteurs RAR joue un rôle déterminant dans leur dégradation par la voie dépendante de l'ubiquitine (Kopf et al., 2000).

La phosphorylation des récepteurs nucléaires permettrait donc de réguler l'amplitude ainsi que la durée des effets provenant de l'activation des récepteurs par leur ligand. Les mutations dans l'hélicase XPD de TFIIH réduisent la transactivation ligand-dépendante du récepteur à l'acide rétinoïque RAR (Rochette-Egly et al., 1997; Keriél et al., 2002c). En effet, ces mutations altèrent l'activité de Cdk7 et par conséquent la phosphorylation des récepteurs nucléaires. Les effets des mutations de XPD sur l'activité des récepteurs nucléaires sont cruciales quant à la compréhension des phénotypes des patients.

Facteurs de transcription	Effet	Référence
FBP interacting factor (FIR)	La partie N-terminale de FIR interagit avec TFIIH. FIR supprime l'activité hélicase 3'-5' de XPB	(Liu et al., 2000)
Breast and ovarian tumor suppressor protein (BRCA1)	TFIIH est nécessaire à l'activation transcriptionnelle observée avec BRCA1	(Haile et Parvin, 1999)
Activateur transcriptionnel qui régule les genes aux bornes G/S (E2F1)	E2F1 a deux domaines de contact avec TFIIH. E2F1 interagit avec TFIIH via la sous-unité p62	(Pearson et al., 1997)
Retinoblastoma tumor suppressor (Rb)	Rb interagit avec la sous-unité p62 de TFIIH	(Pearson et al., 1997)
Proto-oncogene responsable de la pathogénécité du chromosome de Philadelphie (BRC)	BRC interagit avec la sous-unité XPB de TFIIH	(Maru et al., 1999)
Epstein Barr virus nuclear antigen 2 (EBNA2)	TFIIH interagit avec le domaine d'activation de EBNA2 via les sous-unités XPD et p62	(Tong et al., 1995)
Herpes simplex virus transactivator (VP16)	TFIIH interagit avec le domaine d'interaction de VP16 via la sous-unité p62	(Xiao et al., 1994b)
Suppresseur de tumeur p53	Le domaine d'activation de p53 se lie a la sous-unité p62	(Xiao et al., 1994b)
Positive cofactor 4 (PC4)	PC4 interagit avec TFIIH via la sous-unité XPD	(Fukuda et al., 2003)
RAD52	Le domaine C-terminal de RAD52 active la transcription. Ce domaine s'associe avec XPD et XPB et la polymérase II.	(Liu et al., 2002)
Androgen receptor (AR)	Cdk7 et la cycline H interagissent avec le domaine N-terminal de AR	(Lee et al., 2000)
Sug1	Sug1 est une ATPase contenue dans le proteasome 19S. Sug1 interagit avec XPB.	Weeda, 1997
Estrogen receptor (ER )	ER est phosphorylé dans sa région N-terminale par TFIIH. Cette phosphorylation à un rôle crucial dans la transcription médiée par un ligand.	(Chen et al., 2000)
Retinoid receptor (RAR )	Cdk7 phosphoryle RAR dans sa région N-terminale. Cette phosphorylation est importante dans le dérepression de la chromatine.	(Keriel et al., 2002b)
Retinoic acid receptor (RAR )	RAR est phosphorylé par Cdk7. Cette phosphorylation est cruciale pour l'activation des genes cibles et pour la différenciation cellulaire.	(Bastien et al., 2000a)
Metastasis-associated protein 1 (MTA)	MTA interagit avec TFIIH via MAT1. Cette interaction inactive l'activité du complexe CAK.	(Talukder et al., 2003)
Hepatitis virus transcriptional transactivator HBX	Les activités hélicase de TFIIH sont stimulées par HBX.	(Qadri et al., 1996)
Cyclin-dependant kinase 8 (Cdk8)	Cdk8 phosphoryle la cycline, cette phosphorylation réprime l'activation transcriptionnelle de TFIIH.	(Akoulitchev et al., 2000)

Nss (Rift Valley fever virus)	Nss interagit avec TFIID via p44 et bloque la transcription	(Le May et al., 2004)
-------------------------------	---	-----------------------

Tableau 6 : Protéines interagissant avec TFIID. Adapté d'après Zurita (Zurita et al., 2003).

## 5. TFIID à l'interface entre réparation et transcription de l'ADN

Comme nous l'avons vu précédemment, TFIID est à la fois impliqué dans la transcription des gènes de classe II ainsi que dans la réparation de l'ADN par excision resynthèse des nucléotides. Dès lors, une question fondamentale se pose, comment TFIID est-il recruté dans l'un ou dans l'autre de ces processus ? D'une part, il apparaît de plus en plus probable que la cellule utilise divers TFIIDs de composition identique qu'ils soient facteur de transcription ou de réparation mais dotés de modifications post-traductionnelles différentes. Certaines modifications de TFIID ont déjà été identifiées. Ainsi, la kinase mitotique cdc2/MPF phosphoryle la sous unité p62 de TFIID pouvant inhiber ainsi la transcription des gènes de classe II (Long et al., 1998). Il semblerait d'autre part que l'extrémité C-terminale de la sous unité XPB soit la cible d'une kinase de type CKII. Une phosphorylation de XPB favoriserait le rôle transcriptionnel de TFIID au détriment de celui de réparation, sans aucun effet sur l'activité hélicase de XPB. Par contre, une telle mutation à cette extrémité terminale de XPB empêche l'ouverture du promoteur autour du site d'initiation alors qu'elle permet l'ouverture de l'ADN autour de la lésion lors de la réparation. On constate ainsi que la fonction d'ouverture de l'ADN ne résulte pas du seul fait de l'activité hélicase de XPB.

D'autre part, TFIID est libéré du complexe de préinitiation dès que la polymérase entre dans la phase d'élongation (Zawel et al., 1995) (Spangler et al., 2001; Wang et al., 2003b). Récemment, Riedl et ses coauteurs (Riedl et al., 2003b) ont observé dans des tests de reconstitution *in vitro* la possibilité qu'à TFIID de passer de la transcription au NER. Les auteurs ont démontré qu'après l'arrivée de XPF-ERCC1, TFIID est libéré du complexe NER. TFIID est toujours actif et il peut réamorcer un nouveau cycle de NER ou de transcription. TFIID, une fois libéré du complexe

d'élongation de l'ARN polymérase II est capable de participer à un nouveau cycle de NER. De manière très intéressante, les auteurs ont constaté que TFIIH ne subit pas de modifications qui empêcheraient son recrutement dans l'un ou dans l'autre des mécanismes. Des tels résultats ont également été observé *in vivo*. Ce type de recyclage présume la libération directe du complexe de transcription bloqué sur une lésion de l'ADN et l'immédiate reprise de la transcription après le passage du complexe NER et la réparation de la lésion. De plus, l'affinité de TFIIH pour le NER semble être plus importante que celle pour la transcription (Vichi et al., 1997; You et al., 1998).

**E. Synthèse *Médecine et Sciences* : « Les mécanismes de déroulement affectés dans les maladies associées aux hélicases ».**

# **Les mécanismes de déroulement affectés dans les maladies associées aux hélicases**

## *Abréviations*

NER : réparation par excision resynthèse des nucléotides

WRN : syndrome de Werner

BS : syndrome de Bloom

CS : syndrome de Cockayne

FA : anémie de Fanconi

TTD : trichothiodystrophie

XP : xeroderma pigmentosum



## Résumé

Les hélicases sont des enzymes qui utilisent l'énergie libérée par l'hydrolyse de l'ATP pour catalyser le déroulement d'ARN ou d'ADN double brin. Des hélicases sont directement incriminées dans plusieurs désordres génétiques rares dans lesquels, le système de réparation des patients est atteint : le xeroderma pigmentosum, le syndrome de Cockayne, la trichothiodystrophie, le syndrome de Bloom, le syndrome de Rothmund-Thomson et le syndrome de Werner. Ces hélicases jouent un rôle majeur dans la réparation de l'ADN et permettent le maintien de l'intégrité du génome. Les hélicases apparentées à RecQ sont essentielles à la recombinaison de l'ADN et les protéines XPD et XPB sont impliquées dans la réparation par excision de nucléotides. Ces derniers font partie d'un complexe multi-protéique, le facteur TFIIH qui est par ailleurs également impliqué dans la transcription des gènes codant pour les protéines.

## Summary

### **Unwinding defect is incriminated in diseases related to helicases**

Deoxyribonucleic acid (DNA) repair is a fundamental process designed to keep the integrity of genomic DNA that is continuously challenged by intrinsic or environmental induced alterations. Numerous helicases involved in DNA repair have been cloned and are involved in different DNA repair pathways: base excision repair (BER), nucleotide excision repair (NER), mismatch repair, DNA recombination. RecQ helicases are a family of conserved enzymes requiring for maintaining the genome integrity, that function as suppressors of inappropriate recombination. Mutations in RecQ4, BLM and WRN give rise to the disorders: Bloom syndrome, Rothmund-Thomson syndrome, and Werner syndrome characterized by genomic instability and increased cancer susceptibility. Defections in XPD and XPB proteins, other helicases which are incriminated in xeroderma pigmentosum, Cockayne syndrome and Trichothiodystrophy, three genetic disorders with different clinical features but with association of NER defects. XPD and XPB helicases are components of the transcription factor TFIIH which is involved in both basal and activated transcription. To better define the precise roles of these helicases *in vivo*, significant research effort has been devoted to characterizing their biochemical properties and to identifying important protein interactions between these helicases and other characterised proteins. This review will focus primarily on these aspects and summarizes our current knowledge concerning their molecular mechanisms.

## Présentation générale des hélicases

La double hélice d'ADN constitue une structure stable bien adaptée à la transmission et à la sauvegarde de l'information génétique. Paradoxalement, de nombreux aspects du métabolisme de l'ADN nécessitent l'accès à une forme simple brin ou les bases nucléotidiques sont accessibles donc l'intervention de protéines de remodelage parmi lesquelles figurent hélicases et topo-isomérases. Les hélicases agissent au niveau de jonctions simple-brin double-brin et utilisent l'énergie libérée par l'hydrolyse de l'ATP pour catalyser la rupture des liaisons hydrogènes entre paires de bases nucléotiques. Cette famille de protéines qui présente une grande variabilité de séquence est néanmoins caractérisée par un ensemble de motifs caractéristiques connu sous le nom de "motifs hélicases". La fonction hélicase est assurée par un module catalytique composé de 2 domaines / parallèles dont la topologie est identique à celle de RecA, une protéine bactérienne impliquée dans les processus de recombinaisons. Les motifs conservés, au nombre de 7 pour les hélicases des superfamilles SF1 et SF2, sont localisés à l'interface des deux domaines. On peut schématiquement distinguer les motifs impliqués dans fixation de l'ATP/ADP-Mg, ceux dont la fonction est principalement la reconnaissance de l'oligonucléotide et enfin, ceux qui participent au mécanisme de couplage entre hydrolyse de l'ATP/changement conformationnel, ce qui conduit au déroulement de la double hélice et à des phénomènes de déplacement de brins (Caruthers Mc Kay 2002) (figure 1). Les hélicases n'opèrent pas de manière isolée dans la cellule mais leur action est couplée à celle d'autres enzymes du métabolisme des acides nucléiques (polymérase, ligase, endonucléase...) et bon nombre d'entre elles sont intégrées au sein de complexes multi-protéiques.

Des hélicases essentielles aux processus de réplication, transcription, recombinaison et réparation de l'ADN ont été identifiées dans l'ensemble du monde vivant et l'on estime que jusqu'à 1% des gènes eucaryotes pourraient coder pour ces enzymes. Une dizaine de pathologies humaines sont liées à des mutations dans des gènes codant pour une hélicase (voir tableau 1 et (van Brabant et al., 2000) pour une revue). Dans cette liste figure la senataxine, un orthologue de l'ARN hélicase Sen1p à l'origine d'une forme d'ataxie autosomique récessive (Moreira et al., 2004), des ADN hélicases apparentées à RecQ essentielles à la

recombinaison de l'ADN ainsi que les sous-unités XPB et XPD du facteur de transcription/réparation TFIIH.

## **Pathologies associées aux hélicases de la famille RecQ**

Les protéines de la famille RecQ sont impliquées dans les mécanismes de recombinaison et de maintien de la stabilité du génome. Homologues à la protéine RecQ d'*Escherichia coli*, ces protéines ubiquitaires ont été répertoriées dans un grand nombre de génomes bactériens, achaeobactériens et eucaryotes. Chez l'homme, les gènes de cinq hélicases de type RecQ ont été identifiés et des mutations dans trois d'entre eux (*WRN*, *BLM*, *RecQL4* et de localisations chromosomique respectives 8p-12, 15q-26.1 et 8q-24.3) sont à l'origine de maladies autosomales récessives rares (Kitao et al., 1999 ; Yu et al., 1996). Les syndromes associés, connus sous les noms de syndromes de Werner (WS), de Bloom (BS) et de Rothmund-Thomson (RTS) sont caractérisés d'un point de vue clinique par des risques élevés de développer des tumeurs malignes et des signes de vieillissement prématurés (voir tableau 1) (Mohaghegh and Hickson, 2001). Au niveau cellulaire, ces syndromes se caractérisent principalement par une instabilité génétique reliée à des anomalies de la réplication et de la recombinaison : fréquences élevées de recombinaison homologue incluant échanges réciproques entre chromatides sœurs et chromosomes homologues pour le BS, recombinaisons illégitimes et de grandes délétions chromosomiques dans le cas du WRN. Les mutations présentes chez un grand nombre de patients ont été étudiées. Dans le cas du BS, tous les types de mutation sont représentés, y compris des mutations faux sens conduisant au remplacement d'un unique acide aminé et à l'expression d'une protéine inactive. En ce qui concerne le WS, les mutations répertoriées sont exclusivement des mutations non sens, des mutations d'épissage ou des changements de phase, conduisant toutes à l'expression de protéines tronquées avec perte du signal de localisation nucléaire localisé à l'extrémité carboxy-terminale de la protéine et/ou déstabilisation de la protéine (Moser et al., 2000).

Les protéines RecQ de la famille sont des hélicases à ADN de polarité 3'-5'. Leur domaine catalytique est composé d'environ 400 acides aminés et la structure 3D du cœur catalytique de la protéine d'*Escherichia coli* (Bernstein et al., 2003) s'est révélée proche de celle observée pour d'autres protéines de la famille SF2 comme les hélicases de l'hépatite C NS3, le facteur d'initiation eIF4a de levure ou la protéine bactérienne RecG. La majorité d'entre elles

présentent une région C-terminale conservée composée de deux modules : (i) Un domaine RQC (Rec-Q Conserved) impliqué dans le recrutement d'autres protéines du métabolisme des acides nucléiques et dont un motif hélice boucle hélice est capable de fixer l'ADN. (ii) Un domaine HRDC (Helicase-RNaseD-C-terminal) qui n'est pas indispensable à l'activité catalytique mais qui pourrait former des contacts spécifiques avec l'ADN au sein de complexes de recombinaison et/ou de réplication. Deux membres de cette famille, les protéines de Werner (WRN) et la protéine FFA-1 de Xénope possèdent un domaine exonucléase (Shen and Loeb, 2000).

La caractérisation *in vitro* des substrats des hélicases RecQ ainsi que l'identification de partenaires permet d'envisager un rôle clef pour ces protéines dans la résolution de structures d'ADN aberrantes qui apparaissent lors de divers processus du métabolisme de l'ADN. En effet, les hélicases RecQ utilisent préférentiellement des ADN dont la structure secondaire mime soit des intermédiaires de réplication ou de recombinaison comme des fourches, des bulles ou des jonctions Holliday (Bernstein et al., 2003). Leur activité hélicase participe également à la résolution de structures à 4 brins appelées G-quaruplex. De telles structures, formées par des séquences riches en guanine et stabilisées par des liaisons hydrogène de type Hoogsteen sont trouvées entre autre au niveau de l'ADN télomérique et de triplets CGG associés au syndrome X-fragile. Ces données, corrélées à l'étude des réseaux d'interaction protéine-protéine auxquels participent les protéines RecQ (Figure 2A) et à des données fonctionnelles ont permis de proposer des rôles dans : la réplication, la recombinaison homologe, la réparation ainsi que dans les mécanismes de stabilisation des télomères et de surveillance du génome. A titre d'exemple, et concernant ce dernier rôle, il a été montré que la protéine BLM appartient au complexe BASC (*BRCA1-associated genome surveillance complex*) impliqué dans des mécanismes de détection et/ou de réparation des lésions de l'ADN (Meetei et al., 2003). De plus, les protéines BLM et WRN interagissent avec la protéine p53 et la formation du complexe inhibe leurs activités hélicases ainsi que l'activité exonucléase de WRN (Shen and Loeb, 2000). De manière générale, les protéines RecQ pourraient coordonner diverses voies de réponse aux dommages et de réparation et jouer le rôle de transducteur. Par exemple, elles participeraient à la reconnaissance de différents types de lésions (bases modifiées, coupure simple brin ou coupure double brin) et permettre le recrutement d'une voie de réparation appropriée : réparation des coupures single brin (Single Strand Break Repair) ou réparation par excision de bases (Base Excision Repair),

recombinaison homologue (Homologous Recombination) ou réparation par jonction non homologue (Non Homologous End Joining) (Opresko et al., 2004) (figure 2B).

### **XPD et XPB : les hélicases de la réparation par excision resynthèse**

La réparation par excision-resynthèse de nucléotides (NER) permet de corriger les modifications de l'ADN à l'origine d'une distorsion de la double hélice comme des adduits de cisplatine ou des photoproduits engendrés par les rayons UV qui bloquent la réplication de l'ADN et la transcription. Il existe deux voies de réparation par NER : la réparation couplée à la transcription (TCR) qui concerne la réparation rapide des lésions situées sur le brin transcrit et affectant des gènes en cours de transcription, et la réparation globale du génome (GGR) qui concerne la réparation lente de l'ADN pour des gènes qui ne sont pas en cours de transcription.

Ce mécanisme est identique chez les eucaryotes et les procaryotes bien que les enzymes impliquées soient différentes. Chez l'homme, il est maintenant bien élucidé en grande partie grâce à l'étude de lignées cellulaires de patients atteints de maladies génétiques autosomales récessives comme le xeroderma pigmentosum (XP), la trichothiodystrophie (TTD) et le syndrome de Cockayne (CS). Toutes ces pathologies sont caractérisées par une sensibilité très élevée au soleil avec prédisposition précoce et exagérée au développement de cancers de la peau dans le cas du XP, mais non dans les cas CS et TTD. Les analyses de groupe de complémentation ont permis l'identification des protéines impliquées dans la réparation par NER de l'ADN. Les protéines ont été nommées selon leur groupe d'appartenance. Sept groupes de complémentations XP ont ainsi été identifiés (XP-A à XP-G), deux groupes CS (CSA et CSB) et trois TTD (TTD-A, XPB-TTD, XPD-TTD). Chacun de ces groupes correspond à la mutation d'un gène XP (voir le site : <http://www.xpmutations.org/>) entraînant l'altération spécifique du système de réparation. Chez l'homme, une trentaine de protéines sont nécessaires pour réaliser les différentes étapes : reconnaissance de la lésion par le complexe XPC/HHR23B, ouverture locale de la double hélice par les hélicases XPD et XPB, incision du brin endommagé par les endonucléases XPG et XPF/ERCC1 de part et d'autre de la lésion suivie de son excision et finalement re-synthèse par l'ADN polymérase / en

présence des facteurs PCNA et RFC et ligation réalisée par la DNA ligase I (Dip et al., 2004) (figure 3).

Les gènes XPB et XPD codent pour des ADN hélicases et font partie du complexe général de transcription TFIIH. TFIIH, complexe multi-protéique de 10 sous-unités, possède 3 activités enzymatiques dépendantes de l'ATP : deux activités ADN hélicase (Schaeffer et al., 1993) et une activité kinase dépendante des cyclines (Fisher and Morgan, 1994). Le facteur TFIIH est nécessaire au démarrage de la transcription qui consiste en l'avancée de l'ARN Polymérase II libérant le promoteur d'où son nom de "promotor clearance". Dans ce processus, la fonction hélicase de XPB est absolument nécessaire alors que la présence physique de l'hélicase XPD est suffisante (Bradsher et al., 2000). Lors de l'initiation de la transcription, qui nécessite l'hydrolyse des liaisons de l'ATP (Bunick et al., 1982) requise pour la formation d'un complexe ouvert (revue dans (Dvir et al., 2001)), l'activité hélicase XPB joue un rôle déterminant alors que l'hélicase XPD n'a qu'un rôle mineur (Bradsher et al., 2000). Enfin l'activité kinase Cdk7 contenue dans le complexe TFIIH permet de phosphoryler le domaine carboxy-terminal de l'ARN Polymérase II engagée, guidant le processus vers l'étape de l'élongation.

Bien que les hélicases XPD et XPB semblent être impliquées de manière différente dans la transcription, leur rôle respectif dans la réparation par NER est relativement similaire. Ainsi des mutations dans l'hélicase XPB sont rarement viables. La cartographie des mutations trouvées chez plus d'une centaine de patients XP-D révèle que plus de 80% des mutations sont localisées dans la partie C-terminale de XPD (Taylor et al., 1997). Le spectre de mutations de XPD a révélée que la position de la mutation détermine le phénotype du patient. Bien que certaines caractéristiques cliniques soient dues à des déficiences dans le NER, d'autres ont été associés à des défauts dans la transcription (de Boer et al., 2002). A ce jour, aucune explication moléculaire ne permet d'expliquer les différences cliniques générées par les mutations dans le gène XPD. Lorsque les mutations sont positionnées dans les motifs hélicase I et V, la liaison à l'ATP et/ou l'ADN pourrait être affectée, ce qui expliquerait l'inhibition de l'activité hélicase de XPD chez les patients XP. Positionnées dans la partie C-terminale de l'hélicase, les mutations empêchent l'interaction de XPD avec p44, la sous-unité régulatrice également présente dans le facteur TFIIH (Coin et al., 1998). Les mutations n'ont aucun effet sur l'activité hélicase bien que l'activité hélicase de XPD mutée soit diminuée chez ces patients. L'absence d'interaction de XPD avec p44 est à l'origine de la déficience

d'ouverture normale de l'ADN autour de la lésion. L'incision du brin endommagé et son excision ne peuvent se réaliser.

Il existe d'autres maladies associées à des déficiences des hélicases XPD et XPB : la combinaison du xeroderma pigmentosum avec la trichothiodystrophie ou avec le syndrome de Cockayne résultent dans des syndromes mixtes (XP/TTD ou XP/CS).

La trichothiodystrophie est une maladie très rare, caractérisée par des cheveux cassants et une ichtyose. La mutation ponctuelle de résidus ou la délétion de parties des gènes codant pour les hélicases XPD et XPB sont à l'origine de ce syndrome. La plupart des patients portent des mutations sur l'hélicase XPD, les mutations sur XPB ne concernent que peu de patients (Weeda et al., 1997). Ces mutations semblent déstabiliser la structure de TFIIH car la quantité du complexe est considérablement affectée dans les cellules TTD en comparaison aux cellules XP-B et XP-D. De ce fait, la déstabilisation de TFIIH pourrait être à l'origine d'une déficience dans la réparation par NER et dans la transcription de l'ADN (de Boer et al., 1998). Les mutations dans le motif I de l'hélicase XPD empêcheraient l'hydrolyse de l'ATP à l'origine d'un changement conformationnel de XPD et/ou TFIIH (Taylor et al., 1997). L'absence d'activité ATPase affecterait l'interaction avec les composants de la machinerie basale de transcription et empêcherait ainsi la formation du complexe de préinitiation de la transcription. Récemment, la protéine incriminée dans le groupe de complémentation TTD-A a été identifiée (Ranish et al., 2004). C'est une mutation dans la 10<sup>ème</sup> sous-unité de TFIIH, TFB5/p8, qui est à l'origine de cette forme de TTD (Giglia-Mari et al., 2004).



## Conclusion

Les efforts investis ces dix dernières années dans l'étude de la structure et de la fonction des hélicases nous permettent à présent de mieux situer leur importance dans les deux grandes voies du métabolisme de l'ADN que sont la réparation par NER et la réplication de l'ADN. D'autres hélicases interviennent dans le NER par la voie de transcription couplée à la réparation (TCR), par exemple la protéine CSB (ou ERCC6) qui est caractérisée par la présence des sept motifs hélicase mais dont aucune activité hélicase n'a été identifiée à ce jour fait partie de la famille SNF2 des ATPases et des putatives hélicases. La région N-terminale de CSB est très homologue à la putative hélicase SWI2 du complexe de levure SWI/SNF qui aurait une activité de remodelage de la chromatine. Le gène codant pour la protéine CSB est muté chez les patients développant le syndrome de Cockayne. Cette protéine interagit avec le facteur TFIIH et des mutations dans les gènes codant pour les hélicases XPB et XPD de TFIIH conduisent alors à des syndromes mixtes XP/CS (Itoh et al., 1996). Dans tous les cas les mutations, multiples, impliquent une déficience dans la réparation de l'ADN chez les patients (Venema et al., 1990). L'hélicase putative ATRX appartient également à la famille SNF2. Cette protéine est largement exprimée et est associée à la chromatine. L'alpha thalassémie est provoquée par des mutations sur le gène ATRX, elle est récessive et liée à l'X (Ion et al., 1996). Les mutations sont à l'origine de divers changements dans le schéma de méthylation de l'ADN ; cependant, il n'a pas encore été élucidé si elles étaient responsables du phénotype clinique.

Les déficiences d'hélicases sont des pathologies fascinantes aussi bien pour le clinicien que pour le scientifique. Bien que les syndromes soient rares, il apparaît important de préciser les fonctions des hélicases dans la genèse de maladies humaines. Les patients atteints ayant un risque important de développer des pathologies graves liées à l'âge, comme les cancers, le diabète ou les maladies cardiovasculaires, on peut envisager que les facteurs endogènes ou exogènes capables de moduler l'activité des hélicases chez les individus sains puissent aussi augmenter le risque d'apparition de ces maladies.

### **Diagnostic du Xeroderma Pigmentosum**

Pour affirmer le diagnostic de Xeroderma Pigmentosum, la méthode biologique la plus couramment utilisée est l'UDS (unscheduled DNA synthesis) qui repose sur la vérification du fonctionnement du système de réparation de l'ADN du patient par l'incorporation d'un précurseur radioactif de l'ADN (thymidine tritiée) après irradiation cellulaire aux ultraviolets. Une cellule normale incorporera d'autant plus de ce précurseur radioactif pour réparer son ADN que la dose d'UV reçue aura été plus importante. Une cellule XP, incapable de réparer correctement les dommages, n'incorporera que peu de précurseur marqué. Cette méthode est appliquée sur des cellules fœtales cultivées in vitro provenant de biopsie de trophoblaste ou de liquide amniotique. L'efficacité de réparation de l'ADN est comparée à celle de cellules issues de biopsie de peau des parents normaux.

## Bibliographie

**Bernstein, D. A., Zittel, M. C., and Keck, J. L.** High-resolution structure of the E.coli RecQ helicase catalytic core 2. *EMBO J.*22[19], 4910-4921. (1-10-2003).

**Bradsher J, Coin F, Egly JM.** Distinct roles for the helicases of TFIIH in transcript initiation and promoter escape. *J Biol Chem.* 28;275[4], 2532-2538. (2000).

**Brosh, R. M., Jr., Li, J. L., Kenny, M. K., Karow, J. K., Cooper, M. P., Kureekattil, R. P., Hickson, I. D., and Bohr, V. A.** Replication protein A physically interacts with the Bloom's syndrome protein and stimulates its helicase activity. *J.Biol.Chem.*275[31], 23500-23508. (4-8-2000).

**Broughton, B. C., Steingrimsdottir, H., and Lehmann, A. R.** Five polymorphisms in the coding sequence of the xeroderma pigmentosum group D gene. *Mutat.Res.*362[2], 209-211. (15-2-1996).

**Caruthers JM, McKay DB.** Helicase structure and mechanism. *Curr Opin Struct Biol.* 12[1]:123-33. (2002)

**Coin, F., Marinoni, J. C., Rodolfo, C., Fribourg, S., Pedrini, A. M., and Egly, J. M.** Mutations in the XPD helicase gene result in XP and TTD phenotypes, preventing interaction between XPD and the p44 subunit of TFIIH. *Nat.Genet.*20[2], 184-188. (1998).

**Cordin, O., Tanner, N. K., Doere, M., Linder, P., and Banroques, J.** The newly discovered Q motif of DEAD-box RNA helicases regulates RNA-binding and helicase activity. *EMBO J.*23[13], 2478-2487. (7-7-2004).

**de Boer, J., de Wit, J., van Steeg, H., Berg, R. J., Morreau, H., Visser, P., Lehmann, A. R., Duran, M., Hoeijmakers, J. H., and Weeda, G.** A mouse model for the basal transcription/DNA repair syndrome trichothiodystrophy 5. *Mol.Cell*1[7], 981-990. (1998).

**de Boer J, Andressoo JO, de Wit J, Huijmans J, Beems RB, van Steeg H, Weeda G, van der Horst GT, van Leeuwen W, Themmen AP, Meradji M, Hoeijmakers JH.** Premature aging in mice deficient in DNA repair and transcription. *Science.* May 17;296[5571],1276-1279.(2002).

**Dip R, Camenisch U, Naegeli H.** Mechanisms of DNA damage recognition and strand discrimination in human nucleotide excision repair. *DNA Repair* 2;3[11]:1409-23. (2004)

**Giglia-Mari, G., Coin, F., Ranish, J. A., Hoogstraten, D., Theil, A., Wijgers, N., Jaspers, N. G., Raams, A., Argentini, M., van der Spek, P. J., Botta, E., Stefanini, M., Egly, J. M., Aebersold, R., Hoeijmakers, J. H., and Vermeulen, W.** A new, tenth subunit of TFIIH is responsible for the DNA repair syndrome trichothiodystrophy group A. *Nat.Genet.*36[7], 714-719. (2004).

- Gorbalenya, A. E., Koonin, E. V., Donchenko, A. P., and Blinov, V. M.** Two related superfamilies of putative helicases involved in replication, recombination, repair and expression of DNA and RNA genomes. *Nucleic Acids Res.*17[12], 4713-4730. (26-6-1989).
- Ion, A., Telvi, L., Chaussain, J. L., Galacteros, F., Valayer, J., Fellous, M., and McElreavey, K.** A novel mutation in the putative DNA helicase XH2 is responsible for male-to-female sex reversal associated with an atypical form of the ATR-X syndrome 1. *Am.J.Hum.Genet.*58[6], 1185-1191. (1996).
- Itoh, T., Cleaver, J. E., and Yamaizumi, M.** Cockayne syndrome complementation group B associated with xeroderma pigmentosum phenotype. *Hum.Genet.*97[2], 176-179. (1996).
- Kitao, S., Shimamoto, A., Goto, M., Miller, R. W., Smithson, W. A., Lindor, N. M., and Furuichi, Y.** Mutations in RECQL4 cause a subset of cases of Rothmund-Thomson syndrome. *Nat.Genet.*22[1], 82-84. (1999).
- Matson, S. W.** DNA helicases of Escherichia coli. *Prog.Nucleic Acid Res.Mol.Biol.*40, 289-326. (1991).
- Meetei, A. R., Sechi, S., Wallisch, M., Yang, D., Young, M. K., Joenje, H., Hoatlin, M. E., and Wang, W.** A multiprotein nuclear complex connects Fanconi anemia and Bloom syndrome 4. *Mol.Cell Biol.*23[10], 3417-3426. (2003).
- Mohaghegh, P. and Hickson, I. D.** DNA helicase deficiencies associated with cancer predisposition and premature ageing disorders. *Hum.Mol.Genet.*10[7], 741-746. (2001).
- Moreira, M. C., Klur, S., Watanabe, M., Nemeth, A. H., Le, Ber, I, Moniz, J. C., Tranchant, C., Aubourg, P., Tazir, M., Schols, L., Pandolfo, M., Schulz, J. B., Pouget, J., Calvas, P., Shizuka-Ikeda, M., Shoji, M., Tanaka, M., Izatt, L., Shaw, C. E., M'Zahem, A., Dunne, E., Bomont, P., Benhassine, T., Bouslam, N., Stevanin, G., Brice, A., Guimaraes, J., Mendonca, P., Barbot, C., Coutinho, P., Sequeiros, J., Durr, A., Warter, J. M., and Koenig, M.** Senataxin, the ortholog of a yeast RNA helicase, is mutant in ataxia-ocular apraxia 2 1. *Nat.Genet.*36[3], 225-227. (2004).
- Moser, M. J., Kamath-Loeb, A. S., Jacob, J. E., Bennett, S. E., Oshima, J., and Monnat, R. J., Jr.** WRN helicase expression in Werner syndrome cell lines 3. *Nucleic Acids Res.*28[2], 648-654. (15-1-2000).
- Opresko, P. L., Cheng, W. H., and Bohr, V. A.** Junction of RecQ helicase biochemistry and human disease. *J.Biol.Chem.*279[18], 18099-18102. (30-4-2004).
- Pedrazzi, G., Perrera, C., Blaser, H., Kuster, P., Marra, G., Davies, S. L., Ryu, G. H., Freire, R., Hickson, I. D., Jiricny, J., and Stagljar, I.** Direct association of Bloom's syndrome gene product with the human mismatch repair protein MLH1 4. *Nucleic Acids Res.*29[21], 4378-4386. (1-11-2001).
- Ranish, J. A., Hahn, S., Lu, Y., Yi, E. C., Li, X. J., Eng, J., and Aebersold, R.** Identification of TFB5, a new component of general transcription and DNA repair factor IIIH. *Nat.Genet.*36[7], 707-713. (2004).

**Riedl T, Hanaoka F, Egly JM.** The comings and goings of nucleotide excision repair factors on damaged DNA. *EMBO J.* 22[19], 5293-303. (2003).

**Schaeffer, L., Moncollin, V., Roy, R., Staub, A., Mezzina, M., Sarasin, A., Weeda, G., Hoeijmakers, J. H., and Egly, J. M.** The ERCC2/DNA repair protein is associated with the class II BTF2/TFIIH transcription factor 4. *EMBO J.* 13[10], 2388-2392. (15-5-1994).

**Shen, J. C. and Loeb, L. A.** The Werner syndrome gene: the molecular basis of RecQ helicase-deficiency diseases. *Trends Genet.* 16[5], 213-220. (2000).

**Taylor, E. M., Broughton, B. C., Botta, E., Stefanini, M., Sarasin, A., Jaspers, N. G., Fawcett, H., Harcourt, S. A., Arlett, C. F., and Lehmann, A. R.** Xeroderma pigmentosum and trichothiodystrophy are associated with different mutations in the XPD (ERCC2) repair/transcription gene 1. *Proc.Natl.Acad.Sci.U.S.A*94[16], 8658-8663. (5-8-1997).

**van Brabant, A. J., Stan, R., and Ellis, N. A.** DNA helicases, genomic instability, and human genetic disease. *Annu.Rev.Genomics Hum.Genet.* 1, 409-459. (2000).

**Venema, J., Mullenders, L. H., Natarajan, A. T., van Zeeland, A. A., and Mayne, L. V.** The genetic defect in Cockayne syndrome is associated with a defect in repair of UV-induced DNA damage in transcriptionally active DNA. *Proc.Natl.Acad.Sci.U.S.A*87[12], 4707-4711. (1990).

**Weeda, G., Eveno, E., Donker, I., Vermeulen, W., Chevallier-Lagente, O., Taieb, A., Stary, A., Hoeijmakers, J. H., Mezzina, M., and Sarasin, A.** A mutation in the XPB/ERCC3 DNA repair transcription gene, associated with trichothiodystrophy 10. *Am.J.Hum.Genet.* 60[2], 320-329. (1997).

**Yu, C. E., Oshima, J., Fu, Y. H., Wijzman, E. M., Hisama, F., Alisch, R., Matthews, S., Nakura, J., Miki, T., Ouais, S., Martin, G. M., Mulligan, J., and Schellenberg, G. D.** Positional cloning of the Werner's syndrome gene. *Science* 272[5259], 258-262. (12-4-1996).

## Figures et tableaux

Figure 1 : (A) Représentation schématique des boîtes hélicases et assignement des fonctions. Les noms des motifs (Q, I, Ia, II, II, IV, V, VI) sont indiqués dans les boîtes correspondantes. L'espacement entre les motifs est indiqué de manière arbitraire. Les fonctions définies pour chaque motif sont données. (B) Structure de l'hélicase RecQ. Le code couleur représentant les motifs hélicase est identique à (A). (C) Relations entre les motifs au site de liaison à l'ATP et au site de liaison à l'oligonucléotide.

Figure 2 : (A) Partenaires identifiés des protéines BLM et WRN dans la réparation. Les partenaires de BLM sont représentés par une flèche rouge, ceux de WRN par une flèche bleue. Les voies métaboliques dans lesquelles ces protéines participent incluent la réplication, la réparation par excision de base, la recombinaison homologue et la réparation non homologue. (B) Modèle de réponse aux dommages et voies de réparation de l'ADN pour WRN et BLM. SSB : dommage simple brin; DSB : dommage double brin; BER : réparation par excision de base; HR : recombinaison homologue; NHEJ : réparation par jonction non homologue.

Figure 3 : Les hélicases de TFIIH interviennent dans la réparation par excision des nucléotides (NER). La lésion est reconnue par la machinerie de réparation et TFIIH est recruté, les hélicases XPD et XPB ouvrent le double brin d'ADN autour de la lésion et l'ADN est réparé. Si les hélicases XPD et XPB sont mutées, TFIIH n'est plus capable d'ouvrir l'ADN : c'est le cas les maladies autosomiques récessives rares Xeroderma Pigmentosum (XP), Cockayne Syndrome (CD) et trichothiodystrophie (TTD).

Tableau 1 : Syndromes associés à des mutations dans les hélicases

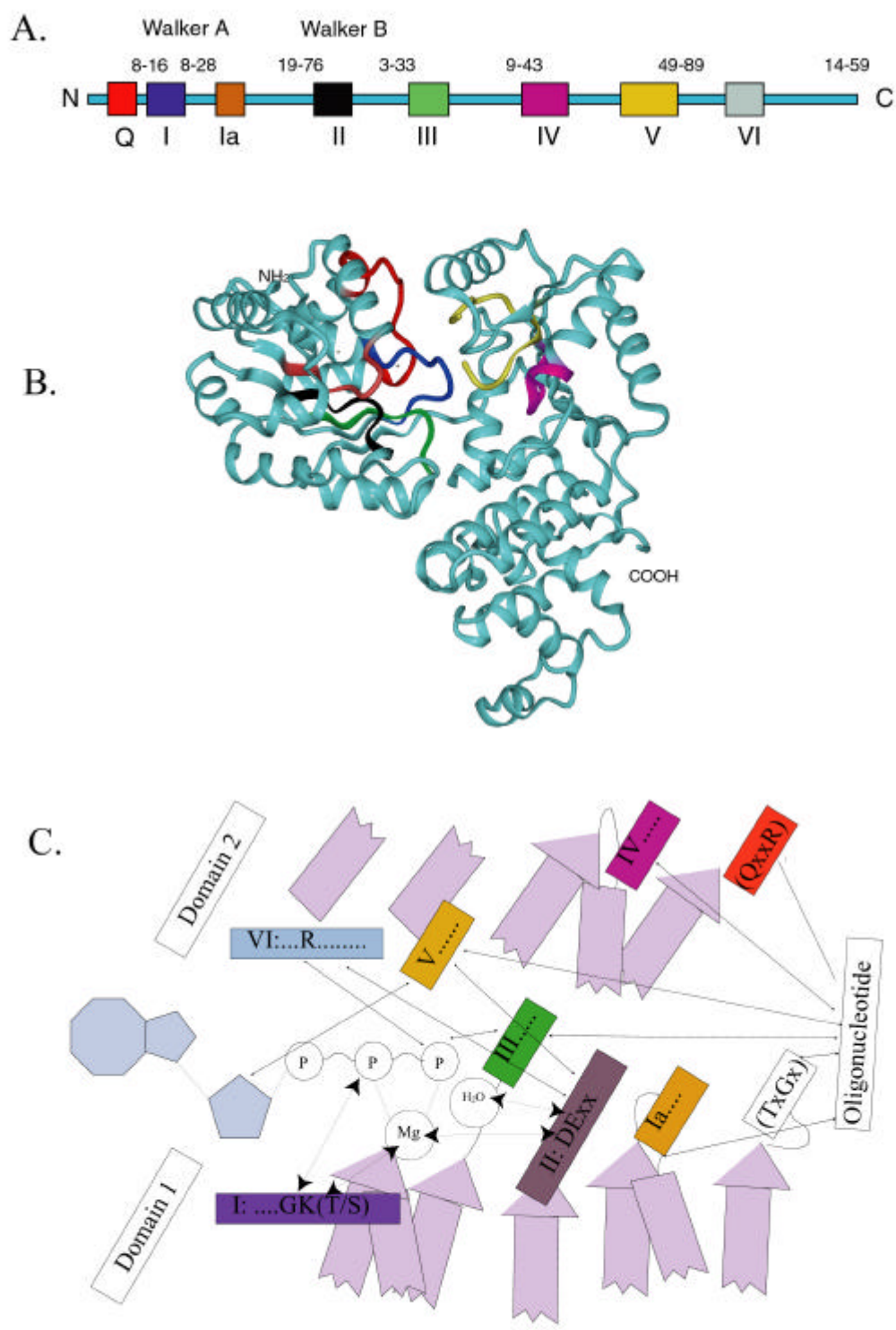
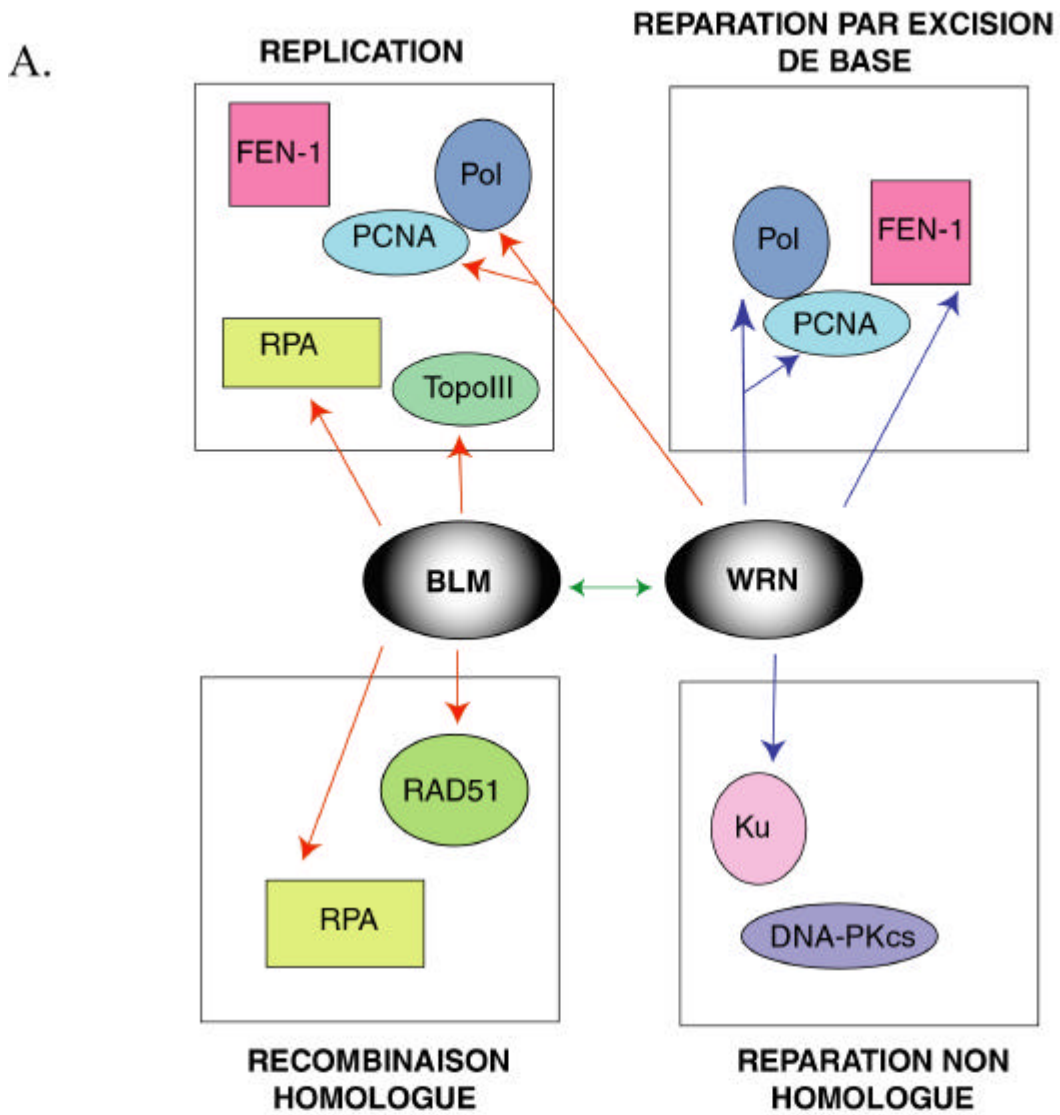


Figure 1



B.

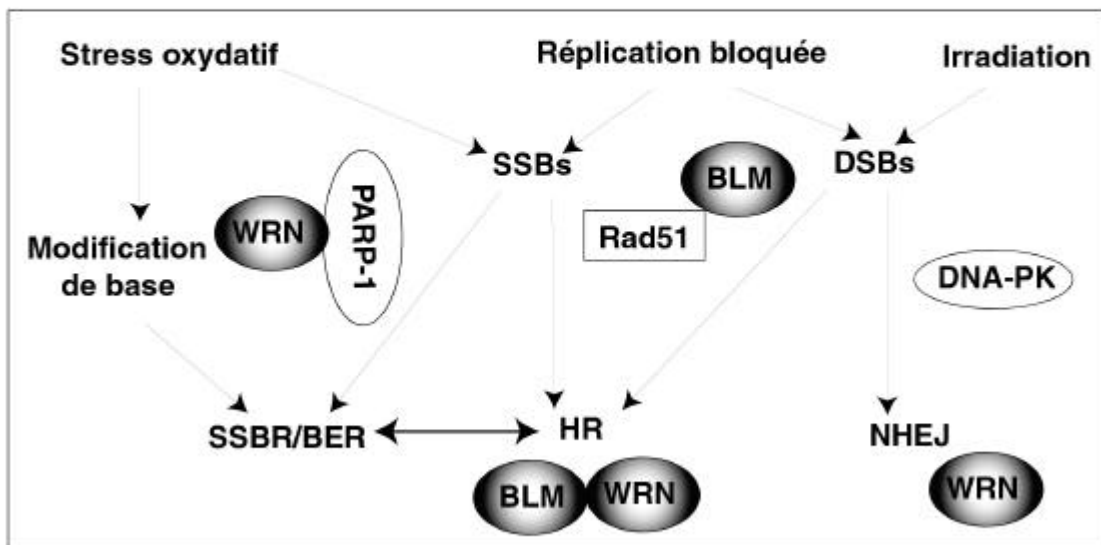
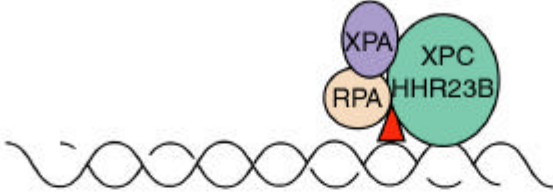


Figure 2

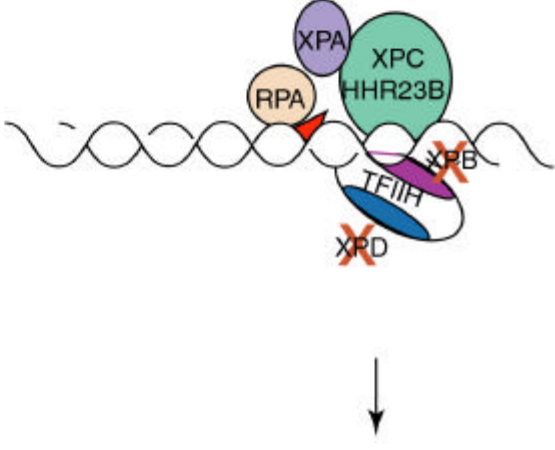
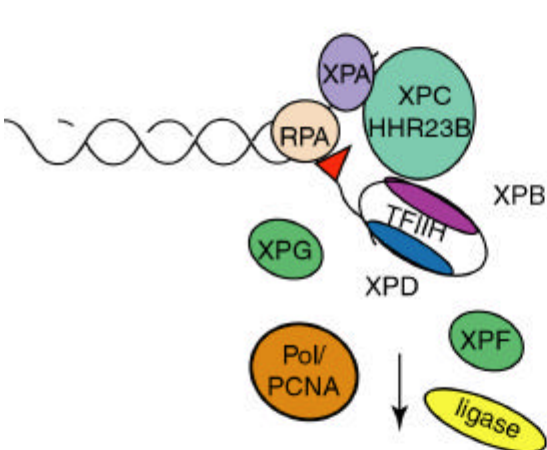


RECONNAISSANCE DE LA LESION



OUVERTURE DE L'ADN  
PAR LES HELICASES XPD ET XPB DE TFIIH

MUTATION DES HELICASES DE TFIIH :  
L'ADN N'EST PAS OUVERT



REPARATION DE L'ADN

MALADIES AUTOSOMALES RECESSIVES :  
TTD, CS, XP

Figure 3

SYNDROME	HELICASES IMPLIQUEES	SIGNES CLINIQUES	FONCTION DE L'HELICASE	COMPLEXE	FAMILLE APPARENTÉE	LOCALISATION CHROMOSOMIQUES
Xeroderma Pigmentosum	XPD XPB	Photosensibilité Cancers de la peau	Transcription couplée au NER	TFIIH	SF2	19q13.2
Syndrome de Cockayne	XPD XPB	Photosensibilité Retard de croissance Retard mental	Transcription couplée au NER	TFIIH	SF2	19q13.2
Trichothiodystrophie	XPD XPB	Photosensibilité	Transcription couplée au NER	TFIIH	SF2	19q13.2
Syndrome de Bloom	BLM	Photosensibilité Cancer Viellissement prématuré	Réplication de l'ADN	BASC	RecQ	15q26.1
Syndrome de Werner	WRN	Cancer Viellissement prématuré	Réplication de l'ADN	BRAFT	RecQ	8p11.1
Syndrome de Rothmund Thomson	RECQL4	Cancer Viellissement prématuré	Réplication de l'ADN	?	RecQ	8q24.3
Anémie de Fanconi	FA	Retard de croissance Cancers	?	SWI/SNF ?	SF2	16q24.3
Thalassémie	ATR-X	Retard mental	Régulation de la transcription		SF2	Xq31
Ataxie oculaire	Senataxine	Atrophie cérébrale	Epissage Réparation de l'ADN ?	?	SF1	9q34.13

Tableau 1

## Chapitre II : ORGANISATION QUATERNAIRE DE TFIIH

### *Introduction*

La biologie structurale regroupe trois techniques différentes permettant l'accès à des informations structurales à des degrés différents. La cristallographie par diffraction des rayons X est la plus résolutive puisqu'elle permet l'accès à des informations du niveau atomique (de 1Å à 2.5Å) voire en au-delà dans les cas les plus favorables (résolution inférieure à 0.7Å dans le cas de l'aldose réductase (Howard, Proteins, 2004)). Cette technique ne souffre *a priori* d'aucune contrainte de taille de la molécule à étudier mais elle nécessite l'obtention sous forme d'un cristal, d'un réseau ordonné macroscopique. Ainsi des données de l'ordre de 2.5Å sont disponibles sur des complexes macromoléculaires tels que la RNA polymérase II de levure, le ribosome (Ban et al., 1998; Ban et al., 2000; Dahlberg, 2001; Joseph, 2003). Toutefois, il faut remarquer que la difficulté de cristalliser augmente avec la taille et la complexité de l'objet.

La seconde technique est la résonance magnétique nucléaire (RMN) qui permet l'obtention d'informations structurales d'une valeur généralement acceptée autour de 3Å de résolution. Cette technique a pour limitation majeure l'intensité des champs magnétiques appliqués (800Mhz en routine, 900MHz ou 1GHz dans 1 ou 2 endroits au monde) permettant d'étudier en routine des protéines d'une taille d'environ 30-35kDa maximum bien que de récents développements dans les techniques d'enregistrements indiquent une probable extension (Salzmann et al., 1998).

La dernière méthode est la microscopie électronique. Ces dernières années, cette technique a vu son champ d'action considérablement s'accroître et notamment pour ce qui est de la limite de résolution. En 1991, la structure à l'échelle atomique de la bactériorhodopsine a été obtenue à partir de cristaux en deux dimensions (Miercke et al., 1991). Mais pendant longtemps, les meilleures résolutions obtenues à partir de molécules isolées étaient confinées aux alentours de 10Å. Les récents développements informatiques de traitement des données (Penczek et al., 1992) (van Heel et al., 1996) ont permis de pousser cette résolution bien au-delà de 10Å

avec les structures du ribosome d'*E. coli* (9Å) (Valle et al., 2003) et de sa sous-unité 50S (7.5Å) (Matadeen et al., 1999) et de la GroEL à 6Å (Ludtke et al., 2004). A cette résolution, les hélices sont clairement visibles et la densité des feuilletés également. Les molécules étudiables par cette technique doivent être d'une taille suffisante (supérieure à 200kDa).

L'étude structurale du facteur de transcription TFIIH entier par microscopie électronique a débuté par la collaboration entre le laboratoire de Génomique et de Biologie Structurale dirigé par le Dr Dino MORAS de l'IGBMC et l'équipe du laboratoire de microscopie électronique du Dr Patrick SCHULTZ. Cette étude a donné lieu à une première publication (Schultz et al., 2000) et ce travail a ouvert la voie à une étude plus systématique du facteur TFIIH endogène et recombinant par coloration négative et cryomicroscopie.

Dans un microscope à transmission, les électrons sont émis à partir d'une cathode puis sont accélérés en traversant un gradient de champs électriques. Au début de mon travail de thèse, nous avons utilisé le microscope Philips CM120 doté d'une cathode de type LaB6 et nous avons principalement observé des échantillons en coloration négative. J'ai ensuite pu utiliser un nouveau microscope de type Tecnai F20 équipé d'une cathode à effet de champ (FEG) et opérant à 200kV. Les qualités de cette source sont, par comparaison à la cathode LaB6, une très forte cohérence et luminosité du faisceau d'électrons ainsi qu'une plus petite taille de la source.

Dans un premier temps, nous verrons les principes de la microscopie électronique, l'analyse d'images et la reconstruction d'un modèle en trois dimensions (3D). Ensuite, nous aborderons les premiers résultats obtenus en cryomicroscopie électronique sur le facteur TFIIH endogène. Nous exposerons les limitations rencontrées lors des nombreux efforts fournis pour étudier le complexe endogène par microscopie. Nous nous focaliserons ensuite sur l'étude du sous complexe recombinant core-TFIIH (rIIH6) et finalement nous présenterons un modèle général sur l'architecture du facteur TFIIH.

## **I. La microscopie électronique**

### **A. Généralités sur la microscopie**

La microscopie électronique moléculaire est un moyen puissant pour observer les particules isolées telles des particules virales ou des complexes moléculaires. De nombreuses techniques de préparation de l'échantillon, d'enregistrements d'images ainsi que différentes méthodes d'analyses de données sont employées en microscopie électronique. Selon la combinaison de méthodes utilisées l'information structurale obtenue est de nature morphologique ou vise à obtenir un modèle en trois dimensions (3D) de l'enveloppe ou de la densité électronique de l'objet étudié.

La formation de l'image en microscopie électronique est la même quelque soit l'objet, sur du matériel biologique, elle est limitée par le manque de contraste de l'objet plus que par le pouvoir résolutif de l'appareil.

Il y a trois facteurs principaux qui limitent la résolution : le contraste qui est la différence d'intensité entre deux points de l'image, la préservation structurale et la sensibilité des objets biologiques à l'irradiation. Louis de Broglie notait dès 1940 la conclusion suivante :

" Quand les corpuscules incidents seront assez rapides (c'est-à-dire auront une longueur d'onde assez petite) pour que l'on puisse discerner les détails de la structure de l'atome, ils seront susceptibles dès les premiers chocs efficaces d'arracher l'atome au corps dont ils font partie et même de bouleverser sa structure en lui enlevant des électrons."

#### **1. Coloration négative**

Le contraste entre deux points est la différence de densité électronique entre ces deux points. Si la densité électronique est élevée, une proportion importante des électrons incidents seront diffusés. La densité électronique dépend fortement du numéro atomique de l'atome irradié. L'interaction électronique au niveau de l'objet dépend donc de son numéro atomique moyen appelé Z. Pour le matériel biologique, essentiellement constitué d'atomes de carbone, d'azote, d'hydrogène et d'oxygène et

de quelques éléments représentés de façon mineure, la densité électronique est faible et le contraste par rapport au vide est par conséquent faible. Diverses méthodes sont utilisées pour amplifier ce contraste, elles utilisent l'addition d'atomes de haut numéro atomique soit sur le matériel biologique (coloration positive : le contraste sera fort entre le vide et les atomes lourds) soit autour de l'objet (coloration négative : le contraste est plus fort entre Z faible-Z fort qu'entre le Z faible et le vide).

La diffusion élastique est un phénomène au cours duquel l'électron incident interagit avec le noyau atomique et sa trajectoire est déviée par la charge positive du noyau. La proportion d'électrons diffusés élastiquement dépend de la taille du noyau et les équations de Schroedinger prévoient une loi en  $Z^{3/2}$ , c'est la raison pour laquelle l'on utilise des sels de métaux lourds tels que l'acétate uranyle ou l'acide phosphotungstique. Ainsi un atome d'uranium ( $Z=92$ ) diffuse 60 fois plus les électrons que le carbone ( $Z=6$ ). Les atomes lourds utilisés dans la coloration négative doivent être solubles dans l'eau et amorphes après séchage, stables sous le faisceau et ne doivent pas interférer avec le matériel biologique. Les particules sont enrobées dans un film constitué de matériel dense aux électrons. Cette technique ne donne cependant qu'une information de surface de l'échantillon car ce qui est observé est la fraction de l'objet excluant le colorant. Aucune information structurale interne ne pourra être obtenue. De plus, la préservation de l'objet n'est pas optimale car l'utilisation de sels d'atome lourd entraîne une augmentation de la concentration saline qui peut dégrader l'échantillon. Mais l'échantillon est toutefois mieux préserver que lors du séchage à l'air car le moule de sel n'est pas sensible aux forces de tensions superficielles. De façon usuelle, sont utilisés, d'acide phosphotungstique (A.P.T.), les sels de molybdène et les sels d'uranium. Nous développerons ici la méthode de coloration négative aux sels d'uranium et plus précisément à l'acétate d'uranyle (Malech et Albert, 1979; Frank et Radermacher, 1992).

Tout d'abord, l'échantillon biologique est absorbé sur une grille recouverte d'un film de carbone pendant 1 minute. Ensuite, une solution contenant l'acétate d'uranyle 2% est ajoutée à la macromolécule à étudier. Puis l'excès de liquide est éliminé à l'aide d'un papier filtre et la grille est séchée. Une meilleure répartition des particules est

obtenue en soumettant les grilles avant adsorption à un effluage. Les grilles sont mises dans une enceinte où l'on peut faire un vide peu poussé, l'effluage sera obtenu par une décharge électrique qui ionisera les molécules d'air résiduelles, la surface des grilles sera alors hydrophile. L'acétate d'uranyle permet une grande rapidité d'infiltration et présente un grain très fin, même avec son pH très acide, les objets sont rarement détruits. Mais celui-ci précipite avec les ions OH en hydroxyde d'ammonium ainsi qu'en présence de lumière et de groupements phosphate, comme tous les sels d'uranyle.

Cette technique de coloration négative reste très utilisée pour identifier des vues caractéristiques de l'objet, obtenir un premier modèle 3D et valider la possibilité de débiter une étude en cryomicroscopie (Brenner et HORNE, 1959).

## **2. Cryomicroscopie**

La cryomicroscopie s'est développée depuis le début des années 1980 (Schwartz et Diller, 1982), (Prasad et al., 1992), (Korber et al., 1986). Elle consiste à piéger rapidement dans l'eau vitrifiée des particules à observer. L'avantage de cette technique est qu'elle s'approche le plus des conditions en solution. Le contraste provient uniquement de la particule avec le milieu, ce qui explique qu'il ne doit pas y avoir de glycérol dans le tampon de conservation de l'échantillon ce qui augmente la densité du milieu et donc diminue le contraste. Le contraste ne dépend donc pas de la répartition d'atomes de métal sur l'objet, évitant tout problème d'anisotropie de coloration, de plus la force ionique est préservée. Les macromolécules se trouvant dans une matrice hydratée, tout phénomène d'effondrement de l'objet sur lui-même est écarté. Enfin, on observe une légère cryoprotection des particules biologiques contre les dommages causés par l'exposition aux radiations. On estime qu'à la température de l'azote liquide, l'échantillon peut être irradié 2 à 3 fois plus avant d'atteindre la même perte de résolution qu'à température ambiante. En cryomicroscopie électronique les molécules peuvent ne pas être adsorbées sur un support, elles adoptent toutes les orientations. La population de molécules couvre virtuellement toutes les orientations abrogeant le besoin de tilter la grille

d'observation pour l'enregistrement de données suffisantes pour la reconstruction en trois dimensions (3D).

En cryomicroscopie les particules sont observées sur des grilles recouvertes de carbone à trous dans lequel l'échantillon va venir former une couche mince propice à la congélation rapide et à la formation de glace. L'échantillon est congelé très rapidement sur la grille afin de former une couche de glace vitreuse. Cet état vitreux correspond à un état amorphe métastable de l'eau et évite la formation de glace cristalline, opaque aux électrons.

Le support technique utilisé est appelé une guillotine (figure 9) (Dubochet et al., 1981) : ce dispositif est composé d'une cuve contenant de l'éthane liquide et d'un réceptacle à grilles. L'éthane (point d'ébullition et de fusion de  $-188.75^{\circ}\text{C}$  et  $-183.45^{\circ}\text{C}$ ) reste liquide car il est entouré de vapeurs d'azote liquide (point d'ébullition et de fusion de  $-189.85^{\circ}\text{C}$  et  $-195.85^{\circ}\text{C}$ ), la température est donc maintenue autour de  $-175^{\circ}\text{C}$ . L'échantillon est déposé sur la grille à trous, le surplus de liquide est absorbé à l'aide d'un papier filtre et la guillotine est activée. La grille plonge rapidement dans l'éthane liquide, entraînant une vitrification immédiate de l'échantillon. La grille est dès lors transférée dans le récipient de stockage. La grille est ensuite transférée du récipient à une station de transfert qui permet de maintenir la chaîne du froid, d'éviter la contamination de glace lors du transfert de la grille vers le microscope et de monter la grille sur le porte-objet. La grille est placée dans son logement dans le porte objet froid où elle est fixée à l'aide d'un anneau de maintien. Un volet en cuivre vient protéger la grille de toute contamination de glace extérieure. Le porte objet est introduit rapidement dans le microscope au niveau du goniomètre, le volet est retiré afin de permettre l'observation.



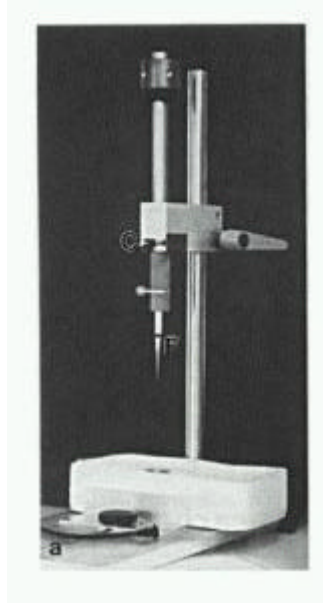


Figure 9 : Congélation rapide de la grille à l'aide d'une « guillotine ». La grille plonge dans l'éthane liquide, entraînant une vitrification immédiate de l'échantillon.

### **3. Cryocoloration négative**

La cryocoloration négative est une méthode tout à fait nouvelle et originale (Adrian et al., 1998; De Carlo et al., 2002). Il s'agit d'ajouter une solution concentrée de molybdate d'ammonium à l'échantillon à observer. La haute teneur saline (16% en masse) peut cependant induire des artefacts dont il faudra tenir compte lors de l'analyse d'images.

Récemment, la structure de l'ARN polymérase I de levure a été obtenue à 18Å de résolution par cryocoloration négative (De Carlo et al., 2003). La structure révèle une excroissance additionnelle non détectée dans les modèles 3D antérieures (Bischler et al., 2002). Toutefois ce module a pu être observé dans des vues non colorées obtenues après cristallisation en 2D de l'ARN Polymérase I (Schultz et al., 1993). Ceci indique que cette nouvelle excroissance n'est pas induite par l'ajout de colorant.

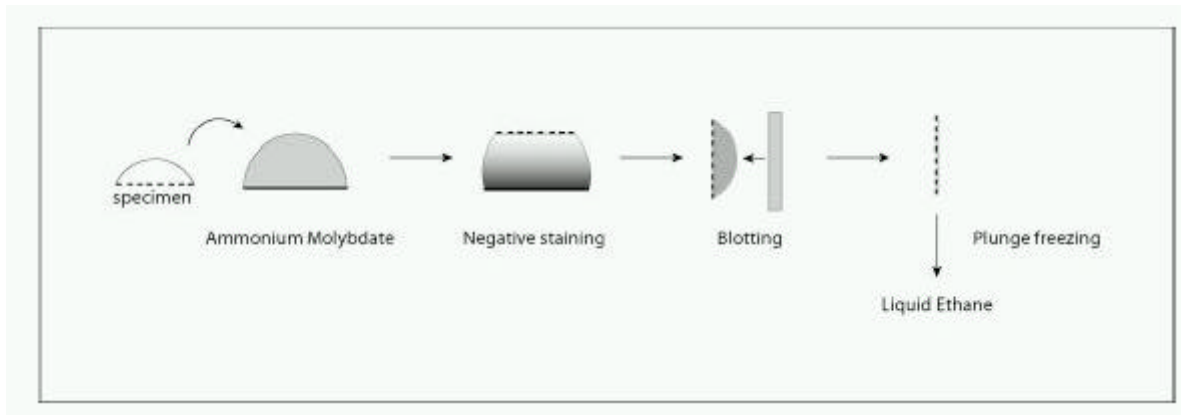


Figure 10 : La cryoloration négative : représentation schématique de la préparation des échantillons

#### 4. Coloration négative à froid

Cette méthode est une combinaison entre la coloration négative par la méthode sandwich et la cryomicroscopie ((Golas et al., 2003) voir dans "supplemented materials"). Les particules sont prises en sandwich entre deux films de carbone fin après coloration au formate d'uranyle 2%. Après environ 2 minutes à température ambiante, l'échantillon est congelé.

On estime que lorsque l'échantillon est scellé entre deux films de carbone il ne se déshydrate pas complètement ce qui explique la bonne préservation structurale.

#### 5. Cristaux bidimensionnels

La cristallisation bidimensionnelle de macromolécules biologiques aspire à former un réseau ordonné en deux dimensions (2D). Cette technique consiste à ordonner sur un film lipidique les molécules par des interactions lipide-protéine et protéine-protéine (Jesior et Wade, 1987). Les cristaux 2D ainsi obtenus sont ensuite transférés sur un support permettant l'observation en microscopie électronique. Dans ces conditions, la macromolécule est toujours visualisée sous le même angle, ce qui facilite l'opération de moyenne et permet d'atteindre plus rapidement des résolutions de 10Å et 15Å. Mais l'accès à l'information 3D nécessite d'enregistrer des images du cristal incliné dans le microscope, ce qui n'est possible que jusqu'à 60° et génère un cône d'informations manquantes.

## B. L'analyse d'image de molécules isolées

Le premier objectif de l'analyse d'image est de séparer le signal contenant l'information structurale du bruit de nature aléatoire afin d'améliorer la résolution interprétable. Pour limiter l'irradiation de l'échantillon, les images sont enregistrées à faible dose d'électrons ce qui résulte en un faible rapport signal sur bruit. Le calcul d'une moyenne de  $N$  images bruitées permet d'augmenter ce rapport, dans la mesure où le bruit gaussien est réduit d'un facteur  $\sqrt{N}$ . Toutefois, l'observation de particules isolées sur une image de microscopie permet de distinguer plusieurs formes dont on ne peut calculer la moyenne. Ces variations morphologiques peuvent être dues, soit à une orientation différente de l'objet, soit à une conformation distincte, soit à des variations d'inclusion dans le colorant. Les méthodes de partitionnement des images seront utilisées pour décrire l'objet selon ces différents facteurs et ainsi de séparer ces images en sous-populations éventuellement reliées entre-elles.

Le deuxième objectif de l'analyse d'image consiste à déterminer la direction d'observation de chaque vue en utilisant le formalisme des angles d'Euler  $\alpha$ ,  $\beta$  et  $\gamma$ . Les images moyennes obtenues correspondent à des projections 2D de la densité électronique d'un objet 3D selon la direction d'observation. Un modèle 3D de la molécule pourra être reconstruit en reprojétant les densités de toutes les images selon les angles déterminés. Le traitement des images a été réalisé avec le logiciel IMAGIC. IMAGIC est un logiciel modulaire qui permet d'analyser des images de molécules isolées ou d'objets organisés. IMAGIC a été optimisé pour la gestion de grands jeux de données correspondant à plusieurs milliers d'images moléculaires. L'analyse de telles populations d'images en 2D se fait grâce à des techniques d'alignement basées sur le principe de la fonction de corrélation croisée. La compression des données se fait par une analyse statistique multivariée et la classification repose sur le principe de la classification hiérarchique ascendante. L'analyse par « corrélation et moyenne d'image » permet donc de calculer à partir d'une population importante et hétérogène d'images de molécules isolées, des images moyennes de la molécule. Ces images moyennes correspondent à différentes orientations, conformations ou espèces moléculaires.

## **1. Pré-traitement des images**

Les négatifs obtenus sur le microscope sont tout d'abord sélectionnés au banc optique qui montre le spectre de puissance de l'image de diffraction du négatif ainsi que la présence d'un éventuel astigmatisme. Les négatifs sont ensuite numérisés et les particules sont directement sélectionnées manuellement sur un écran graphique. Pendant cette opération, une première sélection grossière et visuelle est faite en fonction de l'inclusion de la particule dans le colorant ou de sa dégradation éventuelle. Un nombre important de sous-images centrées sur des objets sont ainsi extraites. Ces images subissent ensuite un premier traitement qui vise à éliminer les variations indépendantes de la structure de la macromolécule. Un filtre de fréquences spatiales est appliqué à chaque image dans l'espace de Fourier qui permet d'éliminer une partie des basses et des hautes fréquences. L'information contenue dans les basses fréquences provient des variations de coloration, d'inclusion dans le colorant et des rampes de colorant. L'élimination des hautes fréquences spatiales n'entraîne aucune perte des détails structuraux fins car ceux-ci ne sont plus accessibles, notamment dans le cas de la coloration négative où la résolution est limitée à environ 20Å. Lors de cette étape, une normalisation sur l'intensité des pixels est également appliquée. Les densités moyennes de chaque image sont dès lors comparables entre elles. Finalement, un masque circulaire est appliqué à chaque image afin d'éliminer une partie des pixels ne décrivant pas l'objet.

## **2. Alignement en rotation et en translation**

La première étape est un centrage des particules. Cette opération ne réalise qu'un alignement en translation. Cette étape est nécessaire car dans l'étape ultérieure un pré-centrage est requis.

Dans un deuxième temps, on procède à l'alignement en rotation et en translation. L'alignement a pour objectif de recalibrer les images sur une ou plusieurs références en utilisant le coefficient de corrélation croisée comme mesure de similarité. Une classification préalable avant l'alignement en rotation vise à obtenir des images significatives du jeu de données et évite de biaiser les résultats. En effet, toutes les

images sont alignées sur une référence. Cette référence est obtenue par sommation de toutes les images individuelles en une image moyenne puis en moyennant cette image en rotation afin que seule la taille de la particule soit responsable du centrage et non sa position. Le choix de la référence est très important car il détermine l'orientation des images recalées. Elle devra être représentative de la population d'images à comparer pour les recalculer de façon à permettre leur répartition en sous-groupes représentant la même orientation. Il faut noter qu'à ce stade il est possible d'injecter des références externes, en effet dans l'analyse par cryomicroscopie de TFIIH nous avons utilisé les références d'alignement du précédent jeu de données de TFIIH en coloration négative. Au fur et à mesure de l'analyse, le biais engendré par l'introduction d'informations externes est apparu et par la suite nous avons suivi un protocole de «reference free » c'est à dire qu'à aucune étape une information externe n'a été utilisée.

### **3. Partitionnement des images**

En général, la particule présente plusieurs orientations, et le calcul direct de l'image moyenne n'améliore pas le rapport signal/bruit. C'est pourquoi des méthodes automatiques vont partitionner la population d'images en classe de projections selon des critères de ressemblance maximale. La classification des images se fait en deux étapes : par analyse factorielle des correspondances et par classification hiérarchique ascendante.

#### **3.1 L'analyse factorielle des correspondances**

Cette analyse permet de décrire une population d'images alignées dans un espace de dimensions restreintes. La détermination des vecteurs propres décrivant le maximum de la variance inter-image de la population d'images alignées permet la description de chaque image bruitée comme une combinaison linéaire des vecteurs de base. L'avantage de cette étape est la diminution du temps de calcul de l'étape de classification suivante. Cette analyse factorielle des correspondances est faite à l'intérieur d'un masque délimitant les variations de la projection à analyser.

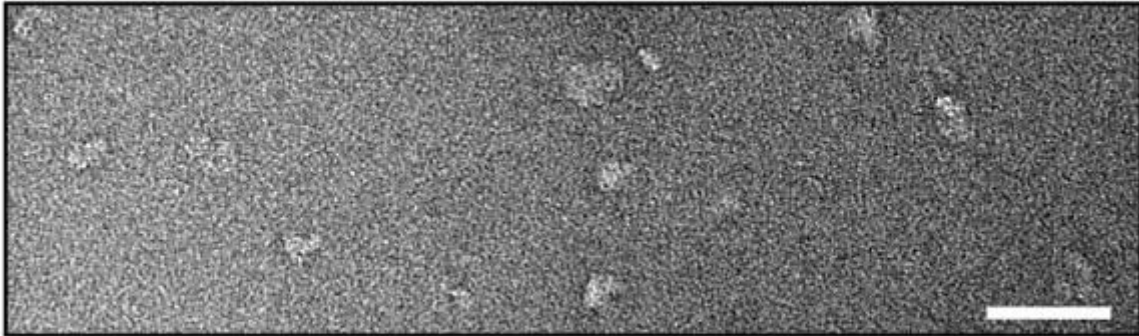
### **3.2 La classification hiérarchique ascendante**

Chaque image de la population est considérée comme une classe à part entière avec ses propres coordonnées dans l'espace de dimensions restreintes calculées par l'analyse factorielle des correspondances. La classification hiérarchique ascendante procède par fusion de deux images (ou de deux classes) selon un critère de proximité. La nouvelle classe ainsi créée est représentée dans l'espace de dimensions réduites par le centre de masse des images contenues dans la classe. Le but de notre analyse est de découvrir les classes les plus compactes (celles dont la variation interne est la plus faible) mais qui sont le plus éloignées possible (correspondant à des projections différentes de l'objet). Un bon partitionnement reflète le nombre de projections (familles) différentes composant la population d'images. Afin de mettre en évidence les différentes familles, une intervention de l'examineur est nécessaire. L'examineur détermine par observation des différentes images moyennes de classe, celles qui correspondent à la même projection de l'objet et il définit ainsi les familles. Il faut alors choisir un membre de chaque famille qui va servir de référence d'alignement en rotation pour le cycle de partitionnement suivant.

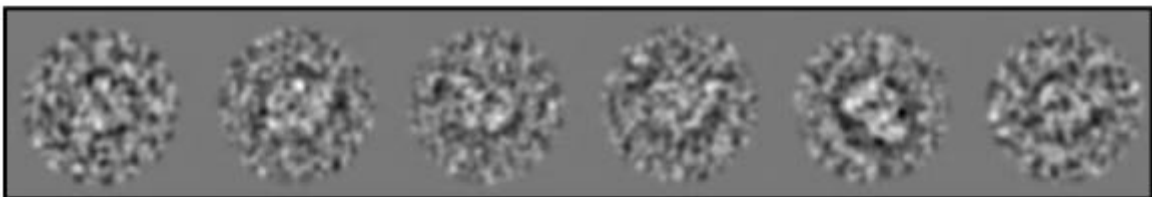
### **4. Alignement multi-références**

Il faut faire en sorte que ces deux étapes d'alignement-classification soient les plus distinctes possibles, bien qu'elles soient liées, sinon les différents cycles ne permettent pas de converger correctement, par exemple de nouvelles classes (formes) ne peuvent émerger. Afin de parfaire l'analyse, les cycles d'alignement-classification sont itérés. Les itérations successives stabilisent les familles : une famille peu représentative va perdre de l'importance alors qu'une nouvelle famille peut apparaître, bien que masquée lors du premier cycle. Cela permet également d'affiner l'alignement des images parce que finalement pour chaque image, l'orientation est unique et plus il y aura de références, plus l'orientation de la référence sera proche de celle de l'image.

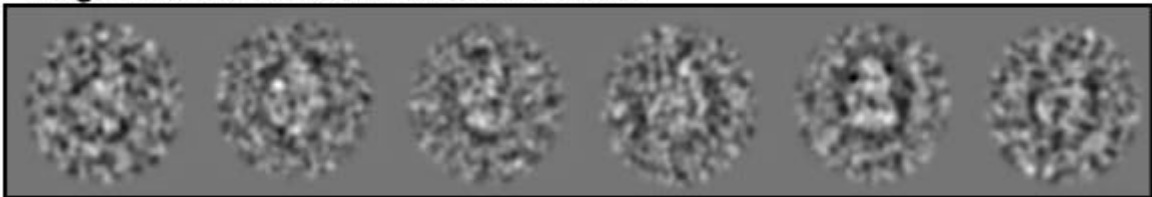
Champ de molécules isolées



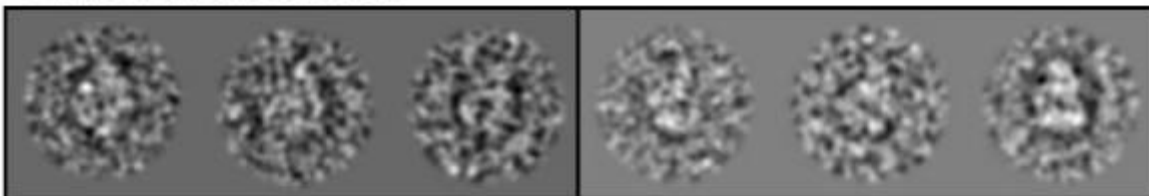
Extraction des molécules



Alignement en translation et en rotation



Partionnement en classe



Calcul d'une image moyenne

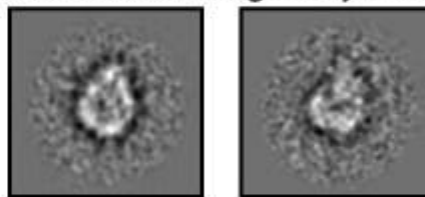


Figure 11 : L'analyse d'images en deux dimensions

Les molécules sont extraites d'un champ de microscopie électronique. Elles sont ensuite centrées, alignées puis partitionnées en classe. Une image moyenne est calculée.

