

2m11.2715.8

Université de Montréal

Interactions de la protéine à multidoigt  
de zinc ZNF74 avec les acides nucléiques

par  
Mélanie Guévremont

Département de Biologie Moléculaire  
Faculté de Médecine

Mémoire présenté à la Faculté des études supérieures  
en vue de l'obtention du grade de  
Maître ès sciences (M. Sc.)  
en Biologie Moléculaire

Février 1999

© Mélanie Guévremont



8.2155.1108

QH  
506  
U54  
1999  
V.010

Université de Montréal

Immunisation de la protéine à multipeptide  
de site ZNF 74 avec les acides nucléiques

Victor G. Gagnon  
Ph.D.

Département de Biologie Moléculaire  
Faculté de Médecine

Mémoire présenté à la Faculté des études supérieures  
en vue de l'obtention du grade de  
Maîtrise en Sciences (M.Sc.)  
en Biologie Moléculaire

Révisé par



Université de Montréal

Université de Montréal  
Faculté des études supérieures

Ce mémoire intitulé:

Interactions de la protéine à multidoigt  
de zinc ZNF74 avec les acides nucléiques

Présenté par:  
Mélanie Guévremont

a été évalué par un jury composé des personnes suivantes:

Dr Martine Raymond

Dr Muriel Aubry

Dr Majambu Mbikay

Mémoire accepté le: 99.04.23

## SOMMAIRE

Le syndrome de DiGeorge (DGS) est caractérisé par un ensemble de défauts congénitaux affectant le thymus, les glandes parathyroïdes, le coeur et la morphologie crânio-faciale. Dans plus de 90% des cas répertoriés de DGS, on observe la perte d'une copie de la région 22q11.2. Par une approche de gène candidat, le gène ZNF74 localisé dans la région critique 22q11.2, connue comme subissant des réarrangements chez les majorité des patients atteints de DGS, a été isolé. Ce gène, dont une seule copie semble présente chez la majorité des patients DGS, est exprimé chez l'embryon et fait partie de la super-famille des gènes à motifs doigt de zinc de type Cys<sub>2</sub>His<sub>2</sub>. ZNF74 comprend douze doigts de zinc et fait partie de la famille des multidoigts de zinc. ZNF74 encode, d'autre part, un domaine conservé de type KRAB présent chez près du tiers des protéines à multidoigt et caractérisé comme un domaine de répression de la transcription. La présence de ce domaine et de motifs doigt de zinc amène à considérer les protéines KRAB à multidoigt comme des facteurs transcriptionnels potentiels. Les objectifs de ce mémoire visaient l'étude des interactions de la protéine ZNF74 avec les acides nucléiques. Nous avons cherché à déterminer si ZNF74 pouvait lier l'ADN à l'instar des facteurs transcriptionnels traditionnels. Par des études de buvardage de type South-Western et de retardement sur gel, nous démontrons qu'en effet la protéine ZNF74 lie l'ADN *via* son domaine doigt de zinc. Par contre, il n'a pas été possible d'isoler, par des études de sélection de sites *in vitro*, des séquences spécifiques d'ADN double brin auxquelles se lie ZNF74. Quoique la liaison spécifique à des séquences d'ADN double brin soit caractéristique de la grande majorité des facteurs transcriptionnels, ce type de liaison n'a pu être démontré jusqu'à présent pour aucune protéine KRAB à multidoigt.

On ne peut exclure à l'heure actuelle que les protéines KRAB à multidoigt aient besoin de co-facteur(s) pour stabiliser leur liaison à l'ADN ou fonctionnent comme des co-régulateurs de la transcription en se liant à des facteurs transcriptionnels eux-mêmes liés à l'ADN. Nos expériences de retardement sur gel démontrent également que ZNF74 lie l'ADN simple brin préférentiellement à l'ADN double brin et hétéroduplex. Au cours des dernières années, il a pu être démontré que certaines protéines se liant à l'ADN simple brin de façon dépendante (ex: FBP et hnRNP K) ou indépendante de la séquence (ex: PC4) sont capables de réguler la transcription en tant que facteurs de transcription *bona fide*. Certaines d'entre elles se lient aussi à l'ARN (FBP, hnRNP K) et pourraient aussi avoir un rôle post-transcriptionnel. Les conséquences fonctionnelles de la liaison de ZNF74 à l'ADN double brin, simple brin ou hétéroduplex telles que détectées *in vitro* de même que ses gènes cibles potentiels demeurent actuellement inconnus.

Compte tenu des résultats présentés dans ce mémoire et de la mise en évidence par d'autres études réalisées dans notre laboratoire d'une interaction de ZNF74 avec l'ARN ainsi qu'avec une forme de l'ARN polymérase II impliquée dans la maturation de l'ARN, il est possible de proposer que ZNF74 puisse avoir un double rôle. En effet, ZNF74 pourrait agir comme facteur transcriptionnel *via* son domaine KRAB et sa liaison potentielle à l'ADN et aussi comme facteur de maturation de l'ARN *via* son interaction avec l'ARN et son interaction avec une forme de l'ARN polymérase retrouvée dans des sites nucléaires enrichis en facteurs d'épissage. On estime actuellement qu'il y aurait de 100 à 300 protéines KRAB à multidoigt encodées par le génome des mammifères. Il est fort probable que d'autres membres de cette famille partagent des caractéristiques communes avec ZNF74.

## TABLE DES MATIÈRES

SOMMAIRE.....	III
TABLE DES MATIÈRES.....	V
LISTE DES TABLEAUX.....	IX
LISTE DES FIGURES.....	X
LISTE DES ABRÉVIATIONS.....	XII
REMERCIEMENTS.....	XIV
DÉDICACE.....	XV
I-INTRODUCTION.....	1
1. LE SYNDROME DE DiGEORGE.....	2
1.1 Phénotype et origine embryologique du syndrome de DiGeorge.....	2
1.2 Défauts génétiques.....	3
1.3 Gènes candidats.....	5
1.4 ZNF74 pourrait-il être responsable des défauts de développement?.....	6
2. LA FAMILLE DES GÈNES À MOTIF DOIGT DE ZINC.....	8
2.1 Le motif doigt de zinc Cys <sub>2</sub> His <sub>2</sub> : ses propriétés de liaison aux acides nucléiques.....	8
2.2 Domaines associés au motif Cys <sub>2</sub> His <sub>2</sub> : un rôle de régulation transcriptionnelle.....	9
3. CARACTÉRISATION DE ZNF74.....	11
3.1 ZNF74 comme prototype des protéines KRAB multidoigt.....	11

3.2 ZNF74 comme protéine doublement impliquée dans la régulation de la transcription et dans la maturation de l'ARN.....	13
4. OBJECTIFS DU PROJET DE RECHERCHE.....	14
II-MATÉRIEL ET MÉTHODES.....	16
1. EXPRESSION ET PURIFICATION DES PROTÉINES DE FUSION.....	17
1.1 Constructions utilisées.....	17
1.2 Expression bactérienne des protéines de fusion avec MBP.....	18
1.3 Purification des protéines de fusion avec MBP sur colonne d'affinité.....	19
2. BUVARDAGE DE TYPE SOUTH-WESTERN.....	19
3. SÉLECTION D'OLIGONUCLÉOTIDES DÉGÉNÉRÉS (N16) SUR COLONNE D'AFFINITÉ.....	21
3.1 Préparation des oligonucléotides dégénérés double brin.....	21
3.2 Purification séquentielle des oligonucléotides dégénérés sur résine d'affinité.....	22
3.2.1 Préparation des résines.....	22
3.2.2 Sélection sur résines d'affinité des oligonucléotides dégénérés double brin et amplification.....	23
3.2.3 Amplification par PCR des oligonucléotides double brin sélectionnés.....	24
3.3 Clonage des oligonucléotides double brin sélectionnés.....	25
3.4 Séquençage.....	26
3.5 Analyse des séquences.....	26
4. ÉTUDES DE LIAISON À L'ADN PAR RETARDEMENT SUR GEL.....	27
4.1 Oligonucléotides utilisés.....	27

4.1.1 Oligonucléotides simple brin.....	27
4.1.2 Oligonucléotides double brin.....	28
4.1.3 Oligonucléotides hétéroduplex.....	28
4.2 Préparation des oligonucléotides double brin et hétéroduplex.....	29
4.3 Marquage des oligonucléotides simple brin, double brin et hétéroduplex.....	29
4.4 Détermination de l'activité spécifique et purification des oligonucléotides.....	30
4.4.1 Détermination de l'activité spécifique.....	30
4.4.2 Isolation des oligonucléotides marqués.....	30
4.4.3 Détermination de la concentration des oligonucléotides.....	31
4.5 Liaison des protéines aux oligonucléotides.....	32
4.6 Séparation des complexes protéine/ADN de l'ADN non lié par retardement sur gel.....	33
<b>III-RÉSULTATS.....</b>	<b>34</b>
1. LIAISON DE ZNF74 À L'ADN: ÉTUDE PAR BUVARDAGE DE TYPE SOUTH-WESTERN.....	35
2. SÉLECTION D'OLIGONUCLÉOTIDES DÉGÉNÉRÉS DOUBLE BRIN AVEC LA PROTÉINE ZNF74 IMMOBILISÉE SUR RÉSINE.....	38
2.1 Sélection d'oligonucléotides double brin par purification/amplification.....	38
2.2 Analyse des séquences d'oligonucléotides double brin sélectionnés.....	48
3. ÉTUDE DE LIAISON DE ZNF74 À L'ADN PAR RETARDEMENT SUR GEL.....	57
3.1 Étude de la liaison à l'ADN simple brin.....	58
3.1.1 Effet de la concentration de protéines pour la formation de complexes protéine/ADN et spécificité d'interaction.....	58

3.1.2 Importance de l'intégrité de la structure doigt de zinc pour la liaison à l'ADN.....	64
3.1.3 Liaison d'oligonucléotides simple brin sélectionnés aléatoirement.....	64
3.2 Étude de la liaison à l'ADN double brin.....	68
3.3 Étude de la liaison à l'ADN hétéroduplex.....	71
IV-DISCUSSION.....	78
1. SÉLECTION DE SÉQUENCES DE LIAISON À L'ADN DOUBLE BRIN.....	79
2. LIAISON DE ZNF74 À L'ADN SIMPLE BRIN.....	86
3. CONCLUSION.....	88
V-RÉFÉRENCES.....	90
VI-ANNEXE I.....	XVI
VII-ANNEXE II.....	XXI

**LISTE DES TABLEAUX**

Tableau I. ADN utilisé pour sélection et récupéré après amplification à chaque étape de purification.....	45
--	----

## LISTE DES FIGURES

### III-RÉSULTATS

Figure 1. Buvar dage de type South-Western.....	37
Figure 2. Représentation schématique de la procédure cyclique de purification/amplification des oligonucléotides double brin.....	40
Figure 3. Analyse par électrophorèse des protéines de fusion provenant d'extraits bactériens.....	43
Figure 4. Détermination du nombre maximum de cycles PCR pour amplification spécifique des oligonucléotides double brin.....	47
Figure 5. Identification sur gel d'acrylamide des clones plasmidiques contenant un oligonucléotide sélectionné comme insert.....	50
Figure 6. Sélection d'oligonucléotides double brin par MBP-ZNF74-R <sub>1</sub> immobilisée sur résine (première expérience).....	53
Figure 7. Sélection d'oligonucléotides double brin par MBP-ZNF74-R <sub>1</sub> immobilisée sur résine (deuxième expérience).....	55
Figure 8. Analyse par électrophorèse des protéines de fusion solubles utilisées pour les essais de retardement.....	60
Figure 9. Étude par retardement sur gel de la liaison des protéines ZNF74-R <sub>1</sub> et Kid <sub>1</sub> à l'ADN simple brin.....	63
Figure 10. Étude par retardement sur gel de la liaison de la protéine ZNF74-R <sub>1</sub> à des oligonucléotides simple brin de taille différente.....	67
Figure 11. Étude par retardement sur gel de la liaison de la protéine ZNF74-R <sub>1</sub> à l'ADN double brin.....	70
Figure 12. Étude par retardement sur gel de la liaison des protéines ZNF74-R <sub>1</sub> et Kid <sub>1</sub> à trois différentes structures de l'ADN.....	73
Figure 13. Étude par retardement sur gel de la liaison de la protéine ZNF74-R <sub>1</sub> à l'ADN hétéroduplex.....	76

IV-DISCUSSION

Figure 14. L'analyse stéréochimique des douze motifs Cys<sub>2</sub>His<sub>2</sub> de ZNF74 permet de prédire lesquels sont probablement capables de lier spécifiquement l'ADN.....83

VI-ANNEXE I

Figure 1. Sélection d'oligonucléotides double brin par MBP-Kid<sub>1</sub> immobilisée sur résine.....XVIII

Figure 2. Sélection d'oligonucléotides double brin par MBP immobilisée sur résine.....XX

VII-ANNEXE II

Figure 1. Fichier d'entrée du programme Gibbs.....XXII

Figure 2. Fichier de sortie du programme Gibbs.....XXIV

## LISTE DES ABRÉVIATIONS

aa: acides aminés

BSA: albumine bovine du sérum

CATCH22: Cardiac defect, Abnormal facies, Thymic hypoplasia, Cleft palate and Hypoparathyroidism resulting from 22q11.2 deletions

CTD: domaine C-terminal

DGS: syndrome de DiGeorge

DTT: dithiothreitol

*E. coli*: *Escherichia coli*

EDTA: disodium ethylenediaminetetra-acetate dihydrate

Hepes: N-[2[Hydroxyethyl] piperazine-N-[2-ethanesulfonic acid]]

IPTG: isopropyl- $\beta$ -D-thiogalactopyranoside

kb: kilobases

LS: large sous-unité

MBP: maltose binding protein

nt: nucléotides

P.M.: poids moléculaire

pb: paires de bases

PCR: réaction en chaîne avec la polymérase

PNK: kinase de polynucléotides T<sub>4</sub>

pol II: ARN polymérase II

pol IIa: ARN polymérase II hypophosphorylée

pol IIo: ARN polymérase II hyperphosphorylée

SDS: dodecyl sulfate de sodium

SSB: single strand binding protein

TBE: tampon Tris-borate/EDTA

TE: tampon Tris-EDTA

VCFS: syndrome vélo-cardio-facial

## REMERCIEMENTS

- Ma plus sincère reconnaissance au Dr Muriel Aubry pour sa constante participation au cours de ma formation. Son support et ses encouragements m'auront aidée à mener à terme cette maîtrise.
- Un merci tout particulier à Martine Bazinet pour son aide technique, ses encouragements et son amitié.
- Merci à Steves Morin, mon ami d'enfance, pour ses conseils techniques.
- Merci à Isabelle Blain pour l'excellence de son travail de secrétariat tout au long de ma maîtrise.
- Merci à la Fondation Dr Georges Phénix pour son support financier.

Soyez tous et toutes assurés de mes plus sincères remerciements.

À toutes les personnes qui m'ont  
aidée, de près ou de loin...

## **I-INTRODUCTION**

## 1. LE SYNDROME DE DIGEORGE

### 1.1 Phénotype et origine embryologique du syndrome de DiGeorge

Le syndrome de DiGeorge (DGS) se caractérise par un défaut du développement de structures embryologiques comme le thymus, les parathyroïdes et les gros vaisseaux du cœur qui dérivent des arcs et poches pharyngés 3 et 4. Les anomalies résultantes incluent l'absence ou l'hypoplasie du thymus et des parathyroïdes, ainsi que des malformations au niveau des gros vaisseaux du cœur, mais plus de 30 différentes manifestations du syndrome ont été rapportées (Goldberg *et al.*, 1993, Wilson *et al.*, 1993). Parmi les manifestations cliniques majeures du DGS, on observe des anomalies au niveau des fonctions cardiaques, des infections récurrentes chez l'enfant dues à une immunité cellulaire déficiente, des signes d'hypoparathyroïdisme pouvant mener à des convulsions hypocalcémiques, des anomalies crânio-faciales et un retard mental. Soixante-quinze pour-cent des nouveau-nés meurent de déficiences cardiaques peu après la naissance. Les premières indications d'un possible mécanisme cellulaire responsable du DGS proviennent des expériences de Kirby *et al.* (1984) qui démontraient que l'ablation des cellules occipitales prémigratoires de la crête neurale chez l'embryon de poulet produisait des malformations analogues à celles du DGS chez l'humain. On pense actuellement que le développement normal des gros vaisseaux du cœur et en particulier la septation normale de l'aorte et de l'artère pulmonaire ainsi que le développement du thymus et des parathyroïdes dépendent d'une migration et d'une interaction normale de cellules de la crête neurale au niveau des arcs et poches

pharyngés 3 et 4, des structures embryonnaires à partir desquelles se développent les principaux organes affectés dans le syndrome de DiGeorge.

## 1.2 Défauts génétiques

Des analyses cytogénétiques et moléculaires avec différentes sondes d'ADN servant de marqueurs pour la localisation chromosomique précise ont révélé que des microdélétions hémizygotiques du chromosome 22q11.2 sont présentes dans plus de 90% des cas de DGS (Aubry *et al.*, 1993, Driscoll *et al.*, 1992, Scambler *et al.*, 1991). Plus récemment, des délétions hémizygotiques du chromosome 22q11.2 ont aussi été décelées chez des patients présentant un désordre de développement, le syndrome vélo-cardio-facial (VCFS) qui présente un recouvrement phénotypique avec le DGS (Driscoll *et al.*, 1992, Kelly *et al.*, 1993); il est possible que le ou les mêmes gènes soient responsables de ces deux syndromes. Pour signifier que des manifestations phénotypiques variables pourraient avoir une origine génétique commune, le DGS, le VCFS et d'autres cas de délétion du chromosome 22 ont récemment été décrits comme faisant partie d'un regroupement clinique unique, le CATCH 22 (Cardiac defect, Abnormal facies, Thymic hypoplasia, Cleft palate and Hypoparathyroidism resulting from 22q11.2 deletions) (Wilson *et al.*, 1993). Le DGS et le VCFS surviennent de façon autosomale dominante sporadique ou familiale avec une incidence combinée de 1:10 000 (incidence de 1:20 000 pour chaque syndrome) (Emanuel *et al.*, 1993, Kelly *et al.*, 1993, Wilson *et al.*, 1991, Wilson *et al.*, 1992). Une étude plus récente quoiqu'encore préliminaire a révélé une incidence de 1/4000 pour les délétions au niveau du 22q11.2 associées à des

déficiences néonatales (étude sur 40 000 naissances enregistrées en Angleterre) (Burn *et al.*, 1995).

Des analyses moléculaires de la région 22q11.2 de nombreux patients atteints de DGS/VCFS ont été réalisées par différents groupes et ont révélé que la région critique de délétion comprend environ 1500 kb (Emanuel *et al.*, 1993, Driscoll *et al.*, 1992, Gong *et al.*, 1996, Glover, 1995). La large taille de la délétion chez les patients atteints de DGS/VCFS a suggéré que ce désordre pourrait être dû à un syndrome de gènes contigus, requérant la délétion de plusieurs gènes physiquement contigus sur le chromosome 22 plutôt qu'un défaut au niveau d'un seul gène. Deux régions critiques de délétions minimales qui ne se chevauchent pas ont été identifiées (Halford *et al.*, 1993, Kurahashi *et al.*, 1997, O'Donnell *et al.*, 1997) et une translocation balancée avec un point de bris à l'extérieur des deux régions critiques de délétion a aussi été identifiée chez un patient unique nommé ADU (Budarf *et al.*, 1995). Ceci suggère que ces réarrangements chromosomiques doivent influencer à distance l'expression du ou des gènes responsables du DGS par effet positionnel. Il a été rapporté que des effets positionnels peuvent affecter l'expression de gènes sur des distances pouvant aller jusqu'à 400 kb (Bedell *et al.*, 1996). Donc, on pense actuellement qu'un ou plusieurs gènes localisés dans la large région critique de 1,5 Mb pourraient être impliqués dans ce désordre complexe de microdélétion (Glover, 1995).

Deux patients uniques sans réarrangement chromosomique détectable ont aussi été identifiés. Ceci suggère la possibilité de mutations ponctuelles au niveau des gènes responsables ou de phénocopie (imitation de défaut génétique par préjudice

environnemental) avec une possible implication d'agents tératogéniques (éthanol, dérivés de l'acide rétinoïque) (Ammann *et al.*, 1982, Lammer *et al.*, 1985).

### 1.3 Gènes candidats

Depuis quelques années, les cartes physiques et de transcription de la région de délétion chez les patients DGS/VCFS ont été construites, permettant l'identification de plusieurs gènes candidats (Aubry *et al.*, 1993, Goldberg *et al.*, 1993, Emanuel *et al.*, 1993, Kelly *et al.*, 1993, Gong *et al.*, 1997). En particulier, des travaux systématiques pour isoler de nouveaux gènes à motifs doigt de zinc et pour positionner ces derniers sur les cartes génétiques disponibles (Aubry *et al.*, 1992, Hoovers *et al.*, 1992, Huebner *et al.*, 1991) ont permis l'isolation de 23 nouveaux gènes à motifs doigt de zinc dans notre laboratoire dont un, ZNF74, qui est localisé sur le chromosome 22q11.2 (Aubry *et al.*, 1992) et est présent en une seule copie chez la majorité des patients atteints du DGS (Aubry *et al.*, 1993). ZNF74 a été un des premiers gènes, absent sur un des deux chromosomes 22 des patients DGS/VCFS, à être isolé. Plus récemment, il a été localisé proche de la région de délétion critique minimum télomérique (Kurahashi *et al.*, 1997, O'Donnell *et al.*, 1997). À l'heure actuelle, on sait qu'il y a au moins 15 gènes dans la large région de délétion de 1,5 Mb. On compte en particulier LZTR1 (gène encodant une séquence leucine zipper), TUPLE1 (un régulateur potentiel de la transcription aussi appelé HIRA), LAN (une protéine de membrane intégrale aussi nommée DGR2 ou IDD), CTP/DGR5 (une protéine de transport du citrate), CLT (protéine reliée à la clathrine) et DGR6 (relié à une protéine gonadale de drosophile et aussi à la laminine), GSCL

("goosecoid-like"), DVL-22 ("dishevelled-like"), ARVCF (un gène à motif *armadillo*), et T10 (un gène à fonction inconnue). De plus, le clonage de DGR3, un gène localisé au point de bris de la translocation chez le patient ADU décrit plus tôt, a été rapporté (Budarf *et al.*, 1995). À cause des caractéristiques génétiques de ce désordre de microdélétion (délétion de large taille, régions de délétions minimales qui ne se chevauchent pas et translocation à l'extérieur des régions de délétions minimales), tous les gènes de la large région critique demeurent des candidats à étudier pour déterminer leur implication possible dans les aspects phénotypiques variés du DGS/VCFS.

#### **1.4 ZNF74 pourrait-il être responsable des défauts de développement?**

Des analyses par Northern blot ont révélé que ZNF74 est exprimé dans plusieurs tissus embryonnaires humains (23<sup>ème</sup> semaine de gestation) tandis que le message est indécelable dans les tissus adultes correspondants (Aubry *et al.*, 1993). Des études d'hybridation *in situ* très récentes indiquent de plus que le profil d'expression de ZNF74 est restreint tôt dans l'embryogenèse chez l'embryon humain (6,5-7 semaines) et se limite à des dérivés de la crête neurale incluant les gros vaisseaux du coeur (aorte et artère pulmonaire) affectés dans le DGS ainsi qu'à des épithéliums des voies respiratoires (Philippe Ravassard *et al.*, manuscrit en préparation).

Compte tenu de son profil d'expression durant l'embryogenèse, de sa localisation à proximité de la seconde région de délétion critique minimum et de son

rôle potentiel comme régulateur de l'expression génique, ZNF74 apparaît comme un très bon candidat pour le DGS.

Dans plusieurs cas, la caractérisation de modèles animaux pathologiques procure d'importantes données en ce qui concerne la nature cellulaire et moléculaire des défauts qui contribuent à la pathogénèse. Un groupe de défauts congénitaux ressemblants au DGS a été observé chez des rats nouveau-nés lorsqu'un agent chélatant le zinc, la bisdiamine, avait été administré à la mère en gestation à un moment critique du développement embryonnaire (E9-E12; période de développement des structures affectées dans le DGS) (Oster *et al.*, 1983). On peut émettre l'hypothèse qu'en chélatant le zinc, la bisdiamine peut réduire la concentration de zinc libre empêchant le bon fonctionnement de protéines dont l'activité requiert du zinc telle ZNF74. Un deuxième modèle animal, présentant des caractéristiques phénotypiques similaires à celles rencontrées dans le DGS, a été généré chez la souris par la destruction du facteur de transcription à homéobox HoxA3 (Hox1.5) par recombinaison homologue, créant ainsi une souris "knock-out" pour ce gène (Chisaka *et al.*, 1991, Manley *et al.*, 1995). Deux différences majeures existent entre ce modèle et le syndrome génétique humain: 1) le DGS est transmis de façon autosomale dominante alors que le phénotype HoxA3<sup>-/-</sup> est autosomal récessif, et 2) l'homologue humain de HoxA3 de la souris se retrouve sur le chromosome humain 7 (et non 22). Les similarités phénotypiques et les différences génétiques entre les souris HoxA3<sup>-/-</sup> et les patients atteints du DGS suggèrent que le facteur de transcription HoxA3 n'est pas le gène impliqué dans la majorité des cas de DGS chez l'humain mais qu'il ait une fonction dans une voie de signalisation développementale

commune au(x) gène(s) impliqué(s) dans le DGS. Donc, il peut être proposé que ZNF74 et/ou les autres gènes candidats pour le DGS puissent agir comme régulateurs en amont ou effecteurs en aval de HoxA3. Comme exemple d'une telle régulation, l'interaction directe de la protéine à motifs doigt de zinc Krox20 avec les séquences régulatrices du gène Hox2.8 a récemment été démontrée dans l'activation de ce gène homéobox (Sham *et al.*, 1993).

## 2. LA FAMILLE DES GÈNES À MOTIF MULTIDOIGT DE ZINC

### 2.1 Le motif doigt de zinc Cys<sub>2</sub>His<sub>2</sub>: ses propriétés de liaison aux acides nucléiques

Les gènes à motifs doigt de zinc encodent des protéines contenant des motifs doigt de zinc répétés, chacun coordonnant un ion de zinc *via* des interactions spécifiques avec des histidines et/ou des cystéines. Les douze motifs doigt de zinc retrouvés chez ZNF74 correspondent à la séquence consensus suivante, **Cys-X<sub>2</sub>-Cys-X<sub>3</sub>-Phe-X<sub>5</sub>-Leu-X<sub>2</sub>-His-X<sub>3</sub>-His**, séquence caractéristique de la famille Cys<sub>2</sub>His<sub>2</sub>. Ce motif consensus a premièrement été identifié dans la protéine TFIIIA de *Xenopus* et plus tard dans la protéine Kruppel chez la drosophile. À ce jour, on estime qu'il y aurait entre 300 et 700 gènes humains encodant une famille de protéines à motifs doigt de zinc Cys<sub>2</sub>His<sub>2</sub> du type TFIIIA/Kruppel. Les protéines de cette famille les mieux caractérisées (telles Egr-1, Sp1, WT1, et GLI3 qui possèdent de trois à cinq motifs doigt de zinc) ont un rôle démontré dans la croissance et la différenciation cellulaire (Mitchell *et al.*, 1989, Pelletier *et al.*, 1991, Mo *et al.*, 1997).

