



UNIVERSITÉ
DE NAMUR

University of Namur

Institutional Repository - Research Portal Dépôt Institutionnel - Portail de la Recherche

researchportal.unamur.be

THESIS / THÈSE

MASTER EN BIOCHIMIE ET BIOLOGIE MOLÉCULAIRE ET CELLULAIRE

Mise au point d'une méthode sans a priori d'identification par spectrométrie de masse des protéines interagissant avec une séquence oligonucléotidique

Michel, Sébastien

Award date:
2009

[Link to publication](#)

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal ?

Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.



**FACULTES UNIVERSITAIRES NOTRE-DAME DE LA PAIX
NAMUR**

Faculté des Sciences

**MISE AU POINT D'UNE METHODE SANS *A PRIORI* D'IDENTIFICATION PAR
SPECTROMETRIE DE MASSE DES PROTEINES INTERAGISSANT AVEC UNE
SEQUENCE OLIGONUCLEOTIDIQUE**

**Mémoire présenté pour l'obtention
du grade académique de master en biochimie et biologie moléculaire et cellulaire**

Sébastien MICHEL

Janvier 2009

Facultés Universitaires Notre-Dame de la Paix
FACULTE DES SCIENCES
Secrétariat du Département de Biologie
Rue de Bruxelles 61 - 5000 NAMUR
Téléphone: + 32(0)81.72.44.18 - Téléfax: + 32(0)81.72.44.20
E-mail: joelle.jonet@fundp.ac.be - <http://www.fundp.ac.be/fundp.html>

Mise au point d'une méthode sans *a priori* d'identification par spectrométrie de masse des protéines interagissant avec une séquence oligonucléotidique

MICHEL Sébastien

Résumé

L'étude des interactions entre protéines et acides nucléiques représente un passage nécessaire vers la compréhension des mécanismes de régulation de l'expression génique. Alors que les techniques classiques comme l'EMSA, le footprinting ou encore le CHIP sont des outils couramment utilisés, ils représentent des approches avec *a priori* qui nécessitent la connaissance d'au moins un des deux partenaires étudiés (facteur de transcription ou site de liaison). C'est la raison pour laquelle nous développons au laboratoire une technique d'identification sans *a priori*, par spectrométrie de masse, des protéines interagissant avec une séquence oligonucléotidique d'intérêt. Alors que le principe général de la méthode a déjà été établi grâce à l'utilisation d'une séquence bien décrite, le premier objectif de ce travail est d'appliquer cette méthode à une séquence encore peu caractérisée du HIV-1, un enhancer intragénique qui pourrait permettre de mieux comprendre la régulation des étapes précoces de la transcription du provirus. Après diverses mises au point, dont la diminution de la complexité des échantillons par fractionnement chromatographique, cette méthode a permis d'identifier une centaine de protéines capturées par cette séquence, parmi lesquelles plusieurs protéines connues pour appartenir à des complexes de remodelage de la chromatine, comme INO80, mais également des facteurs de transcription décrits pour les recruter. Le second objectif est d'évaluer en parallèle par ELISA-like la capacité de liaison à cette séquence de 3 trois facteurs de transcription susceptibles de participer à l'inductibilité de cette région régulatrice. Les résultats obtenus par cette approche suggèrent notamment une éventuelle coopération de NFAT et AP-1 au niveau d'un site composite.

Mémoire de master en biochimie et biologie moléculaire et cellulaire

Janvier 2009

Promoteur: P. Renard

Remerciements

Et nous voici déjà à la fin de 10 mois de mémoire et l'occasion de remercier toutes les personnes qui ont pu contribuer de près ou de loin à son bon déroulement.

Tout d'abord, merci au professeur Raes de m'avoir accueilli à l'URBC.

Je souhaite ensuite remercier ma promotrice, Patsy Renard, pour m'avoir si bien suivi durant ces 10 mois, pour la relecture en profondeur et critique de ce manuscrit jusque bien tard le soir.

Je tiens également à remercier Aurélie pour m'avoir aidé durant ce mémoire, pour l'apprentissage des techniques, d'avoir toujours été de bonne humeur et toujours disponible malgré les TP. Ce fut un plaisir de travailler avec toi, en espérant que les mises au point de la technique avancent à grands pas.

Je tiens d'ailleurs à remercier tout particulièrement 244,13 fois Marc pour son aide précieuse et toutes les informations données sur la spectrométrie de masse. Durant ces 10 mois, j'ai pu constater des similitudes entre le personnage et la machine : ça ne travaille pas souvent mais quand ça marche c'est efficace !! ;-)

Merci aux gens du labo pour tous les bons moments passés en leur compagnie et plus particulièrement à Lio, Guillaume, Emilie, Cindy, Celine, et les autres...

Je ne pourrais oublier les 5 affreux (et affreuses !) du bureau : Alex toujours zen, NicNic la blonde locale toujours de bonne humeur, Casp3 aux blots artistiques hors norme, l'homme qui jouait du UKULELE (t'inquiètes l'euro millions et les palmiers c'est pour bientôt) et la tête de pioche nationale avec qui on a pu avoir de bon fou rire.

Merci aussi aux autres potos de bio et plus particulièrement à Thomas, Arnaud, Caro, Mike.

Merci aux gens que j'ai sûrement oubliés faut pas m'en vouloir c'est la fatigue (si si)

Et pour finir, merci à mes parents qui m'ont permis de faire ces études et m'ont toujours supporté (pas évident ...), ainsi qu'à Celine et Gauthier pour leur contribution !!

Liste des abréviations

ADN	Acide désoxyribonucléique
ADN-PK	DNA-dependent protein kinase catalytic subunit
AP1	Activator protein 1
AR	Androgen receptor
ARID	ARID domain-containing protein
ARN	Acide ribonucléique
ARNm	ARN messenger
ARNP	ARN polymérase ADN dépendante
ARNr	ARN ribosomique
ARNt	ARN de transfert
Arp	Actin-related protein
ARRE	Antigen receptor response element
ATF	Activating transcription factor
ATP	Adénosine triphosphate
BHC	BRAF-HDAC complex
BRE	TFIIB recognition element
bZIP	Leucine zippers
CBP	CREB binding protein
CDK	Cyclin dependent kinase
ChIP	Chromatin immunoprecipitation
CIP	Cytokine induced protein
CK	Casein Kinase
CKAP	Cytoskeleton-associated protein
Cn	Calcineurine
COMM	Copper metabolism gene MURR1
coREST	REST corepressor
CRAC	Calcium release-activated Ca ²⁺
CRE	C-AMP responsive elements
CREB	CRE binding protein
CTD	Carboxy-terminal domain
DAG	Diacylglycérol
Dam	DNA adénine methyltransferase
DCE	Downstream core element
DPE	Downstream promoter element
DRAP	DR1-associated protein
EBV	Epstein-Barr Virus
Ef	Eukaryotic translation elongation factor
EGF	Epidermal growth factor
EMSA	Electromobility shift assay
ER	Estrogen receptor
ERK	Extracellular signal-regulated protein kinase

FRET	Fluorescence resonance energy transfer
FRK	Fyn-Related Kinase
FT	Facteur de transcription
GFP	Green fluorescence protein
GIF	General initiation factor
GM-CSF	Granulocyte Colony Stimulating Factor
GR	Glucocorticoid receptor
GSK	Glycogen synthase kinase
HAT	Histone acétyltransferase
HCF	Host cell factor
HDAC	Histone déacetylase
HDM	Histome Demethylase
HIV-1	Human immunodeficiency virus 1
HMT	Histone méthyltransférase
hnRNP	Heterogeneous nuclear ribonucleoproteins
HPLC-SCX	High performance liquid chromatography
HSV	Herpes simplex virus
HTH	Hélice-tour-hélice
HTLV	Human T-lymphotropic virus
IAP	Intracisternal A-particle
ICAT	Isotope-coded affinity tag
IFN	Interferon
I κ B	Inhibitors of κ B
IL-1b	Interleukine 1b
IN	Intégrase
Inr	Elément initiateur
JNK	Jun N-terminal Kinase
KH	K homologous
KIF	Kinesin superfamily member
LTR	Long terminal repeat
MAPK	Mitogen-activated protein kinase
MHC	Major histocompatibility complex
MTE	Motif ten element
NC	Negative co-factor
NES	Nuclear export sequence
NFAT	Nuclear factor of activated T cells
NFAT	Nuclear factor of activated T cells
NF- κ B	Nuclear factor κ B
NFR κ B	Nuclear factor related to κ B binding protein
NLS	Nuclear localisation sequence
NLS	Nuclear localizing sequence
NuRD	Nucleosome remodeling and DNA methylation

ORC	Origin recognition complex
PARP	Poly [ADP-ribose] polymerase
pb	Paire de base
PCR	Polymerase chain reaction
PI3	Phosphatidylinositol 3 phosphate
PIC	Complexe de préinitiation ou Complexe de préintégration
PIP2	Phosphatidylinositol biphosphate
PKC	Protein kinase C
PLC	Phospholipase C
PMA	Phorbol 12-myristate-13-acetate
PR	Protéase
P-TEFb	Positive Transcription Elongation Factor b
RBP	Retinoblastoma binding protein
RE	Réticulum endoplasmique
RFX	Regulatory factor X
RHD	Rel homology domain
RT	Reverse transcriptase
RTC	Rétrotranscription complex
SIE	Sis inductible element
SRE	Serum responsive element
SRF	Serum responsive factor
STAT	Signal Transducers and Activator of Transcription
SV40	Virus Simien 40
TAF	TATA associated factors
TBP	TATA binding protein
TCF	Ternary complex factor
TK	Thymidine kinase
TLK	Tousled-like kinase
TNF	Tumor necrosis factor
TPA	Phorbol 12-O-tetradecanoate-13-acetate
TRE	TPA responsive element
VRK	Vacciniarelated Kinase

Table des matières

I.	INTRODUCTION.....	10
I.1.	LA TRANSCRIPTION CHEZ LES CELLULES EUCARYOTES.....	11
I.1.1.	<i>Description générale.....</i>	11
I.1.2.	<i>Initiation de la transcription : description et régulation.....</i>	11
I.1.3.	<i>Structure et fonction des facteurs de transcription.....</i>	13
I.2.	LES MOYENS D'ETUDE DES INTERACTIONS ADN-PROTEINES.....	14
I.2.1.	<i>Approche avec a priori.....</i>	14
I.2.1.1.	L'immunoprécipitation chromatiniene.....	15
I.2.1.2.	L'empreinte aux nucléases.....	15
I.2.1.3.	Retard sur gel et méthodes dérivées.....	16
I.2.1.4.	Prédiction <i>in silico</i>	17
I.2.2.	<i>Analyse sans a priori des interactions ADN-protéines.....</i>	18
I.2.2.1.	Intérêt et avancées actuelles d'une approche sans <i>a priori</i>	18
I.2.2.2.	Principe de la méthode développée.....	20
I.3.	LE VIRUS DE L'IMMUNODEFICIENCE ACQUISE (HIV-1).....	21
I.3.1.	<i>Structure.....</i>	21
I.3.2.	<i>Cycle réplcatif du HIV-1.....</i>	21
I.3.3.	<i>Identification et caractérisation d'une région d'hypersensibilité au sein du gène pol 23</i>	
I.4.	DESCRIPTION DES FACTEURS DE TRANSCRIPTION ETUDIÉS.....	24
I.4.1.	<i>NF-κB.....</i>	24
I.4.2.	<i>AP-1.....</i>	25
I.4.3.	<i>NFAT.....</i>	26
I.5.	OBJECTIFS.....	28
II.	MATERIELS ET METHODES.....	29
II.1.	EXTRACTIONS DES PROTEINES NUCLEAIRES.....	30
II.1.1.	<i>Description des types cellulaires utilisés.....</i>	30
II.1.2.	<i>Culture cellulaire.....</i>	30
II.1.2.1.	Matériel.....	30
II.1.2.2.	Méthode.....	30
II.1.3.	<i>Stimulation cellulaire.....</i>	31
II.1.3.1.	Matériel.....	31
II.1.3.2.	Méthode.....	31
II.1.4.	<i>Extraction nucléaire.....</i>	31
II.1.4.1.	Principe.....	31
II.1.4.2.	Matériel.....	31
II.1.4.3.	Méthode.....	32
II.1.4.3.1.	Cellules adhérentes :.....	32