### C. La reconstruction tridimensionnelle d'un objet à partir de ses projections

Une fois qu'une partition stable est obtenue, il est possible de combiner les différentes projections d'un même objet entre elles afin de reconstruire l'objet en 3D. Pour cela, il nous faut déterminer les relations angulaires relatives des différentes projections. Cette relation angulaire est caractérisée par les angles d'Euler  $\alpha$ ,  $\beta$  et  $\gamma$  qui décrivent les trois degrés de liberté en rotation (les images sont centrées pour fixer les translations). L'angle  $\alpha$  représente la rotation de la particule dans le plan de la grille de microscopie alors que  $\beta$  représente l'angle d'inclinaison et  $\gamma$  décrit la rotation dans le plan après inclinaison. Les angles  $\beta$  et  $\gamma$  caractérisent les différentes projections d'un objet. La convention utilisée par le logiciel IMAGIC est décrite dans la figure 12.

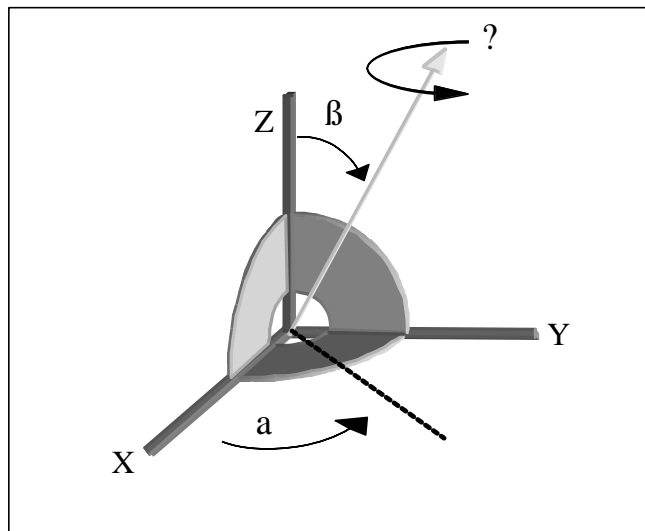


Figure 12 : Convention des angles d'Euler telle qu'elle est établie dans le logiciel IMAGIC

La reconstruction d'un modèle 3D crédible doit satisfaire plusieurs critères :

- Pour couvrir l'ensemble de l'espace de Fourier, nous avons besoin de suffisamment d'orientations différentes de l'objet, sinon le volume est déformé (Boisset et Lamy, 1991),
- La dose totale de radiation infligées à l'échantillon ne doit pas dépasser  $10 \text{ e-} / \text{Å}^2$ ,
- La reconstruction doit être représentative de l'ensemble des molécules,



- L'ensemble des différentes vues doit être décrite suivant une matrice de référence commune.

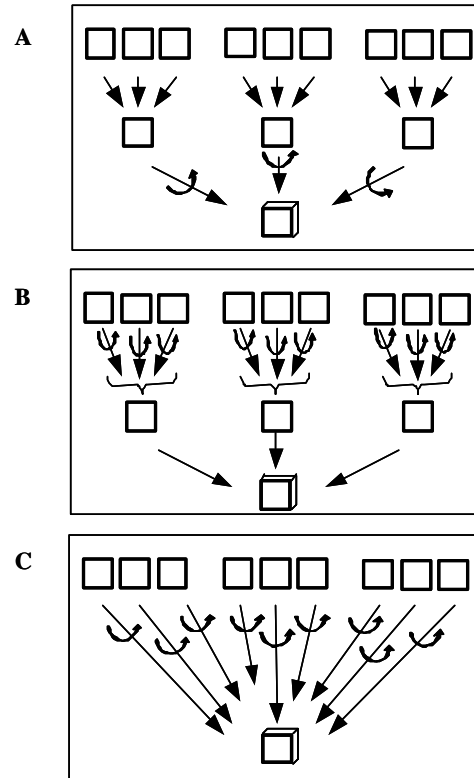


Figure 13 : Trois méthodes pour combiner l'information de projection dans une structure 3D. (A) Les molécules sont séparément reconstruites et moyennées après ré-orientation des unes par rapport aux autres. (B) Les particules sont regroupées en classes puis reconstruites, les orientations relatives sont déterminées et la reconstruction est achevée. (C) Les projections sont directement regroupées dans un modèle après détermination de leur orientation relative (Taylor et Crowther, 1992; Taylor et Crowther, 1991) (Winkler et Taylor, 1994).

## **1. Reconstruction cylindrique moyenne**

Certaines structures ne permettent pas de déterminer l'orientation de la particule. La reconstruction cylindrique est alors la meilleure possibilité. On considère que l'axe de symétrie de l'objet est dans le plan d'observation et l'on utilise une seule vue de la particule pour reconstruire une image moyenne 3D autour de cet axe de symétrie.

## **2. Méthode conique aléatoire de collection des données**

Le problème principal lié à la reconstruction d'un objet 3D réside dans la détermination de la direction de projection de chaque image expérimentale. L'utilisation des angles Eulériens ( $\theta$ ,  $\phi$ ,  $\psi$ ) permet de caractériser des directions de projection.

La technique de reconstruction conique aléatoire permet de préserver la structure de l'échantillon en réduisant à seulement deux le nombre d'expositions de l'objet aux électrons et à produire une série de projections dont les angles eulériens peuvent être déterminés expérimentalement. Cette méthode permet d'enregistrer les données suffisantes pour reconstruire en 3D l'objet observé. Deux séries d'images sont collectées : la première, non tiltée, permet d'extraire les projections pour la reconstruction, la seconde, tiltée, permet de séparer les particules suivant leur orientation (alignement) et de déterminer les azimuts pour les particules (classification). Il faut que les particules soient orientées au départ et on ne sélectionne que celles qui ont le même orientation. Les angles d'alignement en rotation nous renseigne sur l'angle azimuthal, un des angles d'Euler, le deuxième étant donné par l'angle de tilt.

Plusieurs limitations surviennent à cette technique : (i) la classification des images est effectuée à partir des particules observées non tiltées, ce qui peut là donner des résultats ambigus, (ii) les azimuts sont déterminés dans un second temps après ré-exposition des particules et donc après un éventuel dommage de celles-ci. Ces limitations sont levées par l'utilisation de processus d'affinement et de reconstructions itératifs.

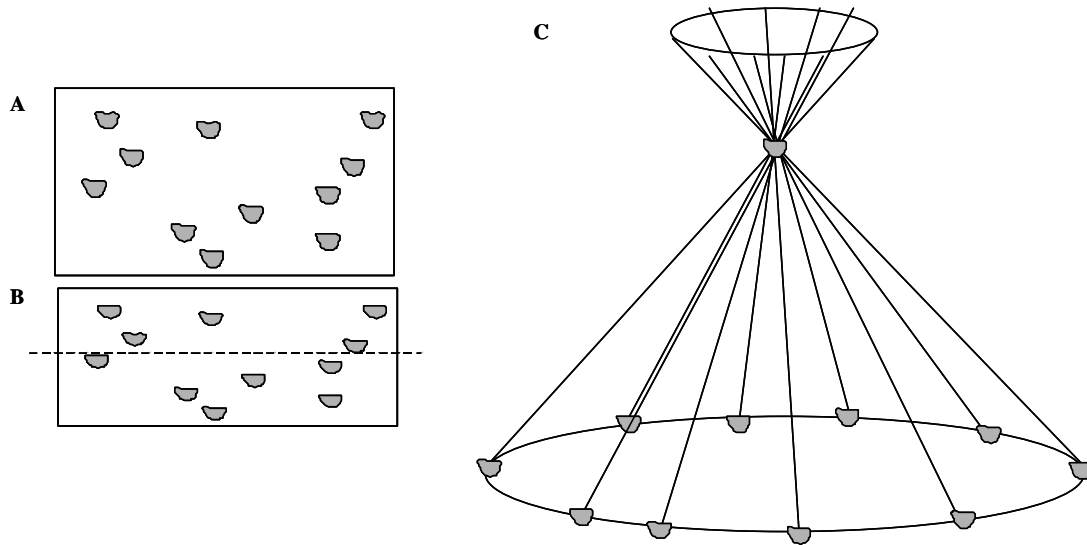


Figure 14 : Principe de la méthode conique aléatoire de collection de données. Une série d'images non tiltée (A) et tiltée (B) est enregistrée. La position équivalente (C) est ensuite déterminée.

### 3. Méthode des lignes communes

La méthode des lignes communes représente des lignes le long desquelles, en accord avec le théorème de projection, la transformée de Fourier des projections est identique en l'absence du bruit (Crowther et al., 1970). Si, arbitrairement, on représente deux projections d'un même objet dans l'espace de Fourier, ces projections se croisent en une ligne commune à travers l'origine (figure 15). Si l'on ignore l'orientation relative de ces projections, nous pouvons trouver la ligne commune en comparant chaque section d'une transformée 2D d'une orientation à chaque section d'une transformée de l'autre orientation. La comparaison est effectuée par une corrélation croisée qui dans l'espace de Fourier est équivalente à la somme de produit conjugué. Ceci se traduit par un maximum lors d'une correspondance (van Heel, 1987).

Cette méthode a été développée à l'origine en espace de Fourier pour la reconstruction de virus à symétrie icosaédrale (DeRosier et Moore, 1970) (Crowther, 1971). Elle a ensuite été généralisée, en espace réel, pour des objets à faible degré de symétrie donnant naissance au concept des sinogrammes .

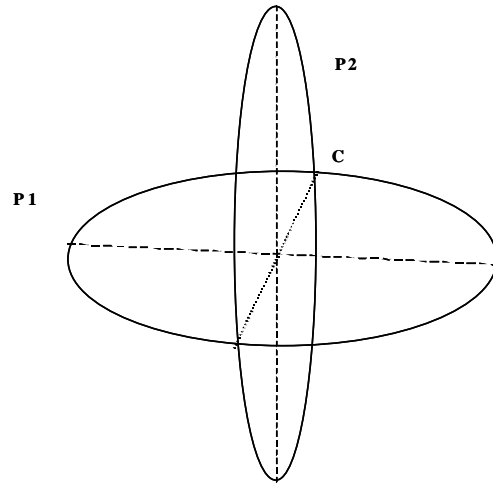


Figure 15 : Les principes des lignes communes : 2 projections P1 et P2 du même objet ont en commun à l'intersection à l'origine, une ligne commune C. Sur cette ligne les transformées de Fourier sont identiques.

#### D. Limitations du microscope et fonction de transfert de contraste

Dans leur principe, les microscopes optiques et électroniques à transmission sont relativement similaires. Dans le cas du microscope électronique, une source produit un faisceau d'électrons qui peut être dévié par des lentilles électromagnétiques à cause de leur charge négative. La microscopie électronique fournit par conséquent une image directe de l'objet comme dans la microscopie photonique. La différence majeure réside dans le fait que les électrons ne se propagent que dans le vide et, dans le cas d'un échantillon mince, ne sont pas absorbés par la matière.

Rappelons maintenant les principes de base de la formation d'une image. La formation d'une image à travers une lentille se fait en deux étapes. Soit un objet ABC, dans un premier temps l'image de diffraction de l'objet est formée dans le plan focal de la lentille, elle correspond à la transformée de Fourier de l'objet (TF), dans un second temps l'image de l'objet A'B'C' est formée par transformée de Fourier inverse (TF<sup>-1</sup>) de l'image de diffraction (figure 16).

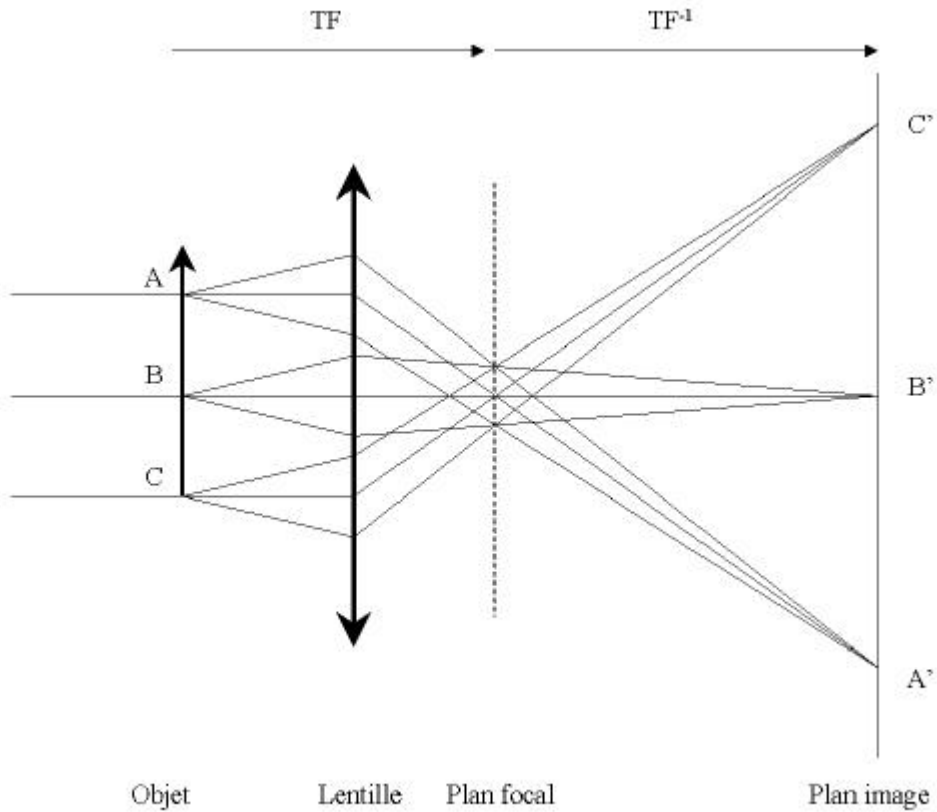


Figure 16 : Principe de formation de l'image d'un objet.

TF= transformée de Fourier,

TF<sup>-1</sup>= transformée de Fourier inverse

Pour préserver la structure des échantillons observés en cryomicroscopie il est impératif de les observer à faible doses d'électrons. Les images obtenues sont alors très bruitées à cause de leur faible rapport signal sur bruit et les techniques actuelles d'analyse d'images permettent de remédier à ce problème. Il s'agit de récupérer le signal perdu et dans un premier temps, en corrigeant les imperfections du microscope.

Les limitations dues aux imperfections des lentilles électromagnétiques sont : l'aberration de sphéricité, l'aberration chromatique et l'astigmatisme. Le premier défaut se traduit par une convergence plus forte des électrons traversant les bords des lentilles qu'au centre des lentilles électromagnétiques. En pratique, pour minimiser ce disque de confusion, il faut réduire l'angle de demi-ouverture. La seconde imperfection est l'aberration chromatique des lentilles et elle est directement

reliée à la qualité du canon à électrons du microscope. L'astigmatisme des lentilles, quant à lui, peut être directement corrigé à fort grandissement en observant un film de carbone.

Il existe une fonction de transfert de contraste (FTC) qui prend en compte à la fois l'aspect ondulatoire du contraste dû aux interactions des électrons avec l'échantillon mais aussi les imperfections du microscope. Les imperfections du microscope conduisent à une FTC oscillante traduisant des inversions de contraste entre certaines gammes de fréquences spatiales et qui équivaut à une perte de signal à haute résolution. L'intérêt de corriger la FTC est donc de récupérer une partie du signal perdu. Les étapes pour parvenir à cette fonction sont décrites grossièrement : tout d'abord les négatifs utilisés pour l'analyse d'image sont observés au banc optique. Le banc optique est un instrument qui permet de visualiser le spectre de puissance (intensité = (amplitude)<sup>2</sup>) de la transformée de Fourier de l'image, et donc la FTC. Celle-ci se manifeste par une alternance d'anneaux concentriques clairs et sombres : les anneaux de Thon. Les anneaux sombres correspondent aux fréquences spatiales non transmises et la position du premier anneau sombre détermine la résolution jusqu'à laquelle les fréquences spatiales sont transmises sans inversion de phase. La forme des anneaux renseigne sur la présence d'astigmatisme et sur une dérive de l'échantillon lors de la prise d'image. Ce type de négatif n'est pas utilisé dans la correction de la fonction de transfert de phase.

La FTC est corrigée pour chaque négatif. Les corrections portent sur les amplitudes et les phases de la fonction. Une première correction correspond à l'inversion des phases négatives qui conduisent, si elles ne sont pas corrigées, à une inversion du contraste des fréquences spatiales concernées. En observant la FTC, les fréquences spatiales ne sont pas toutes transmises avec la même amplitude. Une fonction particulière appelée filtre de Wiener restaure à 1 la transmission des fréquences dont l'amplitude n'est pas nulle. Cependant, des images prises à diverses valeurs de défocalisation sont nécessaires pour combler les fréquences spatiales non-transmises omniprésentes à cause de la nature oscillante de la FTC.

La fréquence des oscillations de la courbe de la FTC est déterminée par le paramètre de défocus. Pour chaque négatif, le défocus déterminé est appliquée à l'ensemble du négatif.

### **E. Mesure de la résolution des reconstructions**

La méthode utilisée est la méthode de corrélation des enveloppes de Fourier pour FSC (Fourier shell correlation) (Harauz et al., 1987). Elle repose sur la comparaison de deux volumes à des fréquences spatiales croissantes et sur la mesure de la corrélation entre les enveloppes de Fourier. Cette corrélation diminue des basses vers les hautes fréquences et c'est cette propriété qui permet d'identifier la limite de résolution. Elle correspond au passage du coefficient de corrélation sous une valeur seuil (en ordonnée) dépendante de la fréquence spatiale (en abscisse). Cette valeur seuil est défini soit par 0.5 soit par 3s. Lorsque la valeur seuil est considérée comme à 0.5, le rapport signal sur bruit est de 1 et la limite de résolution est la fréquence spatiale à laquelle la courbe de FSC passe sous le seuil de 0.5.

Lorsque l'on considère le seuil à 3s, la limite de résolution est la fréquence spatiale à laquelle la courbe de FSC coupe une seconde courbe correspondant à trois écarts types au dessus de la corrélation de deux volumes ne contenant que du bruit (Orlova et al., 1997). Les deux valeurs de résolutions FSC0.5 et FSC 3s sont donnés pour chaque reconstruction 3D.

## II. Etude du facteur TFIIH par microscopie électronique

### A. Objectifs du travail

A mon arrivée au laboratoire, une analyse des images de molécules individuelles de TFIIH endogène colorées négativement à l'acétate d'uranyle avait été réalisée à une résolution de 3,8 nm (Schultz et al., 2000b). Un des objectifs de mon travail a été d'améliorer la résolution de la structure 3-D. A cet effet nous avons utilisé la cryomicroscopie, une méthode qui préserve l'hydratation de l'échantillon dans le vide du microscope. De plus les images ont été acquises sur un microscope électronique équipé d'un canon à effet de champ (FEG) installé au laboratoire fin 2002.

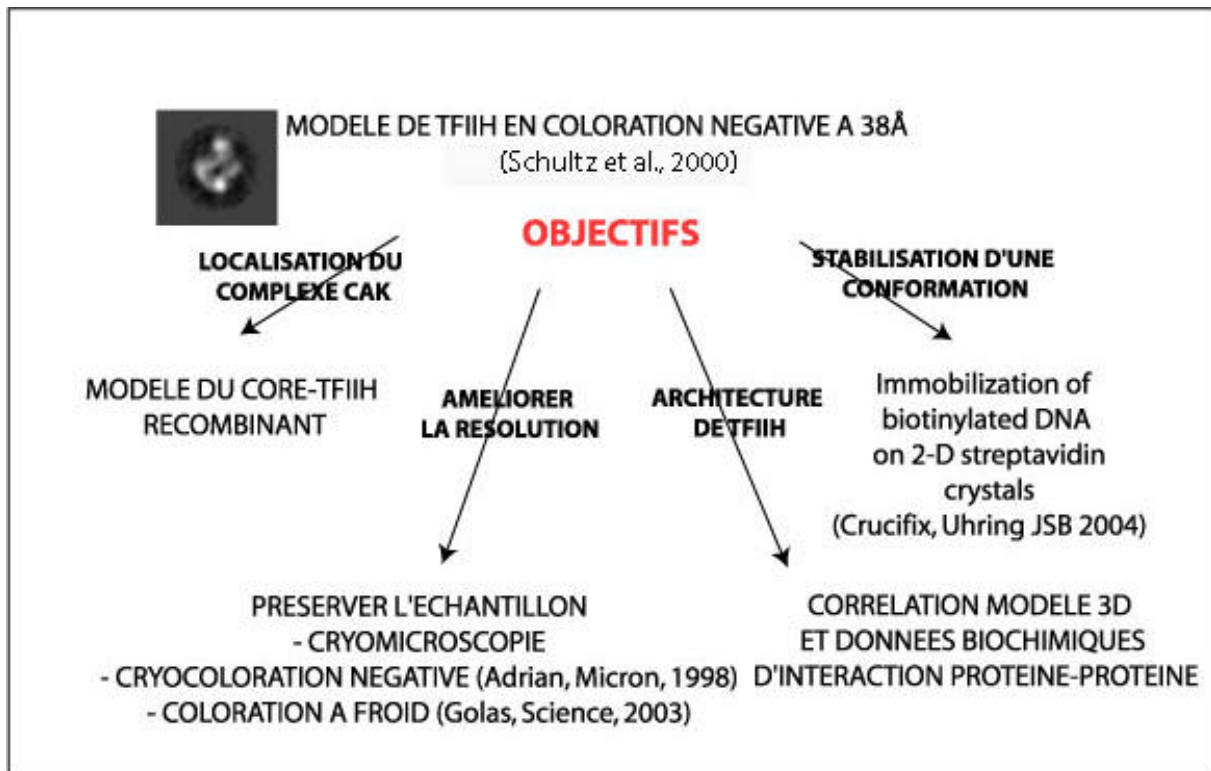






Figure 17 : Le microscope électronique à canon à effet de champ (FEG) Tecnai F20 200kV

## **B. Modèle de TFIIH en microscopie par coloration négative à froid**

### **1. La coloration négative à froid**

Un des objectifs de mon travail a donc été d'améliorer la résolution de la structure 3-D du facteur TFIIH. A cet effet nous avons dans un premier temps utilisé la méthode de choix pour l'amélioration de la résolution d'une structure 3D : la cryomicroscopie qui préserve l'hydratation de l'échantillon dans le vide du microscope. Cette méthode est très utilisée pour la résolution de structures de protéines en microscopie électronique (Ludwig et al., 2003) (Nield et al., 2003). Pourtant, dans le cas de TFIIH nous avons rencontré bon nombre de difficultés. Le comportement du facteur TFIIH n'est pas en adéquation avec cette méthode : les particules de TFIIH congelées présentent un contraste trop faible pour sélectionner des formes caractéristiques et dans l'état actuel nous n'avons pas pu aboutir avec cette méthode de préparation de l'échantillon.

Dès lors, nous avons testé la coloration négative à froid (Golas et al., 2003). Brièvement, l'échantillon contenant le facteur TFIIH endogène actif à une concentration d'environ 300µg/µl est déposé sur une grille à film de carbone continu, fixé à la glutaraldéhyde (0.1%), coloré au formate d'uranyle 2% et partiellement séché avec du papier filtre. Les particules sont ensuite prises en « sandwich » par dépôt d'une nouvelle couche de carbone puis congelées dans l'éthane liquide. Le séchage se fait d'un seul côté, d'après Cyrklaff (Cyrklaff et al., 1990) un effet d'absorption des particules à l'interface air-eau permet de concentrer l'échantillon. La grille est ensuite transférée dans le microscope grâce à une platine froide qui va maintenir la température de la grille à une valeur d'environ -170°C afin qu'il n'y a pas formation de glace cristalline. La prise d'images se fait selon la technique du « low-dose » : aucune irradiation du champ n'est faite avant la prise d'images. L'acquisition d'images se fait en présence d'une fourchette et d'un anti-contaminateur froids ce qui limitent la contamination par la vapeur d'eau résiduelle du microscope et la dégradation de l'échantillon par le faisceau d'électrons.

Les micrographies sont prises à deux valeurs de défocalisation. Une prise à une forte valeur de défocalisation facilite la visualisation des particules mais le transfert d'information est meilleur sur le micrographe pris à faible valeur de défocalisation. Les micrographes sont sélectionnés au banc optique en fonction de l'absence d'astigmatisme puis numérisés à 3Å.

Environ 10303 particules du facteur TFIIH ont été sélectionnées manuellement. L'amélioration du contraste par la méthode de coloration à froid est significative, de plus l'ajout du formate d'uranyle 2% ne semble pas, à première vue, perturber la structure de la molécule (figure 18).

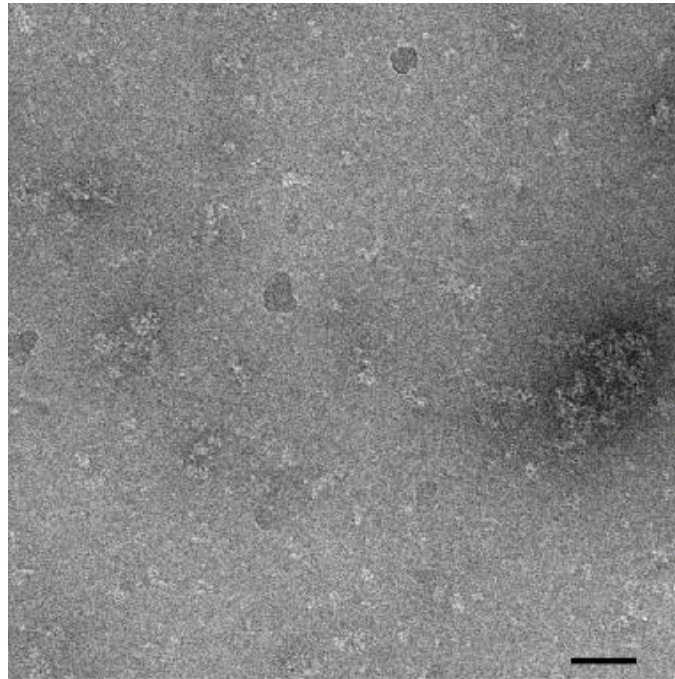


Figure 18 : Champ de molécules de TFIIH après coloration à froid. La barre d'échelle représente 100nm.

## **2. Sélection des particules**

La sélection des particules est une étape très importante. Il est délicat d'identifier et de sélectionner les molécules de TFIIH sur un champ de microscopie électronique. En effet plusieurs formes apparaissent sur les micrographies et il est impératif de sélectionner des formes homogènes en grand nombre. Nous avons trié les particules de TFIIH d'après le critère de taille, les molécules trop petites ou allongées n'ayant pas été sélectionnées (figure 19).

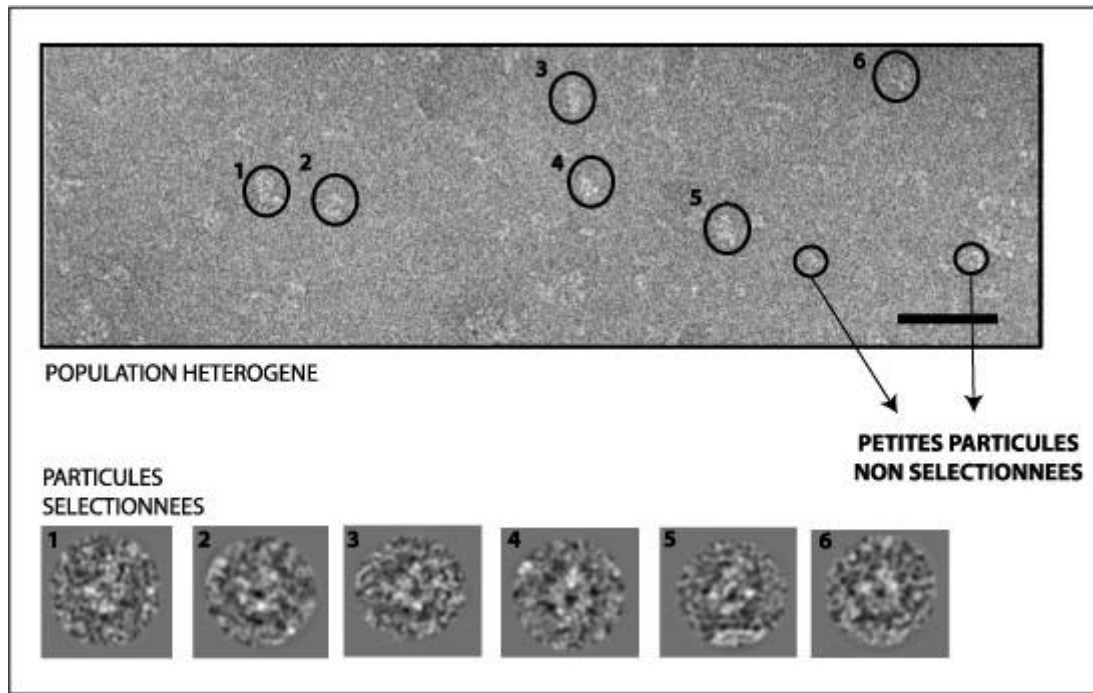


Figure 19 : Extraction des molécules de TFIIH. La population est hétérogène, les particules sont sélectionnées d'après le critère de taille (en corrélation avec le modèle obtenu en coloration négative). Les molécules allongées ou de petites tailles ne sont pas sélectionnées. La barre de taille correspond à 100nm.

### 3. L'analyse d'images

A ce stade, il est impératif de sélectionner un maximum de particules afin de disposer de toutes les vues qui reflètent l'échantillon et ainsi espérer pouvoir identifier les sous-classes représentatives. Le jeu de données a été analysé et après extraction des particules les images sont filtrées dans l'espace de Fourier afin d'éliminer les basses fréquences spatiales. Les hautes fréquences spatiales, contenant les informations structurales, sont conservées. La correction de la fonction de transfert de contraste est appliquée. La distribution de densité des images est normalisée et chaque image est masquée. Plusieurs cycles consécutifs d'alignement-classification sont nécessaires pour stabiliser le partitionnement des images. Cette stabilisation se traduit par le fait qu'aucune nouvelle projection n'apparaît lors du calcul des images moyennes de classe. Les 10303 sont réparties en 78 classes afin d'améliorer le rapport signal sur bruit (Orlova, 2000). Certaines classes obtenues sont caractéristiques du facteur TFIIH, en particulier celles qui

présentent une particule avec un anneau central et une protusion au sommet. Les classes bien définies présentent certains détails structuraux. Les particules ayant été absorbées sur un support, certaines classes adoptent une orientation préférentielle (pour exemple voir les classes 1, 5, 6, 7 (\*) de la figure 20)

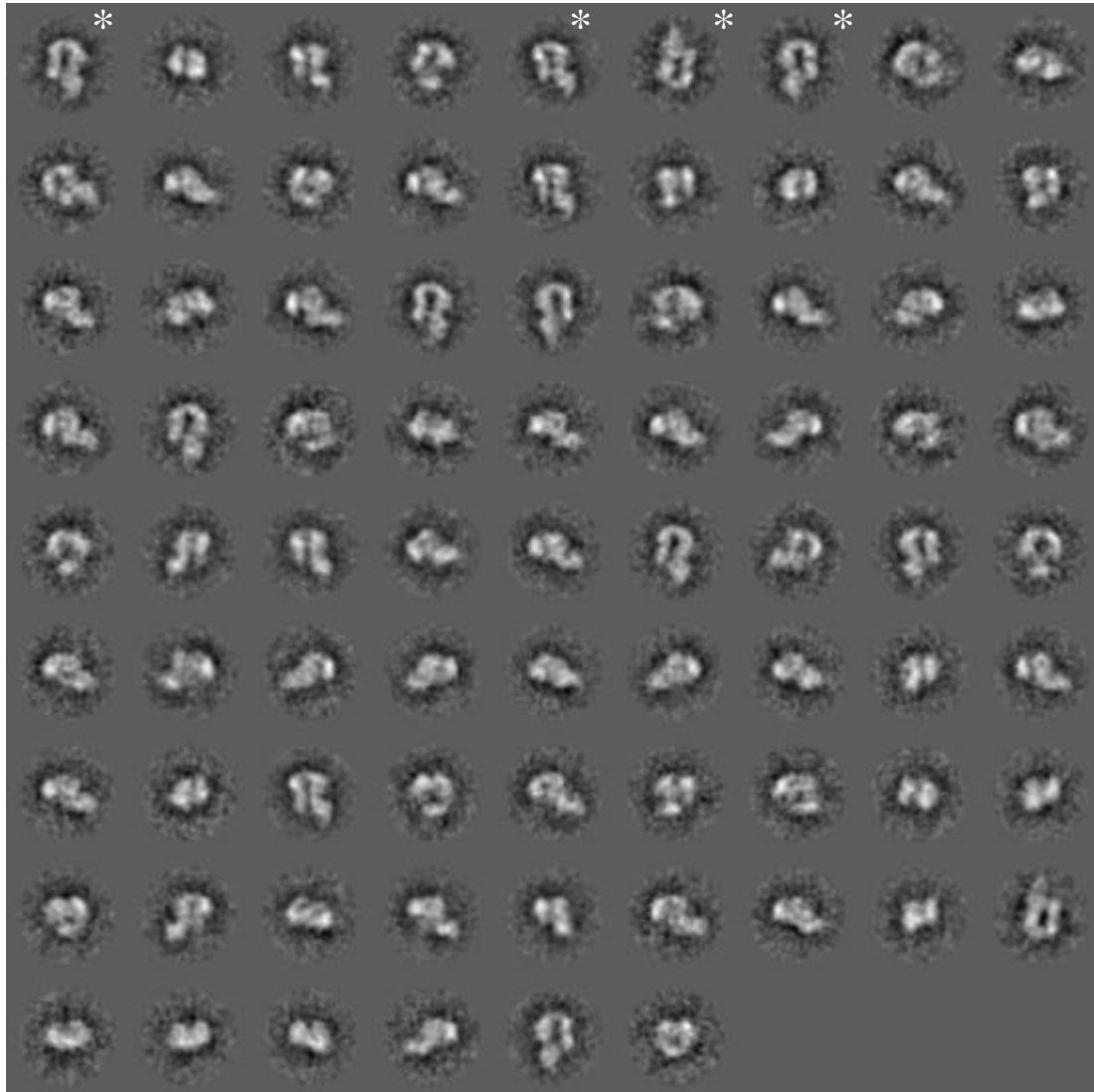


Figure 20 : Galeries d'images moyennes de classe de TFIH observées en coloration à froid.

Il est important de souligner maintenant la méthodologie suivie lors de l'analyse d'image de TFIH. Nous disposons en effet d'un précédent modèle de TFIH en coloration négative (Schultz et al., 2000f) et donc des projections de ce modèle 3D. Dans un premier temps nous avons décidé d'utiliser les projections de ce modèle

comme références d'alignement pour le jeu de données de TFIIH obtenu en coloration à froid. Nous nous sommes rapidement rendu compte que cette introduction de références biaisait le modèle. Il nous est apparu essentiel dans un second temps d'effectuer l'analyse d'image sans introduire d'information externe au jeu de données. L'analyse s'est ainsi faite avec un protocole « reference free ». Après centrage des particules les classes les plus représentatives sont extraites et utilisées pour l'alignement des particules. Un premier modèle est calculé. Les projections du modèle 3D sont ensuite utilisées pour un nouveau cycle d'alignement/classification. Après plusieurs cycles d'affinement un modèle 3D stable est obtenu.

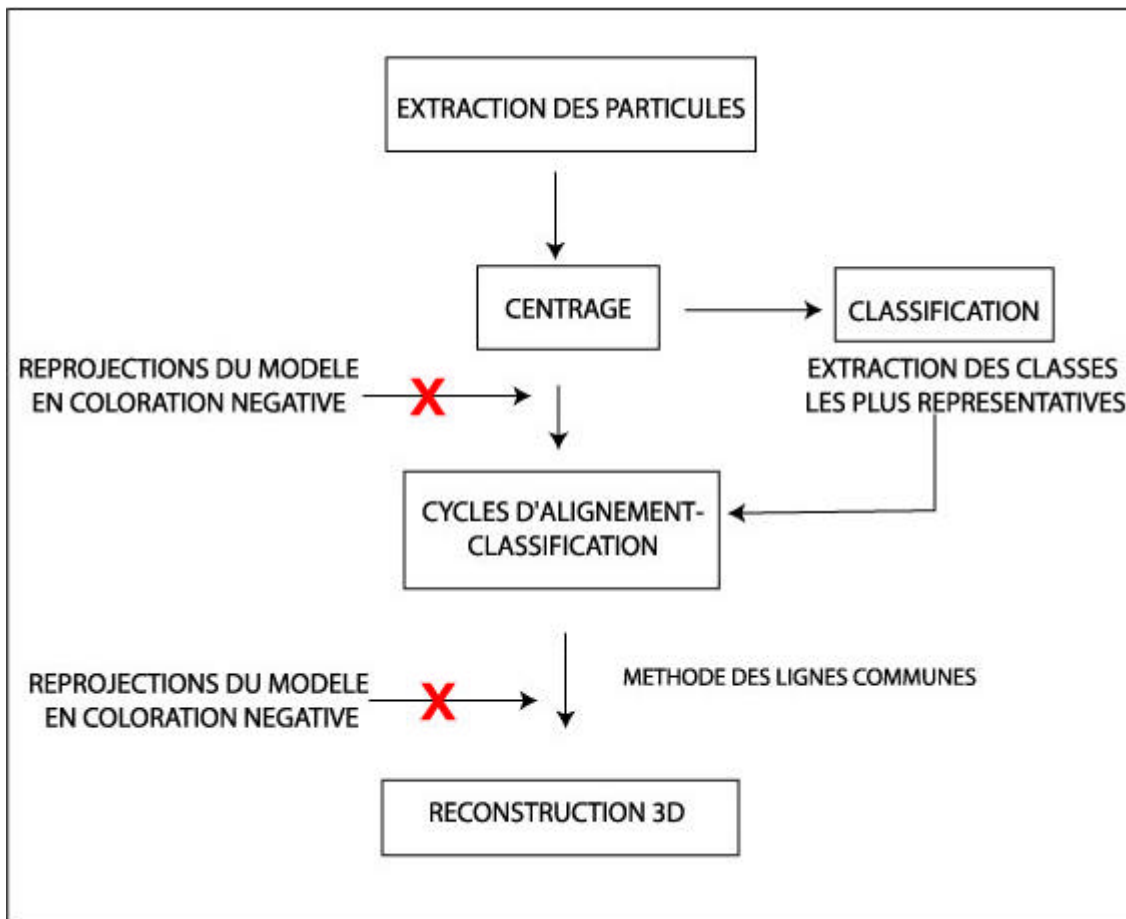


Figure 21 : Analyse d'image de TFIIH. Les différentes étapes de l'analyse sont indiquées. La croix rouge symbolise les étapes à proscrire lors de l'analyse d'images de TFIIH.

Le modèle en trois dimensions a été reconstruit selon la méthode des lignes communes. La figure 22 représente la structure 3D du facteur TFIIH obtenu en coloration à froid. Le facteur TFIIH est une particule allongée de dimension 16x12.5x7.5nm comprenant un anneau central de 2.6 à 3.4nm. La structure est similaire au modèle obtenu en coloration négative. Certains détails structuraux sont mieux définis que dans le précédent modèle et ceci illustre le gain de résolution apporté par cette nouvelle méthode. Précédemment des études d'immunolocalisation des sous-unités de TFIIH ont été effectuées, en combinant ces marquages avec les études biochimiques d'interaction protéine-protéine nous proposons un modèle général sur l'architecture quaternaire du complexe (voir §III).

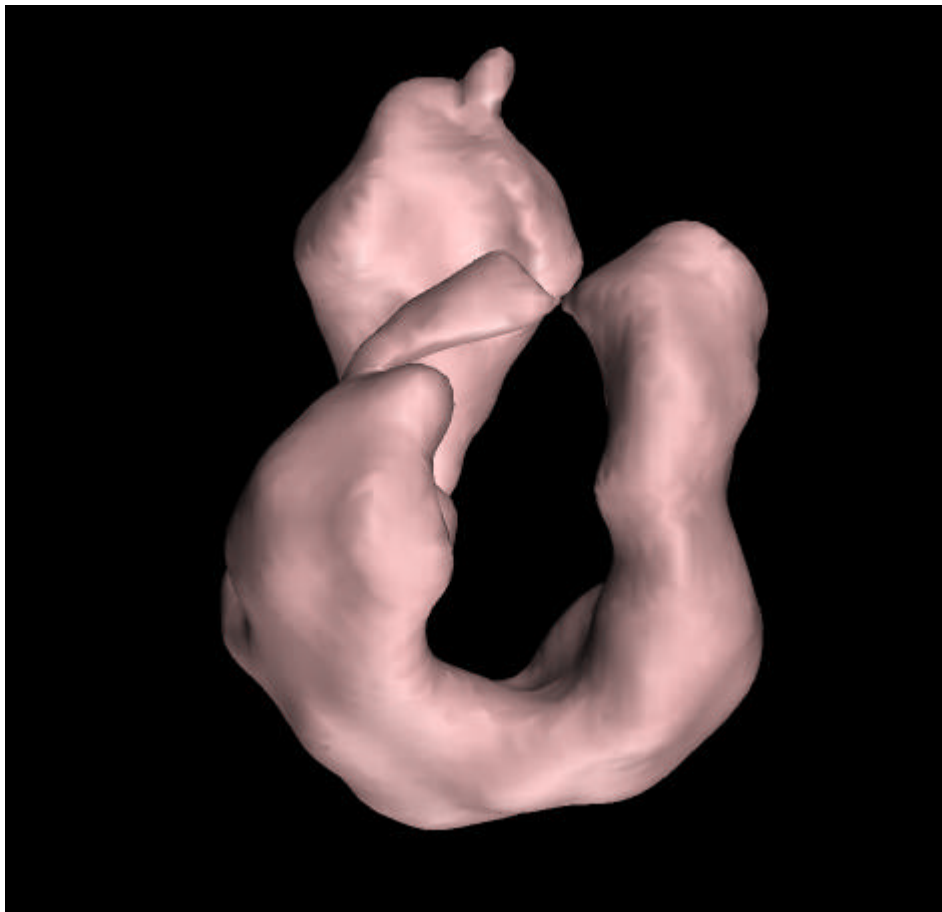


Figure 22 : Reconstruction 3D du modèle de TFIIH obtenu en coloration à froid

L'étude de TFIIH par microscopie électronique nous a confronté à plusieurs problèmes. Le problème majoritaire étant dû au comportement intrinsèque de TFIIH. Nous avons rencontré les plus grandes difficultés pour observer des préparations, par ailleurs hétérogènes, de ces complexes qui ont tendance à se dénaturer. Dans d'autres situations, les protéines se sont agrégées au cours de la préparation de l'échantillon. Ces observations nous ont conduit à penser que la préservation structurale de l'échantillon n'était pas optimale notamment en présence de l'interface air/tampon créée avant la congélation. TFIIH est un complexe fragile et il apparaît crucial d'adapter les méthodes à son observation en microscopie électronique. Dans ce cas de figure une innovation technologique nous semblait indispensable pour (i) prévenir la dénaturation de surface de ce complexe protéique particulièrement sensible (ii) concentrer les protéines peu abondantes et (iii) moduler l'environnement de la protéine afin de préserver son activité biologique. Un développement technologique consistant à immobiliser une molécule d'ADN sur un support adéquat pour l'observation ultérieure du complexe ADN-protéine en cryo-microscopie électronique est présenté dans le Chapitre IV (Article 2).

De plus, un nombre important de particules est nécessaire à la mise en évidence d'une famille faiblement représentée. Dans notre cas, la population de particules est hétérogène et nous n'avons voulu sélectionner que les particules présentant une forme et une taille similaire au précédent modèle. En effet des données biochimiques et des observations microscopiques indiquent que le facteur existe sous au moins deux états (avec et sans le domaine CAK) ce qui complique l'analyse. L'identification de sous-complexes, si tenter qu'il n'en existe que deux, a été abordée par des protocoles de tri. Après classification, il s'agissait de séparer la population en deux formes distinctes : le core-TFIIH et le TFIIH entier. Ce type d'analyse n'a pas abouti, il apparaît nécessaire d'augmenter considérablement le nombre de particules pour espérer différencier deux sous-populations.

Dans le cadre de nos travaux, nous supposons un mouvement de la protubérance située au sommet de la structure et qui a été identifié comme le sous-complexe CAK (Schultz et al., 2000g). L'enjeu sera d'identifier ces différents états au sein d'une population mixte de molécules figées par congélation à un instant  $t$  et de mettre en



évidence de tels mouvements. L'enregistrement direct de paramètres propres à chaque molécule par microscopie électronique suivie d'une analyse statistique rigoureuse devrait permettre l'identification de telles sous-populations. L'analyse de variance des images conduira à une description détaillée des changements conformationnels du complexe pour peu que deux états métastables soient identifiés, comme l'a été montré dans le cas de l'ARN polymérase I de levure (De Carlo et al., 2003).

Les états intermédiaires des complexes macromoléculaires sont présents de manière transitoires dans la cellule. D'où la difficulté d'observer ces états. Zhao et coll. (Zhao et Craig, 2003) ont développé une nouvelle méthode permettant la capture de ces états par ajout d'acétate d'uranyle ou d'acide tannique. L'acétate d'uranyle est connu pour être un bon fixateur, stabilisant les protéines, les acides nucléiques et les lipides auxquels il se lie via des interactions ioniques avec les groupes chargés négativement. La préservation des structures est clairement établie par la comparaison des structures 3D obtenues en coloration négative comparées à la cryomicroscopie et à la diffraction aux rayons X comme dans le cas de l'observation d'un changement conformationnel de l'ARN polymérase II d'*Escherichia coli* (15Å) comparée à la structure atomique de *Thermus* (Darst et al., 2002). Les auteurs décrivent l'utilisation d'une méthodologie utilisant l'acétate d'uranyle et l'acide tannique pour la fixation de structures moléculaires intermédiaires dans un laps de temps extrêmement rapide, de l'ordre de 10ms.

## C. Purification et caractérisation du complexe core-TFIIH recombinant

Je présenterais ci-après l'étude du sous-complexe core-TFIIH (rIIH6) recombinant. Certaines des sous-unités étaient disponibles au laboratoire, pour le clonage des autres, se référer au Chapitre Matériel et Méthodes. La mise en place d'un protocole de purification du complexe rIIH6 actif nous a permis de caractériser le complexe en spectrométrie de masse et de reconstruire un modèle 3D en microscopie électronique.

### 1. Purification de complexe recombinant core-TFIIH dans les cellules d'insectes

Les virus exprimant les différentes sous-unités du core-TFIIH ou rIIH6 (XPB, XPD, p62, p52, p44, p34F) sont co-infectés dans les cellules d'insectes *Sf9*. Les cellules d'insectes infectées, correspondant à 1 litre de culture, sont récoltées après 48 heures. Les cellules sont resuspendues dans un tampon de lyse (Tris-HCl 20mM pH 7.8, NaCl 250mM, mercaptoéthanol 2mM) et cassées mécaniquement à l'aide d'un potter. Après une ultracentrifugation d'une heure à 4°C à 45000g, la fraction soluble contenant les protéines est incubée avec une 1ml de résine Héparine pendant 2 heures à 4°C. Un premier lavage à 0.22M NaCl permet d'éliminer les complexes incomplets. Le complexe est élué à 0.4 M NaCl et la fraction est incubée avec les billes anti-flag (p34 est fusionné à un peptide flag) ou avec les billes préalablement immobilisées avec l'anticorps dirigé contre la sous-unité p34. Dans les deux cas, la complexe est immunopurifié via la sous-unité p34. Le complexe est finalement élué par compétition avec le peptide Flag ou avec le peptide mimant la partie C-terminale de p34.

Les fractions contenant les complexes recombinants sont déposés sur gel SDS 12.5% et révélés au bleu de Coomassie. Les six bandes visualisées sur gel correspondent aux six sous-unités du complexe core-TFIIH : XPB, XPD, p62, p52, p44, p34. Il n'y a pas de présence de contaminants ce qui nous laisse penser que la fraction déposée est relativement pure. La quantité de complexe est approximativement estimée par comparaison avec une quantité connue d'albumine bovine. Nous disposons d'environ 150µg de complexe recombinant core-TFIIH.

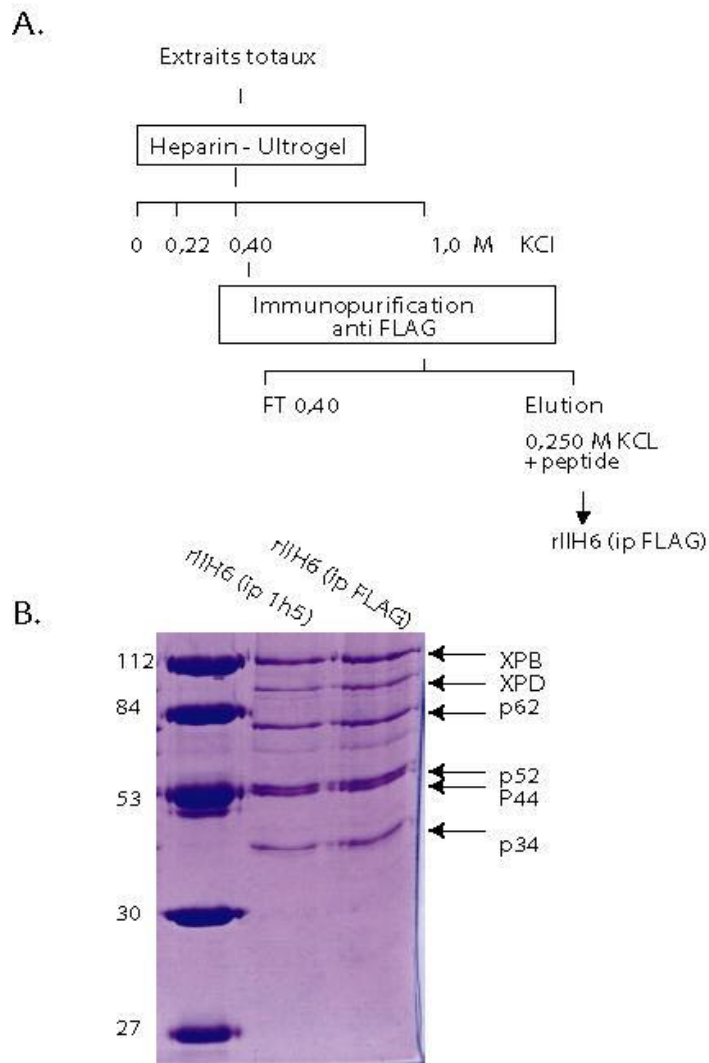


Figure 23 : Purification du complexe recombinant core-TFIIH. (A) Protocole de purification : le complexe est immunopurifié via l'étiquette FLAG fusionnée à la sous-unité p34 (rIIh6 (ip FLAG)) ou par immunopurification avec un anticorps dirigé contre la sous-unité p34 (rIIh6 (ip 1h5)) et précédemment immobilisé sur des billes. (B) Dépôt des complexes recombinants sur gel SDS 12.5% et coloration au bleu de Coomassie.

## 2. Test d'activité sur le complexe recombinant core-TFIIH

Afin de tester l'activité du complexe recombinant rIIH6, nous avons purifié le complexe avec différentes protéines de fusion : Histidine (His) tag sur XPB ou p44 ou Flag sur p44. Ainsi l'influence des différents tags sur l'activité du complexe est évaluée. Dans un premier temps, une comparaison de la production du complexe est effectuée dans deux lignées de cellules d'insectes : les cellules *Sf9* et les cellules High Five. La purification de six complexes composés de différents virus est effectuée dans les deux types de lignées (figure 24). La purification se fait en deux étapes, avec une résine Héparine suivie d'une résine Talon. Les différents complexes rIIH6 sont déposés sur gel SDS 12.5% et colorés au bleu de Coomassie. La quantité de complexes diffère selon la composition virale utilisée, toutes les sous-unités sont présentes. Cette différence de niveau de purification peut être expliquée par le fait que le tag histidine s'est dégradé sur certains vecteurs de transfert, ceci entraînant une baisse de l'affinité du complexe pour la résine talon.

Dans un second temps, les différents sous-complexes rIIH6 sont évalués en test de transcription *in vitro* (voir Matériels et Méthodes) et en test de réparation *in vitro*. Les complexes rIIH6, purifiés à partir de cellules *Sf9* ou de cellules High Five, sont actifs en transcription après ajout du complexe CAK (Cdk activating kinase), ce dernier étant nécessaire à la transcription *in vitro* (Saiz et al., 2002). Le complexe CAK est produit dans les cellules d'insectes et purifié comme décrit (Tirode et al., 1999). Le niveau de transcription des complexes recombinants est équivalent au niveau transcriptionnel du facteur TFIIH endogène. Il n'y a pas d'influence des différents tags d'affinité sur l'activité en transcription des complexes core-TFIIH recombinants.

Dans un test de réparation *in vitro*, les complexes recombinants sont actifs lorsque purifiés à partir des cellules High Five. Il faut noter que le niveau en réparation des complexes recombinants est inférieur à celui du complexe TFIIH endogène. De plus les complexes recombinants purifiés à partir des cellules *Sf9* ne sont pas actifs en réparation. Il peut s'agir des modifications post-traductionnelles qui ne sont pas achevées dans la lignée *Sf9*.

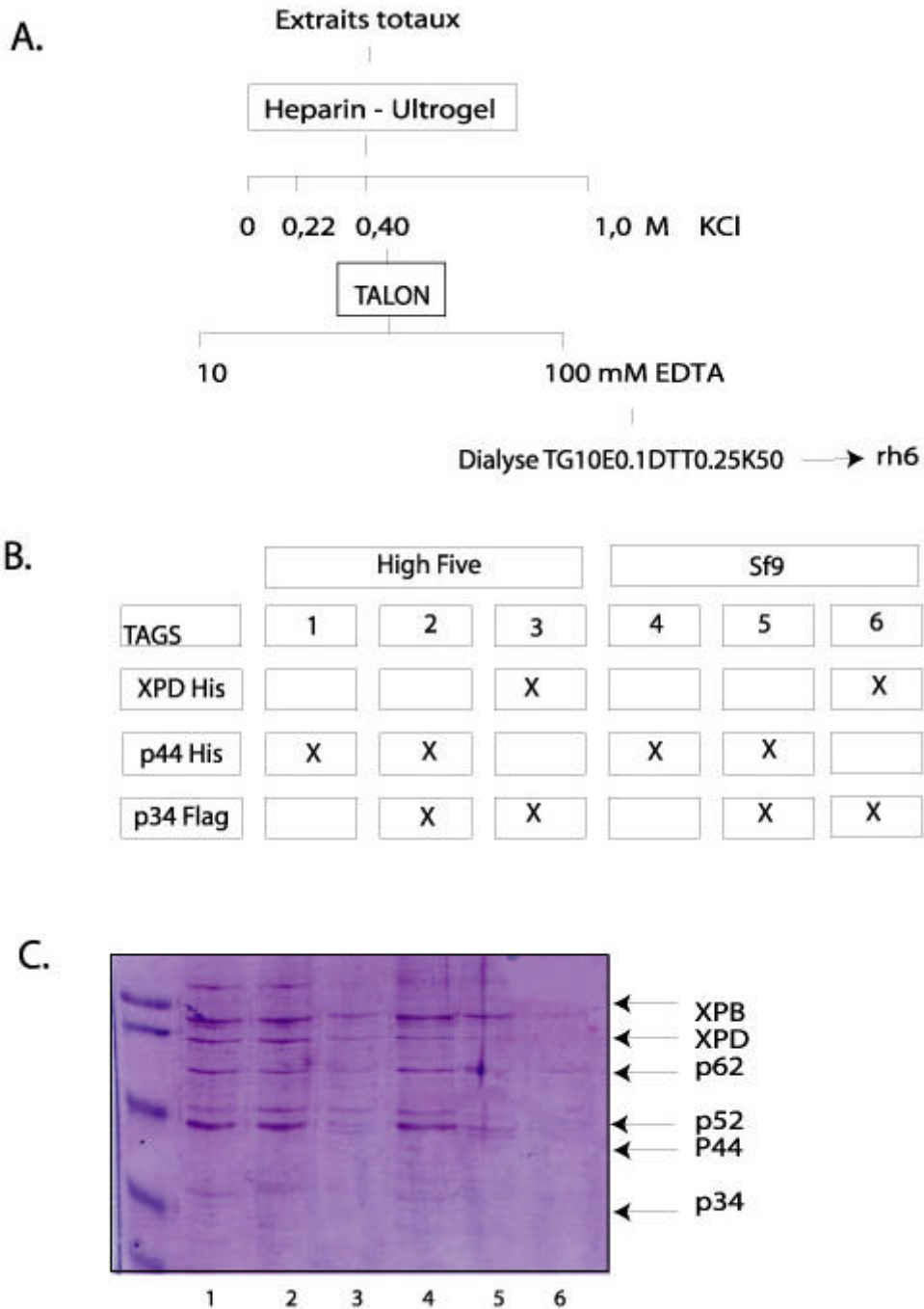
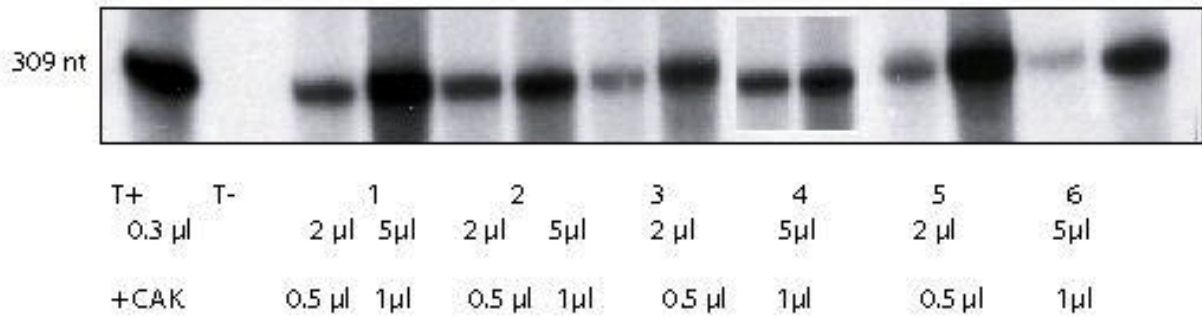


Figure 24 : Comparaison de la production de complexes recombinants core-TFIIH dans les cellules d'insectes High Five ou *Sf9*. (A) Schéma de purification des complexes recombinants. En fin de purification les échantillons sont dialysés. (B) Tableau récapitulatif des sous-unités fusionnées à un tag histine et/ou flag. Les sous-unités non fusionnées ne sont pas représentées. (C) Les complexes purifiés sont déposés sur gel SDS 12.5% et colorés au bleu de Coomassie.

A.



B.

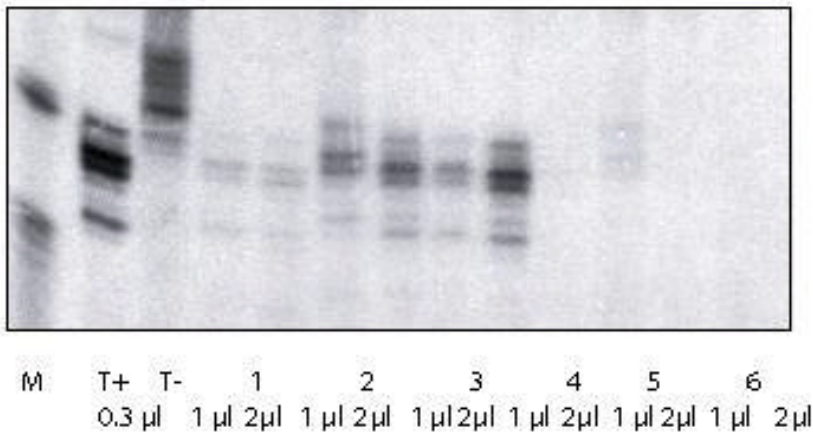


Figure 25 : Comparaison de l'activité des complexes recombinants fusionnés rllh6 purifiés à partir des cellules d'insectes High Five ou *Sf9*. (A) Test de transcription *in vitro*. Différentes quantités d'échantillons sont déposés (2µl ou 5µl). Le témoin positif (0.3µl) est du complexe endogène purifié à partir de cellules Hela. Tous les complexes recombinants sont actifs en transcription après ajout du complexe CAK.

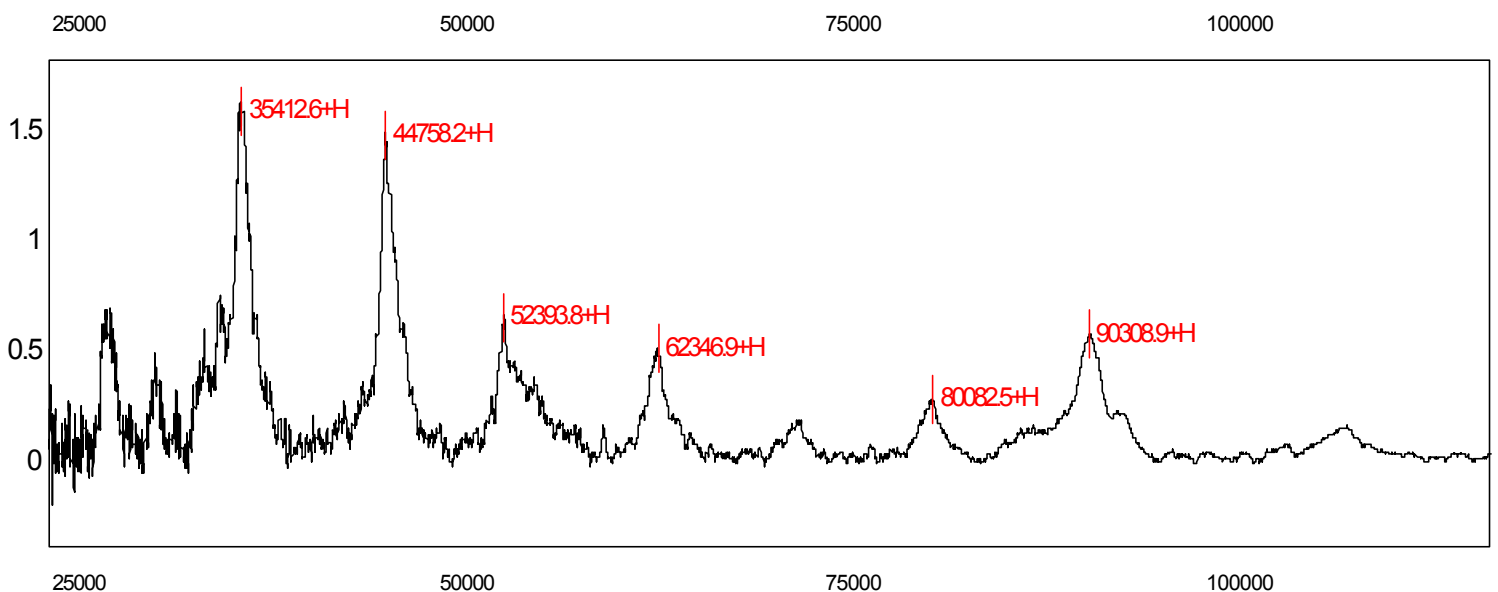
(B) Test de réparation par excision resynthèse des nucléotides *in vitro*. Les complexes recombinants purifiés à partir des cellules d'insectes High Five sont actifs en réparation. Les complexes purifiés à partir des cellules *Sf9* ne sont pas actifs en réparation. Les modifications post-traductionnelles dans les cellules *Sf9* peuvent être inachevées.

### 3. Caractérisation du complexe core-TFIH par spectrométrie de masse

Précédemment nous avons voulu caractériser le complexe core-TFIH par spectrométrie de masse. Malheureusement, tous nos essais ont été infructueux, nous n'avons jamais pu obtenir de spectres complets illustrant la composition du complexe recombinant.

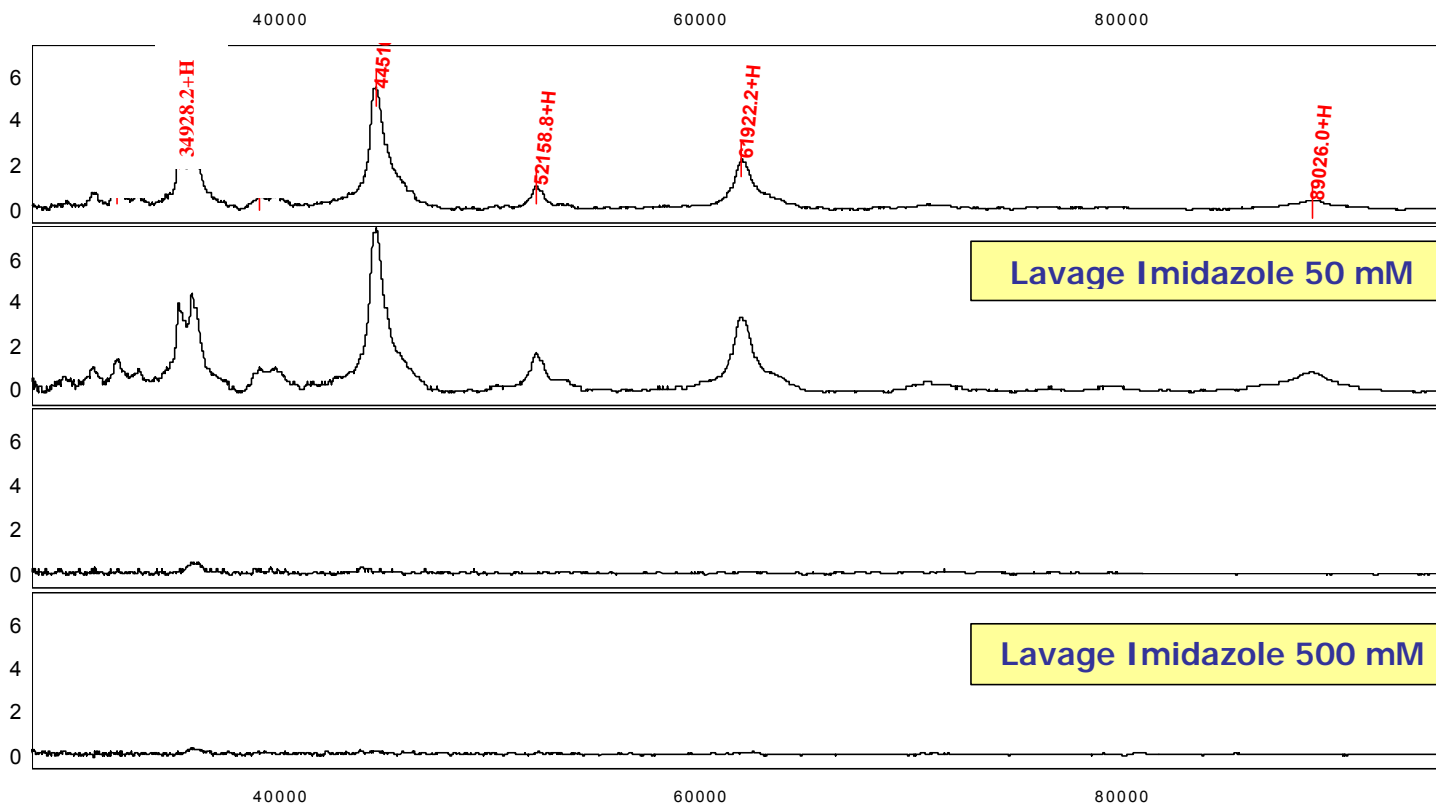
Une visite de la société Cyphergeren nous a donné l'occasion de tester nos complexes recombinants avec cet appareil qui promettait d'être très sensible et très efficace. La société Cyphergeren commercialise un spectromètre de masse assez étonnant du point de vue taille puisqu'il fait à peine 1m<sup>3</sup>. Le spectromètre contient un laser à désorption/ionisation et la détection et la détermination de la masse moléculaire des protéines se fait par TOF-MS (Time-of-Light Mass Spectrometry) (figure 26). La préparation de l'échantillon est assez triviale : il suffit de déposer l'échantillon à analyser sur une Chip (surface) qui existe avec différentes chimies : affinité avec des métaux, liaisons hydrophiles, hydrophobes, interaction avec l'ADN. Des lavages sont réalisés et la chip est directement introduite dans le spectromètre pour la mesure.

Nous avons utilisés deux types de surface, la première permet les interactions hydrophiles entre la protéine et les résidus -OH couplées à la matrice. Après dépôt du complexe recombinant et lavage en conditions stringeantes, nous obtenons le spectre suivant :



Parmi les pics obtenus, et bien que la hauteur des pics soit très faibles, nous distinguons six pics de poids moléculaires : 35412.6 Da, 44758.2 Da, 52393.8 Da, 62345.9 Da, 80082.5 Da et 90308.9 Da qui correspondent aux masses protéiques attendues. Les masses moléculaires mesurées sont les masses des protéines qui forment le complexe rIIH6 : p34, p44, p52, p62, XPD, XPB. Il y a également des pics correspondant à des contaminants mais ceux-ci restent minoritaires. Ce premier spectre est un contrôle qualité du complexe recombinant, de ce point de vue nous sommes assurés de l'expression des six protéines dans les cellules d'insectes.

Nous avons voulu ensuite mettre en évidence l'association du complexe par rétention sélective de ce dernier via la sous-unité His-taguée. La Chip utilisée est la chip IMAC qui assure la liaison des protéines His-taguées sur la matrice. Le complexe est adsorbé sur la surface et élué spécifiquement en présence de 50mM d'imidazole. Les pics obtenus sont les suivants : 34928.2 Da, 44510.1Da, 52158.8 Da, 61922.2 Da, 89026.0 Da. Ils correspondent aux protéines p34, p44, p52, p62, XPB mais XPD n'est pas détectée. En effet, XPD est labile et associée soit au core-TFIIH soit au complexe CAK, elle est souvent sous-stoichiométrique sur un gel SDS comparativement aux autres protéines du complexe (figure 23). Lors du dernier lavage, à une concentration élevée d'imidazole (500mM) plus aucune protéine n'est détectée sur le spectre de masse.





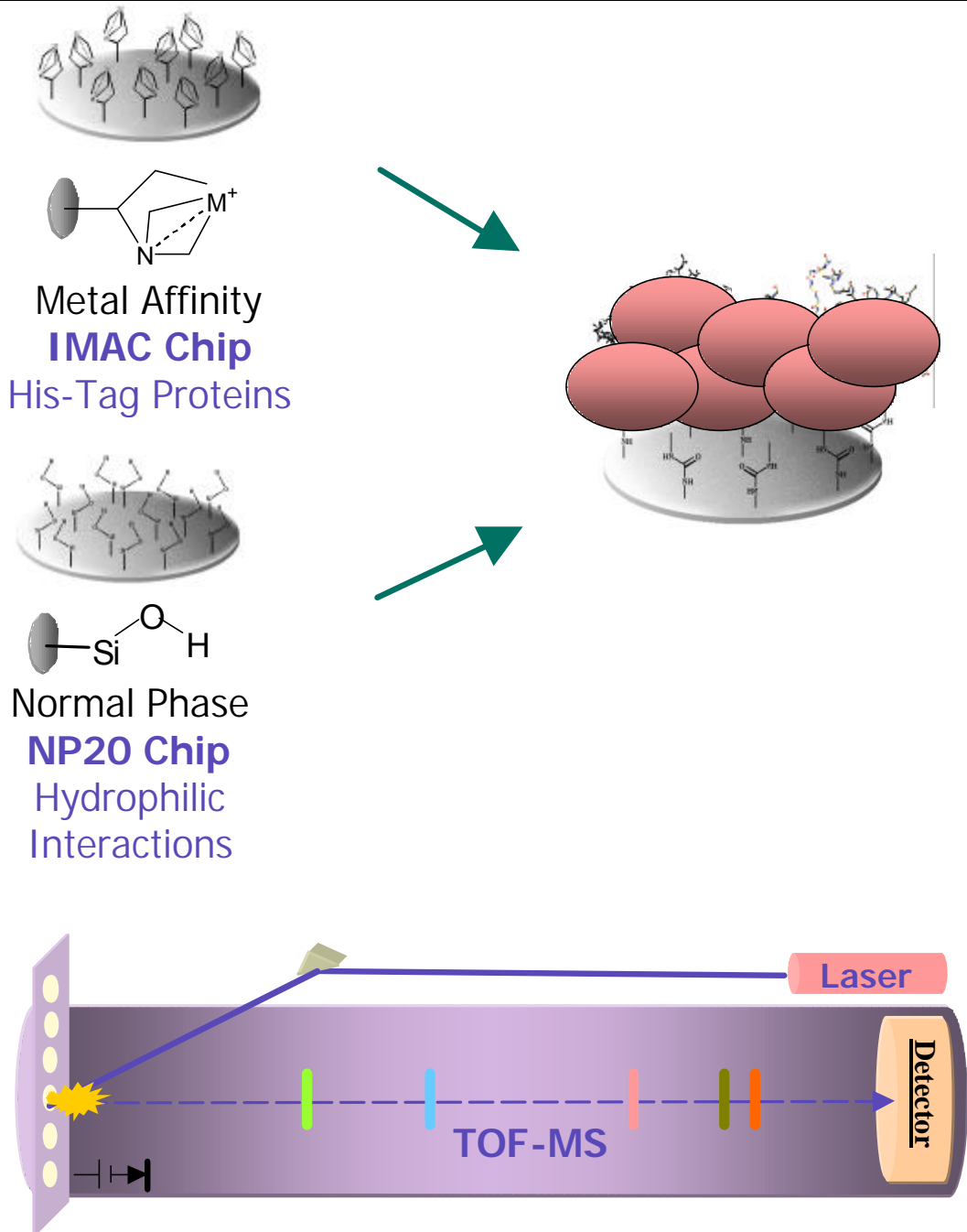


Figure 26 : Principe de la technologie CypherGen : cette technologie est basée sur le couplage entre la chromatographie de rétention sur surface et la spectrométrie de masse. L'utilisation de la surface (Chip) d'affinité au métal permet l'immobilisation de protéines His-taguées ; la surface NP20 est utilisée pour les interactions hydrophiliques entre la protéine et la surface. Après immobilisation, les complexes multiprotéiques sont lavés pour éliminer les contaminants non spécifiques. La chip est ensuite introduite dans le spectromètre qui est équipé d'un laser à désorption/ionisation, la détection et la détermination de la masse moléculaire se fait par Time-of-Light Spectrométrie (TOF-MS).

#### 4. Structure en coloration négative du complexe rIIH6

L'échantillon contenant le complexe rIIH6 est déposé sur une grille à trous, fixé à la glutaraldéhyde (0.1%), coloré à l'acétate d'uranyle, partiellement séché avec du papier filtre puis observé au microscope. Environ 3919 particules sont extraites manuellement à partir des micrographies numérisées. L'analyse d'images a été conduite comme précédemment décrit (voir §II.B).

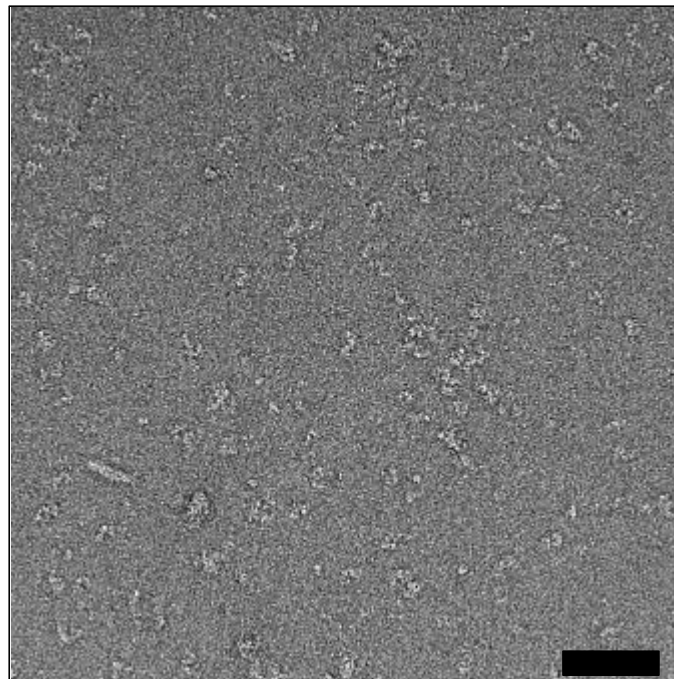


Figure 27 : Champ de particules de core-TFIIH observées en coloration négative. La barre d'échelle correspond à 100nm.

Les 3919 particules sont réparties en 68 classes (figure 28). Certaines classes apparaissent globulaire avec un anneau central. Les cycles d'alignement classification ont été réalisées sans introduction de références externes, dans un protocole « reference free » (voir §II.B). Les projections du modèle ont été ensuite utilisées comme nouvelles références d'alignement pour obtenir un modèle 3D stable.

Le modèle 3D du complexe core-TFIIH a été reconstruit et la taille de la particule est de 12x10nm. Le poids moléculaire du complexe est de 360KDa. Le complexe rIIH6 est un anneau qui a tendance à s'ouvrir. La figure 29 illustre la superposition de la

structure du complexe recombinant entier rIIH9 (Jawhari et al., 2002b) avec la structure du core-TFIIH. Par différence nous pouvons localiser le complexe CAK (100KDa) qui se situe dans la protubérance.

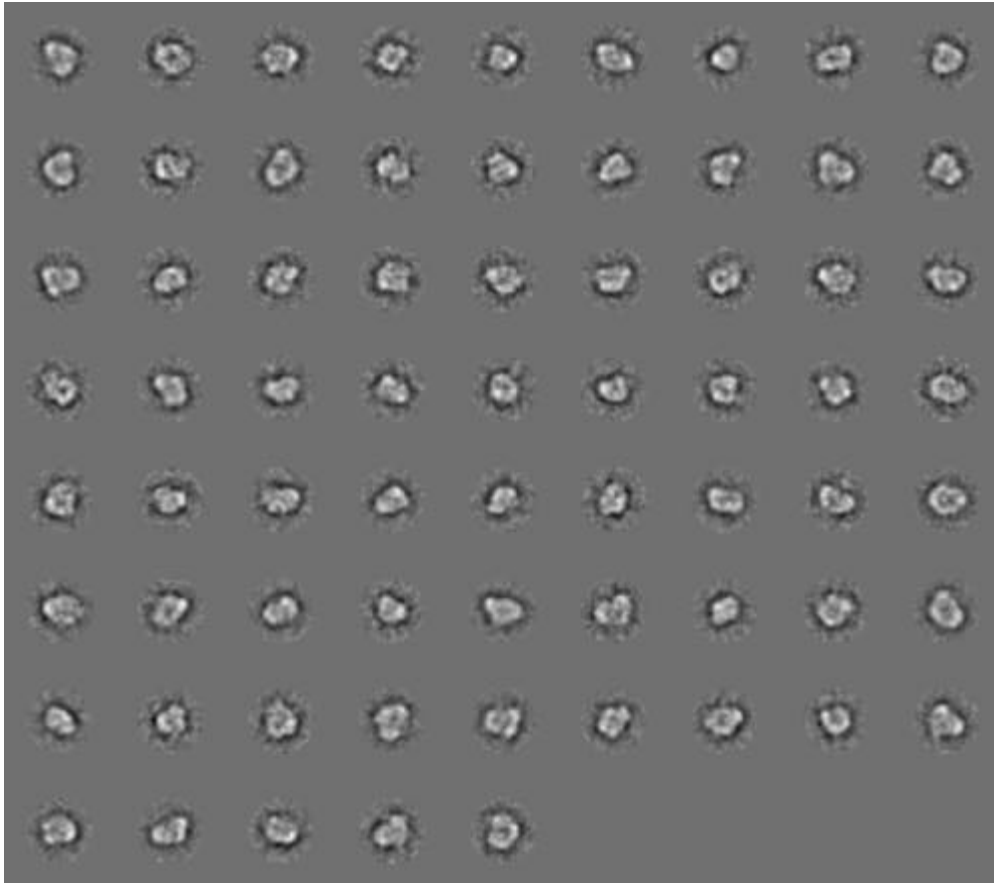


Figure 28 : Classification des particules de core-TFIIH

La perspective évidente de cette étude est l'analyse du complexe rIIH6 en cryomicroscopie. Un premier modèle en coloration négative a été reconstruit, et a ainsi validé la possibilité de poursuivre cette étude. En cryomicroscopie une résolution de l'ordre de 1.5nm peut être atteinte et à cette résolution nous pouvons escompter visualiser les domaines structuraux des six protéines qui forment le complexe rIIH6 et en particulier positionner les deux hélicases XPD et XPB afin d'appréhender le rôle d'ouverture du complexe. Sachant que le complexe CAK n'est pas nécessaire à la réparation de l'ADN, ceci ouvre des perspectives quant aux rôles *in vivo* de différents sous-complexes du facteur TFIIH et ceci dans des mécanismes distincts de transcription/réparation de l'ADN.