Typiquement, ces protéines à motifs doigt de zinc ont une fonction de facteur de transcription, activateurs ou répresseurs, se liant à l'ADN *via* la région à motifs doigt de zinc et régulant la transcription *via* un motif de régulation situé en dehors de la région doigt de zinc. Quelques protéines à motif multidoigt de zinc de la famille Cys<sub>2</sub>His<sub>2</sub> de *Xenopus* avaient préalablement été caractérisées comme des protéines liant l'ARN (c'est-à-dire p43, XFG 5-1, Xfin et TFIIIA) (Theunissen *et al.*, 1992, Andreazzoli *et al.*, 1993). Parmi celles-ci, seule TFIIIA avait été décrite comme ayant à la fois une fonction de liaison à l'ADN et à l'ARN. TFIIIA active la transcription des gènes codant pour la sous-unité 5S de l'ARN ribosomal par liaison à l'ADN de manière séquence-spécifique et participe au transport du transcrit ARN 5S du noyau vers le cytoplasme (Theunissen *et al.*, 1992). Plus récemment, une des isoformes de la protéine à quatre motifs doigt de zinc de la tumeur de Wilms, WT1, a aussi été caractérisée comme liant à la fois l'ADN et l'ARN (Caricasole *et al.*, 1996). Nos études ont aussi permis de déterminer que ZNF74 lie spécifiquement certains ARN (ex: poly(U) et poly(G)) (Grondin *et al.*, 1996) mais son activité de liaison à l'ADN n'a pas été caractérisée en détail dans le laboratoire.

## **2.2 Domaines associés au motif Cys<sub>2</sub>His<sub>2</sub>: un rôle de régulation transcriptionnelle**

Les gènes à motif multidoigt de zinc de la famille Cys<sub>2</sub>His<sub>2</sub> sont divisés en sous-classes s'appuyant sur la présence et l'identité de domaines conservés au cours de l'évolution localisés à l'extrémité N-terminale des protéines encodées (Knochel *et al.*, 1989, Numoto *et al.*, 1993, Klocke *et al.*, 1994, Bardwell *et al.*, 1994, Williams

*et al.*, 1995). Ces différents domaines sont connus comme étant: FAX (**F**inger **A**ssociate **boX**), FAR (**F**inger **A**ssociated **R**epeats), POZ (**P**Ox virus and **Z**inc fingers, aussi connu sous l'appellation **ZiN**), SCAN et KRAB (**K**Ruppel-Associated **B**ox). Quoiqu'ils puissent jouer un rôle important pour dicter les interactions spécifiques avec d'autres protéines régulatrices tel que proposé pour le domaine POZ (Bardwell *et al.* (1994)) et le domaine KRAB (Friedman *et al.*, 1996, Moosmann *et al.*, 1996, Kim *et al.*, 1996), les rôles précis de ces domaines sont encore mal compris. Le domaine KRAB est retrouvé exclusivement en association avec des protéines à motif multidoigt Cys<sub>2</sub>His<sub>2</sub>. En se basant sur des études d'hybridation, on estime que le tiers des 300-700 gènes à motifs doigt de zinc Cys<sub>2</sub>His<sub>2</sub> présents dans le génome humain contiennent un motif KRAB (Bellefroid *et al.*, 1991). Le motif KRAB est composé de deux domaines distincts, généralement encodés par deux exons séparés nommés boîte-A et boîte-B. Il a récemment été démontré que la boîte-A peut conférer une répression transcriptionnelle indépendante de la distance quand elle est fusionnée à un domaine de liaison à l'ADN hétérologue tel que le domaine de liaison Gal4 chez la levure (Pengue *et al.*, 1996). Contrairement à la boîte-A, la boîte-B est elle-même dépourvue d'activité répressive mais elle augmente celle de la boîte-A. Des essais de transcription indiquent que le domaine KRAB est capable de réprimer l'activité basale de transcription et que cette répression requiert la présence d'une boîte TATA sur le promoteur. Ceci suggère l'existence d'interactions directes et indirectes entre le domaine KRAB et la machinerie de transcription basale dépendante de la boîte TATA (Pengue *et al.*, 1996). De plus, il a récemment été proposé que les domaines KRAB puissent agir comme répresseurs par leur

interaction avec un co-répresseur spécifique. En effet, un co-répresseur ubiquiste nommé KAP/TIF1- $\beta$  interagissant avec le domaine KRAB de plusieurs protéines KRAB multidoigt a été identifié (Friedman *et al.*, 1996). Le patron d'expression de KAP/TIF1- $\beta$  contraste avec l'expression spatio-temporelle restreinte qui caractérise généralement les protéines KRAB/Cys<sub>2</sub>His<sub>2</sub> (par exemple, ZNF74 qui est exprimé seulement chez l'embryon) (Aubry *et al.*, 1993). Ceci suggère que KAP/TIF1- $\beta$  pourrait servir comme co-répresseur général pour les protéines contenant un domaine KRAB exprimées à des temps différents et dans différents tissus, les domaines doigt de zinc de ces protéines dictant peut-être les séquences spécifiques d'ADN à être ciblées. Malgré ces nouvelles hypothèses, les protéines KRAB/Cys<sub>2</sub>His<sub>2</sub> appartiennent à une grande famille d'une nouvelle classe de répresseurs potentiels de la transcription qui demeure encore peu explorée.

### 3. CARACTÉRISATION DE ZNF74

#### 3.1 ZNF74 comme prototype des protéines KRAB multidoigt

La plupart des gènes à motif multidoigt de zinc de la sous-famille KRAB/Cys<sub>2</sub>His<sub>2</sub> ont été découverts récemment par l'entremise des différents projets de cartographie du génome et les informations concernant les protéines qu'ils encodent sont vraiment très limitées. Contrairement aux gènes à motifs doigt de zinc déjà bien caractérisés et encodant de un à cinq doigts, toutes les séquences des gènes KRAB/Cys<sub>2</sub>His<sub>2</sub> retrouvées dans les banques de données définissent un motif

multidoigt excédant largement cinq doigts. Des études cristallographiques ont suggéré que la présence d'un motif de plus de cinq doigts spécifierait un type différent d'interactions avec les acides nucléiques, plus de trois doigts ne pouvant s'inscrire de façon continue dans le sillon majeur de l'ADN (Klug *et al.*, 1995). Il est possible que les protéines à motif multidoigt (>cinq doigts) aient des propriétés biochimiques et fonctionnelles différentes des protéines ayant un nombre plus limité de doigts. À titre de premier exemple, il a été montré récemment que ZNF74 (Grondin *et al.*, 1996) ainsi que d'autres protéines KRAB à motif multidoigt, soit la protéine de souris Zfp59 et la protéine de rat Kid<sub>1</sub> (Francine Côté *et al.*, résultats non publiés) sont très fortement associées à une structure fonctionnelle du noyau, la matrice nucléaire, qui correspond à une fraction insoluble obtenue après traitement des noyaux cellulaires avec des détergents, des nucléases et une concentration de sels élevée; ces protéines ne sont pas détectables dans la fraction soluble du noyau. Ces résultats suggèrent que l'association à la matrice est probablement une propriété de la plupart des protéines KRAB/Cys<sub>2</sub>His<sub>2</sub>; cette propriété n'est pas caractéristique des facteurs transcriptionnels en général ni des protéines ayant un nombre restreint de motifs doigt de zinc qui partitionnent entre la fraction soluble et insoluble du noyau. Il a été suggéré que la matrice nucléaire puisse être une structure de base sur laquelle des domaines fonctionnels sous-nucléaires sont organisés (Penman, 1995). Ceux-ci incluent les sites de réplication de l'ADN, de synthèse d'ARN messagers et de stockage des spliceosomes (Hozak *et al.*, 1993, Wansink *et al.*, 1993, Jackson *et al.*, 1993, Carmo-Fonseca *et al.*, 1992, Spector *et al.*, 1991).

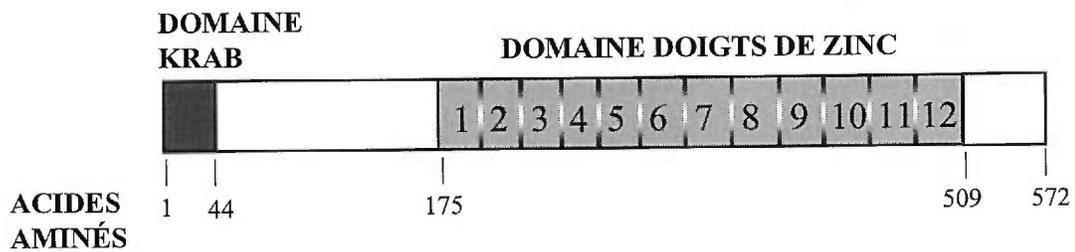
### 3.2 ZNF74 comme protéine doublement impliquée dans la régulation de la transcription et dans la maturation de l'ARN

Alors que la présence d'un motif KRAB et d'un motif doigt de zinc permet de suggérer que ZNF74 pourrait jouer un rôle de facteur transcription, plusieurs résultats récents du laboratoire nous ont amené à proposer que ZNF74 pourrait avoir aussi un rôle dans la maturation de l'ARN. Nous avons récemment démontré par des analyses de Far-Western et de co-immunoprécipitation que ZNF74 interagit, *via* son domaine doigt de zinc, avec une protéine de la matrice nucléaire de 250 kDa. Cette protéine correspond à la forme hyperphosphorylée de la large sous-unité (LS) de l'ARN polymérase II (pol II), l'enzyme impliquée dans la synthèse d'ARN pré-messagers (Grondin *et al.*, 1997). La forme hypophosphorylée de la LS de l'ARN polymérase II (pol IIa) est la forme la plus efficacement recrutée dans le complexe de pré-initiation de la transcription au niveau des promoteurs alors que sa forme hyperphosphorylée, appelée pol IIo, est celle qui est observée dans le complexe d'élongation durant la synthèse d'ARN pré-messagers (Dahmus, 1996, Lu *et al.*, 1991, Chesnut *et al.*, 1992, Weeks *et al.*, 1993, Steinmetz, 1997). L'hyperphosphorylation, qui se produit sur les répétitions d'heptapeptides du domaine C-terminal (CTD) de la pol IIo, est apparemment non essentielle pour l'élongation de la transcription mais est requise pour la maturation des ARN pré-messagers, c'est-à-dire pour l'épissage et le clivage de poly A (Serizawa *et al.*, 1993, Makela *et al.*, 1995, McCracken *et al.*, 1997). Outre ZNF74, les seules autres protéines rapportées pour interagir directement avec la pol IIo appartiennent à trois classes de protéines participant au processus de transformation de l'ARN: des

protéines impliquées dans l'addition de la coiffe à l'extrémité 5' des ARN messagers, des protéines SR impliquées dans l'épissage et les facteurs CstF/CPSF impliqués dans le clivage et la polyadénylation des ARN messagers (Cho *et al.*, 1997, McCracken *et al.*, 1997, Yuryev *et al.*, 1996, Bourquin *et al.*, 1997). De façon intéressante, il a aussi été démontré que ZNF74 colocalise avec la pol II $\alpha$  et un marqueur (SC35) de microdomaines du noyau riches en facteurs d'épissage ("speckles") (Grondin *et al.*, 1997). Ainsi, il est tentant de spéculer que ZNF74 pourrait être doublement impliqué dans la régulation transcriptionnelle *via* sa liaison à l'ADN et son domaine KRAB et dans la maturation de l'ARN *via* sa liaison à l'ARN et son interaction avec la pol II $\alpha$  au niveau des domaines riches en facteurs d'épissage. Il est possible de postuler que ZNF74, comme gène exprimé lors du développement, pourrait réguler d'un côté la transcription de certain(s) gène(s) cible(s) et d'un autre côté, moduler l'épissage alternatif de gène(s) de développement important(s). En fait, la plupart des ARN pré-messagers de mammifères contiennent de multiples introns qui permettent à différentes molécules d'ARN messagers d'être générées à partir du même ARN pré-messager par épissage alternatif et de nombreux gènes ont été identifiés comme utilisant l'épissage alternatif pour réguler leur expression durant le développement ou la différenciation (Yuryev *et al.*, 1996).

#### **4. OBJECTIFS DU PROJET DE RECHERCHE**

Tel que mentionné précédemment, le gène ZNF74, dont une copie est absente chez la majorité des patients atteints du DGS, code pour une protéine à motif multidoigt de zinc contenant un motif conservé de répression, le motif KRAB (schéma ci-dessous).



Cette protéine fait partie de la famille à motifs KRAB/Cys<sub>2</sub>His<sub>2</sub>. La fonction cellulaire et les caractéristiques biochimiques des protéines de cette famille demeurent encore mal connues. Comme la protéine à multidoigt (neuf motifs doigt de zinc) TFIIIA de *Xenopus* semble avoir la capacité de lier de façon dualiste les acides nucléiques ADN et ARN, nous émettons l'hypothèse que ZNF74 puisse de façon similaire lier ces deux type d'acides nucléiques. Nous proposons de déterminer si ZNF74 qui possède 12 motifs doigt de zinc pourrait lier une séquence d'ADN spécifique. Jusqu'à présent, une liaison à des séquences spécifiques d'ADN n'a pu être mise en évidence pour aucune autre protéine à multidoigt possédant plus de cinq doigts que TFIIIA et il n'existe pas d'exemple de ce type de liaison dualiste pour des protéines de mammifères. Sachant que ZNF74 lie l'ARN simple brin, nous proposons aussi de déterminer si ZNF74 peut se lier à l'ADN simple brin.

Les buts spécifiques du projet de recherche présenté dans ce mémoire seront de déterminer:

- 1) si ZNF74, défini précédemment comme une protéine liant l'ARN, peut lier l'ADN double brin;
- 2) si ZNF74 se lie à l'ADN double brin *via* des séquences nucléotidiques spécifiques tel que démontré pour la plupart des facteurs transcriptionnels;
- 3) si ZNF74, qui reconnaît l'ARN simple brin, peut se lier à des structures ou des séquences spécifiques d'ADN simple brin.

## **II-MATÉRIEL ET MÉTHODES**

## 1. EXPRESSION ET PURIFICATION DES PROTÉINES DE FUSION

### 1.1 Constructions utilisées

Les constructions mentionnées ci-dessous, hormis celle encodant la protéine de fusion Kid<sub>1</sub>, ont été décrites précédemment (Grondin *et al.*, 1996). Pour générer la protéine ZNF74 (complète ou tronquée) ou la protéine Kid<sub>1</sub> en fusion avec la protéine liant le maltose (MBP pour "maltose binding protein"), le vecteur pMAL-c (New England Biolabs) exprimable dans des bactéries et encodant MBP est utilisé. Des fragments de restriction encodant les protéines d'intérêt ont été clonés en phase en aval de la séquence de MBP et amplifiés par la réaction en chaîne avec la polymérase (PCR). Les fragments de restriction pour ZNF74 ont été obtenus en utilisant comme matrice l'ADNc de ZNF74-1 précédemment isolé dans le laboratoire (Aubry *et al.*, 1993) (numéro d'accession dans les banques de données: EMBL X71623) et ceux pour Kid<sub>1</sub> ont été obtenus par reverse transcriptase faite à partir d'ARN de rein de rat (numéro d'accession dans les banques de données: M96548). Les fragments de PCR préparés avec des amorces contenant un site *Xba* I ont été digérés avec l'enzyme de restriction *Xba* I et sous-clonés dans le site *Xba* I du vecteur pMAL-c. Les différentes constructions préparées par clonage de fragments de PCR, soient pMAL-c-Kid<sub>1</sub>, pMAL-c-ZNF74, pMAL-c-Zn, correspondent respectivement à la région codante complète de la protéine Kid<sub>1</sub> (nucléotides (nt) 312-2043; MBP-Kid<sub>1</sub> (acides aminés (aa) 1-577), poids moléculaire (P.M.) = 106 500) ainsi qu'à celle de la protéine ZNF74 (nt 163-1881; MBP-ZNF74 (aa 1-572), P.M. = 106 000) et à la région à motifs doigt de zinc de la protéine ZNF74 (nt 685-1689; MBP-Zn (aa 175-509),

P.M.= 80 200). De plus, la construction pMAL-c-ZNF74-R<sub>1</sub>, préparée par clonage d'un fragment de restriction *Eco* RI de 1,9 kilobases (kb) de l'ADNc de ZNF74 (nt 478-2416) dans le site *Eco* RI du vecteur pMAL-c, correspond à la région codante complète de ZNF74 moins les 105 premiers acides aminés (MBP-ZNF74-R<sub>1</sub> (aa 106-572), P.M. = 94 300). Quand aucun insert n'est cloné dans le vecteur pMAL-c en aval de la séquence MBP, ce vecteur génère la protéine de fusion MBP-β-gal-α (P.M. = 51 000) qui a été utilisée comme contrôle pour plusieurs expériences décrites dans ce mémoire.

## 1.2 Expression bactérienne des protéines de fusion avec MBP

En bref, des bactéries compétentes *Escherichia coli* DH5α (Sambrook *et al.*, 1989; chapitre 1) sont transformées avec le vecteur pMAL-c seul pour obtenir la protéine MBP et avec les constructions suivantes pour obtenir les protéines de fusion avec MBP: pMAL-c-ZNF74, pMAL-c-Zn, pMAL-c-ZNF74-R<sub>1</sub> et pMAL-c-Kid<sub>1</sub>. Tel que décrit dans Grondin *et al.* (1997), les bactéries (culture de 500 mL) sont mises à croître jusqu'à une densité optique de 0,3-0,5 à 600 nm. L'expression de la protéine de fusion est ensuite induite par addition d'isopropyl-β-D-thiogalactopyranoside (IPTG) (concentration finale de 0,3 mM) au milieu de culture et par le maintien de la pousse bactérienne pour 3 heures à 37°C. La culture bactérienne est resuspendue dans du tampon de lyse (Tris 200 mM à pH 7,5, NaCl 200 mM, β-mercaptoéthanol 10 mM) (50 mL) et est soniquée pour briser les parois bactériennes et ainsi libérer les

protéines dans le tampon. On obtient ainsi des extraits bactériens contenant les protéines de fusion solubles.

### **1.3 Purification des protéines de fusion avec MBP sur colonne d'affinité**

Tel que décrit dans Grondin *et al.* (1997), les protéines de fusion avec MBP, provenant des extraits bactériens solubles décrits à la section précédente, sont immobilisées sur une résine d'amylose (0,5 mL) (New England Biolabs) en incubant celle-ci pour 60 minutes avec la totalité des extraits bactériens; l'incubation se fait à 4°C avec une légère agitation (sur agitateur de type "Nutator"). Par la suite, la résine est déposée dans une colonne et lavée avec 20 volumes de tampon de lyse et avec 40 volumes de tampon d'élution (N-[2[Hydroxyethyl] piperazine-N-[2-ethanesulfonic acid]] (Hepes) 20 mM à pH 7,5, NaCl 100 mM). Ensuite, les protéines de fusion avec MBP immobilisées sont éluées avec 1,5 mL de tampon d'élution contenant du maltose (concentration finale de 10 mM). Selon la préparation et les protéines de fusion avec MBP, 100-1000 µg de protéines purifiées sont obtenues, tel qu'estimé par la méthode de Bradford ("Protein Assay", Bio-Rad). La pureté et l'intégrité des protéines éluées de la colonne d'affinité et ayant migrées sur un gel de polyacrylamide dénaturant en présence de dodecyl sulfate de sodium (SDS) sont confirmées après coloration avec du bleu de Coomassie (Dunbar, 1987). Les protéines purifiées sont aliquotées et conservées à -80°C.

## **2. BUVARDAGE DE TYPE SOUTH-WESTERN**

L'expérience de buvardage de type South-Western permettant de mettre en évidence des interactions protéine/ADN a été réalisée selon une adaptation des conditions décrites par Sukegawa *et al.* (1993) et Perez *et al.* (1996). En bref, les protéines d'intérêt purifiées (MBP-ZNF74, MBP-Zn et MBP) (0.5-1µg/puits) sont séparées par électrophorèse sur gel de polyacrylamide dénaturant en présence de SDS (Dunbar, 1987) et ensuite transférées sur membranes de nitrocellulose en utilisant un appareil de transfert électrophorétique de Bio-Rad ("Mini Trans-Blot electrophoretic transfer cell") selon le protocole du fabricant. Deux membranes identiques, A et B, sont pré-traitées respectivement dans un tampon de renaturation A (Tris-HCl 50 mM à pH 7,5, KCl 100 mM, Triton X-100 1%, glycérol 10%, ZnCl<sub>2</sub> 1 mM) et dans un tampon de renaturation B sans zinc et contenant un agent chélateur et un agent réducteur (Tris-HCl 50 mM à pH 7,5, KCl 100 mM, Triton X-100 1%, glycérol 10%, disodium ethylenediaminetetra-acetate dihydrate (EDTA) 50 mM, dithiothreitol (DTT) 10 mM) pour une période de 12 heures. Ensuite, elles sont préincubées respectivement dans le tampon de réaction A (Tris-HCl 50 mM à pH 7,5, KCl 100 mM, Triton X-100 0,1%, glycérol 10%, ZnCl<sub>2</sub> 0,1 mM) et dans le tampon de réaction B (Tris-HCl 50 mM à pH 7,5, KCl 100 mM, Triton X-100 0,1%, glycérol 10%, EDTA 10 mM, DTT 2 mM) pour une période de 2 heures avant l'addition de la sonde d'ADN. La sonde d'ADN utilisée correspond à de l'ADN génomique humain total, soniqué de manière à obtenir des fragments de la taille de 1 kb. La sonde est marquée radioactivement avec du [ $\alpha$ -<sup>32</sup>P] dCTP (DuPont NEN) en présence d'hexamères oligonucléotidiques aléatoires et du fragment Klenow de la polymérase I de l'ADN

(Trousse "Oligolabelling kit", Pharmacia Biotech). Les membranes et la sonde d'ADN marquée ( $5 \times 10^4$  cpm/mL; activité spécifique d'environ  $1-2 \times 10^9$  cpm/ $\mu$ g) sont incubées pour une période de 3h30. Les membranes sont ensuite lavées 5 fois (10 minutes par lavage) dans leur tampon de réaction respectif, séchées et exposées en contact avec un film X-Omat Blue XB-1(Kodak) pour autoradiographie à  $-80^\circ\text{C}$ . Toutes les incubations sont faites à  $4^\circ\text{C}$ .

### **3. SÉLECTION D'OLIGONUCLÉOTIDES DÉGÉNÉRÉS (N16) SUR COLONNE D'AFFINITÉ**

La sélection sur résine d'affinité d'oligonucléotides double brin dégénérés en 16 positions (N16) a été faite avec trois protéines immobilisées sur la résine, soit MBP-ZNF74-R<sub>1</sub>, MBP-Kid<sub>1</sub> et MBP à titre de contrôle. Les oligonucléotides sélectionnés une première fois sur résine d'affinité ont été amplifiés par PCR. Sept cycles successifs de purification/amplification des oligonucléotides double brin ont été réalisés avant leur clonage et leur séquençage. Le protocole de sélection d'oligonucléotides dégénérés sur colonne d'affinité décrit ci-dessous correspond à une adaptation des conditions décrites par Kinzler *et al.* (1989), He *et al.* (1995) et Arranz *et al.* (1997).

#### **3.1 Préparation des oligonucléotides dégénérés double brin**

Pour obtenir un mélange représentatif d'oligonucléotides dégénérés double brin à partir d'oligonucléotides dégénérés simple brin, on a procédé par amplification PCR. La réaction d'amplification est faite dans un volume de 25  $\mu$ L de tampon de PCR

(Tris-HCl 20 mM à pH 8,4, KCl 50 mM, MgCl<sub>2</sub> 1,9 mM), contenant des dNTP (concentration finale de 0,26 mM), 0,25 µg de protéine de liaison simple brin de *E. Coli* (SSB pour "single strand binding protein"), 10 ng d'oligonucléotides dégénérés N16 simple brin (oligo 842: 5'-GACAGTTGTTGTCTTCCTACTAGCTGCAG-N16-CTCGAGTGGGCGTAGTACTTCAT-3'), 200 ng de chacune des amorces (oligo 844 correspondant aux 29 premiers nucléotides de l'oligo dégénéré 842: 5'-GACAGTTGTTGTCTTCCTACTAGCTGCAG-3'; oligo 845 correspondant au complémentaire inverse des 23 derniers nucléotides de l'oligo dégénéré 842: 5'-ATGAAGTACTACGCCCACTCGAG-3') et 2,5 unités de la polymérase Taq (BRL). Tel que décrit dans la section *Résultats*, un maximum de 8 cycles de 1 minute à 94°C (dénaturation), 1 minute à 46°C (association de l'amorce et de l'ADN matrice) et 1 minute à 72°C (élongation) sont effectués pour l'amplification des oligonucléotides dégénérés double brin par PCR. Pour estimer la quantité d'oligonucléotides dégénérés amplifiés, un échantillon (5 µL) est déposé directement sur gel de polyacrylamide 12% (acrylamide 29,2%: bisacrylamide 0,8%) non dénaturant (Maniatis, page 6.39) et dosé par comparaison de l'intensité de coloration avec le bromure d'éthidium avec celle d'un standard de phage PhiX174 RF digéré *Hae* III.

## **3.2 Purification séquentielle des oligonucléotides dégénérés sur résine d'affinité**

### **3.2.1 Préparation des résines d'affinité**

Pour sélectionner des oligonucléotides double brin ayant une affinité pour les diverses protéines de fusion avec MBP, on utilise une résine d'amylose sur laquelle on a immobilisé chacune des protéines de fusion respectivement. Ainsi, des extraits bactériens solubles tels que décrits à la section 1.2 exprimant chacun une protéine de fusion donnée sont incubés dans des tubes eppendorfs de 1,5 mL contenant 50  $\mu$ L de résine d'amylose pré-lavée 3 fois avec 750  $\mu$ L de tampon de lyse (décrit à la section 1.2) (lavage par centrifugation). Après une incubation de 60 minutes à 4°C sur un agitateur de type "Nutator", les résines sont lavées 3 fois avec 750  $\mu$ L de tampon de lyse et 1 fois avec 750  $\mu$ L de tampon de liaison (Hepes 50 mM à pH 7,5, KCl 50 mM, MgCl<sub>2</sub> 5 mM, ZnSO<sub>4</sub> 10  $\mu$ M, glycérol 20 %) contenant 1 mM de DTT. Les résines ainsi traitées sont prêtes à être incubées avec des oligonucléotides dégénérés (voir section suivante). À cette étape, des échantillons contrôles de résine d'affinité sont resuspendus dans du tampon de Laemmli (glycérol 15%, SDS 2%, Tris 80 mM à pH 6,8, 0,01% bleu de bromophénol, 5%  $\beta$ -mercaptoéthanol) et soumis à une électrophorèse sur gel de polyacrylamide (Dunbar, 1987) pour évaluer la quantité de protéines de fusion avec MBP retenue par la résine d'affinité. La quantité de protéine de fusion est estimée après coloration avec du bleu de Coomassie par comparaison avec une protéine standard, soit l'albumine bovine du sérum (BSA). De façon générale, pour préparer les résines d'affinité devant servir à la sélection d'oligonucléotides dégénérés, on a utilisé une quantité d'extrait bactérien permettant d'obtenir une résine d'affinité retenant 1  $\mu$ g de protéine de fusion/50  $\mu$ L de résine.

### **3.2.2 Sélection sur résines d'affinité des oligonucléotides dégénérés double brin et amplification**

La réaction de liaison des oligonucléotides dégénérés aux résines d'affinité (50  $\mu\text{L}$ ) est faite dans un volume de 500  $\mu\text{L}$  de tampon de liaison (décrit à la section 3.2.1) contenant 2  $\mu\text{g}$  de poly dIdC, 1  $\mu\text{g}/\mu\text{L}$  de BSA acétylée, 1 mM de DTT et environ 100-400 ng d'oligonucléotides dégénérés double brin tels que préparés à la section 3.1. Les résines d'affinité (1  $\mu\text{g}$  de protéine de fusion/50  $\mu\text{L}$  de résine d'amylose, voir section précédente) et l'ADN sont incubés pendant 30 minutes à la température de la pièce avec une légère agitation (agitateur de type "Nutator"). Les résines d'affinité sont ensuite lavées 1 fois avec 750  $\mu\text{L}$  de tampon de liaison contenant 1mM de DTT, 3 fois avec 750  $\mu\text{L}$  de tampon de liaison modifié (Hepes 50 mM à pH 7,5, KCl 150 mM,  $\text{MgCl}_2$  5 mM,  $\text{ZnSO}_4$  10  $\mu\text{M}$ , glycérol 20 %), 3 fois avec 750  $\mu\text{L}$  de tampon PCR (Tris-HCl 20 mM à pH 8,4, KCl 50 mM,  $\text{MgCl}_2$  1,9 mM) et resuspendues dans 50  $\mu\text{L}$  de tampon PCR. Ensuite, les résines en suspension sont incubées 5 minutes à 95°C afin de dénaturer les protéines. Les surnageants des résines d'affinité (50  $\mu\text{L}$ ) contenant maintenant les oligonucléotides sélectionnés sont récupérés pour amplification subséquente par PCR.