II.1.4.3.2.	Cellules en suspension :	32
II.1.5.	Dosage par la méthode de Bradford (Bradford 1976)	33
II.2.	AMPLIFICATION PCR	33
II.2.1.	Description des séquences utilisées	33
II.2.2.	Méthode	34
II.3.	PURIFICATION DES PRODUITS PCR	34
II.3.1.	Méthode	34
II.4.	DOSAGE DE LA LIAISON A L'ADN PAR ELISA-LIKE (TRANSAM)	35
II.4.1.	Principe	35
II.4.2.	Matériel	35
II.4.3.	Méthode	35
II.5.	PURIFICATION SUR BILLE	35
II.5.1.	Principe	35
II.5.2.	Matériel	35
II.5.3.	Méthode	36
II.6.	SEPARATION DES ECHANTILLONS PAR HPLC-SCX	37
II.7.	ANALYSE EN SPECTROMETRIE DE MASSE EN TANDEM	37
II.7.1.	Principe de la spectrométrie de masse	37
II.7.2.	Matériel	39
II.7.3.	Méthode	39
II.7.4.	Analyse des résultats	39
III.	RESULTATS ET DISCUSSION	40
III.1.	ETUDE DE LA LIAISON DE FACTEURS DE TRANSCRIPTION SUSCEPTIBLES D'EXPLIQUER L'INDUCTIBILITÉ DE LA RÉGION INTRAGÉNÉRIQUE DU HIV-1 : NF- κ B, NFAT ET AP-1	41
III.1.1.	NF- κ B	41
III.1.2.	NFAT	43
III.1.3.	AP-1	44
III.1.4.	Etude de la coopération entre AP-1 et NFAT	46
III.2.	ETUDE SANS A PRIORI DES PROTÉINES LIANT LA RÉGION INTRAGÉNÉRIQUE DU HIV-1	49
III.2.1.	Résultats préliminaires	49
III.2.2.	Mises au point	51
III.2.2.1.	Amélioration de la digestion trypsique	51
III.2.2.2.	Élimination de la biotine	51
III.2.2.3.	Diminution de la complexité des échantillons	51
III.2.2.4.	Etude des protéines liant la région intragénérique du HIV-1 après amélioration de la technique	53

III.2.2.5.	Protéines identifiées dans les conditions stimulées et non stimulées.....	53
III.2.2.6.	Protéines identifiées uniquement dans les conditions non stimulées.....	54
III.2.2.7.	Protéines identifiées uniquement dans les extraits nucléaires provenant de cellules stimulées par le PMA+ionomycine.....	55

IV. CONCLUSIONS ET PERSPECTIVES.....60

IV.1.	ETUDE DE LA LIAISON DE NF- κ B, AP-1 ET NFAT PAR ELISA-LIKE.....	61
-------	---	----

IV.2.	RECHERCHE SANS A PRIORI DE FACTEURS INTERAGISSANT AVEC LA SEQUENCE INTRA1	63
-------	---	----

V. BIBLIOGRAPHIE.....67

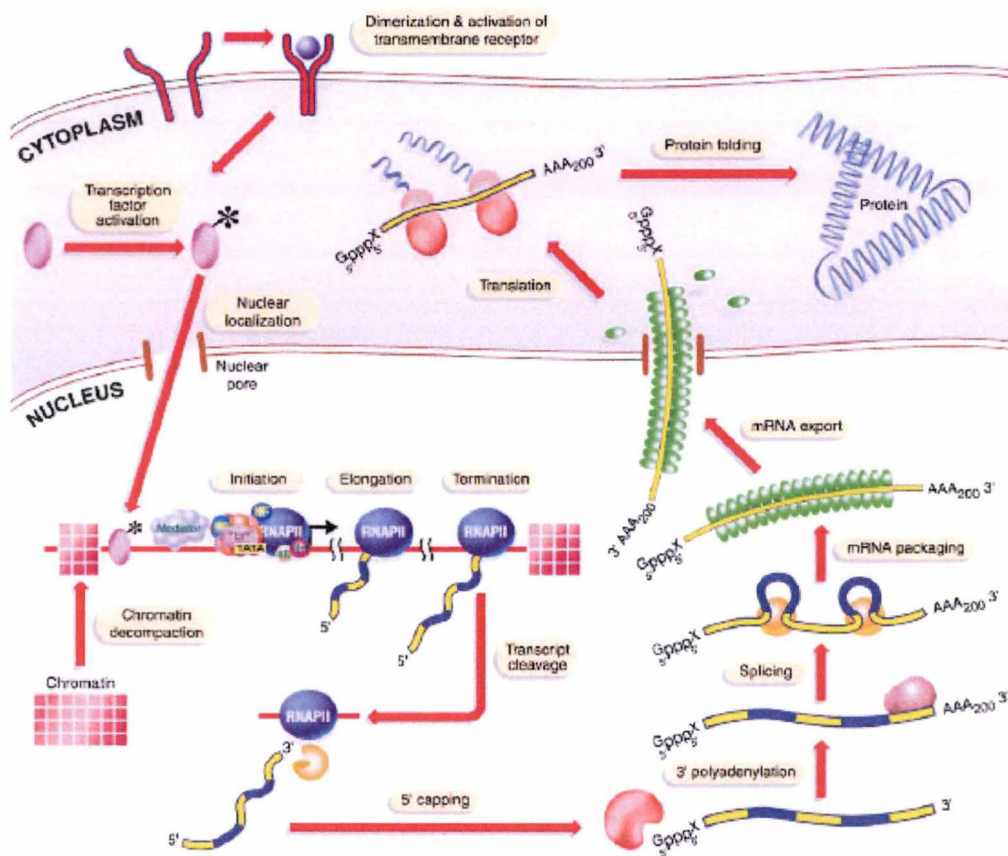


Figure 1.1 : Représentation des différentes étapes menant de l'expression des gènes à la synthèse des protéines. (Orphanides and Reinberg 2002)

I. Introduction

I.1. *La transcription chez les cellules eucaryotes*

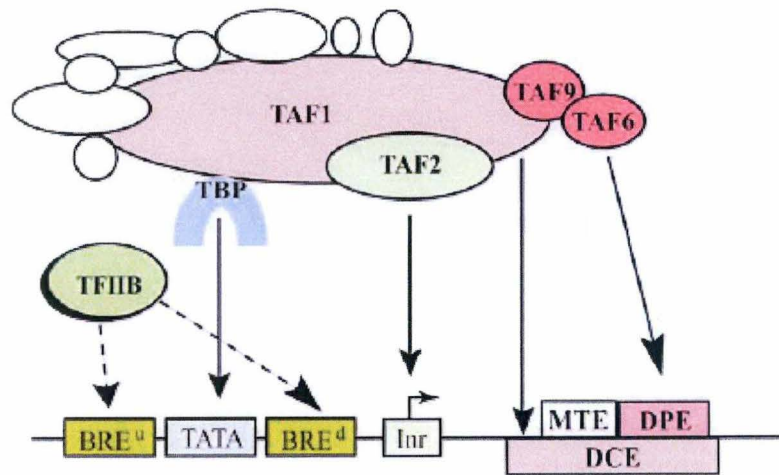
Chaque cellule composant le corps humain contient le même génome. Et pourtant, avec une estimation de 210 types cellulaires différents chez l'homme, comment est-il dès lors possible que deux cellules puissent avoir des caractéristiques morphologiques et fonctionnelles aussi différentes ? Si le séquençage du génome humain a permis d'estimer qu'approximativement 20000 à 25000 gènes codent des protéines, ils ne sont pas systématiquement tous exprimés. Au contraire, seule une partie d'entre eux est spécifiquement exprimée selon le type cellulaire mais aussi en fonction de nombreux stimuli provenant de l'environnement, faisant appel à des systèmes de régulation complexes. Le dogme central introduit par Watson et Crick a longtemps été décrit comme un processus multi-étapes (Fig.1). Cependant, de plus en plus de preuves sont en faveur d'un mécanisme continu où les différentes étapes sont couplées et régulées par des acteurs communs (Orphanides and Reinberg 2002). Lors de ce mémoire, nous nous focaliserons majoritairement sur les événements précoces de régulation de la transcription et plus précisément sur les moyens de les étudier.

I.1.1. Description générale

La transcription peut être définie comme le processus par lequel l'information contenue dans une séquence nucléique va être utilisée pour la synthèse d'un ARNm. Trois étapes peuvent être distinguées : l'initiation, l'élongation et la terminaison. La première étape consiste à lever l'action répressive de la chromatine, importante structure dans la régulation de l'expression génique comme nous le verrons par la suite, pour ensuite pouvoir positionner correctement une ARN polymérase ADN-dépendante (ARN P) au site d'initiation et débiter la transcription. Trois classes d'ARN Ps ont été décrites chez les cellules animales : l'ARNP I, II et III, responsables respectivement, de la synthèse des ARNs ribosomiques (ARNr), des ARNs messagers (ARNm) et des ARNs de transfert (ARNt) en plus de petits ARNs (Roeder and Rutter 1970). Peu après l'initiation, une coiffe va être ajoutée à l'extrémité 5' de l'ARN pré-messager qui aura pour fonction de protéger l'ARN naissant, de participer à l'exportation du transcrit mûri et de favoriser la transition vers l'élongation. Durant cette étape, l'ARNP II va se déplacer de l'extrémité 5' vers l'extrémité 3' de la séquence nucléique afin de transcrire le gène et continuer la synthèse de l'ARN pré-messager ayant commencé à l'étape précédente. Les régions non codantes (introns) intercalées au milieu des zones codantes (exons) sont épissées pour former l'ARNm. Finalement, une fois la fin du gène atteinte, l'ARN polymérase II se détache de l'ADN, l'ARNm est clivé, additionné d'une queue polyA à l'extrémité 3', après quoi il est exporté dans le cytoplasme, par le biais de pores nucléaires, pour être traduit. Chacune de ces étapes est soumise à une régulation hautement contrôlée par des facteurs pouvant être communs à différentes phases de la transcription (pour une revue voir (Orphanides and Reinberg 2002)). Cette fois encore, nous réduirons notre champ de vision à l'étape d'initiation dont le principe et la régulation vont être décrits plus en détail.

I.1.2. Initiation de la transcription : description et régulation

Deux tâches sont à accomplir lors de l'initiation : lever la répression émanant de la structure chromatinienne afin de rendre l'ADN accessible à la machinerie de transcription et le positionnement de l'ARNP II au site de démarrage. La chromatine est une forme condensée de l'ADN dont l'unité de base est le nucléosome. Celui-ci est constitué d'un octamère d'histone autour duquel l'ADN s'enroule deux fois, correspondant approximativement à 146



Core Promoter Element	Position	Consensus Sequence (5' to 3')	Bound Protein
BRE ^u	-38 to -32	(G/C)(G/C)(G/A)CGCC	TFIIB
TATA	-31 to -24	TATA(A/T)A(A/T)(A/G)	TBP
BRE ^d	-23 to -17	(G/A)T(T/G)A(T/G)(G/T)(T/G)(T/G)	TFIIB
Inr	-2 to +5	PyPyAN(T/A)PyPy	TAF1/TAF2
MTE	+18 to +29	C(G/C)A(A/G)C(G/C)(G/C)AACG(G/C)	n.a.
DPE	+28 to +34	(A/G)G(A/T)CGTG	TAF6/TAF9
DCE	3 subelements +6 to +11 +16 to +21 +30 to +34	core sequence: S _I CTTC S _{II} CTGT S _{III} AGC	TAF1

Figure 1.2 : Positionnement relatif par rapport au site d'initiation de la transcription et séquence de quelques-uns des « cores » promoteurs pouvant être présents chez les eucaryotes.

L'ensemble de ces sites n'est jamais présent simultanément dans le promoteur d'un gène. (Thomas and Chiang 2006)

Table 1.1 : Symbole correspondant aux nucléotides dégénérés selon le code IUPAC (<http://www.chem.qmul.ac.uk/iubmb/misc/naseq.html>)

R	Purine (G ou A)
Y	Pyrimidine (T ou C)
W	Adénine ou thymine
S	Guanine ou cytosine
M	Adénine ou cytosine
K	Guanine ou thymine
H	Adénine, thymine ou cytosine
B	Guanine, cytosine ou thymine
V	Guanine, adénine ou cytosine
D	Guanine, adénine ou thymine
N	Guanine, adénine, thymine ou cytosine

paires de bases (pb). Le complexe d'histones est composé de deux fois quatre sous unités : H2A, H2B, H3 et H4, protéines chargées positivement et fortement conservées entre les espèces, accentuant l'idée que ces dernières jouent un rôle capital dans la structure de l'ADN.

L'initiation de la transcription a lieu au niveau du promoteur, situé en amont du site de démarrage de la transcription. Cette zone de régulation peut être divisée en trois parties : le promoteur « basal », la partie proximale et la distale, toutes trois cruciales dans le contrôle de la transcription.