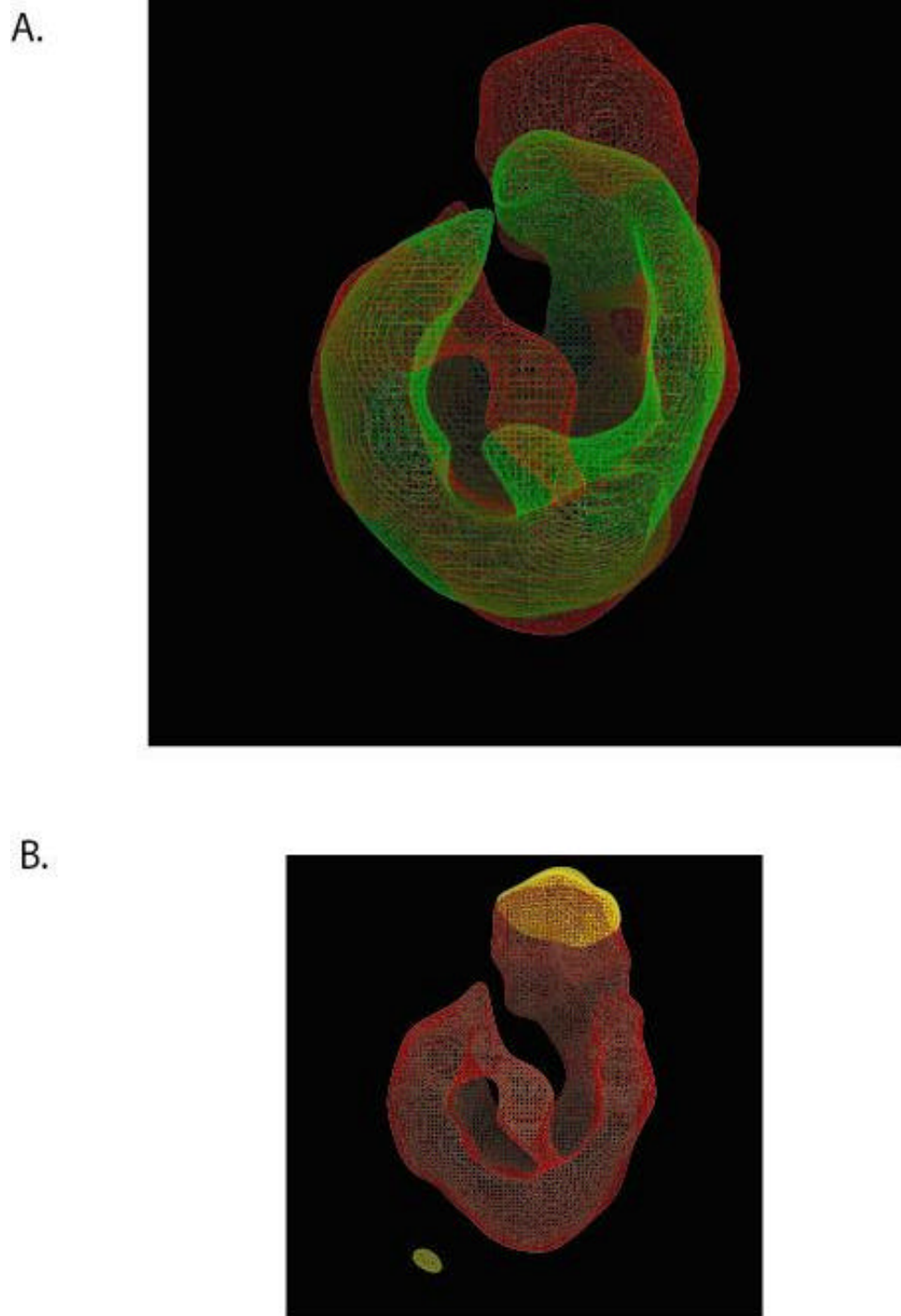


Figure 29 : Structure en microscopie électronique du facteur recombinant core-TFIIH. (A) Superposition des structures du facteurs rIIH9 et rIIH6 recombinants (B) Carte de différence. Le volume en jaune représente la différence de masse entre les deux structures et correspond au complexe CAK.

### **III. Organisation moléculaire du facteur de transcription TFIIH**

#### **A. Objectif de l'étude**

Nous avons cherché à intégrer dans le modèle structural de TFIIH des informations de nature et de résolution différentes. Nous avons utilisé les informations biochimiques décrivant le réseau d'interaction protéine-protéine au sein du complexe, le modèle 3-D de TFIIH servant de base pour l'intégration des informations de manière à proposer un modèle général. Ma contribution à cet égard a été d'étudier plus précisément l'architecture de la protéine XPD. Dans ce sens, une quinzaine de virus exprimant différents fragments de XPD ont été construits (voir Chapitre 3, figure 33) et ont permis, par des expériences de co-infection réalisées avec le système d'expression baculovirus, d'identifier les domaines d'interaction entre les hélicases et les autres sous-unités de TFIIH. Cette connaissance de l'architecture de TFIIH permettra de mieux appréhender la réaction d'ouverture de l'ADN et sa régulation.

#### **B. Modèle général de l'architecture de TFIIH**

Un modèle de l'architecture du facteur TFIIH humain est proposé en se basant sur les études d'interaction protéines-protéines et les immunomarquages des sous-unités XPD, XPB, Cdk7, p44 par microscopie électronique (Schultz et al., 2000). Un certain nombre d'interactions est désormais connu (tableau) : p34 est le partenaire de p44 (Fribourg et al., 2001), p44 interagit lui-même avec p62 (Jawhari et al., 2000) et régule l'activité hélicase de XPD (Coin et al., 1998). La sous-unité MAT1 interagit avec Ckd7, cycline H, XPD et XPB (Busso et al., 2000). Nos études ont permis d'une part, de valider ces interactions et d'autre part, de proposer de nouveaux partenaires. Ainsi les domaines d'interactions de l'hélicase XPD ont été investigués (voir Chapitre 3). Avec l'ensemble de ces informations nous présentons un modèle général de l'architecture de TFIIH (figure 29A).

	p34	p44	p52	p62	XPB	XPB	MAT1
p34							
p44							
p52							
p62							
XPB							
XPB							

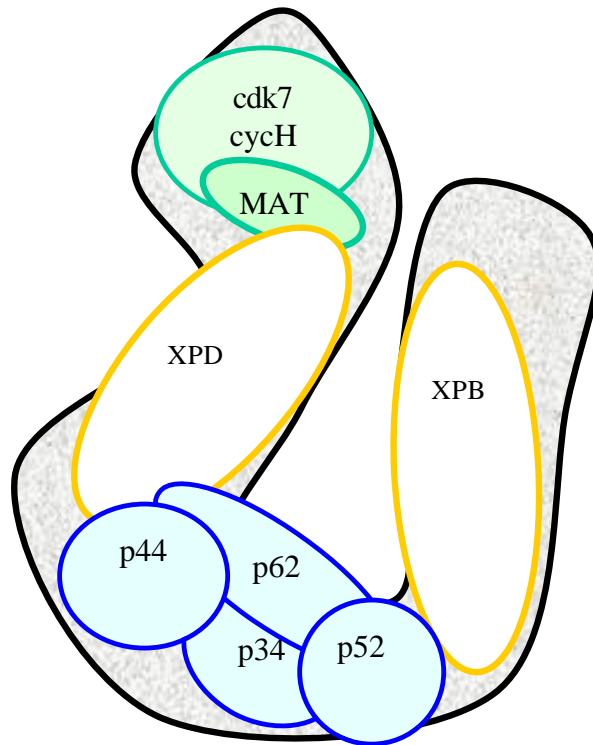


Figure 29A : Modèle général de l'architecture de TFIIH (A) Tableau résumant les interactions entre les sous-unités (B) Représentation schématique du modèle

## Chapitre III : L'hélicase XPD : dissection de la protéine humaine, vers la structure de son homologue archaebactérien

### A. Généralités sur les hélicases

Les hélicases sont des enzymes ubiquitaires qui participent à de nombreux processus métaboliques impliquant l'ADN ou l'ARN. Elles sont essentielles au déroulement normal de la réplication, de la transcription, de l'épissage, de la recombinaison et de la réparation de l'ADN (Matson, 1991), ce qui explique leur grande diversité et leur potentielle redondance dans la cellule. Leur rôle consiste à rompre les liaisons hydrogène qui assurent la cohésion de la double hélice d'ADN et à défaire les structures secondaires qui peuvent s'établir le long d'une molécule d'ARN ou d'ADN simple brin. Parce qu'elles sont impliquées dans bon nombre de processus cellulaires fondamentaux, un nombre croissant de pathologies humaines rares liées au dysfonctionnement d'une hélicase a été identifié ces dix dernières années. Ces maladies sont en général liées à des mutations récessives portées par les autosomes (voir synthèse « Les hélicases et les maladies associées » en Introduction).

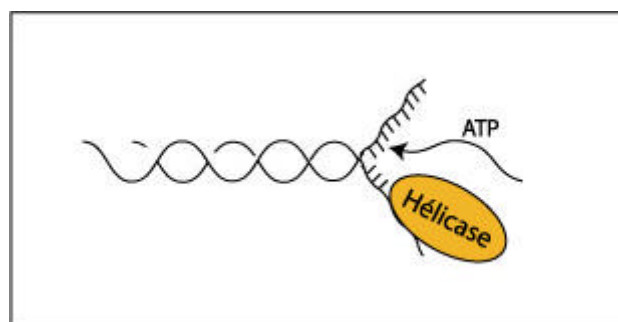


Figure 30 : Mécanisme d'action d'une hélicase : l'hélicase rompt les liaisons hydrogène de la double hélice d'ADN en présence d'ATP

## 1. Les familles d'hélicases, conservation des motifs

Il y a à peine plus d'une dizaine d'années que Gorbalenya et Koonin ont proposé une classification générale des hélicases en cinq groupes d'après la conservation de motifs peptidiques (Gorbalenya et al., 1989). Toutes ces protéines lient l'ATP et disposent de motifs assignés à cette fonction : le motif A de Walker ou boucle de liaison au phosphate et le motif B de Walker ou motif de liaison au magnésium. Les deux groupes les plus importants correspondent aux superfamilles (SF) 1 et 2. Une troisième superfamille, SF3, inclut des domaines hélicase putatifs d'environ une centaine d'acides aminés. Elle dispose de trois motifs conservés, incluant les motifs A et B de Walker. Une quatrième famille contenant les homologues de l'hélicase DnaB d'*Escherichia coli* ont cinq motifs conservés et forment en général des structures hexamériques. Enfin, le dernier groupe qui présente des homologies de séquences avec les ATPases « pompes à protons », regroupe les protéines apparentées à Rho, un facteur de terminaison agissant sur les ARNm naissants.

Les hélicases des superfamilles SF1 et SF2 regroupent un grand nombre d'ADN et ARN hélicases de bactéries, d'archaebactéries, d'eucaryotes et de virus. La séparation entre deux groupes n'est pas justifiée par un substrat particulier, ADN ou ARN ou une activité hélicase préférentielle comme la direction d'ouverture de l'ADN mais plutôt par la présence de séquences particulières. Ces groupes d'hélicases se caractérisent par l'existence de sept motifs (I, Ia, II, III, IV, V, VI) auxquels se rajoutent deux motifs récemment identifiés (le motif Q et le motif TxGx).

Les protéines XPD et XPB de TFIIH sont des ADN hélicases et appartiennent à la famille SF2. Nous allons décrire la fonction de ces motifs ci-après, les principales caractéristiques étant résumées dans le tableau 7.

Par comparaison de séquences, il a été déterminé que la boîte I correspond au motif A de Walker ou boucle de liaison au phosphate (« boucle P »), qui est présente chez toutes les ATPases (Walker, 1982). Ce motif est responsable de la liaison et de l'hydrolyse des nucléotides tri-phosphates dans la plupart des hélicases liant les nucléotides. La lysine de la séquence GKT est extrêmement importante puisqu'elle interagit via son groupement amine avec le phosphate de l'ATP/ADP alors que le



groupement hydroxylique de la thréonine lie l'ion  $Mg^{2+}$ . Une mutation de cette lysine abolit totalement la liaison .

Le motif II ou « DEAD box » correspondant au motif B de l'ATPase défini par Walker est le motif de fixation à l'ion  $Mg^{2+}$ . Les hélicases portant la boîte DEAD sont des ARN hélicases, les ADN-hélicases disposent de la boîte DEAH ou DExH.

Les structures des hélicases cristallisées en présence d'ADN comme Rep, liée à de l'ADN simple brin (Korolev et al., 1997), PcrA liée à un hybride ADN/ADN simple brin et double brin (Velankar et al., 1999) et l'ARN hélicase Ns3 de l'HCV (hepatitis C virus) (Kim et al., 1996a) liée à un ADNss, donnent des indications quant aux mécanismes de liaison des hélicases SF1 et SF2 à l'ADN. Les motifs Ia et IV sont des motifs de contact avec l'ADNss. De plus, une région localisée dans une boucle entre le motif Ia et II est capable de lier l'ADNss. Cette région est référencée comme le motif TxGx. Un motif analogue, le motif QxxR, est spécifique des ARN hélicases à boîte DEAD.

Le rôle des motifs III et V est encore peu connu. Velankar et ses coauteurs proposent l'existence d'un réseau d'interactions entre l'ATP, l'ADNss, le motif I et le motif III. Ceci suggérerait une fonction de relai du motif III et par conséquent la coordination de l'activité hélicase et de l'hydrolyse de l'ATP. De même, le motif V, qui contient des résidus en contact avec les acides nucléiques, serait capable de former une chaîne d'interactions aboutissant à l'activité hélicase.

L'arginine au milieu du motif VI est caractéristique des superfamilles SF1 et SF2. Dans la structure cristallographique de PcrA, le groupement guanidine de l'arginine forme un pont salin avec le phosphate terminal gamma de l'analogue non hydrolysable de l'ATP, l'AMP-PNP (Velankar et al., 1999). L'hydrolyse de l'ATP et la libération du pyrophosphate pourrait détacher l'arginine. Ceci pourrait jouer un rôle sur la déstabilisation de l'ADN et le déplacement du brin. La structure de eIF4A, une ARN hélicase à boîte DEAD, confirme l'interaction de l'ATP avec une arginine du motif VI (Pause et al., 1993).

Récemment un nouveau motif hélicase a été découvert : le motif Q spécifique des ARN hélicases à boîte DEAD (Cordin et al., 2004). Le motif Q est localisé à 17 acides aminés après le motif I. De plus, un groupe aromatique très conservé mais isolé, se trouve également à 17 acides aminés, cette fois du motif Q. Ces deux éléments sont impliqués dans l'activité ATPase des protéines à boîtes DEAD. Cordin et ses coauteurs ont démontré que ce motif est non seulement impliqué dans la liaison et l'hydrolyse de l'ATP mais aussi dans la liaison de l'ARN.

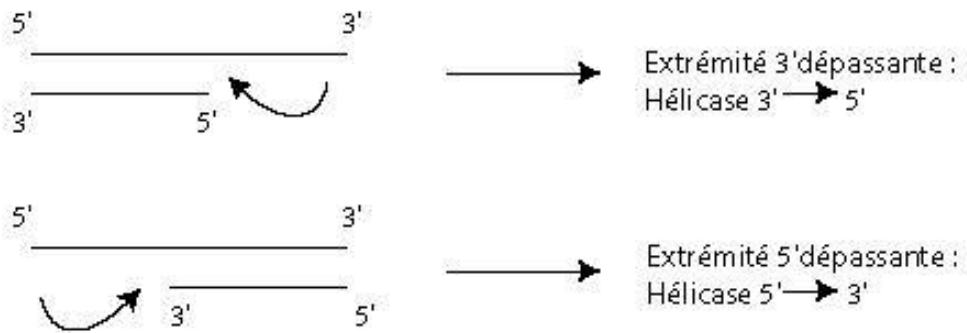
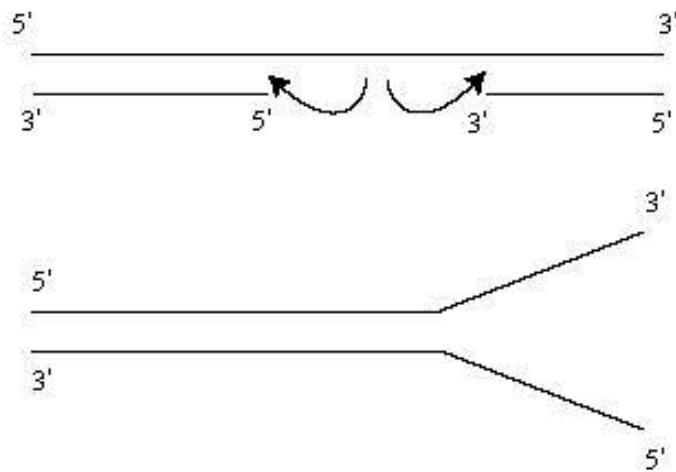
Motifs	Fonction	SF1	SF2
I	Boucle de liaison au phosphate Hydrolyse de l'ATP	++GxAGoGKS	++xxxoGxGKT
Ia	Contact avec l'ADN simple brin	Xx+xxxoo	X+++xPoo
Q	Hydrolyse de l'ATP – Activité hélicase – liaison ARN	GFxxPxPIQ	GFxxPxPIQ
TxGx	Contact avec l'ADN simple brin	TxGx	TxGx
II	"DEAD box" - fixation du Mg <sup>2+</sup>	+++DExo	+++DExH
III	Couplage des activités ATPase et hélicase	++++GDxoQ	+x+SATxxxTGS
IV	Contact avec l'ADN simple brin	Xx+xooxR	++Fxxoxo
V	Contact avec l'ADN Couplage des activités ATPase et hélicase	xxT+xxxQG+o+ooV	+xTxxxxxG+o+xo+
VI	Couplage des activités ATPase et hélicase	VA+TRxoo	QxxGRxxR

Tableau 7 : Tableau récapitulatif des motifs conservés chez les hélicases des superfamilles (SF) 1 et 2 d'après Gorbalenya et Koonin (1993) et Corbin (2004). (x) acide aminé quelconque ; (+) acide aminé hydrophobe ; (o) acide aminé hydrophile.

## 2. Mesure de l'activité déroulante des hélicases

Les hélicases sont capables de dérouler un ADN double brin en présence d'ATP. Cette propriété est directement utilisée pour caractériser les hélicases putatives déduites des alignements de séquences. Pour cela, il existe un test simple : un ADN double brin est mis en présence de la protéine qui est susceptible de posséder une activité hélicase. Un exemple de combinaison de substrats hélicase (ADN avec extrémité libre en 5' ou en 3', duplex ADN) est présenté dans la figure 31. L'utilisation de ces différents substrats permet de déterminer la spécificité et le sens de déplacement de la protéine. Les brins sont séparés sur gel non dénaturant et comparés au double brin non dénaturé et au double brin dénaturé à 90°C. Si la protéine possède la propriété de dissocier le double brin, la forme simple brin apparaîtra (figure 31c).

La méthode la plus utilisée pour marquer le substrat est le marquage radioactif au [ -  $^{32}\text{P}$ ] en 5' de l'oligonucléotide. La T4 polynucléotide kinase catalyse la transfert du phosphate de l'ATP radioactif sur le phosphate en 5' de l'oligonucléotide. L'oligonucléotide marqué est ensuite hybridé au brin complémentaire. De plus en plus, les techniques radioactives sont remplacées par des méthodes moins dangereuses. Deux exemples de marquage non radioactifs sont présentées : (1) Zang et ses coauteurs (Zhang et al., 2001) ont développé un test hélicase basé sur l'électroluminescence. L'oligonucléotide est marqué en 5' par du ruthenium et hybridé à son oligonucléotide complémentaire. Après ouverture du duplex par l'hélicase, le brin libéré marqué au ruthenium est réhybridé à un brin biotinylé. L'hybride biotinylé est capturé par des billes magnétiques couplées à la streptavidine et finalement par une électrode sous un champ magnétique. L'application d'une différence de potentiel électrique excite le ruthénium et induit une libération de lumière par chimioluminescence. (2) Plus récemment, Boguszewska et ses coauteurs (Boguszewska-Chachulska et al., 2004) ont développé un nouveau test hélicase utilisant la technologie FRET (Fluorescence Resonance Energy Transfer). Le substrat utilisé est un duplex ADN marqué par un fluorochrome, l'autre brin étant marqué par un « quencher » qui absorbe la fluorescence lorsqu'il est proche de l'autre brin (distance d'environ 10 à 100Å). Un signal de fluorescence correspond donc directement à l'ouverture du duplex.

**A.** Unidirectionnelle**B.** Bidirectionnelle**C.**

	37°C	90°C					
ATP	+	+	+	-	+	+	+
Hélicase	-	-	-	+			
Double brin							
Simple brin							

Figure 31 : Mesure de l'activité déroulante d'une hélicase : (A) Différents modèles de doubles brins servant de substrats à l'hélicase : l'extrémité 5' ou 3' dépassante sert à tester la polarité de l'hélicase. (B) Substrats bidirectionnels et fourche à extrémité dépassante (C) Exemple de test hélicase. L'effet de la concentration en protéine et la

présence de l'ATP sont étudiés. La séparation de contrôle des 2 brins est effectuée par dénaturation à 90°C.

### **3. Mesure de l'activité ATPase**

Le déroulement des doubles brins d'ADN nécessite la présence d'ATP. La mesure de l'activité ATPase est réalisée en dosant la libération du phosphate inorganique (Pi) en fonction du temps, dans un mélange contenant la protéine, l'ADN à dérouler et de l'ATP radioactif [ $^{32}\text{P}$ -ATP]. Après arrêt de la réaction, du charbon actif est ajouté au milieu réactionnel. Celui-ci a la capacité d'adsorber l'ATP, l'ADP et l'AMP. Après centrifugation du charbon, la mesure de la quantité de radioactivité présente dans le surnageant correspond au Pi libéré.

### **4. Mécanisme d'action**

Deux modèles décrivant le mécanisme d'action des hélicases sont proposées : le mécanisme du roulement actif ou « active rolling » et le mécanisme de translocation ou « inchworm model ». Dans le modèle du « active rolling » l'hélicase est sous forme oligomérique, au minimum dimérique. Elle ne peut pas se lier simultanément à l'ADN simple brin et double brin mais de manière alternée. Dans le modèle du « inchworm model » l'hélicase peut être sous n'importe quelle forme oligomérique, y compris monomérique. Elle peut se lier simultanément à l'ADN simple brin et double brin par deux sites de liaison distincts. Ces deux mécanismes sont tous deux envisageables et semblent liés à l'état oligomérique des hélicases. Certaines hélicases fonctionnent sous forme d'hexamères comme les hélicases des bactériophages T7-gene4 et T4-gp41 et encercler la molécule d'ADN (Egelman et al., 1995).

Il n'est pas toujours facile de trancher entre ces hypothèses de mécanisme d'action. Par exemple, des observations contradictoires ont été publiées concernant UvrD. D'après le groupe de Mechanic, UvrD est active *in vitro* sous forme de monomère (Mechanic et al., 1999), alors que d'autres auteurs ont suggéré que la forme active optimale de UvrD est oligomérique (Runyon et al., 1993). Récemment, Maluf et ses coauteurs ont démontré qu'*in vitro*, la forme active de UvrD est dimérique (Maluf et al., 2003a; Maluf et al., 2003b). Dans un autre exemple, la résolution de la structure

crystallographique de l'hélicase du virus de l'hépatite C Ns3 a permis de proposer que l'état monomérique est actif. D'autres auteurs ont observé l'état oligomérique comme étant la forme active (Levin et Patel, 1999). Ces résultats illustrent la complexité de déterminer l'état oligomérique d'une hélicase et encouragent les méthodes procédant sur des « molécules uniques ».

### **5. Déplacement des protéines par l'hélicase**

Les équipes de Byrd et Kawaoka (Byrd et Raney, 2004) (Kawaoka et al., 2004) mettent en avant un rôle qui n'a rien à voir avec l'ouverture de l'ADN. Certaines hélicases seraient capables d'avancer sur la molécule d'ADN et de « détacher » les protéines qui se trouvent sur leur parcours. Les hélicases avaient précédemment été décrites comme étant capables de rompre les liaisons entre protéines et acides nucléiques et ainsi de relarguer la protéine. Par exemple, l'hélicase RecBCD déplace les nucléosomes pendant l'ouverture du duplex d'ADN (Eggleston et al., 1995). L'hélicase Dda est capable de déplacer une streptavidine liée à un oligonucléotide biotinylé. Le déplacement de la streptavidine pourrait mimer le détachement des protéines liées à l'ADN lors de la réplication. Le déplacement de Dda accroît lorsque la longueur de l'oligonucléotide biotinylé augmente, supposant la liaison d'une multitude de molécules de Dda sur l'ADN. Un modèle similaire de coopération entre plusieurs molécules d'hélicases Ns3 a été reporté (Levin et al., 2004). Ce rôle de détachement des protéines de l'ADN par des hélicases pourrait exister chez de nombreuses hélicases et apparaît comme une fonction biologique fondamentale.

## **B. L'hélicase XPD humaine**

### **1. L'hélicase XPD du complexe TFIIH humain**

L'hélicase XPD est une des dix sous-unités de complexe de transcription TFIIH. L'activité hélicase de XPD est indispensable à la réaction de réparation par excision resynthèse des nucléotides mais non à la transcription (Winkler et al., 2000c). La présence physique de XPD, plutôt que son activité ATPase/hélicase, est nécessaire à l'initiation de la transcription (Bradsher et al., 2000a).

Des mutations dans le gène XPD (figure 32) sont à l'origine du Xeroderma Pigmentosum (XP) du groupe D et de la trichothiodystrophie (TTD) ; des maladies autosomiques récessives, qui entraînent une déficience dans le système de réparation par excision resynthèse des nucléotides (NER). Une analyse des mutations observées chez une cinquantaine de malades montrent que dans 87% de cas XP-D et 50% de cas TTD, les patients portent des mutations dans la région C-terminale de XPD (région 675-730 pour l'XP-D et 713-730 pour le TTD). Cette région de XPD est impliquée dans l'interaction avec le régulateur p44 et une mutation dans la partie C-terminale de XPD empêche l'interaction avec p44 et résulte en une diminution de l'activité hélicase de XPD et ainsi d'une déficience de réparation par NER chez les malades.

On peut émettre l'hypothèse d'un lien entre le développement de cancers des patients XP-D et la régulation du cycle cellulaire par l'hélicase XPD. En 1996, l'équipe de Bartkova avaient déjà montré un lien plausible entre l'accumulation de Cdk7 et le cancer (Bartkova et al., 1996). Dernièrement, Chen et ses coauteurs ont suggéré chez la drosophile que l'interaction entre XPD et le complexe CAK (cdk activating kinase) régule le cycle cellulaire (Chen et al., 2003). Cette étude met en relief un nouveau mécanisme qui pourrait expliquer le risque élevé de cancers des patients XP-D. L'hélicase XPD mutée ne pourrait plus réguler le complexe CAK et ce dernier, seul, s'accumulerait dans la cellule. L'excès de CAK faciliterait la progression prématurée du cycle cellulaire et entrainerait le développement de cancers dans les cellules défectives en système de réparation.

La polarité d'ouverture de brin par l'hélicase XPD est en 5'- 3' (Sung et al., 1993). L'hélicase XPD peut être produite seule dans le système d'expression baculovirus et immunopurifiée par un anticorps anti-XPD. XPD dissociée de TFIIH, conserve son activité hélicase, bien que diminuée par rapport au sein du complexe entier (Coin et al., 1998b). L'hélicase XPD a été localisée par un immunomarquage dans le modèle 3D obtenu par microscopie électronique du complexe TFIIH (Schultz et al., 2000c). La localisation de XPD au sein de l'enveloppe apparaît proche du complexe CAK. L'interaction de XPD avec le complexe CAK via la sous-unité MAT1 a été observé dans une expérience de quadruple hybride (Sandrock et Egly, 2001a).

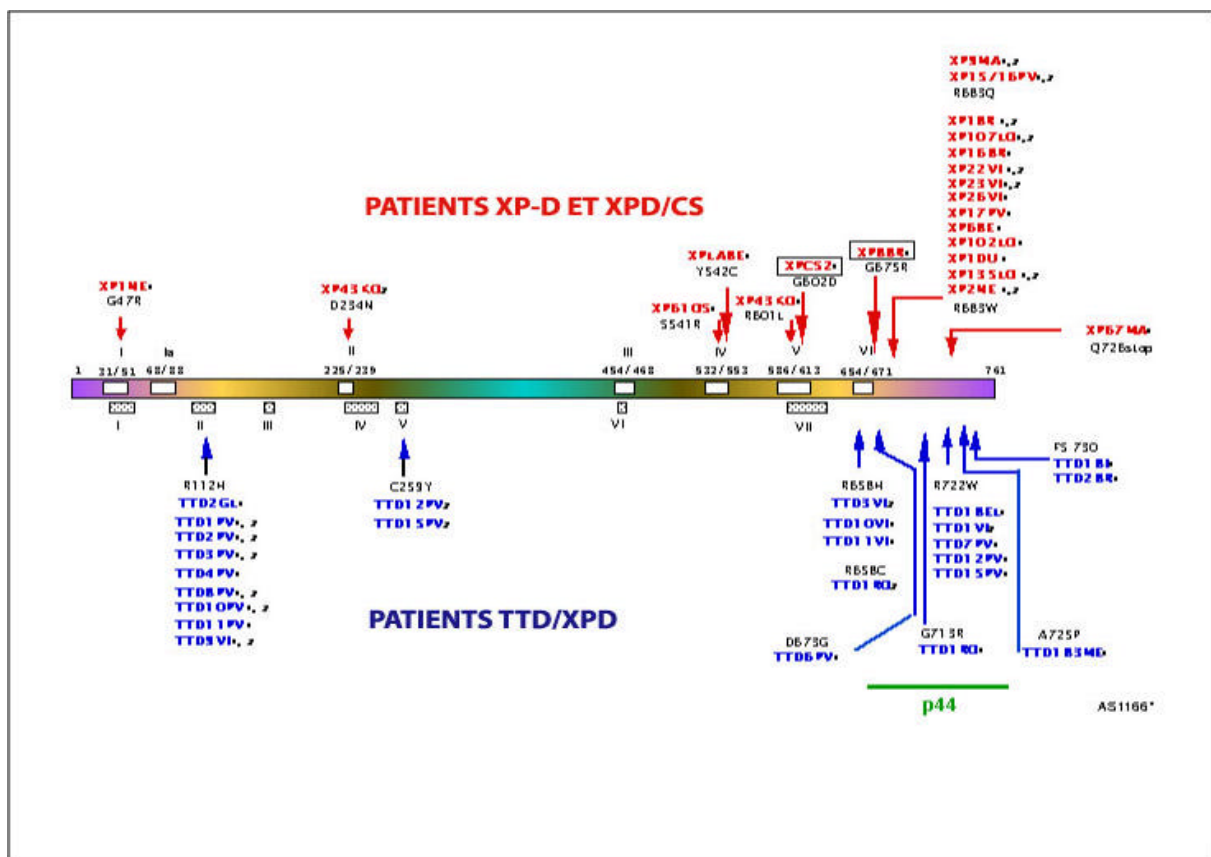


Figure 32 : Les mutations de l'hélicase XPD sont à l'origine des maladies XP-D et CS (en rouge) et TTD/XP-D (en bleu)

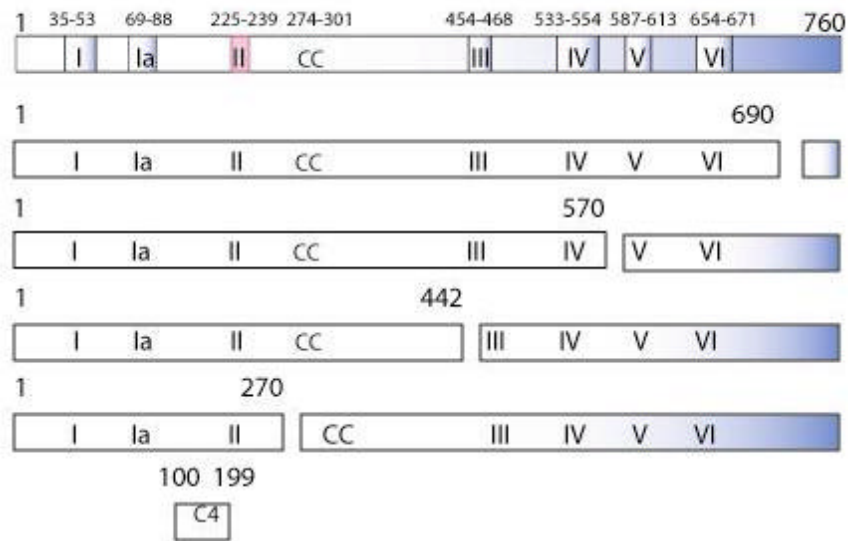


## 2. Dissection de l'hélicase humaine XPD dans le système d'expression baculovirus

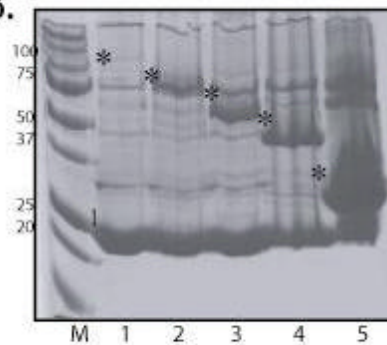
Nous avons amorcé une étude structurale sur l'hélicase humaine XPD. Le niveau d'expression de l'hélicase entière dans *Escherichia coli* est quasi nulle et la protéine est complètement insoluble. C'est pourquoi nous avons choisi d'étudier l'expression de l'hélicase dans le système d'expression baculovirus.

Nos précédentes observations nous indiquent que le taux d'expression de l'hélicase XPD entière est faible dans le système d'expression baculovirus et non suffisante pour amorcer une étude structurale. Nous avons alors décidé de découper l'hélicase en fragments en espérant une meilleure expression de ces derniers. Ces fragments ont été choisis d'après les alignements de séquences en préservant les domaines hélicase et d'après la prédiction de structure secondaire de la protéine et les caractéristiques physico-chimiques des acides aminés. Les constructions des différents fragments sont représentés dans la figure 33a. S'inspirant de notre étude sur l'utilisation des protéines de fusion dans le système d'expression baculovirus (voir publication n°5), nous avons choisi de cloner tous les fragments, y compris la protéine entière, dans le vecteur d'expression pDEST20 (Invitrogen) qui permet la fusion à la Glutathion S transférase (GST) à l'extrémité N-terminale de la protéine. Les tests d'expression des différents fragments, présentés dans la figure 33 montrent que l'expression de la protéine entière n'est pas améliorée (figure 33b, piste 1). Par contre, les fragments sont tous solubles, le taux d'expression des fragments moyens (37, 47, 56, 60, 75, 80, 89 kDa) est correcte et l'expression du petit fragment C-terminal est particulièrement élevée (piste 5).

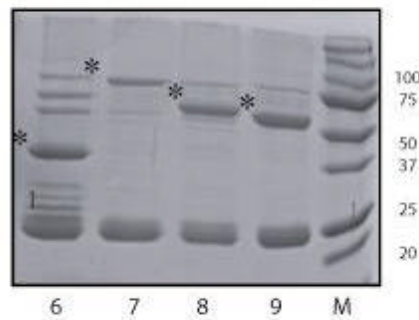
A.



B.



	Fragment XPD	KDa	Expression
1	1-760	110	-
2	271-760	80	++
3	443-760	60	++
4	571-760	47	++
5	691-760	33.5	+++



6	100-199	37	+
7	1-571	89	+
8	1-443	75	++
9	1-271	56	++

Figure 33 : Dissection de l'hélicase humaine XPD et production des fragments dans le système d'expression baculovirus (A) Représentation des différentes constructions de l'hélicase humaine. Les domaines sont découpés en fonction de la position des motifs hélicase (I, Ia, II, III, IV, V, VI) CC:Coiled Coil.(B) Expression des domaines de l'hélicase humaine dans le système d'expression baculovirus. Environ 25ml de cellules d'insectes sont infectées avec le baculovirus exprimant le domaine d'intérêt. Les cellules sont cassées et les extraits solubles obtenus après ultracentrifugation sont

incubés avec la résine GST (Pharmacia) pour retenir spécifiquement les protéines d'intérêt. Après 2h d'incubation à 4°C, les billes sont lavées et du bleu de dépôt est ajouté. Les billes sont déposées sur gel SDS 12.5% et colorées au bleu de Coomassie.

### 3. Optimisation de la production du fragment C-terminal de XPD dans les cellules d'insectes

Après avoir vérifié l'expression des différents domaines de l'hélicase humaine dans le système d'expression baculovirus, nous avons voulu optimiser la production des protéines recombinantes. Les cellules d'insectes habituellement utilisées au laboratoire sont les cellules de papillons *Spodoptera frugiperda* (Sf9). Pour obtenir un maximum de protéine recombinante, les conditions optimales de culture ont été déterminées à la fois pour les cellules Sf9 ainsi que pour les cellules de *Trichoplusia ni* (High Five) une autre lignée d'insecte disponible au laboratoire. Les paramètres de culture étudiés incluent le mode de culture (cellules attachées ou en suspension), le temps d'infection, le type de milieu de culture (présence ou non de sérum) ainsi que le coût. Les différents milieux de culture évalués sont les milieux Grace (Sigma ) supplémenté de 10% de sérum, Insect XPRESS (Cambrex ), Express Five (Life Technologies ) et Sf-900 II (Life Technologies ).

Les modifications post-traductionnelles des protéines sont différentes selon le type de cellules utilisé et il apparaît donc intéressant d'étudier l'expression d'une même protéine dans deux types cellulaires distincts, tout spécialement à des vues structurales. Nous avons optimisés la production du fragment C-terminal (691-760) de XPD (figure 34).

Les conditions de culture optimales pour la production de ce fragment sont : une culture en suspension, car moins laborieuse que la culture sous forme attachée, et un temps d'infection de 72H. Nous avons choisi de cultiver les cellules Sf9 en milieu Express Five, moins onéreux que le milieu Insect Express.

D'un point de vue structural, nous n'avons pas pu produire suffisamment de protéine pour la cristalliser. Le fragment C-terminal de XPD reste tout de même intéressant puisque deux mutations sont présentes chez les patients à ce niveau. Nous avons dès lors utilisé les fragments de XPD pour investiguer le réseau d'interaction de

l'hélicase au sein de TFIIH. L'analyse de deux interactions est présentée dans le paragraphe suivant.

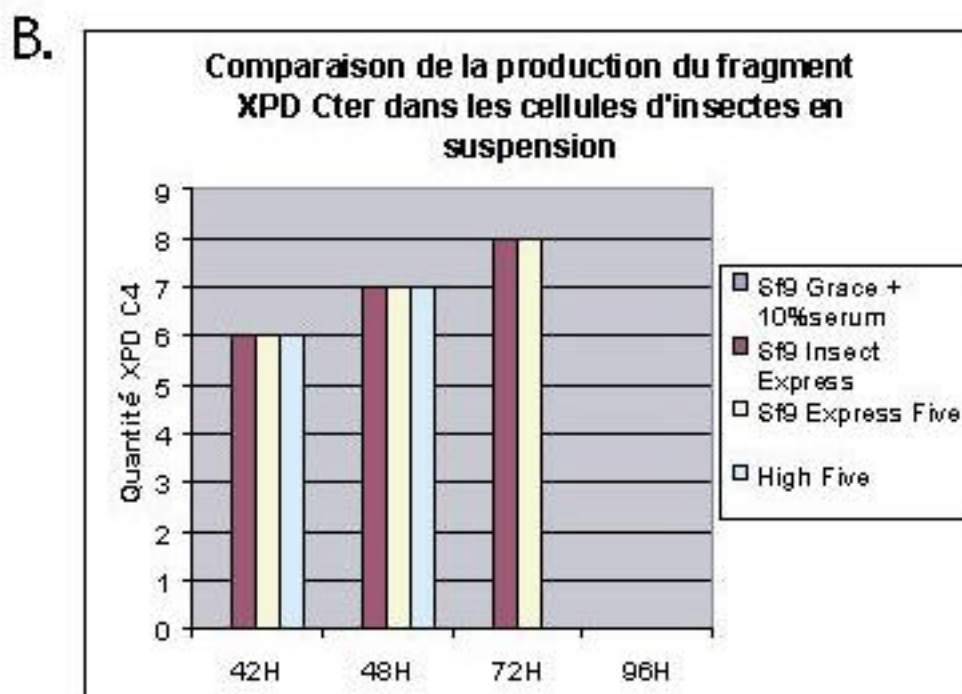
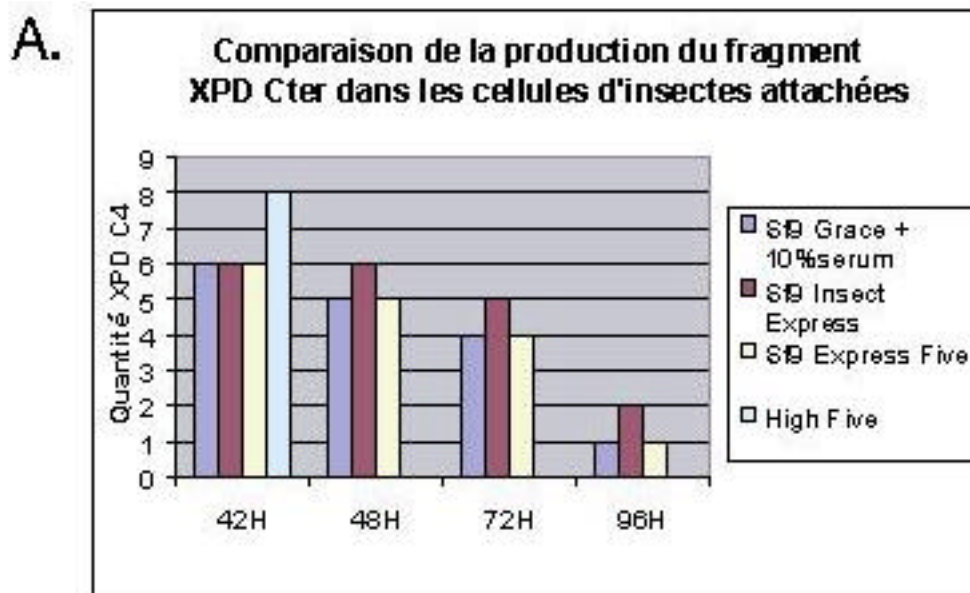


Figure 34 : Comparaison de la production du fragment C-terminal de la protéine XPD en fonction du temps dans les cellules d'insectes (données venant de trois expériences distinctes) (A) Production en cellules d'insectes attachées. L'expression de la protéine est maximale au bout de 42H dans les cellules High Five cultivées en milieu Insect Express sans ajout de sérum (80%). Le taux d'expression est identique dans les milieux Grace, Insect Express et Express Five (60%) en cellules Sf9. Après 48H d'infection, l'expression de la protéine dans les cellules High Five chute brutalement. Ceci peut être expliqué par l'induction de protéases qui dégradent la protéine. L'expression dans les cellules Sf9 diminue de peu dans les milieux Grace et Express Five (50%), elle reste constante dans le milieu Insect Express. Après 72H et 96H d'infection la quantité de protéine diminue progressivement dans les milieux Grace, Insect Express et Express Five en cellules Sf9. (B) Production en cellules d'insectes en suspension. L'expression de la protéine est maximale au bout de 72H dans les cellules Sf9 cultivées en milieu Insect Express ou en Express Five sans ajout de sérum (80%). Les cellules Sf9 ne peuvent se propager en suspension en milieu Grace. Au bout de 72H d'infection les cellules High Five meurent, ces cellules ne sont pas adaptées à la culture en suspension.

#### **4. Coexpression de XPD avec ses partenaires**

Dans les cellules d'insectes, le partenaire p44 de XPD a été identifié par des expériences d'immunopurification utilisant des anticorps spécifiques (Coin et al., 1999). Notre méthode consiste à étudier les protéines qui sont susceptibles d'interagir par le système anti-flag ou par GST pull down (voir Matériel et Méthodes §D.1). Afin de valider cette méthode nous avons coexprimés dans les cellules d'insectes les fragments de l'hélicase (443-760 et 691-760) ainsi que l'hélicase entière (1-760) avec la protéine p44 flaguée. La protéine p44 est immobilisée sur les billes anti-flag, les fragments de XPD qui interagissent avec p44 sont alors retenus. Nous n'avons pas observé d'aspécificité des fragments XPD pour les billes anti-flag. L'expérience confirme que XPD interagit avec p44 via sa partie C-terminale (figure 35, piste 2). Nous pouvons constater que le fragment C-terminal court est suffisant pour permettre l'interaction avec p44 (figure 35, piste 3). Il apparaît que la zone d'interaction de XPD avec le partenaire p44 est localisée dans l'extrémité de la

partie C-terminale (691-760), dans laquelle de nombreuses mutations sont présentes chez les malades.

Nous avons également disséqué, de la même manière, la zone d'interaction de XPD avec les sous-unités MAT1 et p62 (figure 36 et 36A). Les fragments XPD (FL, 1-442, 443-760) sont coexprimés avec la protéine MAT1 ou p62 dans les cellules d'insectes. Les fragments XPD (entier, 1-442 et 443-760) sont fusionnés à la GST alors que MAT1 (entier) et p62 (1-157, 158-389, 390-548) sont fusionnés à un peptide FLAG afin de pouvoir contrôler l'interaction des deux côtés. Dans les deux configurations, XPD interagit avec MAT1 via sa partie N-terminale (figure 36, piste 3). Les partenaires XPDNter/MAT1 forment un complexe stable qui peut être purifié. La protéine XPD (FL) interagit également avec le domaine (158-389) de p62 (figure 36A). Le rôle de cette interaction n'est pas connue.

D'après son pattern d'interaction, XPD peut être définie comme une charnière entre le complexe core-TFIID et le complexe CAK. Elle peut être soit associée au core-TFIID via les sous-unités p44 et p62 soit au CAK via MAT1 et forment alors le CAK-XPD. XPD a un rôle dans l'organisation de l'assemblage du complexe TFIID et c'est ainsi qu'elle régulerait les activités enzymatiques du facteur de transcription.

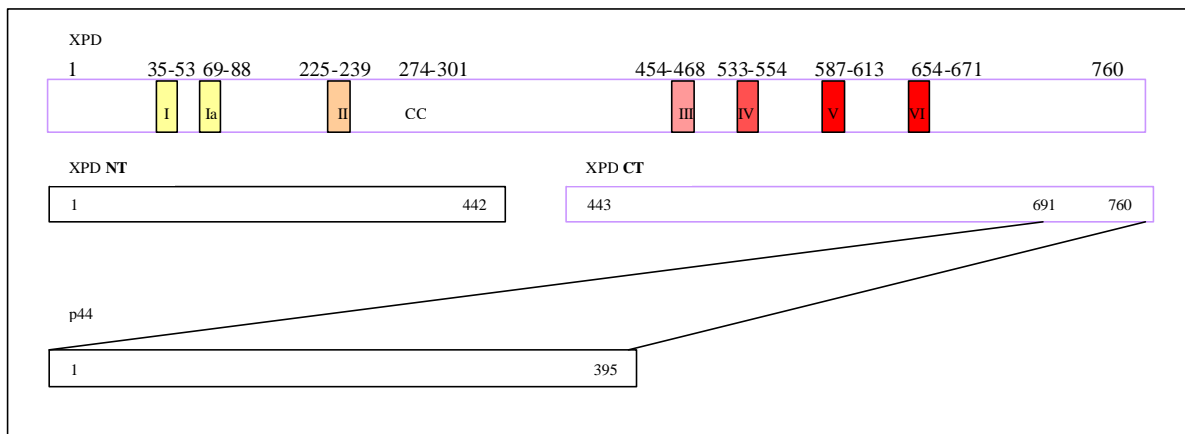
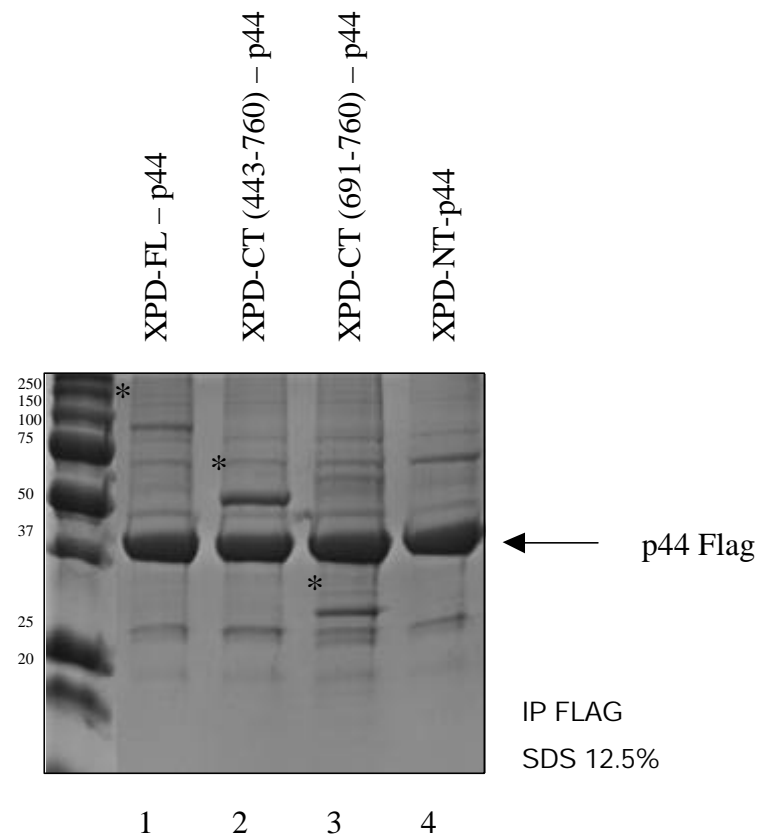


Figure 35 : Coexpression entre XPD et p44 dans les cellules d'insectes : La protéine p44 est immobilisée sur les billes via le peptide Flag. Les fragments de XPD (\*) sont indiqués. Les complexes sont visualisés après dépôt sur gel SDS 12.5% et coloration au bleu de Coomassie (FL: full length; IP:Immunopurification).

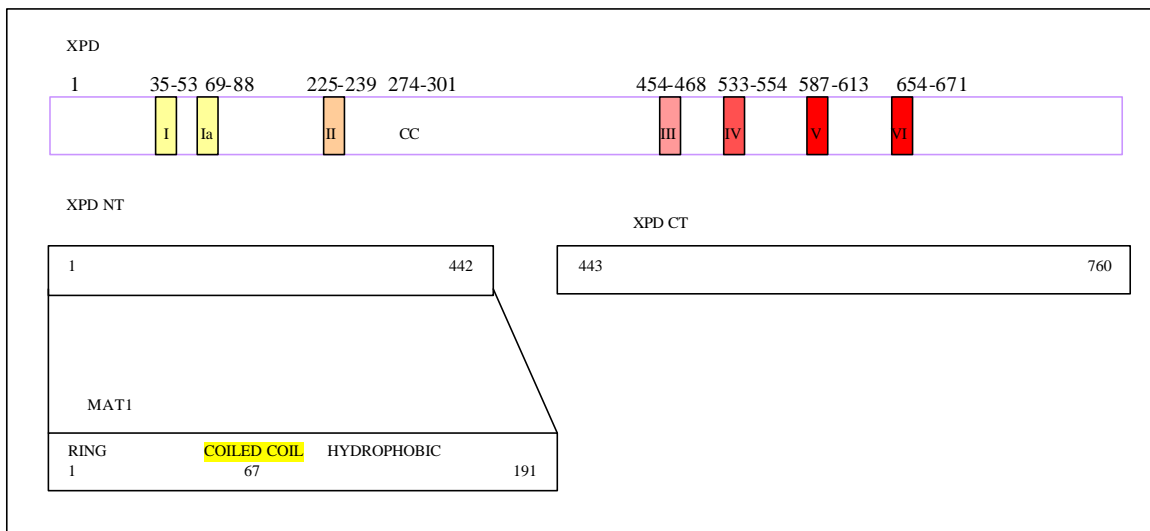
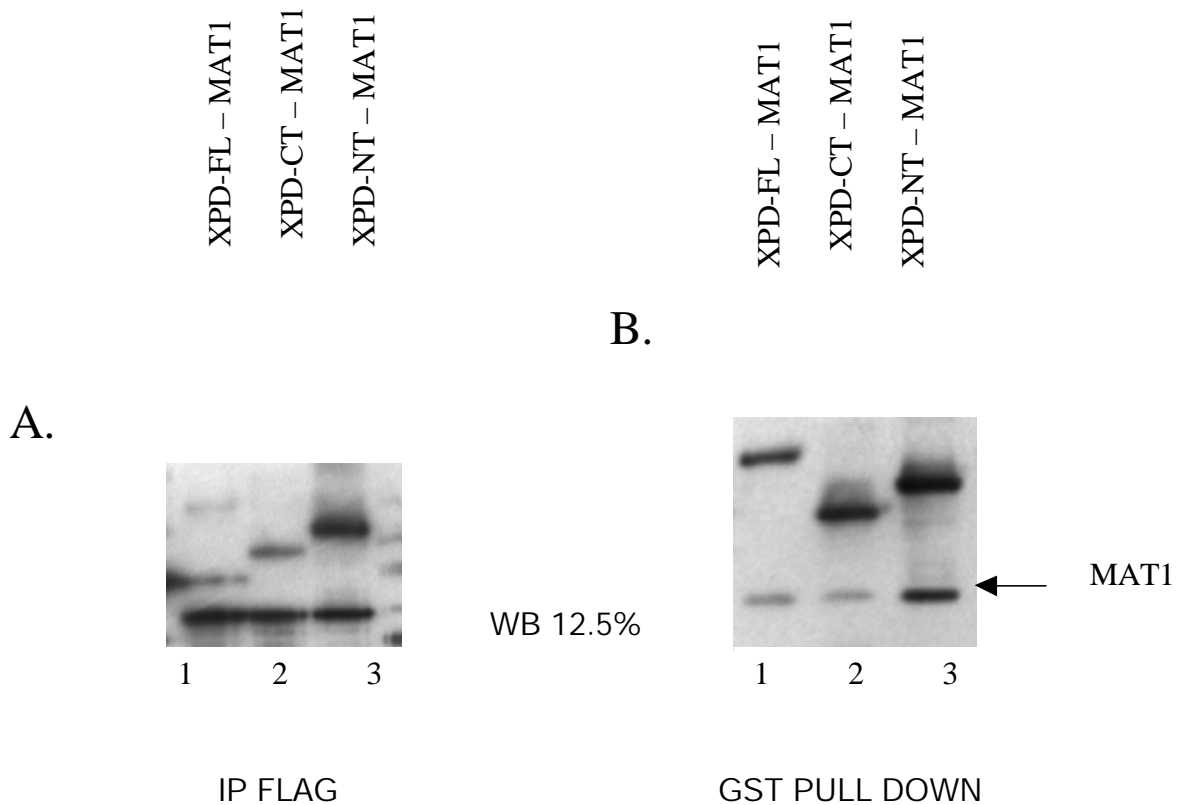


Figure 36 : Coexpression entre XPD et MAT1 dans les cellules d'insectes : Les protéines sont déposées sur gel SDS 12.5% et révélés avec les anti-corps monoclonaux correspondant à XPD et MAT1 par Western Blot. (A) Immunopurification anti-FLAG (IP Flag): MAT1-flag est immobilisé sur la résine, le fragment N-terminal de XPD est retenu. (B) GST Pull-down: les fragments XPD-gst sont immobilisés, MAT1 est retenu. XPD interagit avec MAT1 via sa partie N-terminale (1-442) (IP:Immunopurification).



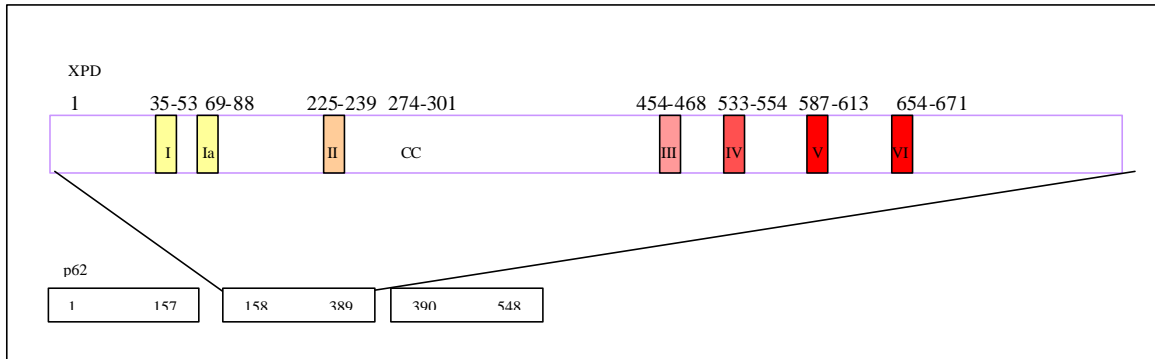
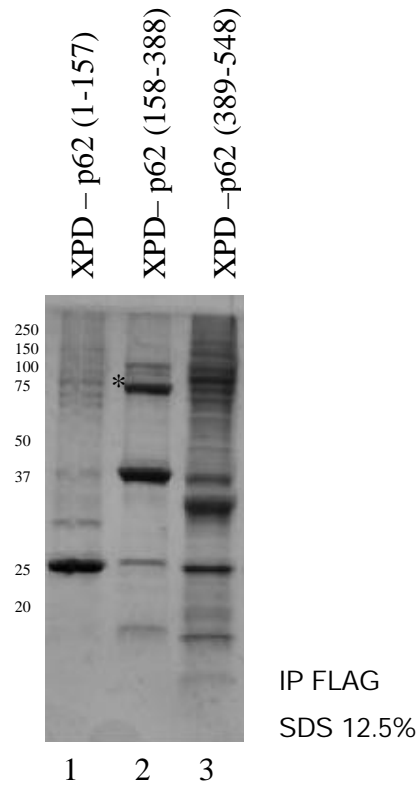


Figure 36A : Coexpression entre XPD et p62 dans les cellules d'insectes : Les fragments p62 sont immobilisés sur les billes via le peptide Flag. La protéine XPD (\*) interagit avec le fragment (158-389) de p62. Les protéines sont visualisées après dépôt sur gel SDS 12.5% et coloration au bleu de Coomassie (IP:Immunopurification).

## C. L'hélicase XPD d'archaebactérie

### 1. Généralités sur les archaebactéries

Les archaebactéries, considérées comme le troisième domaine vivant, présentent des similitudes significatives avec les eucaryotes et les bactéries dans la réplication, la réparation et la transcription du matériel génétique. Ce règne est petit à petit devenu un modèle d'étude pour la compréhension du monde eucaryote, plus complexe. Ceci est tout particulièrement vrai en ce qui concerne les études structurales, comme le corroborent les structures cristallographiques d'archaebactéries de la primase (Augustin et al., 2001), de l'ADN polymérase (Hashimoto et al., 2001), d'une DEAD-box hélicase (Story et al., 2001), de la PCNA sliding clamp (Matsumiya et al., 2001). Nous avons amorcé une étude structurale sur l'orthologue de l'hélicase XPD dans *Sulfolobus solfataricus* afin d'apporter une lumière dans la compréhension de mécanismes complexes que sont la transcription et la réparation de l'ADN et ceci par l'utilisation d'un modèle simple.

Depuis les travaux de Carle Woese (Woese et Fox, 1977) (Sogin et al., 1971) portant sur la phylogénie, le monde vivant est divisé en trois domaines (Woese et al., 1978) (Woese et Olsen, 1986): les Eucarya, les Bacteria et les Archaebactéries (figure 37). Le domaine des archaebactéries regroupe des organismes très divers aux niches écologiques variées et souvent très particulières (Kandler, 1994). On distingue d'un point de vue physiologique :

- les **halophiles** : ces organismes vivent dans des milieux à très forte concentration en sel (plus de 5M NaCl). Ils se développent dans les lacs salés ou dans les mines de potasse.

- les **méthanogènes** : ces organismes vivent en anaérobiose stricte (absence d'oxygène) et ils réduisent le dioxyde de carbone en méthane. Des enzymes et des coenzymes spécifiques de la méthanogénèse caractérisent ce type d'organismes. Ils peuvent être également thermophiles (température de 45°C à 65°C) ou hyperthermophiles (température supérieure à 70°C), halophiles, ou alcaliphiles. Les archaebactéries méthanogènes colonisent de nombreux milieux tels que le sol, les eaux stagnantes ou encore le tractus intestinal des animaux. Les espèces

thermophiles et hyperthermophiles sont présentes dans les sources chaudes et les zones volcaniques sous-marines.

- les **thermophiles extrêmes** (ou **hyperthermophiles**) : ces organismes sont soufre-dépendants et vivent à des températures supérieures à 65°C. Aérobie, ils oxydent le soufre en acide sulfurique. Anaérobie, ils réduisent le soufre en H<sub>2</sub>S. Ils peuvent se développer jusqu'à 113°C ou à pH 0,1.

D'un point de vue phylogénique, le domaine des Archaeobactéries est divisé actuellement en deux règnes (Woese et al., 1984) :

- Les **Crenoarcheota** qui regroupent les thermophiles soufre-dépendant. Ce règne comprend les Thermoprotéales, les Pyrodictiales et les Sulfolobales.

- Les **Euryarchaeota** qui regroupent les Thermococcales, les Méthanococcales, les Méthanobactériales et les halophiles extrêmes.

Le séquençage de différents génomes dans les trois domaines du vivant laisse apparaître une grande similitude au niveau du métabolisme carboné entre les archaeobactéries et les Bacteria. En revanche, le métabolisme de l'ADN (réplication, transcription...) chez les Archaeobactéries semble plus proche de celui des Eucarya.

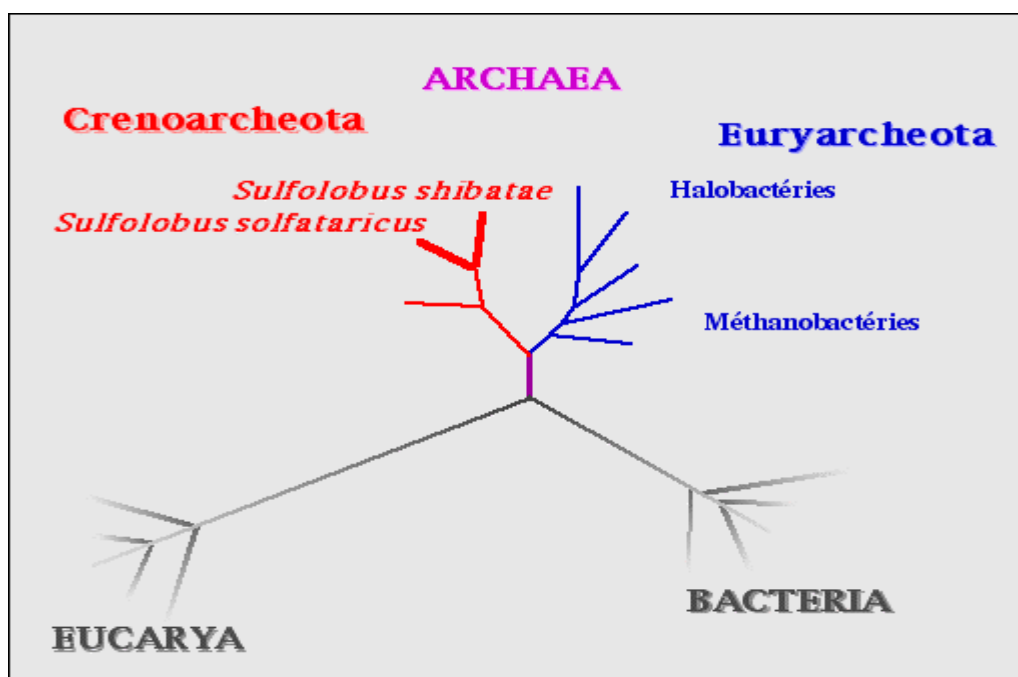


Figure 37 : Arbre phylogénique représentant les trois règnes : les Eucarya, les Bacteria et les Archaeobactéries. Le règne Archaeobactérien est subdivisé en deux domaines : les Crenoarcheota et les Euryarchaeota. *Sulfolobus solfataricus* appartient au règne des Crenoarcheota.

## **2. Les protéines de la chromatine chez les archaeobactéries**

Afin de mieux comprendre la mécanique des processus cellulaires chez les archaeobactéries, faisons un point sur les différentes protéines qui régissent le modelage de la chromatine et permettent ainsi l'accessibilité aux protéines de la transcription ou la réparation sur l'ADN.

Tous les organismes utilisent des protéines liant l'ADN double brin afin de relaxer ou de compacter leur matériel génétique. Chez les eucaryotes, ce mécanisme très complexe fait intervenir le nucléosome (voir Chapitre 1, §A.1). Chez les archaeobactéries, qui contiennent un petit génome circulaire, d'environ 0.5 à 5.5 Mbp, la mécanique protéique paraît moins complexe. La plupart des Euryarchaeota, mais pas les Crenoarcheota, possèdent des histones ; plus d'une trentaine de séquences d'histones ont été déduites des alignements de séquences à partir des génomes d'archaeobactéries connus. De plus, la présence de nucléosomes tétramériques formés à partir d'histones a été démontré chez les Euryarchaeota (Sandman et al., 1998).

En plus des histones, la plupart des archaeobactéries possèdent une protéine liant l'ADN : Alba. Elle a été découverte chez *Sulfolobus solfataricus*, c'est une petite protéine de 10kDa (Forterre et al., 1999). Une structure cristallographique de Alba existe et un modèle du complexe Alba-ADN a été calculé à partir de la structure de la protéine native (Wardleworth et al., 2001; Wardleworth et al., 2002). Alba est un dimère en solution ; le modèle du complexe Alba-ADN suggère que le dimère, chargé positivement, contacte l'ADN sur une région d'environ 12 pb via une hélice de chaque monomère. *In vivo*, Alba est complexé à une déacétylase, Sir2. Cette enzyme est capable de déacétyler Alba *in vitro* et cette déacétylation a pour effet l'augmentation de l'affinité pour l'ADN et l'inhibition de la transcription *in vitro* (Bell et al., 2002). Chez les eucaryotes, la réaction d'acétylation et de déacétylation des histones résultent dans l'activation et la désactivation transcriptionnelle de la

chromatine. Malgré ces filiations enzymologiques, aucune donnée n'explique comment ni quand Alba est acétylée ou désacétylée *in vivo*. Alba reste toutefois un bon candidat pour la régulation de la transcription au niveau de la chromatine.

Une autre nucléoprotéine de 7 kDa, Sul7d, a été découverte chez *Sulfolobus solfataricus* et n'a pas encore été trouvée chez d'autres archaebactéries. De manière surprenante pour une si petite protéine, Sul7d possède une activité ATPase et fonctionne comme une chaperonne ATP dépendante : Guagliardi et ses coauteurs ont constaté sa capacité à renaturer des agrégats protéiques. La structure cristallographique de Sul7d en complexe avec l'ADN a été résolue (Su et al., 2000). Sul7d lie l'ADN sur une région d'environ 4 pb et induit l'ouverture d'une hélice. La présence de Sul7d ne provoque pas une inhibition de la transcription *in vitro* (Fiorentino et al., 2003). De plus Sul7d facilite le réassemblage de brins complémentaires à hautes températures (Guagliardi et al., 2004), permet un superenroulement négatif de l'ADN et compacte l'ADN relaxé (Napoli et al., 2002). Cette protéine de la chromatine semble donc avoir un effet sur le remodelage de la chromatine. Elle faciliterait également l'accès des protéines de la réparation sur l'ADN lésé. En effet, le gène codant pour Sul7d a été amplifié par RT-PCR sur des cellules irradiées aux UV, et son expression est régulée (Salerno et al., 2003).

SSB, pour single strand binding protein ou protéine liant un ADN simple brin est présente chez de nombreux organismes incluant les bactéries et les virus (Krejci et Sung, 2002). Chez les eucaryotes, elle est représentée sous le nom de RPA (pour replication protein A) (Iftode et al., 1999). SSB apparaît nécessaire à de nombreux processus impliquant l'ADN : réplication, recombinaison, réparation de l'ADN. SSB, tout comme RPA, comporte un motif OB connu pour son implication dans la liaison aux oligonucléotides et aux oligosaccharides (Theobald et al., 2003). La structure cristallographique de SSB de *Sulfolobus Solfataricus* a été résolue (Kerr et al., 2003) et confirme des conservations structurales entre archaebactéries et eucaryotes. Le motif OB est fortement similaire à celui de la RPA eucaryote, alors que la partie C-terminale, flexible, est plus proche de celle des bactéries. Cette partie C-terminale n'est pas impliquée dans la liaison à l'ADN mais interagit avec des partenaires protéiques. Récemment Richards et ses coauteurs ont observé une interaction stable de SSB avec l'ARN polymérase via cette partie C-terminale (Richard et al., 2004) .

Cette interaction a pour effet l'activation de la transcription. Dans ce sens, SSB pourrait remodeler la chromatine de manière transitoire autour du promoteur afin de permettre l'accès aux protéines de la transcription. En présence du complexe SSB-ARN polymérase, cet effet est réprimé et la transcription est activée. Ces données mettent en avant un rôle dans la transcription chez les archaebactéries, d'une protéine liant un ADN simple brin.

### **3. La transcription chez les Archaebactéries**

L'étude de la transcription chez les Archaebactéries est largement investiguée dans l'équipe du Dr Bell (Qureshi et al., 1997; Bell et Jackson, 1998a; Bell et Jackson, 1998b; Bell et Jackson, 2001; Bell et al., 1999). Leurs travaux précisent l'homologie fondamentale existant entre les Archaebactéries et les Eucarya au niveau de la transcription. En effet, les Archaebactéries possèdent un homologue de la TATA-binding protein (TBP) des Eucarya ainsi que du facteur de transcription de base TFIIB, nommé TFB ou aTFA. Les Archaebactéries disposent d'une ARN polymérase (RNAP pour RNA polymerase) de complexité et de composition similaires à la RNAP eucaryote. Le recrutement de l'ARN polymérase se fait via l'interaction directe entre TFB et RNAP. La présence d'un doigt de Zinc dans la région N terminal de TFB est requise (Magill et al., 2001b). La figure 38 illustre le complexe de préinitiation présent chez les archaebactéries.

Le complexe ternaire TBP/TFB/DNA a été résolu par cristallographie aux rayons X (Kosa et al., 1997) et démontre une similarité troublante avec la structure du complexe eucaryote TBP/TFIIB/DNA (figure 39). Ceci suggère que le dernier ancêtre commun des règnes Eucarya et des Archaebactéries possède un système de transcription composé de TBP/TFB/TFIIB et d'une ARN polymérase.

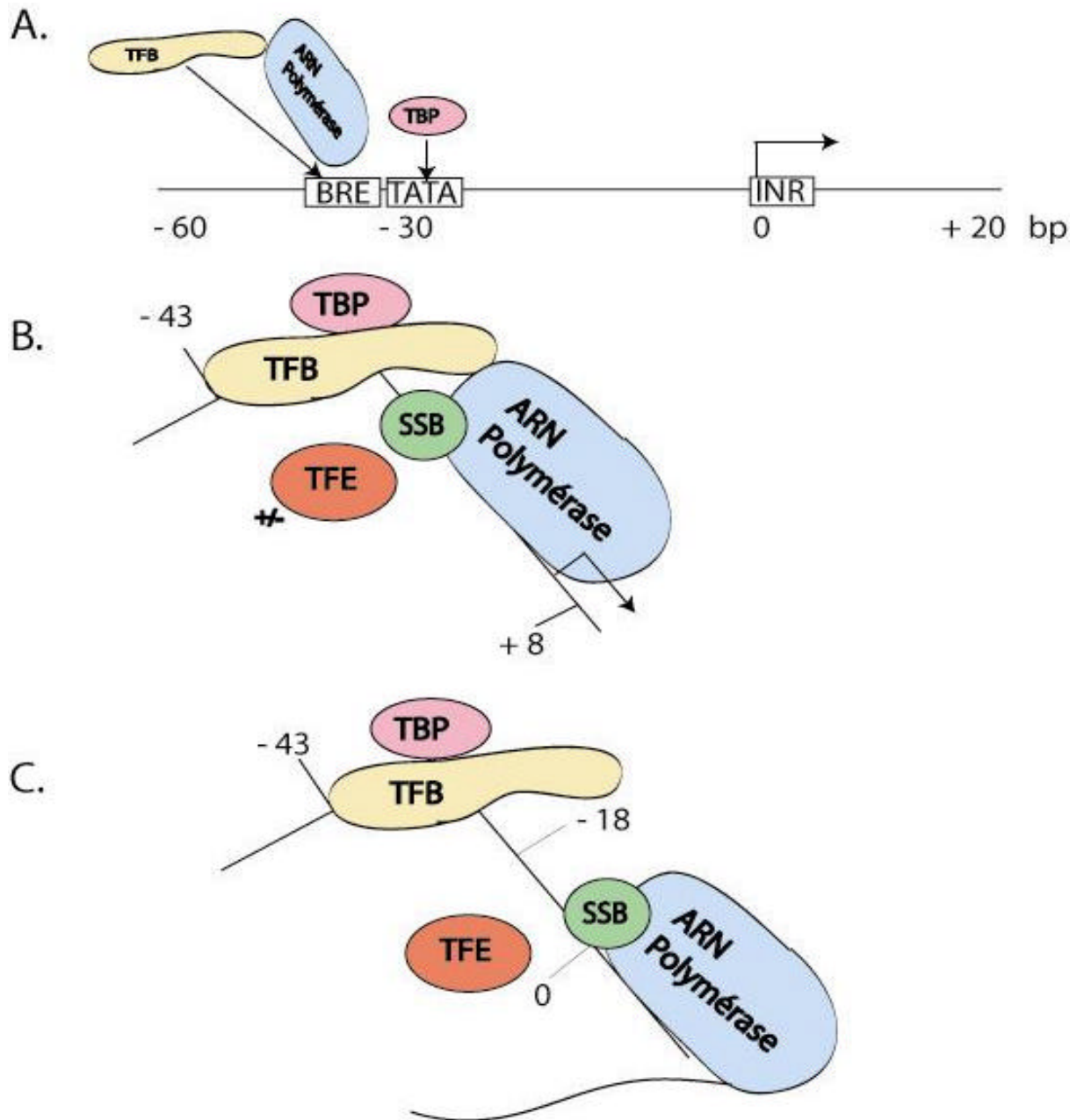


Figure 38 : Le complexe de préinitiation chez les archaebactéries. (A) Une région promotrice chez les archaebactéries : la liaison de TFB (l'homologue du facteur de transcription eucaryote TFIIB) sur l'élément BRE (Eléments de Reconnaissance de TFB) est nécessaire à la stabilisation de la liaison de la TBP (l'homologue de la TATA binding protein eucaryote) sur la boîte TATA *in vitro*. Cette étape ne nécessite pas la présence de l'ARN polymérase. (B) Le complexe de préinitiation de la transcription chez les archaebactéries est formé de la TBP, de TFB et de l'ARN polymérase. TFE (l'homologue du facteur de transcription eucaryote TFIIE) n'est pas présent dans tous

les systèmes de transcription archaens actuellement connus. La single-strand DNA binding protein (ssb) interagit avec la polymérase via sa partie C-terminale. Elle apparaît nécessaire à la formation du complexe de préinitiation (C) L'hybride ADN/ARN généré est de 8 à 12 pb. Nos connaissances actuelles du système de transcription d'archaebactéries ne nous permettent pas de dire si le complexe TBP et TFB reste lié au promoteur ou s'il est libéré.

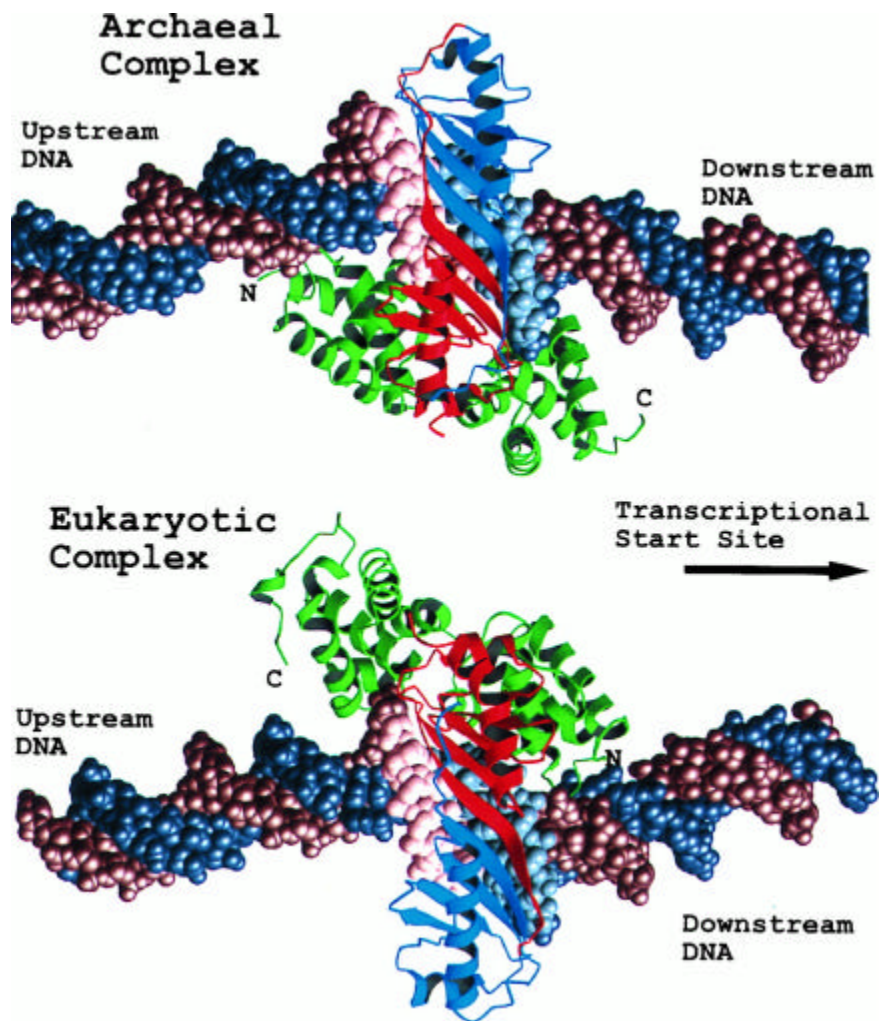


Figure 39 : (A) Structure cristallographique du complexe ternaire TBP/TFB/DNA chez les archaebactéries et (B) Structure du complexe eucaryote TBP/TFIIB/DNA. D'après (Kosa et al., 1997).