### **3.2.3 Amplification par PCR des oligonucléotides double brin sélectionnés**

La totalité de l'ADN double brin sélectionné sur la résine d'affinité tel que récupéré à la section précédente (50  $\mu\text{L}$ ) est soumis à une première amplification

par PCR. Les 50  $\mu\text{L}$  récupérés contenant les oligonucléotides sélectionnés sont divisés dans 5 tubes (10  $\mu\text{L}$ /tube) et les amorces, le tampon, les nucléotides, le SSB et l'enzyme nécessaires pour faire une amplification dans les mêmes conditions que celles décrites à la section 3.1 sont ajoutés. Dans la plupart des cas, après dosage des produits de première amplification (5  $\mu\text{L}$ ) sur gel d'acrylamide non dénaturant (voir section 3.1), on effectue une deuxième amplification pour obtenir suffisamment d'oligonucléotides amplifiés. Généralement, pour la deuxième amplification, on utilise 100  $\mu\text{L}$  du matériel récupéré après la première amplification PCR, répartis dans 4 tubes (25  $\mu\text{L}$ /tube), et une deuxième réaction de PCR est effectuée dans les conditions décrites ci-dessus et à la section 3.1. Les oligonucléotides double brin obtenus par deuxième amplification sont dosés afin de pouvoir déterminer la proportion de cet ADN requise pour les autres cycles de purification sur résines d'affinité. Après la septième étape de purification sur résine, l'ADN récupéré et amplifié par deux étapes successives de PCR est utilisé pour l'étape de clonage décrite ci-dessous.

### **3.3 Clonage des oligonucléotides double brin sélectionnés**

Les oligonucléotides double brin obtenus tel que décrit plus haut après sept cycles de purification/amplification (100 ng) sont clonés (Sambrook *et al.*, 1989; chapitre 1) de façon orientée dans le vecteur pBluescript SK (Stratagene) préalablement digéré avec *Eco* RV et *Pst* I (30 ng). Avant leur clonage, les oligonucléotides sélectionnés sont soumis à l'action de la polymérase T4 (Sambrook *et al.*, 1989; chapitre 5) afin de

s'assurer qu'ils possèdent des extrémités franches, et digérés par l'enzyme *Pst* I, compte tenu de la présence d'un site *Pst* I à leur extrémité 5'. Après ligation du plasmide et des oligonucléotides sélectionnés en présence de ligase d'ADN T4, l'ADN est introduit dans des cellules compétentes *E. coli* DH5 $\alpha$  pour transformation (Sambrook *et al.*, 1989; chapitre 1). À partir des clones bactériens obtenus, on isole l'ADN plasmidique après lyse alcaline en présence de SDS selon le protocole de mini-préparation d'ADN décrit par Sambrook *et al.* (1989; chapitre 1).

### 3.4 Séquençage

La méthode de séquençage en présence de didéoxynucléotides, de [ $\alpha$ -<sup>35</sup>S] dCTP et de T7 DNA polymérase est telle que proposée par le manufacturier (Trousse "T7 Sequencing Kit", Pharmacia Biotech). L'ADN plasmidique circulaire (2 à 3  $\mu$ g) d'une mini-préparation préparée à partir de 1,5 mL de milieu de culture bactérienne est utilisé comme matrice et un oligonucléotide (5'-ATTAACCCTCACTAAAG-3') (20 ng) correspondant au promoteur T<sub>3</sub> est utilisé comme amorce pour l'élongation. Les produits de la réaction de séquençage sont séparés sur un gel de polyacrylamide 5% (acrylamide 38%: bisacrylamide 2%) dénaturant contenant 6,2 M d'urée (Sambrook *et al.*, 1989; chapitre 13). Après séchage le gel est exposé en contact avec un film X-Omat AR (Kodak) à la température de la pièce.

### 3.5 Analyse des séquences

Afin de déterminer si les oligonucléotides sélectionnés renfermaient un consensus, on a analysé les séquences des deux brins arbitrairement nommés a et b pour chaque oligonucléotide. La région centrale des oligonucléotides sélectionnés correspondant à la région N16 de l'oligo 842 (voir section 3.1) ainsi que les régions flanquantes, lorsque nécessaire ont été analysées par inspection visuelle ainsi qu'à l'aide du programme informatique Gibbs (Lawrence *et al.*, 1993).

#### 4. ÉTUDES DE LIAISON À L'ADN PAR RETARDEMENT SUR GEL

Les études de liaison à l'ADN par retardement sur gel ont été faites pour trois protéines, soit MBP-ZNF74-R<sub>1</sub>, MBP-Kid<sub>1</sub> et MBP à titre de contrôle. Le protocole de retardement sur gel décrit ci-dessous correspond à une adaptation des conditions décrites par Elser *et al.* (1997) et Pierrou *et al.* (1995).

##### 4.1 Oligonucléotides utilisés

###### 4.1.1 Oligonucléotides simple brin

Les oligonucléotides simple brin utilisés lors des études de liaison à l'ADN par retardement sur gel sont les suivants. Oligo 844: 5'-GACAGTTGTTGTCTT-CCTACTAGCTGCAG-3', oligo 845: 5'-ATGAAGTACTACGCCCACTCGAG-3', oligo SS1: 5'-CCTCACTAAAGGGAGTCGACTACCTTTGAAATCGAAA-GGCGCCTCGGATCCAGGGAATCTTAGGA-3', oligo SS2: 5'-TCCTAAG-ATTCCCTGGATCCTCTTATAAGGGATCGGGACCCTTGCGTCGACTCCCT

TTAGTGAGG-3', oligo SS5: 5'-CCTCACTAAAGGGAGTCGACTACCTTTGGATCCAGGGAATCTTAGGA-3', oligo SS6: 5'-TCCTAAGATTCCCTGGATCCCCCTTGCCTCGACTCCCTTTAG-TGAGG-3', oligo Jul 1: 5'-GTCTAGTACCCGGGGAATGGACAGGGAAGATCTCTGGCATTCA-3', oligo Jul 2: 5'-CGTTAACCCGGGTCATGTATGAGTCCTCTGATGTCTTGTAAG-3', oligo RNPU1: 5'-CGAGTTTGAGGCAGCGCTAGC-3', oligo RNPU6: 5'-CCCTCAGATTACGCTTCTTTCGGGCAGCA-3'.

#### 4.1.2 Oligonucléotides double brin

Les oligonucléotides double brin utilisés lors des études de liaison à l'ADN par retardement sur gel sont les suivants. Oligo SS2/SS2c, SS2: 5'-TCCTAAGATTCCCTGGATCCTCTTATAAGGGATCGGGACCCTTGCCTCGACTCCCTTTAGTGAGG -3', SS2c: 5'-CCTCACTAAAGGGAGTCGACGCAAGGTCCCGATCCCTTATAAGAGGATCCAGGGAATCTTAGGA-3'; oligo SS5/SS5c, SS5: 5'-CCTCACTAAAGGGAGTCGACTACCTTTGGATCCAGGGAATCTTAGGA-3', SS5c: 5'-TCCTAAGATTCCCTGGATCCAAAGGTAGTCGACTCCCTTTAGTGAGG-3'; oligo SS6/SS6c, SS6: 5'-TCCTAAGATTCCCTGGATCCCCCTTGCCTCGACTCCCTTTAGTGAGG-3', SS6c: 5'-CCTCACTAAAGGGAGTCGACGCAAGGGGGATCCAGGGAATCTTAGGA-3'.

#### 4.1.3 Oligonucléotides hétéroduplex

Les oligonucléotides hétéroduplex utilisés lors des études de liaison à l'ADN par retardement sur gel sont les suivants. Oligo SS5/SS6, SS5: 5'-

CCTCACTAAAGGGAGTCGACT***ACCTTT***GGATCCAGGGAATCTTAGGA-3', SS6: 5'-TCCTAAGATTCCCTGGATCC***CCCTT***GCGTCGACTCCCTTTA-GTGAGG-3' (bulle de 7 nts non-complémentaires, en gras et en italique); oligo SS7/SS8, SS7: 5'-CCTCACTAAAGGGAGTCGACT***ACCTTT***GAAATCGAG-GATCCAGGGAATCTTAGGA-3', SS8: 5'-TCCTAAGATTCCCTGGATCC-***GATCGGGACCCTT***GCGTCGACTCCCTTTAGTGAGG-3' (bulle de 15 nts non-complémentaires, en gras et en italique); oligo SS1/SS2, SS1: 5'-CCTCACTAAAGGGAGTCGACT***ACCTTT***GAAATCGAAAGGCGCCTCGGA-TCCAGGGAATCTTAGGA-3', oligo SS2: 5'-TCCTAAGATTCCCTGGATCCT-***CTTATAAGGGATCGGGACCCTT***GCGTCGACTCCCTTTAGTGAGG-3' (bulle de 25 nts non-complémentaires, en gras et en italique).

#### 4.2 Préparation des oligonucléotides double brin et hétéroduplex

L'appariement des oligonucléotides simple brin pour former des oligonucléotides double brin ou hétéroduplex, selon le cas, est fait dans un volume de 50  $\mu$ L de tampon TE (Tris 100 mM, EDTA 10 mM à pH 8,0) contenant du NaCl 100 mM. Une quantité équimolaire de chacun des deux oligonucléotides (125 nmoles) est ajoutée au tampon pour obtenir une concentration finale d'oligonucléotides double brin (ou hétéroduplex) de 5 pmoles/ $\mu$ L. Le mélange est chauffé 5 minutes à 95°C dans un bloc chauffant et refroidi lentement jusqu'à 25°C.

#### 4.3 Marquage des oligonucléotides simple brin, double brin et hétéroduplex

Les oligonucléotides sont marqués radioactivement avec du [ $\gamma$ - $^{32}\text{P}$ ] dATP (activité spécifique 3000 Ci/mme, DuPont NEN) (Sambrook *et al.*, 1989; chapitre 5). Cinq unités de la kinase de polynucléotides T<sub>4</sub> (T4 PNK) (New England Biolabs) et 50  $\mu\text{Ci}$  de [ $\gamma$ - $^{32}\text{P}$ ] dATP sont utilisés pour marquer 5 pmoles d'oligonucléotides dans 10  $\mu\text{L}$  de tampon tel que fourni par le manufacturier (Tris-HCl 70 mM à pH 7,6,  $\text{MgCl}_2$  10 mM, DTT 5 mM) durant une réaction de 60 minutes à 37°C. L'activité spécifique des oligonucléotides marqués est ensuite déterminée tel que décrit ci-dessous.

#### **4.4 Détermination de l'activité spécifique et purification des oligonucléotides**

##### **4.4.1 Détermination de l'activité spécifique**

Cette étape sert à déterminer le taux de radioactivité incorporée aux oligonucléotides. L'équivalent de 125 fmoles d'oligonucléotides marqués sur un total de 5 pmoles est déposé sur un filtre DE81 (Whatman). Sans laisser sécher le filtre, on le lave 2 fois durant 2 minutes dans du tampon phosphate ( $\text{Na}_2\text{HPO}_4$  0,5 M,  $\text{NaH}_2\text{PO}_4$  0,5 M), 1 minute dans l'eau et 1 minute dans l'éthanol 95%. On laisse sécher le filtre et on compte ensuite la radioactivité fixée (correspondant à la radioactivité incorporée) à l'aide d'un compteur  $\beta$  en l'absence de solution de scintillant. L'activité spécifique est exprimée en cpm/fmole.

##### **4.4.2 Isolation des oligonucléotides marqués**

Pour éliminer la radioactivité non incorporée aux oligonucléotides et purifier ces derniers, on dépose directement sur gel de polyacrylamide 12% (acrylamide 29,2%:bisacrylamide 0,8%) non dénaturant (Sambrook *et al.*, 1989; chapitre 6) la totalité de l'ADN marqué (environ 5 pmoles, voir section précédente). Une piste est réservée pour du bleu de bromophénol à titre de marqueur. L'électrophorèse, est effectuée à 100V (10 V/cm) et est arrêtée lorsque le marqueur a migré jusqu'à la moitié du gel (durée de migration de 40-45 minutes). Le gel est ensuite exposé en contact avec un film X-Omat Blue XB-1 (Kodak) pour un temps très court (environ 15 à 20 sec) afin de positionner la bande d'intérêt. Cette bande est découpée, et l'ADN qu'elle contient est élué grâce à une incubation d'une nuit à 37°C effectuée avec agitation dans 300 µL de tampon TE contenant du NaCl 400 mM. Le tout est ensuite passé sur une mini-colonne contenant un filtre de 45 µm (Costar, numéro de catalogue 8160) et l'éluat contenant les oligonucléotides purifiés est recueilli par centrifugation dans un tube eppendorf de 1,5 mL. L'éluat d'oligonucléotides est concentré par précipitation à l'aide de 2,5 volumes d'éthanol 98% et de 0,1 volume de NaOAc 3M à pH 5,2 (30 minutes sur glace, 30 minutes de centrifugation). Le culot contenant les oligonucléotides précipités est lavé avec de l'éthanol 80% avant d'être séché et resuspendu dans 50 µL de tampon TE contenant du NaCl 400 mM.

#### **4.4.3 Détermination de la concentration des oligonucléotides**

Pour déterminer la quantité d'oligonucléotides récupérée après purification sur gel d'acrylamide et précipitation (section 4.3.2) (rendement de purification), et en déduire sa concentration, on estime par comptage dans un compteur  $\beta$  la radioactivité récupérée (comptage d'un échantillon de  $2\mu\text{L}/50\mu\text{L}$ ). Déduisant ainsi le nombre de cpm/ $\mu\text{L}$  récupérés dans la préparation d'oligonucléotides purifiés et connaissant l'activité spécifique en cpm/fmole déterminée avant purification (section 4.3.1), on peut déduire la concentration de l'ADN récupéré exprimée en fmole/ $\mu\text{L}$ .

#### **4.5 Liaison des protéines aux oligonucléotides**

La réaction de liaison des oligonucléotides aux protéines s'effectue pendant 20 minutes à température de la pièce en présence de 10 fmoles d'oligonucléotides (en général 40 000-60 000 cpm) et 2 pmoles de protéines purifiées (voir section 1.3) dans 20  $\mu\text{L}$  de tampon de liaison (Hepes 12 mM à pH 7,5, Tris 4 mM à pH 8,0, KCl 60 mM,  $\text{MgCl}_2$  5 mM, EDTA 0,6 mM, DTT 5 mM, BSA 0,5  $\mu\text{g}/\mu\text{L}$ , glycérol 12%). Des réactions effectuées en présence de protéines MBP ou en l'absence de protéines servent de contrôles négatifs. Pour les expériences de compétition, un excès molaire d'oligonucléotides non-marqués de 1-, 5-, 10-, ou 100- fois est ajouté aux protéines au même moment que l'ADN marqué. Pour les expériences où l'on souhaite déstabiliser la structure en doigts de zinc en chélatant le zinc, une combinaison d'EDTA 10 mM à pH 8,0, et de 1,10-phénantroline 0,1 mM (concentrations finales) est ajoutée au tampon de liaison au même moment que les protéines et l'ADN. La 1,10-

phénantroline est ajoutée à partir d'une solution stock de 20 mM préparée dans l'éthanol 98%; afin de s'assurer que l'éthanol n'interfère pas avec la réaction, on effectue un contrôle dans lequel un volume égal d'éthanol (concentration finale d'éthanol de 0,5%) est ajouté au tampon de liaison.

#### **4.6 Séparation des complexes protéine/ADN de l'ADN non lié par retardement sur gel**

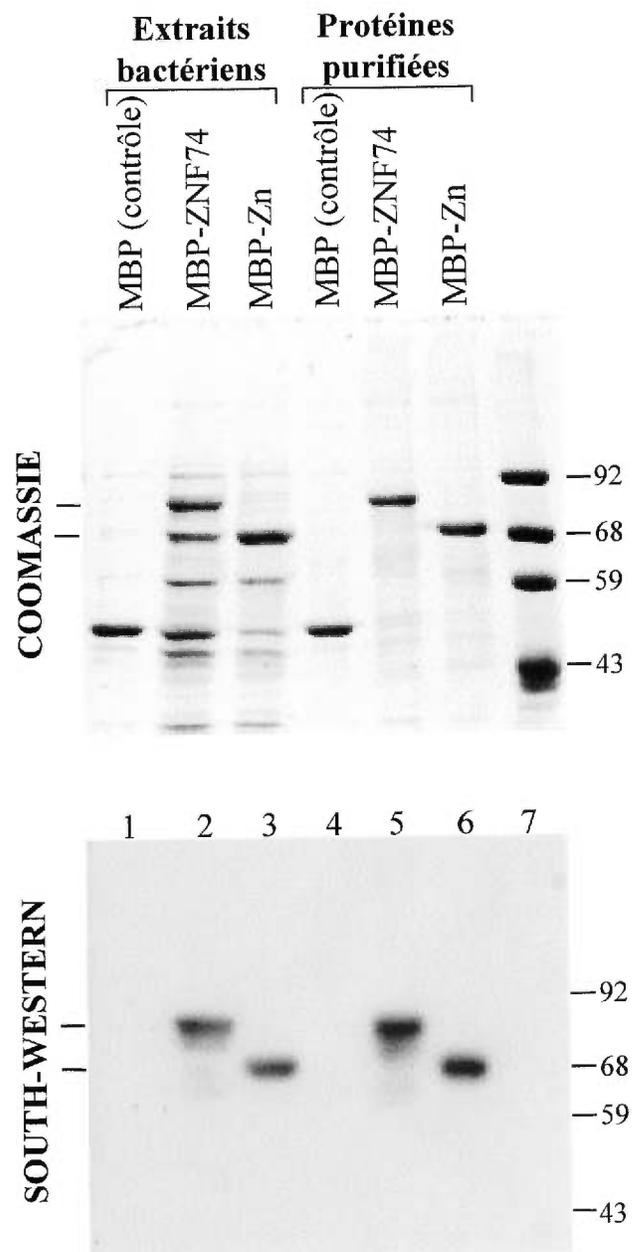
Afin de séparer les complexes protéine/ADN de l'ADN non complexé, la totalité des milieux de réaction de liaison (20  $\mu$ L) est déposée directement sur un gel de polyacrylamide 4% (acrylamide 30%:bisacrylamide 1%) contenant du tampon TBE 0,25X (TBE 1X: Tris-borate 90 mM, EDTA 2 mM) et ayant subi une pré-migration de 30 minutes à 35V. Le tampon de migration est aussi composé de TBE 0,25X. L'électrophorèse est effectuée à 200V (10V/cm) pendant 90 minutes à température de la pièce. Le gel est ensuite séché et exposé en contact avec un film X-Omat Blue XB-1 (Kodak) à -80°C.

### **III-RÉSULTATS**

## 1. LIAISON DE ZNF74 À L'ADN: ÉTUDE PAR BUVARDAGE DE TYPE SOUTH-WESTERN

Par des études de liaison de polyribosides radioactifs, il avait déjà été démontré par Grondin *et al.* (1996) que la protéine à motifs doigt de zinc ZNF74 pouvait lier l'ARN de façon spécifique *via* son domaine multidoigt de zinc (Zn). Afin de déterminer si cette même protéine avait aussi une affinité pour l'ADN comme déjà démontré pour plusieurs protéines à motifs doigt de zinc (Theunissen *et al.*, 1992), une expérience de buvardage de type South-Western a été réalisée. Pour ce faire, la protéine ZNF74 fusionnée avec la protéine MBP a été transférée sur membrane après séparation par électrophorèse sur gel de polyacrylamide dénaturant. L'interaction protéine de fusion/ADN a ensuite été testée en incubant la membrane avec une sonde correspondant à de l'ADN génomique humain fragmenté par sonication (300 à 1000 pb) et marqué radioactivement avec du [ $\alpha$ - $^{32}$ P]. Comme on peut le voir à la Figure 1 (bas), la protéine de fusion MBP-ZNF74 (pistes 2 et 5) lie l'ADN radioactif. Le domaine doigt de zinc de la protéine ZNF74 (MBP-Zn) (pistes 3 et 6) semble suffisant pour obtenir une liaison à l'ADN. Par contre, lorsqu'utilisée en quantité comparable aux protéines MBP-ZNF74 et MBP-Zn (voir coloration au bleu de Coomassie, Figure 1 (haut)), la protéine MBP seule (pistes 1 et 4) ainsi que les différents standards de protéines (piste 7) servant de contrôles négatifs ne lient pas l'ADN radioactif. L'intégrité du domaine doigt de zinc semble requise pour la liaison de l'ADN puisque qu'aucun signal de liaison n'est détecté lorsque que la membrane est pré-incubée avant l'étape d'addition de l'ADN marqué avec un chélateur du zinc (voir Section 2, *Matériel et Méthodes*) (résultats non présentés).

**Figure 1. Buvardage de type South-Western.** Des protéines de fusion à MPB (0,5 à 1  $\mu$ g) provenant d'extraits bactériens (pistes 1 à 3) ou purifiées par résine d'affinité (pistes 4 à 6) sont séparées sur deux gels d'électrophorèse dénaturants (8%) en présence de SDS. Le gel du haut a été coloré au bleu de Coomassie pour permettre de voir que la quantité de protéines est environ la même dans chacune des pistes. Le gel du bas a quant à lui été transféré sur une membrane de nitrocellulose et la membrane a par la suite été hybridée avec une sonde d'ADN génomique humain soniqué et marqué au [ $\alpha$ - $^{32}$ P] dCTP. Ce gel permet de voir que la protéine ZNF74 (aa 1-572) (pistes 2 et 5) et la protéine correspondant à la région doigt de zinc de ZNF74 (aa 175-509) (pistes 3 et 6) sont capables de lier l'ADN; par contre, ni la protéine contrôle MBP (pistes 1 et 4) ni les protéines du standard (piste 7) ne lient l'ADN dans ces conditions. La masse moléculaire des protéines du standard (phosphorylase B, BSA, catalase et albumine d'œuf de poulet) est indiquée à droite des gels en kDa.



Compte tenu du fait que l'ADN génomique utilisé dans cette expérience de buvardage de type South-Western était essentiellement constitué d'ADN double brin, cette expérience suggère que ZNF74 a une affinité pour l'ADN double brin.

## **2. SÉLECTION D'OLIGONUCLÉOTIDES DÉGÉNÉRÉS DOUBLE BRIN AVEC LA PROTÉINE ZNF74 IMMOBILISÉE SUR RÉSINE**

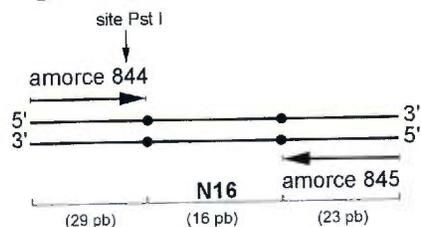
À l'issue de l'expérience de buvardage de type South-Western, on savait que la protéine ZNF74 était capable de lier l'ADN présumément double brin et que son motif doigt de zinc était responsable de cette liaison. On a ensuite cherché à déterminer si ZNF74, comme de nombreux facteurs transcriptionnels, avait une affinité accrue pour certaines séquences spécifiques d'ADN. Pour ce faire, des expériences de sélection d'oligonucléotides dégénérés sur résine d'affinité ont été réalisées.

### **2.1 Sélection d'oligonucléotides double brin par purification/amplification**

La Figure 2 représente schématiquement la procédure cyclique de purification/amplification qui a été employée pour sélectionner à partir d'un mélange d'oligonucléotides dégénérés ceux se liant à ZNF74 ou à des protéines contrôles. Pour préparer les résines d'affinité, on a immobilisé sur résine d'amylose soit la protéine MBP-ZNF74-R<sub>1</sub> provenant d'extraits bactériens et correspondant aux acides aminés 106 à 572 de ZNF74 incluant le domaine doigt zinc, soit une autre protéine à motif multidoigt de zinc, la protéine Kid<sub>1</sub> (Elser *et al.*, 1997) fusionnée à la protéine MBP (MBP-Kid<sub>1</sub>; contrôle positif potentiel), soit MBP à titre de contrôle négatif. L'intégrité et la pureté des protéines retenues sur les résines d'affinité peuvent être évaluées à la

**Figure 2. Représentation schématique de la procédure cyclique de purification/amplification des oligonucléotides double brin.** Des oligonucléotides double brin dégénérés en 16 positions (N16) sont synthétisés (voir *Matériel et Méthodes*, Section 3.1) et sélectionnés par incubation avec des protéines de fusion exprimées par *E. coli* et immobilisées sur une résine d'amylose. Après purification des oligonucléotides par liaison aux protéines de la résine d'affinité, les oligonucléotides sélectionnés sont amplifiés par deux étapes de PCR. Cette procédure de purification/amplification est répétée 7 fois (voir *Matériel et Méthodes*, Section 3.2 pour plus de détails). Enfin, les oligonucléotides sélectionnés sont clonés puis séquencés.

**mélange d'oligonucléotides  
dégénérés (N16) double brin**



**extrait protéique**



**liaison des protéines à  
la résine d'amylose**

**liaison des oligonucléotides  
aux protéines**



**lavages**



**élution des oligonucléotides liés aux protéines**



**première amplification PCR des  
oligonucléotides sélectionnés**



**deuxième amplification PCR des  
oligonucléotides sélectionnés**



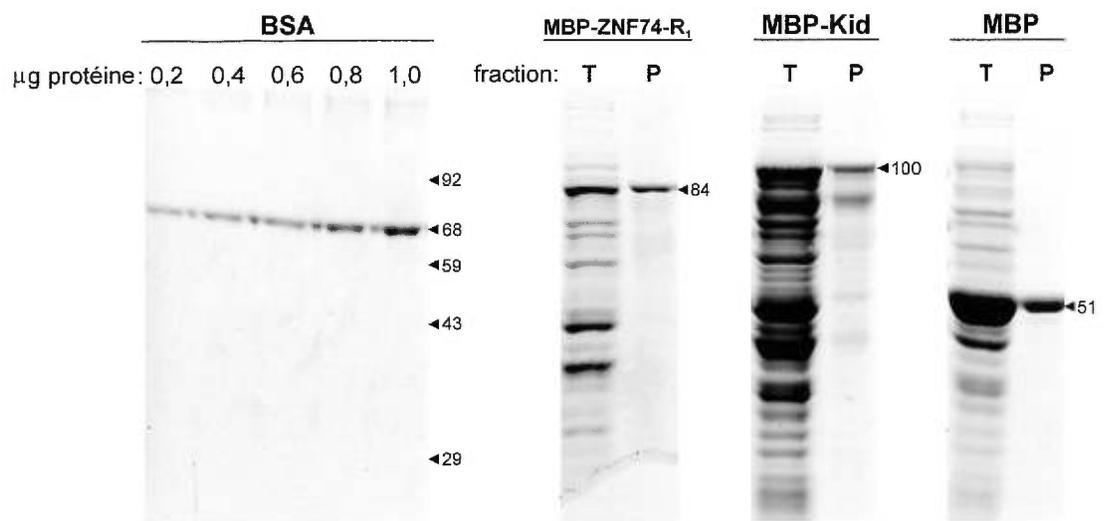
**clonage et séquençage des oligonucléotides sélectionnés**

**7 cycles**

Figure 3. Alors que les protéines MBP-ZNF74-R<sub>1</sub> et MBP sont pures à environ 90%, la protéine MBP-Kid<sub>1</sub> subit une dégradation partielle (environ 50%) tel que souvent observée pour des protéines de fusion de 100 kDa ou plus. Pour les expériences de sélection d'oligonucléotides, on a immobilisé sur résine l'équivalent de 1 µg de chacune des protéines de fusion intacte sur résine.