Le promoteur basal peut être défini comme la région minimale indispensable pour la liaison de la machinerie de la transcription et existe grâce à différentes séquences (Fig. 1.2). La plus caractérisée est la boîte TATA située 25 pb en amont du site de démarrage et ayant pour consensus la séquence : TATAWAAR (les nucléotides dégénérés sont désignés selon le code IUPAC <http://www.chem.qmul.ac.uk/iubmb/misc/naseq.html> et repris dans la table 1.1). L'élément DPE (downstream promoter element, consensus : RGWYV) est un autre exemple de séquence consensus classiquement retrouvée dans les promoteurs ne contenant pas de boîte TATA. L'élément initiateur (Inr, consensus : YYANWYY) est quant à lui situé à proximité du site +1 et présent dans des promoteurs contenant une boîte TATA ou non. D'autres séquences sont également souvent présentes dans le promoteur de base (« core promoter ») telles que le motif MTE (motif ten element, consensus : CSARCSSAACGS), l'élément BRE (TFIIB recognition element, consensus : SSRGCC), le DCE (downstream core element) ou encore les îlots CpG (pour une revue sur ces éléments voir (Smale and Kadonaga 2003)). L'initiation de la transcription liée à ces différentes séquences peut se faire selon deux méthodes : elle peut être concentrée sur quelques nucléotides ou dispersée parmi de multiples sites sur une région plus large (50 à 150 nt). La première stratégie est majoritairement retrouvée au niveau des séquences TATA, Inr ou DPE alors que la seconde est davantage associée aux îlots CpG (Juvén-Gershon, Hsu et al. 2006).

Le promoteur proximal est situé entre 50 et 200 pb en amont du site +1 et contient des sites de liaison pour des régulateurs transcriptionnels. Pour finir, le promoteur distal est constitué de séquences situées à longue distance aussi bien en amont qu'en aval du site d'initiation. Les éléments régulateurs retrouvés à ce niveau sont de type enhancer ou silencer, favorisant et réprimant, respectivement, la transcription. Bien qu'ils soient éloignés au niveau de la séquence nucléotidique du site d'initiation, la courbure de l'ADN peut les rapprocher suffisamment dans l'espace afin qu'ils puissent agir sur la régulation de la transcription (Kadonaga 2004).

Les contacts préliminaires menant à l'activation de l'expression des gènes ont lieu entre des activateurs transcriptionnels, ou facteurs de transcription (FT), et la structure chromatinienne. Cette première phase engendre un remodelage et une décondensation de l'ADN, permettant le recrutement, la liaison à l'ADN de la machinerie de transcription et des facteurs qui lui sont associés (pour une revue voir (Thomas and Chiang 2006)), formant le complexe de préinitiation (PIC, Fig. 1.3). L'assemblage de ce complexe représente une étape cruciale de la régulation de la transcription. Afin de former un PIC fonctionnel, des facteurs généraux d'initiation (GIFs) sont requis pour favoriser le bon placement et l'activité de l'ARNP II. Le premier GIF à lier le core promoteur est le facteur TFIID, qui, dans le cas d'une boîte TATA, va interagir avec l'ADN par une de ses sous-unités conservées, la TBP (TATA binding protein). Les autres sous-unités de ce facteur général sont les TAFs (TATA associated factors) qui participent aussi à la liaison à l'ADN. L'arrivée du second facteur, TFIIB (pour une revue sur ce facteur voir (Deng and Roberts 2007)), forme une plate-forme reconnue par la polymérase qui s'y lie, suivie de l'arrivée des facteurs TFIIF, TFIIH et TFIIE. Une fois le PIC fonctionnel formé, et en présence de dNTPs, les deux brins de l'ADN se

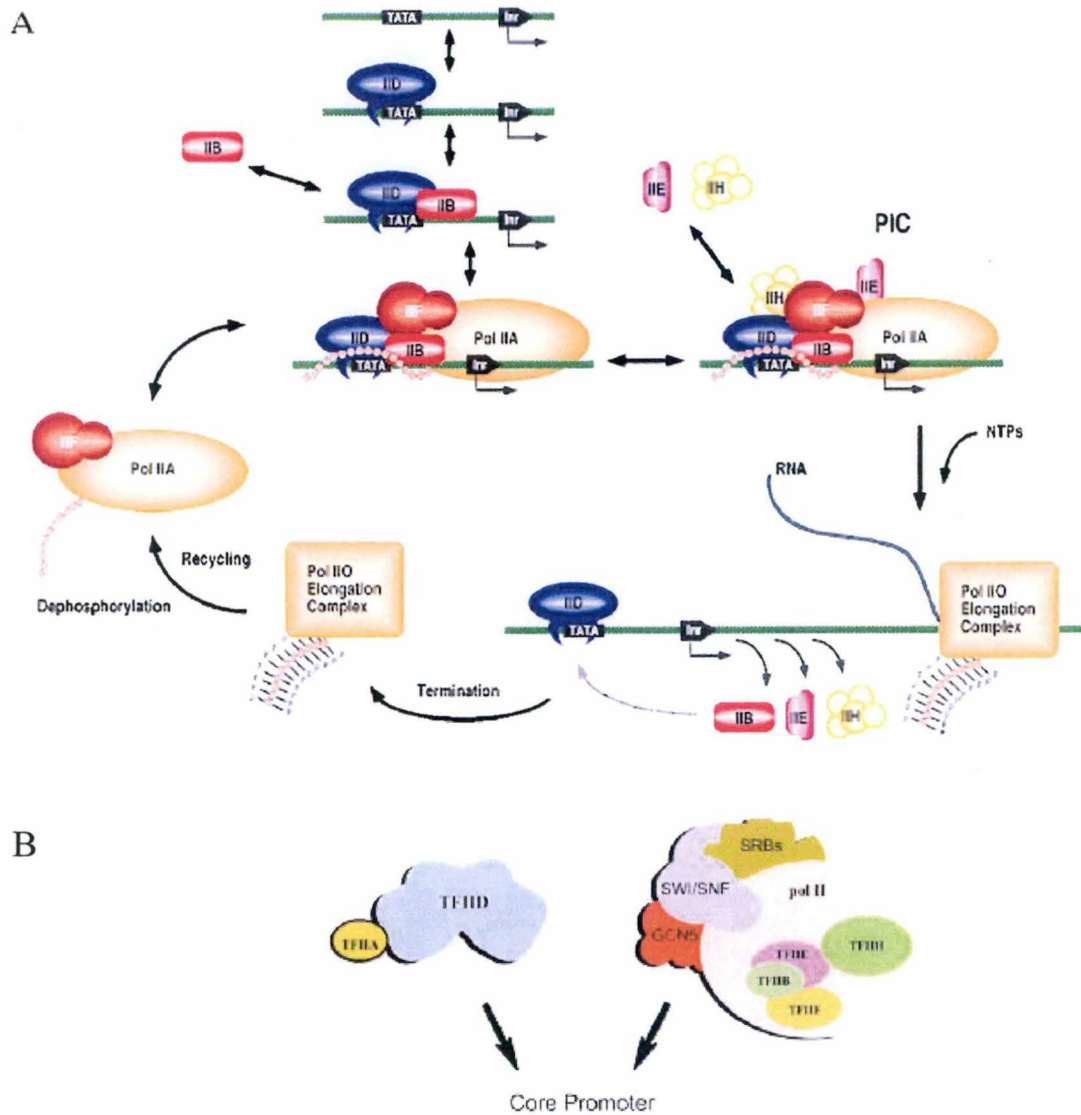


Figure 1.3 : Représentation des différentes étapes de la formation du PIC

(A) Formation du PIC selon le modèle multi-étapes. Les facteurs généraux sont recrutés de manière ordonnée (TFIID, TFIIB, recrutement de la polymérase, TFIIF, TFIIH, TFIIE) afin de placer l'ARN PII au site d'initiation. (Nikolov and Burley 1997).

(B) Un second modèle est également proposé, postulant le recrutement d'un complexe préformé comme cela est décrit pour l'holoenzyme chez les procaryotes. Cependant, ce modèle est controversé car le facteur TFIID n'a pas pu être purifié avec le reste du complexe, nécessitant au minimum deux étapes lors de la formation du PIC (Thomas and Chiang 2006).

séparent au site d'initiation et le domaine C-terminal de la large sous-unité de l'ARNP II est phosphorylé grâce à l'activité kinase du facteur TFIIH. Ce dernier est composé de 9 sous-unités de masse allant de 39 à 89 kDa et outre sa fonction de kinase, il possède également une activité hélicase ATP-dépendante et une activité ATPase ADN-dépendante (Nikolov and Burley 1997). Le contrôle de l'expression génique et la régulation de l'activité de la l'ARN PII se fait également par l'intermédiaire d'interactions avec d'autres protéines comme le médiateur (Lewis and Reinberg 2003), capable d'interagir avec la polymérase et de réguler la transcription d'un point de vue structurel et temporel.

I.1.3. Structure et fonction des facteurs de transcription

Les facteurs de transcription jouent un rôle crucial dans le contrôle de l'expression génique en faisant le lien entre l'information contenue dans la séquence nucléique, les signaux environnementaux et la machinerie de transcription.

En général, leur structure peut être disséquée en trois modules : un domaine de liaison à l'ADN (pouvant également servir à la formation de dimères), un domaine trans-activateur et un régulateur. La présence de ces domaines a été mise en évidence grâce à une étude chez la levure du facteur Gal4. Une protéine de fusion liant Gal4 à lexA, une protéine bactérienne de liaison à l'ADN, a permis de montrer que l'activité transcriptionnelle était indépendante de la capacité de liaison à l'ADN. Différentes structures de ces domaines de liaison à l'ADN existent mais quatre d'entre elles sont souvent représentées : l'hélice-tour-hélice (HTH), le motif en doigt de zinc, la glissière à leucine (bZIP) et l'hélice-boucle-hélice. D'autres domaines existent comme les homéodomains, formes variantes des HTHs.

Généralement, une séquence consensus est définie pour chaque famille de facteurs de transcription, c'est-à-dire une suite de nucléotides conservée entre les différentes espèces, qui est reconnue et liée par la protéine. Ces séquences peuvent être plus ou moins dégénérées, affectant l'affinité et la stabilité de liaison du facteur de transcription à l'ADN. Trois types de contacts peuvent avoir lieu entre les facteurs de transcription et les acides nucléiques : les ponts hydrogènes, les interactions hydrophobes et électrostatiques (Fig. 1.4). Si un point de contact individuel d'un facteur de transcription à son site est relativement faible, l'ensemble des liens formés avec les acides nucléiques permet de renforcer considérablement et de stabiliser l'interaction de la protéine avec sa séquence-cible (Yang 1998). Contrairement au domaine de liaison, il est par contre plus difficile de définir des conformations consensuelles pour les domaines trans-activateurs. Certains sont connus tels que des régions riches en résidus acides, des motifs riche en glutamine ou en proline et des feuilletts β hydrophobes. (Kadonaga 2004)

De manière générale, l'activation d'un facteur de transcription est le point d'arrivée d'une cascade de transduction initiée suite à l'activation d'un récepteur. De nombreux facteurs de transcription sont présents dans la cellule sous une forme latente, permettant une réaction rapide en cas de stimulus favorisant leur activation. Trois types de régulations majeures existent : la régulation de leur localisation spatiale (séquestration dans le cytoplasme par exemple), les modifications post-traductionnelles influençant la structure et/ou certaines caractéristiques des facteurs de transcription, et la liaison d'un ligand. Un exemple bien connu du contrôle spatial sera décrit par après avec NF- κ B (nuclear factor κ B), séquestré constitutivement dans le cytoplasme et dirigé vers le noyau en réponse à une large gamme de stimuli. Des modifications post-traductionnelles peuvent aussi réguler la présence des certains facteurs de transcription dans le noyau, comme c'est le cas par exemple de NFAT (Nuclear Factor of Activated T cells). La déphosphorylation de ce facteur permet de découvrir une séquence NLS (nuclear localisation sequence) d'adressage nucléaire suite à une modification

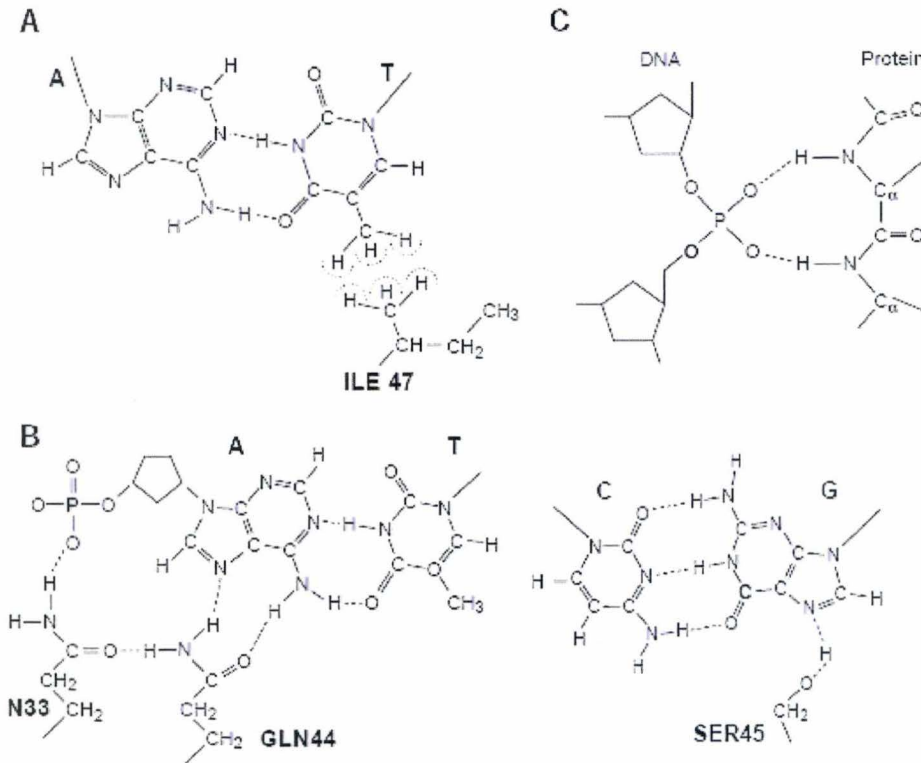


Figure 1.4 : Représentation des certaines interactions pouvant exister entre les acides nucléiques et acides aminés : (A) Liaison de Van der Waals et (B) (C) ponts hydrogène (Carey and Smale, 1999).