Malgré l'homologie fondamentale existant entre le système de transcription chez les Archaeobactéries et les Eucaryotes, il faut noter d'importantes différences. Ainsi, alors que l'hydrolyse de l'ATP est nécessaire lors de l'ouverture du promoteur par la machinerie transcriptionnelle eucaryote, elle n'apparaît pas essentielle chez les Archaeobactéries (Magill et al., 2001a). De même, l'ARN polymérase II eucaryote interagit avec les facteurs de transcription TFIIE et TFIIH alors que TFIIE et TFIIH ne sont pas présents chez les Archaeobactéries. Il en va de même pour les orthologues des facteurs de transcription eucaryotes TFIIA, TFIIIF, également absents.

#### **4. Régulation de la transcription**

La transcription chez les archaeobactéries présente, comme nous l'avons vu, des similitudes avec le système de transcription eucaryote. Pourtant du point de vue de la régulation de la transcription, le système archaéen semble plus proche du système bactérien que du système eucaryote. Ceci suggère que les régulateurs de la transcription existaient avant la bifurcation des lignées archaéennes et bactériennes. Quelques régulateurs de la transcription chez les archaeobactéries ont été caractérisés au niveau moléculaire. Deux homologues bactériens des protéines à leucine (Lsp pour Leucine-responsive regulatory protein), Lsr14 et Sa-Lrp, se lient elles-mêmes sur leur propre région promotrice, ce qui suggère une boucle d'autorégulation négative (Napoli et al., 1999; Enoru-Eta et al., 2000). Le régulateur putatif de la transcription MDR1 a été caractérisé par des tests de transcription *in vitro* chez *Archaeoglobus fulgidus*. MDR1 se positionne directement en amont du promoteur de la transcription reconnu par la TBP et TFB. La liaison de MDR1 sur son promoteur ne perturbe pas la formation du complexe de préinitiation mais permet la stabilisation de la liaison de l'ARN polymérase II sur le promoteur.

Bien que des données considérables soient maintenant disponibles sur la transcription chez les archaeobactéries, de nombreuses questions restent en suspens. L'homologie fondamentale existant entre les règnes archéens et eucaryotes peut permettre d'accéder à la compréhension de ces mécanismes fondamentaux. L'analyse des assemblages moléculaires avec une sous-unité ou un domaine protéique manquant, pourrait élucider le rôle moléculaire de ces sous-

unités au sein du complexe fonctionnel. Quant à la régulation de la transcription, nombre de régulateurs putatifs ont été identifiés, mais quant à savoir si ces régulateurs sont actifs par simple recrutement, comme chez leur homologue bactérien, rien n'est certain à ce jour. Finalement, l'effet des protéines de la chromatine sur la transcription est actuellement étudié mais sur des ADN nu *in vitro* : il serait intéressant d'étudier leur effet *in vivo*.

### **5. La réparation par excision resynthèse des nucléotides (NER pour Nucleotide Excision Repair) chez les archaebactéries**

La réparation des lésions sur l'ADN, incluant les lésions dues à l'irradiation par les Ultra-Violets (UV), est gouvernée par les protéines de la réparation (voir Chap.I, §C La réparation de l'ADN). Chez les Archaebactéries, et plus précisément chez les méthanogènes, le système de réparation de l'ADN par excision resynthèses des nucléotides est homologue au système bactérien UvrABC. Les Archaebactéries qui ne possèdent pas le système UvrABC, disposent des orthologues des nucléases XPF, XPG ainsi que des hélicases XPD et XPB qui sont impliquées dans le système de réparation par excision resynthèse des nucléotides chez les eucaryotes. Aucun homologue des protéines XPA et XPC du NER eucaryote n'ont été détecté chez une archaebactérie, ce qui suggère que ces protéines sont apparues plus tard dans le mécanisme de la réparation de l'ADN par NER. Toutefois, très peu d'évidences sont à ce jour connues concernant le NER chez les Archaebactéries. Le tableau 8 récapitule les protéines liant l'ADN et les protéines de la réparation dans les génomes d'Archaebactéries séquencés à ce jour.

Les travaux de l'équipe du Dr Salerno ont démontré l'induction de dimère de pyrimidine cyclobutane (CPD) après irradiation des cellules de *Sulfolobus solfataricus* aux ultra-violets . Et surtout, que les lésions sont éliminées *in vivo* dans le noir. Ceci suggère l'existence d'un système de réparation par excision resynthèse des nucléotides au sein des hypertermophiles. Dans la figure 40, un modèle du système de réparation par NER est proposé.

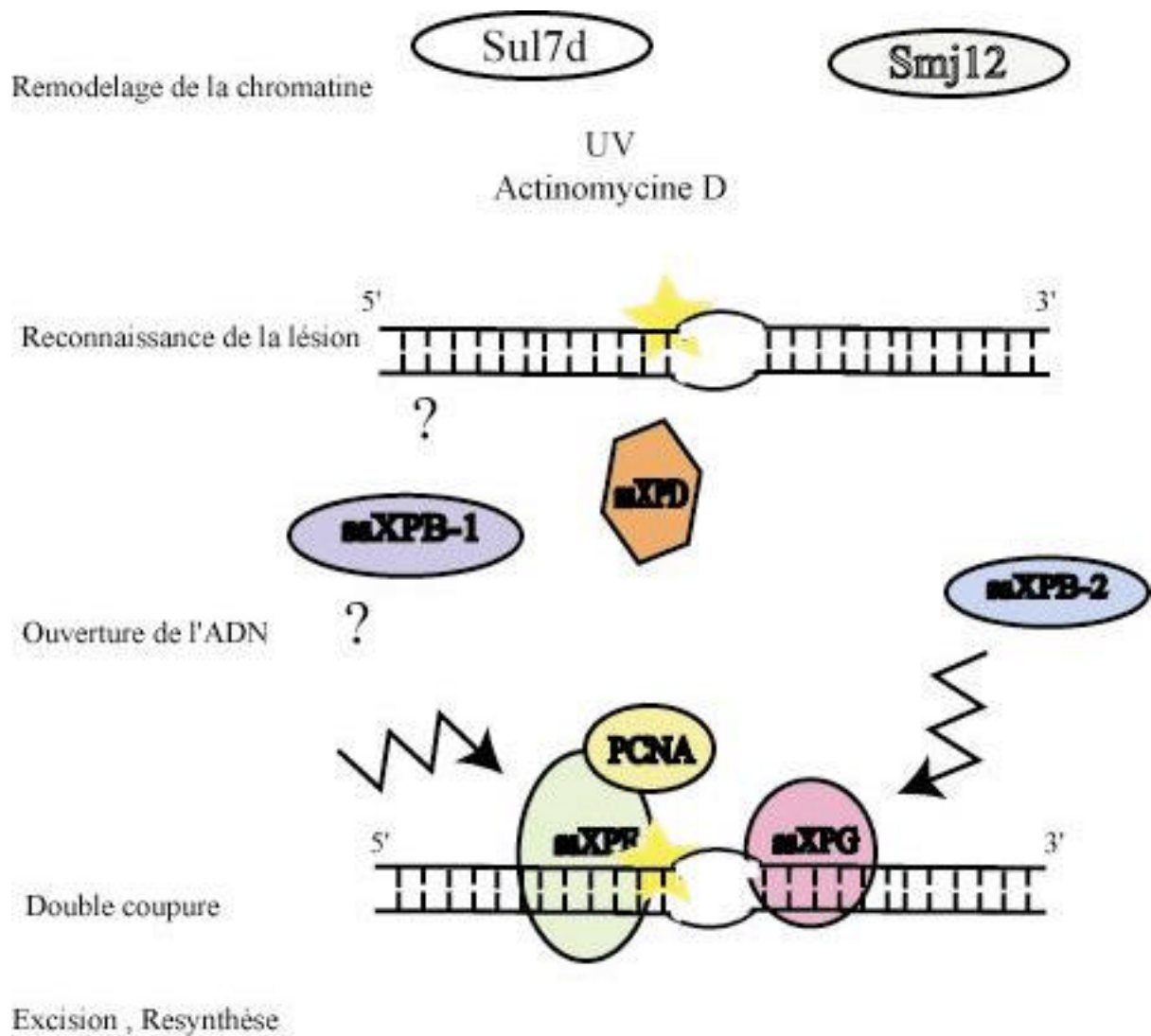


Figure 40 : Les archaebactéries possèdent un système de réparation par excision resynthèse des nucléotides. L'irradiation aux UV des cellules de *Sulfolobus solfataricus* induit la formation de dimères de pyrimidine sur l'ADN. La transcription des gènes des protéines du NER ssXPF, ssXPG et ssXPB est induite. Deux protéines de la chromatine, Sul7d et Smj12 pourraient faciliter l'accès à la lésion des protéines de la réparation en remodelant la chromatine. Sul7d et Smj12 ont des effets opposés sur la chromatine. Alors que l'expression de Sul7d, qui condense l'ADN est réprimée, celle de Smj12, qui relaxe l'ADN, est augmentée. Les acteurs des étapes de reconnaissance de la lésion, d'ouverture de l'ADN et d'excision resynthèse ne sont pas connus. La présence des nucléases XPF/PCNA et XPG suggère une contribution majeure dans la double coupure de l'ADN lésé.

Espèces	XPF	XPD	XPB	PCNA	RPA	SSB	Histones	Alba	UvrABC
<i>Aeropyrum pernix</i>	1	1	1	3	0	1	0	2	0
<i>Pyrobaculum aerophilum</i>	1	1	1	2	0	0	0	1	0
<i>Sulfolobus solfataricus</i>	1	1	2	3	0	1	0	2	0
<i>Sulfolobus tokodaii</i>	1	1	1	3	0	1	0	2	0
<i>Thermoplasma acidophilum</i>	0	0	0	1	1	1	0	1	0
<i>Thermoplasma volcanium</i>	0	0	0	1	1	1	0	1	0
<i>Methanopyrus kandleri</i>	1	1	1	1	1	0	4	2	0
<i>Pyrococcus furiosus</i>	1	1	1	1	1	0	2	1	0
<i>Pyrococcus horikoshii</i>	1	1	1	1	1	0	2	1	0
<i>Pyrococcus abyssi</i>	1	1	1	1	1	0	2	1	0
<i>Methanococcus jannaschii</i>	1	1	1	1	1	0	5	1	0
<i>Archeoglobus fulgidus</i>	1	1	1	1	1	0	2	2	0
<i>Methanothermobacter Thermautotrophicus</i>	1	0	0	1	1	0	3	1	1
<i>Methanosarcina acetivorans</i>	1	0	0	1	1	0	1	0	1
<i>Methanosarcina mazei</i>	1	0	0	1	1	0	1	0	1
<i>Halobacterium</i>	1	0	0	1	1	0	4	0	1

Tableau 8 : Distribution des protéines liant l'ADN et des protéines de la réparation dans les génomes d'Archaeobactéries séquencés à ce jour. Les crenarcheae apparaissent en gras. Les protéines histones sont quantifiées sur la base du nombre total de signatures d'histones présents dans les séquences de chaque organisme.

L'équipe du Dr Roberts ont caractérisé biochimiquement la nucléase XPF de *Sulfolobus solfataricus* (Roberts et al., 2003). Chez l'homme, XPF forme un hétérodimère avec ERCC1 (Enzlin et Scharer, 2002). L'hétérodimère XPF/ERCC1 reconnaît les jonctions double brin/simple brin lorsque l'ADN présente une jonction d'une polarité 5'- 3'. *In vivo*, l'entrée de XPF/ERCC1 au niveau de la machinerie de réparation permet l'incision puis l'élimination de la lésion. L'orthologue d'XPF chez *Sulfolobus solfataricus* est composé de 233 acide aminés (en comparaison à la protéine humaine composée de 367aa ). Toutefois le domaine nucléase en partie C-terminale de la protéine est conservé. Son activité est dépendante de l'interaction via sa partie C terminale avec PCNA, qui est un hétérotrimère dans *Sulfolobus*

*solfatarius* (Dionne et al., 2003). D'après Robert et ses coauteurs XPF-ERCC1 est capable d'ouvrir un jonction simple brin/double brin avec une extension simple brin 3', ce qui suggère fortement que XPF est une endonucléase qui fait partie intégrante d'un système de réparation chez les archaebactéries.

La régulation de la transcription après induction d'un dommage sur l'ADN a été étudiée chez les orthologues des gènes du NER eucaryotes XPF, XPG, XPB. L'expression des gènes codant pour les deux nucléases ssXPF et ssXPG apparaît modérée après une exposition aux UV. Concernant ssXPB, il est intéressant de noter que *Sulfolobus solfataricus* possède deux paralogues de l'hélicase humaine XPB (ss-XPB-1 et ss-XPB-2). L'expression des gènes, après irradiation, de ssXPB-1 est fortement induite alors que celle de ssXPB-2 est normale. Ceci suggère un rôle distinct des deux paralogues qu'il serait très intéressant de caractériser. Des analyses de séquences mettent en lumière une similarité de séquence de l'ordre de 58% ainsi qu'une conservation des motifs et des signatures hélicases entre autres la DEAH/D box, site de liaison à l'ATP. L'hélicase humaine étant à ce jour peu connue, l'étude des paralogues de XPB dans *Sulfolobus solfataricus* est une alternative intéressante pour l'aspect structural, la production de la protéine d'archaebactéries apparaissant plus réalisable que la production de la protéine humaine. En outre, la compréhension du mécanisme de réparation de l'ADN chez les archaebactéries, où ces protéines sont présentes seules et non complexées à d'autres protéines, présente moins de complexité que chez les eucaryotes.

## D. Etude structurale de l'homologue de l'hélicase XPD de *Sulfolobus solfataricus*

Nous avons identifié, d'après les alignements de séquence, une nouvelle hélicase putative chez les archaebactéries. L'étude structurale de l'homologue de l'hélicase humaine XPD chez *Sulfolobus solfataricus* est présentée.

### 1. Motifs conservés

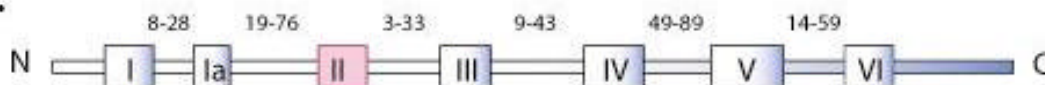
L'identité, c'est à dire la quantité de résidus homologues, est de 24% entre l'hélicase putative d'archaebactérie et l'hélicase humaine XPD. La comparaison des séquences de la protéine humaine aux différents orthologues d'archaebactéries est représentée (figure 42). Le pourcentage de conservation des motifs hélicases (I, Ia, II, III, IV, V et VI) entre la protéine XPD humaine (XPDhs) et la protéine XPD de *Sulfolobus solfataricus* (XPDss) est indiqué dans le tableau B de la figure 41. Quatre motifs hélicase, le motif I, II, V et VI présentent une homologie supérieures à 50% entre les deux espèces.

Le motif I de l'hélicase XPD d'archaebactérie présente une forte homologie avec les différents orthologues *Pyrococcus abyssi* (pa), *Pyrococcus furiosus* (pf), *Aeropyrum pernix* (ap) et *Methanococcus jannaschii* (mj). Par rapport à l'hélicase XPD humaine, la conservation est assez élevée, la moitié des résidus étant identique. Des structures cristallographiques d'hélicase montrent un motif I très proche du motif I des hélicases d'archaebactéries : PcrA(31-38) : AGAGSGKT, Rep(22-29) : AGAGSGKT et Ns3(204-211) : APTGSGKS par rapport à XPDss (28-36) : APTGSGKT. Dans toutes les structures le groupe amino-terminal de la lysine interagit avec le phosphate du MgATP/MgADP et le groupe hydroxyle de la sérine ou de la thréonine lie le Mg<sup>2+</sup> (Hickman et al., 2002) (Yan et al., 1998).

La boîte DEAH, spécifique des hélicases à ADN de la superfamille SF2, est totalement conservée entre XPDhs et XPDss. Dans toutes les structures cristallographiques citées au dessus, l'acide aspartique en position 1 de la boîte lie l'ion Mg<sup>2+</sup>. De plus, dans la structure de l'hélicase Ns3 du virus de l'hépatite C (HCV), l'histidine en position 4 de la boîte DECH (290-293) interagit avec le motif VI de l'hélicase .

Les motifs V et VI présentent une homologie de plus de 50% entre XPDhs et XPDss. Le motif est très conservé parmi les orthologues archaéens. Le motif VI de XPDss(506-518) : QSIGRAIR a également une forte homologie avec le motif VI de la protéine bactérienne UvrB(531-538) : QTIGRAAR. Dans la structure de UvrB, la glutamine en position 1 interagit avec l'histidine 336 du motif II de l'hélicase (Theis et al., 1999). Cette similitude est tout à fait intéressante. Nous avons vu auparavant que le système de réparation chez les archaebactéries a subi une divergence au cours de l'évolution. Certaines archaebactéries possèdent un système de réparation de l'ADN homologue au système bactérien UvrABC et les autres disposent des orthologues des hélicases XPD et XPB qui sont impliquées dans le système de réparation par excision resynthèses des nucléotides chez les eucaryotes. On peut émettre l'hypothèse que l'hélicase putative de *Sulfolobus solfataricus* a évolué dans cette caractéristique fonctionnelle chimérique : peut-être une marque phylogénique de l'évolution des trois règnes ?

A.



B.

Motifs	XPD hs	XPD ss	Homologie
Motif I	MPSGTGKTVS	APTGS <b>G</b> KTLF	50%
Motif Ia	KLIYCSRTVPEIEKVIEELR	KVAFV <b>VR</b> THNEFFPV <b>Y</b> RELS	37%
Motif II	DEAHNID	DEAH <b>N</b> IE	86%
Motif III	PLDIYPKILDFH	PLDYLRN <b>I</b> IGIK	48%
Motif IV	VAFFTSYQY	LAIFPS <b>Y</b> EF	44%
Motif V	ILLSVARGKVSEGI	IIMGI <b>ARG</b> KLAEGI	58%
Motif VI	HAAQCVGRAIRGK	AVK <b>Q</b> SIG <b>R</b> AIRGP	54%

Figure 41 : (A) Représentation schématique de l'hélicase humaine XPDhs. Les sept motifs hélicases, ainsi que leurs positions, sont indiqués. (B) Tableau récapitulatif des motifs hélicases retrouvés chez XPDhs et chez XPD de *Sulfolobus solfataricus*. L'homologie entre les motifs hélicase de XPDhs et XPDss est représentée en pourcentage.

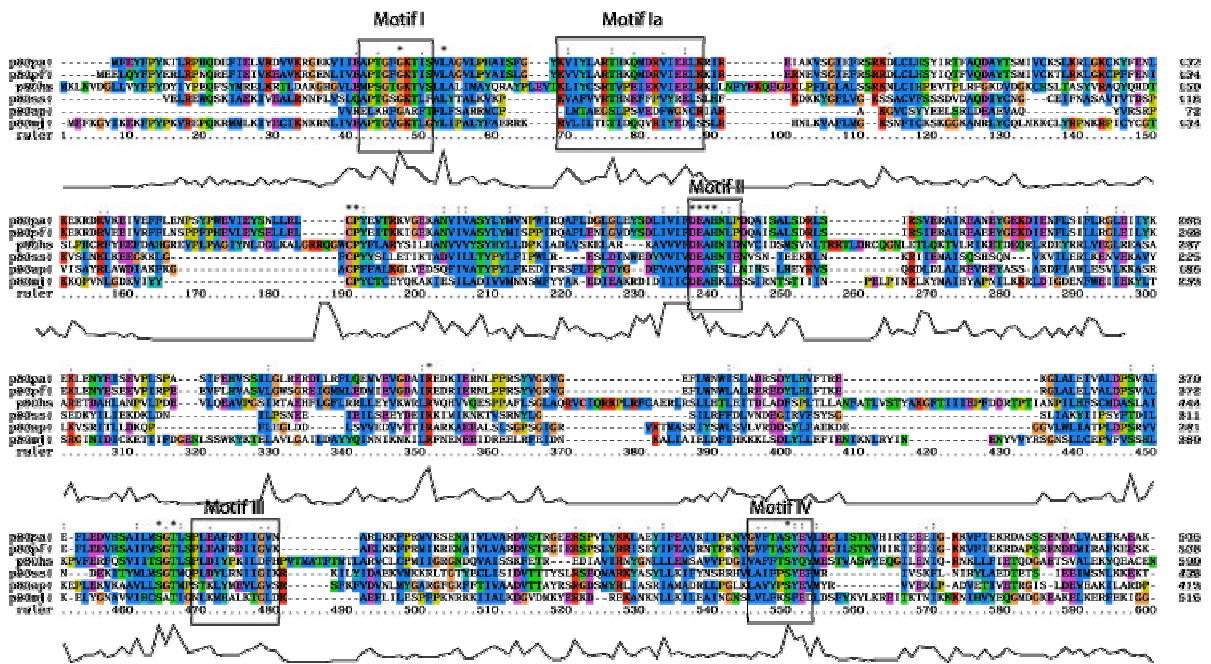


Figure 42 : Alignement des hélicases d'archaebactéries et représentation des motifs hélicase.



## 2. Purification de l'hélicase putative XPD de *Sulfolobus solfataricus*

Le gène codant pour l'hélicase putative a été amplifié à partir de l'ADN génomique de *Sulfolobus solfataricus*. L'ADNc a été introduit dans le vecteur d'expression pDEST17 permettant la fusion de la protéine à une étiquette polyhistidine (His tag) dans *Escherichia coli*. Le clonage a été effectué à l'aide de la technologie Gateway (voir Matériel et Méthodes). L'expression de la protéine est testée en premier lieu dans différentes souches bactériennes. L'expression de l'hélicase nécessite la présence des ARNt codant pour les codons rares AGG, AGA, CCC apportés par la souche BL21-Codon Plus RP. La souche BL21-Codon Plus RP est transformée par le plasmide codant pour l'hélicase et est cultivée en présence d'ampicilline qui sélectionne le pDEST17, et de chloramphénicol qui sélectionne le vecteur-pRare. La protéine est purifiée à partir de 6 litres de culture en milieu d'auto-induction. La figure 43 illustre les schéma de purification de l'hélicase putative de *Sulfolobus solfataricus*. Après sonication des extraits totaux et ultracentrifugation, les extraits solubles contenant la protéine sont incubés avec la résine Talon (Clontech). La résine Talon permet de sélectionner les protéines fusionnées à l'étiquette polyhistidine. Un lavage avec du tampon contenant une faible quantité d'imidazole (10mM) permet d'éliminer une majeure partie de protéines contaminantes. La protéine est ensuite éluée par compétition avec une plus forte concentration d'imidazole (250mM). L'étiquette polyhistidine est ensuite coupée sous l'action de la thrombine pendant une nuit à 37°C. Une deuxième étape de purification consiste à chauffer la protéine à 70°C pendant 15 minutes, les protéines d'*Escherichia coli* sont à cette étape complètement dénaturées alors que la protéine de *Sulfolobus solfataricus* est thermostable et reste dans le surnageant. La protéine subit une dernière étape de purification par passage sur une colonne échangeuse de cations (MonoS) et est éluée par un gradient en sel. La quantité de sels est ajustée à 100mM par trois passages sur Centricon 50K. La protéine est finalement concentrée sur Centricon 50K à environ 2mg/ml.

L'hélicase putative de *Sulfolobus solfataricus* est dès lors utilisable pour des tests biophysiques et des essais de cristallisation.

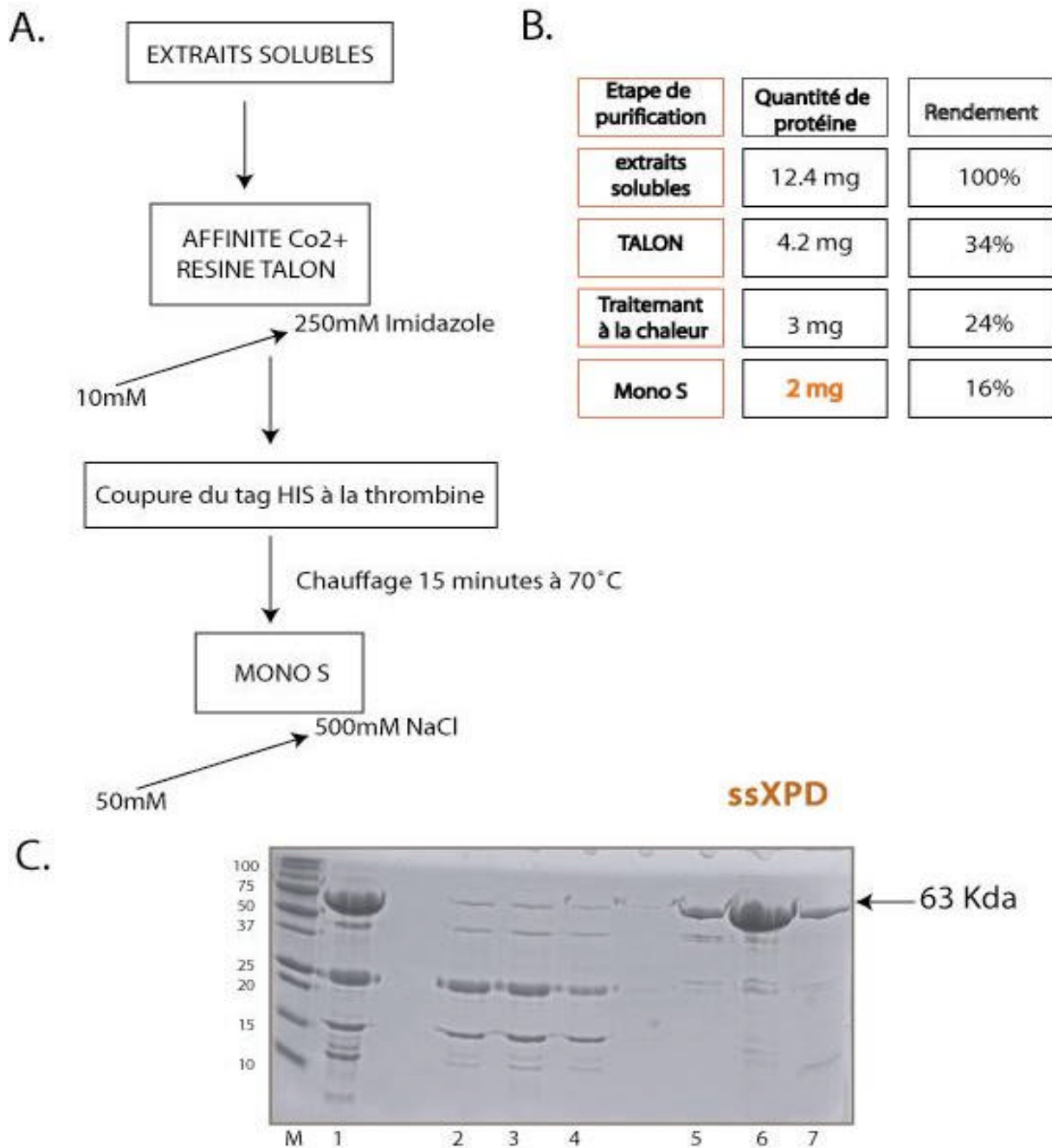


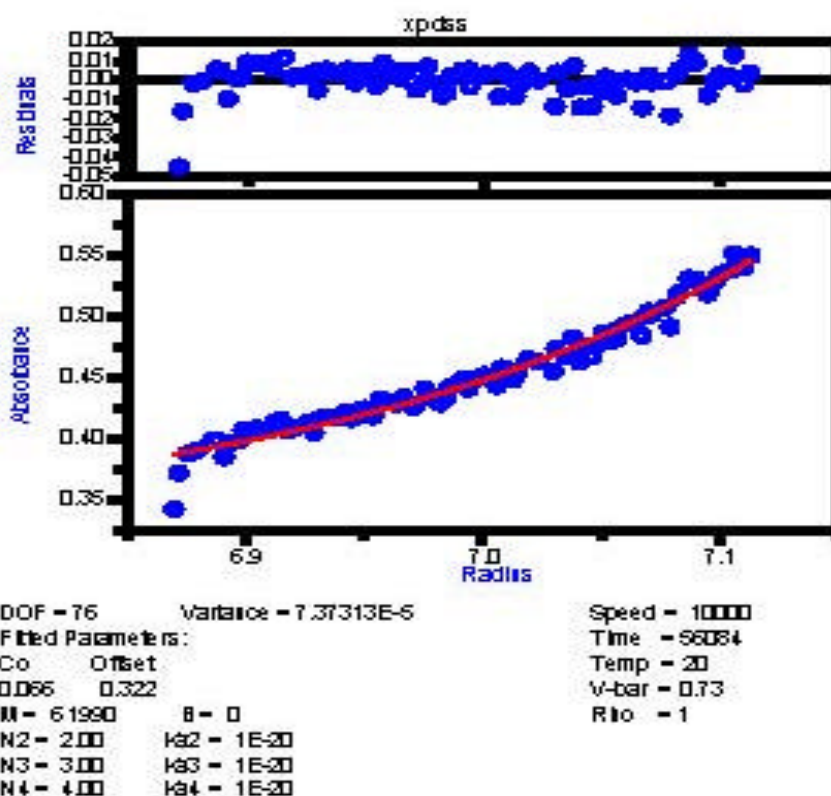
Figure 43 : Purification de l'hélicase XPD<sub>ss</sub> de *Sulfolobus solfataricus* (A) Schéma de purification : la purification se fait en quatre étapes. Les extraits solubles sont chargés sur une colonne Talon (Clontech) et la protéine est éluée en présence de 250mM d'Imidazole. Le tag 6-Histidine est coupé sous l'action de la thrombine pendant une nuit à 37°C. Les protéines contaminantes sont éliminées par chauffage à la chaleur pendant 15 minutes à 70°C. La protéine est chargée sur une colonne échangeuse de cations et éluée par un gradient en NaCl. (B) Tableau récapitulatif de la purification : le rendement

est calculé par rapport à la quantité de protéine présente dans les extraits solubles. (C)  
Purification sur colonne échange de cations (1) charge (2,3,4) lavages (5,6,7) élution.

### **3. Caractérisations biophysiques**

La qualité de l'échantillon est vérifiée par diffusion de lumière qui permet de mesurer rapidement l'état d'agrégation d'une protéine en solution. Une seule forme de population de même masse moléculaire est présente et très peu d'agrégats (< 2%) sont présents en solution. La masse moléculaire d'hélicase putative est de l'ordre de 63kDa comme attendue d'après la séquence.

La masse moléculaire de la protéine a été également mesurée par ultracentrifugation analytique. Cette méthode permet de mesurer la masse de la protéine mais surtout son état d'oligomérisation. Cette manipulation a été effectuée par le Dr Catherine Birck dans notre laboratoire. L'hélicase putative de *Sulfolobus solfataricus* apparaît sous forme monomérique dans les conditions de stockage de la protéine (Tris-HCl 20mM, NaCl 100mM) avec une masse moléculaire d'environ 60KDa. D'autres expériences sont encore à réaliser : définir l'état oligomérique de la protéine en fonction de la concentration en sels, de la présence d'ions bivalents comme le magnésium et de la présence d'un substrat comme l'ATP ou l'AMP-PNP, l'analogue non hydrolysable de l'ATP ou encore de l'ADN. Les constantes d'association de la protéine à son substrat restent également à définir.



#### 4. Cristallisation de l'hélicase putative XPDss

##### 4.1. Généralités sur la cristallisation

Le principe de la cristallographie de macromolécules est de porter progressivement une solution protéique contenant des agents précipitants à un état de sursaturation, provoquant ainsi la nucléation puis la croissance des cristaux (figure 44). Pour obtenir l'état de sursaturation, les précipitants habituellement utilisés sont les sels, les solvants organiques et les polyéthylènes glycol (PEG). Les facteurs qui influencent notablement la croissance des cristaux sont les concentrations du précipitant et de la protéine, le pH, la force ionique, la température et la pureté des agents chimiques et surtout de la protéine. Pour atteindre progressivement l'état de sursaturation, plusieurs méthodes existent. Nous avons utilisé la méthode la plus fréquemment employée, celle de la diffusion en phase vapeur.

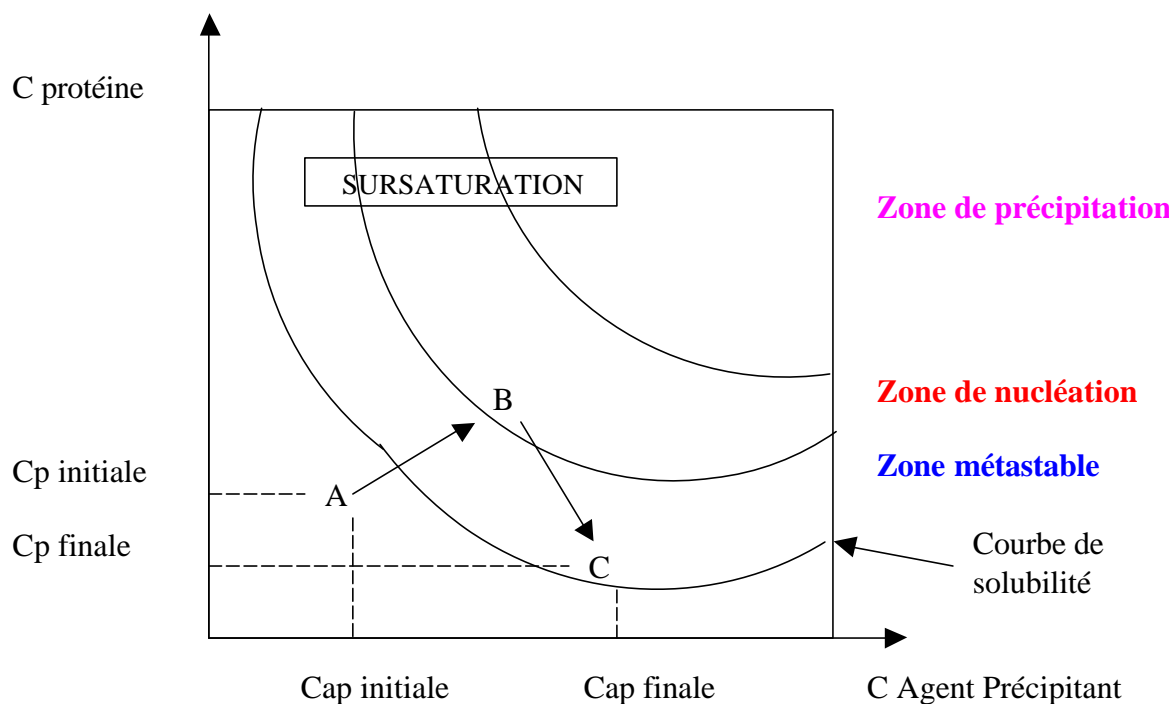


Figure 44 : Diagramme schématisé de solubilité montrant la variation de la concentration d'une protéine en fonction de la concentration en agent précipitant au cours de la cristallisation par diffusion de vapeur.

#### 4.2. La cristallisation par la méthode de diffusion en phase vapeur

Une goutte contenant la protéine, le précipitant, le tampon est mise en équilibre avec le réservoir contenant l'agent précipitant à une concentration plus élevée que dans la goutte.

L'équilibre se fait par diffusion de l'agent volatil qui diffuse entre la goutte et le puits jusqu'à ce que la concentration soit identique dans les deux compartiments. Dans notre cas, la goutte est obtenue à partir de deux volumes égaux (2 $\mu$ l) de solution protéique et de solution du réservoir. Ainsi, la composition du puits fixe les conditions de cristallisation. La vitesse de l'échange d'eau dépend des forces ioniques des deux compartiments.

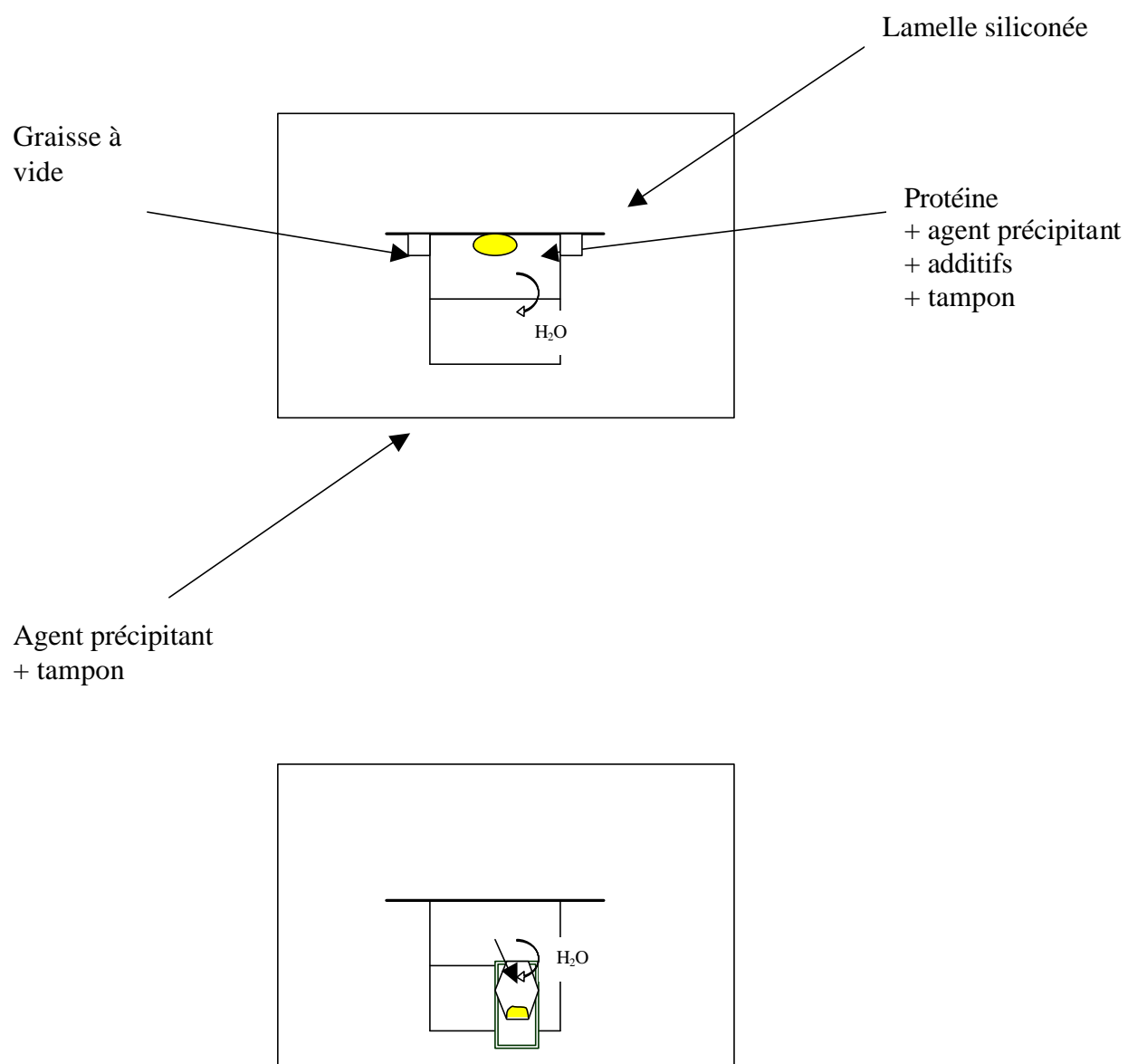


Figure 45 : Principe de la méthode par « goutte suspendue » et de « la goutte assise ».

Avant l'enregistrement des données, le cristal doit être congelé sous flux d'azote gazeux à 100K. La durée de vie du cristal est ainsi augmentée et la dégradation due au rayonnement X est réduite. Comme le cristal contient du solvant, qui est principalement de l'eau, la congélation sans cryoprotectant cristallise cette eau sous forme de microcristaux qui détériorent le cristal et donnent des anneaux de diffraction sur les clichés. Les principaux cryoprotectants couramment utilisés sont le glycérol, l'éthylène glycol, les PEG de faible masse moléculaire. Ils transforment l'eau en glace amorphe, empêchent la formation des cristaux de glace sans détériorer le cristal. A l'aide d'une boucle, le cristal est transféré dans une goutte de solution cryoprotectante. Le cristal peut alors être congelé sous le flux d'azote ou dans l'éthane liquide.

#### 4.3. Enregistrement des données

Dans un premier temps, les données sont enregistrées sur le diffractomètre du laboratoire. Ceci permet de vérifier que le cristal est bien de la protéine et non un cristal de sels (le cliché de diffraction du sel est caractéristique) et d'apprécier le pouvoir diffractant du cristal.

La longueur d'onde du rayonnement utilisé correspond à la raie du cuivre :  $\lambda = 1.54\text{\AA}$ . La boucle qui contient le cristal est placée sous la tête goniométrique du diffractomètre, de sorte que le cristal se trouve au niveau du jet d'azote gazeux. Le cristal est centré en utilisant les réglages de la tête goniométrique de façon à ce qu'il tourne sur lui-même au cours de sa rotation autour de l'axe du goniomètre. Ce réglage garantit que le cristal reste à tout moment dans le faisceau des rayons X ce qui permet de garder une intensité constante du rayonnement frappant le cristal.

Dans un second temps, lorsque l'on est assuré que le cristal est bien un cristal de protéine, une série de cristaux est testée au synchrotron. Le rayonnement synchrotron est un rayonnement blanc très intense, son émission est très focalisée et de structure temporelle pulsée. Son utilisation permet une réduction des temps d'exposition de plusieurs ordre de grandeur, un gain de résolution, une meilleure résolution spatiale des taches de diffraction et une optimisation du rapport signal/bruit. Les cristaux ont été testés aux synchrotrons SLS en Suisse par le Dr Marc Ruff et à l'Advanced Photon Source de Chicago par le Dr André Mitschler.

#### 4.4. Le problème de la phase en cristallographie

Pour résoudre la structure d'une macromolécule en cristallographie il faut trouver la phase et l'amplitude du facteur de structure de la molécule, nombre complexe égal à la transformée de Fourier de la densité électronique. Le cristal est exposé au rayonnement électromagnétique des rayons X, dont les longueurs d'onde (comprises entre 0,01 et 10 nm) sont de l'ordre des distances qui séparent les plans atomiques des réseaux cristallins. Lorsque le cristal à étudier est irradié par un fin faisceau de rayons X, chacun des atomes du cristal réfléchit une onde de faible amplitude, qui se propage dans toutes les directions. Les ondes issues des atomes interfèrent, faisant apparaître sur le film photographique qui les reçoit des «taches» qui correspondent au maximum des ondes en phase; les autres, en opposition de phase, se sont annulées. La mesure de l'intensité diffractée  $I(hkl)$  de chacune des taches de diffraction d'indice de Miller ( $hkl$ ) permet de déterminer l'amplitude du facteur de structure (l'intensité étant proportionnelle au carré du module du facteur de structure) mais la phase de ce dernier reste inconnue. Pour obtenir cette phase, différentes méthodes existent : la diffusion anormale, le remplacement isomorphe, le remplacement moléculaire et les méthodes directes pour les petites molécules. Le remplacement isomorphe consiste à introduire un atome lourd avec un grand nombre d'électron qui va se fixer sur la molécule soit de façon covalente, soit par des liaisons électrostatiques, Van der Waals en un ou plusieurs sites. L'utilisation d'au moins deux dérivés lourds ou d'un dérivé lourd avec une contribution anormale, c'est à dire un changement de phase dû à la nature de l'atome, permet de lever l'indétermination de la phase. La carte de Patterson, qui se calcule à partir des intensités et permet d'obtenir les distances interatomiques, devient rapidement lisible si un métal lourd est présent car la phase et l'amplitude de ce dernier vont dominer. Dans le cas du remplacement moléculaire, la détermination des phases nécessite l'utilisation de structure connues proches de la protéine. L'hélicase XPD de *Sulfolobus solfataricus* présente de faibles homologues de séquence avec les hélicases dont la structure est actuellement résolue comme UvrB (code PDB : 1D9Z) ou PcrA (code PDB : 2PJR). Les sept motifs hélicases, très conservés, pourraient servir de matrice à la reconstruction de la structure.



#### 4.5. Cristallisation de l'hélicase putative XPDss

Les essais de cristallisation sont réalisés par la méthode de diffusion en phase vapeur. Dans une condition, la protéine cristallise en 48 heures à 22°C. Les gouttes (2µl) de cristallisation sont composées d'un volume égal de protéine (2mg/ml) et de réservoir : acide succinique, glycine, phosphate à 100mM, pH8 et 25% de PEG 3350 comme agent précipitant. La forme cristalline obtenue est un cube de dimensions 200x150nm. Les conditions de congélation des cristaux sont testées au laboratoire et l'ajout de 15% de glycérol rend la solution cryoprotectante. Les cristaux sont testés au synchrotron SLS en Suisse.

Le groupe d'espace du cristal est cubique à face centrée I23. Les paramètres de maille sont  $a=b=c=154.29$ . Cette forme cristalline ne diffracte qu'à une résolution d'environ 5Å. Ceci est dû à la très haute symétrie de ce groupe d'espace et la forte proportion de solvant présent dans le cristal. Plusieurs cristaux de la même forme cristalline ont été testés et tous présentent une limite de diffraction à environ 5Å. A cette résolution la reconstruction n'est pas envisageable. Dès lors nous avons démarré de nouveaux essais de cristallisation, cette fois en présence d'un substrat : l'AMP-PNP, l'analogue non hydrolysable de l'ATP. Une nouvelle forme cristalline a été obtenue, en présence du ligand la protéine cristallise sous forme de baguette de dimension 30x330µm au bout de 48 heures à 22°C. Les gouttes de 2µl contiennent un volume équivalent de protéine, préalablement incubée avec 2mM d'AMP-PNP et 2mM de MgCl<sub>2</sub>, et de réservoir constitué de 0.2M NaF et de 20% PEG 3350. Les cristaux sont congelés en présence de 15% glycérol pour rendre la solution cryoprotectante.

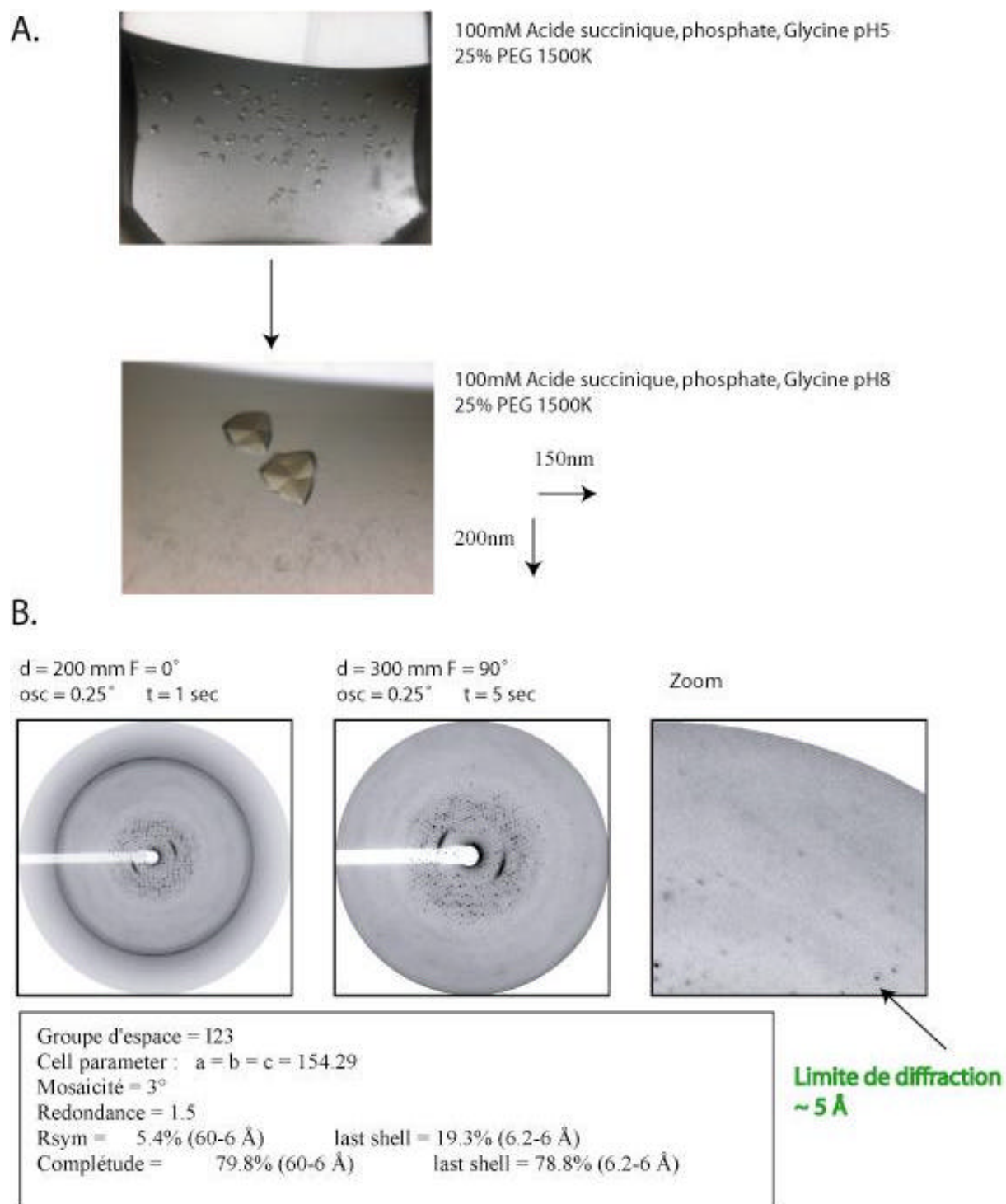


Figure 46 : Cristallisation de l'hélicase putative XPDss sous sa forme native. (A) Cristallogénèse : les cristaux apparaissent au bout de 24heures à 22°C. Les dimensions du cristal sont 150x200nm. (B) Enregistrement des données au synchrotron SLS. La limite de diffraction du cristal est d'environ 5Å.

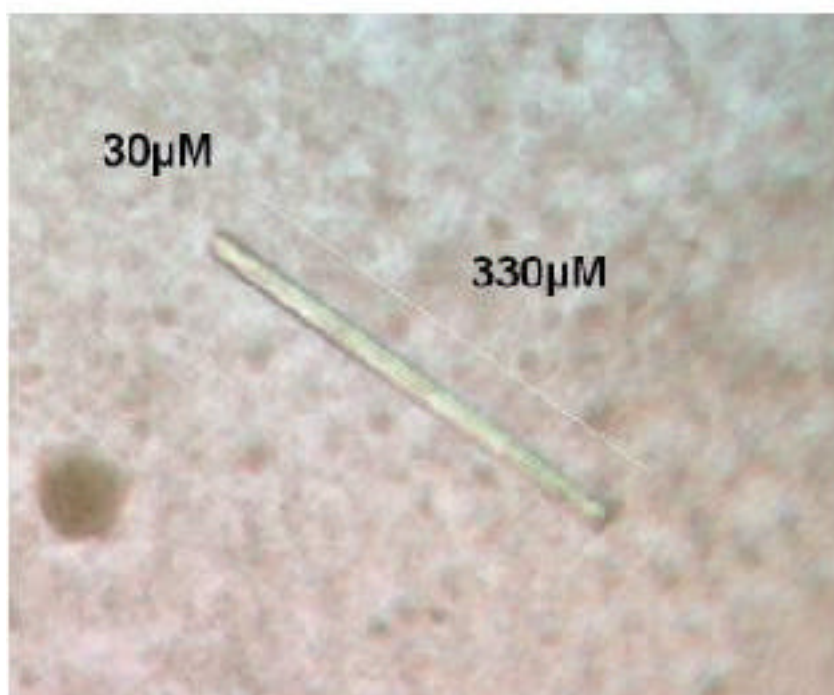
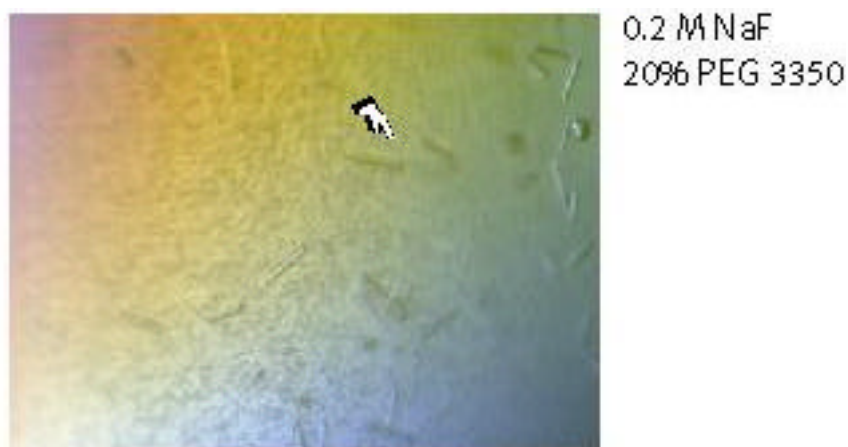


Figure 47 : Obtention d'une nouvelle forme cristalline. XPDs cristallise sous forme de baguette en présence de 2mM AMP-PNP.

Les cristaux de l'hélicase en présence de l'AMP-PNP ne diffractent pas sur une des lignes les plus puissantes au monde, l'Advanced Photon Source à Chicago. L'effet escompté dû à la présence d'un substrat n'a pas été atteint, l'ajout de l'AMP-PNP n'a pu ordonner le cristal et permettre la diffraction. La forme en fine baguette, indicatrice d'une seule direction de croissance corrobore l'idée de faible liaison entre les molécules.

### *Perspectives*

L'hélicase XPD de *Sulfolobus solfataricus* a été cristallisée et que ce soit sous sa forme native ou liée à un substrat elle diffracte mal ou pas du tout. Des essais de cristallisation de l'hélicase en présence d'ADN pourraient être envisageables en espérant qu'un changement conformationnel soit induit. Pour obtenir rapidement la phase des cristaux de dérivés lourds de l'hélicase de *Sulfolobus solfataricus* en présence d'ADN on pourrait substituer une base par une autre base modifiée (par exemple la thymine par un 5-iodourine).

Dernièrement au laboratoire, un nouvel étudiant, Wassim Abdulrahman a amorcé l'étude structurale des hélicases archaebactériennes analogues des hélicases XPD et XPB du complexe humain TFIIH. Plusieurs hélicases putatives ont été clonées et les tests d'expression sont en cours. Ce projet s'inscrit directement dans la perspective de caractériser et de cristalliser des hélicases putatives d'archaebactéries homologues aux protéines humaines. De plus la mise en place d'un test ATPase a permis de déterminer l'activité ATPase de l'hélicase XPDss en présence d'ADN. La fixation de l'hélicase sur un ADNss de 30mers suffit à déclencher l'hydrolyse de l'ATP.

## Chapitre IV : TFIIH EN COMPLEXE AVEC SES PARTENAIRES PHYSIOLOGIQUES

### Article 2: Immobilization of biotinylated DNA on 2-D streptavidine crystals

L'étude structurale des complexes de réparation/transcription contenant TFIIH, ainsi que l'étude de l'interaction de TFIIH avec l'ADN nous a posé d'énormes problèmes liés au fait que ces complexes sont transitoires. Le simple « mélange » des composants produit trop peu de complexes et il n'est possible ni de les purifier ni de les reconnaître directement en microscopie électronique. J'ai participé au développement d'une nouvelle technologie pour immobiliser l'ADN sur un support adéquat permettant l'observation ultérieure du complexe ADN-protéine en cryo-microscopie électronique. Nous nous sommes inspirés du système Dynabeads® qui permet l'immobilisation d'ADN biotinylés sur des billes couplées à la streptavidine. Ce système nous a permis de caractériser les conditions d'interaction de TFIIH avec des molécules d'ADN synthétiques. Ces expériences ont été transposées sur des cristaux 2-D de streptavidine formé au contact de lipides biotinylés. Nous avons montré par des expériences utilisant un marquage radioactif que des molécules d'ADN biotinylées étaient retenues spécifiquement par ces cristaux et pouvaient être transférées sur une grille de microscopie. Par ailleurs nous avons pu visualiser ces molécules d'ADN non colorées et hydratées en cryo-microscopie électronique.

Cette nouvelle approche méthodologique permet de concentrer l'ADN au contact des lipides et sélectionne les complexes ADN/protéines. Il sera possible non seulement d'étudier des échantillons peu abondants mais également d'accroître la probabilité d'observer des interactions faibles en **augmentant** localement la concentration d'un des partenaires. Cette méthode pourra être utilisée pour observer et déterminer la structure des complexes TFIIH/TFIIH lié à un ADN simple brin et TFIIH/XPC-HHR23B lié à un ADN endommagé.

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

Corinne Crucifix, **Muriel Uhring** and Patrick Schultz

**Immobilization of biotinylated DNA on 2-D streptavidin crystals**

Journal of Structural Biology, volume 146 issue 3, June 2004, Pages 441-451

Pages 441-451 :

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

Pour les utilisateurs ULP, il est possible de consulter cette publication sur le site de l'éditeur

<http://dx.doi.org/10.1016/j.jsb.2004.02.001>

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

*Perspective : Etude de TFIIH en liaison à l'ADN*

En premier lieu, j'ai cherché les conditions optimales pour l'observation de TFIIH en liaison avec un oligonucléotide synthétique. Pour ce faire nous avons utilisé le système Dynabeads .

Les Dynabeads M-280 sont des billes en polystyrène uniformes, superparamagnétiques, avec de la streptavidine liée à leur surface de manière covalente. La streptavidine est une protéine d'un poids moléculaire de 66.000 Dalton formée de 4 sous-unités identiques, chacune possédant une affinité pour la biotine. La constante d'affinité de la streptavidine pour la biotine est de :  $K_D = 10^{-15}M$ . Les Dynabeads (1mg) sont capables de lier 200pmoles d'un oligonucléotide simple brin biotinylé ou 20 pmoles d'un ADN double brin de 0.5 à 2 kb.

Dans un premier temps, les billes sont lavées trois fois avec le tampon : 10 mM Tris-HCl (pH 7.5), 1 mM EDTA, 2M NaCl. Les billes sont ensuite resuspendues dans 1 volume de tampon de lavage : 10 mM Tris-HCl (pH 7.5), 1 mM EDTA, 1M NaCl. Les billes peuvent dès lors être mises en présence de l'oligonucléotide biotinylé, dans notre cas il s'agit d'un poly(dT) de 30mers. L'incubation a lieu pendant 15 minutes à température ambiante. Une étape de lavage permet d'éliminer l'oligonucléotide en excès non lié aux billes. Les billes sont resuspendues dans le tampon adéquat. A ce stade, les billes sont prêtes à être incubées avec les protéines pendant 2 heures à 4°C. Après incubation, un dernier passage avec le tampon de lavage permet d'éliminer les protéines en excès et/ou les protéines non liées à l'ADN synthétique.

Nous avons testé les conditions de liaison du facteur TFIIH endogène avec un oligonucléotide poly(dT) de 30 mers. Après préparation des billes, la fraction contenant TFIIH (environ 300µg/ml) est incubée avec les billes préalablement liées à l'oligonucléotide. Différents tampons de liaison sont testés, la quantité de sels étant essentielle pour la liaison d'une protéine à l'ADN, nous avons testé plusieurs tampons et présentons les résultats obtenus avec : tampon (1) 20 mM Tris-HCl (pH 7.5), 150 mM NaCl et tampon (2) 20 mM Tris-HCl (pH 7.5), 50 mM NaCl . Les billes sont lavées et le bleu de dépôt est ajouté, les billes sont mises à bouillir et le tout est déposé sur un gel SDS 12.5%. Les protéines sont transférées sur une membrane afin d'identifier les protéines par des anticorps monoclonaux (Western Blot). Les

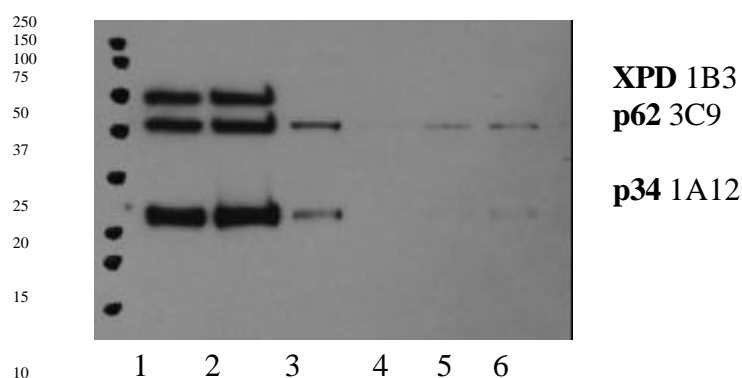
anticorps utilisés sont les anticorps 1B3, 3C9 et 1A12 respectivement dirigés contre les protéines XPD, p62 et p34.

### 1. Efficacité de liaison de TFIIH sur l'ADN avec le tampon (1) : 150 mM NaCl

	1	2	3	4	5	6
TFIIH ng	300	600	600	180	300	600
$\mu$ l	1	2	2	0.6	1	2

Oligo 30mers pmoles	-	-	-	200	200	200
	Témoin Dépôt	Témoin Dépôt	Témoin aspécificité			

#### Western Blot SDS 12.5%



Dans ces conditions, il y a de l'aspécificité de TFIIH pour les billes (non couplées à l'ADN), en comparant la piste 2 et 3, on constate qu'environ 1/3 des molécules de TFIIH se fixent de manière non spécifiques sur les billes. L'utilisation d'un détergent (Tween, Triton) pourrait diminuer cette aspécificité.

La quantité de TFIIH liée sur l'ADN (et/ou billes) est de l'ordre du bruit de fond, il n'y a pas d'interaction visible dans ces conditions. Les conditions de liaison 20 mM Tris-HCl (pH 7.5), 150 mM NaCl ne sont adaptées pour étudier l'interaction TFIIH-ADN.

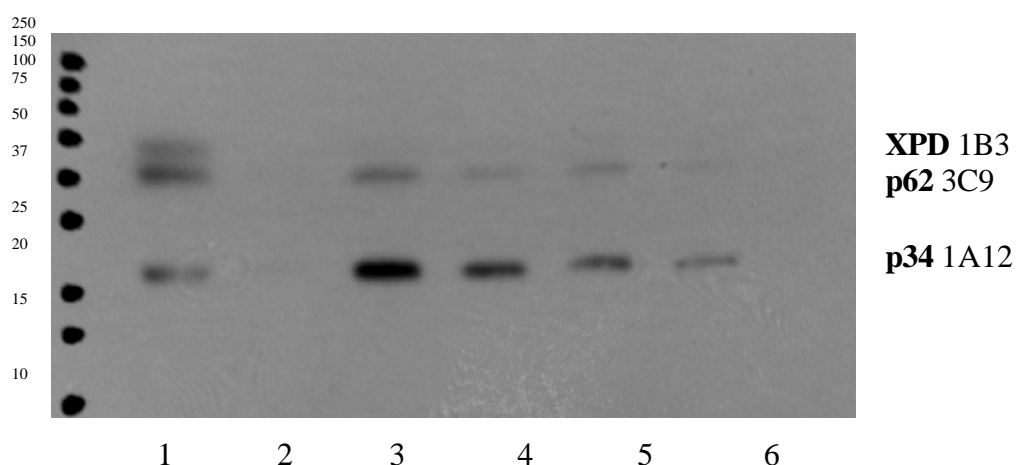


## 2. Efficacité de liaison de TFIIH sur l'ADN avec le tampon (1) : 50 mM NaCl

	1	2	3	4	5	6
TFIIH ng	600	600	600	600	600	600
$\mu$ l	2	2	2	2	2	2

Oligo 30mers pmoles		-	200	200	20	2
	Témoin Dépot	Témoin aspécificité				

WB SDS 12.5%

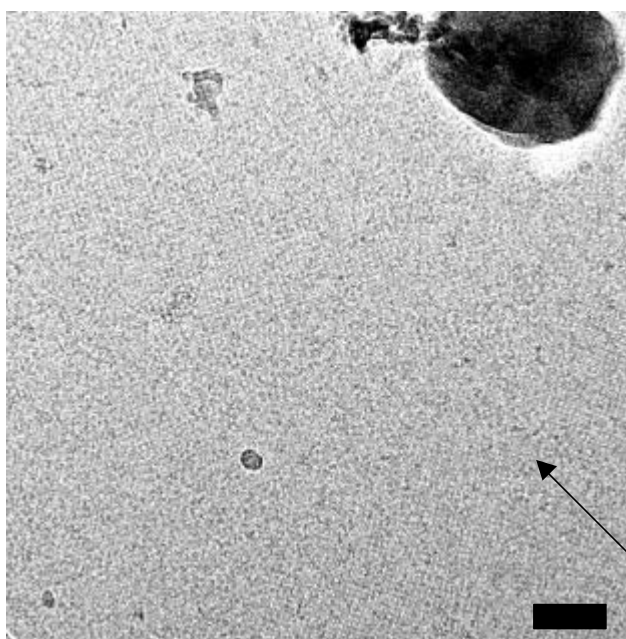


Dans ces conditions, il n'y a pas de liaison aspécifique de molécules de TFIIH sur les billes non couplées à l'oligonucléotide (piste 2). De plus nous pouvons observer une liaison de toutes les molécules de TFIIH sur les billes couplées au poly(dT) de 30mers à une concentration de 200pmoles, à cette concentration toutes les billes sont saturées. Cette liaison va en décroissant lorsque la quantité d'oligonucléotide diminue (piste 4 à 6).

Le tampon adéquat à la liaison de TFIIH sur de l'ADN avec cette méthode est le tampon (2) : 20 mM Tris-HCl (pH 7.5), 50 mM NaCl.

### 3. Observation du complexe TFIIH-ADN en cryomicroscopie

Nous observons maintenant TFIIH en liaison avec ce même oligonucléotide, dans les conditions de liaison établies, par la méthode d'immobilisation d'ADN sur des cristaux de streptavidine. De premiers résultats obtenus en cryomicroscopie sont décrits.



Les conditions de formation des complexes ADN-protéines sont les suivantes :

1  $\mu$ l TFIIH endogène  
 100 fmoles de polydT 30 mers  
 tampon : 20 mM Tris-HCl (pH 7.5), 50 mM NaCl

TFIIH « déroulé »

Figure 47 : Observation de molécules de TFIIH liées à l'ADN par la méthode d'immobilisation d'ADN sur des cristaux de streptavidine. Barre d'échelle = 100nm

Dans ces conditions, toutes les molécules d'ADN-protéines devraient se fixer aux cristaux de streptavidine préformés (se référer à l'article).

Dans la figure 47 nous pouvons visualiser des molécules liées sur les cristaux de streptavidine. Ces molécules apparaissent très allongées et semblent correspondre à des TFIIHs déroulés. La taille de ces particules est d'environ 16nm de long sur 5nm de large. Ces résultats préliminaires valident la méthode d'observation de TFIIH lié à de l'ADN et nous permet d'espérer dans un proche avenir d'élucider le phénomène d'ouverture lié à ce complexe. La signification physique de cette ouverture n'est pas encore connue mais on peut supputer un rôle des hélicases XPB et XPD dans ce phénomène lors de l'ouverture de l'ADN. Il serait intéressant d'observer les

changements conformationnels du complexe liés à son déplacement sur un ADN double brin. Mais conscients que ce mécanisme nécessite la coordination de plusieurs facteurs, il serait envisageable d'étudier le complexe TFIIH dans le contexte de la réparation, TFIIH est rapidement recruté sur le site après reconnaissance de la lésion par le complexe HHR23B/XPC et il est maintenant possible d'observer à l'aide de cette technologie un complexe ADNlésé/XPC-HHR23B/TFIIH puisque tous les composants sont disponibles.

### **Article 3: Molecular organization and oligomeric state of TFIIH**

TFIIH est recruté au sein du complexe de pré-initiation, via TFIIIE, ce qui permet l'ouverture de l'ADN au niveau du promoteur. Les facteurs TFIIIE et TFIIH interagissent physiquement et cette interaction stimule l'activité CTD kinase de TFIIH.

Nous avons déterminé un modèle de TFIIIE dont la structure n'a encore jamais été décrite. L'analyse d'images de molécules individuelles de TFIIIE, préservées en cryo-coloration négative, a été réalisée à une résolution de 16Å. TFIIIE est un hétérodimère contenant une seule copie de chaque sous-unité (TFIIIE $\alpha$  et TFIIIE $\beta$ ) A cette résolution, des domaines structuraux sont clairement discernables.

Disposant de structures expérimentales des facteurs TFIIIE et TFIIH, la poursuite de ce travail consistera naturellement à étudier l'interaction TFIIIE/TFIIH dans le but de déterminer la structure du complexe.

# Structure and oligomeric state of human transcription factor TFIIE

Uhring M.\*, Jawhari A.\*

\*These authors contributed equally to this work.

Keywords: electron microscopy, TFIIE, image analysis, immunolabelling, 3-D model, oligomeric state, co-immunopurification, crosslinking and ultra-centrifugation.

Department of Structural Biology and Genomics, Institut de Génétique et de Biologie Moléculaire et Cellulaire, CNRS/INSERM/ULP, 1, rue Laurent Fries, BP10142, F-67404 Illkirch, France.

## Abstract

The general transcription factor TFIIE plays essential roles in both transcription initiation and the transition from initiation to elongation. Human TFIIE is composed of 56 kDa (TFIIE- $\alpha$ ) and 34-kDa (TFIIE- $\beta$ ) subunits TFIIE and is described as an heterotetramer containing two copies of each  $\alpha$  and  $\beta$  subunits. Here, we report biophysical and structural data indicating that the human TFIIE is merely organized as an  $\alpha\beta$  dimer. Despite an aberrant migration in gel filtration, co-immunopurification and crosslinking assays as well as analytical centrifugation experiments indicate that purified recombinant TFIIE behaves as a dimer in solution. When observed by electron microscopy, most of the molecules appeared as dimers whereas less than 15% of the particles were larger and corresponded to  $(\alpha\beta)_2$  tetramers. A 3-D model of the tetramers, preserved in negative stain was reconstructed from isolated molecules and a higher resolution model of the  $\alpha\beta$  dimer was obtained by cryo-negative staining, a method that preserves the hydration of the specimen at low temperature. At a resolution of 1.7nm, TFIIE is composed of three distinct modules organized in an almost closed ring-like structure. A GST tag fused to the N-terminus of the  $\beta$  subunit was detected after image analysis and mapped the positions of the two subunits.

## Introduction

Transcription of protein-coding genes in eucaryotes is a multienzymatic process requiring RNA polymerase II (RNA Pol II) in addition to a complement of a minimum of five general transcription factors (GTFs), TFIIB, -IID (TBP), -IIF, -IIE and -IIH, which form the preinitiation complex (PIC), near the transcription start site. The GTFs are necessary to assemble the preinitiation complex (PIC) at the promoter and initiate transcription (reviewed of protein coding genes (Orphanides et al., 1996; Roeder, 1996). Order-of-addition experiments demonstrated that TFIIE enters the PIC after RNA Pol II and prior to TFIIH (Buratowski et al., 1989; Flores et al., 1992). TFIIE interacts directly with the unphosphorylated form of RNA Pol II (IIA), with both subunits of TFIIF, and with TFIIH (Flores & al., 1989; Maxon & al., 1994). Functions attributed to TFIIE include recruitment of TFIIH to the PIC, stimulation of TFIIH-dependent phosphorylation of the RNA Pol II CTD, and stimulation of TFIIH-dependent ATP hydrolysis (Lu et al., 1992). Furthermore, two-dimensional crystallography of a TFIIE-RNA Pol II complex suggested that TFIIE promotes a conformational switch at the active centre upon RNA Pol II-DNA interaction (Leuther et al., 1996).

Human TFIIE was initially purified to homogeneity from HeLa cell nuclear extracts (Inostroza et al., 1991; Ohkuma et al., 1990) is composed of 56-kDa (TFIIE $\alpha$ ) and 34-kDa (TFIIE $\beta$ ) subunits present in equimolar amounts. Both subunits are highly charged, with pI values of 4.5 and 9.5 for TFIIE $\alpha$  and  $\beta$ , respectively (Peterson, nature, 1991; Ohkuma, nature). Size exclusion chromatographic studies indicate a molecular mass of approximately 180 kDa, suggesting that TFIIE is a heterotetramer containing 2  $\alpha$  and 2  $\beta$  subunits (Ohkuma et al., 1990). Systematic deletion mutations of TFIIE subunits were used to characterize the function of the TFIIE $\alpha$  and  $\beta$  conserved regions. In fact, human TFIIE $\alpha$  possesses several putative structural motifs (Ohkuma & al., nature, 1991). The N-terminal part of the protein, which contains all the evolutionarily conserved structural motifs, was shown to be essential for basal transcription and phosphorylation as well as TFIIE $\beta$  interaction, whereas the C-terminal part is dispensable even if he can bind specifically to TFIIH factor (Ohkuma & al., MCB, 1995). Concerning the human TFIIE $\beta$ , it also possesses several conserved structural motifs (Sumimoto & al., Nature, 1991; Ohkuma & al., NAR, 1992). In fact, three functional regions were mapped, namely, a central core (from 66 to 146), a basic helix-loop-helix (from

197 to 237), and a C-terminal basic helix loop (from 258 to 291) (Okamoto & al., JBC, 1998; Okuda & al., EMBO, 2000; Sumimoto & al., Nature, 1991). The C-terminal basic region was reported to be essential for physical interaction with TFIIE $\alpha$  and single strand (ss) DNA. The central core-region was found to bind double strand (ds) DNA (Okuda & al., EMBO, 2000).

Interestingly, in the context of transcription initiation, TFIIE was shown to contact the Rpb9 subunit of RNA polymerase II (van Mullem & al., JBC, 2002), and Tfb1 the core subunit of transcription factor TFIIF (Bushnell & al., JBC, 1996). Protein-DNA crosslinking experiments revealed that TFIIE contacts ds DNA between positions -14 and -2 upstream from the transcription start site (+1) including the region around -10 to -8 where the promoter DNA starts to be melted (Roberts & al., 1996). The  $\beta$  subunit of TFIIE contacts the promoter DNA (Roberts, JBC, 1996) and the DNA binding domain was mapped to residues 66-146 (Okuda & al., EMBO, 2000). Interestingly, TFIIE $\beta$  by itself interacts only with ss DNA through its C-terminal part (Okamoto et al. JBC 1998), whereas the presence of TFIIE $\alpha$  which does not bind to DNA is required for TFIIE $\beta$  to bind to ds DNA (Roberts & al., JBC, 1996). The interaction between the two subunits of TFIIE occur through the C-terminal part of TFIIE $\beta$  and the N-terminal part of TFIIE $\alpha$  (Okamoto & al., JBC, 1998; Ohkuma, MCB, 1995). Otherwise, TFIIE is believed to interact with additional partners including transcription activators Sp1 and TFIIA (Peterson & al., Nature 1991; Langelier & al., JBC, 2001) as well as Kruppel gene-specific transcriptional activator (Sauer et al., 1995). Furthermore, TFIIE $\beta$  was shown to contact general chromatin factor Spt16 (Kang & al., Genes to Cells, 2000).

Although no structural information is available on the entire human TFIIE, atomic structures of two conserved modules have been determined by X-ray crystallography or NMR. The first one corresponds to residues 1-97 of TFE, the archaee homologue of the TFIIE, and consists of a winged helix-turn-helix module (three helices with a hairpin at the C-terminus) extended by specific alpha-helices at the N and C-termini (Meinhart & al., JBC, 2003). The second module, encompassing residues 66-146 of the central core domain from TFIIE, resembles to winged helix proteins and possesses a double-stranded DNA-binding activity (Okuda & al., EMBO, 2000).



In this study, we have employed electron microscopy coupled with solution studies to examine the molecular organization of the entire and functional TFIIE complex. A first model of a heterotetrameric  $2 \times 2$  state of the factor was determined in negative stain. In a second step the TFIIE complexes were observed by cryo negative staining in a hydrated state that better preserved their structure. At a resolution of about  $17\text{\AA}$  the independently performed 3D reconstruction represents an almost closed ring-like structure composed of three distinct modules that closely resembles the asymmetric unit of the negatively stained  $2 \times 2$  model. We also report biochemical and biophysical data indicating that the human TFIIE is merely organized as a dimer. These observations provide the first structural data of the human TFIIE complex.

## RESULTS

### Purification of TFIIE

For the structural investigation of human TFIIE, cDNAs encoding its  $\alpha$  and  $\beta$  subunits were cloned into the pet28b and pACYC expression vectors, respectively. These vectors have compatible replication origins and allow the co-expression of the two subunits in *E.coli*, with a typical yield of 1 mg of pure human TFIIE per liter of culture. As indicated in Figure 1A, TFIIE was purified first by metal affinity and then by size exclusion chromatography. Since TFIIE $\alpha$  was expressed in fusion with an hexa-histidine tag at its N-terminus, we used metal affinity chromatography to retain the complex. After extensive washing the complex was eluted by application of an imidazole gradient. At this stage, and despite the fact that only the  $\alpha$  subunit was histidine tagged, an excess of the  $\beta$  subunit retained on the metal affinity column was observed. To remove the excess of the  $\beta$  subunit, the protein containing fractions were further purified using a gel filtration column. Two protein peaks were obtained. One of them corresponded to the excess of the  $\beta$  subunit (data not shown) while the second contained the  $\alpha$  and  $\beta$  TFIIE subunits (Figure 1B).

As a prerequisite for electron microscopy analysis, we further characterized TFIIE preparation in terms of homogeneity. First, the purified human TFIIE complex was analysed by native PAGE. As illustrated in Figure 1B, a single band, indicative of a homogeneous sample was observed. Next, we submitted recombinant TFIIE to an analytical size exclusion chromatography. A single peak containing the  $\alpha$  and  $\beta$  subunits was observed with a retention time corresponding to the molecular mass of a globular protein of approximately 200 kDa (Figure 1C and 1D). In agreement with the previously published data (Ohkuma & al., 1990), this value is close to the molecular weight calculated for a heterotetrameric TFIIE molecule with 2  $\alpha$  and 2  $\beta$  subunits.

### **The 3-D model of negatively stained human TFIIE**

When the purified TFIIE molecules were adsorbed on a carbon film and were negatively stained, the distribution in size of the particles was heterogeneous indicating that the preparation method induced the formation of different forms of particles (Figure 2A). The largest proportion (about 75%) of particles showed a globular shape about 11 by 6 nm in size but did not show any internal symmetry as would be expected for an  $\alpha_2\beta_2$  assembly. Conversely, about 15% of the particles were larger (12 by 11 nm) and showed an internal symmetry in some orientations suggesting that a minor proportion of the observed particles had an  $\alpha_2\beta_2$  stoichiometry. About 10% of the particles appeared significantly smaller and corresponded most probably to disrupted particles that were not further analyzed.