Des oligonucléotides de 68 pb dégénérés en 16 positions (N16) ont été utilisés pour la sélection par affinité. La complexité de la banque d'oligonucléotides dégénérés est de l'ordre de 4<sup>16</sup>. Afin d'avoir un mélange représentatif de toutes les séquences de la banque dégénérée, on a utilisé lors de la première étape de purification 400 ng d'oligonucléotides dégénérés soit l'équivalent de 325 fois la banque d'oligonucléotides dégénérés. Seulement, 100 à 300 ng d'oligonucléotides sélectionnés et amplifiés ont été utilisés lors des étapes subséquentes de purification (Tableau 1). Le nombre maximum de cycles de PCR permettant d'obtenir une amplification spécifique des oligonucléotides sélectionnés a été déterminé. Dans les conditions de PCR décrites à la section *Matériel et Méthodes* (Section 3.1), 8 cycles de PCR se sont révélés adéquats pour une amplification spécifique d'oligonucléotides sélectionnés. À titre d'exemple, la Figure 4 illustre l'amplification des oligonucléotides sélectionnés par les protéines MBP-ZNF74-R<sub>1</sub> et MBP-Kid<sub>1</sub> à la suite de la première étape de purification. Il apparaît qu'une seule bande de taille attendue (68 pb) est obtenue après 8 cycles de PCR alors qu'un halo de bandes non spécifiques de taille supérieure est visible après 10 cycles de PCR. Cette amplification non spécifique obtenue après un trop grand nombre de cycles de PCR est attribuable à l'épuisement des amorces et à l'appariement aléatoires des produits de PCR.

**Figure 3. Analyse par électrophorèse des protéines de fusion provenant d'extraits bactériens.** Les protéines des extraits bactériens (T) et les protéines de fusion purifiées sur 50  $\mu$ L de résine d'amylose (P) sont séparées sur un gel d'électrophorèse dénaturant (8%) en présence de SDS puis colorées au bleu de Coomassie après leur migration. Les extraits bactériens correspondent à l'équivalent de 1/10 000 (MBP-ZNF74-R<sub>1</sub> et MBP) ou 1/5000 (MBP-Kid<sub>1</sub>) des cultures initiales de 500 mL induites pendant 3 heures avec de l'IPTG. Afin d'évaluer la quantité d'extraits bactériens nécessaire pour la rétention de l'équivalent de 1  $\mu$ g de protéine de fusion, les protéines de fusion purifiées sur 50  $\mu$ L de résine d'amylose sont éluées par ajout de tampon Laemmli et déposées sur gel. La quantité des protéines de fusion pures est estimée par coloration au bleu de Coomassie en comparaison avec un standard de BSA et est confirmée par dosage des protéines par la méthode de Bradford. Les fractions purifiées correspondent à 1/1600 (MBP-ZNF74-R<sub>1</sub> et MBP) ou 1/800 (MBP-Kid<sub>1</sub>) des cultures initiales. Pour chacune des protéines purifiées, la masse moléculaire apparente est indiquée. Les protéines qui ont été utilisées comme standard et dont les masses moléculaires sont indiquées sont: phosphorylase B (92 kDa), BSA (68 kDa), catalase (59 kDa), albumine d'œuf de poulet (43 kDa), anhydrase carbonique (29 kDa), inhibiteur de la trypsine (21 kDa).



**Tableau 1. ADN utilisé pour sélection et récupéré après amplification à chaque étape de purification.** Les quantités d'ADN oligonucléotidique double brin utilisées pour chacune des sept purifications ainsi que les quantités d'ADN recueillies après chacune des deux amplifications par PCR pour deux expériences de sélection d'oligonucléotides indépendantes réalisées avec MBP-ZNF74-R<sub>1</sub> (ZNF74-R<sub>1</sub>), MBP-Kid<sub>1</sub> (Kid<sub>1</sub>) et MBP sont indiquées.

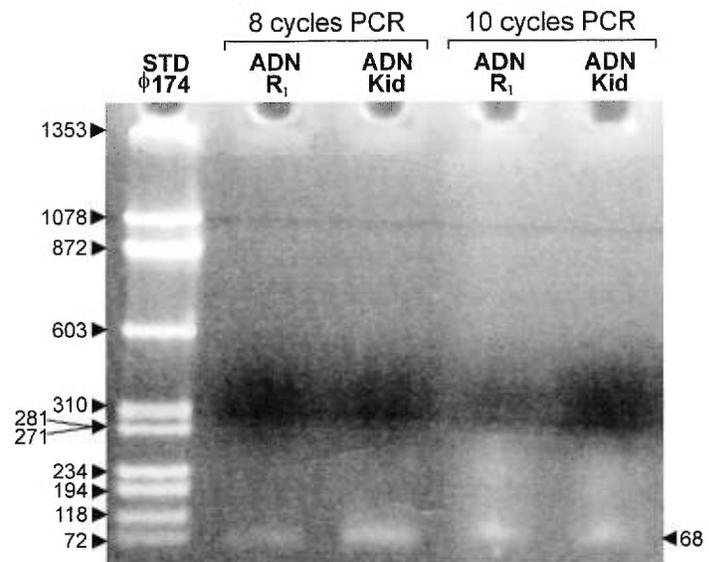
Tableau I. ADN utilisé pour sélection et récupéré après amplification à chaque étape de purification.

	Première expérience			Deuxième expérience		
	ADN incubé avec chacune des protéines	ADN recueilli après le 1 <sup>er</sup> PCR (volume total 250 µL)	ADN recueilli après le 2 <sup>e</sup> PCR (volume total 100 µL)	ADN incubé avec chacune des protéines	ADN recueilli après le 1 <sup>er</sup> PCR (volume total 250 µL)	ADN recueilli après le 2 <sup>e</sup> PCR (volume total 100 µL)
première purification	ZNF74-R1: 400 ng Kid1: 400 ng MBP: 400 ng	ZNF74-R1: 6 ng/µL Kid1: 6 ng/µL MBP: N.D.*	ZNF74-R1: 6 ng/µL Kid1: 11 ng/µL MBP: 7 ng/µL	ZNF74-R1: 400 ng Kid1: 400 ng MBP: 400 ng	ZNF74-R1: N.D. Kid1: N.D. MBP: N.D.	ZNF74-R1: 3 ng/µL Kid1: 6 ng/µL MBP: 1 ng/µL
deuxième purification	ZNF74-R1: 500 ng Kid1: 500 ng MBP: 200 ng	ZNF74-R1: 3 ng/µL Kid1: 3 ng/µL MBP: 1 ng/µL	ZNF74-R1: 2 ng/µL Kid1: 2 ng/µL MBP: 3 ng/µL	ZNF74-R1: 200 ng Kid1: 200 ng MBP: 200 ng	ZNF74-R1: N.D. Kid1: N.D. MBP: N.D.	ZNF74-R1: 1 ng/µL Kid1: 6 ng/µL MBP: 1 ng/µL
troisième purification	ZNF74-R1: 250 ng Kid1: 250 ng MBP: 250 ng	ZNF74-R1: 3 ng/µL Kid1: 3 ng/µL MBP: N.D.	ZNF74-R1: 2 ng/µL Kid1: 2 ng/µL MBP: 6 ng/µL	ZNF74-R1: 200 ng Kid1: 200 ng MBP: 100 ng	ZNF74-R1: N.D. Kid1: N.D. MBP: N.D.	ZNF74-R1: 3 ng/µL Kid1: 6 ng/µL MBP: 1 ng/µL
quatrième purification	ZNF74-R1: 250 ng Kid1: 250 ng MBP: 250 ng	ZNF74-R1: 2 ng/µL Kid1: 2 ng/µL MBP: N.D.	ZNF74-R1: 3 ng/µL Kid1: 3 ng/µL MBP: 3 ng/µL	ZNF74-R1: 300 ng Kid1: 300 ng MBP: 100 ng	ZNF74-R1: N.D. Kid1: 1 ng/µL MBP: N.D.	ZNF74-R1: 6 ng/µL Kid1: 6 ng/µL MBP: 1 ng/µL
cinquième purification	ZNF74-R1: 250 ng Kid1: 250 ng MBP: 250 ng	ZNF74-R1: 7 ng/µL Kid1: 7 ng/µL MBP: N.D.	ZNF74-R1: 3 ng/µL Kid1: 3 ng/µL MBP: 3 ng/µL	ZNF74-R1: 300 ng Kid1: 300 ng MBP: 100 ng	ZNF74-R1: 1 ng/µL Kid1: 7 ng/µL MBP: N.D.	ZNF74-R1: 6 ng/µL Kid1: 3 ng/µL MBP: 1 ng/µL
sixième purification	ZNF74-R1: 250 ng Kid1: 250 ng MBP: 250 ng	ZNF74-R1: 9 ng/µL Kid1: 6 ng/µL MBP: 3 ng/µL	ZNF74-R1: 5 ng/µL Kid1: 5 ng/µL MBP: 5 ng/µL	ZNF74-R1: 300 ng Kid1: 300 ng MBP: 100 ng	ZNF74-R1: N.D. Kid1: 7 ng/µL MBP: N.D.	ZNF74-R1: 10 ng/µL Kid1: ** MBP: N.D.
septième purification	ZNF74-R1: 250 ng Kid1: 250 ng MBP: 250 ng	ZNF74-R1: 6 ng/µL Kid1: 6 ng/µL MBP: N.D.	ZNF74-R1: 3 ng/µL Kid1: 3 ng/µL MBP: 5 ng/µL	ZNF74-R1: 300 ng Kid1: 300 ng MBP: 100 ng	ZNF74-R1: 1 ng/µL Kid1: N.D. MBP: N.D.	ZNF74-R1: 10 ng/µL Kid1: 3 ng/µL MBP: N.D.

\* : non déterminable par coloration au bromure d'éthidium.

\*\* : quantité d'ADN suffisante après une seule amplification pour entreprendre le cycle suivant

**Figure 4. Détermination du nombre maximum de cycles PCR pour amplification spécifique des oligonucléotides double brin.** L'ADN oligonucléotidique double brin retenu par les protéines MBP-ZNF74-R<sub>1</sub> (R<sub>1</sub>) et MBP-Kid<sub>1</sub> (Kid<sub>1</sub>) a été amplifié par 8 ou 10 cycles de PCR. Les produits de PCR sont déposés sur gel d'acrylamide non dénaturant (12%) puis colorés au bromure d'éthidium. Les standards de poids moléculaires correspondent au phage PhiX174 RF digéré avec *Hae* III.



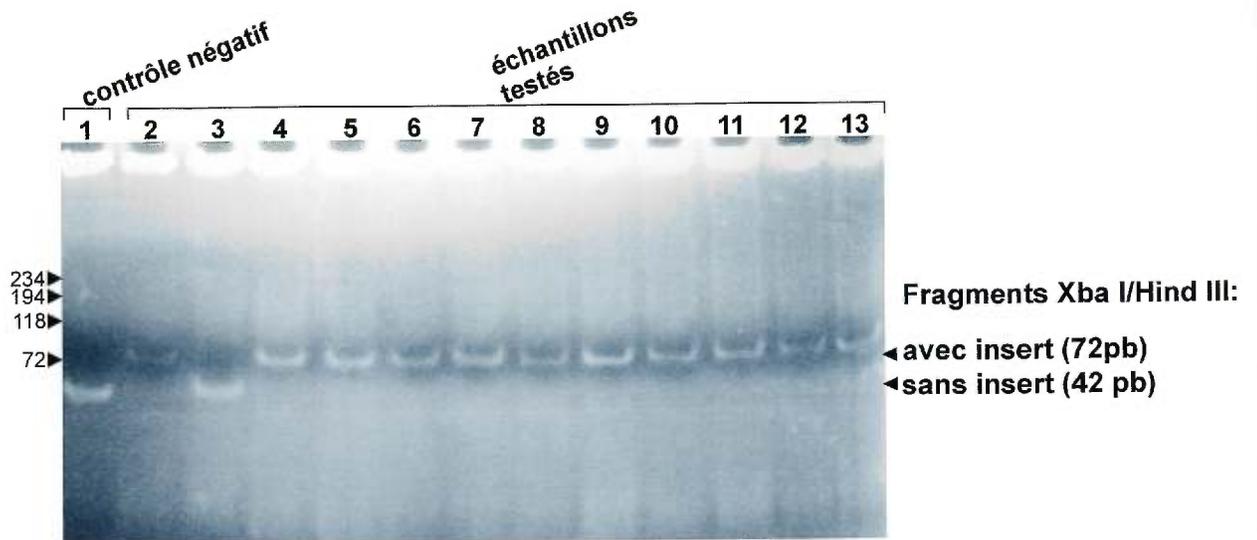
De façon générale, moins de 15 ng d'oligonucléotides de départ doivent être utilisés pour obtenir une amplification adéquate et spécifique après 8 cycles de PCR.

Deux expériences indépendantes de sélection par purification/amplification ont été réalisées avec les protéines MBP-ZNF74-R<sub>1</sub>, MBP-Kid<sub>1</sub> et MBP. Le Tableau 1 présente les quantités d'ADN utilisées pour chaque étape de purification et récupérées après chacune des deux amplifications lors de ces deux expériences de purification/amplification. De façon générale, la comparaison des quantités d'ADN obtenues après amplification (essai semi-quantitatif) révèle que la quantité d'ADN retenue par les protéines MBP-ZNF74-R<sub>1</sub>, MBP-Kid<sub>1</sub> et servant de matrice à l'amplification est supérieure à celle retenue par la protéine MBP qui n'a pas d'affinité particulière pour l'ADN.

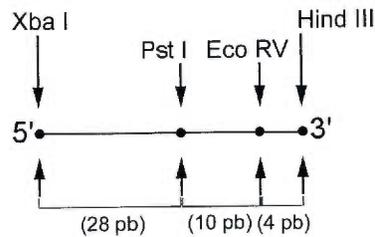
Pour chacune des expériences, les oligonucléotides sélectionnés après sept cycles de purification/amplification sur résine d'affinité ont été digérés avec l'enzyme *Pst* I et clonés de façon orientée dans le vecteur pBluescript SK digérés avec *Pst* I et *Eco* RV (Figure 5); *Eco* RV laisse des extrémités franches qui peuvent s'adapter à l'extrémité du produit PCR non clivé par *Pst* I. Afin de s'assurer de la présence d'inserts dans les plasmides avant l'étape de séquençage, les préparations d'ADN plasmidiques obtenues après transformation de *E. coli* ont été digérées avec *Xba* I/*Hind* III (deux sites de coupure présents dans le vecteur). Tel qu'illustré à la Figure 5, la libération d'un fragment de 72 paires de bases (pb) confirme que le clone contient un insert alors que celle d'un fragment de 42 pb révèle un vecteur sans insert.

## **2.2 Analyse des séquences d'oligonucléotides double brin sélectionnés**

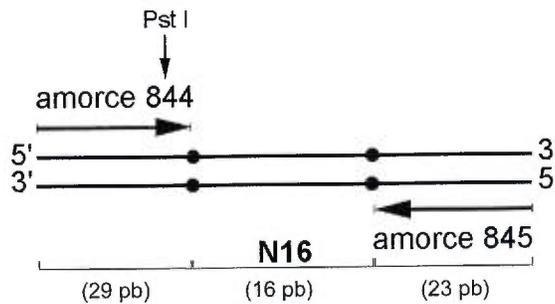
**Figure 5. Identification sur gel d'acrylamide des clones plasmidiques contenant un oligonucléotide sélectionné comme insert.** (A) Un échantillon de chaque préparation d'ADN plasmidique digéré *Xba I/Hind III* est mis sur gel de polyacrylamide non dénaturant (12%) pour déterminer la taille du fragment libéré (pistes 2 à 13). Le vecteur SK coupé *Xba I/Hind III* est utilisé à titre de contrôle négatif (piste 1). Un fragment de 72 pb révèle la présence d'un insert (tel que décrit en B) libéré par digestion (pistes 2 et 4 à 13) alors qu'un fragment de 42 pb indique l'absence d'insert dans le vecteur digéré (résultant d'une digestion incomplète par *Pst I* et *Eco RV* lors du clonage) (piste 3). Les standards de poids moléculaires correspondent au phage PhiX174 RF digéré *Hae III*. (B) Représentation schématique du clonage des oligonucléotides sélectionnés dans le vecteur SK.

**A****B**

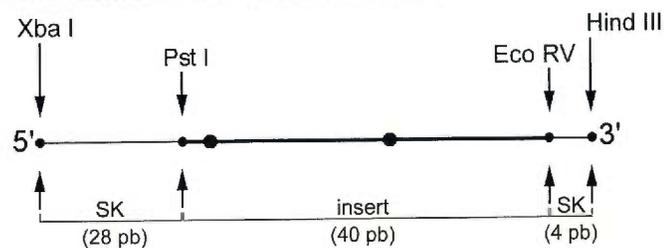
### Région du vecteur SK pour clonage orienté dans le site Pst I/Eco Rv



### Oligonucléotides sélectionnés à digérer avec Pst I pour clonage



### Construction contenant les oligonucléotides sélectionnés clonés dans le site Pst I/Eco RV



Pour chacune des protéines MBP-ZNF74-R<sub>1</sub>, MBP-Kid<sub>1</sub> et MBP, plus de 32 clones contenant des oligonucléotides provenant des deux expériences indépendantes de purification par affinité ont été séquencés (Figures 6 et 7, et Annexe I). Les séquences obtenues de même que leur complémentaire inverse ont été analysées premièrement par inspection visuelle pour la recherche d'un consensus possible. Il est à noter, en effet, que la recherche du consensus doit être effectuée sur les deux brins de la séquence sauf dans le cas où la séquence consensus est de type palindromique. D'autre part, les séquences sélectionnées ont aussi été plus systématiquement analysées à l'aide du programme informatique Gibbs qui utilise un algorithme pour alignement optimal local de séquences multiples (Laurence *et al.*, 1993). Ce programme a été notamment utilisé avec succès pour identifier des motifs exoniques reconnus par des protéines SR impliquées dans l'épissage (Liu *et al.*, 1998). Le programme Gibbs permet de rechercher la présence d'un consensus subtil dans un ensemble de séquences en calculant la fréquence de la présence d'un nucléotide à une position donnée et d'aligner les séquences après avoir déterminé des matrices de probabilités optimales. Le programme ne requiert pas d'avoir des informations initiales ni sur le type de séquence consensus ni sur la position du consensus dans les séquences mais par contre requiert de prédéterminer la taille du consensus comme paramètre de départ. Le programme Gibbs a été utilisé pour analyser les séquences sélectionnées par MBP-ZNF74-R<sub>1</sub>, MBP-Kid<sub>1</sub> et MBP en faisant varier successivement la taille du consensus de 10 à 3 bases comme paramètre de calcul. L'Annexe II présente un exemple d'un fichier d'entrée contenant les séquences à analyser (Annexe II: Figure 1) et d'un fichier de sortie du programme Gibbs (Annexe II: Figure 2) présentant les séquences alignées

**Figure 6. Sélection d'oligonucléotides double brin par MBP-ZNF74-R<sub>1</sub> immobilisée sur résine (première expérience).** (A) Séquences sélectionnées par la protéine ZNF74-R<sub>1</sub> lors de la première expérience. Le brin b correspond au complémentaire inverse du brin a. La région soulignée sur chacune des séquences correspond aux nucléotides inclus dans un consensus à six sites tel qu'optimisé par le programme Gibbs. Un alignement de ces séquences au niveau de la région consensus est aussi présenté dans le fichier de sortie de l'annexe II. (B) Matrices de rapport de probabilités allant de 10 à 3 sites telles que déterminées par le programme informatique Gibbs. Les valeurs numériques indiquant une probabilité égale ou supérieure à 50% pour un nucléotide (A, C, G ou T) de se retrouver à une position donnée sont mentionnées en caractère gras et le nucléotide correspondant est intégré dans le consensus compilé dans la cinquième colonne pour chacune des matrices.

### A. Séquences sélectionnées

#1	a: 5'- TTATCCAGATGGGIGG -3' b: 5'- CCACCCCAICTGGATAA -3'
#2	a: 5'- AGCTTATACAITTATA -3' b: 5'- TATAAATGATAAGCT -3'
#3	a: 5'- GTATATACACAICTCA -3' b: 5'- TGAGAIGIGIATATAC -3'
#4	a: 5'- GTAGGATGGGATTTAAA -3' b: 5'- TTTAATCCCATCCTAC -3'
#5	a: 5'- ACGGCITACATCTTTC -3' b: 5'- GAAAGATGTAAGCCGI -3'
#6	a: 5'- TGGGTCIGCACGATTA -3' b: 5'- TAATCGTGCAGACCCA -3'
#7	a: 5'- GGTAGGCATGTCTCG -3' b: 5'- CGAGACAATGCCTACC -3'
#8	a: 5'- TGGACGCACGGGTAGA -3' b: 5'- TCTACGGGIGCGTCCA -3'
#9	a: 5'- CGATTCAAAGGGGATTA -3' b: 5'- TAAICCCITTTGAATCG -3'
#10	a: 5'- CAGTAACGCITTTTCA -3' b: 5'- TGA AAAAGCGTACTG -3'
#11	a: 5'- AACGAAATGGATGGTA -3' b: 5'- TACCAICCAAATTCGTT -3'
#12	a: 5'- GCICGAGAATCAGCAA -3' b: 5'- TTGCTGATTCIGCAGC -3'

### B. Recherche de sites consensus par le programme Gibbs

position	10 sites			9 sites			8 sites			7 sites										
	A	C	T	A	C	T	A	C	T	A	C	T								
1	15	23	23	49	T	86	6	6	11	T	49	31	2	28	A	57	2	15	36	A
2	65	15	27	3	A	36	23	19	32	N	65	6	19	20	A	74	19	2	15	A
3	61	15	11	24	A	45	36	23	7	A	11	36	19	45	T	11	23	23	53	T
4	70	11	2	28	A	3	48	36	24	C	20	23	65	3	G	11	19	69	11	G
5	3	31	27	49	T	36	27	44	3	N	20	65	11	15	C	7	69	23	11	C
6	15	31	56	7	G	45	2	15	49	T/A	70	11	19	11	A	53	11	36	11	A
7	11	36	27	36	N	20	6	19	65	T	3	2	15	90	T	3	15	6	86	T
8	36	6	40	28	N	3	19	11	78	T	36	31	2	40	N					
9	3	19	2	86	T	53	27	23	7	A										
10	57	11	15	28	A															

position	6 sites			5 sites			4 sites			3 sites										
	A	C	T	A	C	T	A	C	T	A	C	T								
1	90	15	2	3	A	28	40	40	3	N	3	15	90	3	G	3	27	69	11	G
2	11	27	19	53	T	95	2	11	3	A	15	52	27	15	C	45	56	2	7	C
3	15	31	61	3	G	3	23	2	82	T	57	6	27	20	A	36	2	2	70	T
4	7	65	27	11	C	3	40	48	20	G	3	2	6	99	T					
5	53	27	27	3	A	11	44	19	36	N										
6	15	15	11	70	T															

**Figure 7. Sélection d'oligonucléotides double brin par MBP-ZNF74-R<sub>1</sub> immobilisée sur résine (deuxième expérience).** (A) Séquences sélectionnées par la protéine ZNF74-R<sub>1</sub> lors de la deuxième expérience. Le brin b correspond au complémentaire inverse du brin a. Les séquences présentées sur fond gris sont des séquences GC riches. (B) Matrices de rapport de probabilités allant de 10 à 3 sites telles que déterminées par le programme informatique Gibbs. Les valeurs numériques indiquant une probabilité égale ou supérieure à 50% pour un nucléotide (A, C, G ou T) de se retrouver à une position donnée sont mentionnées en caractère gras et le nucléotide correspondant est intégré dans le consensus compilé dans la cinquième colonne pour chacune des matrices.

### A. Séquences sélectionnées

#1	a: 5'- TGAGCCATAATAGAA -3' b: 5'- TTCTATTATAGGCTCA -3'	#12	a: 5'- CTAAGTTCTCGTTTAT -3' b: 5'- ATAAACGAGAACTTAG -3'
#2	a: 5'- TGAGCCATAATAGAA -3' b: 5'- TTCTATTATAGGCTCA -3'	#13	a: 5'- CGATAATTTCTAAAATT -3' b: 5'- AATTTTAGAATTATCG -3'
#3	a: 5'- CGATTGTGTATCTT -3' b: 5'- AAGATACACACAATCG -3'	#14	a: 5'- CTTTTGATTTCAAAAAA -3' b: 5'- TTTTTGAAATCAAAG -3'
#4	a: 5'- ATCGGGAGTCATTCAT -3' b: 5'- ATGAATGACTCCCGAT -3'	#15	a: 5'- CTTTTTACCTGGCACA -3' b: 5'- TGTGCCAGGTGAAAAAG -3'
#5	a: 5'- CTTGGTGTCCCACTT -3' b: 5'- AAGTGGGAACACCAAG -3'	#16	a: 5'- TGAGCCTATAATAGAA -3' b: 5'- TTCTATTATAGGCTCA -3'
#6	a: 5'- CTTTTTCACTGGCACA -3' b: 5'- TGTGCCAGTGAAAAAG -3'	#17	a: 5'- GGTGGTGGGTGGATC -3' b: 5'- GATCCACCCACCCACC -3'
#7	a: 5'- CTATACATCGGTACT -3' b: 5'- AGTACCGAATGTATAG -3'	#18	a: 5'- GGTGGTGGGTGGATC -3' b: 5'- GATCCACCCACCCACC -3'
#8	a: 5'- CGATAATTTCTAAAATT -3' b: 5'- AATTTTAGAATTATCG -3'	#19	a: 5'- GGTGGGGGGTGGATC -3' b: 5'- GATCCACCCACCCACC -3'
#9	a: 5'- CTTTTTCACTGGCACA -3' b: 5'- TGTGCCAGTGAAAAAG -3'	#20	a: 5'- GGTGGGGGGTGGATC -3' b: 5'- GATCCACCCACCCACC -3'
#10	a: 5'- CTCTATTTACCCTTAA -3' b: 5'- TTAAGGGTAAATAGAG -3'	#21	a: 5'- GGTGGGGGGGGGTCT -3' b: 5'- AGACCCCGGGCCACC -3'
#11	a: 5'- TTTAAAGAAATATCTA -3' b: 5'- TAGATATTTCTTTAAA -3'	#22	a: 5'- GGAGGGTGGTGGATC -3' b: 5'- GATCCACCCACCCCTCC -3'

### B. Recherche de sites consensus par le programme Gibbs

position	10 sites			9 sites			8 sites			7 sites										
	A	C	T	A	C	T	A	C	T	A	C	T								
1	4	36	11	57	T	70	4	22	11	A	52	2	47	7	A	27	4	49	27	C
2	25	29	36	18	N	16	8	20	63	T	38	6	38	25	N	50	2	13	43	A
3	66	6	31	4	A	29	18	31	29	N	27	11	6	63	T	2	6	49	50	T
4	23	24	38	23	N	18	33	40	16	N	11	36	40	20	N	23	27	22	36	N
5	25	24	11	48	T	32	8	40	27	N	23	29	29	27	N	50	52	4	2	A
6	9	33	49	16	G	34	43	2	29	N	66	27	4	11	A	50	6	2	50	A/T
7	29	15	31	32	N	34	18	36	20	N	27	22	4	54	G	32	61	2	13	C
8	54	40	6	7	A	43	43	15	7	N	32	72	2	2	C					
9	34	20	2	52	T	86	11	8	2	A										
10	34	68	2	4	C															

position	6 sites			5 sites			4 sites			3 sites										
	A	C	T	A	C	T	A	C	T	A	C	T								
1	29	2	4	73	T	41	2	33	32	N	11	2	63	32	G	88	11	6	2	A
2	11	27	45	25	G	2	11	2	93	T	59	11	4	34	A	16	18	2	73	T
3	29	20	18	41	N	29	29	31	18	N	11	6	15	75	T	48	52	2	7	C/A
4	66	6	18	18	A	20	31	38	18	N	11	93	2	2	C					
5	2	47	2	57	T/C	95	2	8	2	A										
6	4	77	2	25	C															

pour un consensus de taille prédéterminée de 6 bases (6 sites) ainsi que la matrice correspondante incluant les rapports de probabilité pour chaque nucléotide (A, C, G, T) à chaque position du consensus optimal prédit. Pour chacune des protéines, on a analysé les séquences sélectionnées lors de la première puis de la deuxième expérience séparément puis de façon groupée. Une analyse initiale des alignements et des matrices pour des consensus de 10 à 3 bases générés avec le programme Gibbs pour les séquences oligonucléotidiques sélectionnées lors de la première expérience par la protéine MBP-ZNF74-R<sub>1</sub> suggérait la présence d'un consensus relativement dégénéré (Figure 6 et Annexe II). Ce consensus avait pour noyau la séquence hexanucléotidique ATGCAT. Cette séquence se retrouve quoique souvent dégénérée à une ou deux positions sur un des brins de la plupart des 12 séquences double brin étudiées et représentées à la Figure 6A où la région consensus de six sites a été soulignée (selon un autre format, l'alignement du consensus de ces séquences est aussi représenté à la Figure 2 de l'Annexe II). Cependant, le consensus ATGCAT de la première expérience ne semble pas spécifique puisqu'il n'a pu être retrouvé dans les 22 séquences sélectionnées lors de la seconde expérience (Figure 7). Parmi ces 22 séquences, cependant, une inspection visuelle a permis de mettre en évidence la présence de 6 séquences très similaires dégénérées seulement en positions 3, 7, 11 et 14 de l'hexapentanucleotide (séquences #17 à 22, Figure 7). Par contre, ces séquences ne semblent pas avoir une affinité très forte pour MBP-ZNF74-R<sub>1</sub> puisqu'elles ne peuvent efficacement être retardées sur gel par MBP-ZNF74-R<sub>1</sub> dans des expériences de retardement sur gel (données non présentées) similaires à celles rapportées ci-dessous (Section 3.2). Il est possible que ces séquences riches en GC soient dérivées

d'une même séquence fondatrice ayant accumulé, durant les étapes d'amplification par PCR, des mutations et cela de façon atypique compte tenu de la fidélité connue de la polymérase utilisée et de la taille réduite des séquences à amplifier. En excluant les six séquences riches en GC, l'analyse par le programme Gibbs des autres séquences de la deuxième expérience seules ou combinées à celles de la première expérience n'ont pas non plus permis d'identifier un consensus robuste spécifique pour la protéine MBP-ZNF74-R<sub>1</sub> (données non présentées).