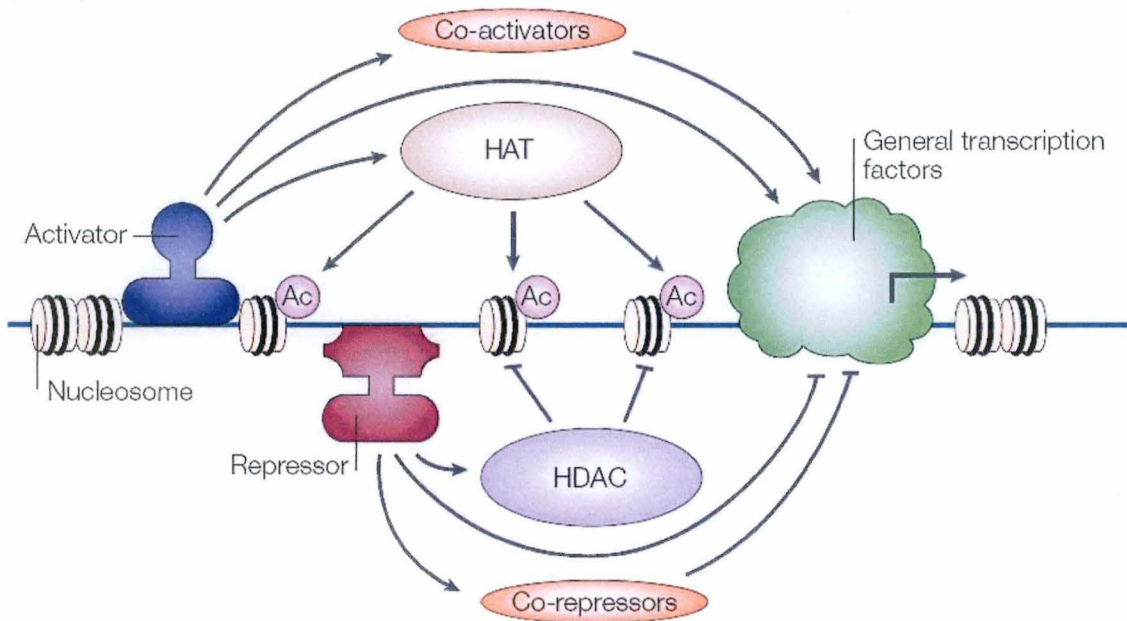


Figure 1.5 : Régulation de l'activité transcriptionnelle d'un gène

Diverses voies de régulation sont employées par les activateurs et les répresseurs transcriptionnels afin de contrôler l'expression génique. En bref, ils peuvent agir de trois manières : par interaction directe avec la machinerie de transcription, par recrutement de co-régulateurs ou en interagissant avec des complexes de remodelage de la chromatine. Le niveau final de l'activité transcriptionnelle résulte de la contribution (positive comme négative) apportée par chacun des régulateurs liés au promoteur du gène transcrit (Muratani and Tansey 2003).

de conformation. Deuxièmement, les modifications post-traductionnelles peuvent influencer certaines propriétés comme l'affinité de certains facteurs de transcription pour leur site de liaison ou encore le contrôle de leur capacité à induire la transcription (activité trans-activatrice). Enfin, le troisième type de régulation des facteurs de transcription repose sur la liaison ou le détachement d'un ligand, comme c'est le cas pour la famille des récepteurs nucléaires (comme le GR (Glucocorticoid Receptor) ou l'ER (Estrogen Receptor)). Pour une revue sur les récepteurs nucléaires voir (McKenna and O'Malley 2002).

Les FTs participent à la régulation de la transcription (Fig. 1.5) en contrôlant l'accès de la machinerie au site d'initiation soit par interaction directe avec les facteurs généraux de la transcription, soit par remodelage de la chromatine et modification des histones. En effet, ils sont capables de recruter des co-régulateurs de deux types : des co-activateurs ou co-répresseurs. Ces derniers peuvent être divisés selon deux classes : les complexes de remodelage ATP-dépendant et les complexes protéiques catalysant les modifications post-traductionnelles au niveau du domaine N-terminal de la queue des histones. Ces derniers sont au nombre de quatre : les HATs (histone acétyl transférases), les HDACs (histone déacetylases), les HMTs (histone méthyle transférases) et pour finir des kinases spécifiques des histones. Cette interaction entre les sous-unités de complexes protéiques altérant la structure de la chromatine permet une décondensation progressive de la chromatine et un accès au site d'initiation pour les facteurs généraux et l'ARNP II, afin de constituer le PIC. Toutefois, si la décondensation de l'ADN nécessite la présence de facteurs de transcription, la chromatine ne peut pas se trouver sous une forme trop condensée afin d'être accessible à ces derniers. Quel est l'événement initiateur : la liaison d'un facteur de transcription ou une diminution préalable du degré de condensation de la chromatine ? Il semblerait que la réponse se trouve dans les deux possibilités. En effet, certains facteurs de transcription seraient capable de reconnaître leur séquence même lorsque la chromatine est condensée, notamment grâce à un positionnement stratégique du facteur de transcription comme, par exemple, dans le grand sillon de la double hélice. Dès lors, ils peuvent recruter des agents modifiant la chromatine, rendant cette dernière accessible à de nouveaux facteurs qui, eux-mêmes, recruteront d'autres complexes protéiques afin d'augmenter progressivement l'accessibilité de l'ADN à la machinerie de transcription (Orphanides and Reinberg 2002).

Si on pensait initialement qu'un facteur de transcription permettait de réguler spécifiquement un seul gène, il est très vite apparu clairement que le schéma de régulation de l'expression génique était bien plus complexe. En effet, avec une estimation de 20000 à 25000 gènes dans le génome humain, un facteur de transcription seul n'est que peu spécifique pour un promoteur donné. La force de la régulation de la transcription provient de l'action coordonnée d'un ensemble de facteurs de transcription dont la composition forme un système hautement spécifique et particulier à un promoteur (ou groupe de promoteurs) (Kadonaga 2004). Comme le montre l'influence de délétions dans les différentes séquences régulatrices d'un promoteur, l'activité transcriptionnelle « nette » est la résultante de l'action collective (aussi bien positive que négative) de chacun des acteurs (machinerie, facteurs de transcription, co-activateurs) liés à la séquence promotrice du gène exprimé.

I.2. Les moyens d'étude des interactions ADN-protéines

I.2.1. Approche avec *a priori*

Les interactions ADN-protéines jouent donc un rôle capital dans le contrôle de l'expression des gènes et leur identification représente une étape indispensable pour une compréhension des régulations transcriptionnelles à la base des fonctions cellulaires. En effet, la régulation de nombreux processus biologiques prend racine au niveau transcriptionnel, de

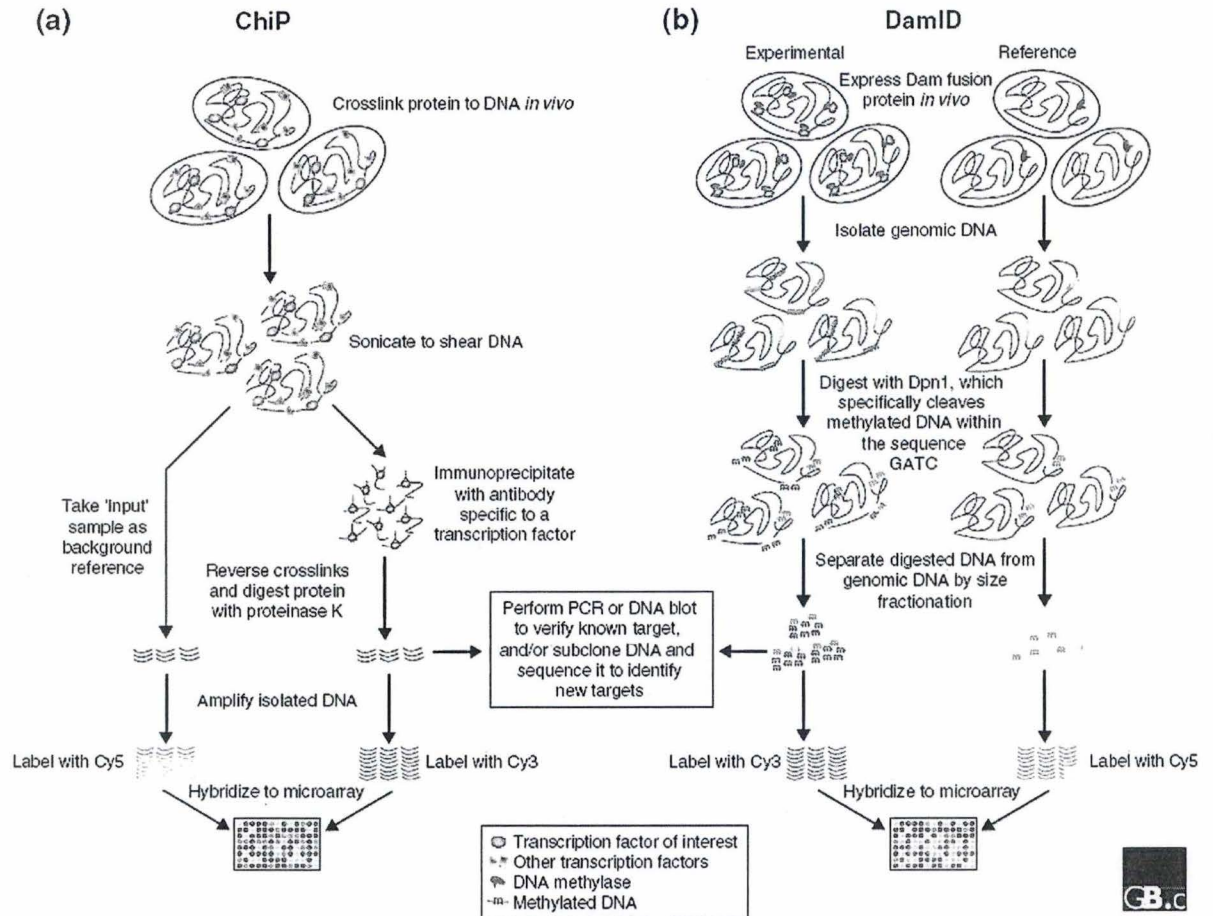


Figure 1.6 : Principe du ChIP on chip (A) et du DamID (B)

Ces deux techniques permettent d'identifier les séquences d'ADN liées *in vivo* par un facteur de transcription d'intérêt. Les différentes étapes sont brièvement développées dans le texte (Taverner, Smith et al. 2004)

par le contrôle du niveau d'expression d'effecteurs-clés. L'importance de ces interactions peut être illustrée par les nombreuses pathologies associées à la liaison anormale de ces facteurs à l'ADN. C'est notamment le cas du facteur de transcription NF- κ B, dont l'activation anormale est souvent associée à des pathologies comme certaines maladies inflammatoires (Atreya, Atreya et al. 2008) ou le développement de cancers (Karin 2006).

I.2.1.1. L'immunoprécipitation chromatinienne

A l'heure actuelle, diverses techniques sont à la disposition du biologiste pour étudier la liaison de ces facteurs aux acides nucléiques. Premièrement, une méthode couramment utilisée pour rechercher les sites de liaison d'un facteur de transcription est l'immunoprécipitation chromatinienne (ChIP, Fig. 1.6). En bref, les cellules vivantes sont fixées avec du paraformaldéhyde, réagissant avec les groupements amines et imines des acides aminés et des acides nucléiques afin de figer, entre autres, les interactions ADN-protéines et protéines-protéines. La chromatine est ensuite isolée et fragmentée en séquences d'environ 500 pb par sonication avant d'être immunoprécipitée grâce à un anticorps spécifiquement dirigé contre la protéine étudiée. Les liens formés avec le paraformaldéhyde sont alors rompus en incubant le précipité à une température supérieure ou égale à 67°C. Finalement, l'ADN est purifié et soit amplifié par PCR (éventuellement par PCR en temps réel) si l'on étudie une séquence nucléique particulière soit hybridés sur des damiers à ADN. Dans cette dernière situation, l'hybridation des séquences purifiées, amplifiées et marquées par une molécule fluorescente permet, par une méthode appelée « ChIP on chip », d'identifier, au niveau de l'ensemble du génome, les séquences nucléiques liant *in vivo* un facteur de transcription d'intérêt.