In order to analyze an homogeneous image data set for three-dimensional reconstruction, a total of 4480 large particles were sorted out. The images were centered and partitioned into classes prior rotational alignment to identify the most representative molecular views to be used as alignment references, without introducing a bias due to the arbitrary selection of alignment references (data not shown). After 4 alignment/classification cycles a stable partition was obtained that revealed characteristic views of the large particles (Fig.2B). Some molecular views clearly showed a two-fold symmetry axis perpendicular to the image or in plane to the image (Figure 2B, 4<sup>th</sup> panel and 5<sup>th</sup> panel, respectively), suggesting that a symmetric assembly consistent with an  $\alpha_2\beta_2$  structure is present on the electron microscopy support. A 3-D model of this heterotetrameric particle was reconstructed and was refined by removing the possibly contaminating dimeric ( $\alpha_2$ ) particles. To this purpose, reprojections of the heterotetramer and of the asymmetric unit were used as alignment references and the contaminating dimer views (about 30% of the data set) were sorted out on a correlation coefficient based criteria. A total of 108 different views of the large particles were used to reconstruct a 3-D model of the  $\alpha_2\beta_2$  assembly and the retroprojection of the model (Figure 2C) were found to be very similar to the class averages (Figure 2B). The two fold symmetry was imposed at a later stage of the refinement.

The 3-D model of the heterotetramer shows that the asymmetric units are elongated along the symmetry axis (11 nm) and about 11 by 12 nm in size when viewed down this axis (Figure 2D). The density threshold was set to represent a volume occupied by a protein of 165

kDa corresponding to the molecular mass of two  $\alpha$  subunits (Mr 49.5) and two  $\beta$  subunits (Mr 33 kDa) as predicted by their amino acid sequences. The asymmetric unit, composed of one  $\alpha$  and one  $\beta$  subunit, reveals at least three distinct modules arranged in an almost closed ring-like structure (Figure 2E). The largest module I appears flat and represents 52.4% of the volume, it makes a thin contact with the other TFIIIE molecule to form the heterotetramer, the central module II represents 16.7% and the distal module III 30.9% of the volume.

### **Refined 3-D model of cryo-negatively stained human TFIIIE**

In order to improve the resolution of the 3-D model, the human TFIIIE molecules were observed at low temperature in a hydrated form and included in a 16% solution of ammonium molybdate in order to increase the contrast of the particles (Figure 3A). As previously observed (Adrian et al., 1998, 2002; De Carlo et al., 2002, Tsitrin et al., 2002), the images show an enhanced contrast as compared to unstained frozen-hydrated samples, particularly when the micrographs are recorded close to focus. In these conditions, little if any large structures were observed indicating that in hydrated conditions the  $\alpha_2\beta_2$  heterotetramers are not present.

Individual particles were clearly distinguished and showed considerable amount of structural details. A total of 12084 molecular images were analysed as described above and a stable 3-D model, included 500 different molecular views, was obtained by angular reconstruction (van Heel et al., 1996), independently of the model derived from the analysis of negatively stained data set (see above). The angular distribution of the data set was homogeneous indicating that the specimen was not preferentially adsorbed (Figure 3B) and the resolution test performed on two independent data sets gave a value of 1.7 nm for the 0.5 FSC criterion and of 1.4 nm for the 3  $\sigma$  criterion (Figure 3C). The density threshold was set to represent a volume occupied by a protein of 82 kDa. The 3-D model of the cryo-negatively stained TFIIIE shows an almost closed circle composed of domains of protein densities (Figure 3D). The superposition of this model with that derived from the  $\alpha_2\beta_2$  data set shows a good fit indicating that the two structures obtained independently are very similar (Figure 3E). The envelop shows much more structural details than the negatively stained model, particularly in module I which represents 59.4% of the volume and can be subdivided into two subdomains

of similar size. The central module II appears slightly smaller (26.1% of the volume) than in negative stained conditions, whereas the distal module III appears larger (14.5% of the volume) and the contacts with module II are not revealed. These differences in size and interdomain contacts probably reflect staining variations.

### **Oligomeric state of TFIIE in solution**

Since the behaviour of TFIIE in gel filtration corresponds to a heterotetramer complex and electron microscopy study show clearly the existence of major amount dimer species, we wondered if TFIIE can exist as dimer or tetramer or both in solution. Then, we investigated its oligomeric state in solution using assays under different and representative sets of conditions. Chemical cross linking experiments were performed to obtain an initial estimate of the degree of oligomerisation that these molecules experience. Purified TFIIE was subjected to the action of glutaraldehyde chemical agent at different concentrations: typically, for each reaction, 10 microliter of TFIIE at 1mg/ml were mixed with 5 microliter of glutaraldehyde at 0, 0.02%, 0.04%, 0.06%, 0.08% or 0.1%. After 30 minutes at room temperature, the reactions were analysed by SDS polyacrylamide gel electrophoresis (Figure 4A). Without addition of glutaraldehyde, bands corresponding to the  $\alpha$  and  $\beta$  subunits of TFIIE, which migrate at 55 and 33 kDa, respectively (See lane 1, Figure 4A) are observed. Reaction with glutaraldehyde resulted in the apparition of a new band with a molecular weight about 90 kDa, according to the molecular weight markers (compare lane 1 with lane 2, Figure 4A). This band should correspond to a crosslinked TFIIE complex. An increase of glutaraldehyde concentration induced a total disappearance of TFIIE  $\alpha$  and  $\beta$  bands, and the apparition of a diffuse TFIIE crosslinked band (compare lane 2 with lane 3, Figure 2). No band at 180 kDa was observed, even at the maximum of glutaraldehyde concentration. If oligomers corresponding to the TFIIE tetramer are formed, they must be present at low concentration, or inefficiently cross linked by glutaraldehyde.

Pull-down interaction assays were carried out to further test if an  $\alpha\beta$  dimer is able to self-associate and forms an  $\alpha_2\beta_2$  heterotetramer. FLAG- and (His)<sub>6</sub>-tagged TFIIE as well as native TFIIE were expressed independently in *E. coli*, mixed and incubated for 3 hours to allow formation of the complexes. The sample was divided. One part was incubated with protein-A sepharose beads cross-linked with anti-FLAG antibody, and after washing with a

buffer containing 250 mM NaCl 0.1% NP40, and elution by Flag peptide competition, proteins were analysed by SDS-PAGE and western blotting using a set of monoclonal antibodies directed against the FLAG-tag, the hexahistidine-tag or TFIIE. If TFIIE forms an  $2 \times 2$  heterotetramer, (His)<sub>6</sub>-tagged TFIIE should co-precipitate with FLAG-TFIIE and TFIIE. On the contrary, TFIIE only forms heterodimers, non histidine-tagged TFIIE is expected to co-precipitate with FLAG-TFIIE and TFIIE. As indicated in Figure 4B (compare lanes 3 and 5), only FLAG-TFIIE and TFIIE were retained on the beads, demonstrating that in solution TFIIE only forms stable heterodimers but not heterotetramers.

The second aliquot was incubated with cobalt affinity resin, and after washing the eluate was subjected to SDS-PAGE with immunoblot detection. As expected, histidine tagged but not FLAG tagged TFIIE was retained by the resin (Figure 4B, compare lanes 3 and 7). Control experiments were extracts from cells expressing either (His)<sub>6</sub>- or FLAG-tagged TFIIE were combined with one from cells containing TFIIE show that none of the protein is a-specifically retained on the beads (Figure 4B, compare lanes 1 to 4 and 2 to 6). Together, these results show that histidine or FLAG-tagged TFIIE forms stable heterodimers with TFIIE, but not heterotetramers.

Finally, we used analytical ultracentrifugation to precisely investigate the oligomeric state of the TFIIE complex. A TFIIE sample concentrated at 1.5 mg/ml in a buffer containing 20 mM Tris HCl-pH 7.5, 250 mM NaCl, 2 mM  $\beta$ -mercaptoethanol was brought to sedimentation equilibrium experiment at 20°C. The solid curve through the data is the global least square fit of the expression for a pure dimer (Figure 5). The uniform distribution of residuals around zero indicates the compatibility of the model with the data. Several other models were tested, including monomers, tetramer, dimer-monomers, dimer-tetramers... They all gave larger residuals than the dimer model and systematic deviation with radius (not shown). The fit performed using a single species returned a molecular weight of 84589 +/- 214 Dalton. This value is almost equal to 84686 Dalton, the calculated molecular weight of a recombinant dimer.

## Mapping of the N-terminus of TFIIE $\alpha$

In order to map the positions of the  $\alpha$  and  $\beta$  subunits within the 3-D model, recombinant TFIIE was produced, in which a 26 kDa GST tag was introduced at the N-terminus of the  $\alpha$  subunit. The GST-tagged TFIIE  $\alpha$  and the TFIIE  $\beta$  subunits were co-expressed in *E. coli* and the complex was purified first by GST affinity and then by size exclusion chromatography. The presence of the GST tag at the amino terminus of the  $\alpha$  subunit does not prevent the formation of the complex as indicated by the purification and also the structure (Figure 6). The mapping of the additional mass due to the GST fusion will allow us to position the interface between the two subunits. The complexes were sandwiched in 2% uranyl acetate between two layers of thin carbon and frozen in liquid nitrogen after 2 minutes at room temperature. A total of 8766 individual particles of negatively stained human TFIIE fused to GST were recorded and analyzed similarly to the two other data sets. In order to avoid any influence from previous models, this analysis was performed independently and in particular no common references were used for alignment or angular assignment. A 3D model was calculated from 187 molecular views and showed a three domain organization when the density threshold is set for a volume occupied by a mass of 108 kDa (Figure 6A). When compared to a similarly oriented model of the negatively stained dimer the modules I and III were similar although the linker between module II and III is not revealed. The central module II is however larger in the GST-tagged TFIIE than in the native complex. The proposed orientation between the two models is the best fit for the relative positions of the three modules. Once pre-oriented manually, the GST-TFIIE model was masked to removed the additional GST volume which perturbs a correlation-based 3-D alignment of the two models. The alignment of the masked TFIIE-GST model was then refined by maximizing the three-dimensional correlation coefficient. A 3-D difference map between the aligned GST-TFIIE model and the native TFIIE model showed that an additional mass (the threshold was set to represents a volume occupied by a mass of 22 kDa) is connected to module II indicating that this module contains the interface between the two subunits. The respective size of the different modules gives some clues about the position of the two subunits. According to the mass of the polypeptides, subunit  $\alpha$  (60% of the mass of TFIIE) is likely to be placed predominantly in module I with its N-termini located in domain II whereas subunit  $\beta$  resides most probably in modules II and III. Proteolytic cleavage of the  $\beta$  subunit revealed two domains resistant to trypsin digestion (Ohkuma & al., 1995). The N-terminal region of

TFIIE was shown to interact with the C-terminal part of TFIIE- (residues 193-240) (Ohkuma & al., 1995) which is thus probably located in module II. The N-terminal domain between residues 74-152 resides most likely in module III.



## DISCUSSION

Transcription factor IIE is an evolutionary conserved factor required for both accurate initiation of transcription by RNA polymerase II and promoter escape (reference, review Hampsey 1998., ..) that has been widely analysed by biochemical means. Further insights into the roles of TFIIE may come from structural studies of the entire complex. To address this issue we have examined its molecular architecture and oligomeric state using a combination of electron microscopy and solution studies. Our studies were performed using human TFIIE produced in *E. coli* by co-expression of the  $\alpha$  and  $\beta$  subunits. The ability of recombinant TFIIE to stimulate transcription in an *in vitro* assay is widely documented (refs). Bacterially expressed TFIIE is used in most functional studies and exhibits biochemical properties identical to those reported for endogenous TFIIE.

We have determined 3 D models of human TFIIE obtained by electron microscopy analysis of negatively and cryo-negatively stained particles. Analysis of negatively stained particles revealed the presence of two groups of particles. The largest proportion (about 75%) exhibited approximate dimensions that do not correspond to a 180 kDa particle and were disregarded. A minor proportion of the particles (about 15%) were larger, some of their molecular views exhibited an internal two-fold symmetry axis and were interpreted as heterotetramers. Given that TFIIE was commonly described as an  $\alpha_2\beta_2$  heterotetramer, the observation of a large population of small particles was initially interpreted as an artefact of the preparation method that induces the dissociation of the heterotetrameric particles into dimers. A 3D envelope was reconstituted at a resolution of 3.0 nm from the particles interpreted as heterotetramers. The envelope is composed of two  $\alpha/\beta$  modules connected by a thin linker that are related by the two-fold symmetry axis. The head to head arrangement of paired dimers (dimer of dimer) particle exhibits limited contacts with approximately xx % of the protein surface buried in subunit contacts. A limited interface is usually observed in complexes where the components associate after they have independently folded and is not typical of oligomers that are permanent assemblies.

When the same sample was analyzed in a frozen-hydrated state by cryo-negative staining, an homogenous distribution of particles was observed and structural model was determined at a resolution of 1.7 nm. Comparison with the envelope of the tetrameric form of

TFIIE indicates that the particle corresponds to an  $\alpha$  /  $\beta$  dimer. The particle is elongated, with approximate dimensions of about 6 by 11 nm in size. It is composed of an almost linear arrangement of distinct modules connected by thin linkers. Three densities corresponding to 60%, 26% and 14% (modules I, II and III) of the total volume of the particle can be distinguished. By comparing native TFIIE with a GST tagged complex, we were able to locate the amino-terminus of the  $\alpha$  subunit in the vicinity of module II (see results). At this resolution, it is however difficult to recognize the structural domains whose 3D structures have been determined at atomic resolution and to identify the limits  $\alpha$  and  $\beta$  subunits.

The previous data, and in particular the observation of an exclusively dimeric TFIIE particles negative staining prompted us to perform additional experiments aimed at investigating hydrodynamic properties and quaternary structure of the complex. TFIIE was commonly described as a heterotetramer composed of two  $\alpha$  and two  $\beta$  subunits. However, this description is based on gel filtration experiments that indicate an apparent molecular weight close to 200 kDa, a mass that is slightly larger than that expected for a heterotetramer. Size exclusion chromatography is suitable for the estimation of the molecular mass of proteins in the case of rigid globular proteins. However, examples describing apparent anomalies in the elution profiles have been reported (ref Junghans JBC 1996, Garnier & al.,2002; Biochemistry). These experiments included co-immunoprecipitation with two different affinity tags, chemical crosslinking experiments as well as equilibrium analytical ultracentrifugation and all demonstrated the existence of an exclusive dimeric form of human TFIIE in solution. These data suggest that the observation by electron microscopy of a large majority of small particles, interpreted as  $\alpha$  /  $\beta$  heterodimers, do not result from a dissociation of heterotetrameric particles. Rather, this observation reflects the native structure of soluble TFIIE and the small proportion of heterotetramers observed in the dried negatively stained sample self association of TFIIE  $\alpha$  /  $\beta$  dimers on the electron microscope grid. The molecular parameters influencing apparent molecular weight on gel filtration are size, shape and charge of the molecule (Hollecker, 1997). The elongated shape of the  $\alpha$  /  $\beta$  dimer can contribute to its behaviour during gel filtration chromatography. Aberrant mobility behaviour have been detected and characterized in the case of the DNA repair protein XPA. It was demonstrated that both the disorder and the Glu-rich region of *Xenopus leavis* XPA were primarily for the discrepancy between its calculated and its apparent molecular weight estimated from SDS-PAGE and gel filtration chromatography. Similar arguments could

explain an aberrant migration of the TFIIE heterodimer: (i) Both TFIIE subunits are highly charged; the  $\alpha$  subunit possesses highly charged Glu- and Asp- rich regions. (ii) Sequence analysis and limited proteolysis (not shown) suggest that the C-terminal of the large  $\alpha$  subunit and the N-terminus of the small  $\beta$  subunit are disordered and/or flexible. Such self association processes Self association processes

The heterodimeric organization of Human TFIIE is consistent with structural evidence for a single Tfa1/Tfa2 heterodimer with yeast RNA Polymerase II (Leuther 1996). Tfa1 and Tfa2 are the respective yeast homologues of human TFIIE  $\alpha$  and  $\beta$ . They exhibit significant sequence homology with their human counterpart (52.2% and 53% similarity for the large and small subunits, respectively) (Feaver & al., JBC, 1994) as well as similar physical properties. Both yeast TFIIE subunits are highly charged and the complex elutes from a superose-12 gel filtration column in 0.5 M KCl with a stockes radius corresponding to a globular protein of 200 kDa. In a glycerol gradient, yeast TFIIE sediments at a rate corresponding to that of a protein that is much smaller than a 200 kDa complex (Sayer & al., 1992). These observations are similar to those presented for Human TFIIE and support an heterodimeric organisation of TFIIE complexes.

Reported crosslinking experiments of human TFIIE on DNA promoter reveal that TFIIE bind both the upstream of the TATA element and downstream of the transcription start site (Forget & al., MCB, 2004). This result might be explained by the extended distribution of TFIIE density and not result from the presence of double copies of TFIIE in the preinitiation complex as previously suggested. In the same way, TFIIF molecular structure solved by electron microscopy in complex with RNA polymerase II, shows clearly an extended distribution of the protein, which is unambiguously at the origin of the multiple promoter interacting sites reported by crosslinking experiments (Chung & al., Mol Cell, 2003). In addition, two-dimensional crystallography of yeast TFIIE-RNA pol II complex, suggest that TFIIE exist as a  $\alpha\beta$  dimer (Leuther et al., 1996). This result and our finding provide evidence for a single TFIIE dimer of 56 kDa and 35 kDa subunits.

Our work cannot exclude the possible existence of equilibrium between the TFIIE tetramer and dimer, but if tetramer oligomers are formed, they are either present at concentrations too low for detection by any of our methods. In that case, we can speculate that

the hetero-tetramer might exist in higher concentration in the cell and might correspond to a modulate self associated TFIIE which may play a role in transcription regulation as postulated in the case of TBP oligomerization (Adams & al., JBC , 2003). Further studies are needed to determine how the TFIIE dimer is able to contact DNA promoter and RNA polymerase II as well as TFIIF factor within the preinitiation complex machinery.

## **MATERIALS and METHODS**

### **Expression of human TFIIE complexes**

The DNA sequences encoding full length human TFIIE-a and TFIIE- were inserted in the *NdeI-BamHI* restriction sites of pET28b, pGEX-NB (a pGEX-4T2 expression vector with a polylinker identical to that of pET15b) and pACYC-11b (ref fribourg), respectively. TFIIE-a *full length* and TFIIEa *full length* were co-expressed in *E.coli* BL21 (DE3) strain.

Cells were grown in LB-Agar medium containing 30 µl/ml kanamycin and 50 µl/ml chloramphenicol at 37°C. Typically, 6 litres of LB or M9 medium were spreading by pre-culture preparation and incubated at 37°C. At 0.8 optic density, the temperature was shifted to room temperature during 2 hours before induction by addition of 0.4 mM isopropyl-D-thiogalactopyranoside overnight.

### **Purification of human TFIIE**

After being harvested by centrifugation (5000g for 20 minutes), cells were resuspended into 10 ml of buffer A (20 mM Tris HCl-pH 7.5, 250 mM, 2 mM β-mercaptoethanol) per liter of culture, disrupted by sonication and then centrifuged again at 50000g for 1 hour. The supernatant containing the histidine tagged TFIIE- subunit was then applied onto a metal affinity chromatography column (1 ml of affinity beads (Talon-Clontech) per liter of cells). After washing with Buffer A 0.1% Nonidet P40, the proteins were eluted by addition of buffer A containing 250 mM imidazol. The fractions of TFIIE were concentrated on Microcon-30 devices (Amicon-Grace). Then, the proteins were loaded on a Superdex S200-16/60 gel filtration (Pharmacia). Finally, the gel filtration fractions were re-concentrated on Microcon-30 devices (Amicon-Grace). Protein purity was assessed using overloaded SDS-PAGE 12.5% gels with coomassie blue staining. Protein integrity was characterized by native-acrylamide gel (gradient 8-25%) with coomassie blue staining.

### **Molecular weight determination.**

The molecular weight of the native protein was estimated by gel filtration with a Superdex S16/60 200pg column from Pharmacia at a flow rate of 1.0 ml/min in 20 mM Tris/HCl, pH 8.0, 250 mM NaCl, 2 mM  $\beta$ -mercaptoethanol. Thyroglobulin, ferritin, catalase, aldolase, albumin, and ovalbumin (from Biorad) were used as calibration standards.

### **Chemical cross-linking**

10  $\mu$ l of purified his-TFIIIEa/ $\beta$  (1 mg/ml) was incubated with 2  $\mu$ l of glutaraldehyde at various percentages (0, 0.02%, 0.04%, 0.06%, 0.08% and 0.1%) during 30 minutes at 30°C. The reactions were stopped by addition of 5  $\mu$ l of Leammli buffer and electrophoresed on a SDS 12% polyacrylamide gel.

### **Protein Pull-down Interaction Assay**

*E. coli* BL21(IDE3) cells were transformed with the plasmid p28b.TFIIIE- , pFLAG.TFIIIE- or with pACYC.TFIIIE- and grown with agitation at 37 °C. When cultures reached A600 = 0.6, temperature was shifted to 22 °C. Two hours after the temperature shift, protein expression was induced with IPTG. After incubation for 12 hours at 22 °C, cells were collected. Pellets were resuspended in 20 mM Tris/HCl pH 7.9, 250 mM NaCl, 1 mM DDT, lysed at 4 °C with lysozyme (0.1 mg/ml) followed by sonication (2 times 30s with a pause for 30s on ice between each burst). Cell extracts were clarified by centrifugation (14 000 rpm in eppendorf tubes at 4 °C for 30 min).

### **Ultracentrifugation**

The experimental buffer (gel filtration) was 20 mM Tris (pH 7.5), 250 mM NaCl and 2 mM  $\beta$ -mercapto-ethanol. Sedimentation equilibrium experiments were performed at a temperature of 20.1°C using a Beckman Optima XL-A analytical ultracentrifuge (Beckman Instruments Inc. Palo Alto, CA) with absorbance monitoring. Protein concentration was 1.5 mg/ml, corresponding to molar concentration of 20  $\mu$ M. In each experiment, three samples were run in a four-place AN60-TI rotor, using 12-mm double-sector cells with quartz windows. For velocity runs, the cells were filled with 400  $\mu$ l of protein solution at 1.5 mg/ml in one sector and 425  $\mu$ l of 20 mM Tris HCl (pH 7.5), 250 mM NaCl and 2 mM  $\beta$ -mercapto-ethanol in the other. For sedimentation equilibrium experiment, samples were spun at 12,000 rpm and systems were first allowed to equilibrate for 12 hours before absorbance profiles and were

compared at different times to ensure that system had reached equilibrium. Using non-linear least-squares analysis, this data set was fitted using single component model and several equilibrium models including monomer/dimer, monomer/dimer/tetramer and dimer/tetramer. The computer program XLAVEL (Beckman) was used to analyse the data and determine the sedimentation coefficient (s). The s determination was based on value estimated for a specific scan with the equation  $s = \ln(r_{\text{inf}}/r_0) / \omega^2 t$ , and calculates the best s value, reported in Svedberg (S,  $10^{-13} \text{ S}^{-1}$ ). The value of s was corrected to standard conditions of water at 20°C.

### **Negative staining**

Ten microliters of this preparation was placed on a 10nm thick carbon film previously treated by a glow discharge in air. After one minute of adsorption the grid was negatively stained with 2% (w/v) uranyl acetate solution. The images were formed on a Philips CM120 Transmission Electron Microscope operating at 120kV with a LaB6 filament. Areas covered with individual molecules were recorded under low dose condition (less than 20 electrons/Å<sup>2</sup>) at a nominal magnification of 45,000 times.

### **Cryo-electron microscopy**

For cryo-negative staining, the purified TFIIE was diluted to a concentration of 0.3 mg/ml in 20 mM Tris-HCl pH 7.4, 150 mM NaCl and a 5 µl droplet was applied on a holey carbon foil. The grid was then put for 20 seconds on a 16% ammonium molybdate solution, previously adjusted to pH 7.4 by the addition of NaOH 10M, before being mounted on the plunger, blotted and vitrified (Adrian et al., 1998). Images were recorded under low-dose condition on Kodak SO163 photographic plates on a FEI Tecnai 20 transmission electron microscope (field emission gun operated at 200 kV). Areas covered with individual molecules were recorded at a magnification of 43,900 x times, as calibrated with the 2.3 nm reflection of tobacco mosaic virus.

### **Image processing**

The original micrographs were checked by optical diffraction for absence of astigmatism and optimal contrast transfer function. The best micrographs were digitized at 13.1 µm raster size, resulting in a pixel spacing of 0.3 nm on the object using a linear CCD array (Super Coolscan ED 8000, from Nikon).

The image analysis was performed using the IMAGIC software package (van Heel et al., 1996). The images of cryo-negatively stained TFIIE molecules were normalized, band-pass filtered and centered onto a rotationally averaged summation of all images. The centered images were analyzed by multivariate statistical methods and hierarchical ascendant classification in order to identify characteristic molecular views (van Heel and Frank, 1981). These views were used as references for multireference rotational and translational alignment of the data set after which the aligned images were again subjected to variance analysis and clustered. Several cycles of the latter process yielded stable, noise-free molecular views whose direction of observation were determined to generate a series of projections which were compared to the cryo-negatively stained views by sinogram correlation functions. Using reprojections of the initial volume as new alignment references further refined the model. The final reconstruction included 12084 molecular images extracted from 26 images recorded at defocus values ranging between 0.5  $\mu\text{m}$  and 1.2  $\mu\text{m}$ . The images were low-pass filtered after the first node of the contrast transfer function of the microscope (CTF). The images were partially corrected for the contrast transfer function of the microscope by phase inversion between the minima.

The resolution of the final reconstructions was estimated from the Fourier shell correlation function obtained by comparing two independent reconstructions, generated by splitting randomly the data set in half. The two data sets were independently aligned and clustered and the resolution was given according to the 0.5 cut-off in the Fourier shell correlation curve (0.5 FSC criterion; Saxton et al., 1984) and the intersection point of the 3  $\sigma$  curve with the FSC curve (3  $\sigma$  criterion; van Heel, 1987).



## Acknowledgements

This work was supported by the Institut National de la Santé et de la Recherche Médicale, the Centre National pour la Recherche Scientifique, the Hôpital Universitaire de Strasbourg (HUS), the Association pour la Recherche sur le Cancer and SPINE.

## LEGENDS TO FIGURES

### **Figure 1** Purification of human TFIIE.

(a) Schematic representation of TFIIE purification steps. (b) Gel filtration of TFIIE with a Sephadex 16/60 200pg (Pharmacia) column. TFIIE containing fractions were analyzed by SDS polyacrylamide gel electrophoresis and coomassie staining (c) Determination of the native molecular weight of human TFIIE . The arrow marks the elution of the TFIIE / complex and circles (?) represent molecular weight standards as described in the text. (d) Native polyacrylamide gel electrophoresis (8-25%) analysis of purified TFIIE.

### **Figure 2** Three-dimensional model of the negatively stained TFIIE molecules

(a) Unprocessed original field of negatively stained TFIIE molecules. (b) Characteristic noise-free views of negatively stained particles obtained upon averaging typically 6 molecular images. (c) and corresponding reprojections (d) Surface representation of the 3D reconstruction of the negatively stained heterotetramer TFIIE. (e) Surface representation of the 3D reconstruction of the negatively stained dimer TFIIE. The arrows indicate the relative orientation of the model. Scale bar represents 30 nm in (a) and 10 nm in (b) and (c), 7.6nm in (d) and (e).

### **Figure 3** Three-dimensional model of the cryo-negatively stained TFIIE molecules

(a) Unprocessed original field of cryo-negatively stained TFIIE molecules. (b) Diagram representing the Fourier shell correlation function between two independent reconstructions (dotted line) and the 3 threshold curve (continuous line). The vertical axis indicates the value of the Fourier shell correlation coefficient versus the resolution in  $1/\text{Å}$  represented on the

horizontal axis. (c) A diagram showing the orientations of the 500 classes averages used in the final reconstruction. Each class average is represented by a point in a ( , ) coordinate system. (d) Surface representation of the 3-D reconstruction of the cryo-negatively stained human TFIIE filtered at a resolution of 1.7 nm. (e) Structure comparison between the cryo-negatively TFIIE model and the model obtained in negative staining. Scale bar represents 30 nm in (a) and 6.6 nm in (b) and 3.6nm in (e).

**Figure 4** Oligomeric state of TFIIE in solution

(a) Chemical crosslinking of TFIIE. Purified TFIIE (10<sup>1</sup> at 1.5 mg/ml) was mixed with 5  $\mu$ l of glyteraldehyde solution at different concentration. Lane 1, 2, 3, 4, 5 and 6 correspond to 0 %, 0.02%, 0.04%, 0.06%, 0.08% and 1%, respectively. The mixed reactions were loaded into SDS electrophoresis gel.

(b) Co-immunoprecipitation studies of TFIIE

**Figure 5** : Sedimentation equilibrium characterization of TFIIE.

TFIIE sample were centrifuged to equilibrium at 20 °C, in 20 mM Tris-HCl, 250 mM NaCl,  $\beta$ -mercaptoethanol buffer. The data set shown here was obtained at 12,000 rpm; the solid line represents the global fit of the single species model (equation 1). The value of Molecular weight was 84589 $\pm$  214 Dalton.

**Figure 6** : Mapping of the TFIIE- N-terminus

(a) Surface representation of the 3-D reconstruction of the cryo-negatively stained human GST-TFIIE (b) Surface representation of the 3-D reconstruction of the model obtained in negative staining. (c) Difference mapping between the aligned GST-TFIIE model and the native TFIIE model shows an additional domain corresponding in mass to the GDT. Scale bar represents 3.3 nm in (a) and (b) and 1.5nm in (d).

## REFERENCES

- Adrian, M., Dubochet, J., Fuller, S. D., and Harris, J. R. (1998). Cryo-negative staining. *Micron* 29, 145-160.
- Adrian, M., Dubochet, J., Lepault, J., and McDowell, A. W. (1984). Cryo-electron microscopy of viruses. *Nature* 308, 32-36.
- Buratowski, S., Hahn, S., Guarente, L., and Sharp, P. A. (1989). Five intermediate complexes in transcription initiation by RNA polymerase II. *Cell* 56, 549-561.
- De Carlo, S., Carles, C., Riva, M., and Schultz, P. (2003). Cryo-negative staining reveals conformational flexibility within yeast RNA polymerase I. *J Mol Biol* 329, 891-902.
- Flores, O., Lu, H., and Reinberg, D. (1992). Factors involved in specific transcription by mammalian RNA polymerase II. Identification and characterization of factor IIIH. *J Biol Chem* 267, 2786-2793.
- Inostroza, J., Flores, O., and Reinberg, D. (1991). Factors involved in specific transcription by mammalian RNA polymerase II. Purification and functional analysis of general transcription factor IIE. *J Biol Chem* 266, 9304-9308.
- Leuther, K. K., Bushnell, D. A., and Kornberg, R. D. (1996). Two-dimensional crystallography of TFIIB- and IIE-RNA polymerase II complexes: implications for start site selection and initiation complex formation. *Cell* 85, 773-779.
- Lu, H., Zawel, L., Fisher, L., Egly, J. M., and Reinberg, D. (1992). Human general transcription factor IIIH phosphorylates the C-terminal domain of RNA polymerase II. *Nature* 358, 641-645.
- Maxon, M. E., Goodrich, J. A., and Tjian, R. (1994). Transcription factor IIE binds preferentially to RNA polymerase IIa and recruits TFIIH: a model for promoter clearance. *Genes Dev* 8, 515-524.

Ohkuma, Y., Sumimoto, H., Hoffmann, A., Shimasaki, S., Horikoshi, M., and Roeder, R. G. (1991). Structural motifs and potential sigma homologies in the large subunit of human general transcription factor TFIIE. *Nature* 354, 398-401.

Ohkuma, Y., Sumimoto, H., Horikoshi, M., and Roeder, R. G. (1990). Factors involved in specific transcription by mammalian RNA polymerase II: purification and characterization of general transcription factor TFIIE. *Proc Natl Acad Sci U S A* 87, 9163-9167.

Orphanides, G., Lagrange, T., and Reinberg, D. (1996). The general transcription factors of RNA polymerase II. *Genes Dev* 10, 2657-2683.

Peterson, M. G., Inostroza, J., Maxon, M. E., Flores, O., Admon, A., Reinberg, D., and Tjian, R. (1991). Structure and functional properties of human general transcription factor IIE. *Nature* 354, 369-373.

Roeder, R. G. (1996). The role of general initiation factors in transcription by RNA polymerase II. *Trends Biochem Sci* 21, 327-335.

Sauer, F., Fondell, J. D., Ohkuma, Y., Roeder, R. G., and Jackle, H. (1995). Control of transcription by Kruppel through interactions with TFIIB and TFIIE beta. *Nature* 375, 162-164.

Sumimoto, H., Ohkuma, Y., Sinn, E., Kato, H., Shimasaki, S., Horikoshi, M., and Roeder, R. G. (1991). Conserved sequence motifs in the small subunit of human general transcription factor TFIIE. *Nature* 354, 401-404.

van Heel, M., Harauz, G., Orlova, E. V., Schmidt, R., and Schatz, M. (1996). A new generation of the IMAGIC image processing system. *J Struct Biol* 116, 17-24.

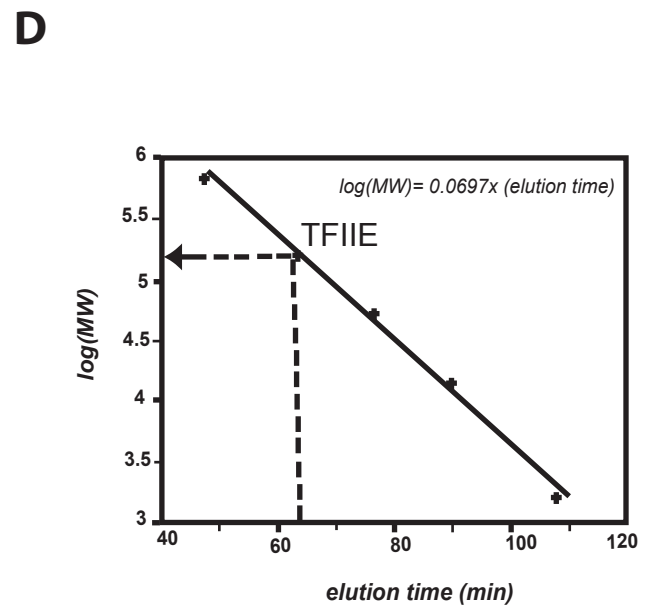
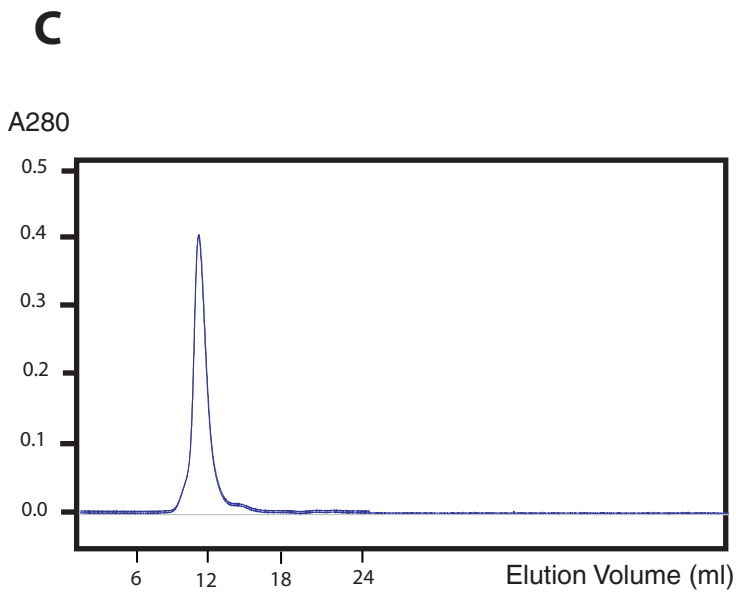
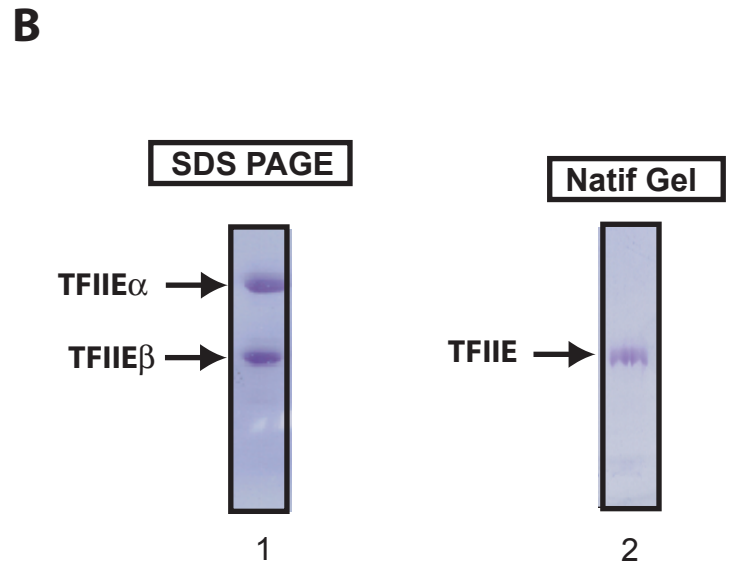
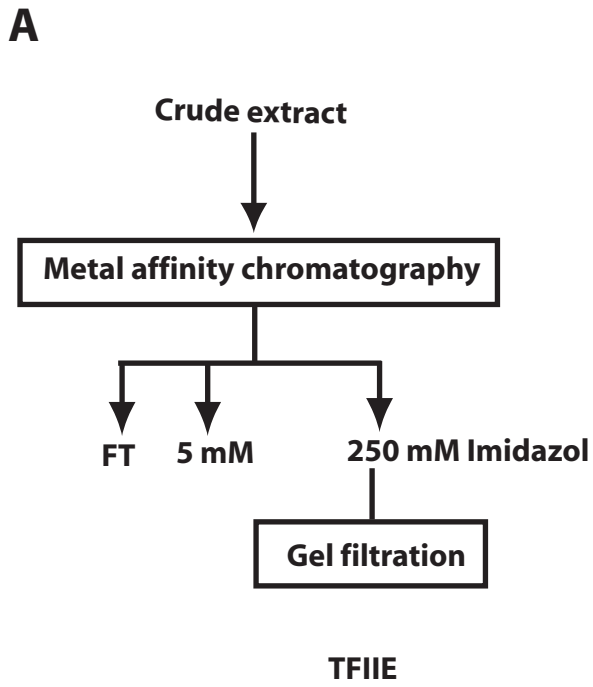


Figure 1

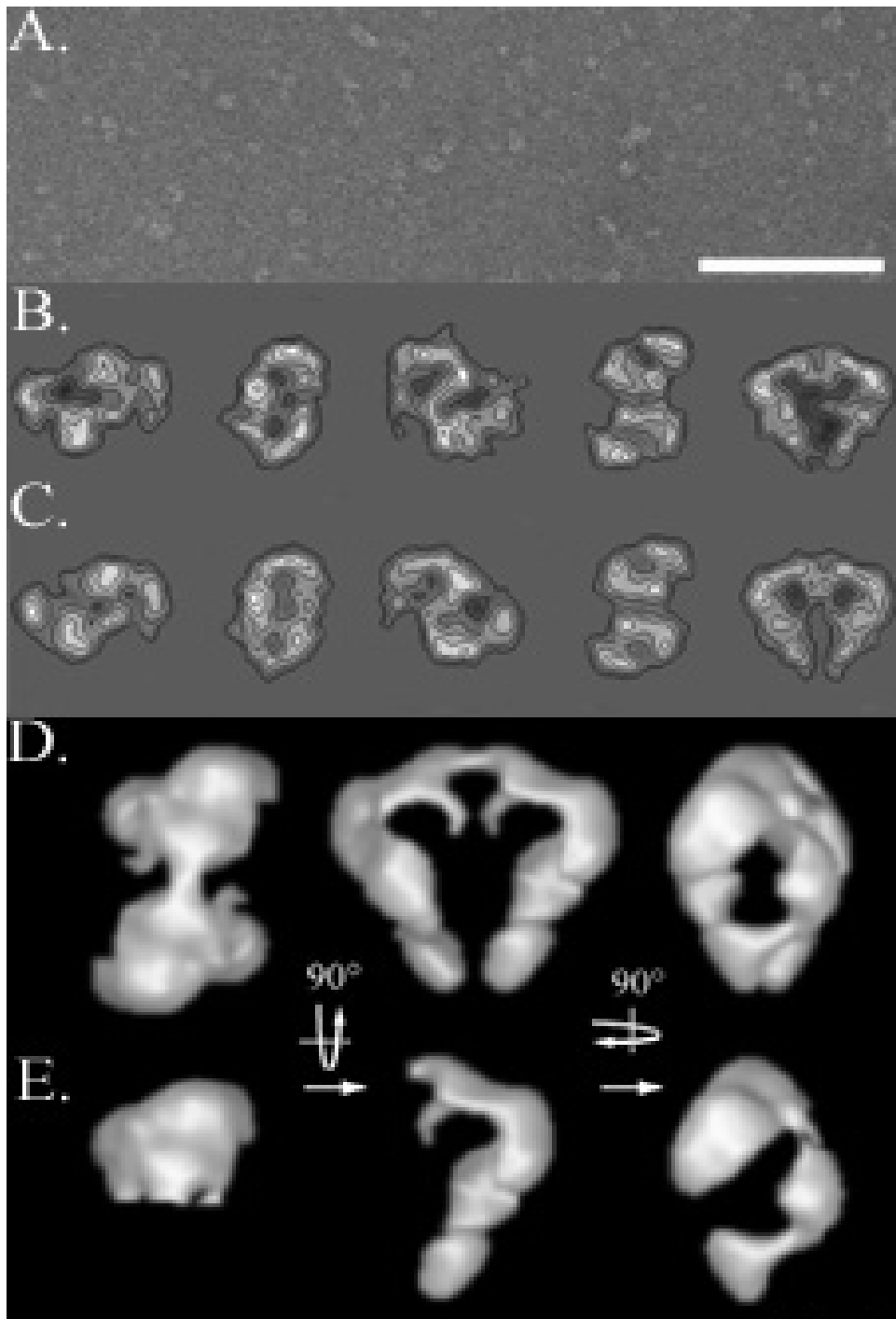


Figure 2

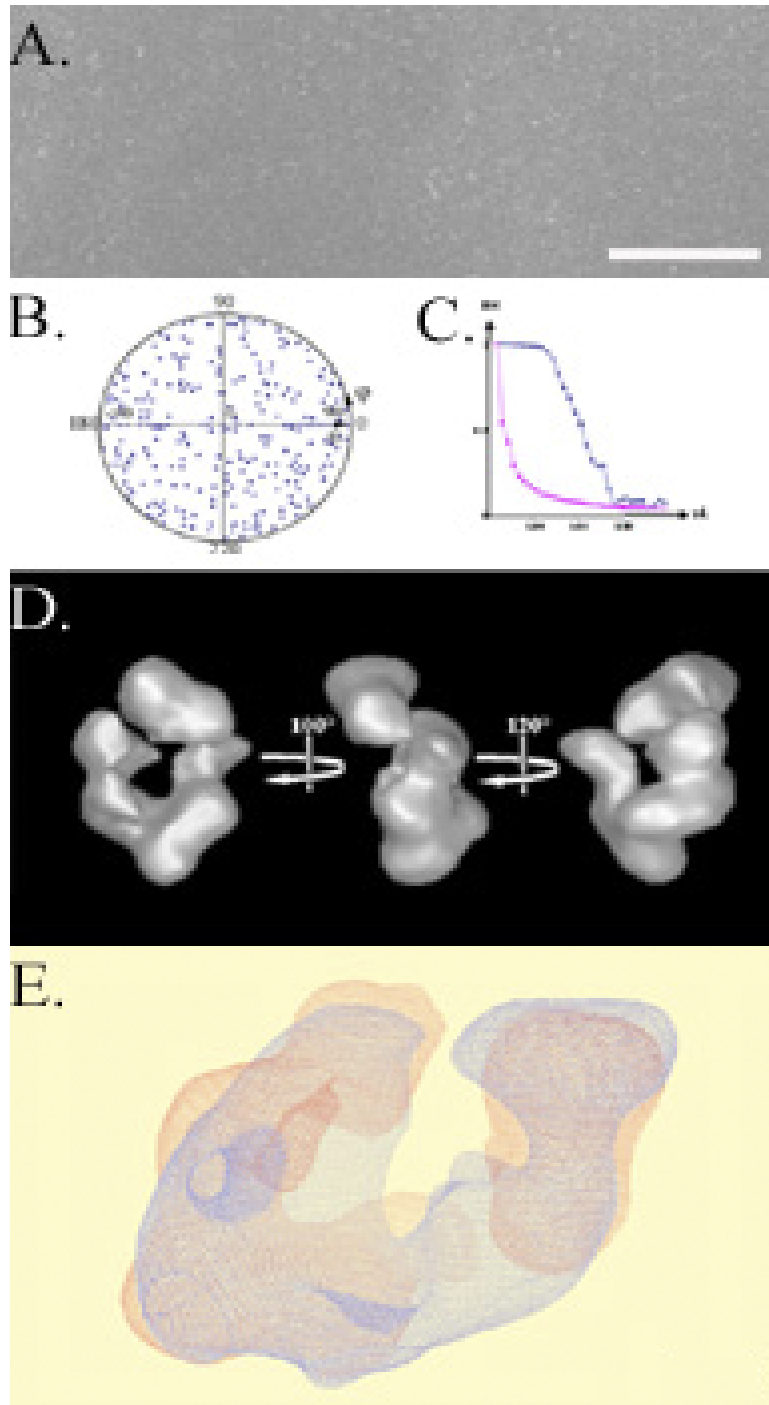


Figure 3

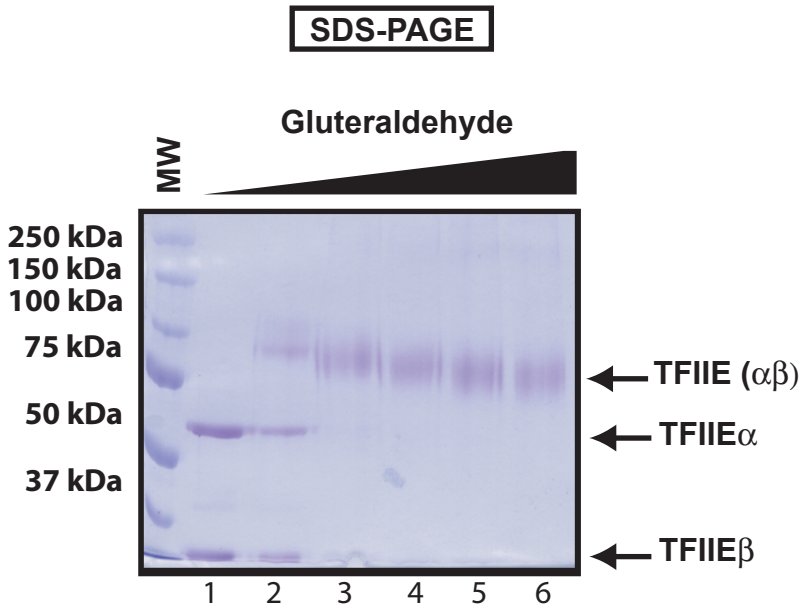
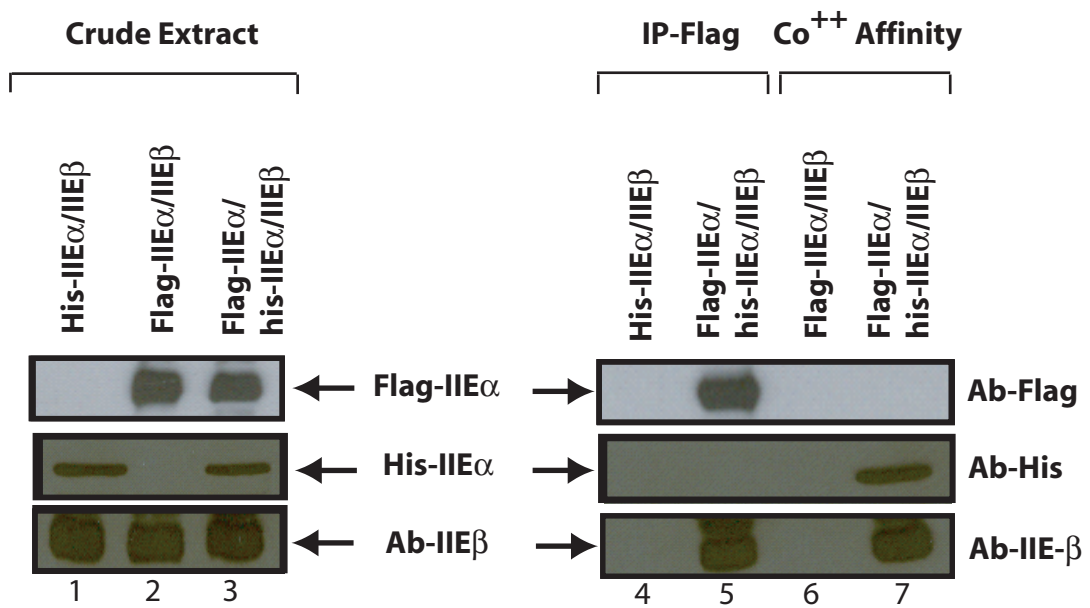
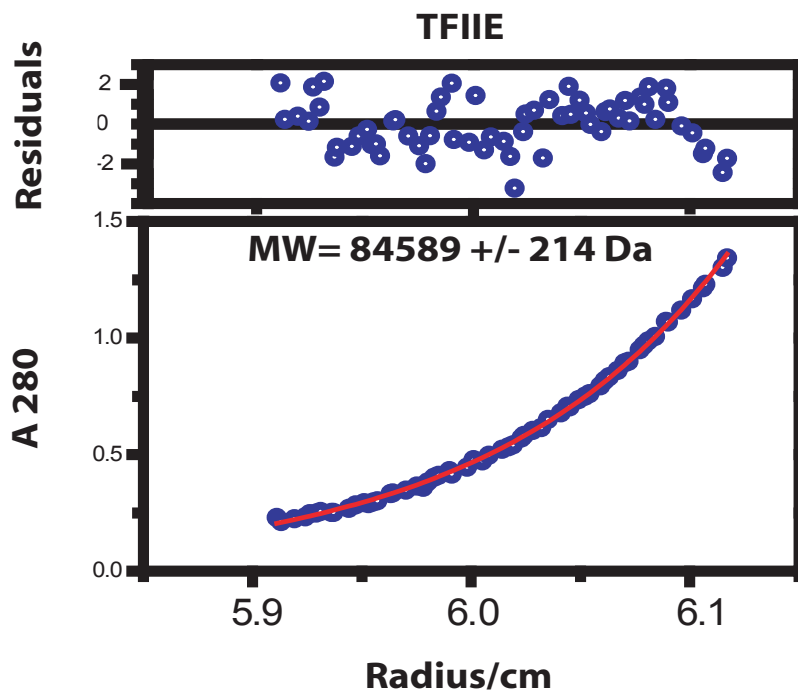
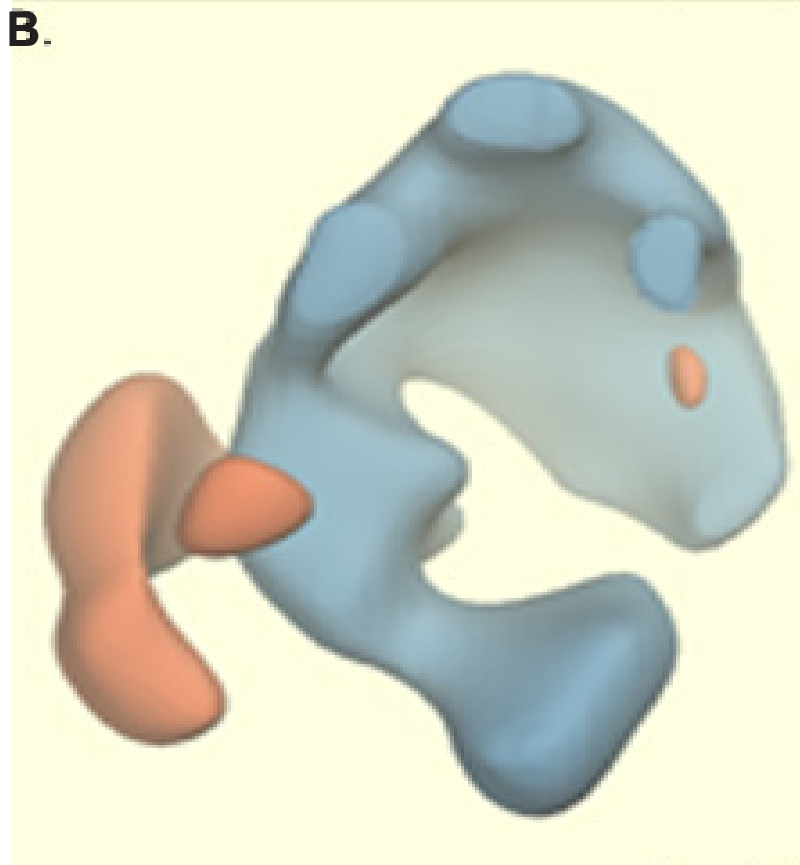
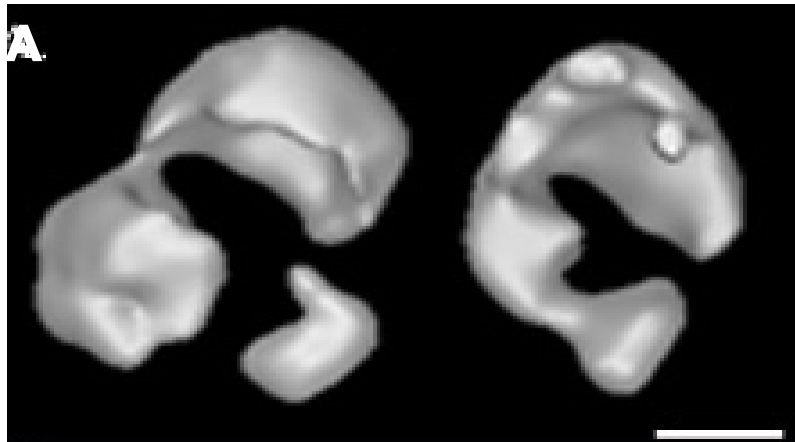
**A****B**

Figure 4





**Figure 5**



**Figure 6**

## Chapitre V : PRODUCTION ET CARACTERISATION DES COMPLEXES RECOMBINANTS

### Article 4: Expression of FLAG Fusion Proteins in Insect Cells : Application to the Multi-Subunit Transcription/DNA Repair Factor TFIIH

Mon travail de thèse a directement contribué à développer un protocole de purification du complexe TFIIH entier produit sous forme recombinante dans des cellules *Sf9* à l'aide du système d'expression baculovirus. Le facteur recombinant est actif dans un test de transcription *in vitro*. Cette étude a montré en outre que la structure moléculaire du facteur recombinant, déterminée par microscopie électronique, était identique à celle du facteur endogène. L'accès à de tels complexes recombinants ouvre des perspectives nouvelles et permettra d'introduire des mutations, identiques à celles trouvées chez les malades, dans certaines sous-unités et d'en analyser les effets sur la structure (et la fonction) de TFIIH.

De plus, nous avons mis en place un système de purification basée sur l'immunoprécipitation anti-Flag qui nous a permis d'étudier les interactions protéine-protéine au sein de complexe.

Dernièrement, la dixième sous-unité de TFIIH a été identifiée chez la levure (Giglia-Mari et al., 2004d). TFB5 (p8 chez l'homme) a été caractérisé par des techniques cellulaires, moléculaires et biochimiques et la mise en évidence de trois mutations du gène qui code TFB5 sont directement incriminées dans la forme A de la trichothiodystrophie. Cette sous-unité du facteur TFIIH pourrait jouer un rôle de stabilisation du complexe recombinant. Il est essentiel de produire du complexe dans les cellules d'insectes en y incorporant un virus exprimant la sous-unité p8. L'influence de p8 sera déterminée de manière fonctionnelle, par des tests de transcription et de réparation *in vitro* (le complexe recombinant à neuf sous-unités présentait une activité plus faible en réparation que le facteur endogène) et de manière structurale, par microscopie électronique.

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

Anass Jawhari, **Muriel Uhring**, Corinne Crucifix, Sébastien Fribourg, Patrick Schultz, Arnaud Poterszman, Jean Marc Egly and Dino Moras

**Expression of FLAG Fusion Proteins in Insect Cells: Application to the Multi-subunit Transcription/DNA Repair Factor TFIID**

Protein expression and purification 24, 513-523 (2002)

Pages 513-523 :

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

Pour les utilisateurs ULP, il est possible de consulter cette publication sur le site de l'éditeur

<http://dx.doi.org/10.1006/prev.2001.1591>

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

## **Article 5: Impact of fusion protein of insect cells expressed proteins**

L'expression des protéines eucaryotes recombinantes dans les cellules d'insectes est devenu un axe de recherche majeur pour la caractérisation structurale et fonctionnelle de ces protéines souvent mal repliées dans les cellules d'*Escherichia coli*. Dans le système d'expression baculovirus les modifications post-traductionnelles sont conservées, la protéine adopte généralement un bon repliement et des niveaux d'expression relativement hauts peuvent être atteints. Il est cependant souvent difficile d'obtenir une protéine pure dans des quantités suffisantes pour débiter une étude structurale. C'est pourquoi l'utilisation de protéines de fusion dans le système d'expression baculovirus tend à se développer. Nous décrivons ici une méthode de purification d'un complexe humain recombinant par l'utilisation de cinq protéines de fusion dans le système d'expression baculovirus. Les cinq protéines de fusions, l'épitope Flag, la Calmodulin Binding Protein, la thioredoxine, la protéine A et la GST ont été comparées en terme de gain de pureté et de quantité de protéine obtenue. Une étude comparative de l'utilisation de ces tags d'affinité dans les cellules d'insectes est présentée.

# **Preliminary results**

## **Influence of affinity tags for protein production in insect cells**

Keyword : affinity tags, baculovirus expression system, recombinant protein, purification

Structural Biology and Genomics Laboratory, Institut de Génétique et de Biologie  
Moléculaire et Cellulaire (IGBMC), CNRS/INSERM/ULP, Illkirch, France

## ABSTRACT

For structural study it is often difficult to obtain pure and correctly folded human proteins. Since post-transcriptional modifications are important for folding and activity, insect cells system has gain interest to produce recombinant proteins at a yield compatible with structural project. In order to easily obtain pure protein, affinity tags are become a commonly used technique. Here, we have compared five different vectors for expression of proteins with N-terminal affinity tags (the flag-tag, the calmodulin binding protein (CBP), the protein A, the thioredoxin and the glutathione S transferase (GST)) in insect cells by testing these on a human complex. The influence of tags on protein solubility and expression yield is estimated. From this comparative study, we conclude that CBP and thioredoxin tags do not enhance yield and solubility of recombinant proteins and proteins are not as pure after one step purification. Protein A and GST tags improve protein solubility and allow high level of purity in one step purification. This comparative study gives insights on the impact of affinity tags on recombinant protein-proteins complexes expressed in insect cells that is helpful to carry out structural investigation.

## INTRODUCTION

Due to fast growth, easy handling and low cost, *Escherichia coli* is often used for expression of recombinant proteins. Moreover, since quantity of expressed proteins can easily reach mg, *Escherichia coli* is the system of choice for structural studies. Consequently, most of the current published structures come from proteins expressed in the *Escherichia coli* system (Wang et al., 2003). Unfortunately, expression of enzymatically active mammalian proteins like membrane proteins or large multi protein complexes in *Escherichia coli* can proven to be a challenging task due to poor solubility, improper folding, and lack of adequate posttranslational modification (Basagoudanavar et al., 2004; Zeghouf et al., 2004; Mohanty and Wiener, 2004). Therefore, implantation of other expression system is necessary. Insect cells, infected by recombinant baculovirus allow post transcriptional modifications and correct folding of proteins expressed at fairly high levels (Dolby et al., 2004; Joel et al., 2004; Cadel et al., 2004). Consequently, expression of recombinant eukaryotic proteins in insect cells has become an important research approach for their structural and functional characterization. However, obtaining pure proteins in quantity sufficient for structural study is often difficult. In order to reduce number of purification steps, the use of affinity tags has become more popular and some epitope peptides and affinity proteins have been developed in insect cells (Darrow et al., 2003; Sheffield et al., 1999). Number of commercial vectors for expression of recombinant proteins harbouring such affinity tags are already developed for *E.coli* (Hewitt and McDonnell, 2004) but their utilization in the insect cells expression system is not so common to allow (i) a one-step purification procedure, (ii) a minimal effect on tertiary structure and biological activity (Zhang et al., 2001; Smyth et al., 2003) and (iii) easily cleaved of by specific proteases .

The human transcription factor IIIH (TFIIH) that consists of ten sub-units (p8, p34, p44, p52, p62, XPD/p80, XPB/p89, cdk7, cyclinH and MAT1) for a total molecular mass of 468Kda (Giglia-Mari et al., 2004) was the first factor found to possess several catalytic activities such as ATPase, helicases and kinase (Coin and Egly, 1998). The recombinant factor has been expressed in insect cells and purified fully active via immunopurification anti-flag in our laboratory (Jawhari et al., 2002). The molecular envelop of the recombinant TFIIH factor (rIIIH9) has been obtained by electronic microscopy and is relatively similar to the endogenous factor . Thus, we have decided to use the insect cells expression system to



identify protein protein interaction in order to characterize stable sub complexes of TFIID. We have demonstrated the feasibility of the technique by producing and by purifying via anti-Flag immunoaffinity the p34/p44 sub complex already characterized by coexpression in *Escherichia coli* (Fribourg et al., 2001).

It is known that tags may influence protein behaviour in term of expression and solubility but no rules exist to define the best tag. Consequently, we have decided to construct expression vectors for insect cells production of recombinant protein fused to calmodulin binding protein (CBP), protein A, thioredoxin and glutathione S transferase tag (GST) and we used the p34/p44 complex as a model. We compare the influence of tags on protein solubility and show that protein A and GST tag play in this way good effect. We next determine which tag is more powerful to purify the recombinant complex near to homogeneity in the baculovirus system. Using GST fusion led to a complex with one contaminant which could be eliminate with another step of purification .

Here, we described a method to express and purify recombinant proteins in insect cells via five different affinity tags that provide the first answer to affinity protein to be used in the baculovirus expression system using a sub-complex of the transcription factor TFIID. This study is directly applicable to many proteins, difficult to be produced in *Escherichia coli*.

## MATERIAL AND METHODS

### *Construction of the baculovirus fusion vectors*

The fusion proteins were constructed and expressed in the plasmid pBacPAK8 (BD Biosciences, Palo Alto, USA). The pBacPAK8 is a transfer vector designed for high-level expression of a cloned gene under control of the strong AcMNPV polyhedrin promoter. The sequences flanking AcMNPV promoter allow recombination with viral DNA to transfer the expression cassette to the polyhedrin locus of the viral DNA. The polyhedrin coding sequences have been replaced by a multiple cloning site (MCS) with 18 unique sites that facilitate insertion genes in the correct orientation for expression. The plasmid was modified as follow: six oligonucleotides (10pmoles/ $\mu$ l) were phosphorylated using 1 $\mu$ l of T4 kinase (10U/ $\mu$ l) (NEB, Beverly, MA), mixed and hybridized over-night in a water bath to form a DNA cassette containing the consensus Kozak sequence, a tag cassette *NcoI-SacII* and a protease cassette *SacII-NdeI* (PreScission Protease, Amersham Biosciences, Little Chalfont, UK) (table 1 (A)). The cloning cassette was ligated into the pBacPAK8 *BamHI-EcoRI* sites using T4 DNA ligase (NEB) resulting in pBac-Tag vector (figure 1A).

The ORF (open reading frame) of the different fusion tags Gluthatione S Transferase (GST), Protein A, Thioredoxine (TRX) and Calmodulin Binding Protein (CBP) (figure 1C) were amplified by PCR using as a template 2xTAP tag plasmid for CBP and Protein A amplification, pGEX (Amersham Biosciences) for GST amplification and the pET32b (Novagen, Darmstadt, DE) for amplification of the thioredoxin using the specific primers (Table 1 (B)). The resulting PCR products corresponding cDNA were initially sub-cloned into the pCRBlunt vector (Invitrogen, Carlsbad, CA).

Thus, the pCRBlunt vectors were digested by *NcoI-SacII* and the resulting cDNA encoding fusion protein were cloned in the pBacpak8 vector previously digested by the *NcoI-SacII* enzymes resulting in the pBac-FLAG, pBac-CBP, pBac-TRX, pBac-PA and pBac-GST vectors (figure 1B). The ORF coding for the p34 protein was obtained by digesting the pET15-p34 vector (unpublished results) with *NdeI-BamHI* enzymes and cloned into the pBacTag fusion vector digested by *NdeI-BamHI*. The correctness of all sequences was verified by DNA sequencing at the different step of the procedure.

### *Insect cells*

*Sf9* (*Spodoptera frugiperda*) cells were grown in Grace's medium (Invitrogen) supplemented with 10% fetal bovine serum (FBS; Invitrogen). Baculovirus infections were carried out as follows: *Sf9* cells were seeded in monolayers at  $10^6$  cells per well, the culture medium was removed and replaced with virus inoculums at a multiply of infection (MOI) of 4 for pVL1392 vector coding for p44-his or p44 (Tirode et al., 1999) and 5 for pBac-Tag vector. After removing media containing the unbound virus, fresh medium was added and the cells were incubated at 27 °C for 48 hours. The cells were harvest by centrifugation, washed in 1x PBS containing 20% glycerol, frozen in liquid nitrogen and stored at -80°C.

### *PAGE analysis and immunoblots*

Protein samples were analyzed after boiling in SDS-loading buffer by SDS-PAGE on 12.5% polyacrylamide gels using Tris–glycine buffer (Dunn, 1986) and stained with Coomassie blue R250 (Serva, Heidelberg, DE). Proteins were electrotransferred onto nitrocellulose membranes (Immobilon Millipore). Membranes were blocked with 5% milk in TBS (50 mM Tris/HCl, pH 7.5, 150 mM NaCl, and 0.05% Tween 20), incubated with anti-p34 or anti-p44 antibodies, and developed using Enhanced Chemiluminescence (Amersham Biosciences).

### *Purification of affinity protein complex*

To over express the affinity fusion complex, cell pellet of infected *Sf9* cells was resuspended in 50 ml ice-cold buffer A containing 20mM Tris-HCl pH 7.5, 250mM NaCl. A cocktail of protease inhibitors (PIC; Sigma-Aldrich, Lyon, France) and  $\beta$  mercaptoethanol were added to a final concentration of 1x and 2mM, respectively. The insect cells were mechanically disrupted using a dounce homogenizer. The cell lysate was centrifuged at 30,000g and supernatant was incubated with the appropriate resin in the following conditions: for GST fused protein purification, approximately 1ml swollen pre-equilibrated glutathione-agarose beads (Sigma-Aldrich) for at least 2hours at 4°C are used to allow GST-fusion proteins binding to beads. Agarose beads were washed three times with in 10ml of buffer A to remove non-specifically bound proteins. Bound GST-fusion proteins were eluted with 0.5ml of 20mM reduced glutathione in buffer A.

For thioredoxin (TRX) fused protein purification, the supernatant was incubated with 1ml swollen Thiobond resin (Invitrogen), pre-washed in 20mM  $\beta$  mercaptoethanol for 30 minutes at room temperature. To allow thioredoxin-fusion proteins binding, the supernatant was incubated for at least 2hours at 4°C. Beads were washed three times with 20mM 10ml of buffer A to remove non-specifically bound proteins. Bound TRX-fusion proteins were eluted with 0.5ml of 20mM  $\beta$  mercaptoethanol in buffer A.

For CBP fused protein purification, the supernatant was incubated 2hours at 4°C with Calmodulin affinity resin (Stratagene, La Jolla, CA) pre-equilibrated in buffer A containing 2mM CaCl<sub>2</sub> (buffer B). Beads were washed three times with 10ml of buffer B. Bound CBP-fusion proteins were eluted with 0.5ml of buffer A containing 2mM EGTA.

For Protein A fused protein purification, the supernatant was incubated 2hours at 4°C with IgG sepharose fast-flow affinity resin (Amersham Biosciences) pre-equilibrated in buffer A. To remove non-specifically bound proteins, beads were washed three times with 10ml of buffer A. Bound protein A-fusion proteins were removed by cleaving off the Precision protease site over-night at 4°C.

For FLAG fusion protein purification, extracts were incubated with anti-Flag M2 monoclonal antibody-conjugated agarose beads (Flag-M2, Sigma-Aldrich) previously equilibrate in buffer A at 4°C for 4hours. The resin was washed three times with buffer A. Bound complexes were eluted in buffer A containing 0.2mg/ml of the Flag peptide (DYKDDDDK) by incubation for 8-12h at 4°C.

#### *Quantification of purified complex*

An aliquot (10 $\mu$ l) of purified proteins were separated by a SDS-PAGE gel further stained with Coomassie blue to determine the recovery and the purity of the complex. The amount of purified complex was estimated using the Quantity One Software 4.1.1 of the ChemiDoc Imaging System (Biorad, Hercules, CA).

## RESULTS and DISCUSSION

### *Construction of the baculovirus fusion vectors*

Expression of proteins in *Escherichia coli* can be achieved by using any one of several commercial fusion protein vectors. Two common fusion protein systems used are the His peptide Tag and glutathione S-transferase (GST). These systems work well for a variety of proteins but problems with mammalian multi protein complexes or membrane protein make necessary to switch to a different expression system. The BEVS (baculovirus expression vector system) allow post-translational modifications for properly folded of enzymatic mammalian complexes. Thus, we designed a transfer vector for expression of protein with affinity tags in insect cells. We modified the pBacPAK8 baculovirus expression vector (BD Biosciences) which has a large selection of restriction sites and introduced a cloning cassette to form the pBac-Tag vector (figure 1A). Family of resulting affinity protein transfer vectors allows cloning of the ORF of the gene of interest in the *NdeI-BamHI* sites with an affinity tag fused to the N-terminus (figure 1B). The five transfer vectors, named pBac-FLAG, pBac-CBP, pBac-TRX, pBac-PA and pBac-GST vectors were designed to express protein of interest with respective affinity tags : flag, calmodulin binding protein (CBP), protein A (PA), thioredoxin (TRX), and glutathione S transferase (GST).

Tags sizes (figure 1B) influence protein behaviour to which is fused. Small tag (such as CBP) does not perturb folding (Klein, 2003; Melkko and Neri, 2003) while larger tags (like thioredoxin, protein A or GST) have a beneficial effect on protein solubility (Terpe, 2003; Mercado-Pimentel et al., 2002). To validate this purpose we introduced into the *NdeI-BamHI* sites of the baculovirus transfer vector the ORF of p34 and obtained the pBac-FLAG-p34, pBac-CBP-p34, pBac-TRX-p34, pBac-PA-p34 and pBac-GST-p34 vectors. With the goal of producing p34/p44, a sub complex of the human transcription factor TFIIF, as a purification model of a mammalian complex expressed in insect cells.

### *Purification of p34/p44 complexes produced in insect cells*

To test the ability to purify the p34/p44 complex we produced in insect cells, we fused a flag-tag at the N-terminus of p34 and characterized complex formation after immunoprecipitation using anti-Flag M2 monoclonal antibody-conjugated agarose beads (Flag-M2, Sigma-Aldrich). The complex was specifically retained on beads and eluted after incubation with peptide mimics the Flag sequence (DYKDDDDK). Both subunits forming a stable p34/p44 sub-complex in insect cells were revealed on Coomassie stained SDS-PAGE (figure 3B, lane1). Thus, we co-infected each of the four baculovirus vectors expressing the p34 subunit fused to the different affinity tags with the pVL1392 vector expressing the p44-His subunit in order to form the complex p34tags/p44-His in insect cells.

To test the impact of the different affinity tags on the formation of the p34/p44 complex, we co-infected the pBac-FLAG-p34, pBac-CBP-p34, pBac-TRX-p34, pBac-PA-p34 and pBac-GST-p34 vectors with the p44-His one in insect cells. The His-tag is commonly used in *Escherichia coli* expression system and allows rapid purification of recombinant protein (Woestenenk et al., 2004). In the baculovirus system, the yield of expressed protein is lower than in *Escherichia coli* and the use of the His-tag is not so extensively used. Purification of p44-His/p34-tags complexes using IMAC (Immobilized metal affinity chromatography) in insect cells was investigated. Expression of p44-His/p34-tags complexes were analyzed after Coomassie staining of SDS gels and by Western blotting using specific monoclonal antibodies raised against both subunits (figure 2). Each one of the p34-tagged subunit is able to form a complex with p44-His in insect cells. However, purification yield of every p44-His/p34-tagged complexes is not good. Multiple bands were visible in complexes eluates purified by affinity chromatography. This phenomenon probably occurred because of partial degradation of His-tag, or some bands were chaperonins involved in the proper folding of nascent proteins in *Sf9* cells. Another important reason could lie in the specificity of the resin, non-specific binding for IMAC (immobilized metal ion affinity chromatography) columns is one of the drawbacks of polyhistidine tags (Wu and Filutowicz, 1999). Theoretically fusion protein can be purified to near homogeneity from crude biological mixtures by a single affinity chromatography step. In order to get highly purified p34/p44 complex, family of affinity tags vectors were tested.

### *Purification yield of p34-tags/p44 complexes*

Baculovirus expression vectors pBac-FLAG-p34, pBac-CBP-p34, pBac-TRX-p34, pBac-PA-p34 and pBac-GST-p34 were co-infected with pVL1392-p44 in insect cells. To evaluate affinity tags influence on solubility, expression level of each p34tags/p44 complex was analyzed in the *Sf9* crude cell extract by Western blotting (figure 3A). GST and protein A affinity tags improve solubility of p34 fused protein in comparison to the other tags, flag-tag, CBP and thioredoxin which show lower influence on protein solubility. (figure 3A, compare lane 3 and 5 to 1,2 and 4). Solubility improvement due to GST and protein A has already been described in *Escherichia coli* expression system (Smith and Johnson, 1988).

Affinity tags purification were carried out in order to determine which vector(s) would be preferred for purification of soluble p34/p44 tagged complex. All tagged complexes could be produced in insect cells (figure 3B). Protein concentrations in the eluates were determined using Bradford assay and by gel band analysis. After Coomassie staining of SDS gels, bands intensity were estimated by the Quantity One Software (Biorad) and compared to protein concentration determined with the Bradford assay (table 2). In cases where band intensity is small compared to the concentration obtained by the Bradford assay, we observed that the protein sample were not pure (see p34-CBP/p44 and p34-TRX/p44). The opposite situation, in which band intensity is larger than would be expected from Bradford measurements, was observed for three complexes p34-FLAG/p44, p34-GST/p44, p34-proteinA-p44). We found that for p34-p44 complex purification yield, pBac-FLAG-p34, pBac-PA-p34 and pBac-GST-p34 perform equally well, while the other two vectors pBac-CBP-p34 and pBac-TRX-p34 give low yield of complex (figure 4).

After a single affinity chromatography step the tagged complexes were only partly purified for CBP and thioredoxin system. Expression of GST-complexes showed on SDS-PAGE gel a major contaminant band, described as intrinsic GST protein of insect cells (Bichet et al., 2000). Single step purification of soluble FLAG and proteinA-complexes were achieved.

## LEGENDS TO FIGURES

**Figure 1 : Construction of fusion protein baculovirus transfer vectors.** (A) The baculovirus transfer vector pBacpak8 contains the polyhedrin promoter P (Ppolyhedrin) derived from *Autographa californica* nuclear polyhedrosis virus (AcMNPV) and a multiple cloning site (MCS). pBacpak8 has been modified in order to introduce a linker between *Bam*HI-*Eco*RI sites creating specific site for (i) tag sequence insertion (ii) protease cleavage site and (iii) ORF insertion (see material and methods) and (B) the resulting vector were named pBac-FLAG, pBac-CBP, pBac-TRX, pBac-PA and pBac-GST . The unique *Nco*I and *Sac*II sites were used to introduce encoding sequence for affinity protein; the sequence encoding the precision protease cleavage site was introduced between the *Sac*II and *Nde*I sites ; finally the cDNA coding for the protein of interest was cloned into the *Nde*I and *Bam*HI sites. The AcMNPV sequences flanking the promoter are used for *in vivo* homologous recombination in insect cells. (C) The number of residues and size in kilo Daltons (kDa) of the fusion proteins are summarized.

**Figure 2 : His tag purification.** *Sf9* insect cells were infected with baculoviruses expression of the following complexes: p44HIS/p34Flag, p44HIS/p34CBP, p44HIS/p34GST, p44HIS/p34thioredoxin, p44HIS/p34proteinA. Complexes expressed in the clarified lysates were subjected to purification using the histidine tag. The eluted proteins (see material and methods) were analyzed by SDS-PAGE (12.5%) (upper panel) followed by Western blotting using specific antibodies (lower panel). p44-His is indicated by (\*) and the different p34-tag are indicated by ( $\Delta$ ).

**Figure 3 : Purification yield of p34tags-p44 complexes.** The insect cell lysates infected with p44/p34Flag, p44/p34CBP, p44/p34GST, p44/p34thioredoxin, p44/p34proteinA baculoviruses are purified using tag fused to p34 moiety on appropriate resin (see material and methods). (A) Crude cell extract were analyzed by Western Blotting (B) Eluted proteins were analysed by SDS-PAGE (12.5%) followed by Coomassie staining.

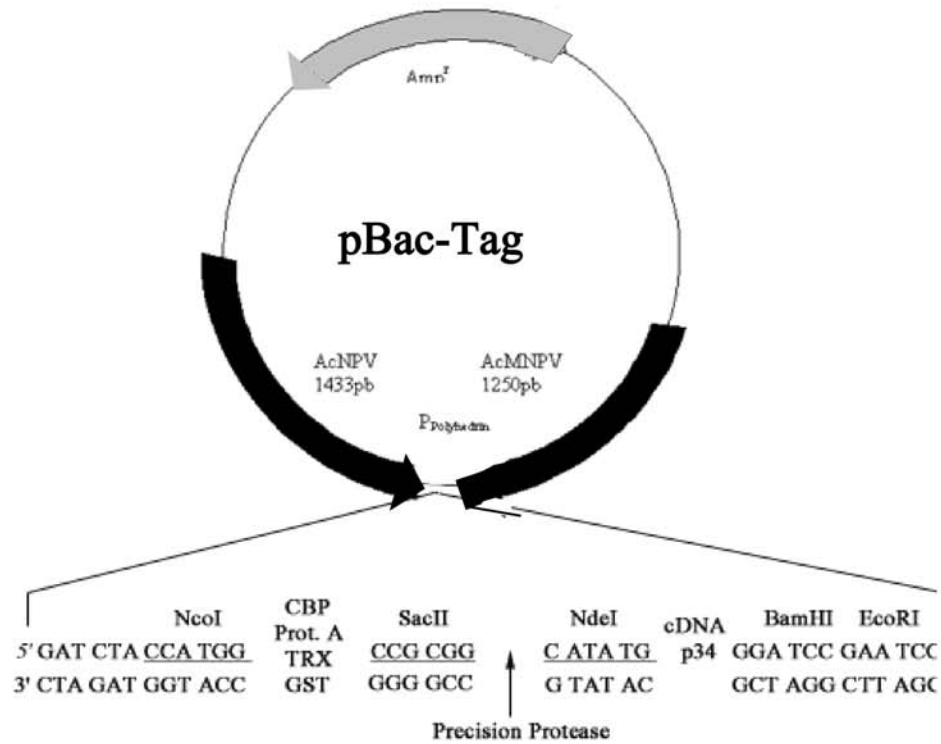


**Figure 4 :** Comparison on expression rates of the five fusion systems. The success rate is represented in percentage for each fusion protein. Based on the results of three experiments.