Quoique l'affinité globale de MBP-ZNF74-R<sub>1</sub> et de MBP-Kid<sub>1</sub> pour l'ADN double brin apparaisse supérieure à celle de la protéine contrôle MBP (Tableau 1), ni l'inspection visuelle, ni le programme Gibbs, ni la combinaison de ces deux méthodes dans un but d'optimisation n'a permis de détecter une séquence consensus spécifique de haute affinité reconnue individuellement par ces protéines dans les conditions de purification par affinité utilisées ici.

### **3. ÉTUDE DE LIAISON DE ZNF74 À L'ADN PAR RETARDEMENT SUR GEL**

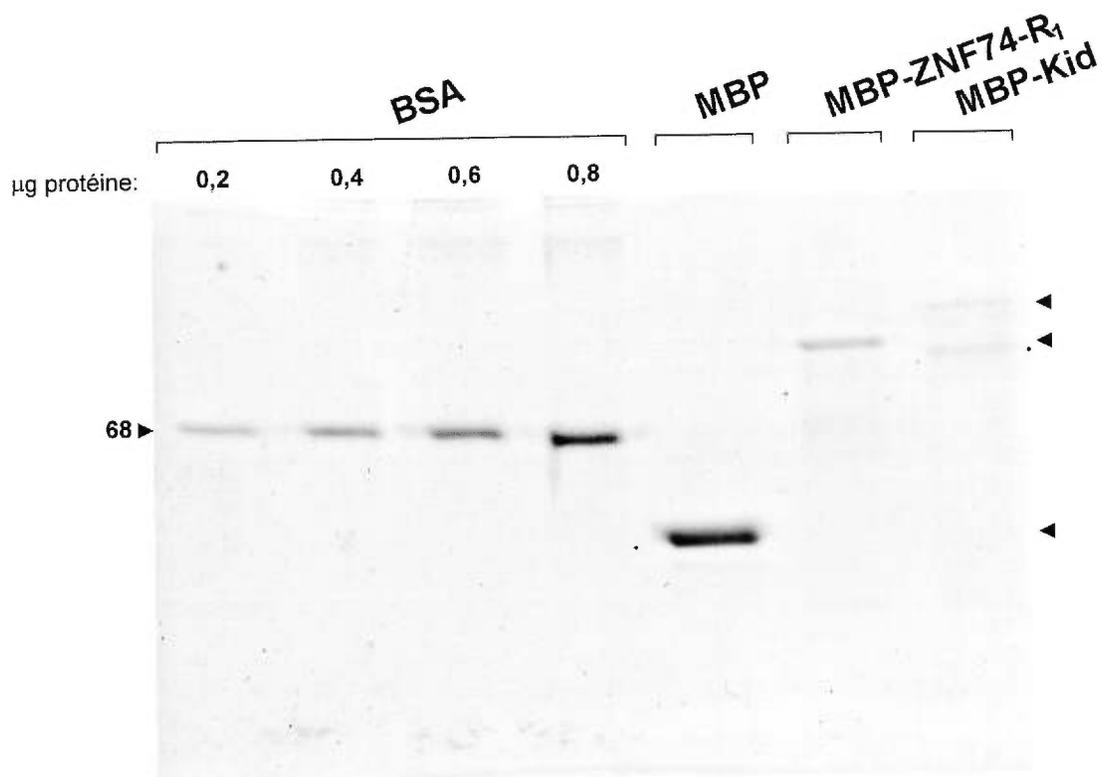
Comme démontré à la section précédente, contrairement aux résultats obtenus avec plusieurs protéines contenant deux à cinq motifs doigt de zinc, les expériences de sélection d'oligonucléotides double brin ne nous ont pas permis d'identifier des séquences spécifiques d'ADN double brin reconnues par la protéine ZNF74. Il semble que cette protéine de la famille KRAB à multidoigt de zinc ait une affinité pour l'ADN qui ne requiert pas une séquence en nucléotides précise. Étant donné que des études précédentes ont révélé que ZNF74 est capable de se lier à certains ARN (Grondin *et al.*, 1996), nous

avons émis l'hypothèse que ZNF74 pouvait avoir aussi une affinité pour l'ADN simple brin et/ou l'ADN hétéroduplex. Alors que nous allions tester cette hypothèse, le groupe de Witzgall (Elser *et al.*, 1997) a rapporté que la protéine Kid<sub>1</sub> de la famille KRAB à multidoigt était capable de lier de l'ADN hétéroduplex de façon indépendante de la séquence. Tout en continuant à mettre l'emphase sur l'étude de ZNF74, nous avons donc cherché à comparer les propriétés de liaison à l'ADN de ZNF74 et de Kid<sub>1</sub>. Pour ce faire, des expériences de retardement sur gel ont été réalisées avec trois types d'ADN, c'est-à-dire de l'ADN simple brin, double brin et hétéroduplex en utilisant entre autre certains des oligonucléotides utilisés dans l'étude du groupe de Witzgall (Elser *et al.*, 1997) (oligonucléotides identifiables par les deux premières lettres de leur nom, soient SS). Pour les expériences de retardement sur gel, les protéines de fusion MBP-ZNF74-R<sub>1</sub> de même que MBP-Kid<sub>1</sub> et la protéine MBP (contrôle négatif) ont été utilisées. Ces protéines ont été préalablement purifiées sur résine d'amylose et récupérées sous forme libre et soluble par élution avec du maltose. La pureté et l'intégrité des protéines solubles utilisées pour les essais de retardement sur gel ont été préalablement vérifiées après migration sur gel d'acrylamide dénaturant (Figure 8). Une seule protéine est éluee des résines d'affinité dans le cas de MBP-ZNF74-R<sub>1</sub> et MBP alors que MBP-Kid<sub>1</sub> apparaît présent en quantité égale avec une seconde protéine de taille réduite correspondant très probablement à un produit de dégradation.

### **3.1 Étude de la liaison à l'ADN simple brin**

#### **3.1.1 Effet de la concentration de protéines pour la formation de complexes protéine/ADN et spécificité d'interaction**

**Figure 8. Analyse par électrophorèse des protéines de fusion solubles utilisées pour les essais de retardement.** La pureté et l'intégrité des protéines de fusion MBP-ZNF74-R<sub>1</sub>, MBP-Kid<sub>1</sub> et MBP, purifiées sur résine d'amylose, éluées en présence de maltose et resuspendues dans du tampon de Laemmli (voir section 1.3 dans *Matériel et Méthodes*) sont vérifiées par électrophorèse sur gel de polyacrylamide (8%) après coloration au bleu de Coomassie. Les protéines de fusion purifiées correspondent respectivement à l'équivalent de 1/300 (MBP-ZNF74-R<sub>1</sub> et MBP-Kid<sub>1</sub>) et 1/1500 (MBP) des cultures initiales de 500 mL induites pendant 3 heures avec de l'IPTG. Les pointes de flèche indiquent la position de migration des trois protéines de fusion purifiées et un point indique un produit de dégradation de la protéine MBP-Kid<sub>1</sub> observable dans toutes les préparations de cette protéine de fusion. Des quantités croissantes de BSA (0,2 µg à 0,8 µg) sont aussi utilisées pour valider le dosage des protéines purifiées effectué en parallèle par la méthode de Bradford.

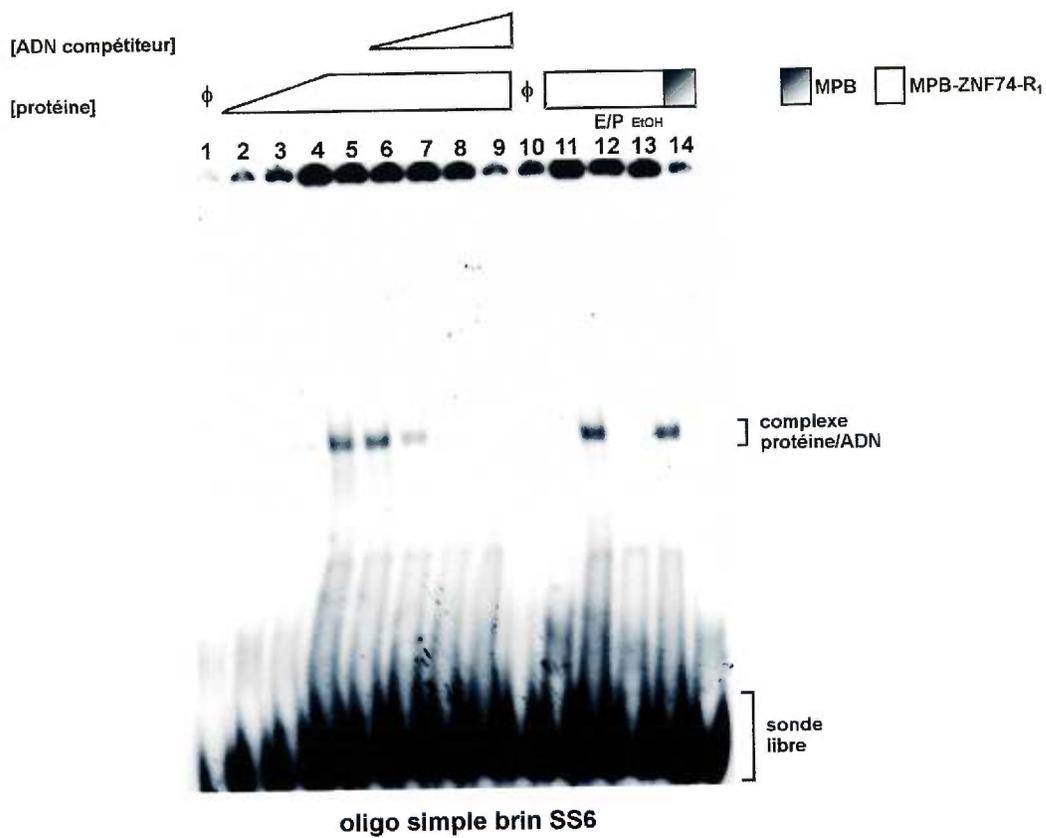


Pour tester la liaison possible de ZNF74 à de l'ADN simple brin, nous avons initialement utilisé un oligonucléotide simple brin (47 pb), SS6, décrit par le groupe de Witzgall (Elser *et al.*, 1997). Il apparaît à la Figure 9 que ZNF74 aussi bien que Kid<sub>1</sub> sont capables de retarder la migration sur gel non dénaturant de l'oligonucléotide marqué SS6 utilisé à une concentration de 0,5 nM. Il est mis en évidence que l'intensité des complexes protéine/ADN augmente graduellement en fonction de la concentration de protéines (Figure 9 A et B, pistes 2 à 5) étant plus difficilement détectable à la concentration minimum de 10 nM (piste 2) et très clairement identifiable à la concentration maximum de 100 nM (piste 5).

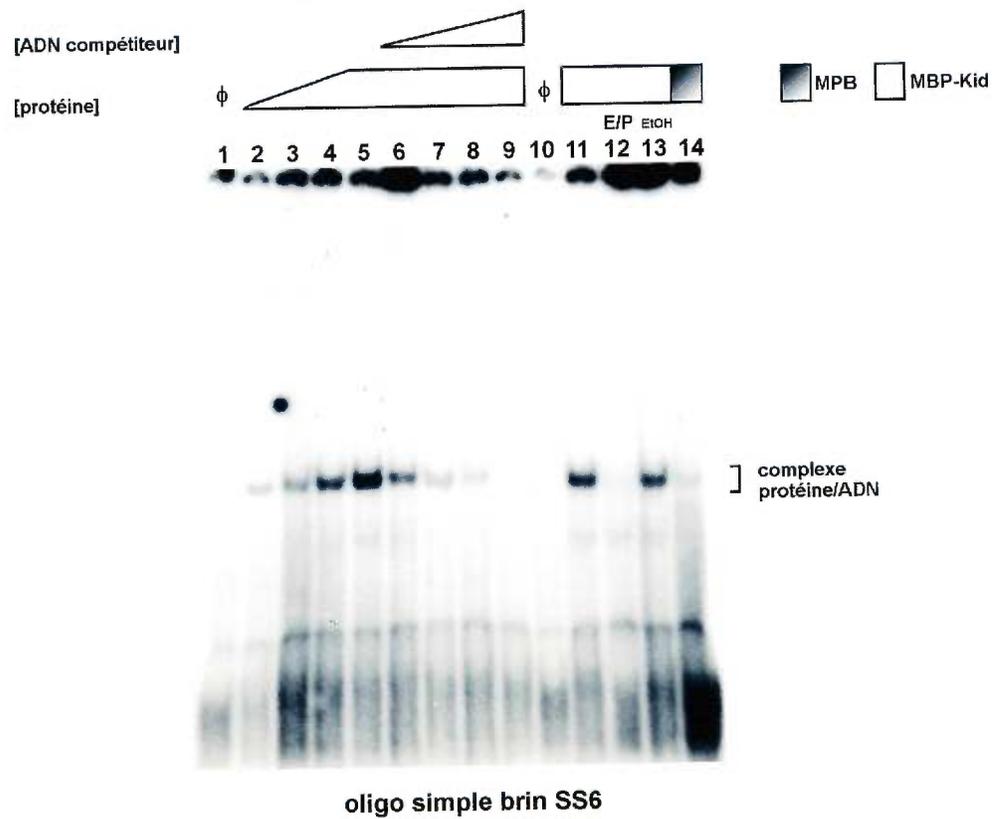
La spécificité de la liaison de l'oligonucléotide SS6 radioactif aux protéines ZNF74 et Kid<sub>1</sub> est confirmé par une étude de compétition en présence de concentrations croissantes d'oligonucléotide SS6 non marqué (Figure 9 A et B, pistes 6 à 9). Dès l'addition d'une concentration d'oligonucléotide non marqué équivalente à la concentration d'oligonucléotide marqué radioactivement, on observe une diminution significative de la quantité de complexes ZNF74/SS6 et Kid<sub>1</sub>/SS6 radioactifs (piste 6) par rapport à la quantité de complexes détectés en absence d'oligonucléotide compétiteur non marqué (piste 5). D'autre part, en présence d'un excès de 100 fois de compétiteur non marqué, les complexes protéine/ADN radioactifs deviennent essentiellement non détectables. L'efficacité du compétiteur non marqué à prévenir la formation de complexes protéine/ADN radioactifs permet de conclure que la liaison des deux protéines de la famille KRAB à multidoigt à SS6 est spécifique. Il est à noter que les complexes protéine/ADN identifiés comme spécifiques ne sont pas détectés

**Figure 9. Étude par retardement sur gel de la liaison des protéines ZNF74-R<sub>1</sub> et Kid<sub>1</sub> à l'ADN simple brin.** (A) La protéine MBP-ZNF74-R<sub>1</sub> est incubée en quantité croissante de 10 nM (correspondant à 0,2 pmoles ou 18,78 ng), 25 nM, 50 nM et 100 nM dans les pistes 2 à 5 respectivement, en présence de 0,5 nM (correspondant à 10 fmoles ou 0,15 ng) d'oligonucléotide simple brin SS6 marqué radioactivement (activité spécifique: 6100 cpm/fmole). La spécificité de liaison est évaluée par l'addition d'ADN non marqué (SS6) à concentration croissante de 1X, 5X, 10X et 100X dans les pistes 6 à 9 respectivement, en présence de 0,5 nM de SS6 marqué et de 100 nM de MBP-ZNF74-R<sub>1</sub>. Le rôle de la structure doigt de zinc est testé par l'addition d'une combinaison d'EDTA/1,10-phénantroline (piste 12) à 0,5 nM de SS6 marqué et 100 nM de MBP-ZNF74-R<sub>1</sub>. Des contrôles sans protéine (pistes 1 et 10), avec addition d'une concentration finale d'éthanol de 0,5% à 100 nM de MBP-ZNF74-R<sub>1</sub> (piste 11) et avec 100 nM de MBP sont inclus. (B) La même expérience qu'en A a été réalisée avec des concentrations équivalentes de la protéine MBP-Kid<sub>1</sub> en remplacement de MBP-ZNF74-R<sub>1</sub>. En A et en B, les complexes protéines/ADN sont séparés de la sonde libre sur gel de polyacrylamide non dénaturant (4%). Les complexes protéines/ADN spécifiques et la sonde libre sont mis en évidence sur la figure (la sonde libre est sortie du gel dans l'expérience présentée en B). Les gels présentés en A et en B ont été exposés respectivement 7 et 9 heures à -80°C.

A



B



dans des essais contrôles effectués en l'absence de protéine (pistes 1 et 10) ou en présence de la protéine MBP (piste 14).

### **3.1.2 Importance de l'intégrité de la structure doigt de zinc pour la liaison à l'ADN**

Pour déterminer l'importance de l'intégrité structurale des motifs doigt de zinc des protéines ZNF74 et Kid<sub>1</sub> pour leur liaison spécifique à l'oligonucléotide simple brin SS6, on a étudié l'effet de la chélation du zinc sur la formation des complexes protéine/ADN. L'ajout durant la réaction de liaison d'un mélange d'EDTA et de 1,10-phénantroline comme agents chélateurs a pour conséquence d'empêcher complètement la formation de complexes protéine/ADN (Figure 9 A et B, piste 12). L'absence de liaison résulte bien de la déstabilisation de la structure doigt de zinc et non d'une interférence au niveau de la liaison due à la présence d'une faible concentration d'éthanol utilisée comme solvant de la 1,10-phénantroline. En effet, on observe la même quantité de complexes protéine/ADN spécifiques en présence (piste 13) et en absence (piste 11) d'éthanol dans le milieu de liaison. Ces résultats confirment que l'intégrité structurale des domaines à multidoigt de zinc de ZNF74 et Kid<sub>1</sub> est requise pour leur liaison spécifique à l'ADN simple brin.

### **3.1.3 Liaison d'oligonucléotides simple brin sélectionnés aléatoirement**

Afin de déterminer si l'affinité et la spécificité de la liaison de la protéine ZNF74 à l'ADN simple brin pourrait dépendre de la séquence nucléotidique, on a sélectionné deux autres oligonucléotides (SS1 et SS2, 65 bases) décrits par le groupe de Witzgall

(Esler *et al.*, 1997) ainsi que des oligonucléotides choisis au hasard parmi une collection d'oligonucléotides disponibles au laboratoire. Il apparaît que la proportion des nucléotides G, A, T et C des séquences des oligonucléotides sélectionnés varie légèrement d'un oligonucléotide à l'autre (voir légende de la Figure 10 ainsi que *Matériel et Méthodes*, Section 4.1.1). D'autre part, les oligonucléotides sélectionnés ont une taille variant entre 21 et 65 nucléotides. Outre l'oligonucléotide SS6 testé précédemment, six des huit nouveaux oligonucléotides testés sont liés par la protéine ZNF74 (Figure 10). ZNF74 ne semble pas par contre avoir d'affinité particulière pour les deux oligonucléotides les plus courts soient l'oligonucléotide RNPU 1 (21 nucléotides) et l'oligonucléotide 845 (23 nucléotides). Dans le cas des l'oligonucléotides 844 (pistes 7 et 9) et RNPU 6 (pistes 10 et 12), deux complexes majeurs sont détectables mais, contrairement au complexe ayant la plus grande mobilité électrophorétique dont la formation dépend spécifiquement de l'intégrité du domaine doigt de zinc de ZNF74, celui migrant plus haut dans le gel se forme malgré la présence d'agents chélatants déstabilisant la structure doigt de zinc et perturbant la structure native de ZNF74 (pistes 8 et 11). La spécificité de liaison de ces complexes de plus faible mobilité électrophorétique n'a pas été étudiée ici en détail (on peut suggérer qu'il se soit formé dans ces deux cas des complexes multimériques stables de la protéine ZNF74 dont le zinc aurait été rendu inaccessible aux agents chélatants, tels qu'ajoutés dans cet essai au moment de la réaction de liaison).

En tenant compte du fait que le gel de retardement des oligonucléotides SS1 et SS2 (pistes 19 à 24) a été exposé 2,5 fois moins longtemps que les autres gels présentés à la Figure 10 (pistes 1 à 18), ZNF74 semble avoir une plus grande affinité

**Figure 10. Étude par retardement sur gel de la liaison de la protéine ZNF74-R<sub>1</sub> à des oligonucléotides simple brin de taille différente.** Pour chacun des oligonucléotides utilisés, la première piste correspond au complexe protéine/ADN formé par l'incubation de 0,1 nM (2 fmoles) d'oligonucléotide marqué avec 100 nM de protéine MBP-ZNF74-R<sub>1</sub>. Le rôle de la structure doigt de zinc (piste centrale) est testé par l'addition d'une combinaison EDTA/1,10-phénantroline à une même concentration d'oligonucléotide et de protéine. La piste de droite correspond au contrôle éthanol effectué par l'addition de 0,5% d'éthanol (concentration finale). L'activité spécifique de chacun des oligonucléotides est très similaire: RNPU1 (8400 cpm/fmole), 845 (4700 cpm/fmole), 844 (9400 cpm/fmole), RNPU6 (7200 cpm/fmole), Jul1 (10 300 cpm/fmole), Jul2 (10 000 cpm/fmole), SS1 (6700 cpm/fmole), SS2 (9500 cpm/fmole). La composition en G, A, T, C de chacun des oligonucléotides varie sensiblement et le nombre absolu de chacun des quatre nucléotides est indiqué entre parenthèse ci-après: RNPU1 (8,4,4,5), 845 (5,7,4,7), 844 (7,5,10,7), RNPU6 (6,5,9,10), Jul1 (14,11,9,9), Jul2 (12,8,14,9), SS1 (17,19,13,16), SS2 (16,12,19,18). Les complexes protéine/ADN sont séparés de la sonde libre sur gel de polyacrylamide non dénaturant (4%). Les pistes 1 à 9 provenant d'un premier gel de même que les pistes 10 à 18 provenant d'un deuxième gel ont été exposées pendant 120h à -80°C . Les pistes 19 à 24 proviennent d'un troisième gel qui a été exposé 2,5 fois moins longtemps (48h à -80°C).

oligo  
simple brin:

21 mers	23 mers	29 mers	30 mers	43 mers	43 mers	65 mers	65 mers
RNPU 1	845	844	RNPU 6	Jul 1	Jul 2	SS 1	SS 2
E/P <sub>EIOH</sub>							
1 2 3	4 5 6	7 8 9	10 11 12	13 14 15	16 17 18	19 20 21	22 23 24

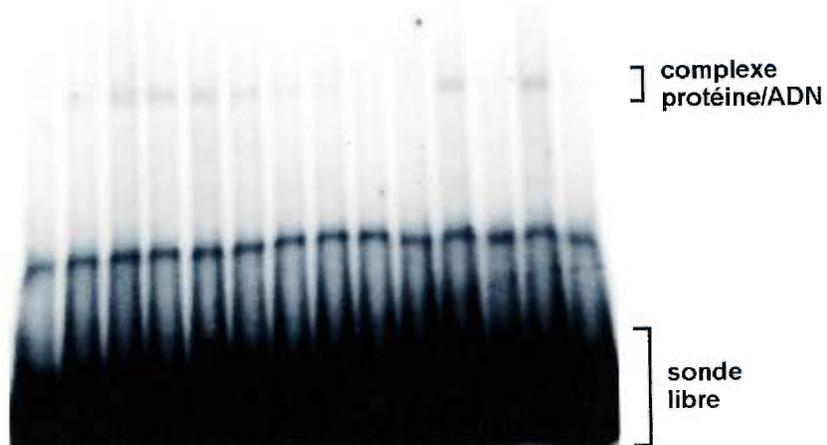
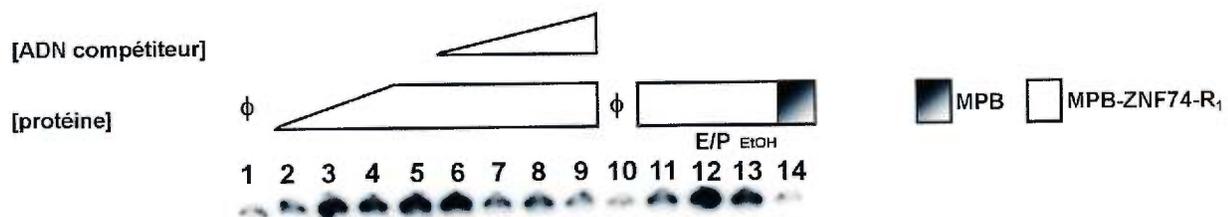


pour les oligonucléotides les plus longs. Cependant, seule une étude beaucoup plus exhaustive pourrait confirmer cette tendance et permettre de déterminer si ZNF74 reconnaît avec plus d'affinité des séquences spécifiques consensus ayant plus de chance d'être représentées dans des séquences plus longues ou bien si la reconnaissance spécifique d'oligonucléotides simple brin par la protéine à multidoigt ZNF74, de façon dépendante ou indépendante de la séquence, est favorisée par la présence d'une plate-forme oligonucléotidique de plus grande taille.