Une variante de cette technique de ChIP on chip est le DamID, permettant également d'identifier les séquences liées par un facteur au niveau génomique, mais reposant sur un principe différent (Fig. 1.6). Cette fois, la protéine d'intérêt fusionnée à la Dam (DNA adénine méthyltransferase) est surexprimée à partir d'un vecteur d'expression. La méthode est basée sur l'activité de la Dam qui va spécifiquement méthyler les adénines des sites GATC à proximité de la position où la protéine de fusion surexprimée s'est liée. L'ADN purifié est ensuite digéré par la *DpnI*, enzyme coupant spécifiquement au niveau des adénines méthylées des sites GATC. Finalement les fragments contenant les sites méthylés sont purifiés, amplifiés et marqués pour être hybridés sur les sondes des damiers à ADN. Cette technique a l'avantage qu'elle ne nécessite pas d'anticorps reconnaissant le facteur de transcription lié à l'ADN, mais elle implique que la fusion de la protéine avec la DAM n'interfère pas avec la capacité de reconnaissance et de liaison du site reconnu par le facteur étudié (Taverner, Smith et al. 2004; Vigneault and Guerin 2005; Heintzman and Ren 2007)

Finalement, lors d'immunoprécipitation chromatinienne classique, les protéines coprécipitées interagissant avec le facteur étudié peuvent également être identifiées par western blot ou spectrométrie de masse. Le ChIP est un outil puissant car il permet de connaître quelles sont les séquences liées par un facteur de transcription *in vivo*, prenant donc en compte la structure chromatinienne de l'ADN, à un moment et dans des conditions expérimentales précises. Néanmoins, il nécessite d'utiliser un anticorps ayant une forte affinité pour la conformation native liée à l'ADN de la protéine étudiée, sans quoi l'immunoprécipitation ne peut avoir lieu.

I.2.1.2. L'empreinte aux nucléases

La seconde stratégie communément utilisée a pour but de rechercher la liaison de facteurs de transcription à une séquence d'intérêt. L'empreinte aux nucléases (Galas and

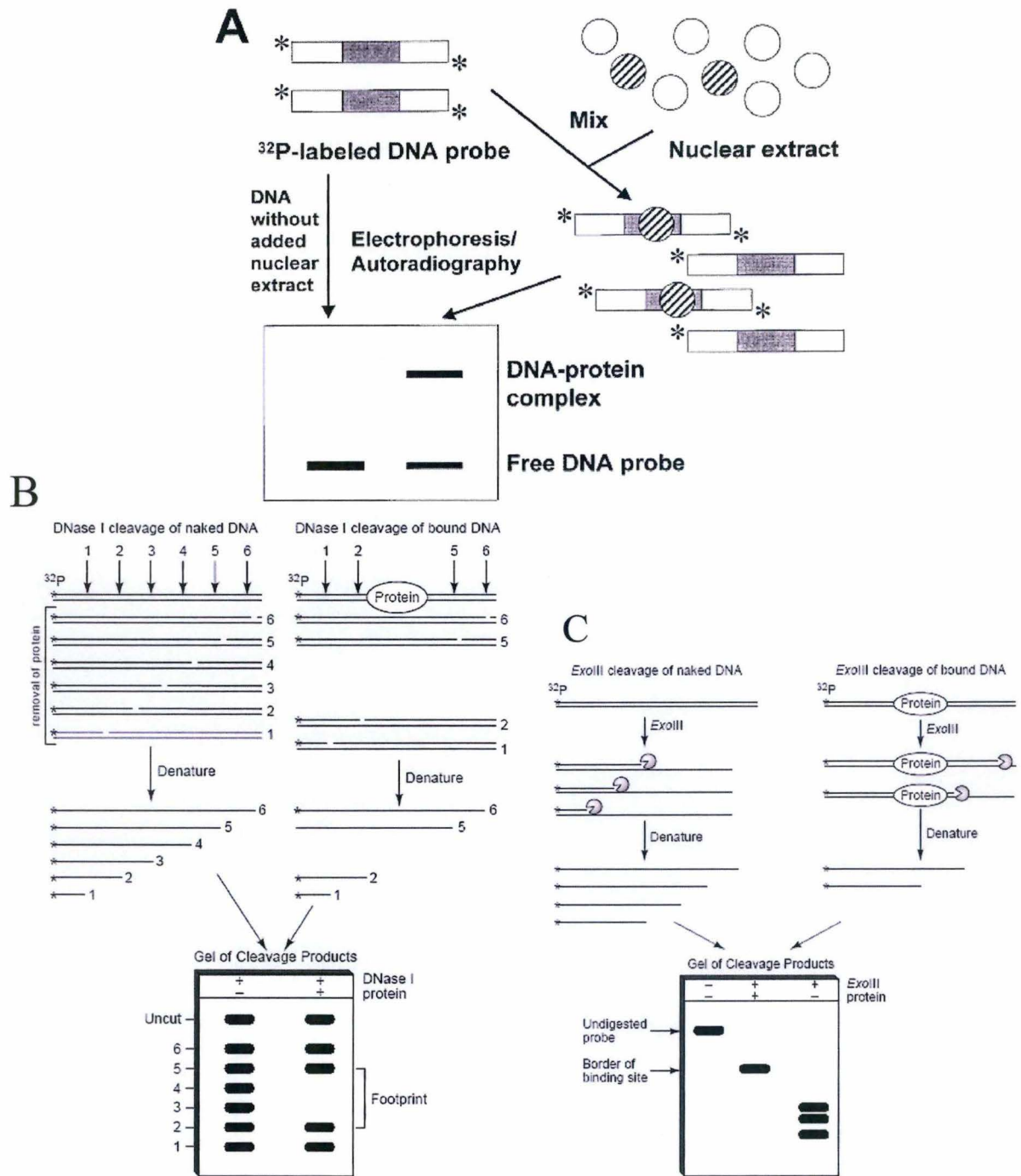


Figure 1.7 : Principe de quelques méthodes d'analyse des interactions entre acides nucléiques et protéines : (A) EMSA (Yang 1998), (B) empreinte à la DNase et (C) empreinte à l'exonucléase (Carey and Smale, 1999).

Schmitz 1978), et le retard sur gel (EMSA pour electromobility shift assay, (Fried 1989)) représentent les deux techniques les plus utilisées dans ce domaine.

Lors d'expériences d'empreinte aux nucléases (« footprinting »), la liaison d'un facteur de transcription à la séquence qui lui est spécifique permet de protéger le site lié d'une digestion enzymatique, marquant d'une « empreinte » le site de fixation. Des sondes, dont une seule extrémité a été marquée avec un isotope radioactif, sont mises en présence d'extraits nucléaires avant d'être incubées avec une enzyme de restriction (par exemple la DNase I (Fig. 1.7 B) ou l'exonucléase III (Fig. 1.7 C)) afin de digérer l'ADN au niveau des sites reconnus. Après séparation par électrophorèse sur gel dénaturant et révélation par autoradiographie, les profils de digestion des séquences mises en présence d'extraits nucléaires ou non permettent de déterminer la liaison d'un facteur de transcription au fragment étudié et d'identifier plus précisément le site de fixation de celui-ci. Si le footprinting possède une résolution plus élevée que l'EMSA, il représente une méthode moins sensible, étant donné qu'une majorité des sondes doivent être liées par le facteur pour qu'une protection à la digestion soit détectable. De plus, certaines enzymes comme l'Exo III nécessitent une forte affinité de la protéine pour son site de liaison et un temps de demi-vie du complexe élevé, sans quoi la digestion peut avoir lieu. Pour finir, différentes améliorations ont été apportées à la technique d'origine comme le couplage du footprinting avec le FRET (fluorescence resonance energy transfer, (Wang, Li et al. 2005)). L'empreinte à l'Exo III peut également être couplée à la PCR en temps réel, permettant une quantification de l'abondance des fragments non digérés afin de mettre en évidence la liaison de protéine présente à des concentrations allant jusqu'au femtomolaire (Hou, Chen et al. 2007).

I.2.1.3. Retard sur gel et méthodes dérivées

La seconde technique couramment utilisée est le retard sur gel. L'EMSA (Fig. 1.7 A) est basé sur le principe qu'une sonde nucléique liée par un facteur de transcription aura une mobilité électrophorétique, déterminée par la forme, la taille et la charge du complexe, réduite par rapport à cette même sonde si elle était libre. D'un point de vue pratique, de courtes sondes d'ADN, de 20 à 70 paires de bases, marquées radioactivement à leurs extrémités, sont incubées en présence de protéines ou d'extraits nucléaires, puis séparées au cours d'une électrophorèse en conditions non dénaturantes. Après révélation par autoradiographie, la comparaison des profils de migration entre les sondes seules ou incubées en présence de protéines permet de déterminer s'il y a eu liaison à l'ADN. La spécificité de ces liaisons peut être ensuite testée par des expériences de compétitions où des sondes non radioactives sont ajoutées en excès croissant, celles-ci correspondant soit à la séquence sauvage étudiée ou à son homologue muté. L'EMSA représente une méthode *in vitro* très sensible étant donné que seuls quelques complexes retardés suffisent pour être détectés après autoradiographie. Toutefois, l'EMSA classique ne permet pas de déterminer quel est le facteur de transcription lié à l'ADN. Afin de répondre à cette question, une variante a été développée, appelée supershift. Cette fois, les sondes marquées sont incubées en présence d'extraits nucléaires mais également avec un anticorps dirigé contre l'hypothétique facteur lié à l'ADN. Si l'hypothèse s'avère correcte, l'anticorps accentuera le retard du complexe ADN-protéines en diminuant davantage la mobilité de ce dernier. Cependant, si cette variante permet d'identifier le facteur lié à la sonde nucléique, l'anticorps doit être de haute qualité, reconnaissant la protéine native liée à l'ADN. De plus, dans une situation où un complexe multiprotéique est lié à l'ADN, l'accentuation du retard généré par l'anticorps n'est pas toujours facilement distinguable du complexe retardé lui-même (Laniel, Beliveau et al. 2001).

C'est notamment pour les raisons citées précédemment que diverses alternatives au supershift ont vu le jour, comme des méthodes basées sur le principe du test ELISA (Fig. 1.8).

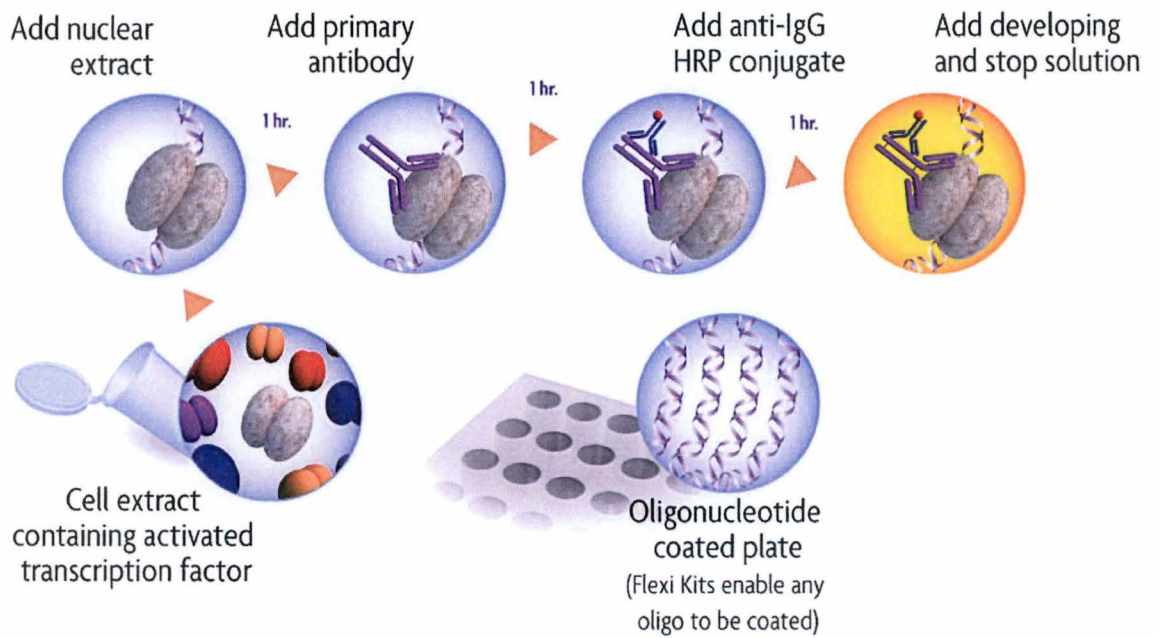


Figure 1.8 : Principe général du dosage colorimétrique de liaison à l'ADN (ELISA-like) commercialisé sous le terme TransAM™ (Active motif). www.activemotif.com

En bref, des sondes nucléiques biotinylées sont tout d'abord immobilisées dans des plaques 96 puits tapissées de streptavidine. Les extraits nucléaires sont ensuite incubés dans celles-ci, en solution avec un tampon de liaison adapté au facteur étudié, avant d'être mis en présence d'un anticorps primaire dirigé contre la protéine d'intérêt. Finalement, la liaison du facteur à la sonde nucléique peut être révélée soit par chémoluminescence (Rosenau, Emery et al. 2004) ou colorimétrie (Renard, Ernest et al. 2001) grâce à un anticorps secondaire couplé à la peroxydase. Ces méthodes sont très sensibles et semi quantitatives mais nécessitent de connaître préalablement le facteur se liant à la séquence et d'utiliser un anticorps spécifique à celui-ci.