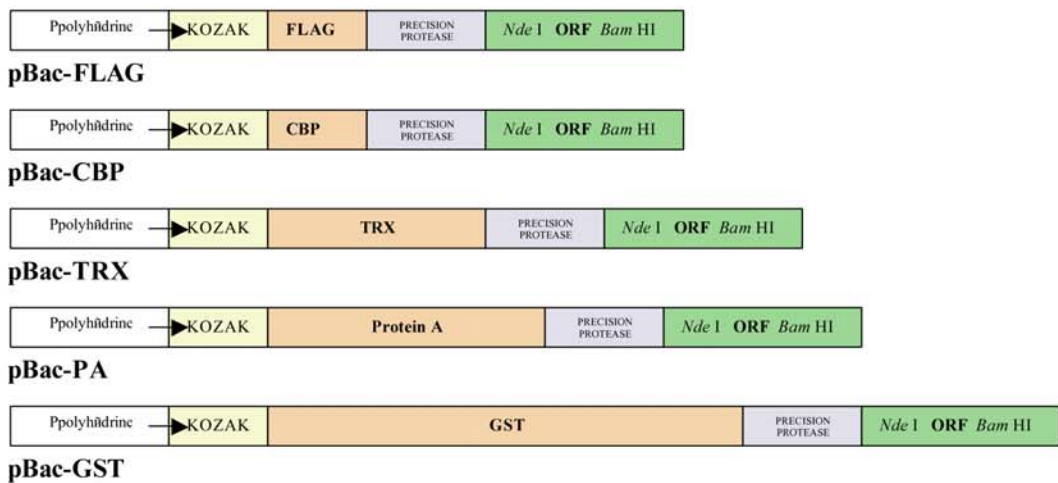
**Table 1: Fusion vector construction** (A) Oligonucleotides used for the construction of the cloning cassette (B) Sequences of DNA primers used for PCR cloning. The underlined sequence is the *NcoI* enzyme recognition site used for cloning in primer 1 and the *SacII* enzyme recognition site in primer 2.

**Table 2 : Purification of p34/p44 complexes.** <sup>a</sup> Total protein was determined using Bradford assay. <sup>b</sup> To estimate protein concentration, 10 $\mu$ l of eluate were loaded onto 12.5% SDS-PAGE, after Commasie staining, band intensities were analyzed by the Quantity One Software (Biorad) and compared to protein concentrations determined with Bradford assay <sup>c</sup> The volume of purified complex is 300 $\mu$ l.

A.



B.



C.

Tags	Residues	Size (kDa)
Flag	8	1.01
CBP	26	2.96
Thioredoxin	93	10
Protein A	137	15
GST	211	26

Figure 1

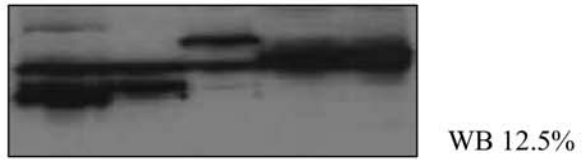
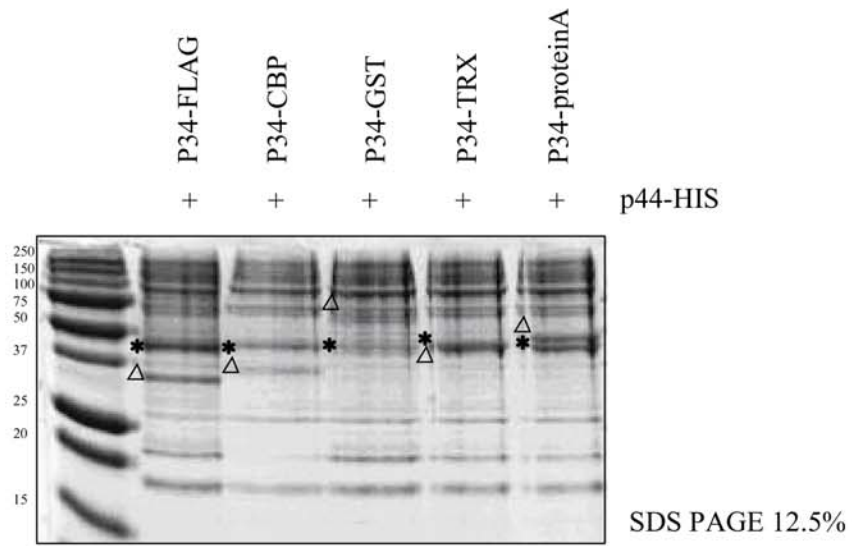
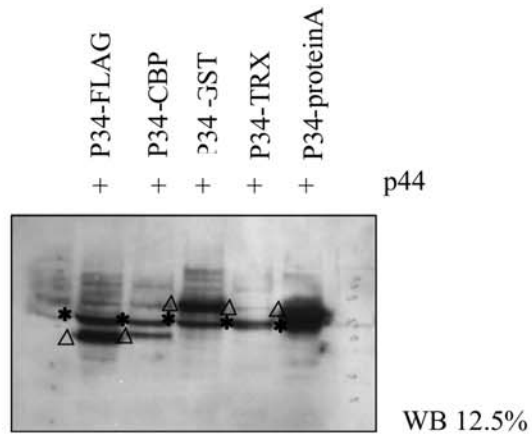


Figure 2

A.



B.

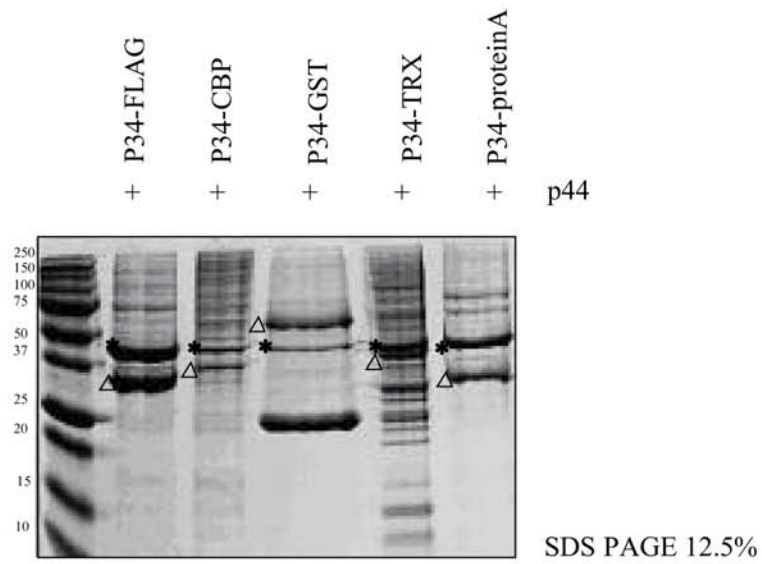


Figure 3

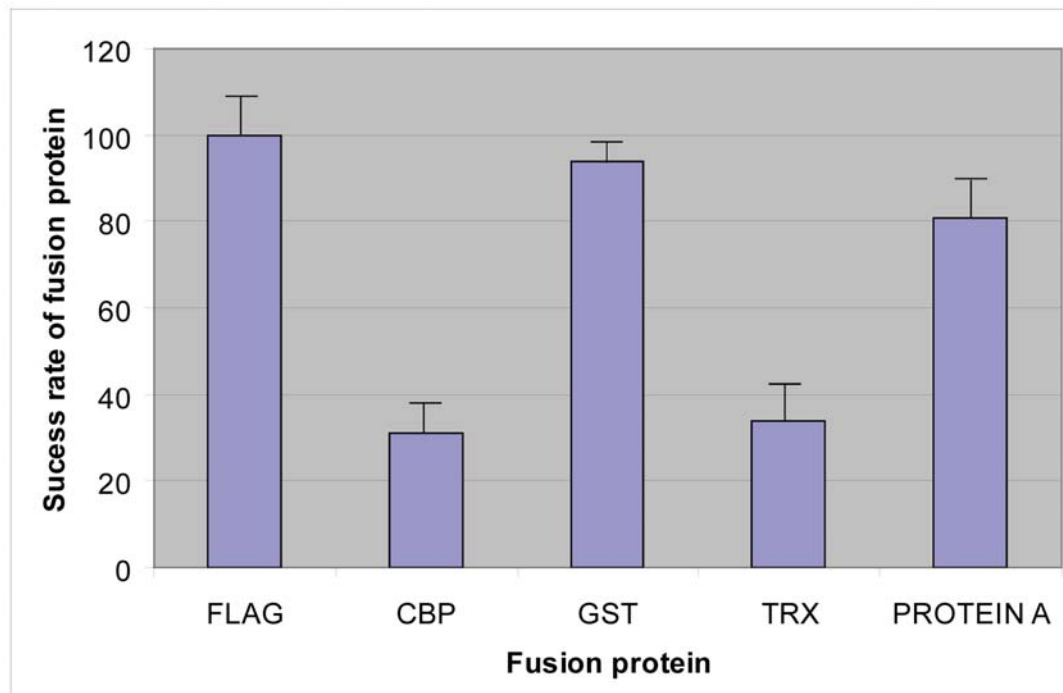


Figure 4

A.

	Primer
1	5'GATCTACCATGGACTACAAGGACTCCGCGG 3'
2	5' GTCTGGAAGTTCTGTTCCAGGGGCCCCATA 3'
3	5' TGCTCGAGGGATCCG 3'
4	5' CCCCAGGGTATACGAGCTCCCTAGGCTTAA 3'
5	5' CCTGAGGCGCCCAGACCTTCAAGACAAGGT 3'
6	5' ATGGTACCTGATGTT 3'

B.

Fusion protein	Primer	
GST	5' GGG <u>CCA TGG</u> CCT CCC CTA TAC TAG GTT ATT 3'	Primer 1
	5' TTT <u>CCG CGG</u> AAT CCG ATT TTG GAG GAT GGT 3'	Primer 2
Protein A	5' GGG <u>CCA TGG</u> AGC TCA AAA CCG CGG CTC TTG 3'	Primer 1
	5' TTT <u>CGG CCG</u> GGT TGA CTT CCC CGC GGA ATT 3'	Primer 2
CBP	5' GGG <u>CCA TGG</u> AAA AGA GAA GAT GGA AAA AGA 3'	Primer 1
	5' TTT <u>CCG CGG</u> AGC AGT AGT TGG AAT ATC ATA 3'	Primer 2
TRX	5' GGG <u>CCA TGG</u> CCA GCG ATA AAA TTA TTC ACC 3'	Primer 1
	5' TTT <u>CCG CGG</u> TAC CAG AAC CGG CCA GGT TAG 3'	Primer 2

Table 1

	<b>Crude cell extract (mg)<sup>a</sup></b>	<b>One-step affinity purification (mg/ml)<sup>b c</sup></b>	<b>Purification (percentage of enrichment)</b>
Flag immunoaffinity	92	2	60
CBP	108	0.4	12
Thioredoxin	101	0.7	21
Protein A	94	1.2	41
GST	90	1.4	42

Table 2

## Bibliography

- Basagoudanavar, S. H., Goswami, P. P., Tiwari, V., Pandey, A. K., and Singh, N.** Heterologous expression of a gene encoding a 35 kDa protein of *Mycobacterium avium* paratuberculosis in *Escherichia coli* 1. *Vet.Res.Commun.*28[3], 209-224. (2004).
- Bichet, P., Mollat, P., Capdevila, C., and Sarubbi, E.** Endogenous glutathione-binding proteins of insect cell lines: characterization and removal from glutathione S-transferase (GST) fusion proteins. *Protein Expr.Purif.*19[1], 197-201. (2000).
- Cadel, S., Gouzy-Darmon, C., Petres, S., Piesse, C., Pham, V. L., Beinfeld, M. C., Cohen, P., and Foulon, T.** Expression and purification of rat recombinant aminopeptidase B secreted from baculovirus-infected insect cells. *Protein Expr.Purif.*36[1], 19-30. (2004).
- Coin, F. and Egly, J. M.** Ten years of TFIIH. *Cold Spring Harb.Symp.Quant.Biol.*63, 105-110. (1998).
- Darrow, A. L., Conway, K. A., Vaidya, A. H., Rosenthal, D., Wildey, M. J., Minor, L., Itkin, Z., Kong, Y., Piesvaux, J., Qi, J., Mercken, M., Andrade-Gordon, P., Plata-Salaman, C., and Ilyin, S. E.** Virus-based expression systems facilitate rapid target in vivo functionality validation and high-throughput screening. *J.Biomol.Screen.*8[1], 65-71. (2003).
- Dolby, V., Collen, A., Lundqvist, A., and Cronet, P.** Overexpression and functional characterisation of the human melanocortin 4 receptor in Sf9 cells. *Protein Expr.Purif.*37[2], 455-461. (2004).
- Dunn, S. D.** Effects of the modification of transfer buffer composition and the renaturation of proteins in gels on the recognition of proteins on Western blots by monoclonal antibodies. *Anal.Biochem.*157[1], 144-153. (15-8-1986).
- Fribourg, S., Romier, C., Werten, S., Gangloff, Y. G., Poterszman, A., and Moras, D.** Dissecting the interaction network of multiprotein complexes by pairwise coexpression of subunits in *E. coli* 2. *J.Mol.Biol.*306[2], 363-373. (16-2-2001).
- Giglia-Mari, G., Coin, F., Ranish, J. A., Hoogstraten, D., Theil, A., Wijgers, N., Jaspers, N. G., Raams, A., Argentini, M., van der Spek, P. J., Botta, E., Stefanini, M., Egly, J. M., Aebersold, R., Hoeijmakers, J. H., and Vermeulen, W.** A new, tenth subunit of TFIIH is responsible for the DNA repair syndrome trichothiodystrophy group A. *Nat.Genet.*36[7], 714-719. (2004).
- Hewitt, L. and McDonnell, J. M.** Screening and optimizing protein production in *E. coli*. *Methods Mol.Biol.*278, 1-16. (2004).
- Jawhari, A., Uhring, M., Crucifix, C., Fribourg, S., Schultz, P., Poterszman, A., Egly, J. M., and Moras, D.** Expression of FLAG fusion proteins in insect cells: application to the multi-subunit transcription/DNA repair factor TFIIH. *Protein Expr.Purif.*24[3], 513-523. (2002).



- Joel, P. B., Fagnant, P. M., and Trybus, K. M.** Expression of a nonpolymerizable actin mutant in Sf9 cells. *Biochemistry*43[36], 11554-11559. (14-9-2004).
- Klein, W.** Calmodulin-binding peptide as a removable affinity tag for protein purification. *Methods Mol.Biol.*205, 79-97. (2003).
- Melkko, S. and Neri, D.** Calmodulin as an affinity purification tag. *Methods Mol.Biol.*205, 69-77. (2003).
- Mercado-Pimentel, M. E., Jordan, N. C., and Aisemberg, G. O.** Affinity purification of GST fusion proteins for immunohistochemical studies of gene expression. *Protein Expr.Purif.*26[2], 260-265. (2002).
- Mohanty, A. K. and Wiener, M. C.** Membrane protein expression and production: effects of polyhistidine tag length and position. *Protein Expr.Purif.*33[2], 311-325. (2004).
- Sheffield, P., Garrard, S., and Derewenda, Z.** Overcoming expression and purification problems of RhoGDI using a family of "parallel" expression vectors. *Protein Expr.Purif.*15[1], 34-39. (1999).
- Smith, D. B. and Johnson, K. S.** Single-step purification of polypeptides expressed in *Escherichia coli* as fusions with glutathione S-transferase. *Gene*67[1], 31-40. (15-8-1988).
- Smyth, D. R., Mrozkiewicz, M. K., McGrath, W. J., Listwan, P., and Kobe, B.** Crystal structures of fusion proteins with large-affinity tags. *Protein Sci.*12[7], 1313-1322. (2003).
- Terpe, K.** Overview of tag protein fusions: from molecular and biochemical fundamentals to commercial systems. *Appl.Microbiol.Biotechnol.*60[5], 523-533. (2003).
- Tirode, F., Busso, D., Coin, F., and Egly, J. M.** Reconstitution of the transcription factor TFIIH: assignment of functions for the three enzymatic subunits, XPB, XPD, and cdk7. *Mol.Cell*3[1], 87-95. (1999).
- Wang, A., Clapper, J., Guderian, J. A., Foy, T. M., Fanger, G. R., Retter, M. W., and Skeiky, Y. A.** A novel method for increasing the expression level of recombinant proteins 1. *Protein Expr.Purif.*30[1], 124-133. (2003).
- Woestenenk, E. A., Hammarstrom, M., van den, Berg S., Hard, T., and Berglund, H.** His tag effect on solubility of human proteins produced in *Escherichia coli*: a comparison between four expression vectors. *J.Struct.Funct.Genomics*5[3], 217-229. (2004).
- Wu, J. and Filutowicz, M.** Hexahistidine (His6)-tag dependent protein dimerization: a cautionary tale. *Acta Biochim.Pol.*46[3], 591-599. (1999).
- Zeghouf, M., Li, J., Butland, G., Borkowska, A., Canadien, V., Richards, D., Beattie, B., Emili, A., and Greenblatt, J. F.** Sequential Peptide Affinity (SPA) system for the identification of mammalian and bacterial protein complexes 2. *J.Proteome.Res.*3[3], 463-468. (2004).
- Zhang, L., Hernan, R., and Brizzard, B.** Multiple tandem epitope tagging for enhanced detection of protein expressed in mammalian cells. *Mol.Biotechnol.*19[3], 313-321. (2001).

## Chapitre VI : Discussion

Au cours de mes études doctorales, mon travail a porté sur le facteur de transcription/réparation TFIIH. Initialement identifié comme un facteur nécessaire à l'initiation de la transcription des gènes de classe II (c'est à dire des gènes transcrits par l'ARN polymérase II et codant pour les protéines), TFIIH est également impliqué dans la réparation de l'ADN par excision-resynthèse des nucléotides (NER : Nucleotide Excision Repair).

TFIIH est un facteur multi-protéique composé de dix sous unités possédant trois activités enzymatiques, une activité kinase principalement destinée à phosphoryler le domaine carboxy-terminal de l'ARN polymérase II, étape nécessaire à la transition d'un complexe de pré-initiation à un complexe d'initiation, et deux activités hélicases de polarité inversées XPD (5'–3') et XPB (3'–5') impliquées dans l'ouverture de l'ADN au niveau du site d'initiation (dans le cadre de l'initiation de la transcription) et au niveau des lésions de l'ADN (dans le cadre de la réparation de l'ADN) (Bradsher et al., 2000b).

Des études *in vivo* ont confirmé le rôle de TFIIH, et en particulier des sous-unités hélicases, dans la transcription et dans la réparation de l'ADN (Volker et al., 2001a). A cet égard, il a été montré chez l'homme que des mutations dans les gènes codant pour les hélicases XPB et XPD étaient à l'origine de trois désordres génétiques : le xeroderma pigmentosum, le syndrome de Cockayne et la trichothiodystrophie, dont les phénotypes sont expliqués par des déficiences spécifiques dans la réparation (voir Synthèse *Médecine et Science* en Introduction).

### **I. Les enjeux d'une étude structurale**

L'étude du facteur TFIIH a fait l'objet de nombreuses études biochimiques visant à définir son rôle exact dans les processus cellulaires que sont la transcription et la réparation de l'ADN. Une étude structurale a été entreprise afin de comprendre à l'échelle atomique ses mécanismes d'action et la façon dont ses fonctions sont régulées par les acteurs de la vie cellulaire. Il en a naturellement découlé une

---

collaboration entre les groupes du Dr Jean-Marc Egly (régulation de la transcription), du Dr Dino Moras (Génomique structurale) et du Dr Patrick Schultz (microscopie électronique). Ce travail se situe ainsi à l'interface d'une étude structurale et fonctionnelle.

Actuellement trois approches sont suivies au laboratoire pour obtenir des informations structurales sur le facteur de transcription/réparation TFIIH :

Premièrement, nous disséquons le complexe afin d'identifier et d'isoler des sous ensembles stables pouvant être sur-exprimés et faire l'objet d'une étude structurale par R.M.N. (Résonance magnétique nucléaire) ou diffraction des rayons X. Une telle approche a débouché sur l'identification de sous complexes stables et sur les déterminations de plusieurs structures 3D dans notre laboratoire (p44-cter, MAT1-Ring, p62 PH et cycline H)(Fribourg et al., 2001 ; Gervais et al., 2001, Gervais et al., 2004 ; Anderson et al., 1996).

Deuxièmement, nous cherchons à isoler des homologues des sous-unités catalytiques du complexe, les hélicases, dans le système archaebactérien afin d'en déterminer la structure par diffraction aux rayons X. Cette approche a donné lieu à l'identification d'une première hélicase, homologue de XPD, chez *Sulfolobus solfataricus*.

Troisièmement, le facteur entier est étudié par microscopie électronique. Cette technique est particulièrement bien adaptée à l'étude d'objets de masse importante, ne nécessite pas de grandes quantités de matériel et grâce aux développements technologiques récents permet d'atteindre une résolution de l'ordre de 10-15 Å.

## **II. Production de l'échantillon**

Compte tenu des difficultés liées à la production et à la purification du facteur TFIIH humain, sa cristallisation et la détermination de sa structure par diffraction des rayons X n'est pas encore envisageable. La production et la caractérisation des échantillons constituent l'étape limitante dans la plupart des études structurales et

ceci a été le cas pour ce projet. Au cours de ma thèse une partie de mon travail a consisté à produire et à caractériser des échantillons en utilisant le système d'expression baculovirus. Le schéma de purification mis en place sur le facteur TFIIH recombinant produit à partir de cellules d'insectes a permis l'obtention de complexe recombinant en quantité suffisante pour une étude structurale (publication n°4).

#### Purification et caractérisation de complexes recombinants

Lorsque j'ai débuté ce travail, les complexes TFIIH recombinants produits dans les cellules d'insectes avec le système d'expression baculovirus dont nous disposions étaient actifs mais les préparations étaient très hétérogènes et ne se prêtaient pas à des études structurales. Une grande partie de l'étude a donc consisté à améliorer le protocole de purification des complexes recombinants afin d'augmenter l'homogénéité des préparations et de permettre des études en microscopie électronique. Cet objectif a été atteint et un modèle 3D du complexe recombinant à neuf sous-unités sensiblement identique au complexe endogène a pu être obtenu (publication n°4).

Une autre application de la technologie recombinante avec expression de protéines fusionnées à un peptide FLAG a été la production du complexe core-TFIIH humain actif en transcription et en réparation *in vitro*. Nous avons déterminé la structure de ce sous-complexe de TFIIH purifié à partir de cellules d'insectes.

La comparaison du modèle 3D du core-TFIIH humain avec le core TFIIH de levure (figure 48) (Chang et al., 2000) soulève plusieurs observations : (i) le modèle de levure a été obtenu à partir de cristaux 2D alors que le modèle humain a été calculé sur des molécules isolées après coloration négative, la résolution des deux modèles n'est par conséquent pas la même (environ 10Å d'écart), (ii) l'hélicase Ssl2/XPD n'est pas présente dans le core-TFIIH de levure, elle n'est pas essentielle aux activités de réparation et de transcription *in vitro*, (iii) les 4 sous-unités Tfb1/p62, Tfb2/p52, Tfb4/p34 et Ssl1/p44 de levure forment une structure annulaire relativement similaire à l'homme exceptée l'hélicase Rad3/XPB, se trouvant en dehors de l'anneau.

D'ores et déjà il s'agit de poursuivre l'analyse du core-TFIIH en cryomicroscopie afin d'améliorer la résolution et d'espérer localiser les sous-unités hélicases XPD et XPB.

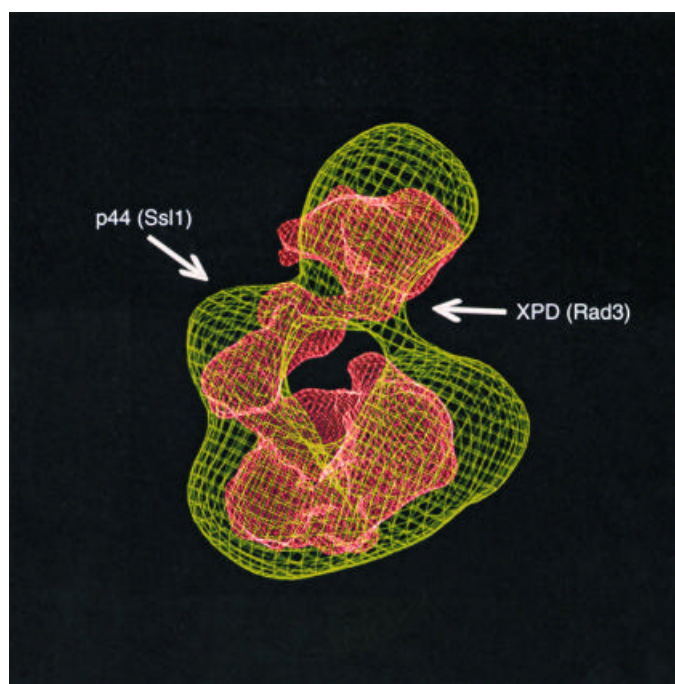


Figure 48 : Comparaison des structures du core-TFIIH de levure (rouge) au facteur TFIIH humain (jaune)

A la lumière de nos résultats, il semble que des efforts sur la production de complexes recombinants à des concentrations très élevées et homogènes restent encore à fournir.

Premièrement, pour l'expression du facteur recombinant, les cellules *Sf9* sont co-infectées avec 9 virus distincts, chaque virus codant pour une sous-unité de TFIIH. Cette approche est très flexible et parfaitement adaptée à des études d'interaction (ou fonctionnelles) dans la mesure où, pour introduire une mutation, il suffit de remplacer le virus exprimant la protéine sauvage par un virus exprimant la protéine mutée. En revanche, pour des raisons statistiques, le nombre de cellules co-infectées par les 9 virus, et donc co-exprimant effectivement les 9 sous-unités est relativement faible. Ceci affecte évidemment les taux d'expression ainsi que

l'homogénéité des complexes produits. Pour palier à ces problèmes, liées à l'utilisation de 9 virus différents, un ensemble de 3 virus exprimant simultanément chacun 3 sous-unités de TFIIH ont été dernièrement construits, espérant ainsi augmenter les taux d'expression. De plus l'identification récente d'une dixième sous-unité TTD-A/p8 laisse suggérer que cette entité est nécessaire à l'architecture du complexe. La production d'un complexe recombinant à 10 sous-unités est à présent indispensable.

Deuxièmement, pour faciliter la purification et améliorer les rendements, différentes combinaisons de tags d'affinité (Protéine A, Thioredoxine, CBP et GST), positionnés aux extrémités N-terminales de la sous unité p34 ont été construits (publication n°5). La production de sous-complexes recombinants rIIH6, rIIH5 et rIIH3 est désormais envisageable par l'utilisation d'un tag d'affinité qui facilitera la purification. L'efficacité des fusions a été évaluée sur un complexe stable produit dans les cellules d'insectes : le complexe p34/p44. Une étude comparative des différentes fusions, en terme d'efficacité de purification et de quantité de matériel obtenu a été réalisée. Depuis, de premiers résultats encourageants ont été obtenus quant à la purification du sous-complexe rIIH3 (p34/p44/p62), les trois protéines forment un sous-complexe stable dans les cellules d'insectes. J'ai pu purifier le sous-complexe via la fusion GST sur p34 dans des quantités raisonnables (de l'ordre de 1mg). En diffusion de lumière, le complexe p34p44/p62 n'est pas agrégé, une seule population d'un poids moléculaire d'environ 140Kda a été détectée. A ce jour la quantité d'échantillon n'a pas été suffisante pour débiter la cristallogénèse. Le complexe rIIH3 reste toutefois un sérieux candidat à la cristallisation.

Enfin, l'amélioration de la qualité de l'échantillon permettra l'accès à la caractérisation biophysique de l'échantillon. Ainsi, dans le cas du facteur TFIIIE, de récentes analyses en SAXS (Diffusion aux petits angles) nous ont confirmé la forme très allongée adoptée par le complexe, ce qui explique assurément sa migration aberrante en gel filtration (publication n°3). Des mesures de l'échantillon en solution par des méthodes biophysiques (ultra-centrifugation analytique, diffusion des rayons X aux petits angles et spectrométrie de masse) permettront d'identifier des paramètres agissant sur l'homogénéité conformationnelle de l'échantillon, facilitant

ainsi la cristallisation et l'étude en microscopie électronique des complexes.

L'approche de technologie recombinante est un bon système pour la production de complexes multi-protéiques humains. En effet, les modifications post-traductionnelles sont achevées dans les cellules d'insectes et l'obtention de matériel pur en quantité souhaitable pour la cristallisation est à tout fait réalisable. Il s'agit de coordonner les efforts de production de cellules en masse (utilisation de fermenteur par exemple) à la mise en place de protocoles efficaces, sans multiplier les étapes de purification. De plus, le système d'expression baculovirus présente une certaine flexibilité qui nous permet d'introduire des points de mutations (identiques à celles des malades), des protéines tronquées, des protéines de fusion et d'en apprécier les effets sur la structure et la fonction.

#### Apport de la spectrométrie de masse à la caractérisation de l'échantillon

L'apport de la spectrométrie pour la compréhension du comportement de TFIIH est essentielle. En particulier, l'étude des états de phosphorylations de TFIIH est un volet à exploiter. En effet, certains récepteurs de l'acide rétinoïque (RAR) (Bastien et al., 2000b ; Ito et al., 2004) sont capables d'interagir physiquement avec le complexe TFIIH et de les phosphoryler via sa kinase cdk7. Dans les cellules provenant de patients présentant certaines mutations de la sous-unité XPD l'activité transactivatrice des récepteurs RAR est altérée (Keriel et al., 2002a). Parallèlement à l'altération de l'activité transactivatrice de ces récepteurs, leur phosphorylation par TFIIH est fortement perturbée par ces mutations. L'utilisation du système d'expression baculovirus (dans lequel les éléments de phosphorylation sont conservés) permettra la production d'un facteur recombinant portant ces mêmes mutations afin de caractériser les états de phosphorylation et d'étudier l'influence de la phosphorylation sur les activités de TFIIH. La connaissance du pattern de phosphorylations de TFIIH servira à mieux appréhender l'étude structurale à venir.

### Obtention de lignées stables

Conjointement à l'étude du facteur recombinant, des efforts de production d'un échantillon endogène sont de mise. Il s'agit d'exploiter l'utilisation de lignées stables de cellules humaines dont l'une des sous unités portera un tag d'affinité pour la production du complexe TFIIH endogène. Dans ce sens, le protocole du « Tap-Tag » est actuellement très utilisé dans la purification de complexes multi-protéiques (Puig et al., 2001; Graumann et al., 2004). L'utilisation de souches tap-taguées permettra l'obtention de facteur en très peu d'étapes de purification (2 au maximum). La construction de telles souches est en cours au laboratoire.

### **III. Etude moléculaire du complexe TFIIH par microscopie électronique**

Un autre volet de mon travail a été l'étude du complexe dans son ensemble par microscopie électronique et analyse d'images. Cette étude a préalablement débuté sur le complexe TFIIH endogène (purifié dans le groupe du Dr Egly) et sur un sous-complexe recombinant et a donné les premières images à basse résolution (de l'ordre de 38 Å) de ce facteur (Schultz et al., 2000d). Ceci constitue la première description structurale du facteur TFIIH entier mais la résolution est insuffisante pour raisonner en termes de domaines structuraux et pour localiser, en particulier, les sites catalytiques des deux hélicases et l'entité kinase constituée du complexe cdk7/cycline H.

### Cryomicroscopie sur le facteur endogène

L'objectif majeur de l'analyse du facteur TFIIH endogène par microscopie électronique a été l'amélioration de la résolution de la structure 3D. Pour ce faire nous avons utilisé les récentes améliorations méthodologiques et technologiques que sont la cryomicroscopie combinée à l'analyse d'images sur un microscope à canon à effet de champ.

La cryomicroscopie électronique consiste à observer l'échantillon préservé sous forme hydratée-congelée dans une mince couche d'eau vitreuse. La molécule est



observée dans un état proche de son environnement natif et sa structure est conservée à l'échelle atomique. Les images obtenues sont ensuite analysées et cette opération permet d'accéder à l'information volumétrique. Ainsi, l'analyse statistique multivariée est utilisée pour identifier différentes sous-populations. L'utilisation d'un microscope équipé d'un canon à émission de champ (FEG) avec un faisceau plus cohérent permet de produire des images avec un meilleur transfert d'information.

Nos premiers résultats en cryomicroscopie électronique sur le facteur endogène montrent une structure 3D assez similaire à celle obtenue en coloration négative avec une légère amélioration de la résolution. La résolution escomptée n'a pas encore été atteinte mais nous pouvons tout de même observer quelques nouveaux détails structuraux correspondants à des masses protéiques.

L'analyse d'image a également montré que TFIIH pouvait adopter différentes conformations et notamment qu'il pouvait partiellement se dérouler. Il est possible que ce phénomène dynamique soit d'une importance primordiale dans le mécanisme d'action du facteur lorsque par exemple, les deux hélicases XPB et XPD fusionnent l'ADN autour de la lésion sur plusieurs dizaines de paires de bases. Il s'agit dès lors de décrire le mouvement en solution des différents domaines de TFIIH à partir de l'analyse des images de cryomicroscopie électronique. L'analyse de variance conduira à une description détaillée des changements conformationnels du complexe pour peu que deux états métastables soient identifiés.

Nos observations microscopiques ont montré que, malgré l'utilisation de méthodes nouvelles, l'hétérogénéité de l'échantillon et son comportement intrinsèque sont un frein à l'amélioration de la résolution. L'enjeu des années à venir sera d'identifier différents états au sein d'une population mixte de molécules figées par congélation à un instant  $t$  et de mettre en évidence des états transitoires. L'enregistrement direct de paramètres propres à chaque molécule par microscopie électronique suivie d'une analyse statistique rigoureuse devrait permettre l'identification de telles sous-populations.

#### IV. Organisation quaternaire de TFIIH

Un aspect fort intéressant de ce travail a été de corrélérer la structure du facteur TFIIH endogène obtenu en microscopie électronique avec les données biochimiques dont nous disposons au laboratoire. L'analyse systématique des interactions protéine-protéine au sein du complexe a dégagé plusieurs hypothèses : (i) l'existence d'un complexe minimal p34/p44/p62 assimilé au cœur structural mais dont la fonction reste inconnue (ii) la mise en évidence d'un réseau d'interaction qui s'articule autour de ce cœur structural par des interactions structurées avec les protéines régulatrices de TFIIH, les hélicases XPB et XPD ainsi que le complexe trimérique CAK, pour former le facteur entier (figure 49). On peut se demander quels sont les domaines responsables dans l'assemblage du complexe. Il faut à ce propos noter la présence de domaines RING finger dans le facteur TFIIH. Ces domaines sont souvent impliqués dans des interactions protéine-protéine et leur intégrité est intimement liée à l'assemblage de certains complexes macromoléculaires (pour une revue (Matthews et al., 2002)).

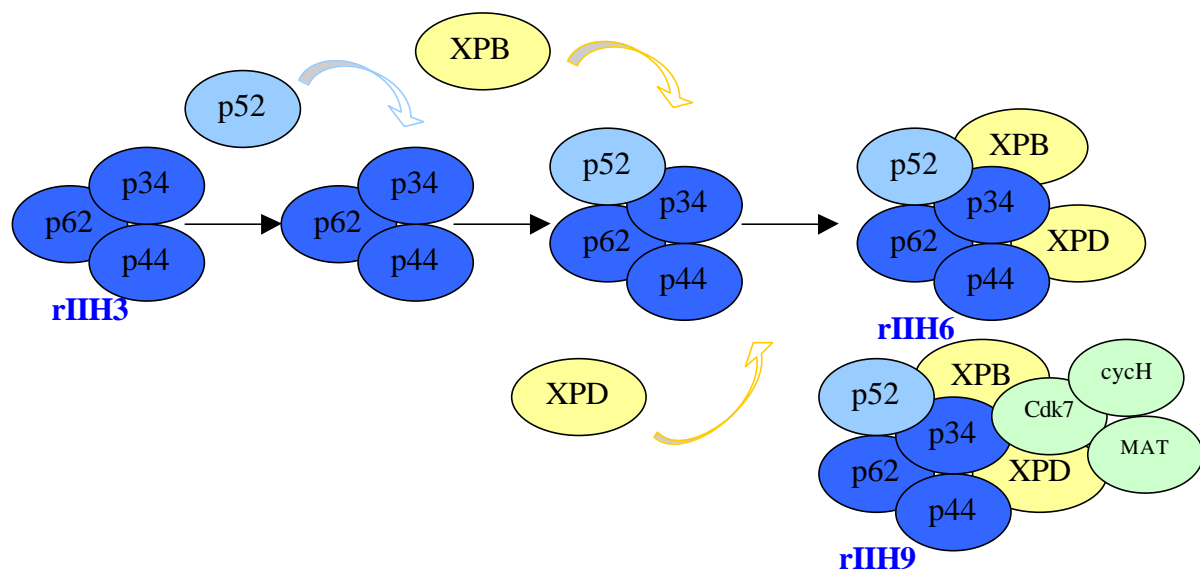


Figure 49 : Modèle d'assemblage du complexe TFIIH

Au cours de cette thèse nous avons particulièrement analysé les interactions de l'hélicase XPD avec ses partenaires. Un des partenaires connu de l'hélicase XPD est la sous-unité régulatrice p44 (Coin et al., 2000). En plus de former un complexe stable, l'interaction entre XPD et p44 est fondamentale puisqu'une mutation dans la partie C-terminale de XPD empêche toute interaction avec p44 et ceci est à l'origine de maladies.

Dans ce sens, plusieurs questions se posent : en quoi des mutations dans XPD fragilisent-elles l'architecture de TFIIH ainsi que son intégration dans un complexe d'activation de la transcription ? Se pourrait-il que des modifications dans les domaines de contact XPD/p44 perturbent ou remodelent certains domaines d'autres sous unités de TFIIH empêchant ce dernier de les contacter ? L'hélicase XPD joue un rôle majeur dans la cohérence de l'architecture du complexe par ces interactions régulatrices. Un autre partenaire de XPD, cette fois du côté N-terminal, est la sous-unité MAT1. De plus de ce même côté, une interaction entre XPD et la protéine p8 est supputée (résultat non publié, groupe du Dr Egly). Quelle est le rôle de cette interaction, sachant qu'une mutation dans la sous-unité p8 est à l'origine de la forme A de trichothiodystrophie (TTD-A) ? La protéine p62 interagit également avec XPD mais le rôle de cette interaction n'est pas connue. A telle une fonction dans la stabilisation du complexe ? Permet-elle de réguler l'activité de l'hélicase ?

L'hélicase apparaît indispensable au maintien et à l'intégrité du complexe entier, c'est une charnière entre le complexe core-TFIIH (via p44 et p62) et le complexe CAK (via MAT1). Des analyses similaires ont été effectuées au laboratoire sur les autres sous-unités et ont permis l'obtention des différentes interactions protéine-protéine au sein du complexe. Cette vision globale du réseau d'interaction en corrélation avec la structure du facteur endogène obtenue en microscopie électronique a donné lieu à un modèle général de l'architecture moléculaire de TFIIH. Etant donné que la résolution de structure 3D est limitée à 3.8nm, les sous-unités sont positionnées de manière grossière, à nouveau un gain en terme de résolution nous permettra certainement de corroborer ce premier modèle moléculaire. De même, l'immunomarquage du complexe avec les anti-corps p62, p52, p8 et les domaines N-terminales des hélicases est nécessaire pour valider le modèle. Ainsi il deviendra possible de combiner les structures atomiques obtenues par cristallographie à partir de fragments

ou de sous-unités avec le modèle à moyenne résolution du complexe TFIIH entier déterminé par cryomicroscopie électronique.

A titre d'exemple, la comparaison de l'enveloppe de l'ARN polymérase I obtenue en cryomicroscopie avec la structure atomique de l'ARN polymérase II, une enzyme homologue dépourvue de quatre sous-unités, a montré la grande homologie structurale entre les deux modèles. Cette comparaison a permis de localiser très précisément, par carte de différence, les quatre sous-unités additionnelles présentes dans l'enzyme de classe I. Ces résultats illustrent l'intérêt de comparer les structures déterminées par deux méthodes complémentaires pour mettre en évidence avec précision les domaines protéiques en interaction.

## V. TFIIH en complexes avec ses partenaires physiologiques

### Structure 3D du facteur de transcription TFIIIE

L'analyse d'images de molécules individuelles du facteur de transcription TFIIIE, qui interagit physiquement avec TFIIH, a été réalisée en coloration négative à froid à une résolution de 16Å (publication n°3). TFIIIE est un hétérodimère contenant une seule copie de chaque sous-unité (TFIIIE $\alpha$  et TFIIIE $\beta$ ). Disposant des facteurs TFIIIE et TFIIH, la poursuite de ce travail consistera naturellement à étudier l'interaction TFIIIE/TFIIH dans le but de déterminer la structure du complexe. Afin d'établir une relation entre les activités de TFIIH et de TFIIIE, il serait intéressant de comprendre l'organisation moléculaire du complexe de transcription au niveau du promoteur et notamment le positionnement des facteurs au sein de ce complexe. En conjonction avec les expériences de photocrosslinking (Forget et al., 2004; Langelier et al., 2001), l'analyse des points de contact des sous-unités du facteur TFIIIE a permis de superposer la structure du dimère obtenue par microscopie électronique sur celle de la polymérase (figure 50).

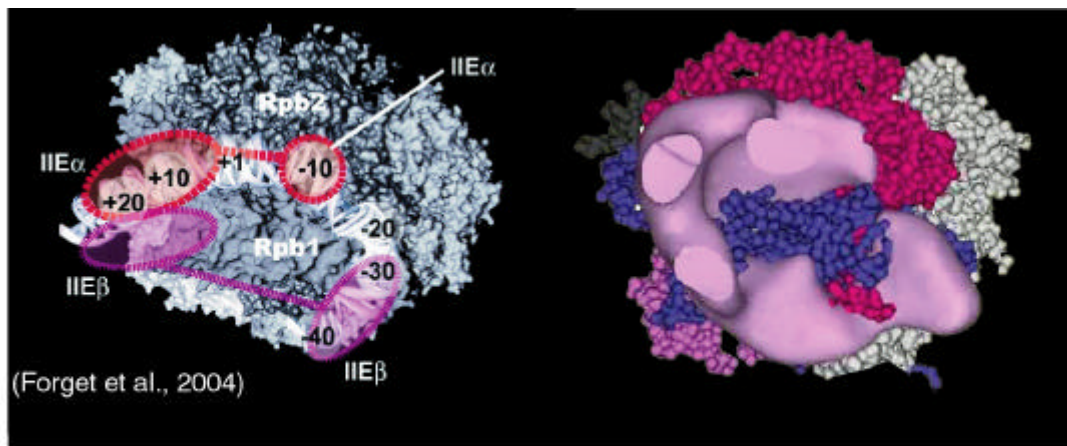


Figure 50 : Localisation du facteur TFIIIE sur la polymérase (A) Point de contact obtenu après photocrosslinking (Forget et al., 2004) (B) Superposition du modèle 3D de TFIIIE sur la polymérase.

#### TFIIH en liaison avec l'ADN

TFIIH est un facteur de transcription impliqué dans l'ouverture de l'ADN possédant des activités enzymatiques ATP-dépendantes et il se pose la question de savoir comment est reconnu l'acide nucléique. L'étude structurale de TFIIH en liaison avec l'ADN n'est pas aisée du fait que ces complexes soient transitoires dans la cellule. Une nouvelle technologie basée sur l'immobilisation de l'ADN sur un support adéquat et permettant l'observation ultérieure du complexe en microscopie a été développée (publication n°2). Cette technologie a été utilisée pour de premières observations du facteur endogène lié un oligonucléotide de 30 mers. Les particules de TFIIH apparaissent ouvertes et ces premières observations soulignent à nouveau l'importance d'étudier TFIIH dans environnement modulé comme au contact de l'ADN.

Une compréhension globale des mécanismes de régulation cellulaire passe par l'étude du comportement du facteur TFIIH isolé lié à l'ADN. En présence d'un petit fragment d'ADN et à une résolution d'environ 15Å tout changement conformationnel sera visualisable et la localisation des domaines structuraux comme les sites actifs

---

des hélicases et/ou du complexe cdk7/cycline H sera envisageable. L'étude du complexe TFIIH lié à de l'ADN impliquera la corrélation de sa structure à sa fonction pour élucider son mécanisme d'action.

Les deux processus cellulaires dans lesquels TFIIH est impliqué (transcription et réparation) font intervenir une cascade de protéines autour de l'ADN. Les acteurs interagissant avec TFIIH lors de l'initiation de la transcription sont disponibles mais la réaction est fortement complexe. La détermination de la structure de ce complexe de préinitiation à l'aide d'un support sur lequel est immobilisé l'ADN contenant la région promotrice n'est envisageable qu'avec un effort biochimique au préalable. La possibilité d'arrêter la transcription à un moment donné, lors de l'initiation par exemple, et de purifier le complexe nous permettrait l'accès à des informations fondamentales sur l'architecture du complexe et la localisation des différents facteurs. L'observation de tels complexes par cryomicroscopie va permettre de définir le trajet suivi par l'ADN au sein du complexe de transcription.

De même, la reconnaissance de la lésion lors du mécanisme de réparation de l'ADN fait intervenir le facteur de TFIIH. L'observation des premières étapes de cette cascade moléculaire paraît plus réalisable. Le complexe binaire XPC/HHR23B reconnaît différents types de lésions sur la molécule d'ADN (photoproduits induits par l'irradiation par les UV, ADN cis-platiné, modifications introduites par les carcinogènes). Le complexe XPC/HHR23B peut être produit en grande quantité et il a été montré qu'il se liait de façon efficace aux lésions *in vitro* (Yokoi et al., 2000). Ainsi, dans une deuxième étape, le complexe XPC/HHR23B recrute TFIIH sur la lésion afin, semble-t-il, de propager l'ouverture simple brin de part et d'autre. Des observations microscopiques avec différents substrats (double brin avec une bulle désappariée, double brin avec cis-platine) peuvent être utilisés et permettront de visualiser ces différents intermédiaires structuraux afin de mieux comprendre le mécanisme de reconnaissance du dommage.

---

## VI. Génomique structurale des hélicases

Le système de transcription des archaebactéries, incluant *Sulfolobus solfataricus*, est très proche du système eucaryote. Le complexe TFIIH n'existe pas sous forme de complexe multi protéique à dix sous-unités chez ces organismes, mais les homologues des hélicases humaines XPD et XPB sont présents. Au cours de cette thèse, j'ai pu cloner, purifier et cristalliser l'homologue de l'hélicase XPD chez *Sulfolobus solfataricus*. Malgré l'obtention de cristaux de l'hélicase sous sa forme native et sous forme complexée à un substrat, aucun des cristaux testés au synchrotron ne diffractent au delà de 5Å.

Dès lors, les nouveaux outils de la génomique structurale permettront d'ouvrir ce projet à d'autres organismes. Des hélicases orthologues aux sous-unités XPB et XPD de TFIIH sont également présentes chez d'autres nombreuses archaebactéries et bactéries pathogènes. On peut raisonnablement évaluer à plus d'une centaine, le nombre de cibles définies dans le cadre de cette étude. Forts de l'expérience et de l'infrastructure dont disposent le laboratoire pour aller du gène à sa structure dans un contexte d'étude à haut débit, l'utilisation de plates-formes technologiques de clonage, d'expression, de purification et de cristallisation seront allouées à ce projet et permettront l'identification de cibles pouvant faire l'objet d'études physico-chimiques et structurales. Ces études permettront non seulement d'aborder l'aspect fonctionnel et les bases structurales de l'organisation de TFIIH mais replaceront également TFIIH dans un contexte d'évolution.

## Chapitre VII : Matériel et Méthodes

### A. BIOLOGIE MOLECULAIRE

#### 1. Matériel

##### 1.1. Souches bactériennes

- **TOP10** : F<sup>-</sup> *mrcAD* (*mrr-hsdRMS-mcrBC*) *f80 lac ZDM15 DlacX74 deoR recA1 araD139 D(ara-leu)7697 galU galK rpsL endA1 nupG*. Cette souche a été utilisée dans les étapes de sous clonage avec le Kit PCR Blunt (Stratagene ).
- **XL1-Blue** : F' *Tn10 proA+B+ lacI<sup>q</sup> D(lacZ)M15/recA1endA1gyrAA96(Nat) thi hsdR17 (r<sub>k</sub>-m<sub>k+</sub>) supE44 relA1 lac*. Cette souche a été utilisée dans les étapes de clonages et d'amplification des vecteurs.
- **BL21(DE3)** : F<sup>-</sup> *ompT hsdSB (r<sub>B</sub>-m<sub>B</sub>-) gal dcm (lclts857, ind1, Sam7 nin5 lac UV5-T7 gene 1)*. Cette souche a été utilisée lors des étapes de production.
- **BL21(DE3)-pLysRARE** : Cette souche est dérivée de la souche BL21 (DE3). Elle est utilisée pour améliorer l'expression de protéines eucaryotes dans *Escherichia coli*. Elle contient des codons rarement utilisés par *Escherichia coli* : AGG, AGA, AUA, CUA, CCC, GGA et un plasmide portant un gène de résistance chloramphénicol.
- **BL21(DE3)-RP** : Cette souche est dérivée de la souche BL21 (DE3). Elle est utilisée pour améliorer l'expression de protéines eucaryotes dans *Escherichia coli*. Elle contient les ARNt argU et proL pour rétablir l'expression des gènes eucaryotes nécessitant les codons AGG/AGA ou CCC.
- **BL21-CodonPlus-RIL** : Cette souche est dérivée de la souche BL21 (DE3). Elle est utilisée pour améliorer l'expression de protéines eucaryotes dans *Escherichia coli*. Elle contient les ARNt argU, ileY, et leuW. La présence de ces ARNt additionnels résout le biais de rencontré dans les organismes qui sont riches en codons AT.



## 1.2. Vecteurs de clonage et d'expression

- les vecteurs plasmidiques pET15b (amp<sup>R</sup>) et pET11b (amp<sup>R</sup>), promoteur T7 (Novagen )
- le vecteur plasmidique pCRBlunt (Kan<sup>R</sup>), promoteur lac (Stratagen )
- le vecteur plasmidique pSK277 (amp<sup>R</sup>), promoteur (Novagen )
- le vecteur plasmidique pVL13092, promoteur polyhédrine (amp<sup>R</sup>) ( Pharmingen )

Les vecteurs d'expression de type pET et pSKB3 permettent la production de protéines fusionnées à un polypeptide hexahistidine à leur extrémité N-terminale. Ce polypeptide peut être éliminé par protéolyse spécifique en utilisant la thrombine dans le cas du pET et la TEV (tobacco etch virus) dans le cas du pSKB3.

Les vecteurs d'expression de type pSK277 permettent la production de protéines fusionnées à un polypeptide FLAG. L'épitope FLAG consiste en un peptide de 8 amino-acides hautement hydrophile de séquence DYKDDDDK facilement détectable par les anti-corps anti-FLAG<sup>®</sup>.

Le vecteur pVL13092 est un vecteur de transfert pour Baculovirus.

## 2. Méthodes

### 2.1. Amplification par PCR « Polymerase Chain Reaction »

Cette méthode enzymatique très utilisée est très efficace pour l'amplification de fragments d'ADN *in vitro*. Elle est basée sur la répétition de trois phases :

1. **dénaturation** : le tube est chauffé quelques secondes à 94°C. Les double-brins d'ADN se séparent et l'ADN est dénaturé.
2. **hybridation** : la température est rapidement abaissée à 55°C. Les amorces "reconnaissent" leur séquence complémentaire sur les brins d'ADN cibles. Elles s'hybrident chacune sur son brin respectif. Cette étape dure une minute.

3. **élongation** : La température du tube est ensuite augmentée à 72°C, ce qui permet à la Taq Polymérase d'ajouter des nucléotides aux amorces hybridées, dans le sens 5' vers 3'. Les nucléotides ne sont pas incorporés de façon aléatoire mais en fonction de la séquence cible (nucléotide complémentaire). Cette étape dure une minute.

La réaction de PCR est préparée comme suit :

Le mélange des 4 dNTPs contient les dATP, dGTP, dCTP et dTTP. L'enzyme Deep-vent Polymerase (New England Biolabs®) est ajoutée en dernier. De l'huile de paraffine permet de recouvrir la surface du mélange réactionnel et de limiter l'évaporation.

Composition du mélange réactionnel de PCR

Composant	Volume	Concentration ou quantité
Tampon Vent-polymerase 10x	10 µl	1x
Matrice (100 ng/µl)	1 µl	100 ng
Mélange des 4 dNTP à 2.5 mMol/L	12 µl	300 µmol/L
Oligonucléotide amorce		
Amorce 1	5 µl	100 pmoles
Amorce 2	5 µl	100 pmoles
Enzyme Deep-Vent polymerase (2unités/µl)	1 µl	2 unités
Eau millipore stérile	66 µl	-

Les paramètres généralement utilisés sont :

Cycles de PCR

Etape	Temps	Température	Cycles
Incubation	2 minutes	94°C	1
Dénaturation	1 minute	94°C	1
Hybridation	1 minute	56°C	<b>25</b>
Extension	1 minute	72°C	1
Extension finale	10 minutes	72°C	1

A l'issue de la réaction un aliquote de la réaction est analysé sur gel d'agarose 1% (m/v) en tampon TAE.

## 2.2. Mutagenèse dirigée

La mutagenèse dirigée par PCR se réalise en deux étapes : une première série de réactions est réalisée sur le fragment à modifier avec un jeu d'oligonucléotides complémentaires, comprenant le ou les résidus à modifier, et un second jeu délimitant la région à amplifier. Les oligonucléotides de mutagenèse sont des 30 à 40 mers avec un minimum de 10 bases s'appariant strictement à la matrice de part et d'autre de la mutation.

Les produits de la première PCR servent de matrice à une seconde réaction ne comprenant cette fois-ci que les oligonucléotides externes. Cette technique est connue sous le nom de la méthode d' « Overlap extension » (figure 51).

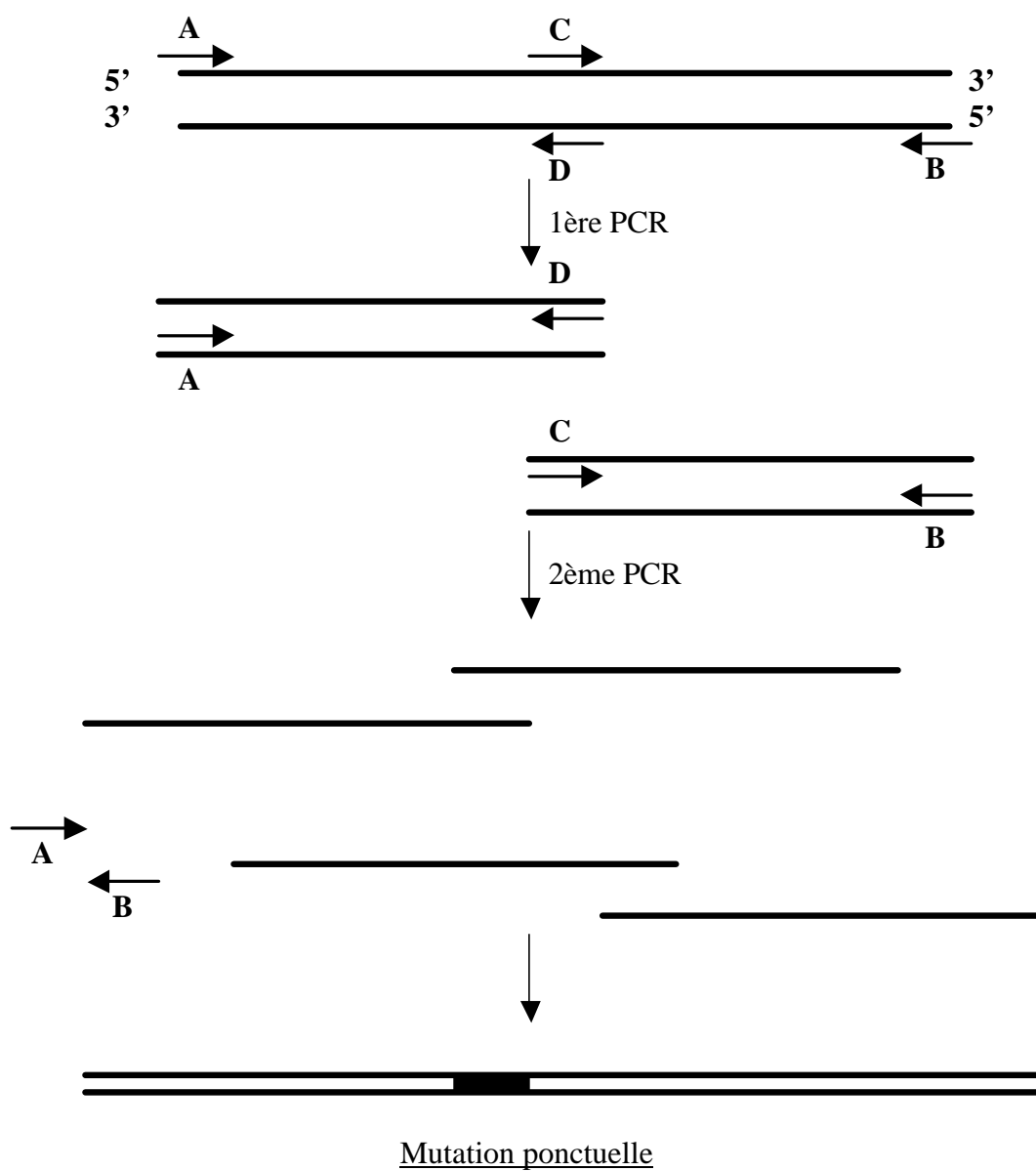


Figure 51 : Méthode d'«Overlap extension »

### 2.3. Electrophorèse en gel d'agarose

Les gels d'agarose ont été les plus couramment utilisés lors des différents étapes de clonage, que ce soit lors de la purification d'ADN, de dépôts de produits de digestion d'ADN ou tout autre manipulation nécessitant la visualisation de l'ADN.

#### Composition du tampon de dépôt

Composition	Quantité
Tris pH=8	10 mMol/L
EDTA pH=8	1 mMol/L
Bleu de bromophénol	0.25% m/v
Bleu de xylène cyanol	0.25% m/v
glycérol	40% v/v

#### Composition du tampon TAE 50x

Composition	Quantité
Tris base	242 g
Acide acétique glacial	57.1 ml
EDTA 0.5 mol/L pH=8	100 ml
Eau millipore	qsp 1 l

L'ADN est repéré par fluorescence aux UV après avoir été mis en contact avec du bromure d'éthidium (BET stock à 10 mg/ml). La masse adéquate d'agarose, typiquement de 0.5 à 1.5% m/v, selon la taille de l'ADN à visualiser est dissoute par chauffage dans du tampon TAE. Le BET est ajouté à une concentration finale de 10 µg/ml. Le gel solidifie par refroidissement.

## 2.4. Estimation de la quantité d'ADN

Une lecture de l'absorbance à 260 nm permet de déterminer la pureté de l'ADN, en effet le rapport de densités optiques (DO) 260 nm/280 nm est normalement voisin d'1.8. Un rapport plus élevé indique une contamination par des ARN. Un rapport plus faible indique une contamination par les protéines (280 nm) ou du phénol (270 nm). Lorsque l'ADN est pure, la quantité d'ADN peut être appréciée d'après la lecture de la DO à 260 nm :

1 DO (à 260 nm) = 50 ng/μl pour l'ADN double brin

1 DO (à 260 nm) = 40 ng/μl pour l'ADN simple brin

## 2.5. Précipitation des ADN

La précipitation des ADN se fait par l'acétate de sodium (stock de sodium acétate 3.3 mol/L pH=5 ) en présence d'éthanol froid à -20°C. Pour un volume de solution d'ADN à précipiter la procédure est la suivante :

- ajout de 0.1 volume d'acétate de sodium stock et de 2 volumes d'éthanol pur
- centrifugation du mélange 14000 tpm, 15 minutes à 4°C
- élimination du surnageant
- lavage du culot par 3 volumes d'éthanol 70%
- centrifugation du mélange 14000 tpm, 15 minutes à 4°C
- élimination du surnageant
- séchage du culot sous vide et resuspension dans du tampon TE ou de l'eau millipore stérile.

## B. CLONAGE

### 1. Stratégies de clonage

#### 1.1. Digestion de l'ADN par des enzymes des restriction

Les plasmides ou les fragments de PCR sont digérés par les enzymes de restriction en les incubant de 2 à 4 heures en présence de 2 à 10 unités d'enzyme par  $\mu\text{g}$  d'ADN à la température optimale pour l'enzyme et dans le tampon adéquat (proposés par le fournisseur) avec ou sans présence de Bovine Sérum Albumine (BSA stock à 10 mg/ml). Le résultat de la digestion est contrôlé sur électrophorèse en gel d'agarose.

#### 1.2. Déphosphorylation des extrémités 5' phosphate du vecteur

Les phosphatases dites alcalines, car actives à pH alcalin, catalysent l'élimination d'un groupement phosphate en 5' d'une chaîne d'ADN. On utilise des phosphatases alcalines d'origine bovine (« CIP » : calf intestinal phosphatase) ou d'origine bactérienne (« BAP »).

Après linéarisation du vecteur plasmidique, ce dernier est déphosphorylé par l'enzyme CIP (Quatum Appligène) afin d'éviter une fermeture du plasmide sur lui-même : ce qui est appelé l'auto-ligation. Lors de l'insertion du fragment d'ADN double brin à étudier, celui-ci apporte ces deux groupements phosphate en 5' et participe à deux des quatre liaisons esters qui seraient nécessaires à la formation d'un ADN double brin recombinant circulaire clos.

Le réaction se fait directement sur le produit de digestion en présence de *tampon CIP* 10x (Tris-HCl 5 mM,  $\text{MgCl}_2$  1mM,  $\text{ZnCl}_2$  0,1mM, pH=9) et de 200 unités de CIP à 200U/  $\mu\text{l}$ . Le volume réactionnel est ajusté à 50  $\mu\text{l}$  et la réaction est conduite à 37°C pendant 1 heure. Il est extrêmement important d'inactiver efficacement l'enzyme pour que celle-ci n'interfère pas avec la suite des étapes de clonage. En effet les traces de CIP peuvent déphosphoryler le fragment PCR et empêcher toute ligation. L'inactivation se fait en deux temps.

- inactivation par chauffage à 65°C pendant 30 minutes
- extraction au phénol-chloroforme

### **1.3. Extraction au phénol-chloroforme**

Cette opération inactive toutes les enzymes résiduelles et élimine les protéines qui précipitent à l'interphase. Un volume de phénol:chloroforme (1:1) est ajouté au volume à extraire. Après une centrifugation de 2 minutes à 12000 tpm, la phase supérieure est prélevée et un volume de chloroforme : isoamylalcool (24:1) est ajouté afin d'éliminer toutes traces de phénol. Une nouvelle centrifugation de 2 minutes à 12000 tpm permet d'extraire la phase supérieure et l'ADN est précipité par l'ajout de 0,1 volume de NaCl/glycogène (1,5M NaCl/ 0,5 mg/ml glycogène) et de 2 volumes d'éthanol 95% froid. Le tout est placé à -80°C pendant 20 minutes puis centrifugé à 4°C, 12000 tpm pendant 10 minutes. Le culot est lavé, séché et repris dans le volume d'H<sub>2</sub>O adéquat.

### **1.4. Purification et extraction de fragment PCR**

Les fragments PCR sont extraits et purifiés sur gel d'agarose.

L'ensemble de la réaction PCR est déposé sur gel d'agarose 1%. Après migration, la bande de gel contenant l'insert à purifier est découpée sous lumière UV avec une lame de scalpel stérile. Le fragment PCR est ensuite purifié par le système de purification de fragment PCR sur colonne (Nucleospin Column Biorad®).

Après élution de l'ADN dans du TE ou de l'eau millipore stérile, la fragment est traité comme suit :

- précipitation à l'acétate de sodium et à l'éthanol
- extraction phénol-chloroforme
- digestion avec les enzymes correspondantes
- extraction phénol-chloroforme
- resuspension dans de l'eau millipore stérile



Le fragment ainsi préparé est utilisable pour sa ligation dans le vecteur de choix préalablement digéré et déphosphorylé.

### 1.5. Ligation

Cette étape est la plus délicate du clonage. L'insertion du fragment PCR préalablement digéré et purifié dans un vecteur préalablement déphosphorylé est catalysée par la ligase du phage T4 ; elle requiert la présence d'ATP et s'effectue dans les conditions suivantes :

Les quantités d'ADN d'à la fois l'insert et le vecteur sont au préalable appréciées sur un gel d'agarose.

#### Condition de ligation

Composants	Quantité ou volume
Vecteur linéarisé et déphosphorylé	25 à 250 ng
Fragment PCR linéarisé et digéré	200 à 750 ng
Tampon ligase 10x	1 µl
Ligase T4 ADN	1 unité
Eau millipore stérile	qsq 10 µl

La réaction est incubée 12 heures à 16°C. La stœchiométrie relative entre fragment à insérer et vecteur est déterminante ainsi que la pureté et la qualité du fragment digéré purifié et du vecteur linéarisé déphosphorylé ; on teste donc systématiquement plusieurs rapports stœchiométriques entre insert à liguer et vecteur plasmidique linéarisé. L'absence de phosphatase ou de toute enzyme de restriction est absolument nécessaire à la réussite de la ligation. Les traces de phosphatases seraient susceptibles de déphosphoryler l'insert ce qui empêcherait toute ligation. Des traces d'enzymes de restriction pourraient digérer les vecteurs qui auraient intégré l'insert au cours de la ligation.

Parallèlement, on réalise trois témoins de ligation :

- un témoin avec le vecteur linéarisé mais en absence de ligase
- un témoin avec l'insert et le vecteur toujours en absence de ligase
- un témoin contenant uniquement du vecteur linéarisé et de la ligase

### **1.6. Transformation à partir du mélange de ligation**

Le mélange peut être utilisé tel quel pour transformer directement les bactéries soit par méthode chimique ou par électroporation. L'électroporation est plus efficace mais requiert soit de diluer le mélange avant électroporation soit d'effectuer une étape supplémentaire de précipitation du mélange de ligation pour enlever toutes traces de sels.

Après transformation des aliquotes sont étalés sur milieu solide avec l'antibiotique adéquat.

### **1.7. Obtention du vecteur de clonage et vérification de la séquence**

Quelques clones issus de la transformation sont prélevés à l'aide de pics stériles directement sur le milieu solide et amplifiés en présence de l'antibiotique adéquat, la purification est le plus souvent faite par minipréparations d'ADN plasmidique à l'aide du kit Biorad® par élution sur de petites colonne de type Nucleospin®.

La présence de l'insert de bonne taille est vérifié soit par digestion avec les enzymes de restriction adéquates et électrophorèse sur gel d'agarose soit par PCR en utilisant pour amorces les deux oligonucléotides amorces de clonage par PCR.

Afin de vérifier l'efficacité de l'amplification à l'étape de PCR on contrôle la séquence de la zone amplifiée par séquençage des deux brins afin de s'assurer que la polymérase n'a pas introduit de mutation. Le séquençage a été effectué au service commun de séquençage de l'IGBMC dirigé par Serge Vicaire. Une fois la séquence vérifiée, la plasmide est amplifié soit par élution sur colonne soit par double gradient de chlorure de Césium ; cette préparation permettant l'obtention d'un matériel beaucoup plus pure, elle est utilisée dans le cas de production de baculovirus recombinant.

## 2. Amplification des vecteurs et préparation d'ADN plasmidique

### 2.1. Maxipréparation d'ADN plasmidique sur colonne Nucleobond<sup>®</sup>

Les vecteurs sont amplifiés dans la souche XL1-Blue cultivée en milieu LB en présence de l'antibiotique adéquat. Environ 250 ml de suspension bactérienne est centrifugée 15 minutes à 6500 tpm. Le culot bactérien est remis en suspension dans 10 ml de *solution de resuspension* (Tris-HCl 50mM pH=8, EDTA 10mM), les cellules sont resuspendues jusqu'à obtention d'une suspension homogène. La lyse se fait par ajout de 10 ml de *solution de lyse* (NaOH 200mM, SDS 1%) pendant exactement 5 minutes à température ambiante. Puis 10 ml de *solution de neutralisation* (acétate de potassium 3,1M pH=5,5) est ajoutée au mélange et incubé 5 minutes à température ambiante. Le lysat est alors centrifugé pendant 40 minutes à 12000tpm. La colonne a été préalablement équilibrée avec 40 ml de *solution d'équilibrage* (NaCl 600mM, acétate de sodium 100mM pH=5, Triton 15%). Le surnageant contenant l'ADN plasmidique est chargé au sommet de la colonne et la solution y passe par gravité. La colonne est ensuite lavée par deux fois avec 30 ml de *solution de lavage* ( NaCl 800mM, acétate de sodium 100mM pH=5). Le système de purification en kit sur colonne Nucleobond (Macherey-Nagel ) a été utilisé. L'ADN plasmidique est élué en ajoutant 15 ml de *solution d'éluion* (NaCl 1,25 M, Tris-HCl 100mM pH=8,5) et précipité par addition de 10,5 ml d'isopropanol. Le mélange est centrifugé pendant 30 minutes à 12000 tpm à 4°C et le culot d'ADN est lavé avec 5 ml d'éthanol 70% froid et centrifugé 10 minutes à 12000tpm à 4°C. La dissolution du culot se fait avec 500µl de tampon TE ou d'eau millipore stérile. Les plasmides purifiés sont conservés à -80°C.

#### Composition du tampon TE

Composition	Concentration
Tris-HCl	10 mMol/L
EDTA	1 mMol/L

## 2.2. Maxipréparation d'ADN plasmidique par double gradient de chlorure de césium

Une culture de 500 ml de LB estensemencée la veille au soir et poussée à 37°C durant la nuit. Après centrifugation à 4000 rpm, 20 minutes, 4°C le culot de cellules est repris dans 10 ml de tampon TE5x (50 mM Tris-HCl, 5 mM EDTA) et laissé 5 minutes à température ambiante. Puis 20 ml de solution de lyse (1% SDS, 0.2M NaOH, préparée extemporanément) est ajoutée et la lyse se fait précisément 5 minutes dans la glace. 15 ml de solution de neutralisation (7.5 M acétate d'ammonium, 3% acide acétique) est additionnée et laissée 5 minutes sur glace. Le tout est homogénéisé puis centrifugé à température ambiante, 15 minutes à 4000 rpm. Une solution limpide doit être obtenue après filtration sur gaz. Cette solution est complétée à 90 ml avec de l'isopropanol et laissée 10 minutes sur glace. Après centrifugation 20 minutes à 4000 rpm, 4°C, le culot d'ADN est séché et repris dans 1 ml de TE1x. L'ADN est transféré dans un tube d'ultracentrifugation Beckmann® dans lequel étaient au préalable ajoutés 200 µl de BET à 10 mg/ml et 4.4 g de CsCl. Les tubes d'ultracentrifugation sont équilibrés avec du TE1x et ultracentrifugés 4 heures à 65k, à température ambiante. Les bandes de plasmide sont prélevées sous UV à l'aide d'une seringue. L'ADN est placé dans un nouveau tube d'ultracentrifugation complété et équilibré avec du TE1x, 0.8 g/ml CsCl. Après ultracentrifugation 16 heures à 50k à température ambiante, la bande de plasmide est à nouveau prélevée. Trois extractions au butanol saturé en eau sont effectuées (le butanol constitue la phase supérieure). Environ 8 ml d'isopropanol sont ajoutés puis le tout est centrifugé 15 minutes à 5000 rpm à température ambiante. Le culot est séché puis dissout dans 100 µl de TE ou d'eau millipore stérile.

### 3. Milieux de culture

#### Composition des milieux LB liquide et LB solide

Composition	milieu LB liquide	milieu LB solide
Bacto-tryptone ou peptone	10 g	10 g
Extrait de levure	5 g	5 g
NaCl	5 g	5 g
Bacto-agar	----	15 g
Eau millipore	qsp 11	qsp 11

Les milieux sont stérilisés par autoclave 2 heures à 110°C en atmosphère liquide et sous une pression de 1 atm pour les milieux liquides ou 0.5 atm pour les milieux solides. Après refroidissement les suppléments nécessaires (antibiotiques) sont ajoutés stérilement.

#### ***Antibiotiques et solutions pour milieux de culture***

Toutes ces solutions sont stérilisées par filtration sur membrane de porosité 0.22µm et stockées à -20°C et à l'abri de la lumière.

- Solution d'ampicilline à 100mg/ml dans l'eau.
- Solution de kanamycine à 500mg/ml dans l'eau.
- Solution d'IPTG à 0.8 mol/L dans l'eau.

## 4. Transformation des bactéries

Deux techniques sont décrites : la méthode chimique et la méthode par électroporation.

### 4.1. Méthode chimique

Cette technique a été principalement utilisée dans le cas de tests d'expression des protéines et dans la production de protéines en masse. En présence de certains cations métalliques tels  $Mn^{2+}$ ,  $Ca^{2+}$  et de certaines molécules organiques telles que le DMSO et le PEG, les parois et membranes bactériennes deviennent perméables par formation de pores autorisant la pénétration de l'ADN.

#### 4.1.1. Préparation des bactéries chimio-compétentes

Toutes les étapes sont effectuées stérilement.

- culture de 50 ml de milieu LB à 37°Censemencée avec une colonie prélevée sur gélose. Les cellules sont récoltées lorsque  $Abs_{600} = 0.5$  par centrifugation (15 minutes à 4000g et à 4°C)
- reprise du culot dans 25 ml de tampon de transformation glacé puis centrifugation (5 minutes à 1200g et à 4°C)
- resuspension douce du culot dans 3 ml de tampon de transformation glacé
- ajout goutte à goutte de DMSO à la concentration finale de 7% v/v
- les cellules sont aliquotées par 250  $\mu$ l et conservées à -80°C après trempage dans l'azote liquide.

#### Composition du tampon de transformation

Composition	Concentration
Tampon PIPES	10 mM/L pH=6.7
KCl	250 mM/L
CaCl <sub>2</sub>	15 mM/L
MnCl <sub>2</sub>	55 mM/L

#### 4.1.2. Transformation chimique

Toutes les étapes sont effectuées stérilement.

- adjonction de 25 à 250 ng de plasmide à 250 µl de bactéries chimio-compétentes préalablement décongelées sur la glace. Incubation de 30 à 60 minutes sur glace
- transformation par choc thermique 1 minute à 42°C
- incubation 5 minutes sur glace et ajout de 1 ml de milieu LB
- incubation d'environ 1 heure à 37°C et étalement sur milieu gélosé de 50, 100 et 250 µl de mélange de transformation

#### 4.1. Méthode par électroporation

Cette technique a été utilisée au cours du sous-clonage ; l'efficacité de transformation est au moins 100 fois supérieure à celle de la technique précédente. Dans ce cas la transformation est causée par un choc électrique bref et brutal qui provoque la formation transitoire de pores dans les parois et les membranes bactériennes. Le milieu doit être résistant au courant électrique et donc contenir aussi peu d'ions (provenant principalement des sels) que possible sous risque de voir se former des arcs électriques.

#### 4.2.1. Préparations des bactéries électro-compétentes

Toutes les étapes sont effectuées stérilement .

- culture de 1 l de milieu LB à 37°Censemencée avec une colonie prélevée sur gélose. Les cellules sont récoltées lorsque Abs<sub>600</sub> = 0.5 par centrifugation (15 minutes à 4000g et à 4°C)
- reprise des cellules dans 1 l d'eau millipore stérile glacée puis centrifugation (15 minutes à 4000g et à 4°C). Cette étape est répétée plusieurs fois afin d'éliminer toutes traces de sels
- resuspension des cellules dans 50 ml d'une solution stérile et glacée de glycérol à 10% (v/v) dans l'eau millipore puis centrifugation (15 minutes à 4000g et à 4°C)

- resuspension des cellules dans 5 ml d'une solution stérile et glacée de glycérol à 10% (v/v) dans l'eau millipore
- les cellules sont aliquotées par 50 µl et conservées à -80°C après trempage dans l'azote liquide

#### 4.2.2. Transformation par électroporation

Toutes les étapes sont effectuées stérilement .

- adjonction du volume adéquat d'ADN à 50 µl de bactéries électro-compétentes préalablement décongelées sur la glace. Une incubation d'environ 5 minutes se fait sur glace et le mélange est placé dans une cuvette d'électroporation stérile, propre et sèche.
- transformation par choc électrique avec un électroporateur Biorad-Gene Pulser II . Paramètres d'électroporation : Tension 2500 volts, Capacité 25 µFarads. La constante de temps est de l'ordre de 5 ms. La transformation se fait par choc électrique au cours de la décharge d'un condensateur. Le champ électrique associé provoque la formation de pores au travers des parois et membranes bactériennes permettant ainsi la pénétration des molécules d'ADN .
- resuspension immédiate des cellules dans 1 ml de LB et incubation d'environ 1 heure à 37°C et étalement sur milieu gélosé de 50, 100 et 250 µl de mélange de transformation.

Les ADN complémentaires codant pour les protéines d'intérêt sont ainsi clonés dans un vecteurs d'expression de choix. Elles pourront dès lors être surexprimées dans le système d'expression adéquat.



## **5. La technologie Gateway**

Depuis peu au laboratoire une nouvelle technologie de clonage s'est développée, c'est la technologie Gateway. La technologie Gateway (Invitrogen®) est une technologie de clonage basée sur les propriétés de recombinaison du bactériophage lambda (Landy, 1989). Un transfert rapide de séquences d'ADN dans de multiples systèmes est possible et permet une analyse de l'expression protéique.

### **5.1. La recombinaison**

Le système de recombinaison du bactériophage lambda est un système de recombinaison site spécifique qui facilite l'intégration du bactériophage lambda dans le chromosome d'*Escherichia coli* et permet le passage du cycle lytique au lysogénique du bactériophage. L'intégration dans le chromosome d'*Escherichia coli* a lieu via une recombinaison intermoléculaire de l'ADN entre le bactériophage lambda et les protéines recombinantes d'*Escherichia coli*. Cette recombinaison se fait par l'intermédiaire de séquences dites d'attachement spécifiques (*att*) : attB sur le chromosome d'*Escherichia coli* et attP sur le bactériophage lambda. Ces sites att servent de site d'interaction pour les protéines de recombinaison. Dès l'intégration du bactériophage lambda, la recombinaison se fait entre les sites attB et attP pour donner les sites attL et attR. L'ADN flanquant les sites de recombinaison est échangé, et après recombinaison, les sites att sont inclus dans la séquence hybride. La recombinaison est conservatrice et ne nécessite aucune synthèse d'ADN, elle peut être utilisée avec n'importe quel type d'ADN (superenroulé, linéaire, relaxé).