### **3.2 Étude de la liaison à l'ADN double brin**

Les études de buvardage de type South-Western (Section 1) et de sélection de sites (Section 2) ont mis en évidence que ZNF74 avait une affinité pour l'ADN double brin. Pour permettre une comparaison entre l'affinité de ZNF74 pour l'ADN double brin et son affinité pour l'ADN simple brin, nous avons analysé la liaison de ZNF74 à un oligonucléotide double brin de 47 pb (SS6/SS6c) sur gel de retardement. Afin de faciliter la comparaison, nous avons employé des conditions de liaison tout à fait similaires dans les deux types d'expérience en utilisant notamment les mêmes quantités de protéine ainsi qu'une même quantité des oligonucléotides à une activité spécifique comparable. Il apparaît à la Figure 11 que ZNF74 est capable de retarder la migration sur gel non dénaturant de l'oligonucléotide marqué SS6/SS6c double brin utilisé à une concentration de 0,5 nM. L'intensité des complexes protéine/ADN augmente faiblement en fonction de la concentration de ZNF74 variant de 10 à 100 nM (Figure 11, pistes 2 à 5). La spécificité de la liaison de l'oligonucléotide SS6/SS6c double brin radioactif à la protéine ZNF74 est confirmé par une étude de

**Figure 11. Étude par retardement sur gel de la liaison de la protéine ZNF74-R<sub>1</sub> à l'ADN double brin.** Pour étudier l'effet de la concentration de protéines pour la détection de complexes protéine/ADN, la protéine MBP-ZNF74-R<sub>1</sub> est incubée en quantité croissante de 10 nM (correspondant à 0,2 pmoles ou 18,78 ng), 25 nM, 50 nM et 100 nM dans les pistes 2 à 5 respectivement, en présence de 0,5 nM (correspondant à 10 fmoles ou 0,31 ng) d'oligonucléotide double brin SS6/SS6c marqué radioactivement (activité spécifique: 4600 cpm/fmole). La spécificité de liaison est évaluée par l'addition d'ADN non marqué (SS6/SS6c) à concentration croissante de 1X, 5X, 10X et 100X dans les pistes 6 à 9 respectivement, en présence de 0,5 nM de SS6/SS6c marqué et de 100 nM de MBP-ZNF74-R<sub>1</sub>. Le rôle de la structure doigt de zinc est testé par l'addition d'une combinaison EDTA/1,10-phénantroline (piste 12) à 0,5 nM de SS6/SS6c marqué et 100 nM de MBP-ZNF74-R<sub>1</sub>. Des contrôles sans protéine (pistes 1 et 10), avec addition d'une concentration finale d'éthanol de 0,5% à 100 nM de MBP-ZNF74-R<sub>1</sub> (piste 11) et avec 100 nM de MBP (piste 14) sont inclus. Les complexes protéine/ADN sont séparés de la sonde libre sur un gel de polyacrylamide non dénaturant (4%) qui a été exposé 22 heures à -80°C. Les complexes protéine/ADN spécifiques et la sonde libre sont mis en évidence sur la figure.



oligo double brin SS6/SS6c

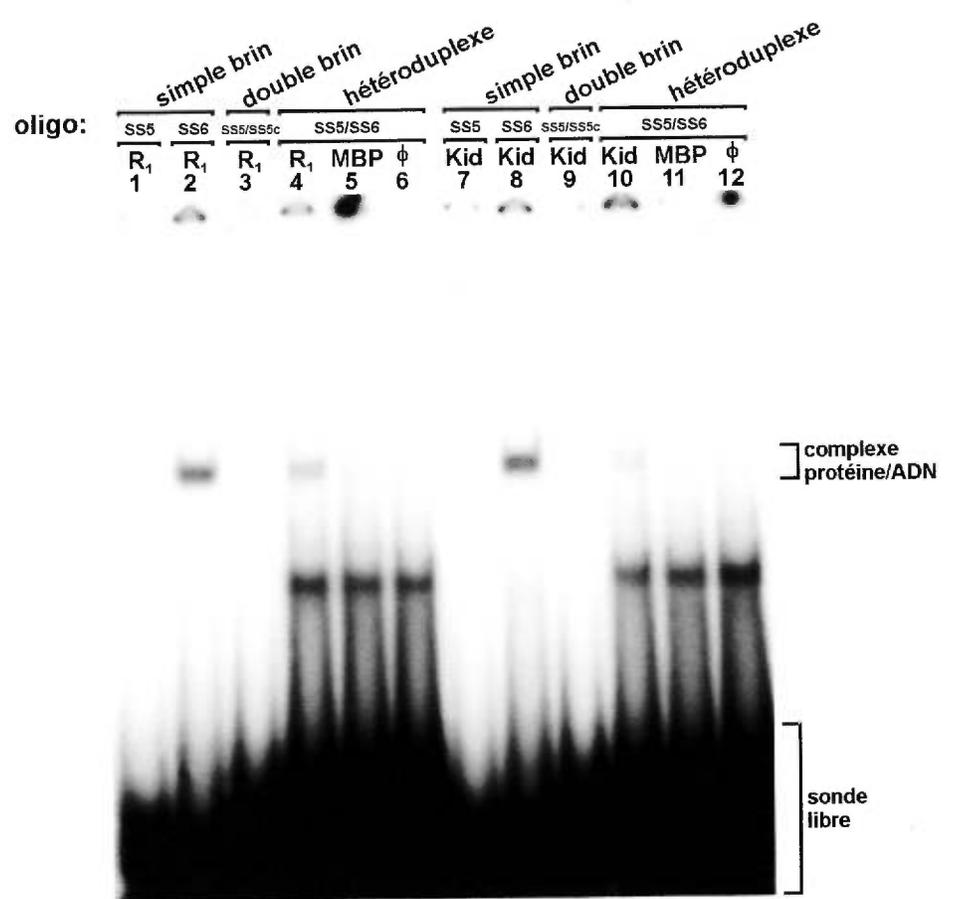
compétition en présence de concentrations croissantes d'oligonucléotide SS6/SS6c non marqué en excès de 1 à 100 fois (Figure 11, pistes 6 à 9). La formation de complexes protéine/ADN identifiés comme spécifiques est prévenue lorsque l'on déstabilise la structure doigt de zinc de ZNF74 par chélation du zinc (piste 12) et n'est pas détectée dans des essais contrôles effectués en l'absence de protéine (pistes 1 et 10) ou en présence de la protéine MBP (piste 14). Des résultats similaires ont été obtenus avec deux autres oligonucléotides double brin de 47 pb (SS5/SS5c) et 65 pb (SS2/SS2c) (Figure 12 piste 3 et résultats non présentés).

L'affinité de ZNF74 pour les trois oligonucléotides double brin testés par retardement sur gel apparaît plus faible que pour les oligonucléotides simple brin tel que déduit de la comparaison de l'intensité des complexes protéine/ADN simple brin (Figure 9, Figure 10, et Figure 12 pistes 1 et 2) avec celle des complexes protéine/ADN double brin (Figure 11, et Figure 12 piste 3, et résultats non présentés).

### **3.3 Étude de la liaison à l'ADN hétéroduplex**

Le groupe de Witzgall (Esler *et al.*, 1997) ayant rapporté que la protéine Kid<sub>1</sub> de la famille KRAB à multidoigt se liait à l'ADN hétéroduplex, nous avons étudié en parallèle la liaison de ZNF74 et de Kid<sub>1</sub> à l'ADN hétéroduplex. Trois oligonucléotides de type hétéroduplex ont été testés afin de vérifier si les protéines à motifs doigt de zinc pouvaient se lier à ce type de structure d'ADN. Les oligonucléotides hétéroduplex testés consistent en deux extrémités de 20 nucléotides pairés (en 5' et en 3') séparés par une bulle de 7 (SS5/SS6), de 15 (SS7/SS8) ou de 25 nucléotides (SS1/SS2). Les oligonucléotides sont identiques dans la région pairée pour les trois hétéroduplex. La

**Figure 12. Étude par retardement sur gel de la liaison des protéines ZNF74-R<sub>1</sub> et Kid<sub>1</sub> à trois différentes structures de l'ADN.** Les pistes 1 à 4 correspondent à l'incubation de la protéine MBP-ZNF74-R<sub>1</sub> (100 nM) avec une même concentration (0,5 nM, 10 fmole) respectivement d'oligonucléotide simple brin SS5 (activité spécifique: 4100cpm/fmole), d'oligonucléotide simple brin SS6 (activité spécifique: 4300cpm/fmole), d'oligonucléotide double brin SS5/SS5c (activité spécifique: 5500cpm/fmole) et d'oligonucléotide hétéroduplex SS5/SS6 (activité spécifique: 6100cpm/fmole). Les pistes 7 à 10 représentent l'incubation de la protéine MBP-Kid<sub>1</sub> (100 nM) avec les mêmes oligonucléotides utilisés aux pistes 1 à 4. Des contrôles avec 100 nM de MBP (pistes 5 et 11) et sans protéine (pistes 6 et 12) sont inclus. Les complexes protéine/ADN sont séparés de la sonde libre sur un gel de polyacrylamide non dénaturant (4%) qui a été exposé 3 heures à -80°C. Les complexes protéine/ADN spécifiques et la sonde libre sont mis en évidence sur la figure.

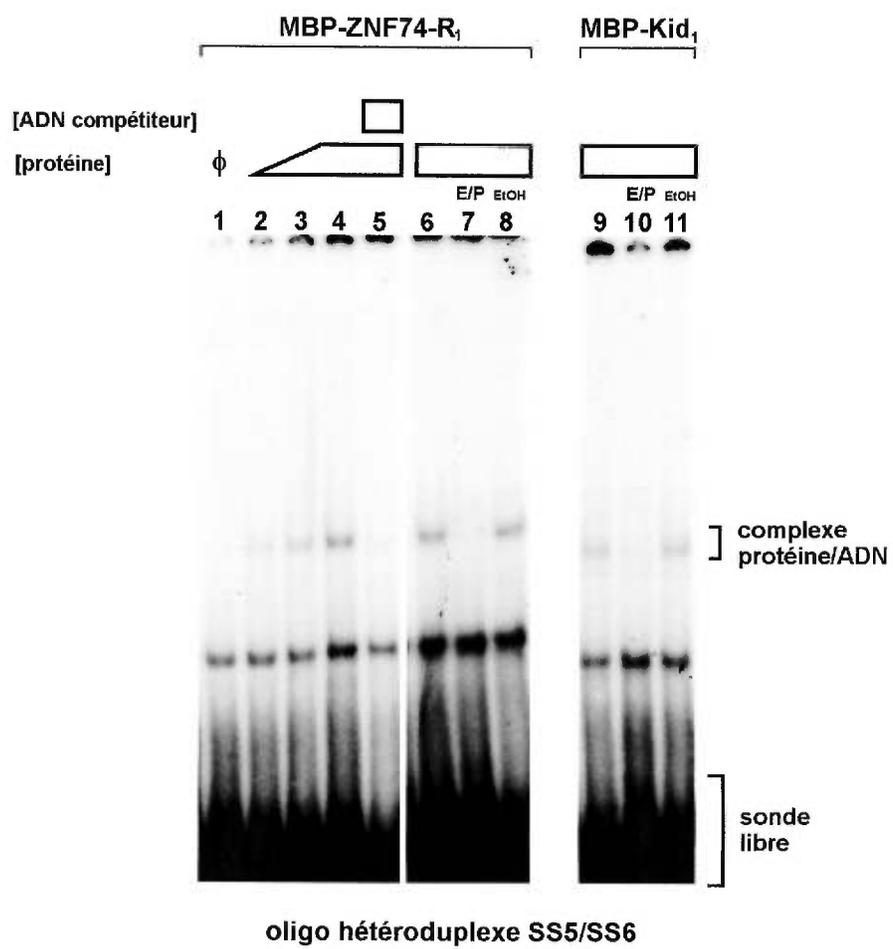


bulle est créée par la présence de bases purine-purine ou pyrimidine-pyrimidine placées les unes en face des autres, ce qui va à l'encontre de l'appariement de Watson&Crick ou de tout autre appariement possible. Il apparaît à la Figure 13 que ZNF74 est capable de lier un ADN hétéroduplex contenant une bulle de 7 nts (SS5/SS6) de façon dépendante de la concentration de protéine (pistes 2 à 4) et spécifique (voir compétition à la piste 5). La liaison de cet oligonucléotide hétéroduplex à ZNF74 (piste 7 et 9) ainsi qu'à Kid<sub>1</sub> (pistes 10 et 11) est dépendante de l'intégrité de la structure doigt de zinc.

La comparaison de la liaison d'ADN double brin correspondant à l'oligo SS6 et son complémentaire (SS6/SS6c, Figure 11) et d'ADN hétéroduplex SS5/SS6 (Figure 13) indique une affinité supérieure de ZNF74 pour l'ADN hétéroduplex SS5/SS6. De même, il apparaît à la Figure 12 que ZNF74 et Kid<sub>1</sub> ont une affinité plus grande pour l'ADN hétéroduplex SS5/SS6 (pistes 4 et 10) que pour l'ADN double brin SS5/SS5c (pistes 3 et 9). Il est à remarquer que les deux oligonucléotides qui composent l'hétéroduplex, soient SS5 (pistes 1 et 7) et SS6 (pistes 2 et 8) lient aussi ZNF74 et Kid avec une affinité plus grande que l'ADN double brin.

La liaison des deux protéines de la famille KRAB-multidoigt à un hétéroduplex avec une boucle de 15 nts (SS7/SS8) est tout à fait comparable à celle de l'hétéroduplex SS5/SS6 (résultats non présentés). Par contre, contrairement aux résultats rapportés par le groupe de Witzgall, on n'a pu détecter de liaison ni de Kid<sub>1</sub> ni de ZNF74 (résultats non présentés) à l'hétéroduplex avec une boucle de 25 nts (SS1/SS2) alors que les oligonucléotides simple brin les composants sont liés de façon significative et comparable par les deux protéines (Figure 10 et résultats non présentés) (voir Section *Discussion*).

**Figure 13. Étude par retardement sur gel de la liaison de la protéine ZNF74-R<sub>1</sub> à l'ADN hétéroduplex.** Les protéines MBP-ZNF74-R<sub>1</sub> (pistes 1 à 8) et Kid<sub>1</sub> sont incubées à une concentration de 25 nM (piste 2), 50 nM (piste 3), 100 nM (pistes 4 à 11) en présence de 65,000 cpm (~ 0.5 nM) d'oligonucléotide hétéroduplex SS5/SS6. La spécificité de liaison est évaluée en présence d'un excès de 50 fois d'ADN non marqué (piste 5). Aux pistes 7 et 10, une combinaison d'EDTA/1,10-phénantroline est ajoutée pendant la réaction de liaison. La piste 1 correspond à un contrôle sans protéine. Les complexes protéine/ADN sont séparés de la sonde libre sur un gel de polyacrylamide non dénaturant (4%) qui a été exposé 3 heures à -80°C. Les complexes protéine/ADN spécifiques et la sonde libre sont mis en évidence sur la figure.



Il est à noter qu'au niveau des gels de retardement présentant des oligonucléotides simple brin (Figure 9) et plus particulièrement double brin (Figure 11) et hétéroduplex (Figure 13), on observe la présence d'une bande de taille intermédiaire, située entre ce qui est identifié comme la sonde libre et le complexe protéine/ADN, dont on ne connaît pas l'origine. Il est peu probable que, dans le cas des expériences avec les oligonucléotides double brin et hétéroduplex, cette bande corresponde aux oligonucléotides simple brin non appariés. En effet, les oligonucléotides double brin et hétéroduplex ont été préalablement purifiés sur gel de façon à éliminer la radioactivité libre et les oligonucléotides non appariés. Dans tous les cas, les oligonucléotides simple brin non appariés étaient indétectables suggérant l'appariement complet des oligonucléotides en présence et la stabilité des oligonucléotides appariés lors de la migration sur gel.

De façon générale, nos résultats indiquent que ZNF74 et Kid<sub>1</sub> ont une affinité pour l'ADN simple brin, double brin et hétéroduplex avec une préférence pour l'ADN simple brin. Il est d'autre part particulièrement frappant de constater que les affinités de ZNF74 et de Kid<sub>1</sub> pour chacun des oligonucléotides testés sont tout à fait comparables suggérant que ces deux protéines KRAB à motifs multidoigt sont régies par des règles d'interaction avec les acides nucléiques comparables. Aucune différence majeure n'a pu être observée, d'autre part, en utilisant des préparations différentes de chacune des protéines ou des préparations différentes de chacun des oligonucléotides marqués.

## **IV-DISCUSSION**

Dans ce mémoire, nous avons mis en évidence que la protéine KRAB à multidoigt codée par *ZNF74*, un gène situé dans la région de délétion critique associée au syndrome de DiGeorge, a non seulement une affinité pour l'ARN tel que démontré précédemment (Grondin *et al.*, 1996) mais aussi pour l'ADN. Alors que la majorité des facteurs transcriptionnels se lient à des séquences d'ADN double brin, aucune séquence spécifique d'ADN double brin reconnue par *ZNF74* n'a pu être identifiée. Par contre, nos études ont montré que *ZNF74* reconnaît l'ADN simple brin préférentiellement à l'ADN double brin et hétéroduplex. Récemment, quelques facteurs transcriptionnels se liant à l'ADN simple brin ont été caractérisés (Tomonaga and Levens, 1995, Duncan *et al.*, 1996, Werter *et al.*, 1998) et certains d'entre eux jouent aussi un rôle de régulateur post-transcriptionnel grâce à leur capacité de se lier à l'ARN (Tomonaga and Levens, 1995, Grossman *et al.*, 1998). Nous suggérons que *ZNF74* puisse de façon similaire fonctionner comme un double régulateur de l'expression génique soit comme facteur transcriptionnel *via* son domaine KRAB et sa liaison possible à l'ADN ainsi que, tel que proposé précédemment, comme facteur de maturation de l'ARN *via* sa liaison à l'ARN (Grondin *et al.*, 1996) et à la forme hyperphosphorylée de l'ARN polymérase II au niveau de sites nucléaires riches en facteurs d'épissage (Grondin *et al.*, 1997).

## **1. SÉLECTION DE SÉQUENCES DE LIAISON À L'ADN DOUBLE BRIN**

Afin d'identifier des séquences reconnues par un facteur de transcription, il est possible d'effectuer des sélections de sites *in vitro* en utilisant soit des banques d'oligonucléotides dégénérés et soit des banques de fragments d'ADN génomique de

faible taille (200-400pb). Pour effectuer la sélection des séquences se liant spécifiquement à des facteurs transcriptionnels d'intérêt, plusieurs méthodes de purification ont été utilisées soit la rétention sur résine d'affinité (Oliphant *et al.*, 1989, Funabiki *et al.*, 1994, Gong *et al.*, 1994, Baron *et al.*, 1995, Zweidler-McKay *et al.*, 1996, Li *et al.*, 1997, Mai and Breeden, 1997) ou sur membrane de nitrocellulose (Swirnoff and Milbrandt, 1995, Tang and Breitman, 1995), l'immunoprécipitation (Pollock and Treisman, 1990) ou le retardement sur gel (Mavrothalassitis *et al.*, 1990, Morris *et al.*, 1994, Vortkamp *et al.*, 1995, Harada *et al.*, 1995, Tsai and Reed, 1998). Comparativement à l'utilisation de banques de fragments d'ADN génomique, l'utilisation de banques d'oligonucléotides dégénérés a comme principaux avantages 1) de limiter le séquençage vu la petite taille des fragments utilisés ainsi que 2) de faciliter la recherche d'un consensus. Cette méthode a permis par exemple d'identifier des séquences d'ADN auxquelles se lient les protéines à motifs doigt de zinc Evi-1 (Funabiki *et al.*, 1994), MZF1 (Morris *et al.*, 1994), NGF1-A (Swirnoff and Milbrandt, 1995), GLI3 (Vortkamp *et al.*, 1995), BCL6 (Baron *et al.*, 1995), Gfi-1 (Zweidler-McKay *et al.*, 1996) et PLZF (Li *et al.*, 1997). L'utilisation des banques d'ADN génomique a, quant à elle, permis l'identification des séquences d'ADN auxquelles se lient les protéines à motifs doigt de zinc TFIIIA (Kinzler and Vogelstein, 1989), GLI (Kinzler and Vogelstein, 1990), WT1 (Nakagama *et al.*, 1995) et MOK2 (Arranz *et al.*, 1997). Comme approche initiale, nous avons choisi de tenter d'isoler des séquences spécifiques reconnues par ZNF74 par sélection d'oligonucléotides dégénérés.

Par des études cristallographiques, il a été montré que la protéine Egr1 à motifs doigt de zinc de type Cys<sub>2</sub>His<sub>2</sub> (dont les trois motifs ont une structure similaire aux motifs

doigt de zinc de ZNF74) est capable de reconnaître un total de 9 pb consécutives (Pavletich et Pabo, 1991). Cette reconnaissance se fait par insertion de l'hélice  $\alpha$  de chaque motif doigt de zinc dans le sillon majeur formé par trois nucléotides consécutifs, chaque doigt de zinc reconnaissant ainsi 3 pb. La protéine ZNF74 possédant 12 doigts de zinc, on serait à même de penser que le site de liaison à l'ADN pourrait être composé d'un maximum de 36 nucléotides. Il est à noter cependant que les résultats de cristallographie (Pavletich et Pabo, 1991) et de résonance magnétique nucléaire ainsi que l'énoncé de règles stéréochimiques précises ont permis au groupe de Suzuki (Suzuki, 1993, Suzuki *et al.*, 1994) de proposer des règles empiriques permettant de prédire quels sont les motifs d'une protéine à motifs Cys<sub>2</sub>His<sub>2</sub> qui sont capables de lier spécifiquement l'ADN. L'interaction spécifique des doigts de zinc avec l'ADN se ferait *via* les chaînes latérales des acides aminés de la région N-terminal de l'hélice  $\alpha$  soit en position  $-1$  de l'hélice  $\alpha$ ,  $+2$ ,  $+3$  et  $+6$  (Figure 14). De plus, plusieurs liaisons non spécifiques obtenues par ponts hydrogène entre les acides aminés des positions  $-5$ ,  $+7$ ,  $+9$  et  $+12$  et les phosphates du squelette de l'ADN semblent stabiliser l'interaction. L'analyse de la séquence primaire de ZNF74 selon les règles de Suzuki révèle que seulement 7 des 12 doigts de zinc de ZNF74 (soient les doigts 1, 2, 4, 5, 7, 11 et 12) seraient susceptibles de se lier spécifiquement à l'ADN (Figure 14).

Nous avons initialement décidé de tester une banque d'oligonucléotides dégénérés de 16 pb (complexité de la banque  $4^{16}$ ) ce qui nous permettait d'avoir une bonne représentativité de toutes les séquences. Nous réalisons cependant que des séquences de 16 nucléotides ne nous permettraient d'accommoder que la liaison potentielle de 5 motifs

**Figure 14. L'analyse stéréochimique des douze motifs Cys<sub>2</sub>His<sub>2</sub> de ZNF74 permet de prédire lesquels sont probablement capables de lier spécifiquement l'ADN.** Les doublets conservés de Cystéine et d'Histidine des douze motifs Cys<sub>2</sub>His<sub>2</sub> de ZNF74 représentés en caractères gras sont encadrés par des rectangles tandis que les acides aminés hydrophobes conservés sont représentés en caractères gras seulement. Les acides aminés aux positions -1, +2, +3 et +6 ont été analysés en fonction de la grosseur de leurs chaînes latérales. Ceux qui respectent les règles stéréochimiques permettant une liaison spécifique à l'ADN sont encadrés par un carré. La présence d'au moins 3 acides aminés respectant ces règles dans un même motif suggère que celui-ci peut lier spécifiquement l'ADN.

		-1		+2 +3		+6		H R R W		S R E K A		a.a. favorables	liaison probable													
Y	K	C	G	E	C	G	K	A	F	R	Q	S	S	L	T	L	H	S	R	E	K	A	1	3/4	+	
F	F	C	D	E	C	G	K	A	F	T	W	S	T	N	L	L	E	H	T	G	E	K	P	2	4/4	+
Y	K	C	G	E	C	G	K	A	F	S	C	H	S	S	L	N	V	H	T	G	E	R	P	3	1/4	-
Y	R	C	S	A	C	E	K	A	F	S	C	S	S	L	L	S	M	H	T	G	E	K	P	4	3/4	+
Y	Q	C	G	E	C	G	K	A	F	N	Q	R	T	H	L	T	R	H	T	G	E	K	P	5	4/4	+
F	K	C	G	S	C	G	K	A	F	T	C	H	S	S	L	T	V	H	S	G	D	K	P	6	1/4	-
F	K	C	S	D	C	E	K	A	F	N	S	R	S	R	L	T	I	H	T	G	E	K	P	7	3/4	+
F	K	C	A	D	C	G	K	G	F	S	C	H	A	Y	L	L	V	H	S	G	E	K	P	8	2/4	-
F	K	C	N	E	C	G	K	A	F	S	S	H	A	Y	L	I	V	H	T	G	E	K	P	9	2/4	-
Y	K	C	S	Q	C	W	K	A	F	S	C	H	S	S	L	I	M	H	T	G	E	K	P	10	2/4	-
F	K	C	S	E	C	G	R	A	F	S	Q	N	H	C	L	I	K	H	S	G	E	K	S	11	3/4	+
L	A	C	E	K	C	G	E	M	F	N	W	S	S	H	L	T	E	H	S	E	G	K	P	12	4/4	+

Espaceur

Hélice  $\alpha$

Feuillet  $\beta 2$

Feuillet  $\beta 1$

doigt de zinc de ZNF74 seulement. Cette étude ne nous a pas permis d'isoler une séquence spécifique reconnue par ZNF74. Il est à noter que si le site consensus contenait une séquence reconnue par l'enzyme de restriction *Pst I*, notre méthode de sélection ne nous aurait pas permis d'identifier une telle séquence car les oligonucléotides sélectionnés ont été digérés avec *Pst I* avant le clonage (section 3.3 de *Matériel et Méthodes*). Cependant, les sites consensus reconnus par les facteurs de transcription sont rarement très stricts dans leurs séquences. Il est donc peu probable que l'identification de séquence(s) consensus ait été prévenue pour cette cause. Notre intention était d'utiliser par la suite une librairie dégénérée plus complexe de 24 pb (complexité  $4^{24}$ ). Cette étude n'a pas été entreprise en raison de la publication d'un article rapportant qu'il n'avait pas été possible d'isoler une séquence spécifique d'ADN reconnue par la protéine KRAB à multidoigt Kid<sub>1</sub> que nous utilisons notamment en parallèle avec ZNF74 dans nos essais de sélection (Elser *et al.*, 1997). Elser et collaborateurs (1997) rapportaient que des études de sélection de site utilisant des librairies d'oligonucléotides dégénérés de 15, 35 et 75 pb n'avaient pas permis d'isoler une séquence spécifique d'ADN reconnue par Kid<sub>1</sub> (il est à noter que toutes les séquences d'une banque de 35 et 75 pb ne peuvent être représentées lors de l'étape initiale de sélection compte tenu des quantités d'ADN de départ utilisées). D'autre part, d'autres groupes nous ont mentionné leurs essais infructueux lors d'expériences visant à isoler des séquences de liaison spécifique à l'ADN reconnues par des protéines KRAB à multidoigt par sélection d'oligonucléotides dégénérés ou de fragments d'ADN génomiques (E.J. Bellefroid, communication personnelle).