1.2.1.4. Prédiction *in silico*

De nos jours, une troisième option en pleine essor se présente au biologiste lorsqu'il souhaite étudier les interactions entre acides nucléiques et protéines. En effet, les prédictions *in silico* font l'objet de nombreuses mises au point depuis quelques années et peuvent être un outil précieux lorsqu'elles sont utilisées en tenant compte de leurs limites. Cette fois encore, diverses méthodologies ont été développées. Une première approche se base directement sur la séquence nucléique afin d'identifier la présence de sites putatifs de liaison en utilisant le site consensus propre à un facteur ou des matrices de scores prenant en considération l'importance de chaque nucléotide aux différentes positions du site de liaison. La seconde méthodologie repose sur l'identification de motifs conservés, comme les sites de liaison, sélectionnés au cours de l'évolution et représentant une empreinte phylogénétique qui peut être recherchée en comparant les séquences homologues chez deux ou plusieurs espèces. Le désavantage majeur des prédictions bioinformatiques est qu'elles génèrent de nombreux faux positifs et négatifs. En effet, la redondance des séquences reconnues par différents facteurs de transcription peut conduire à l'identification d'un site de liaison qui en réalité n'est pas pertinent *in vivo*. Inversement, les prédictions *in silico* fonctionnent à partir des informations contenues dans les bases de données (telles que TRANSFAC (Matys, Kel-Margoulis et al. 2006)) provenant d'expériences antérieures et générées à partir de méthodes classiques, telles que celles décrites précédemment, ce qui rend la bioinformatique « aveugle » face à certains sites de liaison et facteurs de transcription encore non décrits. Il est aussi possible qu'un facteur lie des sites atypiques contenant des séquences non consensus, ce qui le rend encore une fois non détectable à l'aide de certains algorithmes. Finalement, la liaison d'un facteur à son site ne se fait qu'au niveau de quelques nucléotides et la possibilité d'éventuelles substitutions peut permettre à certaines méthodes, comme les matrices de scores, d'identifier de nombreux sites non pertinents avec des paramètres et des seuils de tolérance inadéquats. A l'heure actuelle, la recherche bioinformatique de sites de liaison pour des facteurs de transcription peut s'avérer très utile, mais elle doit être considérée comme un outil parmi d'autres, capable d'identifier au sein de longues séquences la présence de sites putatifs, permettant de la sorte de restreindre les zones étudiées qui doivent être, par la suite, confirmées *in vitro* et *in vivo* par des méthodes plus classiques. (Heintzman and Ren 2007)

Toutes les méthodes décrites ci-dessus ne permettent d'étudier que la liaison d'une protéine à l'ADN. Toutefois, ce n'est pas parce qu'un facteur de transcription est lié au promoteur d'un gène qu'il va pour autant réguler l'expression de ce dernier. En effet, comme nous l'avons déjà vu précédemment des modifications post-traductionnelles peuvent par exemple réguler la capacité de trans-activation d'un facteur. Des études fonctionnelles complémentaires sont donc indispensables lors de la compréhension de la régulation d'une séquence nucléique. Parmi celles-ci, le système de gène rapporteur est couramment utilisé. En bref, la séquence nucléotidique est clonée en amont d'un gène rapporteur, comme la luciférase ou la GFP (green fluorescence protein), au sein d'un plasmide recombinant qui sera transfecté

dans les cellules vivantes. Une méthode de détection adaptée au type de gène rapporteur utilisé permet ensuite de quantifier l'expression du gène rapporteur régulée par la séquence d'intérêt (Cullen 1987). Ce système permet également d'évaluer la contribution d'un site particulier en étudiant l'impact de la délétion de ce site sur l'expression du gène rapporteur. e

I.2.2. Analyse sans *a priori* des interactions ADN-protéines

I.2.2.1. Intérêt et avancées actuelles d'une approche sans *a priori*

De nos jours, aucune technique ne permet d'identifier, sans connaissance préalable, tous les acteurs protéiques interagissant directement ou indirectement avec une séquence oligonucléotidique. Un chercheur souhaitant étudier les protéines liant et régulant l'expression d'un gène d'intérêt encore non décrit n'a pas d'autre alternative que de recourir aux prédictions *in silico* d'éléments *cis* présents dans la région promotrice du gène étudié. Ce type d'analyse bioinformatique génère de nombreux sites candidats, et le chercheur doit souvent choisir arbitrairement les sites qui seront validés expérimentalement par des méthodes telles que décrites ci-dessus. Suite aux nombreux sites pouvant être prédits et aux risques d'erreurs liés à ce type d'approche, une telle démarche peut être longue et n'aboutir à aucun résultat satisfaisant.

Dans ce contexte, des tentatives d'isoler les protéines qui interagissent physiquement avec une séquence oligonucléotidique ont vu le jour ces dernières années. La difficulté majeure rencontrée lors du développement d'une telle méthode émane de l'infime abondance des régulateurs transcriptionnels en comparaison avec l'abondante présence d'autres protéines impliquées par exemple dans la maintenance de l'ADN. De plus, une amplification de la séquence en acides aminés ne peut être envisagée comme cela peut l'être pour les séquences nucléiques via la réaction de polymérisation en chaîne. Nous détaillerons dans les paragraphes suivants les techniques de chromatographie visant à purifier les protéines ayant une affinité pour une séquence d'ADN particulière (« DNA-affinity »).

La chromatographie d'affinité s'est révélée comme un moyen simple et efficace pour directement purifier le complexe protéique interagissant avec l'ADN. Concrètement, une séquence nucléotidique d'intérêt est fixée à un support chromatographique et utilisée comme sonde afin de capturer les protéines interagissant avec cette séquence. Les facteurs sont ensuite élués généralement via un gradient de sel ou de détergent. En 1989 déjà, Gabrielsen et Hornes ont utilisé ce système en immobilisant une séquence oligonucléotidique, constituée d'une répétition du site de liaison de la protéine recherchée, sur des billes magnétiques afin de purifier avec un rendement élevé le facteur de transcription τ de levure (Gabrielsen, Hornes et al. 1989). Cette procédure a permis de déterminer le poids moléculaire de différents peptides composant le complexe protéique liant cette séquence.

Par la suite, le principe de purification par « DNA-affinity » a fréquemment été couplé à une analyse des protéines par spectrométrie de masse (pour une revue sur la contribution de la spectrométrie de masse dans le cadre des interactions entre acides nucléiques et protéines, voir (Rusconi, Guillonnet et al. 2002)). Par exemple, la combinaison de ces deux techniques a permis d'identifier 4 nouvelles protéines liant la séquence SRE du promoteur du gène codant le facteur de transcription c-Fos, en plus des SRFs (serum responsive factor) et Elk-1 connus pour lier cette séquence (Drewett, Molina et al. 2001).

Toutefois, ces approches sont relativement peu sensibles, car l'abondance des facteurs de transcription recherchés est largement inférieure à celles d'autres protéines nucléaires. En

effet, certaines de ces protéines, que nous appellerons « contaminants », se lient soit à l'ADN de manière non spécifique, soit au support solide utilisé pour immobiliser l'oligonucléotide (colonne chromatographique ou billes magnétiques). Lors des étapes d'éluions des protéines liées à l'ADN, typiquement pas le biais d'un gradient de concentration en sels ou de détergent, les contaminants adsorbés sur le support solide sont également récoltés.

Diverses stratégies ont vu le jour afin de pallier à ce problème. Gygi et ses collaborateurs ont utilisé une approche, ICAT (isotope-coded affinity tag), permettant une analyse à la fois quantitative et qualitative des interactions, en soustrayant l'information spécifique du bruit de fond généré par les liaisons aspécifiques (Gygi, Rist et al. 1999). Après extraction des protéines nucléaires, les extraits sont purifiés grâce à une colonne portant une séquence nucléique contenant soit le site spécifique sauvage ou son homologue muté. Après purification, l'éluat est récolté et marqué radioactivement de manière distincte selon qu'il soit passé sur la colonne présentant le site spécifique (marquage à l'hydrogène) ou muté (marquage au deutérium). Finalement, une analyse différentielle des résultats par spectrométrie de masse permet de soustraire le bruit de fond généré par les protéines contaminantes liant spécifiquement le site d'intérêt sauvage. Bien qu'efficace, cette méthode nécessite une bonne connaissance préalable de la séquence oligonucléotidique d'intérêt dans laquelle le(s) site(s) de liaison a (ont) déjà été définis avec précision, ainsi que les mutations permettant d'abolir la liaison des facteurs de transcription. Une telle approche ne peut donc pas être envisagée lorsque l'on souhaite travailler sur une séquence non caractérisée, pour laquelle aucun site de liaison n'a été identifié. La meilleure solution à ce jour pour éliminer le bruit de fond généré par les protéines « contaminantes » semble être l'utilisation d'ADN compétiteur de type ADN de sperme de saumon ou PolydIdC. Cependant, si la nature du compétiteur ne semble pas importer, la concentration utilisée peut influencer les résultats obtenus : à concentration trop élevée, les ADN compétiteurs peuvent également faire chuter la liaison spécifique à la sonde nucléique (Park, Ko et al. 2005).

Diverses solutions ont également été envisagées pour éviter d'identifier les contaminants adsorbés sur le support solide. La première stratégie pour contrer cette difficulté consiste à faire passer les extraits par une étape préliminaire de chromatographie afin d'éliminer ces facteurs avant la purification proprement dite. Néanmoins, ce pré-fractionnement peut être à l'origine d'une perte de matériel et donc d'informations intéressantes. Une seconde approche pour éviter de récolter les protéines adsorbées sur le support solide, tout en assurant une perte de matériel minimale, semble être le décrochage des complexes ADN-protéines du support où ils étaient immobilisés. Diverses stratégies sont envisageables, basées soit sur une compétition permettant le déplacement du complexe ADN-protéines du support solide, soit sur le clivage d'un lien entre les deux.

Par exemple, l'équipe de Shimkus a eu recours à une approche où ils ont introduit un lien contenant un pont disulfure entre la sonde nucléique et le support solide, clivable suite à l'incubation du complexe en présence d'un agent réducteur (Shimkus, Levy et al. 1985). Cependant, les conditions réductrices peuvent influencer la liaison des facteurs de transcription à l'ADN et conduire à la perte d'une partie des informations recherchées. (Park, Ko et al. 2005) Un lien photoclivable peut également être introduit entre l'oligonucléotide et le support solide. Suite à l'exposition aux UVs, les complexes ADN-protéines sont libérés du support solide tandis que les contaminants restent adsorbés sur les billes. Cependant, une exposition aux UVs peut entraîner la liaison covalente de certains facteurs de transcription à l'oligonucléotide empêchant par la suite l'identification de ce facteur par spectrométrie de masse (L. Payen, communication personnelle). Enfin, un site de restriction peut également être introduit dans la séquence afin de détacher par la suite l'oligonucléotide du support (Rieger, Zaika et al. 2006).