### **5.2. Les enzymes recombinantes**

La recombinaison du bactériophage lambda est catalysée par une variété d'enzymes qui se lie à des sites spécifiques (*att*). Ces enzymes recombinantes diffèrent selon le cycle du bactériophage lambda. Le cycle lysogénique du bactériophage lambda est catalysé par l'intégrase lambda et le facteur d'intégration d'*Escherichia coli*. Ces enzymes constituent le mélange BP Clonase®. Le cycle lytique du bactériophage

lambda est catalysé par l'intégrase lambda et l'excisionase et le facteur d'intégration d'*Escherichia coli*. Ces enzymes constituent le mélange LR Clonase®.

### 5.3. Les réactions de recombinaison

La réaction de recombinaison Gateway® se fait en deux étapes :

- **La réaction BP** : a lieu la recombinaison entre un substrat attB (produit PCR attB ou clone d'expression attB linéarisé) et un substrat attP (vecteur donneur) pour créer un clone d'entrée contenant les sites attL. Cette réaction est catalysée par l'enzyme BP clonase®.



- **La réaction LR** : a lieu la recombinaison entre un substrat attL (vecteur d'entrée) et un substrat attR (vecteur de destination) pour créer un clone d'expression contenant les sites attB. Cette réaction est catalysée par l'enzyme LR clonase®.



## C. SYSTEME D'EXPRESSION

### 1. Système d'expression dans *Escherichia coli*

Le gène correspondant à la protéine à produire est inséré dans un vecteur pET (Novagen ) sous le contrôle des signaux de transcription et de traduction du bactériophage T7. Les plasmides sont transformés dans la souche bactérienne BL21 ( DE3) contenant une copie du gène de l'ARN T7 polymérase sous contrôle du promoteur *lacUV5* et l'expression de la protéine est induite par ajout d'IPTG (isopropyl galactosidase, Euromedex ). L'expression de la protéine avant induction sous contrôle du promoteur T7/*lac* est réprimée par le répresseur *lac* codé par le gène *lacI*, contenu à la fois dans le plasmide pET et dans le génome d'*E.coli*, permettant une double répression (figure 52).

### 2. Système d'expression dans les cellules d'insectes infectées par un baculovirus

Le baculovirus utilisé pour infecter les cellules d'insectes est l'AcNPV (*Autographa californica nuclear polyhedrosis virus*) isolé des larves d'*Autographa californica*. Il se multiplie dans les noyaux des cellules de plus de trente espèces d'insectes lépidoptères. Ce virus contient dans son génome la séquence codant pour la polyhédrine, une protéine fortement exprimée lors de son cycle naturel mais qui n'est pas nécessaire à sa réplication, c'est une protéine essentielle de l'enveloppe et des inclusions virales. La stratégie consiste donc à utiliser le promoteur fort de la polyhédrine pour produire une autre protéine, en remplaçant par recombinaison homologue dans le génome viral la séquence codant pour la polyhédrine par la séquence codant pour la protéine à produire. Pour obtenir les virus recombinants, il s'agit d'insérer le gène à exprimer dans un vecteur de transfert bactérien (pVL13092 ou pSK277) au niveau d'un site de restriction entouré de séquences virales, en aval du promoteur de la polyhédrine. Les cellules d'insectes sont cotransfectées avec le vecteur de transfert et l'ADN de baculovirus sauvage. Une recombinaison entre les régions homologues du plasmide et du génome viral peut s'effectuer permettant l'obtention d'un virus recombinant portant le gène à exprimer. Le virus purifié est utilisé pour infecter en masse des cellules d'insectes (cellules de *Spodoptera*

*frugiperda* ou cellules *Sf9*) et la protéine est ainsi exprimée (figure 53). Dernièrement, la méthodologie «Bac to Bac® » a été développée. Elle consiste à utiliser *Escherichia coli* comme intermédiaire de transposition. Le Bacmid recombinant est généré et directement transfecté dans les cellules d'insectes (figure 54).

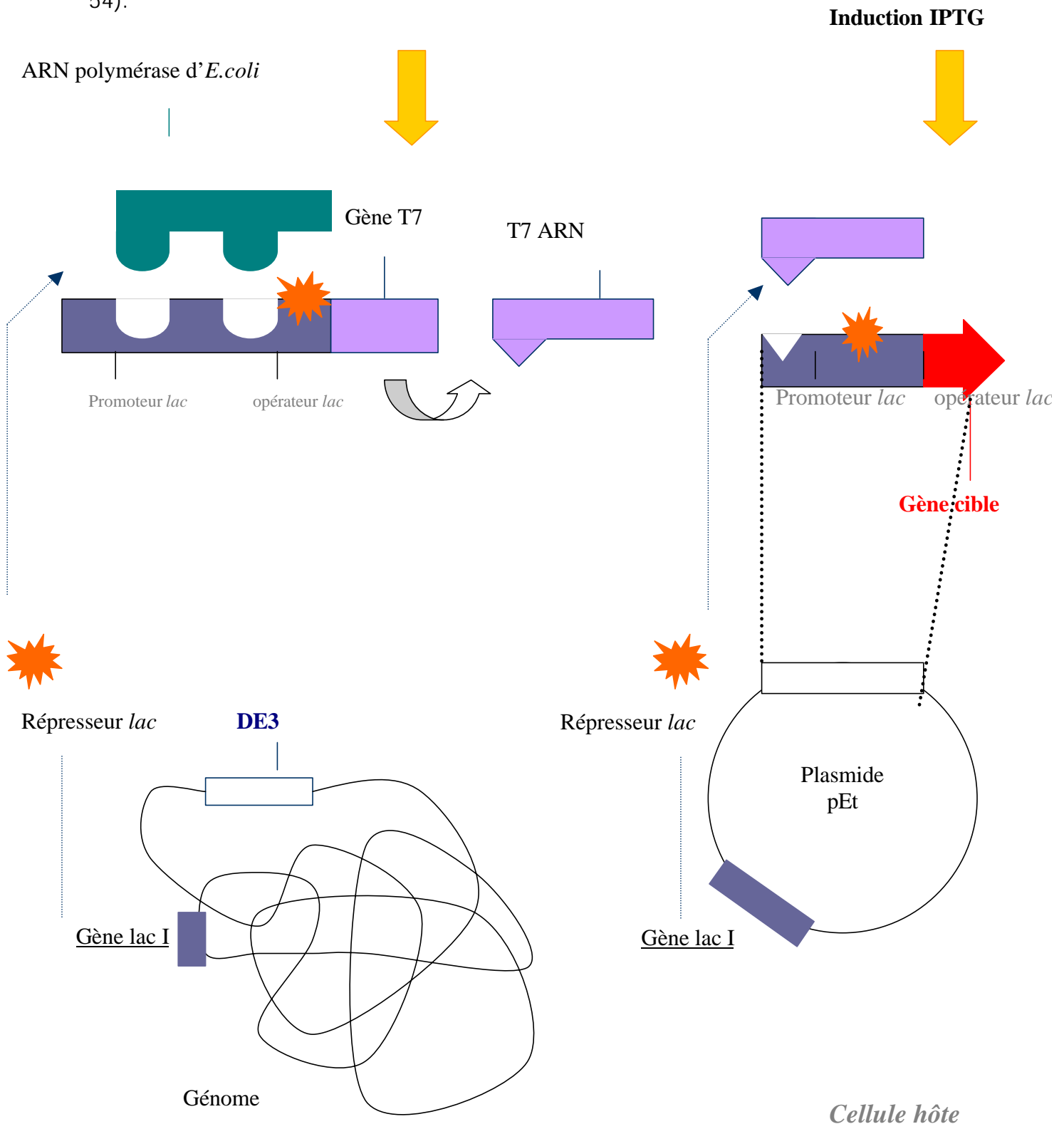


Figure 52 : Système d'expression dans *Escherichia coli*

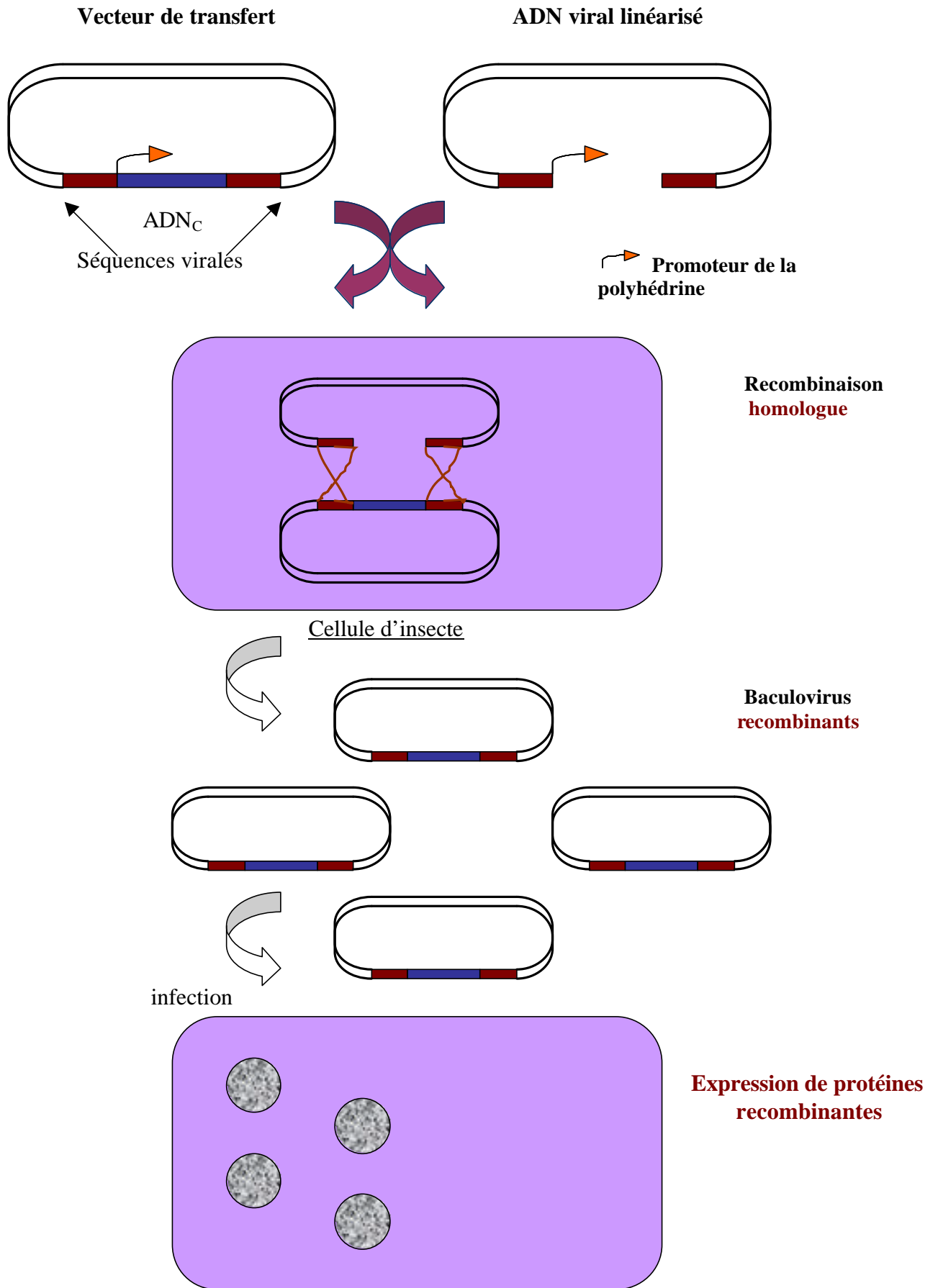


Figure 53 : Système d'expression dans les cellules d'insectes

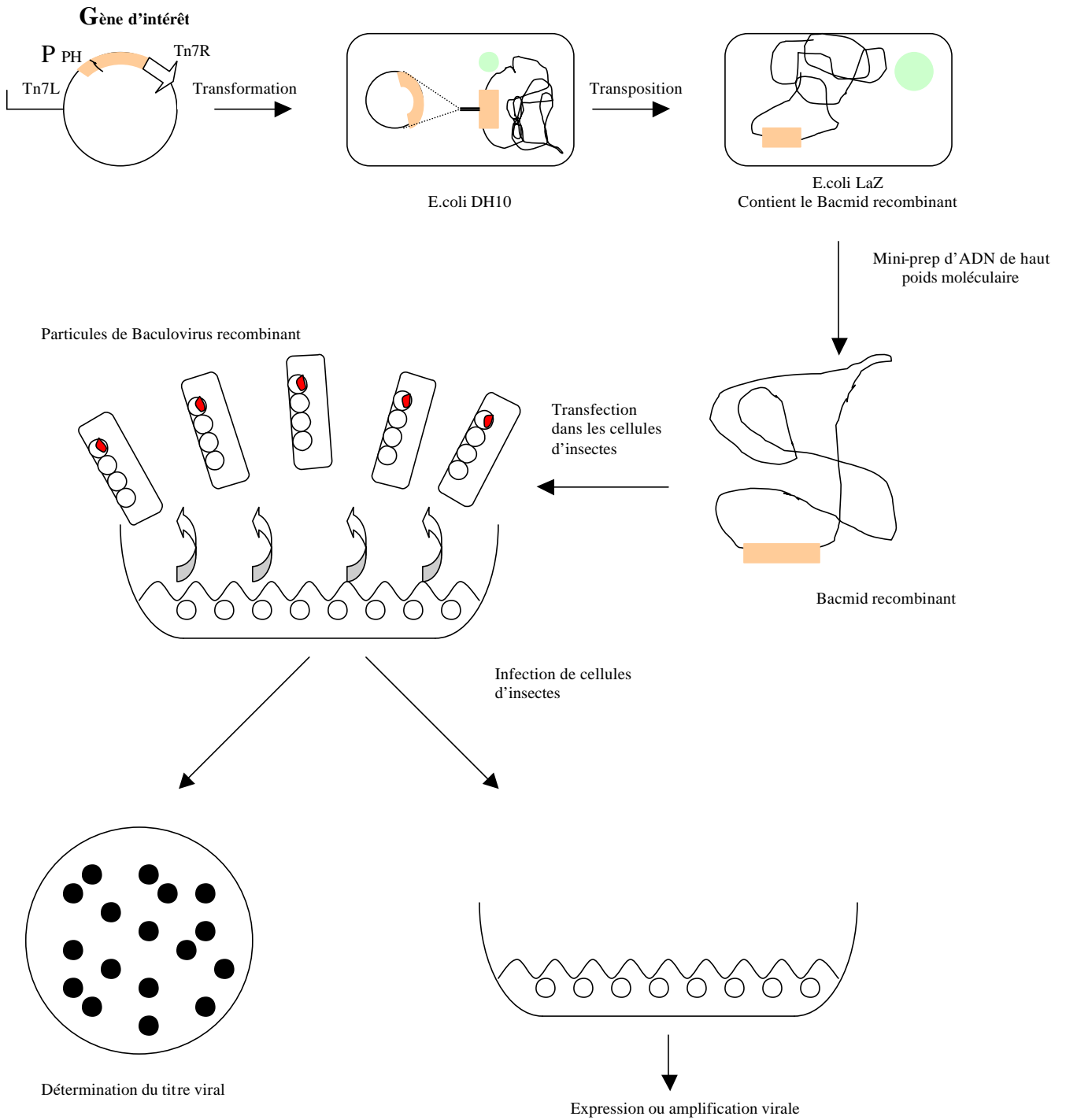


Figure 54 : Le système d'expression Bac-to-Bac®

## D. BIOCHIMIE

### ***Cultures et production de protéines dans Escherichia coli***

Les conditions d'expression sont identiques quelque soit la protéine considérée. Une préculture de 25 ml de milieu LB est directementensemencée avec un mélange de bactéries *Escherichia coli* souche BL21(DE3) préalablement transformées par méthode chimique ou électroporation avec les plasmides correspondants. La totalité de cette suspension est utilisée pour ensemencer 1L de milieu LB. La première phase de culture a lieu à 37°C sous vigoureuse agitation. La progression de la culture est suivie par lecture de l'absorbance à 600nm. Lorsque  $Abs_{600} = 0.4$  à  $0.6$  la biosynthèse de la protéine recombinante est induite par adjonction d'IPTG à la concentration finale de 0.8 mMol/L. La température de culture est alors abaissée à 20°C pendant toute la phase d'induction qui dure environ 4 heures. Les bactéries sont récoltées et soigneusement lavées par 50 ml de PBS1x et centrifugées à 4000g, 4°C. A l'issue de la centrifugation les culots bactériens secs sont directement congelés dans l'azote liquide et conservés à -80°C.

### ***Cultures et production de protéines dans les cellules d'insectes***

Les cellules d'insectes Sf9 (20.10<sup>6</sup> cellules par boîte) sont infectées par le baculovirus recombinant codant pour la protéine d'intérêt et cultivées soit en monocouche dans des grands flacons, soit en belcos ou encore en fermenteur en présence de milieu de culture Grace Sigma complémenté de 10% de sérum de veau fœtal, d'1% de glutamine et de 0.5% de gentamycine. Après élimination du milieu, les cellules sont mises en présence de 5 ml de solution virale dont la concentration est proportionnelle à la multiplicité d'infection par cellule (pfu pour *plaque forming unit*) qui varie de 1 à 10 en fonction du pouvoir infectieux des baculovirus recombinant. Les cellules infectées sont cultivées 48 heures à 27°C. les cellules sont récoltées et soigneusement lavées par 50 ml de PBS1x et centrifugées à 1500g, 4°C. A l'issue de la centrifugation les culots cellulaires secs sont directement congelés dans l'azote liquide et conservés à -80°C.

## 1. Purification

### 1.1. Sonication

Les conditions de lyse sont à adapter selon les protéines étudiées. Pour un premier test d'expression les conditions de lyse qui ont été utilisées sont les suivantes. Les culots bactériens ou de cellules d'insectes sont resuspendus grossièrement dans le volume adéquat de tampon de lyse.

#### Tampon de lyse

Composition	Concentration et pH
Tampon Tris-HCl	20 mMol/L et pH=7.5
Glycérol	10% v/v
NaCl	150 mMol/L
-mercaptoéthanol ou DTT	5 mMol/L ou 1 mMol/L
PMSF	0.1 mMol/L
Cocktail d'inhibiteurs de protéases	Au 1/100 <sup>ème</sup>
Leupeptine	10 µg/ml
Pepstatine	10 µg/ml
Chymostatine	10 µg/ml
Cystatine	10 µg/ml
Antipaine	10 µg/ml
Aprotinine	10 µg/ml

La lyse s'effectue principalement par sonication ou par la méthode de douce mais également par choc thermique lié à la décongélation des cellules. On utilise un sonicateur de type Vibracell 72412 avec une sonde de diamètre 13 mm et une sonde de température. Les paramètres de sonication sont une intensité de 40%, des impulsions de 4 secondes de sonication et 2 secondes de repos pendant au total 30 minutes. Les cellules d'insectes sont lysées mécaniquement par 50 passages au Potter.



## 1.2. Ultracentrifugation

L'extrait total de lyse est ensuite clarifié par ultracentrifugation (1 heures à 45000g et à 4°C). Le surnageant contenant le fraction soluble est conservé et immédiatement utilisé.

## 1.3. Chromatographie d'affinité sur résine de chélation de cations métalliques

Le principe de cette chromatographie dite IMAC (pour « Immobilized Metal Affinity Chromatography) est basé sur l'emploi de cations métalliques divalents ( $\text{Co}^{2+}$ ,  $\text{Zn}^{2+}$ ,  $\text{Ni}^{2+}$ ,  $\text{Cu}^{2+}$ ) chélatés par des groupements de type imido-acétate greffés sur une matrice de type sépharose pour séparer des protéines portant des polyhistidines. Ces cations immobilisés vont interagir avec les polyhistidines de la protéine produite et permettre de séparer rapidement et efficacement la protéine de fusion en utilisant l'imidazole comme agent d'élution.

La résine Talon (Clontech ) est conditionnée avec des cations  $\text{Co}^{2+}$  déjà fixés sur la résine. L'avantage de cette résine réside dans la nature du cation fixé et dans le type de groupement chélatant. Le cobalt interagit moins fortement avec les protéines à polyhistidines, la fixation est moins forte mais plus spécifique. Il y a donc moins de fixation aspécifique et on sépare encore certains contaminants. De plus l'élution se fait à de faibles concentrations en imidazole, ce qui est préférable, celui ci a déjà été décrit comme entraînant l'agrégation et la précipitation des protéines. Enfin le type de groupement chélatant réduit le risque de coélution accidentelle du cation métallique avec la protéine.

## 1.4. Protéolyse ménagée par la thrombine ou par la TEV du peptide hexahistidine

L'élimination par thrombinolyse ménagée du peptide polyhistidine situé en N-terminal des protéines surproduites et purifiées à partir du système pET est réalisée en utilisant la thrombine bovine plasmatique (Sigma Biochemicals ). Cette protéase à sérine requiert la présence de cations divalents calcium et un pH de 8.0-8.5. Dès que la thrombinolyse du peptide polyhistidine s'est avérée complète par visualisation sur

gel de protéines, l'arrêt de la réaction est faite par adjonction d'un inhibiteur : le Pefabloc (Boehringer Mannheim ) qui est soluble dans l'eau. L'élimination de la thrombine et de l'inhibiteur est faite ultérieurement par chromatographie sur colonne analytique de filtration sur gel (Superdex ou HR 10/30 Pharmacia ).

L'élimination par protéolyse ménagée du peptide polyhistidine situé en N-terminal des protéines surproduites et purifiées à partir du système pSKB3 est réalisée en utilisant la TEV (tobacco etch virus). Cette protéase présente l'avantage de couper plus spécifiquement le peptide polyhistidine. De plus elle est elle-même fusionnée à un peptide polyhistidine ce qui permet son élimination après coupure par une chromatographie d'affinité sur résine de chélation de cations métalliques.

### **1.5. Héparine**

L'héparine est un glycosaminoglycane. Le complexe multi-protéique se lie à l'héparine grâce à une affinité pour le polymère mimant l'ADN et par un échange par interactions cationiques. La purification se fait principalement selon l'affinité des protéines vis à vis de l'ADN. Les protéines sont éluées par un gradient en sels. La résine utilisée est l'Heparine Sepharose 6 Fast Flow (Pharmacia ).

### **1.6. Autres protéines de fusion**

Un tableau comparatif présente les principales caractéristiques des principales protéines de fusion utilisées pour l'expression des protéines dans le système de production baculovirus ou dans *Escherichia coli*.

	HIS	FLAG	CBP	PROTEIN A	GST	THIOREDOXIN
TAG PETIT	X	X	X			
ELUTION DOUCE			X		X	
UNE ETAPE		X	X	X	X	X
SPECIFICITE	HAUTE	HAUTE	MOYEN	HAUTE	HAUTE	MOYEN
SOLUBILITE				AMELIORE	AMELIORE	AMELIORE
ELUTION	imidazole	Compétition peptide	pH neutre EGTA	Compétition	Compétition glutathion	mercaptoéthanol
ENZY. TEST					X	
PRIX		ELEVE		ELEVE		

Tableau 9 : Tableau récapitulatif des principales caractéristiques des différentes protéines de fusion

### 1.7. Gel Filtration

La gel filtration, ou chromatographie d'exclusion moléculaire permet de séparer des molécules en fonction de leur poids moléculaire (PM) et de leur forme. Les grosses molécules (dont le diamètre est supérieur à celui des pores) sont exclues et sont donc éluées les premières, au niveau du volume mort ( $V_m$  ou  $V_0$ ). Les petites et moyennes molécules sont éluées plus tardivement, car incluses dans le gel, leur migration est freinée.

Le kit de calibration (Biorad ) a été utilisé afin de déterminer la courbe de calibration de la gel filtration S75 (tableau 10). Il existe une relation linéaire entre le volume d'éluion et le logarithme de la masse moléculaire. La fonction reliant la taille d'une protéine globulaire et le volume d'éluion sur une colonne de filtration est la suivante :

Avec  $K_{AV}$ , le coefficient de partage entre la phase liquide et la phase gel :

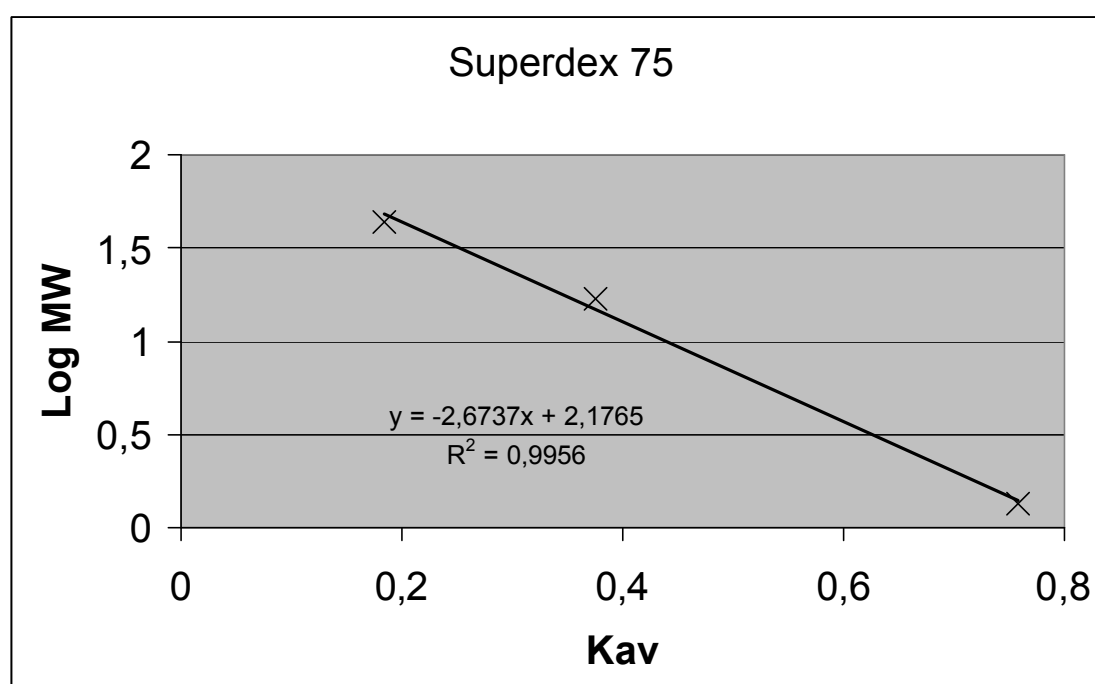
$$\text{Log (PM)} = f(K_{AV})$$

$$K_{AV} = \frac{(\text{volume d'élution} - \text{volume total})}{\text{volume total} - \text{volume exclus}}$$

$$K_{AV} = \frac{(\text{volume d'élution} - \text{volume total})}{(\text{volume total} - \text{volume exclus})}$$

	PM	Volume d'élution	$K_{AV}$
Bleu dextran	2Ma	9	-
Thyroglobuline	670kDa	9.6	0.05
? globuline de poulet	158kDa	12.8	0.322
Ovalbumine de poulet	44kDa	14.8	0.491
Myoglobine de cheval	17.5kDa	17.4	0.711
Vitamine B12	1.35kDa	20.8	1

Tableau 10 : Composition du kit de calibration pour gel filtration et calcul du  $K_{AV}$  pour la gel filtration G75



## 1.8. Immunopurification par le système FLAG

Pour immunopurifier les protéines fusionnées au peptide FLAG, le système d'affinité des billes anti-FLAG-M2 est utilisé. Il s'agit d'anticorps monoclonaux IgG1 de souris attachés à des billes d'agarose par pontage chimique à l'hydrazine. Le Flag est un peptide de 8 amino-acides de séquence DYKDDDDK fusionné à la protéine d'intérêt en position N-terminale. Il permet à la protéine d'être retenue par les billes Flag-M2. L'extrait cellulaire brut obtenu est incubé 4 heures sous agitation à 4°C en présence des billes Flag-M2 (1/10 v/v). Le surnageant est éliminé après une courte centrifugation à 1000g. Les billes Flag-M2 sont lavées trois fois dans un tampon de lavage à haut sels. Les protéines retenues de manière spécifique sur les billes sont éluées sous agitation à 4°C en présence de peptide compétiteur correspondant à la séquence du peptide Flag (1/5 v/v). Le pH du peptide est au préalable ajusté à pH=7 et sa concentration à 1 mg/ml. L'éluion a lieu sous agitation à 4°C pendant une nuit.

### Tampon de lavage

Composition	Quantités
Tris-HCl pH 7.5	20 mMol/L
NaCl	400 mMol/L
NP40	0.1%
Glycérol	20%
-mercaptoéthanol ou DTT	5mMol/L ou 1 mMol/L

## 1.9. Interaction de protéines recombinantes

Pour tester les interactions entre deux protéines recombinantes deux méthodes ont été utilisées : soit l'immunopurification anti-FLAG soit le GST-pull down.

Pour tester l'interaction de la protéine A avec la protéine B, on réalise deux co-infections différentes des cellules d'insectes *Sf9*. La première utilise le baculovirus codant pour la protéine A fusionnée à un peptide Flag en N-terminal et le baculovirus codant pour la protéine B. La deuxième, qui constitue le témoin négatif, utilise le baculovirus codant pour la protéine A non fusionnée ainsi que le baculovirus codant pour la protéine B. Après infection, les extraits cellulaires sont préparés comme décrits. On réalise une immunopurification par le système anti-flag-M2. La protéine A fusionnée au peptide Flag est retenue spécifiquement. Les protéines non fusionnées ne sont théoriquement pas retenues. Cependant, dans le cas où il y a interaction entre la protéine A et la protéine B, la protéine B sera retenue par l'intermédiaire de A.

### ***GST-pull down***

Pour tester l'interaction de la protéine A avec la protéine B par GST-pull down, on réalise une co-infection en cellules d'insectes *Sf9* avec le baculovirus codant pour la protéine A fusionnée à la GST (Glutathion-S-transférase) en N-terminal et le baculovirus codant pour la protéine B. L'extrait cellulaire est mélangé à une suspension de billes d'agarose couplées au Glutathion. La partie GST de la protéine de fusion se fixe à son substrat (le glutathion) mais est incapable de le métaboliser. Le complexe ainsi formé (Billes-Glutathion-GST-protéine) peut être facilement séparé du reste des protéines solubles par simple centrifugation à faible vitesse. Après trois lavages, ce complexe peut être rompu soit par compétition (en ajoutant une dose massive de Glutathion réduit) soit par action de la thrombine. En effet, le plasmide pDEST20 introduit un site de clivage protéique par la thrombine au niveau de son site multiple de clonage.

### Élution par compétition

L'ajout de glutathion réduit provoque un déplacement de l'équilibre enzyme-substrat par compétition. En effet la protéine de fusion peut alors se fixer soit sur le glutathion lié aux billes d'agarose soit au glutathion réduit libre. Lorsque le glutathion réduit est ajouté en large excès, cet équilibre est totalement déplacé vers le décrochage de la protéine de fusion des billes d'agarose. Une centrifugation à faible vitesse permet de séparer le culot de billes du surnageant contenant la protéine de fusion en suspension.

### Élution par clivage protéolytique

L'action de la thrombine pendant une nuit à 16°C sur le complexe billes-Glutathion-GST permet de libérer la protéine ou le complexe d'intérêt. La coupure à la thrombine se fait directement sur les billes, ce qui permet d'éliminer la GST recombinante. Une centrifugation permet de séparer le culot du surnageant qui contient la protéine ou le complexe.

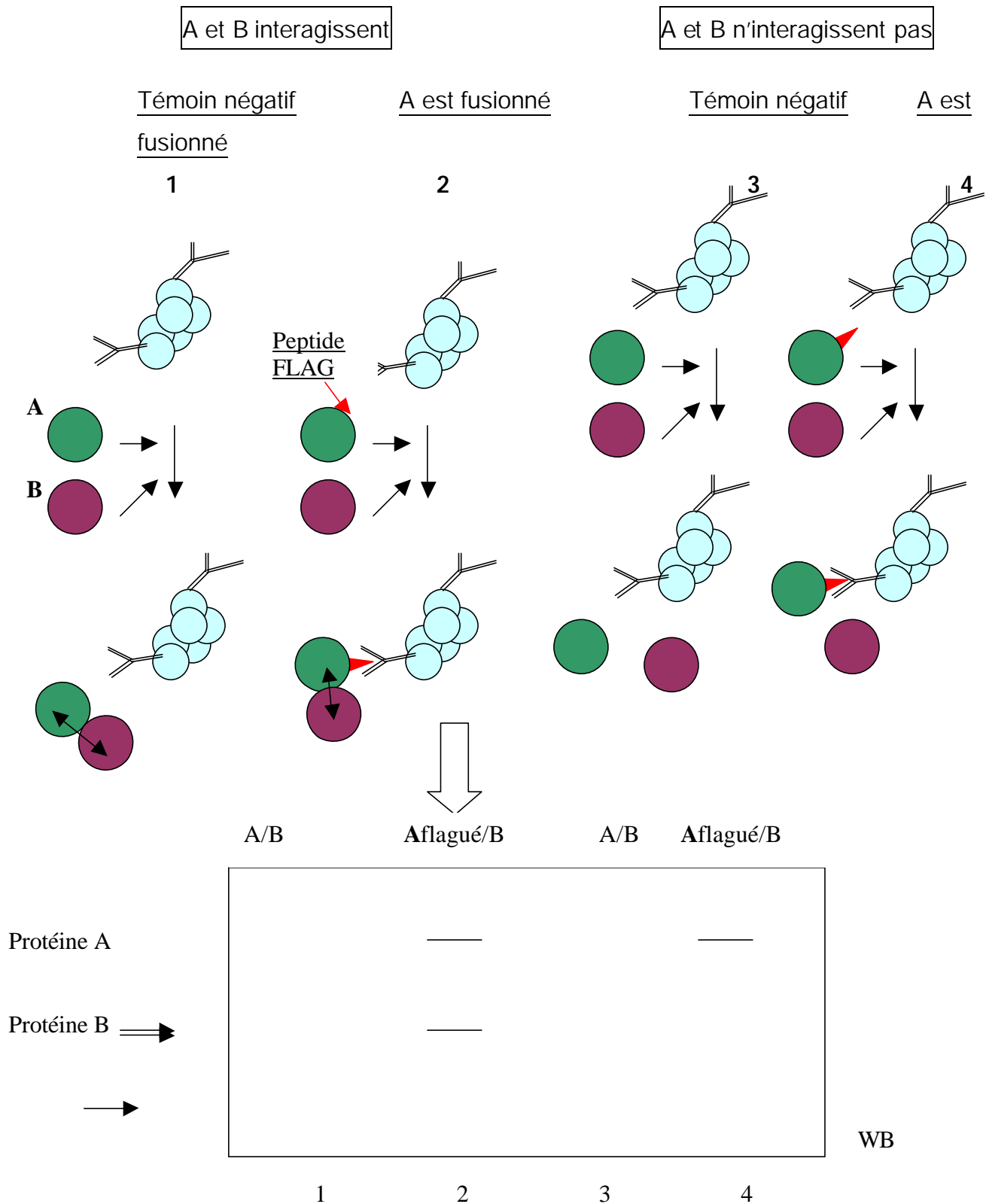


Figure 55 : Immunopurification par le système Flag et analyse par Western-Blot

1. La protéine A n'est pas fusionnée au peptide Flag, A et B ne sont pas retenues
2. La protéine A est fusionnée au peptide Flag, A et B sont retenues par l'intermédiaire de A
3. La protéine A n'est pas fusionnée au peptide Flag, A et B ne sont pas retenues
4. La protéine A est fusionnée au peptide Flag, A est retenue, B n'est pas retenue



## 2. Caractérisation des protéines purifiées

### 2.1. Electrophorèse sur gel de polyacrylamide en conditions dénaturantes

Gels de séparation à 12,5% et 15% (pour 4 gels)

<u>Type de gel</u>	<u>12.5%</u>	<u>15%</u>
<b>Composition</b>	<b>Volume</b>	<b>Volume</b>
Tris-HCl 1.5 mol/L pH=8.8 et SDS0.4% m/v	6.3 ml	6.3 ml
Acrylamide/Bis-Acrylamide 30/0.8 m/v	10.4 ml	12.5 ml
Eau millipore	8.3 ml	6.2 ml
Persulfate d'ammonium 10% m/v	250 µl	250 µl
TEMED	20 µl	20 µl

Pré-gel de résolution à 5%

<b>Composition</b>	<b>volume</b>
Tris-HCl 1.5 mol/L pH=6.8 et SDS0.4% m/v	3.2 ml
Acrylamide/Bis-Acrylamide 30/0.8 m/v	2.0 ml
Eau millipore	7.2 ml
Persulfate d'ammonium 10% m/v	125 µl
TEMED	10 µl

Tampon dissociant concentré 5%

<b>Composition</b>	<b>Concentration</b>
Tris-HCl pH 6.8	250 mol/L
Glycérol	50% v/v
SDS	10% m/v
-mercaptoéthanol	140 mMol/L
DTT	100 mMol/L
Bleu de bromophénol	0.05 % m/v
Eau millipore	qsp

Les échantillons mélangés au tampon dissociant sont dénaturés par chauffage au bain-marie (5 minutes à 100°C). La migration s'effectue à une intensité constante de 80 mA.

Tampon de migration 5%

Composition	Quantité
Tris-Base	30 g
Glycine	188 g
SDS	10 g
Eau millipore	qsp 2000 ml

## 2.2. Révélation au bleu de Coomassie

Le gel est démoulé de son support et mis à fixer et colorer par immersion dans une solution de coloration au bleu de Coomassie pendant 15 minutes à température ambiante et sous agitation. Le gel est ensuite décoloré dans le solution de décoloration pour révéler les protéines précédemment fixées et colorées. Cette solution filtrée est stable.

Solution stock de bleu de Coomassie concentrée 2x

Composition	Quantité
Bleu de Coomassie	4 g
EtOH ou MetOH	120 ml
Eau millipore	qsp 200 ml

Solution de fixation et coloration au bleu de Coomassie

Composition	Quantité
Bleu de Coomassie stock 2x	50 ml
AcOH	10 ml
Eau millipore	qsp 100 ml

Solution de fixation et de décoloration

Composition	Quantité
AcOH	10%
EtOH ou MetOH	30% m/v
Eau millipore	qsp

**2.3. Révélation au nitrate d'argent**

Le gel est démoulé de son support et mis successivement dans les solutions suivantes :

- Fixation dans une solution à 50% v/v EtOH (ou MetOH) pendant 15 minutes
- Lavage en eau millipore pendant 5 minutes
- Coloration dans une solution de complexe argent ammoniacale préparée extemporanément pendant 15 minutes
- Lavage en eau millipore pendant 5 minutes
- Révélation dans une solution citrate/formaldéhyde préparée extemporanément
- Solution d'arrêt de la réaction de révélation

Solution argentique de coloration

Composition	Quantité
NaOH 2.5 Mol/L	800 µl
NH <sub>3</sub> aqueux 25% v/v	1.6 ml
AgNO <sub>3</sub> 100% m/v	800 µl
Eau millipore	qsp 100 ml

Inclure les ingrédients dans cet ordre et ajouter AgNO<sub>3</sub> progressivement en agitant pour éliminer le précipité.

Solution de révélation

Composition	Quantité
Acide citrique 20% m/v	800 µl
Formaldéhyde 27% m/v	800 µl
Eau millipore	qsp 250 ml

Solution de fixation

Composition	Quantité
EtOH ou MeOH	50% v/v
AcOH	10% v/v
Eau millipore	qsp

**2.4. Western Blot**

Les éluats protéiques sont déposés sur gel de polyacrylamide 12.5% en condition dénaturante. Après migration, ce gel est transféré sur une membrane de nitrocellulose qui est incubée 30 minutes dans une solution de TBS-Tween à 10% de lait afin de saturer les sites non spécifiques.

Solution de TBS-Tween

Composant	Quantité
Tris-HCl pH=7.5	50 mMol/L
NaCl	150 mMol/L
Tween	0.05%

Ensuite, la membrane est incubée pendant 1 heure dans une solution de TBS-Tween à 10% de lait contenant l'anticorps primaire, à une dilution précise. Les ascites ont été produits dans le service commun d'anticorps monoclonaux de l'IGBMC dirigé par le Dr Mustapha Abdelghali .

Anticorps utilisés et dilution

<b>Protéine</b>	<b>Anticorps monoclonal</b>	<b>Région reconnue</b>	<b>Dilution</b>
XPB	2G12	C-terminale	1/1000
XPB	2F6	C-terminale	1/2000
P62	3C9	Milieu et N-terminale	1/10000
P52	1D11	C-terminale	1/1000
P44	1H5	N-terminale	1/2000
P34	2B1	Résidu 233 à 303	1/2000
Cycline H	2D4	Résidu 291 à 311	1/2000
Cdk7	2F8	Résidu 313 à 330	1/2000
MAT1	2D3	Résidu 241 à 257	1/1000
Peptide flag	2FL-1B11	DYKDDDDK	1/1000
Peptide his tag	1H7-7-1 (1DM)	Hexa-histidine	1/1000

Trois lavages de trois fois 5 minutes sont ensuite réalisés dans une solution de TBS-Tween. La membrane est alors incubée dans une solution de TBS-Tween contenant l'anticorps secondaire anti-souris dilué au 1/10000. Après trois lavages de trois fois 5 minutes dans une solution de TBS-Tween, les protéines fixées sur la membrane reconnues par l'anticorps primaire sont révélées par le kit ECL pour la détection des protéines en Western Blot par chimioluminescence.

### 3. Dosage des protéines

Les protéines sont dosées :

- par spectrométrie d'absorption à 280 nm en utilisant les coefficients d'extinction molaire linéique calculés à partir de la composition en acide aminés.
- par méthode de Bradford en dosage à spectro photo colorimétrique à 595 nm. La protéine utilisée pour l'étalonnage est la sérum albumine bovine.

### 3.1. Spectrométrie de masse

Les échantillons pour spectrométrie de masse sont conditionnés dans un tampon d'acétate d'ammonium à 500 mM pH=7. Ce tampon volatil permet d'effectuer des mesures de spectrométrie de masse aussi bien en conditions dissociantes que non dissociantes. Ce conditionnement peut s'effectuer de deux manières :

- par échange de tampon sur Centricon (Grace ) ou dialyse en Micro-Collodion (Sartorius )
- par chromatographie sur colonne analytique de filtration sur gel (Superdex ou HR 10/30 Pharmacia ) équilibrée dans ce même tampon.

### 3.2. Digestion trypsique

La digestion trypsique a été conditionnée dans un tampon Tris-HCl pH=8, CaCl<sub>2</sub> 20mM avec un ratio enzyme/substrat de 1/20 pendant 4 heures à 37°C. Le mélange de peptides de l'hydrolyse des protéines en C-terminale des lysines et des arginines est analysé par LC-MS (Liquid Chromatography Mass Spectrometry).

### 3.3. Microséquencage

La dégradation d'Edman des peptides obtenus après coupure enzymatique et la détection des dérivés phenylthiohydantoïne a été observé sur un séquenceur de type Applied Biosystems , modèle 473A.

Les mesures de spectrométrie de masse ont été effectués par Noëlle Potier et Hélène Nierengarten du service commun de spectrométrie de masse de l'IGBMC.

## 4. Tests d'activité

### 4.1. Marquage radioactif d'un oligonucléotide au [ $\gamma$ - $^{32}\text{P}$ ]ATP

La T4 polynucléotide kinase catalyse le transfert du phosphate [ $^{32}\text{P}$ ] de [ $^{32}\text{P}$ ]ATP sur l'extrémité 5' de l'oligonucléotide synthétique (les oligonucléotides sont synthétisés sans phosphate à leur extrémité 5'). Le protocole ci-après permet le marquage de 10 pmoles d'oligonucléotide synthétique. Pour marquer une plus grande quantité d'oligonucléotide synthétique il suffit d'augmenter les concentrations de chaque produit de manière constante. Si la réaction de marquage est effectuée de manière efficace, l'activité spécifique (AS) de la sonde radioactive peut être tout aussi importante que l'activité spécifique de l' [ $^{32}\text{P}$ ]ATP lui-même.

#### Réaction de transfert du phosphate radioactif :

Oligonucléotide (10 pmoles/ $\mu\text{l}$ ) : 1 $\mu\text{l}$

Tampon T4 polynucléotide kinase 10x : 2 $\mu\text{l}$

[ $^{32}\text{P}$ ] ATP (AS : 3000 Ci/mmoles) 10 pmoles : 5 $\mu\text{l}$

H<sub>2</sub>O : 11 $\mu\text{l}$

T4 polynucléotide kinase (10U/ $\mu\text{l}$ ) : 1 $\mu\text{l}$

Le transfert se fait pendant 45 minutes à 37°C. La réaction est stoppée par un chauffage de 10 minutes à 68°C.

Dans ces conditions, 50% de [ $^{32}\text{P}$ ] ATP est transféré sur l'oligonucléotide, soit une activité spécifique de l'oligonucléotide de 1500 Ci/mmoles. L'efficacité du transfert peut être améliorée en augmentant la concentration d'oligonucléotide d'un facteur 10. Le transfert de radioactivité sur l'oligonucléotide est alors de 90% mais l'activité spécifique est réduite d'un facteur 5. La sonde radioactive est ensuite purifiée pour éliminer la radioactivité non incorporée. La méthode la plus simple est la précipitation à l'éthanol (pour des oligonucléotides d'une taille supérieure à 18 mers), mais cette méthode n'est pas suffisante pour éliminer toutes traces de radioactivité. La purification sur Micro BioSpin P30 (Biorad), une colonne chromatographique comprenant une résine d'exclusion, permet d'éliminer environ 90% de radioactivité non incorporée. Le protocole d'utilisation est disponible sur le site du fabricant.

L'efficacité du marquage ainsi que l'élimination de la radioactivité non incorporée sont vérifiées sur gel non dénaturant.

#### Gels non dénaturant

<b>Composants</b>	<b>10%</b>	<b>14%</b>	<b>15%</b>
Acryl 40%	17.5 ml	24.5 ml	26 ml
Bis Acryl 1.5%	14 ml	19.6 ml	21 ml
TBE 10x	3.5 ml	3.5 ml	3.5 ml
H2O	35 ml	22.4 ml	19.6 ml
TEMED	70 µl	70 µl	70 µl
APS	700 µl	700 µl	700 µl

La migration s'effectue à 150 volts pendant environ 3 heures dans du TBE 0.5x. Le gel est ensuite démoulé et exposé O/N à  $-80^{\circ}\text{C}$  avec un film autoradiographique.

#### **4.2. Hybridation des oligonucléotides et réaction de déroulement de brins**

L'oligonucléotide peut être dès lors utilisé dans des hybridations. Les hybridations entre deux oligonucléotides complémentaires se font en quantité équimolaires et en présence de 50mM NaCl et 5mM MgCl<sub>2</sub>. La réaction est chauffée à 95°C et lentement refroidit O/N sur la paillasse. La formation de l'hybride est vérifié sur gel non dénaturant (figure 56). Lors de la désybridation de l'hybride un excès d'oligonucléotide froid (identique à l'oligonucléotide chaud) est rajouté afin d'éviter la réhybridation des deux brins.



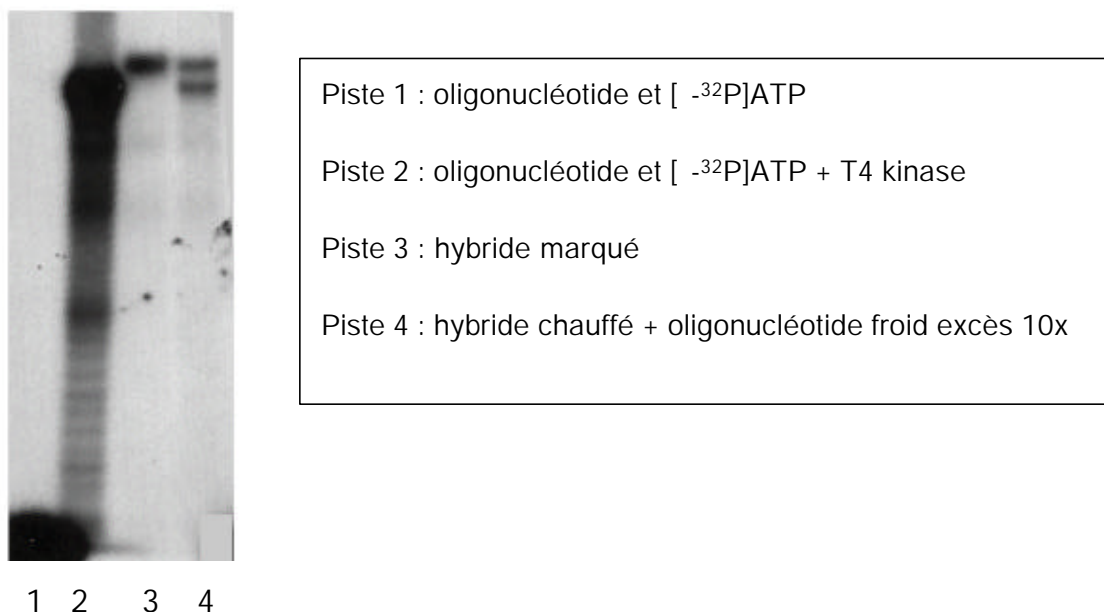


Figure 56 : Purification de l'oligonucléotide marqué au [ <sup>32</sup>P]ATP et hybridation

### 4.3. Test hélicase

La sonde hélicase est construite à partir du phage M13 simple brin auquel est hybridé un oligonucléotide complémentaire de la région 6218-6251. Après digestion par *EcoRI* et marquage radioactif des extrémités, la matrice linéaire présente deux zones doubles brins de taille différente et de polarité opposée. Une fois marquée, la sonde est déposée sur une colonne de gel filtration Sepharose CL4B (Pharmacia ). L'ouverture de l'ADN est mise en évidence par le déplacement des oligonucléotides séparés sur gel non dénaturant. Un  $\mu$ l de chacune des fractions 9 à 19 est dénaturé par la chaleur et déposé sur gel de polyacrylamide 14% non dénaturant. Les fractions les plus radioactives sont mélangées (figure 57).

La fraction contenant le facteur TFIIH purifié est incubée 45 minutes à 37°C en présence de la sonde hélicase (30000-40000 cpm ; 1-3ng d'ADN) dans 20 mM Tris-HCl, pH=7.9, 1mM DTT, 4 mM MgCl<sub>2</sub> et 4 mM ATP, contenant 100  $\mu$ g/ml de BSA, pour un volume final de 25  $\mu$ l. La réaction est arrêtée par adjonction de 10  $\mu$ l de 60 mM EDTA, 50% glycérol, 0.75% SDS et 0.1% de bleu bromophénol. Les échantillons sont déposés sur gel de polyacrylamide 10%. Le gel est fixé 10 minutes dans 10% acide acétique, 10% méthanol et autoradiographié de 6 à 12 heures à -80°C.

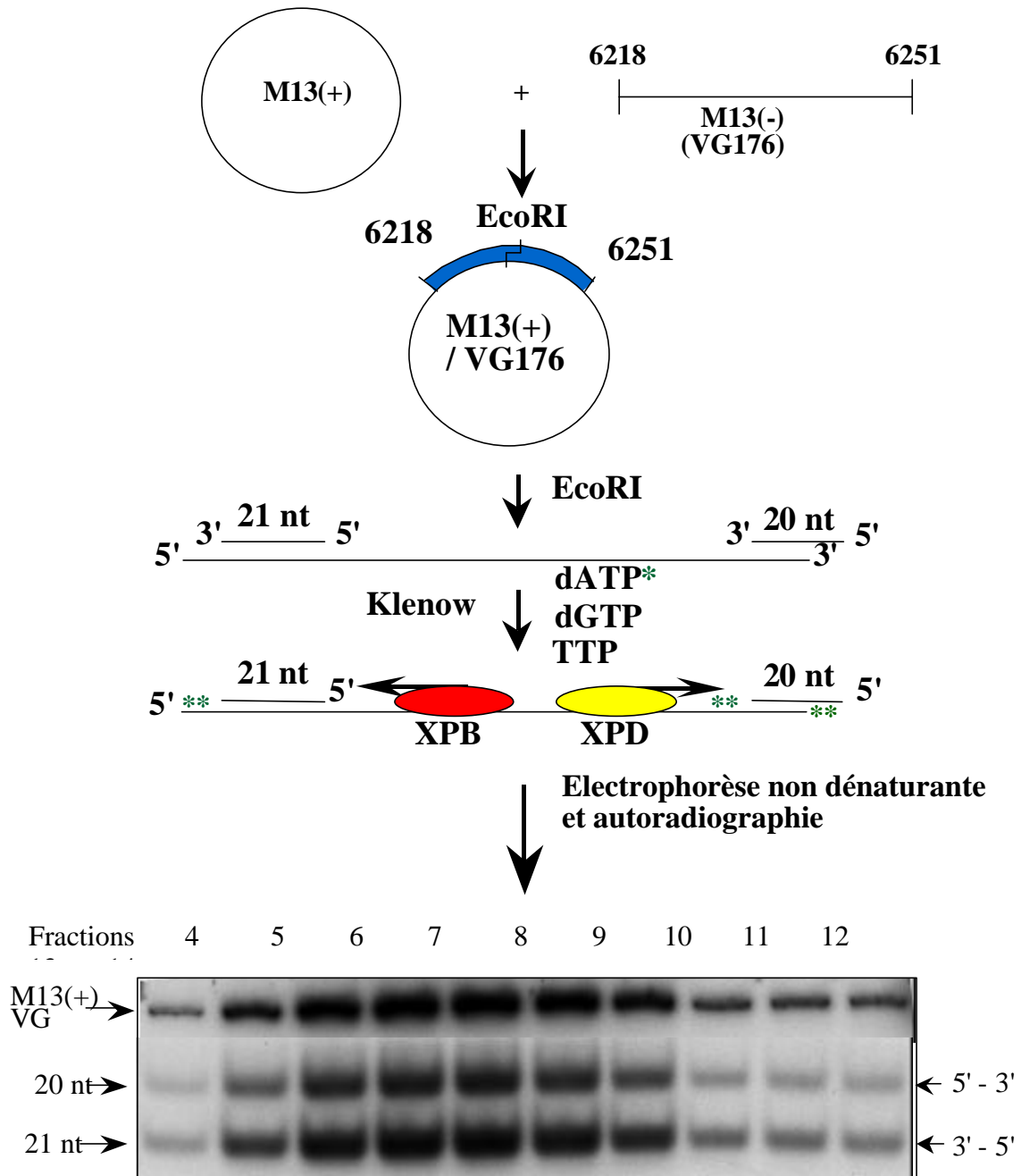


Figure 57 : Construction et purification de la sonde hélicase

#### 4.4. Test de transcription *in vitro*

La matrice utilisée est le promoteur majeur tardif de l'adénovirus (Ad2MLP). La réaction de transcription se fait en deux étapes (figure 58):

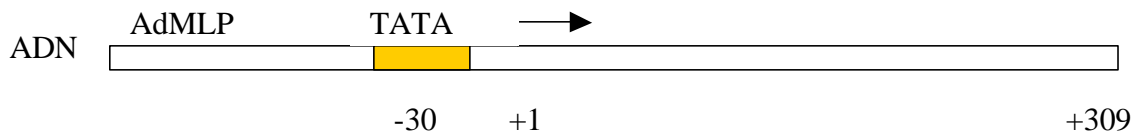
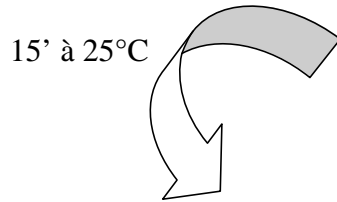
- une étape de préincubation dans laquelle la matrice d'ADN est préincubée avec tous les composants de la machinerie transcriptionnelle de base (TFIIA, TFIIB, TFIID/TBP, TFIIIE, TFIIF, TFIIH et l'ARN polymérase II) afin de permettre la formation du complexe de préinitiation
- une étape de synthèse de l'ARN qui débute par l'adjonction de tous les ribonucléotides (incluant [  $^{-32}\text{P}$ ]CTP)

La réaction se déroule comme suit :

- l'extrait cellulaire brut (correspondant à 50-75  $\mu\text{g}$  de protéines totales) est incubé 15 minutes à 25°C avec 50-70 ng de matrice en présence de 5 mM de  $\text{MgCl}_2$  pour un volume final de 20  $\mu\text{l}$ . Puis 250  $\mu\text{M}$  d'ATP, de GTP et d'UTP, 10  $\mu\text{M}$  de CTP et 4  $\mu\text{Ci}$  de [  $^{-32}\text{P}$ ]CTP sont ajoutés en présence de 6.5 mM de  $\text{MgCl}_2$  et le tout est incubé 45 minutes à 25°C.
- la réaction est arrêtée par adjonction de 400  $\mu\text{l}$  d'une solution contenant 50 mM Na acétate pH=5.2, 0.5% SDS et 50  $\mu\text{g/ml}$  d'ARNt. Une extraction au phénol/chloroforme est réalisée et les acides nucléiques contenus dans la phase supérieure sont précipités par 1/20 d'ammonium acétate et de 2.5 volumes d'éthanol. Le culot est séché et resuspendu dans un tampon de dépôt pour gel de polyacrylamide 5% en conditions dénaturantes. Après électrophorèse le gel est autoradiographié une nuit à -80°C.

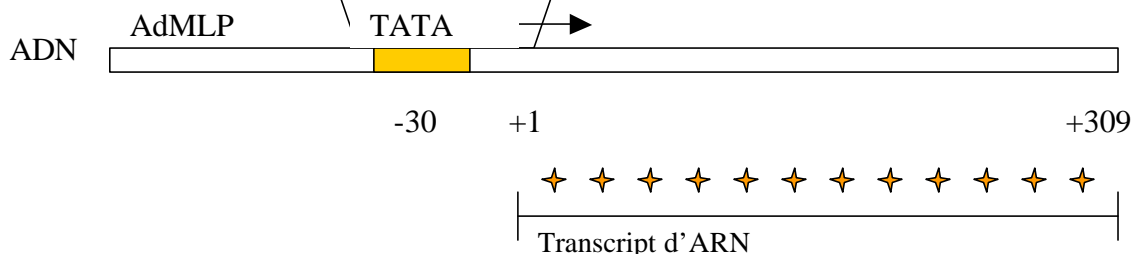
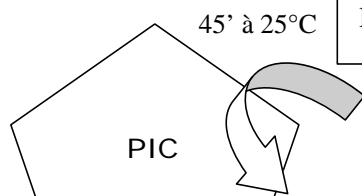
Formation du complexe de préinitiation

- TFIIA ● **Machinerie**
- TFIIB ● **transcriptionnelle**
- TFIID/TBP ● **de base**
- TFIIE ●
- TFIIF ●
- TFIIH ●
- RNA Pol II ■



Synthèse du transcript d'ARN

- NTPs + **CTP\***
- MgCl<sub>2</sub> : 5-10 mM
- KCl : 40-80 mM



Electrophorèse dénaturante et autoradiographie

Figure 58 : Test de transcription *in vitro*

## Bibliographie

A.

- Adachi, N., Matsumoto, M., Hasegawa, S., Yamamoto, T., and Horikoshi, M.** Analysis of TFIIH subunit through isolation of the gene from *Schizosaccharomyces pombe* corresponding to that of *Saccharomyces cerevisiae* SSL1, reveals the presence of conserved structural motifs 2. *Yeast*15[3], 255-262. (1999).
- Adamczewski, J. P., Rossignol, M., Tassan, J. P., Nigg, E. A., Moncollin, V., and Egly, J. M.** MAT1, cdk7 and cyclin H form a kinase complex which is UV light-sensitive upon association with TFIIH 58. *EMBO J.*15[8], 1877-1884. (15-4-1996).
- Adrian, M., Dubochet, J., Fuller, S. D., and Harris, J. R.** Cryo-negative staining 16. *Micron.*29[2-3], 145-160. (1998).
- Akoulitchev, S., Chuikov, S., and Reinberg, D.** TFIIH is negatively regulated by cdk8-containing mediator complexes 4. *Nature*407[6800], 102-106. (7-9-2000).
- Andersen, G., Poterszman, A., Egly, J. M., Moras, D., and Thierry, J. C.** The crystal structure of human cyclin H 10. *FEBS Lett.*397[1], 65-69. (11-11-1996).
- Aparicio, S. A. J. R.** How to count ... human genes. *Nature Genetics*25[2], 129-130. (2000).
- Araujo, S. J., Nigg, E. A., and Wood, R. D.** Strong functional interactions of TFIIH with XPC and XPG in human DNA nucleotide excision repair, without a preassembled repairosome 10. *Mol.Cell Biol.*21[7], 2281-2291. (2001).
- Augustin, M. A., Huber, R., and Kaiser, J. T.** Crystal structure of a DNA-dependent RNA polymerase (DNA primase) 1. *Nat.Struct.Biol.*8[1], 57-61. (2001).

B.

- Ban, N., Freeborn, B., Nissen, P., Penczek, P., Grassucci, R. A., Sweet, R., Frank, J., Moore, P. B., and Steitz, T. A.** A 9 Å resolution X-ray crystallographic map of the large ribosomal subunit 10. *Cell*93[7], 1105-1115. (26-6-1998).
- Ban, N., Nissen, P., Hansen, J., Moore, P. B., and Steitz, T. A.** The complete atomic structure of the large ribosomal subunit at 2.4 Å resolution 8. *Science*289[5481], 905-920. (11-8-2000).
- Bartkova, J., Zemanova, M., and Bartek, J.** Expression of CDK7/CAK in normal and tumor cells of diverse histogenesis, cell-cycle position and differentiation 1. *Int.J.Cancer*66[6], 732-737. (11-6-1996).
- Bastien, J., Adam-Stitah, S., Riedl, T., Egly, J. M., Chambon, P., and Rochette-Egly, C.** TFIIH interacts with the retinoic acid receptor gamma and phosphorylates its AF-1-activating domain through cdk7 1. *J.Biol.Chem.*275[29], 21896-21904. (21-7-2000a).
- Bell, S. D., Botting, C. H., Wardleworth, B. N., Jackson, S. P., and White, M. F.** The interaction of Alba, a conserved archaeal chromatin protein, with Sir2 and its regulation by acetylation 10. *Science*296[5565], 148-151. (5-4-2002).
- Bell, S. D. and Jackson, S. P.** Transcription and translation in Archaea: a mosaic of eukaryal and bacterial features 21. *Trends Microbiol.*6[6], 222-228. (1998a).
- Bell, S. D. and Jackson, S. P.** Transcription in Archaea 19. *Cold Spring Harb.Symp.Quant.Biol.*63, 41-51. (1998b).
- Bell, S. D. and Jackson, S. P.** Mechanism and regulation of transcription in archaea 14. *Curr.Opin.Microbiol.*4[2], 208-213. (2001).
- Bell, S. D., Kosa, P. L., Sigler, P. B., and Jackson, S. P.** Orientation of the transcription preinitiation complex in archaea 18. *Proc.Natl.Acad.Sci.U.S.A*96[24], 13662-13667. (23-11-1999).

- Bentley, D. R.** Decoding the human genome sequence. *Human Molecular Genetics*9[16], 2353-2358. (2000a).
- Bentley, D. R.** The human genome project - An overview. *Medicinal Research Reviews*20[3], 189-196. (2000b).
- Bergmann, E. and Egly, J. M.** Trichothiodystrophy, a transcription syndrome 16. *Trends Genet.*17[5], 279-286. (2001).
- Berneburg, M. and Lehmann, A. R.** Xeroderma pigmentosum and related disorders: defects in DNA repair and transcription 14. *Adv.Genet.*43, 71-102. (2001).
- Bhatia, P. K., Wang, Z., and Friedberg, E. C.** DNA repair and transcription 1. *Curr.Opin.Genet.Dev.*6[2], 146-150. (1996).
- Birck, C., Poch, O., Romier, C., Ruff, M., Mengus, G., Lavigne, A. C., Davidson, I., and Moras, D.** Human TAF(II)28 and TAF(II)18 interact through a histone fold encoded by atypical evolutionary conserved motifs also found in the SPT3 family 1. *Cell*94[2], 239-249. (24-7-1998).
- Bird, G., Zorio, D. A., and Bentley, D. L.** RNA Polymerase II Carboxy-Terminal Domain Phosphorylation Is Required for Cotranscriptional Pre-mRNA Splicing and 3'-End Formation 3. *Mol.Cell Biol.*24[20], 8963-8969. (2004).
- Biros, I. and Forrest, S.** Spinal muscular atrophy: untangling the knot? 2. *J.Med.Genet.*36[1], 1-8. (1999).
- Biros, I., I, Forshaw, K., Sheffield, L. J., Kornberg, A. J., and Forrest, S.** Spinal Muscular Atrophies: An Ongoing Diagnostic Dilemma? 1. *Mol.Diagn.*2[4], 241-249. (1997).
- Bischler, N., Brino, L., Carles, C., Riva, M., Tschochner, H., Mallouh, V., and Schultz, P.** Localization of the yeast RNA polymerase I-specific subunits 1. *EMBO J.*21[15], 4136-4144. (1-8-2002). **Boguszewska-Chachulska, A. M., Krawczyk, M., Stankiewicz, A., Gozdek, A., Haenni, A. L., and Strokovskaya, L.** Direct

- fluorometric measurement of hepatitis C virus helicase activity 1. *FEBS Lett.*567[2-3], 253-258. (4-6-2004).
- Boisset, N. and Lamy, J. N.** Immunoelectron microscopy and image processing for epitope mapping 39. *Methods Enzymol.*203, 274-295. (1991).
- Bootsma, D., Weeda, G., Vermeulen, W., van Vuuren, H., Troelstra, C., van der Spek P., and Hoeijmakers, J.** Nucleotide excision repair syndromes: molecular basis and clinical symptoms. *Philos.Trans.R.Soc.Lond B Biol.Sci.*347[1319], 75-81. (30-1-1995).
- Borukhov, S. and Nudler, E.** RNA polymerase holoenzyme: structure, function and biological implications 17. *Curr.Opin.Microbiol.*6[2], 93-100. (2003).
- Botta, E., Nardo, T., Lehmann, A. R., Egly, J. M., Pedrini, A. M., and Stefanini, M.** Reduced level of the repair/transcription factor TFIIH in trichothiodystrophy 9. *Hum.Mol.Genet.*11[23], 2919-2928. (1-11-2002).
- Boulikas, T.** Nuclear import of protein kinases and cyclins 3. *J.Cell Biochem.*60[1], 61-82. (1996).
- Bradsher, J., Coin, F., and Egly, J. M.** Distinct roles for the helicases of TFIIH in transcript initiation and promoter escape 2. *J.Biol.Chem.*275[4], 2532-2538. (28-1-2000a).
- Brenner, S. and HORNE, R. W.** A negative staining method for high resolution electron microscopy of viruses 2. *Biochim.Biophys.Acta*34, 103-110. (1959).
- Broder, S. and Venter, J. C.** Sequencing the entire genomes of free-living organisms: The foundation of pharmacology in the new millennium. *Annual Review of Pharmacology and Toxicology*40, 97-132. (2000).
- Brown, N. R., Noble, M. E., Endicott, J. A., Garman, E. F., Wakatsuki, S., Mitchell, E., Rasmussen, B., Hunt, T., and Johnson, L. N.** The crystal structure of cyclin A 12. *Structure*3[11], 1235-1247. (15-11-1995).



**Bushnell, D. A., Bamdad, C., and Kornberg, R. D.** A minimal set of RNA polymerase II transcription protein interactions 10. *J.Biol.Chem.*271[33], 20170-20174. (16-8-1996).

**Byrd, A. K. and Raney, K. D.** Protein displacement by an assembly of helicase molecules aligned along single-stranded DNA 1. *Nat.Struct.Mol.Biol.*11[6], 531-538. (2004).

C.

**Cabrejos, M. E., Allende, C. C., and Maldonado, E.** Effects of phosphorylation by protein kinase CK2 on the human basal components of the RNA polymerase II transcription machinery 16. *J.Cell Biochem.*93[1], 2-10. (1-9-2004).

**Chalkley, G. E. and Verrijzer, C. P.** DNA binding site selection by RNA polymerase II TAFs: a TAF(II)250-TAF(II)150 complex recognizes the initiator 1. *EMBO J.*18[17], 4835-4845. (1-9-1999).

**Chang and Roger D. Kornberg.** Electron Crystal Structure of the Transcription Factor and DNA Repair Complex, Core TFIIH. *Cell.*;102[5],609-13. (2000).

**Chen, D., Riedl, T., Washbrook, E., Pace, P. E., Coombes, R. C., Egly, J. M., and Ali, S.** Activation of estrogen receptor alpha by S118 phosphorylation involves a ligand-dependent interaction with TFIIH and participation of CDK7 10. *Mol.Cell*6[1], 127-137. (2000).

**Chen, J., Larochelle, S., Li, X., and Suter, B.** Xpd/Ercc2 regulates CAK activity and mitotic progression 2. *Nature*424[6945], 228-232. (10-7-2003).

**Chiurazzi, P., Pomponi, M. G., Willemsen, R., Oostra, B. A., and Neri, G.** In vitro reactivation of the FMR1 gene involved in fragile X syndrome 5. *Hum.Mol.Genet.*7[1], 109-113. (1998).

**Clarkson, S. G.** The XPG story 12. *Biochimie*85[11], 1113-1121. (2003).

- Cobb, B. S., Morales-Alcelay, S., Kleiger, G., Brown, K. E., Fisher, A. G., and Smale, S. T.** Targeting of Ikaros to pericentromeric heterochromatin by direct DNA binding. *Genes Dev.*14[17], 2146-2160. (1-9-2000).
- Coin, F., Bergmann, E., Tremeau-Bravard, A., and Egly, J. M.** Mutations in XPB and XPD helicases found in xeroderma pigmentosum patients impair the transcription function of TFIIH. *EMBO J.*18[5], 1357-1366. (1-3-1999a).
- Coin, F. and Egly, J. M.** Ten years of TFIIH. *Cold Spring Harb. Symp. Quant. Biol.*63, 105-110. (1998).
- Coin, F., Marinoni, J. C., and Egly, J. M.** Mutations in XPD helicase prevent its interaction and regulation by p44, another subunit of TFIIH, resulting in Xeroderma pigmentosum (XP) and trichothiodystrophy (TTD) phenotypes. *Pathol. Biol. (Paris)*46[9], 679-680. (1998a).
- Coin, F., Marinoni, J. C., Rodolfo, C., Fribourg, S., Pedrini, A. M., and Egly, J. M.** Mutations in the XPD helicase gene result in XP and TTD phenotypes, preventing interaction between XPD and the p44 subunit of TFIIH 5. *Nat. Genet.*20[2], 184-188. (1998b).
- Conaway, R. C., Reines, D., Garrett, K. P., Powell, W., and Conaway, J. W.** Purification of RNA polymerase II general transcription factors from rat liver. *Methods Enzymol.*273, 194-207. (1996).
- Cordin, O., Tanner, N. K., Doere, M., Linder, P., and Banroques, J.** The newly discovered Q motif of DEAD-box RNA helicases regulates RNA-binding and helicase activity 2. *EMBO J.*23[13], 2478-2487. (7-7-2004).
- Cramer, P., Bushnell, D. A., and Kornberg, R. D.** Structural basis of transcription: RNA polymerase II at 2.8 angstrom resolution 16. *Science*292[5523], 1863-1876. (8-6-2001).
- Crowther, R. A.** Three-dimensional reconstruction and the architecture of spherical viruses 30. *Endeavour*30[111], 124-129. (1971).

**Crowther, R. A., Amos, L. A., Finch, J. T., De Rosier, D. J., and Klug, A.** Three dimensional reconstructions of spherical viruses by fourier synthesis from electron micrographs 1. *Nature*226[5244], 421-425. (2-5-1970).

**Cyrklaff, M., Adrian, M., and Dubochet, J.** Evaporation during preparation of unsupported thin vitrified aqueous layers for cryo-electron microscopy 1. *J.Electron Microscop.Tech.*16[4], 351-355. (1990).

D.

**Dahlberg, A. E.** Ribosome structure. The ribosome in action 5. *Science*292[5518], 868-869. (4-5-2001).

**Darst, S. A., Opalka, N., Chacon, P., Polyakov, A., Richter, C., Zhang, G., and Wriggers, W.** Conformational flexibility of bacterial RNA polymerase 7. *Proc.Natl.Acad.Sci.U.S.A*99[7], 4296-4301. (2-4-2002).

**Davie, J. R. and Chadee, D. N.** Regulation and regulatory parameters of histone modifications 3. *J.Cell Biochem.Suppl*30-31, 203-213. (1998).

**De Carlo, S., Carles, C., Riva, M., and Schultz, P.** Cryo-negative staining reveals conformational flexibility within yeast RNA polymerase I 6. *J.Mol.Biol.*329[5], 891-902. (20-6-2003).

**De Carlo, S., El Bez, C., Alvarez-Rua, C., Borge, J., and Dubochet, J.** Cryo-negative staining reduces electron-beam sensitivity of vitrified biological particles 7. *J.Struct.Biol.*138[3], 216-226. (2002).

**de Laat, W. L., Jaspers, N. G., and Hoeijmakers, J. H.** Molecular mechanism of nucleotide excision repair 11. *Genes Dev.*13[7], 768-785. (1-4-1999).

**Della, Coletta L., Rolig, R. L., Fossey, S., Morizot, D. C., Nairn, R. S., and Walter, R. B.** Characterization of the Xiphophorus fish (Teleostei: Poeciliidae) ERCC2/XPD locus 1. *Genomics*26[1], 70-76. (1-3-1995).

**DeRosier, D. J. and Moore, P. B.** Reconstruction of three-dimensional images from electron micrographs of structures with helical symmetry. *J.Mol.Biol.*52[2], 355-369. (14-9-1970).

**Devault, A., Martinez, A. M., Fesquet, D., Labbe, J. C., Morin, N., Tassan, J. P., Nigg, E. A., Cavadore, J. C., and Doree, M.** MAT1 ('menage a trois') a new RING finger protein subunit stabilizing cyclin Hcdk7 complexes in starfish and *Xenopus* CAK 2. *EMBO J.*14[20], 5027-5036. (16-10-1995).

**Dionne, I., Nookala, R. K., Jackson, S. P., Doherty, A. J., and Bell, S. D.** A heterotrimeric PCNA in the hyperthermophilic archaeon *Sulfolobus solfataricus* 1. *Mol.Cell*11[1], 275-282. (2003).

**Drapkin, R., Le Roy, G., Cho, H., Akoulitchev, S., and Reinberg, D.** Human cyclin-dependent kinase-activating kinase exists in three distinct complexes 2. *Proc.Natl.Acad.Sci.U.S.A*93[13], 6488-6493. (25-6-1996).

**Drapkin, R. and Reinberg, D.** The multifunctional TFIIH complex and transcriptional control. *Trends Biochem.Sci.*19[11], 504-508. (1994).

**Dubochet, J., Booy, F. P., Freeman, R., Jones, A. V., and Walter, C. A.** Low temperature electron microscopy 51. *Annu.Rev.Biophys.Bioeng.*10, 133-149. (1981).

E.

**Eberharter, A. and Becker, P. B.** Histone acetylation: a switch between repressive and permissive chromatin. Second in review series on chromatin dynamics 2. *EMBO Rep.*3[3], 224-229. (2002).

**Egelman, E. H., Yu, X., Wild, R., Hingorani, M. M., and Patel, S. S.** Bacteriophage T7 helicase/primase proteins form rings around single-stranded DNA that suggest a general structure for hexameric helicases 13. *Proc.Natl.Acad.Sci.U.S.A*92[9], 3869-3873. (25-4-1995).

- Eggleston, A. K., O'Neill, T. E., Bradbury, E. M., and Kowalczykowski, S. C.** Unwinding of nucleosomal DNA by a DNA helicase 6. *J.Biol.Chem.*270[5], 2024-2031. (3-2-1995).
- Ellison, V. and Stillman, B.** Biochemical characterization of DNA damage checkpoint complexes: clamp loader and clamp complexes with specificity for 5' recessed DNA 1. *PLoS.Biol.*1[2], E33. (2003).
- Emami, K. H., Burke, T. W., and Smale, S. T.** Sp1 activation of a TATA-less promoter requires a species-specific interaction involving transcription factor IID 1. *Nucleic Acids Res.*26[3], 839-846. (1-2-1998).
- Enoru-Eta, J., Gigot, D., Thia-Toong, T. L., Glansdorff, N., and Charlier, D.** Purification and characterization of Sa-Irp, a DNA-binding protein from the extreme thermoacidophilic archaeon *Sulfolobus acidocaldarius* homologous to the bacterial global transcriptional regulator Lrp 1. *J.Bacteriol.*182[13], 3661-3672. (2000).
- Enzlin, J. H. and Scharer, O. D.** The active site of the DNA repair endonuclease XPF-ERCC1 forms a highly conserved nuclease motif 2. *EMBO J.*21[8], 2045-2053. (15-4-2002).
- Erdem, H., Pehlivan, S., Topaloglu, H., and Ozguc, M.** Deletion analysis in Turkish patients with spinal muscular atrophy 2. *Brain Dev.*21[2], 86-89. (1999).
- Evans, E., Moggs, J. G., Hwang, J. R., Egly, J. M., and Wood, R. D.** Mechanism of open complex and dual incision formation by human nucleotide excision repair factors 18. *EMBO J.*16[21], 6559-6573. (3-11-1997).

## F.

- Fan, X., Shi, H., Adelman, K., and Lis, J. T.** Probing TBP interactions in transcription initiation and reinitiation with RNA aptamers that act in distinct modes 6. *Proc.Natl.Acad.Sci.U.S.A*101[18], 6934-6939. (4-5-2004).
- Feaver, W. J., Henry, N. L., Wang, Z., Wu, X., Svejstrup, J. Q., Bushnell, D. A., Friedberg, E. C., and Kornberg, R. D.** Genes for Tfb2, Tfb3, and Tfb4 subunits of

yeast transcription/repair factor IIH. Homology to human cyclin-dependent kinase activating kinase and IIH subunits 6. *J.Biol.Chem.*272[31], 19319-19327. (1-8-1997).

**Feaver, W. J., Huang, W., and Friedberg, E. C.** The TFB4 subunit of yeast TFIIH is required for both nucleotide excision repair and RNA polymerase II transcription 2. *J.Biol.Chem.*274[41], 29564-29567. (8-10-1999).

**Feaver, W. J., Huang, W., Gileadi, O., Myers, L., Gustafsson, C. M., Kornberg, R. D., and Friedberg, E. C.** Subunit interactions in yeast transcription/repair factor TFIIH. Requirement for Tfb3 subunit in nucleotide excision repair 23. *J.Biol.Chem.*275[8], 5941-5946. (25-2-2000).

**Feaver, W. J., Svejstrup, J. Q., Henry, N. L., and Kornberg, R. D.** Relationship of CDK-activating kinase and RNA polymerase II CTD kinase TFIIH/TFIIK. *Cell*79[6], 1103-1109. (16-12-1994a).

**Felsenfeld, G., Boyes, J., Chung, J., Clark, D., and Studitsky, V.** Chromatin structure and gene expression 13. *Proc.Natl.Acad.Sci.U.S.A*93[18], 9384-9388. (3-9-1996).

**Felsenfeld, G., Clark, D., and Studitsky, V.** Transcription through nucleosomes. *Biophys.Chem.*86[2-3], 231-237. (30-8-2000).

**Fiorentino, G., Cannio, R., Rossi, M., and Bartolucci, S.** Transcriptional regulation of the gene encoding an alcohol dehydrogenase in the archaeon *Sulfolobus solfataricus* involves multiple factors and control elements 10. *J.Bacteriol.*185[13], 3926-3934. (2003).