Même si à l'heure actuelle, la liaison spécifique à l'ADN double brin n'a pu être démontrée pour aucune protéine KRAB à multidoigt, il n'est pas possible de conclure que

ZNF74, Kid<sub>1</sub> ou d'autres protéines de cette famille n'ont pas une affinité particulière pour des séquences spécifiques d'ADN double brin à l'instar de la plupart des facteurs transcriptionnels. Plusieurs raisons pourraient avoir prévenu l'identification de telles séquences par les approches *in vitro*. On peut tout d'abord mentionner les limitations des études de sélections d'oligonucléotides dégénérés pour des protéines qui pourraient nécessiter une plate-forme relativement large (>24 nucléotides) pour une liaison à l'ADN dépendante de la séquence. En effet, il n'est pas possible de tester, compte tenu de leur complexité, des banques vraiment représentatives d'oligonucléotides dégénérés sur plus de 20-21 positions. L'utilisation de fragments de 200-300 pb d'ADN représentatifs des séquences présentes dans l'ADN génomique limite la complexité des banques mais rend plus difficile la recherche de consensus. D'autre part, il faut mentionner que la stabilité des interactions protéine/ADN peut être affectée par plusieurs facteurs comme la concentration ionique, la présence de magnésium et le pH. Les conditions optimales de liaison d'une protéine à l'ADN ne pouvant être déterminées sans connaître à priori les séquences cibles, on doit recourir à l'utilisation des tampons lors des étapes de liaison qui, quoiqu'éventuellement sub-optimaux, permettent habituellement la liaison de la majorité des facteurs transcriptionnels connus. Ces conditions peuvent cependant ne pas être favorables pour stabiliser certains complexes protéine/ADN. Par ailleurs, la majorité des études incluant la nôtre ont été réalisées en utilisant des protéines KRAB à multidoigt synthétisées dans des bactéries sous forme de protéines de fusion. D'une part, le repliement adéquat du large domaine multidoigt pourrait être inefficace dans des systèmes bactériens et d'autre part l'interaction des protéines multidoigt avec l'ADN pourrait dépendre de modifications post-traductionnelles uniquement possibles dans des systèmes eucaryotiques; la phosphorylation confère par exemple une meilleure liaison à

l'ADN au facteur de transcription Egr-1 (Huang *et al.*, 1994). L'utilisation de protéines purifiées à partir d'extraits de cellules eucaryotes pourrait donc être suggérée. Enfin, il n'est pas exclu que la liaison de ZNF74 à l'ADN puisse requérir la présence d'un cofacteur qui permettrait de stabiliser son interaction avec l'ADN. L'addition d'extraits nucléaires lors de la préparation des résines d'affinité pourrait permettre de reconstituer un potentiel complexe actif ZNF74-cofacteur c'est à dire capable de lier l'ADN. À l'heure actuelle, on ne peut éliminer la possibilité que ZNF74 puisse aussi réguler la transcription *via* son domaine KRAB non pas en tant que protéine se liant à l'ADN mais en tant que co-répresseur se liant à un facteur de transcription lié lui-même à l'ADN; ainsi par exemple WT1, une protéine à quatre motifs Cys<sub>2</sub>His<sub>2</sub>, peut fonctionner dans certaines circonstances comme un co-régulateur grâce à son association au récepteur nucléaire SF-1 lui-même lié à l'ADN (Nachtigal *et al.*, 1998).

Dans le futur, des approches *in vivo* pourraient être envisagées pour identifier des gènes cibles et par la suite les séquences d'ADN double brin (mais aussi éventuellement les séquences simple brin ou les structures spécifiques de l'ADN) auxquelles ZNF74 pourrait se lier. Deux approches sont envisageables soient: 1) l'immunopurification de complexes protéine/ADN natifs préfixés à la paraformaldéhyde et le clonage des séquences isolées par une méthode similaire à celle initialement décrite par Tomotsune *et al.* (1993) pour isoler les séquences de liaison à l'ADN de la protéine à homéobox hox C8 et 2) des études de co-transfection de ZNF74 avec un gène rapporteur sous le contrôle de promoteurs susceptibles d'être des cibles de ZNF74 car co-exprimés dans les mêmes tissus que ce dernier chez l'embryon (exemples des cibles possibles co-exprimées durant l'embryogenèse: Hox A3, dHAND, Pax 3, Shh, etc....).

## **2. LIAISON DE ZNF74 À L'ADN SIMPLE BRIN**

Par des études de retardement sur gel, nous avons mis en évidence que ZNF74 peut se lier à des oligonucléotides simple brin, double brin et hétéroduplex. Une liaison préférentielle à des séquences aléatoires de l'ADN simple brin est de plus observée. La plupart des facteurs transcriptionnels connus reconnaissent des séquences spécifiques d'ADN double brin mais n'ont pas d'affinité particulière pour l'ADN simple brin. L'affinité observable de ZNF74 pour l'ADN simple brin n'est sans doute pas surprenante considérant le fait que cette protéine a aussi une affinité pour l'ARN monocaténaire tel que démontré précédemment dans notre laboratoire (Grondin *et al.*, 1996). Il est à noter que Esler *et al.* (1997) ont rapporté récemment que Kid<sub>1</sub>, une autre protéine KRAB à motifs multidoigt liait préférentiellement des ADN hétéroduplex de façon indépendante de la séquence alors que la liaison à l'ADN simple brin semblait plus restreinte. De façon intrigante, nos résultats indiquent que ZNF74 aussi bien que Kid<sub>1</sub> semblent se lier plus fortement à l'ADN simple brin qu'à l'ADN hétéroduplex. Nous ne pouvons expliquer actuellement avec certitude l'origine de cette différence. Les conditions de liaison de l'ADN *in vitro* et de retardement sur gel étant très comparables, il est possible que cette différence soit attribuable au fait que nous ayons utilisé des protéines de fusion (MBP-ZNF74 et MBP-Kid<sub>1</sub>) différentes de celles du groupe de Esler (GST-Kid<sub>1</sub>). Quoiqu'il en soit, il est à noter que les conditions de liaison utilisées *in vitro* pouvant différer des conditions *in vivo*, on ne peut conclure actuellement de façon définitive sur le type de séquences ou de structures d'ADN reconnues préférentiellement par ZNF74 *in vivo*.

Au cours des dernières années, quelques protéines ont été caractérisées comme des facteurs transcriptionnels se liant plus spécifiquement à l'ADN simple brin. Une première protéine est PC4 ("Positive Cofacteur 4") (Werter *et al.*, 1998). PC4 peut

réprimer la transcription en se liant de façon indépendante de la séquence à l'ADN simple brin. Cette liaison survient au niveau de promoteurs ayant subi une séparation des deux brins d'ADN. Il est à noter que PC4 a aussi été décrit comme un co-activateur de la transcription dont la fonction de co-activation pourrait dépendre d'une interaction préalable avec l'ADN double brin de façon indépendante de la séquence. D'autre part, les protéines hnRNP K ("heterogeneous Nuclear RibonucleoProtein K") (Takimoto *et al.*, 1993, Siomi *et al.*, 1993, Tomonaga *et al.*, 1995, Siomi *et al.*, 1994) et FBP ("Far-upstream Binding Protein") (Duncan *et al.*, 1994, Duncan *et al.*, 1996, Bazar *et al.*, 1995, Michelotti *et al.*, 1996) ont été décrites comme des protéines capables d'activer la transcription du promoteur c-myc en se liant à des séquences spécifiques simple brin présentes sur ce promoteur. De façon intéressante, hnRNP K et FBP sont aussi capables de lier l'ARN et semblent impliquées dans la maturation de l'ARN (Siomi *et al.*, 1993, Siomi *et al.*, 1994, Grossman *et al.*, 1998). Il est actuellement possible de suggérer que ZNF74 puisse fonctionner comme ce nouveau type de régulateur de l'expression génique en se liant à l'ADN simple brin et à l'ARN. L'identification de séquences cibles *in vivo* de ZNF74 devrait permettre de mieux comprendre la signification fonctionnelle de son affinité pour l'ADN simple brin, double brin et hétéroduplex ainsi que pour l'ARN et de déterminer si la liaison aux acides nucléiques est dépendante d'une séquence ou d'une structure particulière.

### 3. CONCLUSION

En conclusion, ZNF74 représente un régulateur de l'expression génique complexe qui utilise un motif structural multifonctionnel, le motif multidoigt Cys<sub>2</sub>His<sub>2</sub>, ayant des propriétés de liaison à l'ADN (ce mémoire), l'ARN (Grondin *et al.*, 1996) et aux protéines

(Grondin *et al.*, 1997). Le domaine doigt de zinc est de plus responsable du ciblage au noyau cellulaire ainsi qu'à la matrice nucléaire (Grondin *et al.*, (1996) (1997)). La contribution respective de chacun des douze doigts de zinc de ZNF74 à ces interactions demeure à déterminer.

ZNF74 est une protéine de la famille KRAB Cys<sub>2</sub>His<sub>2</sub> qui pourrait jouer un rôle durant l'embryogenèse comme régulateur de la transcription tel que suggéré par la présence du motif KRAB de répression et du motif multidoigt ayant une affinité pour l'ADN et comme régulateur de la maturation de l'ARN tel que suggéré par l'interaction mise en évidence avec l'ARN et la forme hyperphosphorylée de l'ARN polymérase II qui joue un rôle dans la maturation de l'ARN (addition de la coiffe, épissage, clivage et addition de polyA). Des études récentes ont montré qu'un certain nombre de facteurs ayant un rôle transcriptionnel (bicoid, WT1, FBP) ou post-transcriptionnel (hnRNP K) initialement démontré auraient en fait un double rôle soit à la fois transcriptionnel et post-transcriptionnel (Dubneau et Struhl, 1996, Caricasole *et al.*, 1996, Grossman *et al.*, 1998, Tomonaga et Levens 1995). Il est possible que la dualité fonctionnelle de ces protéines favorise une association étroite entre la transcription et la maturation des ARNm et donc une meilleure intégration de ces deux mécanismes.

On estime actuellement qu'il y aurait de 100 à 300 protéines à motifs KRAB Cys<sub>2</sub>His<sub>2</sub> encodées par le génome des mammifères. Les résultats obtenus lors des études de ZNF74 devraient ouvrir des voies pour tenter de mieux comprendre la fonction des différents membres de cette grande famille qui partagent fort probablement des caractéristiques communes avec ZNF74.

## **V-RÉFÉRENCES**

Amman, A.J., Wara, D.W., Cowan, M.J., Barrett, D.J., Stiehm, E.R. (1982) The DiGeorge syndrome and the fetal alcohol syndrome. *Am J Dis Child*, **136**, 906-908.

Andreazzoli, M., De Lucchini, S., Costa, M., Barsacchi, G. (1993) RNA binding properties and evolutionary conservation of the *Xenopus* multifinger protein Xfin. *Nuc Acids Res*, **11**, 4218-4225.

Arranz, V., Harper, F., Florentin, Y., Puvion, E., Kress, M., Ernoult-Lange, M. (1997) Human and mouse MOK2 proteins are associated with nuclear ribonucleoprotein components and bind specifically to RNA and DNA through their zinc finger domains. *Mol Cell Biol*, **17**, 2116-2126.

Aubry, M., Marineau, C., Zhang, F.R., Zahed, L., Figlewicz, D., Delattre, O., Thomas, G., de Jong, P.J., Julien, J.P., Rouleau, G.A. (1992), Cloning of six new genes with zinc finger motifs mapping to short and long arms of human acrocentric chromosome 22 (p and q11.2). *Genomics*, **13**, 641-648.

Aubry, M., Demczuk, S., Desmaze, C., Aikem, M., Aurias, A., Julien, J.P., Rouleau, G.A. (1993) Isolation of zinc finger gene consistently deleted in DiGeorge syndrome. *Hum Mol Genet*, **2**, 1583-1587.

Bardwell, V.J., Treisman, R. (1994) The POZ domain: a conserved protein-protein interaction motif. *Genes Dev*, **8**, 1664-1677.

Baron, B.W., Stanger, R.R., Hume, E., Sadu, A., Mick, R., Kerchaert, J.-P., Deweindt, C., Bastard, C., Nucifora, G., Zeleznik-Le, N., McKeithan, T.W. (1995) BCL6 encodes a sequence-specific DNA-binding protein. *Genes, Chromosomes & Cancer*, **13**, 221-224.

Bazar, L., Meighen, D., Harris, V., Duncan, R., Levens, D., Avignan, M. (1995) Targeted melting and DNA binding of a DNA regulatory element by a transactivator of c-myc. *J Biol Chem*, **270**, 8241-8248.

Bedell, M.A., Jenkins, N.A., Copeland, N.G. (1996) Good genes in bad neighbourhoods. *Nature Genet*, **12**, 229-232.

Bellefroid, E.J., Poncelet, D.A., Lecocq, P.J., Relevant, O., Martial, J.A. (1991) The evolutionarily conserved Kruppel-associated box domain defines a subfamily of eukaryotic multifingered proteins. *Proc Natl Acad Sci USA*, **88**, 3608-3612.

Bockman, D.E., Kirby, M.L. (1984) Dependence of thymus development on derivatives of the neural crest. *Science*, **223**, 498-500.

Budarf, M.L., Collins, J., Gong, W.L., Roe, B., Wang, Z.L., Bailey, L.C., Sellinger, B., Michaud, D., Driscoll, D.A., Emanuel, B.S. (1995) Cloning a balanced translocation

associated with DiGeorge syndrome and identification of a disrupted gene. . *Nature Genet*, **10**, 269-278.

Carey, A.H., Kelly, D., Halford, S., Wadey, R., Wilson, D., Goodship, J., Burn, J., Paul, T., Sharkey, A., Dumanski, J., Nordenskjold, M., Williamson, R., Scambler, P.J. (1992) Molecular genetic study of the frequency of monosomy 22q11 in DiGeorge syndrome. *Am J Hum Genet*, **51**, 964-970.

Caricasole, A., Duarte, A., Larsson, S.H., Hastie, N.D., Little, M., Holmes, G., Todorov, I, Ward, A. (1996) RNA binding by the Wilms Tumor suppressor zinc finger proteins. *Proc Natl Acad Sci USA*, **93**, 7562-7566.

Carmo-Fonseca, M., Pepperkok, R., Carvalho, M.T., Lamond, A.I. (1992) Transcription-dependent colocalization of the U1, U2, U4/U6, and U5 snRNPs in coiled bodies. *J Cell Biol* **117**, 1-14.

Chesnut, J.D., Stephens, J.H., Dahmus, M.E. (1992) The interaction of RNA polymerase II with the adenovirus-2 major late promoter is precluded by phosphorylation of the C-terminal domain of subunit IIa. . *J Biol Chem*, **267**, 10500-10506.

Chisaka, O., Capecchi, M.R. (1991) Regionally restricted developmental defects resulting from targeted disruption of the mouse homeobox gene *hox-1.5*. *Nature*, **350**, 473-479.

Cho, E.J., Takagi, T., Moore, C.R., Buratowski, S. (1997) mRNA capping enzyme is recruited to the transcription complex by phosphorylation of the RNA polymerase II carboxy-terminal domain [see comments]. *Genes Dev*, **11**, 3319-3326.

Dahmus, M.E. (1996) Reversible phosphorylation of the C-terminal of RNA polymerase II. *J Biol Chem*, **271**, 19009-19012.

Driscoll, D.A., Budarf, M.L., Emanuel, B.S. (1992) A genetic etiology for DiGeorge syndrome: consistent deletions and microdeletions of 22q11. *Am J Hum Genet*, **50**, 924-933.

Driscoll, D.A., Spinner, N.B., Budarf, M.L., McDonald-McGinn, D.M., Zackai, E.H., Goldberg, R.B., Shprintzen, R.J., Saal, H.M., Zonana, J., Jones, M.C., *et al.* (1992) Deletions and microdeletions of 22q11.2 in velo-cardio-facial syndrome. *Am J Hum Genet*, **44**, 261-268.

Dubnau, J., Struhl, G. (1996) RNA recognition and translational regulation by a homeodomain protein. *Nature*, **379**, 694-699.

Dunbar, B.S. (1987) Two-dimensional electrophoresis and immunological techniques.

Plenum Press, New York.

Duncan, R., Bazar, L., Michelotti, G., Tomonaga, T., Krutzsch, H., Avignan, M., Levensl, D. (1994) A sequence-specific , single-strand binding protein activates the far upstream element of c-myc and defines a new DNA-binding motif. *Genes Dev*, **8**, 465-480.

Duncan, R., Collins, I., Tomonaga, T., Zhang, T. and Levens, D. A (1996) A unique transactivation sequence motif is found in the carboxyl-terminal domain of the single-strand binding protein FBP. *Mol Cell Biol* **16**, 2274-2282.

Elser, B., Kriz, W., Bonventre, J.V., Englert, C., Witzgall, R. (1997) The Kruppel-associated box (KRAB)-zinc finger protein Kid-1 and the Wilms' tumor protein WT1, two transcriptional repressor proteins, bind to heteroduplex DNA. *J Biol Chem*, **272**, 27908-27912.

Emanuel, B.S., Driscoll, D., Goldmuntz, E., Baldwin, S., Biegel, J., Zackai, E.H., McDonald-McGinn, D., Sellinger, B., Gorman, N., Williams, S., *et al.* (1993) Molecular and phenotypic analysis of the chromosom 22 microdeletion syndromes. *Prog Clin Biol Res*, **384**, 207-224.

Friedman, J.R., Fredericks, W.J., Jensen, D.E., Speicher, D.W., Huang, X.P., Neilson, E.G., Rauscher III, F.J. (1996) KAP-1, a novel corepressor for the highly conserved KRAB repression domain. *Genes Dev*, **10**, 2067-2978.

Funabiki, T., Kreider, B.L., Ihle J.N. (1994) The carboxyl domain of zinc fingers of the Evi-1 myeloid transforming gene binds a consensus sequence of GAAGATAGAG. *Oncogene*, **9**, 1575-1581.

Glover, T.W. (1995) Catching a break on 22. *Nature Genet*, **10**, 257-258.

Goldberg, R., Motzkin, B., Marion, R., Scambler, P.J., Shprintzen, R.J. (1993) Velo-cardio-facial syndrome: a review of 120 patients. *Am J Hum Genet*, **45**, 313-319.

Gong, W., Emanuel, B.S., Collins, J., Kim, D.H., Wang, Z., Chen, F., Zhang, G., Roe, B., Budarf, M.I. (1996) A transcription map of the DiGeorge and velo-cardio-facial syndrome minimal critical region on 22q11. *Hum Mol Genet*, **5**, 789-800.

Gong, W., Emanuel, B.S., Galili, N., Kim, D.H., Roe, B., Driscoll, D.A., Budarf, M.L. (1997) Structural and mutational analysis of a conserved gene (DGSI) from the minimal DiGeorge syndrome critical region. *Hum Mol Genet* **6**,:267-276.

Gong, Z., Hew, C.L. (1994) Zinc and DNA binding properties of a novel LIM homeodomain protein Isl-2. *Biochem*, **33**, 15149-15158.

Grondin, B. (1997) Études structurales et fonctionnelles du gène ZNF74; une protéine encodée par un gène candidat pour le syndrome de DiGeorge. Thèse de doctorat. Université de Montréal.

Grondin, B., Bazinet, M., Aubry, M. (1996) The KRAB zinc finger gene ZNF74 encodes an RNA binding protein tightly associated with the nuclear matrix. *J Biol Chem*, **271**, 15458-15467.

Grondin, B., Côté, F., Bazinet, M., Vincent, M., Aubry, M. (1997) Direct interaction of the KRAB/Cys<sub>2</sub>-His<sub>2</sub> zinc finger protein ZNF74 with a hyperphosphorylated form of the RNA polymerase II largest subunit. *J Biol Chem*, **272**, 27877-27885.

Grossman, J.S., Meyer, M.I., Wang, Y.C., Mulligan, G.J., Kobayashi, R., Helfman D.M. (1998) The use of antibodies to the polypyrimidine tract binding protein (PTB) to analyze the protein components that assemble on alternatively spliced pre-mRNAs that use distant branch points. *RNA*, **4**, 613-25.

Halford, S., Wadey, R., Roberts, C., Daw, S.C., Whiting, J.A., O'Donnell, H., Dunham, I., Bentley, D., Lindsay, E., Baldini, A., *et al.* (1993) Isolation of a putative transcriptional regulator from the region of 22q11 deleted in DiGeorge syndrome, Shprintzen syndrome and familial congenital heart disease. *Hum Mol Genet*, **2**, 2099-2107.

Hanna-Rose, W., Hansen, U. (1996) Active repression mechanisms of eukaryotic transcription repressors. *Trends Genet*, **12**, 229-243.

Harada, R., Bérubé, G., Tamplin, O.J., Denis-Larose, C. and Nepveu, A (1995) DNA-binding specificity of the cut repeats from the human cu-like protein. *Mol Cell Biol*, **15**, 129-140.

He, J., Furmanski, P. (1995) Sequence specificity and transcriptional activation in the binding of lactoferrin to DNA. *Nature*, **373**, 721-724.

Hoovers, J.M.N., Mannen, M., John, R., Blik, J., Van Heyningen, V., Proteous, D.J., Leschot, N.J., Westerveld, A., Little, P.F.R. (1992) High-resolution localization of 69 potential human zinc finger genes: a number are clustered. *Genomics*, **12**, 254-263.

Hozak, P., Hassan, A.B., Jackson, D.A., Cook, P.R. (1993) Visualization of replication factories attached to nucleoskeleton. *Cell*, **73**, 361-373.

Huang, R.P., Adamson, E.D. (1994) The phosphorylated forms of the transcription factor, Egr-1, bind to DNA more efficiently than non-phosphorylated. *BBRC*, **200**, 1271-1276.

Huebner, K., Druck, T., Croce, C.M., Thisen, H.J. (1991) Twenty-seven

nonoverlapping zinc finger cDNAs from human T cells map to nine different chromosomes with apparent clustering. *Am J Hum Genet*, **48**, 726-740.

Jackson, D.A., Hassan, A.B., Errington, R.J., Cook, P.R. (1993) Visualization of focal sites of transcription within human nuclei. *EMBO J*, **12**, 1059-1065.

Kelly, D., Goldberg, R., Wilson, D., Lindsay, E., Carey, A., Goodship, J., Burn, J., Cross, I., Shprintz, R.J., Scambler, P.J. (1993) Confirmation that the velo-cardio-facial syndrome is associated with haplo-insufficiency of genes at chromosome 22q11. *Am J Hum Genet*, **45**, 308-312.

Kim, S.-S., Chen, Y.-M., O'Leary, E., Witzgall, R., Vidal, M., Bonventre, J.V. (1996) A novel member of the RING finger family, KRIP-1, associates with the KRAB-A transcriptional repressor domain of zinc finger proteins. *Proc Natl Acad Sci USA*, **93**, 15299-15304.

Kinzler, K.W., Vogelstein, B. (1989) Whole genome PCR: application to the identification of sequences bound by gene regulatory proteins. *Nucleic Acids Research*, **17**, 3645-3653.

Kinzler, K.W., Vogelstein, B. (1990) The GLI gene encodes a nuclear protein which binds specific sequences in the human genome. *Mol Cell Biol*, **10**, 634-642.

Klocke, B., Koster, M., Hille, S., Bouwmeester, T., Bohm, S., Pieler, T., Knochel, W. (1994) The FAR domain defines a new *Xenopus laevis* zinc finger protein subfamily with specific RNA homopolymer binding activity. *Biochem Biophys Acta*, **1217**, 81-89.

Klug, A., Schwabe, J.W. (1995) Protein motifs 5. Zinc Fingers. *FASEB J*, **9**, 597-604.

Knochel, W., Potting, A., Koster, M., el Baradi, T., Nietfeld, W., Bouwmeester, T., Pieler, T. (1989) Evolutionary conserved modules associated with zinc fingers in *Xenopus laevis*. *Proc Natl Acad Sci USA*, **86**, 6097-6100.

Kurahashi, H., Tsuda, E., Kohama, R., Nakayama, T., Masuno, M., Imaizumi, K., Kamiya, T., Sano, T., Okada, S., Nishisho, I. (1997) Another critical region for deletion of 22q11: a study of 100 patients. *Am J Med Genet*, **72**, 180-185.

Ladomery M. (1997) Multifunctional proteins suggest connections between transcriptional and post-transcriptional processes. *Bioessays*, **19**, 903-909.

Lammer, E.J., Chen, D.T., Hoar, R.M., Agnish, N.D., Benke, P.J., Braun, J.T., Curry, C.I., Fernhoff, P.M., Griz, A.W.J., Lott, I.T., Trichard, J.M., Sun, S.C. (1985) Retinoic acid embryopathy. *New Engl J Med*, **313**, 837-841.

Lawrence, C.E., Altshul, S.F., Boguski, M.S., Liu, J.S., Neuwald, A.F., Wootton, J.C. (1993)

Detecting subtle sequence signals: a Gibbs sampling strategy for multiple alignment. *Science*, **262**, 208-14.

Li, J.Y., English, M.A., Ball, H.J., Yeyati, P.L., Waxman, S., Licht, J.D. (1997) Sequence-specific DNA binding and transcriptional regulation by the promyelocytic leukemia zinc finger protein. *J Biol Chem*, **272**, 22447-22455.

Liu, H.-X., Zhang, M., Krainer, A.R. (1998) Identification of functional exonic splicing enhancer motifs recognized by individual SR proteins. *Genes Dev*, **12**, 1998-2012.

Lu, H., Flores, O., Weinmann, R., Reinberg, D. (1991), The nonphosphorylated form of RNA polymerase II preferentially associated with the preinitiation complex. *Proc Natl Acad Sci USA*, **88**, 10004-10008.

Mai, B. and Breeden, L. (1997) Xbp1, a stress-induced transcriptional repressor of the *Saccharomyces cerevisiae* Swi4/Mbp1 family. *Mol Cell Biol*, **17**, 6491-6501.

Makela, T.P., Parvin, J.D., Kim, J., Huber, L.J., Sharp, P.A., Weinberg, R.A. (1995) A kinase-deficient transcription factor TFIIH is functional in basal and activated transcription. *Proc Natl Acad Sci USA*, **92**, 5174-5178.

Manley, N.R., Capecchi, M.R. (1995) The role of Hoxa-3 in mouse thymus and thyroid development. *Development*, **121**, 1989-2003.

Mavrothalassitis, G., Beal, G., Takis, S.P. (1990) Defining target sequences of DNA-binding proteins by random selection and PCR: determination of the GCN4 binding sequence repertoire. *DNA Cell Biol*, **9**, 783-788.

McCracken, S., Fong, N., Yankulov, K., Ballantyne, S., Pan, G., Greenblatt, J., Patterson, S.D., Wickens, M., Bentley, D.L. (1997) The C-terminal domain of RNA polymerase processing to transcription. *Nature*, **385**, 357-361.

Michelotti, G.A., Michelotti, E.F., Pullner, A., Duncan, R.C., Eick D., Levens, D. (1996) Multiple single-stranded *cis* elements are associated with activated chromatin of the human c-myc gene *in vivo*. *Mol Cell Biol*, **16**, 2656-2669.

Mitchell, P.J., Tjian, R. (1989) Transcriptional regulation in mammalian cells by sequence-specific DNA binding proteins. *Science*, **245**, 371-378.

Mo, R., Freer, A.M., Zinyk, D.L., Crackower, M.A., Michaud, J., Heng, H.H., Chik, K.W., Shi, X.M., Tsui, L.C., Cheng, S.H., Joyner, A.L., Hui, C. (1997) Specific and redundant functions of Gli2 and Gli3 zinc finger genes in skeletal patterning and development. *Development*, **124**, 113-123.

Moosmann, P., Georgiev, O., Le Douarin, B., Bourquin, J.P., Schaffner, W. (1996)

Transcriptional repression by RING finger protein TF1-Beta that interacts with the KRAB repressor domain of KOX1. *Nuc Acids Res*, **24**, 4859-4867.

Morris, J.F., Hromas, R., Rauscher III, F.J. (1994) Characterization of the DNA-binding properties of the myeloid zinc finger protein MZF1: Two independent DNA-binding domains recognize two DNA consensus sequences with a common G-rich core. *Mol Cell Biol*, **14**, 1786-1795.