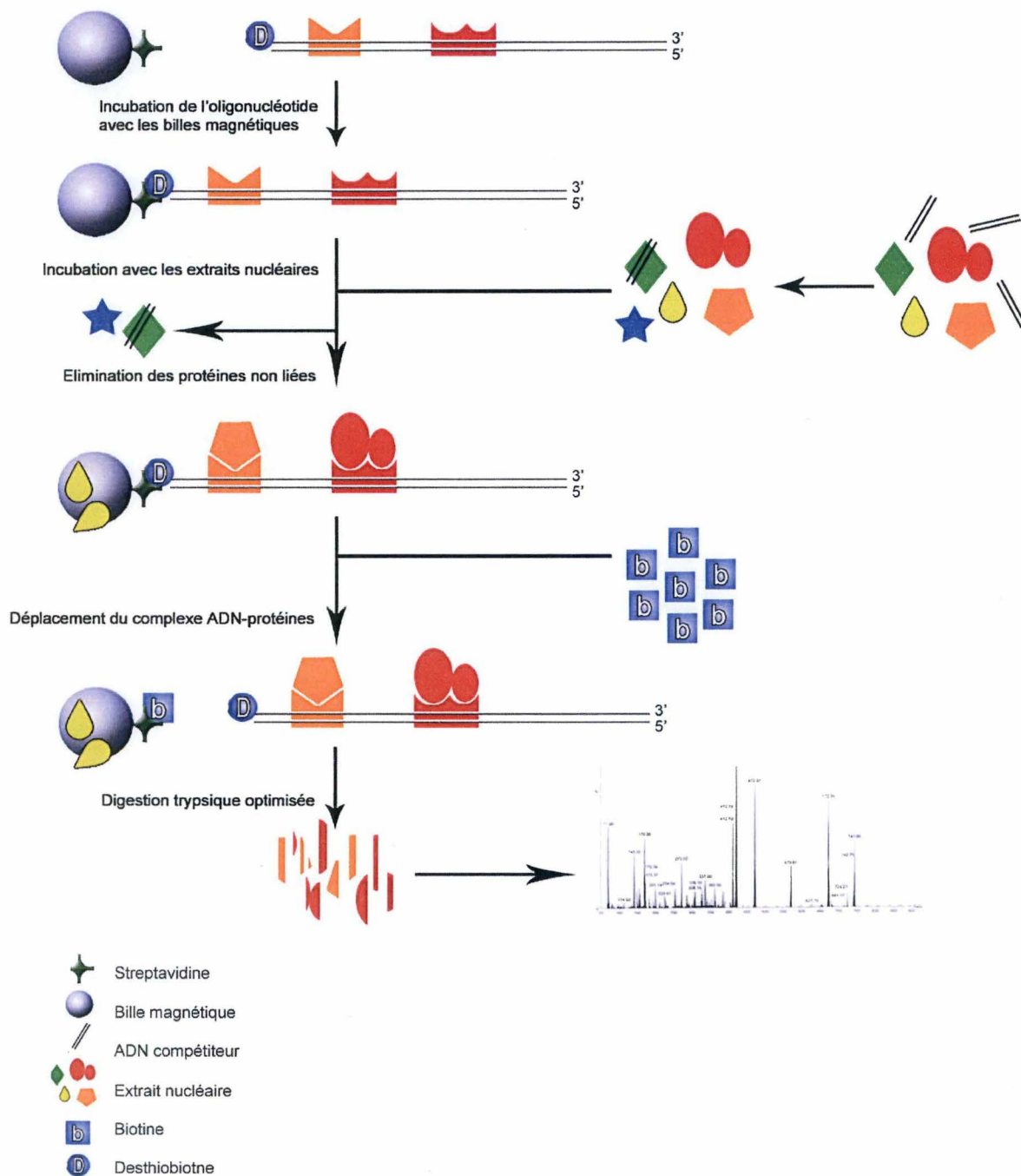


Figure 1.9 : Principe de la méthode d'identification sans *a priori* des protéines liant une séquence nucléique d'intérêt.

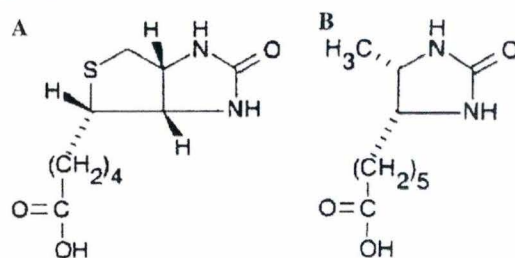


Figure 1.10 : Structure chimique de la biotine (A) et de la desthiobiotine (B)

La deuxième possibilité pour éviter de récolter les protéines adsorbées sur le support solide repose sur des réactions de compétitions où les complexes ADN-protéines sont déplacés grâce à l'ajout d'un excès de molécules ayant une plus grande affinité pour le support solide que celle permettant de lier la sonde nucléique à ce dernier. Cette stratégie a déjà été appliquée à la purification par affinité de protéines liant spécifiquement l'ARN. Dans ce cas, un ARN hybride est synthétisé, possédant un tag liant avec une grande affinité la streptomycine, en plus de la séquence nécessaire à la capture des protéines. Après incubation avec les extraits protéiques, les complexes ARN-protéines sont purifiés sur une colonne sur laquelle de la streptomycine est immobilisée. Finalement, les complexes ARN-protéines sont élués grâce à un excès de streptomycine libre (Bachler, Schroeder et al. 1999). Le désavantage majeur de ce procédé provient de l'ajout de nouvelles molécules en excès qui peuvent interférer avec l'analyse en spectrométrie de masse.

I.2.2.2. Principe de la méthode développée

Au laboratoire, nous travaillons depuis quelques années au développement d'une méthode d'identification des protéines capturées par une séquence oligonucléotidique d'intérêt, en suivant le principe de purification de ces protéines par chromatographie d'affinité pour cette séquence, suivie d'une identification par spectrométrie de masse. Les différentes étapes de cette méthode sont reprises à la Fig. 1.9:

1. Production par PCR de l'oligonucléotide marqué à son extrémité 5' par une desthiobiotine, grâce à l'utilisation d'une amorce sens desthiobiotinylée. La desthiobiotine est une forme modifiée chimiquement de la biotine (Fig. 1.10) ayant pour conséquence une plus faible affinité pour la streptavidine que la vitamine originale. Cette molécule va permettre dans un premier temps de lier les oligonucléotides aux billes magnétiques et ensuite de les décrocher en exploitant la différence d'affinité pour la streptavidine existant entre les deux molécules.
2. Immobilisation de l'oligonucléotide sur le support solide, par interaction desthiobiotine-streptavidine. Des billes magnétiques ont été préférées à un support chromatographique classique car elles permettent un meilleur accès des protéines à l'ADN. Le rapport entre la quantité de billes (1 mg) et l'ADN (20 pmoles) incubé a déjà été l'objet d'optimisations dans le passé.
3. Les séquences oligonucléotidiques, liées aux billes magnétiques, sont incubées avec les extraits nucléaires en présence de quantités modérées d'ADN compétiteurs. Une étape de préincubation des protéines nucléaires avec ces compétiteurs est effectuée afin de lier les facteurs liant non spécifiquement la séquence nucléique d'intérêt.
4. Après différentes étapes de lavage visant à éliminer les protéines non liées à l'oligonucléotide (lavages dont les conditions de salinité ont été mises au point), les complexes ADN-protéines sont déplacés des billes magnétiques suite à une incubation en présence d'un excès de biotine.
5. Finalement, les complexes ADN-protéines sont dénaturés de manière à garantir une digestion trypsique optimale afin d'être analysés en spectrométrie de masse en tandem, permettant l'identification des facteurs qui se sont liés à la séquence d'intérêt étudiée.

Les mises au point effectuées avant le début de ce mémoire ont été réalisées dans le cadre de la thèse d'Aurélien Tacheny, sur une portion bien décrite du LTR-5' (nt 228 à 463) du HIV-1, contenant deux sites de liaison pour le facteur de transcription NF- κ B et dont la liaison a été étudiée à partir d'extraits nucléaires de cellules HeLa stimulées 45 minutes à l'IL-1 β . Cette stimulation permet effectivement d'activer ce facteur de transcription, dont la détection

a servi de point de repère pour les mises au point. Ces travaux ont permis d'identifier différentes protéines dont p50 et p65 à partir d'extraits provenant de cellules stimulées. Ce signal semble spécifique puisque aucun peptide de p50 ou de p65 n'a été détecté dans des conditions où les sites de liaison pour NF- κ B étaient mutés, ni lorsque la séquence sauvage était incubée en présence d'extraits de cellules non traitées. Cependant, aucun autre facteur de transcription n'a été identifié, alors que cette séquence est bien connue pour capturer de nombreux régulateurs.

L'objectif est maintenant d'étudier une séquence intragénique encore peu décrite du HIV-1, en collaboration avec le laboratoire de C. Van Lint (ULB), afin de pouvoir mettre en évidence la liaison de facteurs de transcription et de nouvelles améliorations à cette technique. Nous décrivons dans les paragraphes suivants en quoi la séquence choisie revêt un intérêt particulier pour la régulation du HIV-1.

I.3. Le virus de l'immunodéficience acquise (HIV-1)

I.3.1. Structure

Le HIV-1 (Human immunodeficiency virus) est un lentivirus identifié comme agent étiologique du syndrome d'immunodéficience acquise (SIDA).

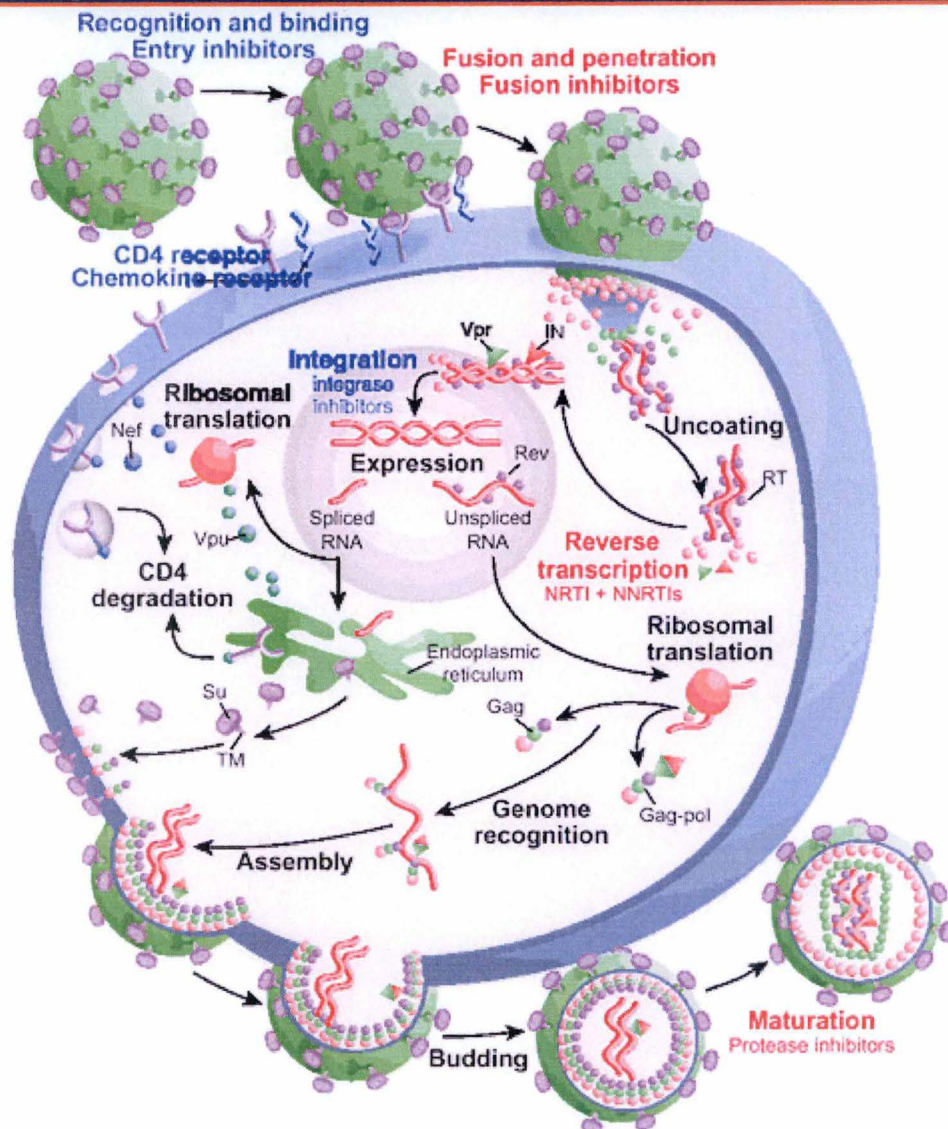
Le HIV-1 est formé d'une structure icosaédrique (Fig. 1.11) constituée d'une enveloppe, d'une matrice et d'une capsidie contenant deux molécules d'ARN monocaténaire de 9,5 kb. L'enveloppe est une bicouche lipidique où sont insérées des glycoprotéines, gp120 et gp41, nécessaires à la reconnaissance et à la fusion avec la membrane hôte, respectivement, ainsi que des protéines provenant de l'hôte cellulaire. La matrice assure quant à elle la bonne cohésion entre l'enveloppe et la capsidie.

Du point de vue de son génome (Fig. 1.12), le HIV-1 contient les 3 gènes communs à tous les rétrovirus, *gag*, *pol*, *env*, ainsi que 6 autres gènes supplémentaires : *tat*, *rev*, *nef* codant pour des protéines de régulation, et *vif*, *vpr*, *vpu*, codant pour des protéines dites accessoires. Le gène *gag* encode des précurseurs qui, une fois maturés, participeront à l'assemblage et la libération des nouvelles particules virales, ainsi qu'à la formation de protéines formant la matrice (p17) et la capsidie (p24). Le gène *pol* code pour la transcriptase inverse (RT), l'intégrase (IN) et la protéase (PR) virale. Les glycoprotéines formant l'enveloppe du virion et permettant la reconnaissance de la cellule hôte, les acteurs permettant la fusion des membranes plasmiques et virales sont encodés par le gène *env*.

La région codante du génome du HIV-1 est entourée de deux régions non codantes, identiques, formée de l'U5 (extrémité 5' unique), R (répétition) et l'U3 (extrémité 3' unique). Ces régions formeront des zones de régulation transcriptionnelle appelées LTR (long terminal repeat) aux extrémités 5' et 3' de la molécule d'ADN bicaténaire rétrotranscrite à partir de l'ARN viral. Alors que le LTR situé en 5' (LTR-5') constitue un promoteur régulant l'expression du génome viral, le LTR-3' joue davantage un rôle de contrôle de la polyadénylation des transcrits générés.