**Fischer, L., Gerard, M., Chalut, C., Lutz, Y., Humbert, S., Kanno, M., Chambon, P., and Egly, J. M.** Cloning of the 62-kilodalton component of basic transcription factor BTF2 1. *Science*257[5075], 1392-1395. (4-9-1992).

**Fisher, R. P. and Morgan, D. O.** A novel cyclin associates with MO15/CDK7 to form the CDK-activating kinase 3. *Cell*78[4], 713-724. (26-8-1994).

- Fitch, M. E., Nakajima, S., Yasui, A., and Ford, J. M.** In vivo recruitment of XPC to UV-induced cyclobutane pyrimidine dimers by the DDB2 gene product 1. *J.Biol.Chem.*278[47], 46906-46910. (21-11-2003).
- Follows, G. A., Tagoh, H., Lefevre, P., Morgan, G. J., and Bonifer, C.** Differential transcription factor occupancy but evolutionarily conserved chromatin features at the human and mouse M-CSF (CSF-1) receptor loci 10. *Nucleic Acids Res.*31[20], 5805-5816. (15-10-2003).
- Forget, D., Langelier, M. F., Therien, C., Trinh, V., and Coulombe, B.** Photo-cross-linking of a purified preinitiation complex reveals central roles for the RNA polymerase II mobile clamp and TFIIE in initiation mechanisms 1. *Mol.Cell Biol.*24[3], 1122-1131. (2004).
- Forterre, P., Confalonieri, F., and Knapp, S.** Identification of the gene encoding archeal-specific DNA-binding proteins of the Sac10b family 11. *Mol.Microbiol.*32[3], 669-670. (1999).
- Frank, J. and Radermacher, M.** Three-dimensional reconstruction of single particles negatively stained or in vitreous ice. *Ultramicroscopy*46[1-4], 241-262. (1992).
- Fribourg, S., Kellenberger, E., Rogniaux, H., Poterszman, A., Van Dorselaer, A., Thierry, J. C., Egly, J. M., Moras, D., and Kieffer, B.** Structural characterization of the cysteine-rich domain of TFIIH p44 subunit. *J.Biol.Chem.*275[41], 31963-31971. (13-10-2000).
- Fukuda, A., Tokonabe, S., Hamada, M., Matsumoto, M., Tsukui, T., Nogi, Y., and Hisatake, K.** Alleviation of PC4-mediated transcriptional repression by the ERCC3 helicase activity of general transcription factor TFIIH 2. *J.Biol.Chem.*278[17], 14827-14831. (25-4-2003).
- Furukawa, T. and Tanese, N.** Assembly of partial TFIID complexes in mammalian cells reveals distinct activities associated with individual TATA box-binding protein-associated factors. *J.Biol.Chem.*275[38], 29847-29856. (22-9-2000).

## G.

- Gaillard, P. H. and Wood, R. D.** Activity of individual ERCC1 and XPF subunits in DNA nucleotide excision repair 13. *Nucleic Acids Res.*29[4], 872-879. (15-2-2001).
- Gaston, K. and Jayaraman, P. S.** Transcriptional repression in eukaryotes: repressors and repression mechanisms 20. *Cell Mol.Life Sci.*60[4], 721-741. (2003).
- Geiger, J. H., Hahn, S., Lee, S., and Sigler, P. B.** Crystal structure of the yeast TFIIA/TBP/DNA complex. *Science*272[5263], 830-836. (10-5-1996).
- Gervais, V., Busso, D., Wasielewski, E., Poterszman, A., Egly, J. M., Thierry, J. C., and Kieffer, B.** Solution structure of the N-terminal domain of the human TFIIH MAT1 subunit: new insights into the RING finger family 3. *J.Biol.Chem.*276[10], 7457-7464. (9-3-2001).
- Gervais, V., Lamour, V., Jawhari, A., Frindel, F., Wasielewski, E., Dubaele, S., Egly, J. M., Thierry, J. C., Kieffer, B., and Poterszman, A.** TFIIH contains a PH domain involved in DNA nucleotide excision repair 1. *Nat.Struct.Mol.Biol.*11[7], 616-622. (2004a).
- Giglia-Mari, G., Coin, F., Ranish, J. A., Hoogstraten, D., Theil, A., Wijgers, N., Jaspers, N. G., Raams, A., Argentini, M., van der Spek, P. J., Botta, E., Stefanini, M., Egly, J. M., Aebersold, R., Hoeijmakers, J. H., and Vermeulen, W.** A new, tenth subunit of TFIIH is responsible for the DNA repair syndrome trichothiodystrophy group A 1. *Nat.Genet.*36[7], 714-719. (2004c).
- Gill, G.** Regulation of the initiation of eukaryotic transcription. *Essays Biochem.*37, 33-43. (2001a).
- Golas, M. M., Sander, B., Will, C. L., Luhrmann, R., and Stark, H.** Molecular architecture of the multiprotein splicing factor SF3b 3. *Science*300[5621], 980-984. (9-5-2003).



**Gorbalenya, A. E., Koonin, E. V., Donchenko, A. P., and Blinov, V. M.** Two related superfamilies of putative helicases involved in replication, recombination, repair and expression of DNA and RNA genomes 14. *Nucleic Acids Res.*17[12], 4713-4730. (26-6-1989).

**Graumann, J., Dunipace, L. A., Seol, J. H., McDonald, W. H., Yates, J. R., III, Wold, B. J., and Deshaies, R. J.** Applicability of tandem affinity purification MudPIT to pathway proteomics in yeast 3. *Mol.Cell Proteomics.*3[3], 226-237. (2004).

**Groft, C. M., Uljon, S. N., Wang, R., and Werner, M. H.** Structural homology between the Rap30 DNA-binding domain and linker histone H5: implications for preinitiation complex assembly 1. *Proc.Natl.Acad.Sci.U.S.A*95[16], 9117-9122. (4-8-1998).

**Gruetzner, F., Roest-Crollius, H., Lutjens, G., Barnes, D. W., Ropers, H. H., Weissenbach, J., and Haaf, T.** Comparative gene and genome mapping in *Tetraodon nigroviridis* and *Fugu rubripes*. *Cytogenetics and Cell Genetics*85[1-2], 32. (1999).

**Guagliardi, A., Mancusi, L., and Rossi, M.** Reversion of protein aggregation mediated by Sso7d in cell extracts of *Sulfolobus solfataricus* 5. *Biochem.J.*381[Pt 1], 249-255. (1-7-2004).

**Gulyas, K. D. and Donahue, T. F.** SSL2, a suppressor of a stem-loop mutation in the HIS4 leader encodes the yeast homolog of human ERCC-3 4. *Cell*69[6], 1031-1042. (12-6-1992).

**Guzder, S. N., Habraken, Y., Sung, P., Prakash, L., and Prakash, S.** RAD26, the yeast homolog of human Cockayne's syndrome group B gene, encodes a DNA-dependent ATPase 2. *J.Biol.Chem.*271[31], 18314-18317. (2-8-1996).

H.

**Haile, D. T. and Parvin, J. D.** Activation of transcription in vitro by the BRCA1 carboxyl-terminal domain 1. *J.Biol.Chem.*274[4], 2113-2117. (22-1-1999).

- Harauz, G., Borland, L., Bahr, G. F., Zeitler, E., and van Heel, M.** Three-dimensional reconstruction of a human metaphase chromosome from electron micrographs 5. *Chromosoma*95[5], 366-374. (1987).
- Hashimoto, H., Nishioka, M., Fujiwara, S., Takagi, M., Imanaka, T., Inoue, T., and Kai, Y.** Crystal structure of DNA polymerase from hyperthermophilic archaeon *Pyrococcus kodakaraensis* KOD1 8. *J.Mol.Biol.*306[3], 469-477. (23-2-2001).
- He, H. and Lehming, N.** Global effects of histone modifications. *Brief.Funct.Genomic.Proteomic.*2[3], 234-243. (2003).
- Henning, K. A., Li, L., Iyer, N., McDaniel, L. D., Reagan, M. S., Legerski, R., Schultz, R. A., Stefanini, M., Lehmann, A. R., and Mayne, L. V.** The Cockayne syndrome group A gene encodes a WD repeat protein that interacts with CSB protein and a subunit of RNA polymerase II TFIIH 1. *Cell*82[4], 555-564. (25-8-1995).
- Hickman, A. B., Ronning, D. R., Kotin, R. M., and Dyda, F.** Structural unity among viral origin binding proteins: crystal structure of the nuclease domain of adeno-associated virus Rep 2. *Mol.Cell*10[2], 327-337. (2002).
- Hoffmann, A., Chiang, C. M., Oelgeschlager, T., Xie, X., Burley, S. K., Nakatani, Y., and Roeder, R. G.** A histone octamer-like structure within TFIID 1. *Nature*380[6572], 356-359. (28-3-1996).
- Hoogstraten, D., Nigg, A. L., Heath, H., Mullenders, L. H., van Driel, R., Hoeijmakers, J. H., Vermeulen, W., and Houtsmuller, A. B.** Rapid switching of TFIIH between RNA polymerase I and II transcription and DNA repair in vivo 2. *Mol.Cell*10[5], 1163-1174. (2002).
- Horn, E. J., Albor, A., Liu, Y., El Hizawi, S., Vanderbeek, G. E., Babcock, M., Bowden, G. T., Hennings, H., Lozano, G., Weinberg, W. C., and Kulesz-Martin, M.** RING protein Trim32 associated with skin carcinogenesis has anti-apoptotic and E3-ubiquitin ligase properties 1. *Carcinogenesis*25[2], 157-167. (2004).
- Humbert, S., van Vuuren, H., Lutz, Y., Hoeijmakers, J. H., Egly, J. M., and Moncollin, V.** p44 and p34 subunits of the BTF2/TFIIH transcription factor have homologies

with SSL1, a yeast protein involved in DNA repair. *EMBO J.*13[10], 2393-2398. (15-5-1994).

**Hwang, J. R., Moncollin, V., Vermeulen, W., Seroz, T., van Vuuren, H., Hoeijmakers, J. H., and Egly, J. M.** A 3' --> 5' XPB helicase defect in repair/transcription factor TFIIH of xeroderma pigmentosum group B affects both DNA repair and transcription 3. *J.Biol.Chem.*271[27], 15898-15904. (5-7-1996a).

I.

**Iben, S., Tschochner, H., Bier, M., Hoogstraten, D., Hozak, P., Egly, J. M., and Grummt, I.** TFIIH plays an essential role in RNA polymerase I transcription 2. *Cell*109[3], 297-306. (3-5-2002).

**Iftode, C., Daniely, Y., and Borowiec, J. A.** Replication protein A (RPA): the eukaryotic SSB 16. *Crit Rev.Biochem.Mol.Biol.*34[3], 141-180. (1999).

**Ikegami, T., Kuraoka, I., Saijo, M., Kodo, N., Kyogoku, Y., Morikawa, K., Tanaka, K., and Shirakawa, M.** Solution structure of the DNA- and RPA-binding domain of the human repair factor XPA 2. *Nat.Struct.Biol.*5[8], 701-706. (1998).

**Inamoto, S., Segil, N., Pan, Z. Q., Kimura, M., and Roeder, R. G.** The cyclin-dependent kinase-activating kinase (CAK) assembly factor, MAT1, targets and enhances CAK activity on the POU domains of octamer transcription factors 1. *J.Biol.Chem.*272[47], 29852-29858. (21-11-1997).

**Itin, P. H., Sarasin, A., and Pittelkow, M. R.** Trichothiodystrophy: update on the sulfur-deficient brittle hair syndromes 15. *J.Am.Acad.Dermatol.*44[6], 891-920. (2001).

**Ito, S., Takeyama, K., Yamamoto, A., Sawatsubashi, S., Shirode, Y., Kouzmenko, A., Tabata, T., and Kato, S.** In vivo potentiation of human oestrogen receptor alpha by Cdk7-mediated phosphorylation 1. *Genes Cells*9[10], 983-992. (2004).

**Iyer, N., Reagan, M. S., Wu, K. J., Canagarajah, B., and Friedberg, E. C.** Interactions involving the human RNA polymerase II transcription/nucleotide excision repair

complex TFIIH, the nucleotide excision repair protein XPG, and Cockayne syndrome group B (CSB) protein 2. *Biochemistry*35[7], 2157-2167. (20-2-1996).

J.

**Jacobson, R. H., Ladurner, A. G., King, D. S., and Tjian, R.** Structure and function of a human TAFII250 double bromodomain module 6. *Science*288[5470], 1422-1425. (26-5-2000).

**Jawhari, A., Laine, J. P., Dubaele, S., Lamour, V., Poterszman, A., Coin, F., Moras, D., and Egly, J. M.** p52 Mediates XPB function within the transcription/repair factor TFIIH 2. *J.Biol.Chem.*277[35], 31761-31767. (30-8-2002a).

**Jawhari, A., Uhring, M., Crucifix, C., Fribourg, S., Schultz, P., Poterszman, A., Egly, J. M., and Moras, D.** Expression of FLAG fusion proteins in insect cells: application to the multi-subunit transcription/DNA repair factor TFIIH 1. *Protein Expr.Purif.*24[3], 513-523. (2002b).

**Jesior, J. C. and Wade, R. H.** Electron-irradiation-induced flattening of negatively stained 2D protein crystals 73. *Ultramicroscopy*21[4], 313-319. (1987).

**Johnson, C. A., O'Neill, L. P., Mitchell, A., and Turner, B. M.** Distinctive patterns of histone H4 acetylation are associated with defined sequence elements within both heterochromatic and euchromatic regions of the human genome. *Nucleic Acids Res.*26[4], 994-1001. (15-2-1998).

**Johnson, K. D., Grass, J. A., Park, C., Im, H., Choi, K., and Bresnick, E. H.** Highly restricted localization of RNA polymerase II within a locus control region of a tissue-specific chromatin domain 13. *Mol.Cell Biol.*23[18], 6484-6493. (2003).

**Joseph, S.** After the ribosome structure: how does translocation work? 1. *RNA*9[2], 160-164. (2003).

K.

- Kaina, B., Ochs, K., Grosch, S., Fritz, G., Lips, J., Tomicic, M., Dunkern, T., and Christmann, M.** BER, MGMT, and MMR in defense against alkylation-induced genotoxicity and apoptosis 20. *Prog.Nucleic Acid Res.Mol.Biol.*68, 41-54. (2001).
- Kandler, O.** [Festival lecture. The position of microorganisms in the global phylogenetic system of three domains] 3. *Mycoses*37 Suppl 1, 13-27. (1994).
- Kang, Z., Janne, O. A., and Palvimo, J. J.** Coregulator Recruitment and Histone Modifications in Transcriptional Regulation by the Androgen Receptor 1. *Mol.Endocrinol.* (12-8-2004).
- Kaufmann, J., Ahrens, K., Koop, R., Smale, S. T., and Muller, R.** CIF150, a human cofactor for transcription factor IID-dependent initiator function 1. *Mol.Cell Biol.*18[1], 233-239. (1998).
- Kaufmann, J. and Smale, S. T.** Direct recognition of initiator elements by a component of the transcription factor IID complex 11. *Genes Dev.*8[7], 821-829. (1-4-1994).
- Kaufmann, J., Verrijzer, C. P., Shao, J., and Smale, S. T.** CIF, an essential cofactor for TFIID-dependent initiator function. *Genes Dev.*10[7], 873-886. (1-4-1996).
- Kawaoka, J., Jankowsky, E., and Pyle, A. M.** Backbone tracking by the SF2 helicase NPH-II 1. *Nat.Struct.Mol.Biol.*11[6], 526-530. (2004).
- Keriel, A., Stary, A., Sarasin, A., Rochette-Egly, C., and Egly, J. M.** XPD mutations prevent TFIIH-dependent transactivation by nuclear receptors and phosphorylation of RARalpha 2. *Cell*109[1], 125-135. (5-4-2002a).
- Kerr, I. D., Wadsworth, R. I., Cubeddu, L., Blankenfeldt, W., Naismith, J. H., and White, M. F.** Insights into ssDNA recognition by the OB fold from a structural and thermodynamic study of Sulfolobus SSB protein 1. *EMBO J.*22[11], 2561-2570. (2-6-2003).

- Kim, J. and Iyer, V. R.** Global role of TATA box-binding protein recruitment to promoters in mediating gene expression profiles 5. *Mol.Cell Biol.*24[18], 8104-8112. (2004).
- Kim, J. L., Morgenstern, K. A., Lin, C., Fox, T., Dwyer, M. D., Landro, J. A., Chambers, S. P., Markland, W., Lepre, C. A., O'Malley, E. T., Harbeson, S. L., Rice, C. M., Murcko, M. A., Caron, P. R., and Thomson, J. A.** Crystal structure of the hepatitis C virus NS3 protease domain complexed with a synthetic NS4A cofactor peptide 12. *Cell*87[2], 343-355. (18-10-1996a).
- Kim, J. L., Nikolov, D. B., and Burley, S. K.** Co-crystal structure of TBP recognizing the minor groove of a TATA element 13. *Nature*365[6446], 520-527. (7-10-1993a).
- Kim, K. K., Chamberlin, H. M., Morgan, D. O., and Kim, S. H.** Three-dimensional structure of human cyclin H, a positive regulator of the CDK-activating kinase5. *Nat.Struct.Biol.*3[10], 849-855. (1996b).
- Kim, Y., Geiger, J. H., Hahn, S., and Sigler, P. B.** Crystal structure of a yeast TBP/TATA-box complex 14. *Nature*365[6446], 512-520. (7-10-1993b).
- Kirchner, J. M., Salazar, E. P., Lamerdin, J. E., Montgomery, M. A., Carrano, A. V., and Weber, C. A.** Cloning and molecular characterization of the Chinese hamster ERCC2 nucleotide excision repair gene 18. *Genomics*23[3], 592-599. (1994).
- Kopf E, Plassat JL, Vivat V, de The H, Chambon P, Rochette-Egly C.** Dimerization with retinoid X receptors and phosphorylation modulate the retinoic acid-induced degradation of retinoic acid receptors alpha and gamma through the ubiquitin-proteasome pathway. *J.Biol.Chem.*275[43], 33280-33288. (27-10-2000).
- Korber, C., English, S., Schwindke, P., Scheiwe, M. W., Rau, G., Hubel, A., and Cravalho, E. G.** Low temperature light microscopy and its application to study freezing in aqueous solutions and biological cell suspensions 1. *J.Microsc.*141 ( Pt 3), 263-276. (1986).

- Korolev, S., Hsieh, J., Gauss, G. H., Lohman, T. M., and Waksman, G.** Major domain swiveling revealed by the crystal structures of complexes of E. coli Rep helicase bound to single-stranded DNA and ADP 3. *Cell*90[4], 635-647. (22-8-1997).
- Kosa, P. F., Ghosh, G., DeDecker, B. S., and Sigler, P. B.** The 2.1-Å crystal structure of an archaeal preinitiation complex: TATA-box-binding protein/transcription factor (II)B core/TATA-box 4. *Proc.Natl.Acad.Sci.U.S.A*94[12], 6042-6047. (10-6-1997).
- Krapp, S., Kelly, G., Reischl, J., Weinzierl, R. O., and Matthews, S.** Eukaryotic RNA polymerase subunit RPB8 is a new relative of the OB family 2. *Nat.Struct.Biol.*5[2], 110-114. (1998).
- Krejci, L. and Sung, P.** RPA not that different from SSB 10. *Structure (Camb.)*10[5], 601-602. (2002).
- Krishna, T. S., Kong, X. P., Gary, S., Burgers, P. M., and Kuriyan, J.** Crystal structure of the eukaryotic DNA polymerase processivity factor PCNA 2. *Cell*79[7], 1233-1243. (30-12-1994).
- Kurdistani, S. K., Tavazoie, S., and Grunstein, M.** Mapping global histone acetylation patterns to gene expression 1. *Cell*117[6], 721-733. (11-6-2004).
- Kwek, K. Y., O'Gorman, W., and Akoulitchev, A.** Transcription meets DNA repair at a PH domain 15. *Nat.Struct.Mol.Biol.*11[7], 588-589. (2004).

L.

**Lander, E. S., Linton, L. M., Birren, B., Nusbaum, C., Zody, M. C., Baldwin, J., Devon, K., Dewar, K., Doyle, M., FitzHugh, W., Funke, R., Gage, D., Harris, K., Heaford, A., Howland, J., Kann, L., Lehoczky, J., LeVine, R., McEwan, P., McKernan, K., Meldrim, J., Mesirov, J. P., Miranda, C., Morris, W., Naylor, J., Raymond, C., Rosetti, M., Santos, R., Sheridan, A., Sougnez, C., Stange-Thomann, N., Stojanovic, N., Subramanian, A., Wyman, D., Rogers, J., Sulston, J., Ainscough, R., Beck, S., Bentley, D., Burton, J., Clee, C., Carter, N., Coulson, A., Deadman, R., Deloukas, P., Dunham, A., Dunham, I., Durbin, R., French, L., Grafham, D., Gregory, S., Hubbard, T., Humphray, S., Hunt, A., Jones, M., Lloyd,**

C., McMurray, A., Matthews, L., Mercer, S., Milne, S., Mullikin, J. C., Mungall, A., Plumb, R., Ross, M., Shownkeen, R., Sims, S., Waterston, R. H., Wilson, R. K., Hillier, L. W., McPherson, J. D., Marra, M. A., Mardis, E. R., Fulton, L. A., Chinwalla, A. T., Pepin, K. H., Gish, W. R., Chissoe, S. L., Wendl, M. C., Delehaunty, K. D., Miner, T. L., Delehaunty, A., Kramer, J. B., Cook, L. L., Fulton, R. S., Johnson, D. L., Minx, P. J., Clifton, S. W., Hawkins, T., Branscomb, E., Predki, P., Richardson, P., Wenning, S., Slezak, T., Doggett, N., Cheng, J. F., Olsen, A., Lucas, S., Elkin, C., Uberbacher, E., Frazier, M., Gibbs, R. A., Muzny, D. M., Scherer, S. E., Bouck, J. B., Sodergren, E. J., Worley, K. C., Rives, C. M., Gorrell, J. H., Metzker, M. L., Naylor, S. L., Kucherlapati, R. S., Nelson, D. L., Weinstock, G. M., Sakaki, Y., Fujiyama, A., Hattori, M., Yada, T., Toyoda, A., Itoh, T., Kawagoe, C., Watanabe, H., Totoki, Y., Taylor, T., Weissenbach, J., Heilig, R., Saurin, W., Artiguenave, F., Brottier, P., Bruls, T., Pelletier, E., Robert, C., Wincker, P., Rosenthal, A., Platzer, M., Nyakatura, G., Taudien, S., Rump, A., Yang, H. M., Yu, J., Wang, J., Huang, G. Y., Gu, J., Hood, L., Rowen, L., Madan, A., Qin, S. Z., Davis, R. W., Federspiel, N. A., Abola, A. P., Proctor, M. J., Myers, R. M., Schmutz, J., Dickson, M., Grimwood, J., Cox, D. R., Olson, M. V., Kaul, R., Raymond, C., Shimizu, N., Kawasaki, K., Minoshima, S., Evans, G. A., Athanasiou, M., Schultz, R., Roe, B. A., Chen, F., Pan, H. Q., Ramser, J., Lehrach, H., Reinhardt, R., McCombie, W. R., de la Bastide, M., Dedhia, N., Blocker, H., Hornischer, K., Nordsiek, G., Agarwala, R., Aravind, L., Bailey, J. A., Bateman, A., Batzoglou, S., Birney, E., Bork, P., Brown, D. G., Burge, C. B., Cerutti, L., Chen, H. C., Church, D., Clamp, M., Copley, R. R., Doerks, T., Eddy, S. R., Eichler, E. E., Furey, T. S., Galagan, J., Gilbert, J. G. R., Harmon, C., Hayashizaki, Y., Haussler, D., Hermjakob, H., Hokamp, K., Jang, W. H., Johnson, L. S., Jones, T. A., Kasif, S., Kasprzyk, A., Kennedy, S., Kent, W. J., Kitts, P., Koonin, E. V., Korf, I., Kulp, D., Lancet, D., Lowe, T. M., McLysaght, A., Mikkelsen, T., Moran, J. V., Mulder, N., Pollara, V. J., Ponting, C. P., Schuler, G., Schultz, J. R., Slater, G., Smit, A. F. A., Stupka, E., Szustakowki, J., Thierry-Mieg, D., Thierry-Mieg, J., Wagner, L., Wallis, J., Wheeler, R., Williams, A., Wolf, Y. I., Wolfe, K. H., Yang, S. P., Yeh, R. F., Collins, F., Guyer, M. S., Peterson, J., Felsenfeld, A., Wetterstrand, K. A., Patrinos, A., and Morgan, M. J. Initial sequencing and analysis of the human genome. *Nature*409[6822], 860-921. (15-2-2001).



- Langelier, M. F., Forget, D., Rojas, A., Porlier, Y., Burton, Z. F., and Coulombe, B.** Structural and functional interactions of transcription factor (TF) IIA with TFIIE and TFIIIF in transcription initiation by RNA polymerase II 2. *J.Biol.Chem.*276[42], 38652-38657. (19-10-2001).
- Le May, N., Dubaele, S., Proietti, De Santis, Billecocq, A., Bouloy, M., and Egly, J. M.** TFIIH transcription factor, a target for the Rift Valley hemorrhagic fever virus. *Cell*116[4], 541-550. (20-2-2004).
- Lee, D. K., Duan, H. O., and Chang, C.** From androgen receptor to the general transcription factor TFIIH. Identification of cdk activating kinase (CAK) as an androgen receptor NH(2)-terminal associated coactivator 5. *J.Biol.Chem.*275[13], 9308-9313. (31-3-2000).
- Lee, D. Y., Teyssier, C., Strahl, B. D., and Stallcup, M. R.** Role of Protein Methylation in Regulation of Transcription. *Endocr.Rev.* (12-10-2004).
- Lehmann, A. R.** DNA repair-deficient diseases, xeroderma pigmentosum, Cockayne syndrome and trichothiodystrophy 2. *Biochimie*85[11], 1101-1111. (2003).
- Lescure, A., Lutz, Y., Eberhard, D., Jacq, X., Krol, A., Grummt, I., Davidson, I., Chambon, P., and Tora, L.** The N-terminal domain of the human TATA-binding protein plays a role in transcription from TATA-containing RNA polymerase II and III promoters 12. *EMBO J.*13[5], 1166-1175. (1-3-1994).
- Levin, M. K. and Patel, S. S.** The helicase from hepatitis C virus is active as an oligomer 6. *J.Biol.Chem.*274[45], 31839-31846. (5-11-1999).
- Levin, M. K., Wang, Y. H., and Patel, S. S.** The functional interaction of the hepatitis C virus helicase molecules is responsible for unwinding processivity 2. *J.Biol.Chem.*279[25], 26005-26012. (18-6-2004).
- Lewis, B. A. and Reinberg, D.** The mediator coactivator complex: functional and physical roles in transcriptional regulation 3. *J.Cell Sci.*116[Pt 18], 3667-3675. (15-9-2003).

- Liang, F., Holt, I., Pertea, G., Karamycheva, S., Salzberg, S. L., and Quackenbush, J.** Gene Index analysis of the human genome estimates approximately 120,000 genes. *Nature Genetics*25[2], 239-240. (2000).
- Liu, D., Ishima, R., Tong, K. I., Bagby, S., Kokubo, T., Muhandiram, D. R., Kay, L. E., Nakatani, Y., and Ikura, M.** Solution structure of a TBP-TAF(II)230 complex: protein mimicry of the minor groove surface of the TATA box unwound by TBP 2. *Cell*94[5], 573-583. (4-9-1998).
- Liu, J., He, L., Collins, I., Ge, H., Libutti, D., Li, J., Egly, J. M., and Levens, D.** The FBP interacting repressor targets TFIIH to inhibit activated transcription 4. *Mol.Cell*5[2], 331-341. (2000).
- Liu, J., Meng, X., and Shen, Z.** Association of human RAD52 protein with transcription factors 1. *Biochem.Biophys.Res.Commun.*297[5], 1191-1196. (11-10-2002).
- Lolli G, Lowe ED, Brown NR, Johnson LN.** The Crystal Structure of Human CDK7 and Its Protein Recognition Properties.12[11], 2067-79. (2004).
- Long, J. J., Leresche, A., Kriwacki, R. W., and Gottesfeld, J. M.** Repression of TFIIH transcriptional activity and TFIIH-associated cdk7 kinase activity at mitosis 1. *Mol.Cell Biol.*18[3], 1467-1476. (1998).
- Lu, H., Zawel, L., Fisher, L., Egly, J. M., and Reinberg, D.** Human general transcription factor IIH phosphorylates the C-terminal domain of RNA polymerase II. *Nature*358[6388], 641-645. (20-8-1992).
- Ludtke, S. J., Chen, D. H., Song, J. L., Chuang, D. T., and Chiu, W.** Seeing GroEL at 6 Å resolution by single particle electron cryomicroscopy 1. *Structure (Camb.)*12[7], 1129-1136. (2004).
- Ludwig, K., Baljinnyam, B., Herrmann, A., and Bottcher, C.** The 3D structure of the fusion primed Sendai F-protein determined by electron cryomicroscopy 8. *EMBO J.*22[15], 3761-3771. (1-8-2003).

M.

- Ma, L., Siemssen, E. D., Noteborn, H. M., and van der Eb, A. J.** The xeroderma pigmentosum group B protein ERCC3 produced in the baculovirus system exhibits DNA helicase activity 1. *Nucleic Acids Res.*22[20], 4095-4102. (11-10-1994).
- Magill, C. P., Jackson, S. P., and Bell, S. D.** Identification of a conserved archaeal RNA polymerase subunit contacted by the basal transcription factor TFB 1. *J.Biol.Chem.*276[50], 46693-46696. (14-12-2001a).
- Makela, T. P., Tassan, J. P., Nigg, E. A., Frutiger, S., Hughes, G. J., and Weinberg, R. A.** A cyclin associated with the CDK-activating kinase MO15 22. *Nature*371[6494], 254-257. (15-9-1994).
- Malech, H. L. and Albert, J. P.** Negative staining of protein macromolecules: a simple rapid method 80. *J.Ultrastruct.Res.*69[2], 191-195. (1979).
- Maluf, N. K., Ali, J. A., and Lohman, T. M.** Kinetic mechanism for formation of the active, dimeric UvrD helicase-DNA complex 1. *J.Biol.Chem.*278[34], 31930-31940. (22-8-2003a).
- Maluf, N. K., Fischer, C. J., and Lohman, T. M.** A Dimer of Escherichia coli UvrD is the active form of the helicase in vitro 2. *J.Mol.Biol.*325[5], 913-935. (31-1-2003b).
- Marinoni, J. C., Rossignol, M., and Egly, J. M.** Purification of the transcription/repair factor TFIIH and evaluation of its associated activities in vitro 5. *Methods*12[3], 235-253. (1997a).
- Marinoni, J. C., Roy, R., Vermeulen, W., Miniou, P., Lutz, Y., Weeda, G., Seroz, T., Gomez, D. M., Hoeijmakers, J. H., and Egly, J. M.** Cloning and characterization of p52, the fifth subunit of the core of the transcription/DNA repair factor TFIIH 7. *EMBO J.*16[5], 1093-1102. (3-3-1997b).
- Martinez, E., Chiang, C. M., Ge, H., and Roeder, R. G.** TATA-binding protein-associated factor(s) in TFIIID function through the initiator to direct basal

- transcription from a TATA-less class II promoter 2. *EMBO J.*13[13], 3115-3126. (1-7-1994).
- Maru, Y., Kobayashi, T., Tanaka, K., and Shibuya, M.** BCR binds to the xeroderma pigmentosum group B protein 6. *Biochem.Biophys.Res.Commun.*260[2], 309-312. (5-7-1999).
- Matadeen, R., Patwardhan, A., Gowen, B., Orlova, E. V., Pape, T., Cuff, M., Mueller, F., Brimacombe, R., and van Heel, M.** The Escherichia coli large ribosomal subunit at 7.5 A resolution 1. *Structure Fold.Des*7[12], 1575-1583. (15-12-1999).
- Matson, S. W.** DNA helicases of Escherichia coli. *Prog.Nucleic Acid Res.Mol.Biol.*40, 289-326. (1991).
- Matsui, P., DePaulo, J., and Buratowski, S.** An interaction between the Tfb1 and Ssl1 subunits of yeast TFIIH correlates with DNA repair activity 1. *Nucleic Acids Res.*23[5], 767-772. (11-3-1995).
- Matsumiya, S., Ishino, Y., and Morikawa, K.** Crystal structure of an archaeal DNA sliding clamp: proliferating cell nuclear antigen from Pyrococcus furiosus 3. *Protein Sci.*10[1], 17-23. (2001).
- Matthews JM, Sunde M.** Zinc fingers--folds for many occasions. *IUBMB Life.* 54[6],351-5. (2002)
- Maxon, M. E., Goodrich, J. A., and Tjian, R.** Transcription factor IIE binds preferentially to RNA polymerase IIa and recruits TFIIH: a model for promoter clearance 12. *Genes Dev.*8[5], 515-524. (1-3-1994).
- Mechanic, L. E., Hall, M. C., and Matson, S. W.** Escherichia coli DNA helicase II is active as a monomer 2. *J.Biol.Chem.*274[18], 12488-12498. (30-4-1999).

- Meinhart, A., Blobel, J., and Cramer, P.** An extended winged helix domain in general transcription factor E/IEF $\alpha$  4. *J.Biol.Chem.*278[48], 48267-48274. (28-11-2003).
- Meininghaus, M., Chapman, R. D., Horndasch, M., and Eick, D.** Conditional expression of RNA polymerase II in mammalian cells. Deletion of the carboxyl-terminal domain of the large subunit affects early steps in transcription 1. *J.Biol.Chem.*275[32], 24375-24382. (11-8-2000).
- Miercke, L. J., Betlach, M. C., Mitra, A. K., Shand, R. F., Fong, S. K., and Stroud, R. M.** Wild-type and mutant bacteriorhodopsins D85N, D96N, and R82Q: purification to homogeneity, pH dependence of pumping, and electron diffraction 2. *Biochemistry*30[12], 3088-3098. (26-3-1991).
- Moreland, R. J., Tirode, F., Yan, Q., Conaway, J. W., Egly, J. M., and Conaway, R. C.** A role for the TFIIH XPB DNA helicase in promoter escape by RNA polymerase II 1. *J.Biol.Chem.*274[32], 22127-22130. (6-8-1999).
- Mounkes, L. C., Jones, R. S., Liang, B. C., Gelbart, W., and Fuller, M. T.** A Drosophila model for xeroderma pigmentosum and Cockayne's syndrome: haywire encodes the fly homolog of ERCC3, a human excision repair gene 1. *Cell*71[6], 925-937. (11-12-1992).
- Mu, D., Hsu, D. S., and Sancar, A.** Reaction mechanism of human DNA repair excision nuclease 2. *J.Biol.Chem.*271[14], 8285-8294. (5-4-1996).
- Murakami, K. S., Masuda, S., Campbell, E. A., Muzzin, O., and Darst, S. A.** Structural basis of transcription initiation: an RNA polymerase holoenzyme-DNA complex. *Science*296[5571], 1285-1290. (17-5-2002).
- Murray, J. M., Doe, C. L., Schenk, P., Carr, A. M., Lehmann, A. R., and Watts, F. Z.** Cloning and characterisation of the *S. pombe* rad15 gene, a homologue to the *S. cerevisiae* RAD3 and human ERCC2 genes 22. *Nucleic Acids Res.*20[11], 2673-2678. (11-6-1992).

## N.

**Nan, X., Cross, S., and Bird, A.** Gene silencing by methyl-CpG-binding proteins 5. *Novartis.Found.Symp.*214, 6-16. (1998).

**Napoli, A., van der, Oost J., Sensen, C. W., Charlebois, R. L., Rossi, M., and Ciaramella, M.** An Lrp-like protein of the hyperthermophilic archaeon *Sulfolobus solfataricus* which binds to its own promoter 4. *J.Bacteriol.*181[5], 1474-1480. (1999).

**Napoli, A., Zivanovic, Y., Bocs, C., Buhler, C., Rossi, M., Forterre, P., and Ciaramella, M.** DNA bending, compaction and negative supercoiling by the architectural protein Sso7d of *Sulfolobus solfataricus*1. *Nucleic Acids Res.*30[12], 2656-2662. (15-6-2002).

**Ng, J. M., Vermeulen, W., van der Horst, G. T., Bergink, S., Sugawara, K., Vrieling, H., and Hoeijmakers, J. H.** A novel regulation mechanism of DNA repair by damage-induced and RAD23-dependent stabilization of xeroderma pigmentosum group C protein1. *Genes Dev.*17[13], 1630-1645. (1-7-2003).

**Nield, J., Morris, E. P., Bibby, T. S., and Barber, J.** Structural analysis of the photosystem I supercomplex of cyanobacteria induced by iron deficiency 3. *Biochemistry*42[11], 3180-3188. (25-3-2003).

**Nikolov, D. B., Chen, H., Halay, E. D., Usheva, A. A., Hisatake, K., Lee, D. K., Roeder, R. G., and Burley, S. K.** Crystal structure of a TFIIB-TBP-TATA-element ternary complex 1. *Nature*377[6545], 119-128. (14-9-1995).

## O.

**Okuda, M., Watanabe, Y., Okamura, H., Hanaoka, F., Ohkuma, Y., and Nishimura, Y.** Structure of the central core domain of TFIIEbeta with a novel double-stranded DNA-binding surface 12. *EMBO J.*19[6], 1346-1356. (15-3-2000).

**Orlova, E. V.** Structural analysis of non-crystalline macromolecules: the ribosome 4. *Acta Crystallogr.D.Biol.Crystallogr.*56 ( Pt 10), 1253-1258. (2000).

**Orlova, E. V., Dube, P., Harris, J. R., Beckman, E., Zemlin, F., Markl, J., and van Heel, M.** Structure of keyhole limpet hemocyanin type 1 (KLH1) at 15 Å resolution by electron cryomicroscopy and angular reconstitution 1. *J.Mol.Biol.*271[3], 417-437. (22-8-1997).

P.

**Park, C. H., Mu, D., Reardon, J. T., and Sancar, A.** The general transcription-repair factor TFIIH is recruited to the excision repair complex by the XPA protein independent of the TFIIIE transcription factor 1. *J.Biol.Chem.*270[9], 4896-4902. (3-3-1995).

**Parvin, J. D. and Young, R. A.** Regulatory targets in the RNA polymerase II holoenzyme. *Curr.Opin.Genet.Dev.*8[5], 565-570. (1998).

**Pause, A., Methot, N., and Sonenberg, N.** The HRIGRXXR region of the DEAD box RNA helicase eukaryotic translation initiation factor 4A is required for RNA binding and ATP hydrolysis 4. *Mol.Cell Biol.*13[11], 6789-6798. (1993).

**Pearson, A. and Greenblatt, J.** Modular organization of the E2F1 activation domain and its interaction with general transcription factors TBP and TFIIH 2. *Oncogene*15[22], 2643-2658. (27-11-1997).

**Penczek, P., Radermacher, M., and Frank, J.** Three-dimensional reconstruction of single particles embedded in ice 2. *Ultramicroscopy*40[1], 33-53. (1992).

**Pereira, L. A., Klejman, M. P., Ruhlmann, C., Kavelaars, F., Oulad-Abdelghani, M., Timmers, H. T., and Schultz, P.** Molecular architecture of the basal transcription factor B-TFIID 7. *J.Biol.Chem.*279[21], 21802-21807. (21-5-2004).

**Plosky, B., Samson, L., Engelward, B. P., Gold, B., Schlaen, B., Millas, T., Magnotti, M., Schor, J., and Scicchitano, D. A.** Base excision repair and nucleotide excision repair contribute to the removal of N-methylpurines from active genes 12. *DNA Repair (Amst)*1[8], 683-696. (6-8-2002).

**Poterszman, A., Lamour, V., Egly, J. M., Moras, D., Thierry, J. C., and Poch, O.** A eukaryotic XPB/ERCC3-like helicase in *Mycobacterium leprae*? 14. *Trends Biochem.Sci.*22[11], 418-419. (1997).

**Prasad, B. V., Yamaguchi, S., and Roy, P.** Three-dimensional structure of single-shelled bluetongue virus 20. *J.Virol.*66[4], 2135-2142. (1992).

**Puig, O., Caspary, F., Rigaut, G., Rutz, B., Bouveret, E., Bragado-Nilsson, E., Wilm, M., and Seraphin, B.** The tandem affinity purification (TAP) method: a general procedure of protein complex purification 5. *Methods*24[3], 218-229. (2001).

#### Q.

**Qadri, I., Conaway, J. W., Conaway, R. C., Schaack, J., and Siddiqui, A.** Hepatitis B virus transactivator protein, HBx, associates with the components of TFIIH and stimulates the DNA helicase activity of TFIIH 1. *Proc.Natl.Acad.Sci.U.S.A*93[20], 10578-10583. (1-10-1996).

**Qureshi, S. A., Bell, S. D., and Jackson, S. P.** Factor requirements for transcription in the Archaeon *Sulfolobus shibatae* 22. *EMBO J.*16[10], 2927-2936. (15-5-1997).

#### R.

**Ranish, J. A., Hahn, S., Lu, Y., Yi, E. C., Li, X. J., Eng, J., and Aebersold, R.** Identification of TFB5, a new component of general transcription and DNA repair factor IIH 2. *Nat.Genet.*36[7], 707-713. (2004).

**Razin, A.** CpG methylation, chromatin structure and gene silencing-a three-way connection 4. *EMBO J.*17[17], 4905-4908. (1-9-1998).

**Reardon, J. T., Ge, H., Gibbs, E., Sancar, A., Hurwitz, J., and Pan, Z. Q.** Isolation and characterization of two human transcription factor IIH (TFIIH)-related complexes: ERCC2/CAK and TFIIH 4. *Proc.Natl.Acad.Sci.U.S.A*93[13], 6482-6487. (25-6-1996).



- Ribeiro, D. T., Machado, C. R., Costa, R. M., Praekelt, U. M., Van Sluys, M. A., and Menck, C. F.** Cloning of a cDNA from *Arabidopsis thaliana* homologous to the human XPB gene 1. *Gene*208[2], 207-213. (27-2-1998).
- Richard, D. J., Bell, S. D., and White, M. F.** Physical and functional interaction of the archaeal single-stranded DNA-binding protein SSB with RNA polymerase 4. *Nucleic Acids Res.*32[3], 1065-1074. (2004).
- Riedl, T. and Egly, J. M.** Phosphorylation in transcription: the CTD and more. *Gene Expr.*9[1-2], 3-13. (2000).
- Riedl, T., Hanaoka, F., and Egly, J. M.** The comings and goings of nucleotide excision repair factors on damaged DNA. *EMBO J.*22[19], 5293-5303. (1-10-2003a).
- Rio-Portilla, F., Gaskell, A., Gilbert, D., Ldias, J. A., and Wagner, G.** Solution structure of the hRPABC14.4 subunit of human RNA polymerases 1. *Nat.Struct.Biol.*6[11], 1039-1042. (1999).
- Roberts, J. A., Bell, S. D., and White, M. F.** An archaeal XPF repair endonuclease dependent on a heterotrimeric PCNA 2. *Mol.Microbiol.*48[2], 361-371. (2003).
- Rochette-Egly, C., Adam, S., Rossignol, M., Egly, J. M., and Chambon, P.** Stimulation of RAR alpha activation function AF-1 through binding to the general transcription factor TFIIH and phosphorylation by CDK7 2. *Cell*90[1], 97-107. (11-7-1997).
- Rojo, F.** Mechanisms of transcriptional repression. *Curr.Opin.Microbiol.*4[2], 145-151. (2001).
- Rossignol, M., Kolb-Cheynel, I., and Egly, J. M.** Substrate specificity of the cdk-activating kinase (CAK) is altered upon association with TFIIH 4. *EMBO J.*16[7], 1628-1637. (1-4-1997).
- Roth, S. Y., Shimizu, M., Johnson, L., Grunstein, M., and Simpson, R. T.** Stable nucleosome positioning and complete repression by the yeast alpha 2 repressor

are disrupted by amino-terminal mutations in histone H4 2. *Genes Dev.*6[3], 411-425. (1992).

**Runyon, G. T., Wong, I., and Lohman, T. M.** Overexpression, purification, DNA binding, and dimerization of the Escherichia coli uvrD gene product (helicase II) 2. *Biochemistry*32[2], 602-612. (19-1-1993).

**Russo, A. A., Jeffrey, P. D., Patten, A. K., Massague, J., and Pavletich, N. P.** Crystal structure of the p27Kip1 cyclin-dependent-kinase inhibitor bound to the cyclin A Cdk2 complex 1. *Nature*382[6589], 325-331. (25-7-1996).

S.

**Salerno, V., Napoli, A., White, M. F., Rossi, M., and Ciaramella, M.** Transcriptional response to DNA damage in the archaeon Sulfolobus solfataricus 2. *Nucleic Acids Res.*31[21], 6127-6138. (1-11-2003).

**Salzmann, M., Pervushin, K., Wider, G., Senn, H., and Wuthrich, K.** TROSY in triple-resonance experiments: new perspectives for sequential NMR assignment of large proteins 4. *Proc.Natl.Acad.Sci.U.S.A*95[23], 13585-13590. (10-11-1998).

**Sancar, A.** DNA excision repair 23. *Annu.Rev.Biochem.*65, 43-81. (1996).

**Sandman, K., Pereira, S. L., and Reeve, J. N.** Diversity of prokaryotic chromosomal proteins and the origin of the nucleosome 1. *Cell Mol.Life Sci.*54[12], 1350-1364. (1998).

**Sandrock, B. and Egly, J. M.** A yeast four-hybrid system identifies Cdk-activating kinase as a regulator of the XPD helicase, a subunit of transcription factor IIH 4. *J.Biol.Chem.*276[38], 35328-35333. (21-9-2001a).

**Schaeffer, L., Moncollin, V., Roy, R., Staub, A., Mezzina, M., Sarasin, A., Weeda, G., Hoeijmakers, J. H., and Egly, J. M.** The ERCC2/DNA repair protein is associated with the class II BTF2/TFIIH transcription factor 2. *EMBO J.*13[10], 2388-2392. (15-5-1994).

- Schneider, E., Montenarh, M., and Wagner, P.** Regulation of CAK kinase activity by p53 1. *Oncogene*17[21], 2733-2741. (26-11-1998).
- Schubeler, D., Lorincz, M. C., Cimborá, D. M., Telling, A., Feng, Y. Q., Bouhassira, E. E., and Groudine, M.** Genomic targeting of methylated DNA: influence of methylation on transcription, replication, chromatin structure, and histone acetylation. *Mol.Cell Biol.*20[24], 9103-9112. (2000).
- Schultz, P., Celia, H., Riva, M., Sentenac, A., and Oudet, P.** Three-dimensional model of yeast RNA polymerase I determined by electron microscopy of two-dimensional crystals 43. *EMBO J.*12[7], 2601-2607. (1993).
- Schultz, P., Fribourg, S., Poterszman, A., Mallouh, V., Moras, D., and Egly, J. M.** Molecular structure of human TFIIH. *Cell*102[5], 599-607. (1-9-2000a).
- Schwartz, G. J. and Diller, K. R.** Design and fabrication of a simple, versatile cryomicroscopy stage 1. *Cryobiology*19[5], 529-538. (1982).
- Seroz, T., Hwang, J. R., Moncollin, V., and Egly, J. M.** TFIIH: a link between transcription, DNA repair and cell cycle regulation 7. *Curr.Opin.Genet.Dev.*5[2], 217-221. (1995).
- Shiekhatar, R., Mermelstein, F., Fisher, R. P., Drapkin, R., Dynlacht, B., Wessling, H. C., Morgan, D. O., and Reinberg, D.** Cdk-activating kinase complex is a component of human transcription factor TFIIH 66. *Nature*374[6519], 283-287. (16-3-1995).
- Shivji, M. K., Podust, V. N., Hubscher, U., and Wood, R. D.** Nucleotide excision repair DNA synthesis by DNA polymerase epsilon in the presence of PCNA, RFC, and RPA. *Biochemistry*34[15], 5011-5017. (18-4-1995).
- Shuttleworth, J.** The regulation and functions of cdk7. *Prog.Cell Cycle Res.*1, 229-240. (1995).

- Slupphaug, G., Kavli, B., and Krokan, H. E.** The interacting pathways for prevention and repair of oxidative DNA damage 6. *Mutat.Res.*531[1-2], 231-251. (29-10-2003).
- Smale, S. T.** Transcription initiation from TATA-less promoters within eukaryotic protein-coding genes. *Biochim.Biophys.Acta*1351[1-2], 73-88. (20-3-1997).
- Sogin, S. J., Sogin, M. L., and Woese, C. R.** Phylogenetic measurement in procaryotes by primary structural characterization 21. *J.Mol.Evol.*1[1], 173-184. (1971).
- Spangler, L., Wang, X., Conaway, J. W., Conaway, R. C., and Dvir, A.** TFIIF action in transcription initiation and promoter escape requires distinct regions of downstream promoter DNA 2. *Proc.Natl.Acad.Sci.U.S.A*98[10], 5544-5549. (8-5-2001).
- Story, R. M., Li, H., and Abelson, J. N.** Crystal structure of a DEAD box protein from the hyperthermophile *Methanococcus jannaschii* 2. *Proc.Natl.Acad.Sci.U.S.A*98[4], 1465-1470. (13-2-2001).
- Su, S., Gao, Y. G., Robinson, H., Liaw, Y. C., Edmondson, S. P., Shriver, J. W., and Wang, A. H.** Crystal structures of the chromosomal proteins Sso7d/Sac7d bound to DNA containing T-G mismatched base-pairs 1. *J.Mol.Biol.*303[3], 395-403. (27-10-2000).
- Subramaniam, V., Li, H., Wong, M., Kitching, R., Attisano, L., Wrana, J., Zubovits, J., Burger, A. M., and Seth, A.** The RING-H2 protein RNF11 is overexpressed in breast cancer and is a target of Smurf2 E3 ligase 2. *Br.J.Cancer*89[8], 1538-1544. (20-10-2003).
- Sung, P., Bailly, V., Weber, C., Thompson, L. H., Prakash, L., and Prakash, S.** Human xeroderma pigmentosum group D gene encodes a DNA helicase 10. *Nature*365[6449], 852-855. (28-10-1993).

**Sung, P., Guzder, S. N., Prakash, L., and Prakash, S.** Reconstitution of TFIIH and requirement of its DNA helicase subunits, Rad3 and Rad25, in the incision step of nucleotide excision repair. *J.Biol.Chem.*271[18], 10821-10826. (3-5-1996a).

**Svejstrup, J. Q., Feaver, W. J., LaPointe, J., and Kornberg, R. D.** RNA polymerase transcription factor IIH holoenzyme from yeast 11. *J.Biol.Chem.*269[45], 28044-28048. (11-11-1994).

**Sweder, K. S., Chun, R., Mori, T., and Hanawalt, P. C.** DNA repair deficiencies associated with mutations in genes encoding subunits of transcription initiation factor TFIIH in yeast 2. *Nucleic Acids Res.*24[8], 1540-1546. (15-4-1996).

T.

**Talukder, A. H., Mishra, S. K., Mandal, M., Balasenthil, S., Mehta, S., Sahin, A. A., Barnes, C. J., and Kumar, R.** MTA1 interacts with MAT1, a cyclin-dependent kinase-activating kinase complex ring finger factor, and regulates estrogen receptor transactivation functions 1. *J.Biol.Chem.*278[13], 11676-11685. (28-3-2003).

**Tan, S., Hunziker, Y., Sargent, D. F., and Richmond, T. J.** Crystal structure of a yeast TFIIA/TBP/DNA complex 3. *Nature*381[6578], 127-151. (9-5-1996).

**Tapias, A., Auriol, J., Forget, D., Enzlin, J. H., Scharer, O. D., Coin, F., Coulombe, B., and Egly, J. M.** Ordered conformational changes in damaged DNA induced by nucleotide excision repair factors 1. *J.Biol.Chem.*279[18], 19074-19083. (30-4-2004).

**Tassan, J. P., Jaquenoud, M., Fry, A. M., Frutiger, S., Hughes, G. J., and Nigg, E. A.** In vitro assembly of a functional human CDK7-cyclin H complex requires MAT1, a novel 36 kDa RING finger protein 2. *EMBO J.*14[22], 5608-5617. (15-11-1995).

**Taylor, K. A. and Crowther, R. A.** A protocol for 3D image reconstruction from a single image of an oblique section 7. *Ultramicroscopy*38[1], 85-103. (1991).

- Taylor, K. A. and Crowther, R. A.** 3D reconstruction from the Fourier transform of a single superlattice image of an oblique section 25. *Ultramicroscopy*41[1-3], 153-167. (1992).
- Theis, K., Chen, P. J., Skorvaga, M., Van Houten, B., and Kisker, C.** Crystal structure of UvrB, a DNA helicase adapted for nucleotide excision repair 5. *EMBO J.*18[24], 6899-6907. (15-12-1999).
- Theobald, D. L., Mitton-Fry, R. M., and Wuttke, D. S.** Nucleic acid recognition by OB-fold proteins 4. *Annu.Rev.Biophys.Biomol.Struct.*32, 115-133. (2003).
- Tirode, F., Busso, D., Coin, F., and Egly, J. M.** Reconstitution of the transcription factor TFIIH: assignment of functions for the three enzymatic subunits, XPB, XPD, and cdk7 3. *Mol.Cell*3[1], 87-95. (1999).
- Tong, X., Drapkin, R., Reinberg, D., and Kieff, E.** The 62- and 80-kDa subunits of transcription factor IIH mediate the interaction with Epstein-Barr virus nuclear protein 2 2. *Proc.Natl.Acad.Sci.U.S.A*92[8], 3259-3263. (11-4-1995).
- Trievel, R. C.** Structure and function of histone methyltransferases 16. *Crit Rev.Eukaryot.Gene Expr.*14[3], 147-169. (2004).
- Tsai, C. C. and Fondell, J. D.** Nuclear receptor recruitment of histone-modifying enzymes to target gene promoters 4. *Vitam.Horm.*68, 93-122. (2004).
- Tsai, F. T. and Sigler, P. B.** Structural basis of preinitiation complex assembly on human pol II promoters 1. *EMBO J.*19[1], 25-36. (4-1-2000).

## V.

- Valay, J. G., Dubois, M. F., Bensaude, O., and Faye, G.** Ccl1, a cyclin associated with protein kinase Kin28, controls the phosphorylation of RNA polymerase II largest subunit and mRNA transcription 3. *C R.Acad.Sci.III*319[3], 183-189. (1996).
- Valle, M., Zavialov, A., Li, W., Stagg, S. M., Sengupta, J., Nielsen, R. C., Nissen, P., Harvey, S. C., Ehrenberg, M., and Frank, J.** Incorporation of aminoacyl-tRNA into

- the ribosome as seen by cryo-electron microscopy 1. *Nat.Struct.Biol.*10[11], 899-906. (2003).
- van Brabant, A. J., Stan, R., and Ellis, N. A.** DNA helicases, genomic instability, and human genetic disease. *Annu.Rev.Genomics Hum.Genet.*1, 409-459. (2000).
- van Heel, M.** Angular reconstitution: a posteriori assignment of projection directions for 3D reconstruction 1. *Ultramicroscopy*21[2], 111-123. (1987).
- van Heel, M., Harauz, G., Orlova, E. V., Schmidt, R., and Schatz, M.** A new generation of the IMAGIC image processing system 2. *J.Struct.Biol.*116[1], 17-24. (1996).
- van Hoffen, A., Balajee, A. S., van Zeeland, A. A., and Mullenders, L. H.** Nucleotide excision repair and its interplay with transcription 13. *Toxicology*193[1-2], 79-90. (15-11-2003).
- Vandel, L. and Kouzarides, T.** Residues phosphorylated by TFIIH are required for E2F-1 degradation during S-phase 1. *EMBO J.*18[15], 4280-4291. (2-8-1999).
- Velankar, S. S., Soultanas, P., Dillingham, M. S., Subramanya, H. S., and Wigley, D. B.** Crystal structures of complexes of PcrA DNA helicase with a DNA substrate indicate an inchworm mechanism 2. *Cell*97[1], 75-84. (2-4-1999).
- Venter, J. C., Adams, M. D., Myers, E. W., Li, P. W., Mural, R. J., Sutton, G. G., Smith, H. O., Yandell, M., Evans, C. A., Holt, R. A., Gocayne, J. D., Amanatides, P., Ballew, R. M., Huson, D. H., Wortman, J. R., Zhang, Q., Kodira, C. D., Zheng, X. Q. H., Chen, L., Skupski, M., Subramanian, G., Thomas, P. D., Zhang, J. H., Miklos, G. L. G., Nelson, C., Broder, S., Clark, A. G., Nadeau, C., McKusick, V. A., Zinder, N., Levine, A. J., Roberts, R. J., Simon, M., Slayman, C., Hunkapiller, M., Bolanos, R., Delcher, A., Dew, I., Fasulo, D., Flanigan, M., Florea, L., Halpern, A., Hannenhalli, S., Kravitz, S., Levy, S., Mobarry, C., Reinert, K., Remington, K., Abu-Threideh, J., Beasley, E., Biddick, K., Bonazzi, V., Brandon, R., Cargill, M., Chandramouliswaran, I., Charlab, R., Chaturvedi, K., Deng, Z. M., Di Francesco, V., Dunn, P., Eilbeck, K., Evangelista, C., Gabrielian, A. E., Gan, W., Ge, W. M.,**

Gong, F. C., Gu, Z. P., Guan, P., Heiman, T. J., Higgins, M. E., Ji, R. R., Ke, Z. X., Ketchum, K. A., Lai, Z. W., Lei, Y. D., Li, Z. Y., Li, J. Y., Liang, Y., Lin, X. Y., Lu, F., Merkulov, G. V., Milshina, N., Moore, H. M., Naik, A. K., Narayan, V. A., Neelam, B., Nusskern, D., Rusch, D. B., Salzberg, S., Shao, W., Shue, B. X., Sun, J. T., Wang, Z. Y., Wang, A. H., Wang, X., Wang, J., Wei, M. H., Wides, R., Xiao, C. L., Yan, C. H., Yao, A., Ye, J., Zhan, M., Zhang, W. Q., Zhang, H. Y., Zhao, Q., Zheng, L. S., Zhong, F., Zhong, W. Y., Zhu, S. P. C., Zhao, S. Y., Gilbert, D., Baumhueter, S., Spier, G., Carter, C., Cravchik, A., Woodage, T., Ali, F., An, H. J., Awe, A., Baldwin, D., Baden, H., Barnstead, M., Barrow, I., Beeson, K., Busam, D., Carver, A., Center, A., Cheng, M. L., Curry, L., Danaher, S., Davenport, L., Desilets, R., Dietz, S., Dodson, K., Doup, L., Ferriera, S., Garg, N., Gluecksmann, A., Hart, B., Haynes, J., Haynes, C., Heiner, C., Hladun, S., Hostin, D., Houck, J., Howland, T., Ibegwam, C., Johnson, J., Kalush, F., Kline, L., Koduru, S., Love, A., Mann, F., May, D., McCawley, S., McIntosh, T., McMullen, I., Moy, M., Moy, L., Murphy, B., Nelson, K., Pfannkoch, C., Pratts, E., Puri, V., Qureshi, H., Reardon, M., Rodriguez, R., Rogers, Y. H., Romblad, D., Ruhfel, B., Scott, R., Sitter, C., Smallwood, M., Stewart, E., Strong, R., Suh, E., Thomas, R., Tint, N. N., Tse, S., Vech, C., Wang, G., Wetter, J., Williams, S., Williams, M., Windsor, S., Winn-Deen, E., Wolfe, K., Zaveri, J., Zaveri, K., Abril, J. F., Guigo, R., Campbell, M. J., Sjolander, K. V., Karlak, B., Kejariwal, A., Mi, H. Y., Lazareva, B., Hatton, T., Narechania, A., Diemer, K., Muruganujan, A., Guo, N., Sato, S., Bafna, V., Istrail, S., Lippert, R., Schwartz, R., Walenz, B., Yooseph, S., Allen, D., Basu, A., Baxendale, J., Blick, L., Caminha, M., Carnes-Stine, J., Caulk, P., Chiang, Y. H., Coyne, M., Dahlke, C., Mays, A. D., Dombroski, M., Donnelly, M., Ely, D., Esparham, S., Fosler, C., Gire, H., Glanowski, S., Glasser, K., Glodek, A., Gorokhov, M., Graham, K., Gropman, B., Harris, M., Heil, J., Henderson, S., Hoover, J., Jennings, D., Jordan, C., Jordan, J., Kasha, J., Kagan, L., Kraft, C., Levitsky, A., Lewis, M., Liu, X. J., Lopez, J., Ma, D., Majoros, W., McDaniel, J., Murphy, S., Newman, M., Nguyen, T., Nguyen, N., and Nodell, M. The sequence of the human genome. *Science*291[5507], 1304-+. (16-2-2001).



**Verrijzer, C. P., Chen, J. L., Yokomori, K., and Tjian, R.** Binding of TAFs to core elements directs promoter selectivity by RNA polymerase II 3. *Cell*81[7], 1115-1125. (30-6-1995).

**Vichi, P., Coin, F., Renaud, J. P., Vermeulen, W., Hoeijmakers, J. H., Moras, D., and Egly, J. M.** Cisplatin- and UV-damaged DNA lure the basal transcription factor TFIID/TBP 1. *EMBO J.*16[24], 7444-7456. (15-12-1997).

**Volker, M., Mone, M. J., Karmakar, P., van Hoffen, A., Schul, W., Vermeulen, W., Hoeijmakers, J. H., van Driel, R., van Zeeland, A. A., and Mullenders, L. H.** Sequential assembly of the nucleotide excision repair factors in vivo 5. *Mol.Cell*8[1], 213-224. (2001a).

W.

**Wang, Q. E., Zhu, Q., Wani, M. A., Wani, G., Chen, J., and Wani, A. A.** Tumor suppressor p53 dependent recruitment of nucleotide excision repair factors XPC and TFIIH to DNA damage 3. *DNA Repair (Amst)*2[5], 483-499. (13-5-2003a).

**Wang, X., Spangler, L., and Dvir, A.** Promoter escape by RNA polymerase II. Downstream promoter DNA is required during multiple steps of early transcription 1. *J.Biol.Chem.*278[12], 10250-10256. (21-3-2003b).

**Wang, X. W., Yeh, H., Schaeffer, L., Roy, R., Moncollin, V., Egly, J. M., Wang, Z., Freidberg, E. C., Evans, M. K., and Taffe, B. G.** p53 modulation of TFIIH-associated nucleotide excision repair activity. *Nat.Genet.*10[2], 188-195. (1995a).

**Wang, Z., Buratowski, S., Svejstrup, J. Q., Feaver, W. J., Wu, X., Kornberg, R. D., Donahue, T. F., and Friedberg, E. C.** The yeast TFB1 and SSL1 genes, which encode subunits of transcription factor IIH, are required for nucleotide excision repair and RNA polymerase II transcription. *Mol.Cell Biol.*15[4], 2288-2293. (1995b).

- Wang, Z., Svejstrup, J. O., Feaver, W. J., Wu, X., Kornberg, R. D., and Friedberg, E. C.** Transcription factor b (TFIIH) is required during nucleotide-excision repair in yeast. *Nature*368[6466], 74-76. (3-3-1994a).
- Wardleworth, B. N., Russell, R. J., Bell, S. D., Taylor, G. L., and White, M. F.** Structure of Alba: an archaeal chromatin protein modulated by acetylation. *EMBO J.*21[17], 4654-4662. (2-9-2002).
- Wardleworth, B. N., Russell, R. J., White, M. F., and Taylor, G. L.** Preliminary crystallographic studies of the double-stranded DNA-binding protein Sso10b from *Sulfolobus solfataricus* 3. *Acta Crystallogr.D.Biol.Crystallogr.*57[Pt 12], 1893-1894. (2001).
- Warren, E. H., Otterud, B. E., Linterman, R. W., Brickner, A. G., Engelhard, V. H., Leppert, M. F., Martin, P. J., and Riddell, S. R.** Feasibility of using genetic linkage analysis to identify the genes encoding T cell-defined minor histocompatibility antigens. *Tissue Antigens*59[4], 293-303. (2002).
- Watanabe, Y., Fujimoto, H., Watanabe, T., Maekawa, T., Masutani, C., Hanaoka, F., and Ohkuma, Y.** Modulation of TFIIH-associated kinase activity by complex formation and its relationship with CTD phosphorylation of RNA polymerase II 3. *Genes Cells*5[5], 407-423. (2000).
- Winkler, G. S., Araujo, S. J., Fiedler, U., Vermeulen, W., Coin, F., Egly, J. M., Hoeijmakers, J. H., Wood, R. D., Timmers, H. T., and Weeda, G.** TFIIH with inactive XPD helicase functions in transcription initiation but is defective in DNA repair 2. *J.Biol.Chem.*275[6], 4258-4266. (11-2-2000c).
- Winkler, G. S., Sugasawa, K., Eker, A. P., de Laat, W. L., and Hoeijmakers, J. H.** Novel functional interactions between nucleotide excision DNA repair proteins influencing the enzymatic activities of TFIIH, XPG, and ERCC1-XPF 7. *Biochemistry*40[1], 160-165. (9-1-2001).
- Winkler, G. S., Vermeulen, W., Coin, F., Egly, J. M., Hoeijmakers, J. H., and Weeda, G.** Affinity purification of human DNA repair/transcription factor TFIIH using

epitope-tagged xeroderma pigmentosum B protein 5. *J.Biol.Chem.*273[2], 1092-1098. (9-1-1998).

**Winkler, H. and Taylor, K. A.** 3D reconstruction by combining data from sections cut oblique to different unit cell axes 2. *Ultramicroscopy*55[4], 357-371. (1994).

**Woese, C. R. and Fox, G. E.** Phylogenetic structure of the prokaryotic domain: the primary kingdoms 18. *Proc.Natl.Acad.Sci.U.S.A*74[11], 5088-5090. (1977).

**Woese, C. R., Gupta, R., Hahn, C. M., Zillig, W., and Tu, J.** The phylogenetic relationships of three sulfur dependent archaebacteria 7. *Syst.Appl.Microbiol.*5, 97-105. (1984).

**Woese, C. R., Magrum, L. J., and Fox, G. E.** Archaebacteria 16. *J.Mol.Evol.*11[3], 245-251. (2-8-1978).

**Woese, C. R. and Olsen, G. J.** Archaebacterial phylogeny: perspectives on the urkingdoms 5. *Syst.Appl.Microbiol.*7, 161-177. (1986).

**Wu, X., Braithwaite, E., and Wang, Z.** DNA ligation during excision repair in yeast cell-free extracts is specifically catalyzed by the CDC9 gene product 4. *Biochemistry*38[9], 2628-2635. (2-3-1999).

X.

**Xiao, H., Pearson, A., Coulombe, B., Truant, R., Zhang, S., Regier, J. L., Triezenberg, S. J., Reinberg, D., Flores, O., and Ingles, C. J.** Binding of basal transcription factor TFIIH to the acidic activation domains of VP16 and p53. *Mol.Cell Biol.*14[10], 7013-7024. (1994a).

Y.

**Yan, Y., Li, Y., Munshi, S., Sardana, V., Cole, J. L., Sardana, M., Steinkuehler, C., Tomei, L., De Francesco, R., Kuo, L. C., and Chen, Z.** Complex of NS3 protease and NS4A peptide of BK strain hepatitis C virus: a 2.2 Å resolution structure in a hexagonal crystal form 7. *Protein Sci.*7[4], 837-847. (1998).

- Yankulov, K. Y. and Bentley, D. L.** Regulation of CDK7 substrate specificity by MAT1 and TFIIH 3. *EMBO J.*16[7], 1638-1646. (1-4-1997).
- Yee, A., Booth, V., Dharamsi, A., Engel, A., Edwards, A. M., and Arrowsmith, C. H.** Solution structure of the RNA polymerase subunit RPB5 from *Methanobacterium thermoautotrophicum* 5. *Proc.Natl.Acad.Sci.U.S.A*97[12], 6311-6315. (6-6-2000).
- Yee, A., Nichols, M. A., Wu, L., Hall, F. L., Kobayashi, R., and Xiong, Y.** Molecular cloning of CDK7-associated human MAT1, a cyclin-dependent kinase-activating kinase (CAK) assembly factor 2. *Cancer Res.*55[24], 6058-6062. (15-12-1995).
- Yildirim, Y. and Doruker, P.** Collective motions of RNA polymerases. Analysis of core enzyme, elongation complex and holoenzyme 1. *J.Biomol.Struct.Dyn.*22[3], 267-280. (2004).
- Yokoi, M., Masutani, C., Maekawa, T., Sugasawa, K., Ohkuma, Y., and Hanaoka, F.** The xeroderma pigmentosum group C protein complex XPC-HR23B plays an important role in the recruitment of transcription factor IIH to damaged DNA 14. *J.Biol.Chem.*275[13], 9870-9875. (31-3-2000).
- Yoon, H., Miller, S. P., Pabich, E. K., and Donahue, T. F.** SSL1, a suppressor of a HIS4 5'-UTR stem-loop mutation, is essential for translation initiation and affects UV resistance in yeast 1. *Genes Dev.*6[12B], 2463-2477. (1992).
- You, J. S., Wang, M., and Lee, S. H.** Biochemical analysis of the damage recognition process in nucleotide excision repair 1. *J.Biol.Chem.*278[9], 7476-7485. (28-2-2003).
- You, Z., Feaver, W. J., and Friedberg, E. C.** Yeast RNA polymerase II transcription in vitro is inhibited in the presence of nucleotide excision repair: complementation of inhibition by Holo-TFIIH and requirement for RAD26 2. *Mol.Cell Biol.*18[5], 2668-2676. (1998).

Z.

- Zawel, L., Kumar, K. P., and Reinberg, D.** Recycling of the general transcription factors during RNA polymerase II transcription 1. *Genes Dev.*9[12], 1479-1490. (15-6-1995).
- Zhang, D. Y., Carson, D. J., and Ma, J.** The role of TFIIB-RNA polymerase II interaction in start site selection in yeast cells. *Nucleic Acids Res.*30[14], 3078-3085. (15-7-2002).
- Zhang, K., Siino, J. S., Jones, P. R., Yau, P. M., and Bradbury, E. M.** A mass spectrometric "Western blot" to evaluate the correlations between histone methylation and histone acetylation 6. *Proteomics.* (27-8-2004).
- Zhang, L., Schwartz, G., O'Donnell, M., and Harrison, R. K.** Development of a novel helicase assay using electrochemiluminescence 2. *Anal.Biochem.*293[1], 31-37. (1-6-2001).
- Zhao, F. Q. and Craig, R.** Capturing time-resolved changes in molecular structure by negative staining 1. *J.Struct.Biol.*141[1], 43-52. (2003).
- Zurita, M. and Merino, C.** The transcriptional complexity of the TFIIH complex 9. *Trends Genet.*19[10], 578-584. (2003).

## Communications personnelles

Jawhari A, **Uhring M**, Crucifix C, Fribourg S, Schultz P, Poterszman A, Egly JM, Moras D.

Expression of FLAG fusion proteins in insect cells: application to the multi-subunit transcription/DNA repair factor TFIIH.

Protein Expr Purif. 2002 Apr;24(3):513-23.

Crucifix C, **Uhring M**, Schultz P.

Immobilization of biotinylated DNA on 2-D streptavidin crystals.

J Struct Biol. 2004 Jun;146(3):441-51.

**Uhring M\***, Jawhari A\*, Schultz P, Poterszman A. (\*co-premiers auteurs)

Molecular organization and oligomeric state of TFIIIE. Soumis EMBO Journal.

**Uhring M**, Poterszman A. Les hélicases et maladies associées : quand la double hélice ne s'ouvre plus. Soumis Médecine et Sciences.

**Uhring M**, : Influence of affinity tags for protein production in insect cells. En préparation

## Digression : "Connaître la structure d'une molécule pour comprendre sa fonction"

En février 1953, deux mois avant la publication du célèbre article dans *Nature*, on raconte que Francis Crick aurait surgi dans un pub à Cambridge en annonçant qu'il avait, avec James Watson, découvert rien de moins que le secret de la vie.

En quoi la structure de l'ADN révèle-t-elle le secret de la vie? En biologie, la fonction est souvent liée à la structure et c'est vrai plus que jamais pour l'ADN. C'est la découverte de la structure de l'ADN qui allait permettre, dans les années suivantes, de comprendre les bases moléculaires de l'hérédité.

La découverte de la structure de l'ADN a permis de comprendre sous quelle forme est l'information dans la cellule, comment cette information peut se reproduire, comment elle peut se modifier permettant ainsi l'évolution et comment elle peut être traduite en caractéristiques physiques transmissibles à travers les générations.

Le traitement de l'information génétique peut être comparé au traitement de l'information numérique. Dans la réplication ou duplication, l'information génétique est copiée durant la division cellulaire pour être transmise aux générations suivantes. Le contenu de l'information de l'ADN dupliqué est identique à celui de la matrice d'ADN ayant produit la copie: de la même manière, lorsque l'on duplique une disquette informatique, les contenus de l'original et de la copie sont identiques.

Dans le processus complexe de la transcription puis de la traduction de l'information génétique codée de l'ADN, via la messagerie de l'ARN, l'information génétique est traitée pour produire des protéines (macromolécules composées d'acides aminés) entrant dans la constitution des organismes vivants et déterminant leurs structures et leurs fonctions. La transcription puis la traduction correspondraient au traitement d'un fichier informatique non pas pour le dupliquer tel quel, mais pour le transformer, l'étoffer en un document élaboré, destiné à des applications spécifiques.

Les protéines sont les molécules les plus complexes et les plus variées des êtres vivants. Un être humain fabriquerait au total quelque chose comme 100 000 sortes

différentes de protéines. Chaque cellule en fabrique en moyenne 15 000 sortes différentes. Près de 50% du poids sec d'un être vivant est fait de protéines.

Une protéine, c'est un polymère d'acides aminés, c'est à dire une grande molécule formée de l'union de plus petites, les acides aminés. La plupart des protéines sont formées de l'union de 100 à 200 acides aminés.

La protéine que nous avez étudiée est appelée hélicase. Dans la cellule, c'est une enzyme du noyau, l'ADN hélicase, qui sépare l'ADN en deux brins. L'ADN hélicase agit un peu comme la tirette d'une fermeture éclair qui sépare les deux parties de la fermeture.

Cette hélicase est incriminée dans l'apparition de cancers de la peau. Leur incidence augmente avec l'exposition solaire : les ultraviolets (UV) contenus dans la lumière solaire en sont les principaux responsables. Ils provoquent des dommages dans l'ADN des cellules exposées, en particulier celles qui composent la couche superficielle de notre peau, l'épiderme. Dans la population normale, la réparation de ces dommages par un système spécialisé de réparation de l'ADN, assure le maintien du patrimoine génétique. La persistance de ces dommages peut, en revanche, conduire à l'apparition de mutations (modification du patrimoine génétique) et, potentiellement, au développement de tumeurs cutanées. Chez les patients atteints d'une maladie génétique rare appelée Xeroderma pigmentosum (XP), la structure de certains gènes impliqués dans le processus de réparation de l'ADN est altérée. La persistance des dommages dans l'ADN, induits par les UV solaires, conduit à une intolérance sévère au soleil, accompagnée d'une prédisposition dramatique aux cancers de la peau, presque systématiquement à l'origine de la mort prématurée des patients.

Nous cherchons à déterminer la forme (structure) de cette hélicase, une grosse molécule très complexe que l'on trouve dans les cellules des organismes vivants, pour comprendre comment elle agit. Pour l'instant, on connaît la forme et la fonction d'une centaine de protéines seulement, alors qu'il en existe des milliers! Et il ne suffit pas d'une loupe pour observer une protéine de près. Trois techniques permettent de déterminer la structure de ces molécules. Pour utiliser la première, qu'on appelle «



cristallographie des rayons X », il faut d'abord figer la fragile structure de la molécule, c'est-à-dire la mettre sous la forme d'un cristal solide, un peu comme l'archéologue doit faire un moulage de l'ossement qu'il vient de découvrir. L'opération est délicate. La deuxième technique, la résonance magnétique nucléaire à très haut champ, permet d'étudier les protéines dans leur état naturel, sans avoir à les modifier. La dernière technique est la microscopie électronique, dans son principe, elle utilise un faisceau d'électron émis par un canon, focalisé sur la préparation à l'aide de lentilles électromagnétiques, l'image se forme sur un écran.

En fait, le simple changement de forme de cette molécule la rend dangereuse et provoque la maladie. Sous une forme, la protéine est donc mortelle, alors qu'elle est inoffensive sous une autre. Dans ces conditions, on comprend pourquoi il peut être aussi utile de déterminer la forme d'une molécule.

## Résumé

Le facteur TFIIH est un facteur multi-protéique impliqué, via ses dix sous-unités, dans la transcription des gènes de classe II et dans la réparation de l'ADN. Chez l'homme, des mutations dans les gènes codant pour des hélicases XPB et XPD de TFIIH sont à l'origine de trois maladies: le Xeroderma Pigmentosum, la Trichothiodystrophie et le Syndrome de Cockayne.

Au cours de mes études doctorales, je me suis intéressée aux liens entre la structure et la fonction du facteur TFIIH humain. Nous avons observé le complexe TFIIH produit sous forme recombinante dans les cellules d'insectes en microscopie électronique et sa structure est identique à celle du facteur endogène. Nous y avons ensuite intégré les données biochimiques d'interaction protéine-protéine et nous proposons un modèle général de l'architecture du complexe. Afin de comprendre le rôle de TFIIH dans l'initiation de la transcription ou dans la réparation de l'ADN, nous avons développé une nouvelle technologie visant à observer les complexes ADN-protéine en microscopie. Nous avons également déterminé un modèle en cryomicroscopie du facteur TFIIIE, qui interagit physiquement avec TFIIH, à une résolution de 1.6nm. Finalement, nous avons obtenu des cristaux de l'homologue de l'hélicase XPD chez une archaebactérie.

## Summary

TFIIH is a general transcription factor involved in class II gene transcription as well as nucleotide excision repair. In human, mutations in both helicases of TFIIH, XPD and XPB, are directly incriminated in three diseases: the Xeroderma Pigmentosum, the Cockayne Syndrome and trichothiodystrophie.

During my PhD studies, I have been interested in the structure and function relationships of the TFIIH human factor. We have characterized the recombinant TFIIH complex in electronic microscopic and its structure was identical to the endogenous one. We have used this structure to dock the biochemical informations obtain by protein-protein interaction studies and we propose a general model of the molecular organization of the human TFIIH. Finally a study aimed to put TFIIH with natural substrates, DNA or partners, was done. I was implicated in the development of a new technology designed to immobilize DNA on a solid support suitable for electron microscopy observation. We also reconstruct a cryo-negative staining model of TFIIIE, which interacts physically with TFIIH, at 1.6nm resolution. Finally, we obtain 3D crystals of the homologue of XPD helicase in an archaebacteria.

---

Mots clefs : **transcription, réparation de l'ADN, TFIIH, hélicase, protéines recombinantes, microscopie électronique**

Discipline : **Sciences du Vivant. Spécialité : Aspects Moléculaires et Cellulaires de la Biologie Mention : Biologie Structurale**

Lieu de préparation de la thèse : **Laboratoire de Génomique et de Biologie Structurales IGBMC, 1, rue Laurent Fries, Parc d'Innovation, 67404 Illkirch Cedex.**

---