Nachtigal, M.W., Hirokawa, Y., Enyeart-VanHouten, D.L., Flanagan, J.N., Hammer, G.D., Ingraham, H.A. (1998) Wilms' tumor 1 and Dax-1 modulate the orphan nuclear receptor SF-1 in sex-specific gene expression. *Cell*, **93**, 445-454.

Nakagama, H., Heinrich, G., Pelletier, J., Housman, D.E. (1995) Sequence and structural requirements for high-affinity DNA binding by the WT1 gene product. *Mol Cell Biol*, **15**, 1489-1498.

Numoto, M., Niwa, O., Kaplan, J., Wong, K.K., Kamiya, K., Yanagihara, K., Calame, K. (1993) Transcriptional repressor ZF5 identifies a new conserved domain in zinc finger proteins. *Nuc Acids Res*, **21**, 3767-3775.

O'Donnell, H., McKeown, C., Gould, C., Morrow, B., Scambler, P. (1997) Detection of an atypical 22q11 deletion that has no overlap with the DiGeorge syndrome critical region [letter]. *Am J Hum Genet*, **60**, 1544-1548.

Oliphant, A.R., Brandl, C.J., Struhl, K. (1989) Defining the sequence specificity of DNA-binding proteins by selecting sites from random-sequence oligonucleotides: Analysis of yeast GCN4 protein. *Mol Cell Biol*, **9**, 2944-2949.

Oster, G., Kilburn, K.H., Siegal, F.P. (1983) Chemically induced congenital thymic dysgenesis in the rat: a model of the DiGeorge syndrome. *Clin Immunol Immunopathol*, **28**, 128-134.

Passantini, C., Corbi, N., Paggi, M.G., Perez, M., Cotelli, F., Stefani, M., Amati, P. (1995) The product of Zfp59 (Mfg2), a mouse gene expressed at the spermatid stage of spermatogenesis, accumulates in spermatogenesis nuclei. *Cell growth and differentiation*, **6**, 1037-1044.

Pavletich, N.P. and Pabo, C.O. (1991). Zinc finger-DNA recognition: crystal structure of a Zif268-DNA complex at 2.1 Å. *Science* **252**, 809-817.

Payre, F., Vincent, A. (1991) Genomics targets of the serendipity beta and delta zinc finger proteins and their respective DNA recognition sites. *EMBO J*, **10**, 2533-2541.

Pelletier, J., Bruening, W., Li, F.P., Haber, D.A., Glaser, T., Housman, D.E. (1991) WT1 mutations contribute to abnormal genital system development and hereditary

Wilms' tumor. *Nature*, **353**, 431-444.

Pengue, G., Lania, L. (1996) Kruppel-associated box repression of RNA polymerase II is influenced by the arrangement of basal promoters elements. *Proc Natl Acad Sci USA*, **93**, 1015-1020.

Penman, S. (1995) Rethinking cell structure. *Proc Natl Acad Sci USA*, **92**, 5251-5257.

Perez, M., Rompato, G., Corbi, N., De Gregorio, L., Dragani TA., Passananti C. (1996) Zfp60, a mouse zinc finger gene expressed transiently during in vitro muscle differentiation. *FEBS Letters*, **387**, 117-121.

Pierrou, S., Enerback, S., Carlsson, P. (1995) Selection of high-affinity binding sites for sequence-specific, DNA binding proteins from random sequence oligonucleotides. *Anal Biochem*, **229**, 99-105.

Pollock, R., Treisman, R. (1990) A sensitive method for the determination of protein-DNA binding specificities. *Nuc Acids Res.*, **18**, 6197-6204.

Rauscher III, F.J., Morris, J.F., Tournay, O.E., Cook, D.M., Curran, T. (1990) Binding of the Wilm's tumor locus zinc finger protein to the EGR-1 consensus sequence. *Nature*, **250**, 1259-1262.

Sambrook, J., Fritsch, E.F., Maniatis, T. (1989) *Molecular Cloning: a laboratory manual* (second edition). Cold Spring Harbor Laboratory Press, USA.

Scambler, P.J., Carey, A.H., Wyse, R.K., Roach, S., Dumanski, J.P., Nordenskjold, M., Williamson, R. (1991) Microdeletions within 22q11 associated with sporadic and familial DiGeorge syndrome. *Genomics*, **10**, 201-206.

Serizawa, H., Conaway, J.W., Conaway, R.C. (1993) Phosphorylation of C-terminal domain of RNA polymerase II is not required in basal transcription. *Nature*, **363**, 371-374.

Sham, M.H., Vesque, C., Nonchev, S., Marshall, H., Frain, M., Gupta, R.D., Whiting, D., Charnay, P., Krumlauf, R. (1993) The zinc finger gene Krox20 regulates HoxB2 (Hox2.8) during hindbrain segmentation. *Cell*, **72**, 183-196.

Siomi, H., Choi, M., Siomi, M.C., Nussbaum, R.L., Dreyfuss, G. (1994) Essential role for KH domains in RNA binding: Impaired RNA binding by a mutation in the KH domain of FMR1 that causes Fragile X syndrome. *Cell*, **77**, 33-39.

Siomi, H., Matunis, M.J., Michael, W.M., Dreyfuss, G. (1993). The pre-mRNA binding K protein contains a novel evolutionarily conserved motif. *Nuc Acids Res*, **21**, 1193-1198.

- Spector, D.L., Fu, X.D., Maniatis, T. (1991) Associations between distinct pre-mRNA splicing components and the cell nucleus. *EMBO J*, **10**, 3467-3481.
- Steinmetz, E.J. (1997) Pre-mRNA processing and the CTD of RNA polymerase II: the tail that wags the dog. *Cell*, **89**, 491-494.
- Sukegawa, J., Blobel, G. (1993) A nuclear pore complex protein that contains zinc finger motifs, binds DNA, and faces the nucleoplasm. *Cell*, **72**, 29-38.
- Suzuki, M. (1993) Common features in DNA recognition helices of eukaryotic transcription factors. *EMBO J*, **12**, 3221-3226.
- Suzuki, M., Gerstein, M., Yagi, N. (1994) Stereochemical basis of DNA recognition by Zn fingers. *Nuc Acid Res*, **22**, 3397-3405
- Swirnoff, A.H., Milbrandt, J. (1995) DNA-binding specificity of NGF1-A and related zinc finger transcription factors. *Mol Cell Biol*, **15**, 2275-2287.
- Takimoto, M., Tomonaga, T., Matunis, M., Avignan, M., Krutzsch, H., Dreyfuss, G., Levens, D. (1993). Specific binding of heterogeneous ribonucleoprotein particle protein K to the human *c-myc* promoter, *in vitro*. *J Biol Chem*, **268**, 18249-18258.
- Tang, S., Breitman, M.L. (1995) The optimal sequence of the Hox11 protein contains a predicted recognition core motif. *Nuc Acids Res*, **23**, 1928-1935.
- Theunissen, O., Rudt, F., Guddat, U., Mentzel, H., Pieler, T. (1992) RNA and DNA binding zinc fingers in Xenopus TFIIIA. *Cell*, **71**, 679-690.
- Tomonaga, T., Levens, D. (1995). Heterogeneous nuclear ribonucleoprotein K is a DNA-binding transactivator. *J Bio. Chem*, **270**, 4875-4881
- Tomotsune, D., Shoji, H., Wakamatsu, Y., Kondoh, H., Takahashi, N. (1993) A mouse homologue of the drosophila tumor-suppressor gene 1 (2)gl controlled by Hox-C8 in vivo. *Nature*, **365**, 69-72.
- Treisman and Desplan (1989) The products of the Drosophila gap genes hunchback and Kruppel bind to the Hunchback promoters. *Nature*, **341**, 335-340.
- Tsai, R.Y.L., Reed, R.R. (1998) Identification of DNA recognition sequences and protein interaction domains of the multiple-Zn-finger protein Roaz. *Mol Cell Biol*, **18**, 6447-6456.
- Vortkamp, A., Gessler, M., Grzeschik, K.-H. (1995) Identification of optimized target sequences for the GLI3 zinc finger protein. *DNA and Cell Biol*, **14**, 629-634.

Wansink, D.G., Schul, W., van der Kraan, I., van Steensel, B., van Driel, R., de Jong, L. (1993) Fluorescent labeling of nascent RNA reveals transcription by RNA polymerase II in domains scattered throughout the nucleus. *J Cell Biol*, **122**, 283-293.

Weeks, J.R., Hardin, S.E., Shen, J., Lee, J.M. Greenleaf, A.L. (1993) Locus-specific variation in phosphorylation state of RNA polymerase II in vivo: correlations with gene activity and transcript processing. *Genes Dev*, **7**, 2329-2344.

Werten, S., Stelzer, G., Goppelt, A., Langen, F.M., Gros, P., Timmers, H.T.M., Vand der Vliet, P. C., Meisterernst, M. (1998) Interaction of PC4 with melted DNA inhibits transcription. *EMBO J*, **17**, 5103-5111.

Williams, A.J., Khachigian, L.M., Shows, T., Collins, T. (1995) Isolation and characterization of a novel zinc finger protein with transcription activity. *J Biol Chem*, **270**, 22143-22152.

Wilson, D.I., Burn, J., Scambler, P., Goodship, J. (1993) DiGeorge syndrome: part of CATCH22. *J Med Genet*, **30**, 852-856.

Wilson, D.I., Cross, I.E., Goodship, J.A., Coulhard, S., Carey, A.H., Scambler, P.J., Bain, H.H., Hunter, A.S., Carter, P.E., Burn, J. (1991) DiGeorge syndrome with isolated aortic coarctation and isolated ventricular septal defect in three sibs with 22q11 deletion of maternal origin. *Birt Heart J*, **66**, 308-312.

Wilson, D.I., Goodship, J.A., Burn, J., Cross, I.E., Scambler, P.J. (1992) Deletions within chromosome 22q11 in familial congenital heart disease. *Lancet*, **340**, 573-575.

Witzgall, R., O'Leary, E., Gessner, R., Ouellette, A.J., Bonventre, J.V. (1993) Kid-1, a putative renal transcription factor: regulation during ontogeny and in response to ischemia and toxic injury. *Mol Cell Biol*, **13**, 1933-1942.

Yuryev, A., Corden, J.L. (1996) Suppression Analysis reveals a functional difference between the serines in positions two and five in the consensus sequence of the C-terminal domain of yeast RNA polymerase II. *Genetics*, **143**, 661-671.

Zweidler-McKay, P.A., Grimes, H.L., Flubacher, M.M., Tschlis, P.N. (1996) Gfi-1 encodes a nuclear zinc finger protein that binds DNA and functions as a transcriptional repressor. *Mol Cell Biol*, **16**, 4024-4034.

## **VI-ANNEXE I**

**Figure 1. Sélection d'oligonucléotides double brin par MBP-Kid<sub>1</sub> immobilisée sur résine.** Séquences sélectionnées pour la protéine Kid<sub>1</sub> lors de la première (A) et de la deuxième expérience (B). Le brin b correspond au complémentaire inverse du brin a.

**B. Séquences sélectionnées par Kid1  
(deuxième expérience)**

<b>#1</b>	a: 5'- CAAACCCTCACATTCG -3' b: 5'- CGAATGTGAGGGTTTG -3'	a: 5'- CACAGCCCAATATCCCA -3' b: 5'- TGGAAATATGGCTGTG -3'
<b>#2</b>	a: 5'- TACTTTGTAGTTTTTCG -3' b: 5'- CGAAAACTACAAAAGTA -3'	a: 5'- GTGTCATTGTTTAAAC -3' b: 5'- GTTTAAACAATGACAC -3'
<b>#3</b>	a: 5'- CTTAATAAAAAAGAATA -3' b: 5'- TATTCCTTTTTTATTAAG -3'	a: 5'- ACAGGCGTAATCATGG -3' b: 5'- CCATGATTACGCCCTGT -3'
<b>#4</b>	a: 5'- CACTATAAACATTCCTC -3' b: 5'- GGGAATGTTTATAGTG -3'	a: 5'- CTAGTAGGAAGAACA -3' b: 5'- TTGTTCTTCCCTACTAG -3'
<b>#5</b>	a: 5'- CTTCCATCAGGCATCA -3' b: 5'- TGATGCCCTGATGGAAG -3'	a: 5'- TGCCCGTAAITGAGTG -3' b: 5'- CACTCAATACC GGCA -3'
<b>#6</b>	a: 5'- CTTTATTTTCTATTCA -3' b: 5'- TGAATAGAAAAATAAAG -3'	a: 5'- AACGTCATAAATAGAA -3' b: 5'- TTCATTTATGACGTT -3'
<b>#7</b>	a: 5'- CTATTCACATATCTTG -3' b: 5'- CAAGATAGTGGAAATAG -3'	a: 5'- TTACATATGAATCGAG -3' b: 5'- CTCGATTCATATGTAA -3'
<b>#8</b>	a: 5'- CTCTTTATATATCTAT -3' b: 5'- ATAGATATATAAAGAG -3'	a: 5'- TCGAAACCCTCAAITTA -3' b: 5'- TAATTGAGGGTTTCGA -3'
<b>#9</b>	a: 5'- CTTTTACCTATTTTTA -3' b: 5'- TAAAAATAGGTAAAAAG -3'	a: 5'- TCGTAAGGACGTCAAA -3' b: 5'- TTTGACGTCCTTACGA -3'
<b>#10</b>	a: 5'- CGATTCITTAATACCAT -3' b: 5'- ATGGTATTAAGAATCG -3'	a: 5'- GCTGTGGGATTATATG -3' b: 5'- CATATAATCCCACAGC -3'
<b>#11</b>	a: 5'- CCCTCTTAGGCCTACA -3' b: 5'- TGTAGGCCCTAAGAGGG -3'	
<b>#12</b>	a: 5'- GGATGGTATGAGTTGT -3' b: 5'- ACAACTCATACCATCC -3'	

**A. Séquences sélectionnées par Kid1  
(première expérience)**

<b>#1</b>	a: 5'- TGCACCTCACTTAACAA -3' b: 5'- TTGTTAAGTGAGTGCA -3'	a: 5'- CACAGCCCAATATCCCA -3' b: 5'- TGGAAATATGGCTGTG -3'
<b>#2</b>	a: 5'- GATTTGAGATATTGGT -3' b: 5'- ACCAATATCTCAATC -3'	a: 5'- GTGTCATTGTTTAAAC -3' b: 5'- GTTTAAACAATGACAC -3'
<b>#3</b>	a: 5'- TAAAAACATAGGGTGA -3' b: 5'- TCACCCTATGTTTTTA -3'	a: 5'- ACAGGCGTAATCATGG -3' b: 5'- CCATGATTACGCCCTGT -3'
<b>#4</b>	a: 5'- TTCGTGATAATGTGAA -3' b: 5'- TTCACATTATCAGGAA -3'	a: 5'- CTAGTAGGAAGAACA -3' b: 5'- TTGTTCTTCCCTACTAG -3'
<b>#5</b>	a: 5'- TTATGAGTGGATTCCG -3' b: 5'- GCGAATCCACTCATAA -3'	a: 5'- TGCCCGTAAITGAGTG -3' b: 5'- CACTCAATACC GGCA -3'
<b>#6</b>	a: 5'- CAATTCACCTGATATC -3' b: 5'- GATATCAGGTGAATTG -3'	a: 5'- AACGTCATAAATAGAA -3' b: 5'- TTCATTTATGACGTT -3'
<b>#7</b>	a: 5'- GCATTCACCATTCGGT -3' b: 5'- ACCGAATGGTGAATGC -3'	a: 5'- TTACATATGAATCGAG -3' b: 5'- CTCGATTCATATGTAA -3'
<b>#8</b>	a: 5'- GTCCTATCTCTAACAT -3' b: 5'- ATGTTAGAGATAGGAC -3'	a: 5'- TCGAAACCCTCAAITTA -3' b: 5'- TAATTGAGGGTTTCGA -3'
<b>#9</b>	a: 5'- CGTGTATCCATTGTTA -3' b: 5'- TAACAATGGATACACG -3'	a: 5'- TCGTAAGGACGTCAAA -3' b: 5'- TTTGACGTCCTTACGA -3'
<b>#10</b>	a: 5'- CTCCTTGATTGGGAT -3' b: 5'- ATCCCAATACAAGGAG -3'	a: 5'- GCTGTGGGATTATATG -3' b: 5'- CATATAATCCCACAGC -3'

**Figure 2. Sélection d'oligonucléotides double brin par MBP immobilisée sur résine.**  
Séquences sélectionnées pour la protéine MBP lors de la première (A) et de la deuxième expérience (B). Le brin b correspond au complémentaire inverse du brin a.

**A. Séquences sélectionnées par MBP  
(première expérience)**

#1	a: 5'- GAATCAAGATTTCTTT -3' b: 5'- AAAGAAATCTTGATTC -3'
#2	a: 5'- CTAGGACACAGTTTCC -3' b: 5'- GGAAACTGTGTCCTAG -3'
#3	a: 5'- GTAAATAACATACGAC -3' b: 5'- GTCGTATGTTATTTAC -3'
#4	a: 5'- TCTTGGATTGAGTCCC -3' b: 5'- GGGACTCAATCCAAGA -3'
#5	a: 5'- GTTCACCATTGCCATC -3' b: 5'- GATGGCAATGGTGAAC -3'
#6	a: 5'- ATACACTAGCATAGTA -3' b: 5'- TACTATGCTAGTGTAT -3'
#7	a: 5'- ATTCATCCAAGTATGA -3' b: 5'- TCATACTTGATGAAT -3'
#8	a: 5'- GCTTACAGGCGTGATA -3' b: 5'- TATCACGCCTGTAAGC -3'
#9	a: 5'- CGTGTATCCATTGTTA -3' b: 5'- TAACAATGGATACACG -3'
#10	a: 5'- TTACCAAATTAATATC -3' b: 5'- GATATTAATTTGGTAA -3'
#11	a: 5'- ATTGCTGAGTCAACTG -3' b: 5'- CAGTTGACTCAGCAAT -3'
#12	a: 5'- ATTGTTAAATAAATCG -3' b: 5'- CGATTTATTTAACAAT -3'
#13	a: 5'- AACTTGGAACACGGGA -3' b: 5'- TCCCGTGTTCCAAGTT -3'
#14	a: 5'- ACCAGATGGATATCGC -3' b: 5'- GCGATATCCATCTGGT -3'
#15	a: 5'- CAATCGGATGTTTGGC -3' b: 5'- GCCAAACATCCGATTG -3'
#16	a: 5'- GCTATGAACCTTTCAA -3' b: 5'- TTGAAAGTTCATAGC -3'
#17	a: 5'- ACTCTTTAGTTGGTCT -3' b: 5'- AGACCAACTAAAGAGT -3'

**B. Séquences sélectionnées par MBP  
(deuxième expérience)**

#1	a: 5'- CGCGGCAGTTGCAGTA -3' b: 5'- TACTGCAACTGCCGCG -3'
#2	a: 5'- TTTTAAATTGATAACC -3' b: 5'- GGTTATCAATTTAAAA -3'
#3	a: 5'- CAGTGTATATACCAAC -3' b: 5'- GTTGGTATATACACTG -3'
#4	a: 5'- TCGTCGTAATAAGTTG -3' b: 5'- CAACTTATTACGACGA -3'
#5	a: 5'- CCATCAGGTTGATATG -3' b: 5'- CATATCAACCTGATGG -3'
#6	a: 5'- CTTTATTTATATTAAT -3' b: 5'- ATTAATATAAATAAAG -3'
#7	a: 5'- GCCTCAATTCCTTGA -3' b: 5'- TCAAGGGAATTGAGGC -3'
#8	a: 5'- ATCGTAAACAGCTTTA -3' b: 5'- TAAAGCTGTTTACGAT -3'
#9	a: 5'- CATGTAGTATGAATAT -3' b: 5'- ATATTCATACTACATG -3'
#10	a: 5'- CTATAACTCATGTTTT -3' b: 5'- AAAACATGAGTTATAG -3'
#11	a: 5'- TACCTGGAAAGGAGAA -3' b: 5'- TTCTCCTTCCAGGTA -3'
#12	a: 5'- GCCTCAATTCCTTGA -3' b: 5'- TCAAGGGAATTGAGGC -3'
#13	a: 5'- ATATTATTATATTTTA -3' b: 5'- TAAAATATAATAATAT -3'
#14	a: 5'- CTAATTTACTGATAAT -3' b: 5'- ATTATCAGTAAATTAG -3'
#15	a: 5'- ATTTTGAAATTTTCATT -3' b: 5'- AATGAAATTTCAAAT -3'
#16	a: 5'- CAACAGGAATCTTCAG -3' b: 5'- CTGAAGATTCCTGTTG -3'
#17	a: 5'- CTCAAAAATTTAACGG -3' b: 5'- CCGTTAAATTTTTGAG -3'
#18	a: 5'- CGCGGCAGTTGCAGTA -3' b: 5'- TACTGCAACTGCCGCG -3'
#19	a: 5'- ATAGTTCCATTTTTTC -3' b: 5'- GAAAAAATGGAACAT -3'

## **VII-ANNEXE II**

**Figure 1. Fichier d'entrée du programme Gibbs.** Ce fichier présente les séquences à analyser par le programme Gibbs. Les séquences ont été entrées à l'aide d'un programme de traitement de texte (pico).

```
>R1 #1 strand +:
TTATCCAGATGGGTGG
>R1 #1 strand -:
CCACCCATCTGGATAA
>R1 #2 strand +:
AGCTTATACATTTATA
>R1 #2 strand -:
TATAAATGTATAAGCT
>R1 #3 strand +:
GTATATACACATCTCA
>R1 #3 strand -:
TGAGATGTGTATATAC
>R1 #4 strand +:
GTAGGATGGGATTA
>R1 #4 strand -:
TTAATCCCATCCTAC
>R1 #5 strand +:
ACGGCTTACATCTTC
>R1 #5 strand -:
GAAAGATGTAAGCCGT
>R1 #6 strand +:
TGGGTCTGCACGATTA
>R1 #6 strand -:
TAATCGTGCAGACCCA
>R1 #7 strand +:
GGTAGGCATTGTCTCG
>R1 #7 strand -:
CGAGACAATGCCTACC
>R1 #8 strand +:
TGGACGCACGCGTAGA
>R1 #8 strand -:
TCTACGCGTGCGTCCA
>R1 #9 strand +:
CGATTCAAAGGGATTA
>R1 #9 strand -:
TAATCCCTTTGAATCG
>R1 #10 strand +:
CAGTAACGCTTTTTCA
>R1 #10 strand -:
```

```
TGAAAAAGCGTTACTG
>R1 #11 strand +:
AACGAATTGGATGGTA
>R1 #11 strand -:
TACCATCCAATTCGTT
>R1 #12 strand +:
GCTCGAGAATCAGCAA
>R1 #12 strand -:
TTGCTGATTCTGCAGC
```

**Figure 2. Fichier de sortie du programme Gibbs.** Ce fichier présente les séquences alignées pour un consensus de taille prédéterminée de 6 bases.

```

terre{guevrem}110: gibbs melanier1.n16 -n
How many types of elements? 1
For element 1...
    How long are the sites? 6
    Do all seqs have the same # of sites (Y/N)? y
    How many sites in each seq? 1
1 sites for each

#1   R1 #1 strand +:
#2   R1 #1 strand -:
#3   R1 #2 strand +:
#4   R1 #2 strand -:
#5   R1 #3 strand +:
#6   R1 #3 strand -:
#7   R1 #4 strand +:
#8   R1 #4 strand -:
#9   R1 #5 strand +:
#10  R1 #5 strand -:
#11  R1 #6 strand +:
#12  R1 #6 strand -:
#13  R1 #7 strand +:
#14  R1 #7 strand -:
#15  R1 #8 strand +:
#16  R1 #8 strand -:
#17  R1 #9 strand +:
#18  R1 #9 strand -:
#19  R1 #10 strand +:
#20  R1 #10 strand -:
#21  R1 #11 strand +:
#22  R1 #11 strand -:
#23  R1 #12 strand +:
#24  R1 #12 strand -:

running gibbs...
nconv = 500
** 1 **
** 2 **
** 3 **
[-] 2.25367 motif A cycle 6
** 4 **
** 5 **
** 6 **
** 7 **

```

```

[-] 0.4044 motif A cycle 3
** 8 **
[-] 11.118 motif A cycle 7
** 9 **
[+] 0.316337 motif A cycle 15
** 10 **
20

```

## MOTIF A

```

1-1      9      ttatccag ATGGGT gg          14
2-1      3          cc ACCCAT ctggataa      8
3-1      6          agctt ATACAT ttata        11
4-1      6          tataa ATGTAT aagct        11
5-1      7          gtatat ACACAT ctca         12
6-1      5          tgag  ATGTGT atatac        10
7-1      6          gtagg ATGGGA ttaaa        11
8-1      5          ttta  ATCCCA tcctac        10
9-1      1          ACGGCT tacatctttc      6
10-1     11  gaaagatgta AGCCGT          16
11-1     6          tgggt CTGCAC gatta         11
12-1    10  taatcgtgc AGACCC a             15
13-1     4          ggt  AGGCAT tgtctcg        9
14-1     8          cgagaca ATGCCT acc          13
15-1     4          tgg  ACGCAC gcgtaga        9
16-1     4          tct  ACGCGT gcgtcca        9
17-1     9          cgattcaa AGGGAT ta           14
18-1     3          ta  ATCCCT ttgaatcg        8
19-1     5          cagt  AACGCT ttttca        10
20-1     6          tgaaa AAGCGT tactg         11
21-1     2          a  ACGAAT tggatggta       7
22-1     5          tacc  ATCCAA ttcgtt        10
23-1     2          g  CTCGAG aatcagcaa       7
24-1    10  ttgctgatt CTGCAG c             15

```

5

```

POS      A  C  G  T
  1     90 15  2  3
  2     11 27 19 53
  3     15 31 61  3
  4       7 65 27 11
  5     53 27 27  3
  6     15 15 11 70
non-
site    29 21 21 29

```

```

Complete log-likelihood ratio = 97 bits
Missing position information = 61 bits
Log-likelihood ratio statistic = 36 bits
Degrees of freedom = 18

```

Information per parameter = 2 bits

Best Sites (std. dev. above mean)

1-1	9	ttatccag	ATGGGT	gg	14	(2.22)
2-1	3	cc	ACCCAT	ctggataa	8	(2.09)
3-1	6	agctt	ATACAT	ttata	11	(3.03)
4-1	6	tataa	ATGTAT	aagct	11	(1.96)
5-1	7	gtatat	ACACAT	ctca	12	(1.56)
6-1	5	tgag	ATGTGT	atatac	10	(1.67)
7-1	6	gtagg	ATGGGA	ttaa	11	(2.50)
8-1	5	ttta	ATCCCA	tcctac	10	(2.23)
9-1	1		ACGGCT	tacatctttc	6	(2.24)
10-1	11	gaaagatgta	AGCCGT		16	(1.73)
11-1	6	tggt	CTGCAC	gatta	11	(2.51)
12-1	10	taatcgtgc	AGACCC	a	15	(2.18)
13-1	4	ggt	AGGCAT	tgtctcg	9	(4.13)
14-1	8	cgagaca	ATGCCT	acc	13	(4.15)
15-1	8	tgacgc	ACGCGT	aga	13	(2.74)
16-1	4	tct	ACGCGT	gcgtcca	9	(3.12)
17-1	9	cgattcaa	AGGGAT	ta	14	(2.43)
18-1	3	ta	ATCCCT	ttgaatcg	8	(3.68)
19-1	6	cagta	ACGCTT	ttca	11	(1.94)
20-1	6	tgaaa	AAGCGT	tactg	11	(3.37)
21-1	2	a	ACGAAT	tggatggta	7	(1.85)
22-1	5	tacc	ATCCAA	ttcgtt	10	(2.89)
23-1	11	gctcgagaat	CAGCAA		16	(1.61)
24-1	10	ttgctgatt	CTGCAG	c	15	(3.88)

5

Total log-likelihood ratio = 36 bits

seed: 456653125

time: 1 seconds (0.02 minutes)