I.3.2. Cycle réplcatif du HIV-1

Le cycle réplcatif du HIV-1 (Fig. 1.13) est semblable à celui des autres lentivirus et peut être divisé en deux phases distinctes. Une première dite précoce, allant de l'attachement du virus à la cellule hôte jusqu'à son intégration dans le génome de celle-ci. La seconde phase est dite tardive et constitue l'ensemble des étapes comprises de la transcription du virus intégré jusqu'au bourgeonnement des virions (pour une revue complète sur l'ensemble du cycle réplcatif du HIV-1 voir (Freed 2001))



Source: Nat Med © 2003 Nature Publishing Group

Figure 1.13 : Cycle répliatif du HIV-1

Les étapes ayant lieu jusqu'à l'intégration du virus représentent la phase précoce du cycle répliatif, alors que les étapes allant de l'expression du provirus jusqu'au bourgeonnement des virions appartiennent à la phase tardive (Pomerantz and Horn 2003).

En bref, la première étape du cycle du HIV-1 commence par la liaison de la glycoprotéine gp120 au récepteur cellulaire CD40, ce qui va induire un changement de conformation de cette protéine, lui permettant d'agir avec un co-récepteur. Deux membres de la famille des récepteurs couplés aux protéines G sont ciblés : CXCR4 et CCR5, co-récepteurs aux chimiokines α et β , respectivement, et dont le tropisme des isolats pour les lignées cellulaires monocytaires/macrophagiques et T provient de la capacité d'interaction avec ceux-ci (Berger, Murphy et al. 1999). Alors que les isolats M-tropiques interagissant avec CCR5 seront plutôt retrouvés en début d'infection, les isolats T-tropiques liant CXCR4 sont détectés chez les patients atteints du SIDA. Un changement de conformation de la seconde glycoprotéine (gp41) insérée dans l'enveloppe virale va stimuler la fusion et l'internalisation du virus grâce au peptide de fusion présent dans son domaine N-terminal. Une fois le virus entré dans la cellule, le complexe de rétrotranscription (RTC) va se former, permettant la rétrotranscription de l'ARN en ADN par un processus commun à tous les rétrovirus. Par après, cette molécule va être transportée vers le noyau sous la forme d'un complexe de préintégration (PIC) comprenant entre autre l'intégrase virale. Une fois la membrane nucléaire franchie, le virus est finalement intégré dans le génome cellulaire, préférentiellement au sein d'introns de gènes transcriptionnellement actifs. Le provirus intégré est dès lors considéré comme les autres gènes cellulaires, ce qui met fin à la première phase du cycle répliatif. Le provirus intégré peut rester silencieux durant des années.

Durant le début de la phase tardive, la transcription du provirus par l'ARNP II à partir du LTR-5' permet la synthèse de transcrits qui seront ensuite exportés dans le cytoplasme afin d'être traduits ou encapsidés dans les nouvelles particules virales. Le LTR-5' comprend de nombreux éléments en *cis* parmi lesquels deux sites NF- κ B, jouant un rôle majeur dans l'activation de la transcription du provirus (Griffin, Leung et al. 1989). Approximativement 30 messagers différents peuvent être générés à partir du virus intégré, par le biais d'épissage alternatif et l'exploitation des divers cadres ouverts de lecture possibles. Ces transcrits peuvent être répartis selon 3 classes : les non épissés (précurseurs de Gag et GagPol), les partiellement épissés (Env, Vif, Vpr) et les multi-épissés (Tat, Rev, Nef). Cette fois encore, la transcription du virus peut être décrite en deux temps. Tout d'abord, les transcrits multiépissés vont permettre la synthèse des protéines virales : Tat, Rev, Nef durant une phase dite « Tat-indépendante » (Fig. 1.14). Les deux premières protéines vont ensuite retourner dans le noyau grâce à leur séquence de localisation nucléaire. Des éléments répresseurs sont responsables de la faible efficacité de l'ARNP II, incapable de passer en phase d'élongation, et qui ne générera que de courts transcrits d'environ 60 nucléotides. Une séquence TAR (Transactivation response region), située directement en aval du site d'initiation, va former une boucle au niveau de ces ARNs naissants qui va être reconnue par la protéine Tat. Le recrutement de la cycline T1 et de la Cdk9 par la protéine virale au niveau de ce site va permettre une hyperphosphorylation du domaine CTD de la large sous-unité de l'ARNP II, qui sera alors capable de passer l'étape d'élongation et de générer des transcrits plus longs. C'est la phase dite « tat-dépendante ». L'origine des premières molécules de Tat dans la cellule infectée est toujours méconnue à l'heure actuelle. L'hypothèse la mieux acceptée est qu'occasionnellement, un transcrit long est tout de même généré durant la phase « tat-indépendante » de la transcription, permettant la synthèse de protéines Tat qui ensuite augmentent très fortement l'efficacité de la transcription de l'ensemble des gènes viraux (Marciniak and Sharp 1991; Karn 1999).

La seconde protéine exportée, Rev, va également reconnaître une séquence RRE (Rev responsive element) au niveau du transcrit et permettre l'exportation des ARNm non épissés ou partiellement épissés dans le cytoplasme. Une fois que l'ensemble des constituants

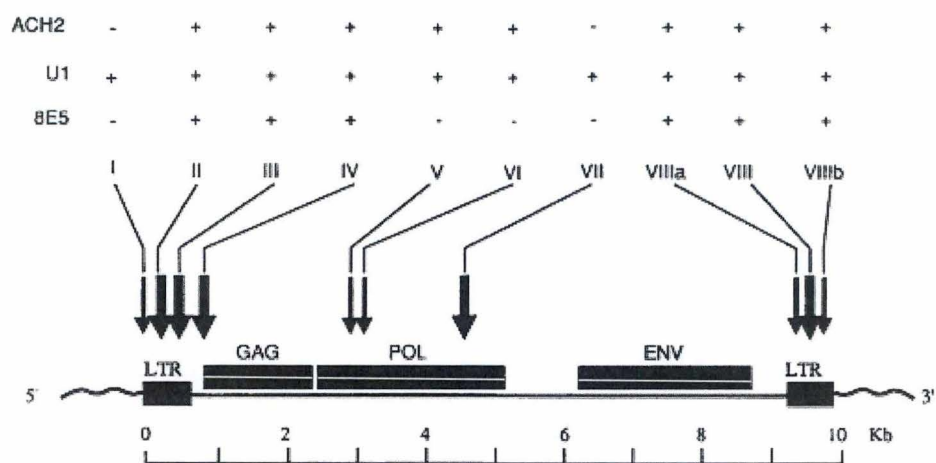


Figure 1.15 : Identification des zones d'hypersensibilités à la DNase I dans le génome du HIV-1 intégré.

Les zones II, III, IV, VII et VIII représentent des zones d'hypersensibilité majeure alors que les régions I, V, VI, VIIa, VIIIB sont des zones d'hypersensibilité mineure.

ACH2, 8E5 (cellules T CD4+) et U1 (promonocytes) sont des lignées cellulaires contenant deux copies du HIV-1 intégré (Verdin 1991)

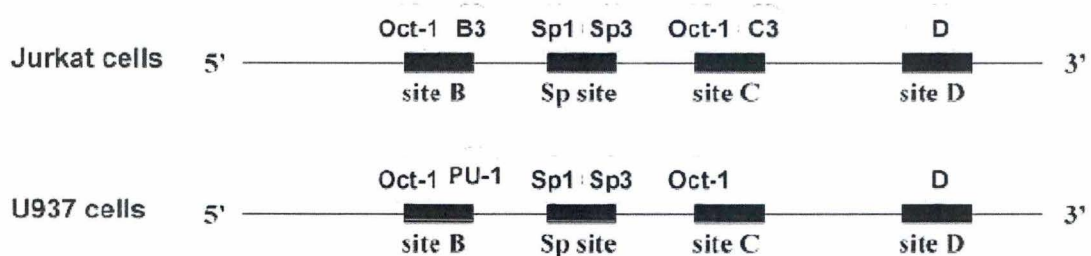


Figure 1.16 : Représentation des différents facteurs de transcription liant la région d'hypersensibilité SH7 dans les cellules Jurkat et U937. La liaison de Oct-1, Sp1, Sp3 et PU.1 a pu être montrée *in vivo* (ChIP). Le facteur D représente un facteur non identifié, ubiquiste, dont la liaison nécessite la présence de zinc. (Goffin, Demonte et al. 2005).

nécessaires pour la formation de virions fonctionnels sont disponibles, l'assemblage des nouvelles particules va avoir lieu en membrane plasmique grâce au complexe formé par Gag et GagPol. La présence d'un signal d'emballage permet d'assurer l'importation d'un génome complet au sein des capsides. Finalement, le cycle finit par le bourgeonnement et la maturation des nouvelles particules virales formées.

I.3.3. Identification et caractérisation d'une région d'hypersensibilité au sein du gène *pol*

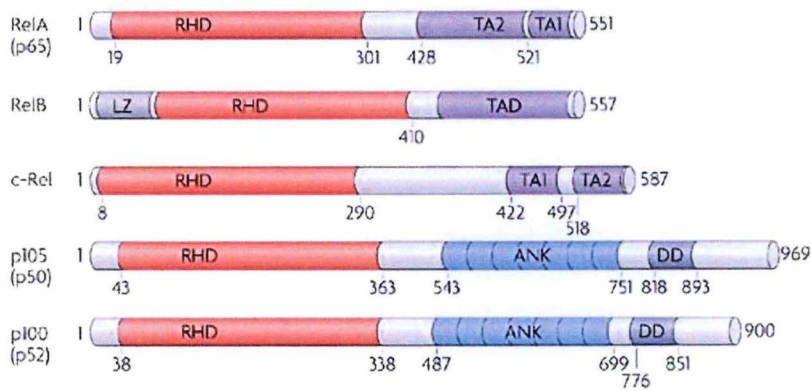
Comme nous venons de le voir, la majorité de la transcription du provirus est régulée au niveau du LTR-5', contenant du nombreux éléments en *cis*. Cependant, l'équipe de A. Burny a également identifié la présence d'un éventuel enhancer intragénique situé dans la gène *pol* du HIV-1 (Verdin, Becker et al. 1990). En effet, lorsque ce fragment est cloné en amont du promoteur de la thymidine kinase (TK) du virus HSV (Herpes simplex virus), il présente des propriétés d'enhancer inductible par le PMA + ionomycine dans des cellules HeLa. La localisation plus précise de cette activité enhancer a permis de mettre en évidence deux sous-domaines fonctionnels situés du nucléotide 4079 au 4342 et du nucléotide 4781 au 6028 (avec le +1 situé au start de l'U3 du LTR-5'). Ces derniers sont tous deux inductibles au TPA lorsqu'ils sont clonés en amont du promoteur TK en cellule HeLa. Par après, l'étude de l'organisation de la structure chromatinienne du virus intégré a permis d'identifier 5 zones d'hypersensibilité majeure, notamment 3 au niveau du LTR-5', une dans le LTR-3' et, de manière inattendue, également une située au centre de la région codante du virus intégré dans une lignée monocyttaire U1 (Fig. 1.15) (Verdin 1991). Plus précisément, elle est localisée entre les deux enhancers mis en évidence précédemment. Les zones d'hypersensibilités, caractérisées par l'absence de structure nucléosomale, sont souvent associées à des zones de régulation qui pourraient être dues à la liaison de protéines régulatrices à l'ADN.

La caractérisation physique de cette région, par EMSA et footprinting, a permis la mise en évidence de la présence de 4 sites de liaison dans les lignées T et monocytaires (Fig. 1.16) (Van Lint, Ghysdael et al. 1994, Goffin, 2005 #29):

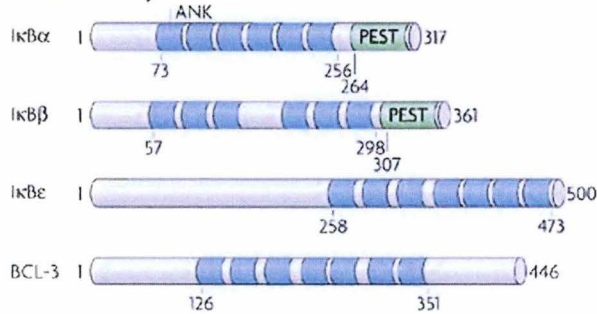
- Site B (Nt 4519 à 4545) : lie Oct-1, PU.1 *in vivo* (ChIP) dans les cellules monocytaires ainsi que Oct-1 et un complexe spécifique aux cellules T *in vitro* (EMSA).
- GC box (Nt 4623 à 4631) : des expériences de ChIP ont montré la liaison de Sp1 et Sp3 à cette boîte dans les lignées cellulaires T et monocytaires.
- Site C (Nt 4681 à 4701) : lie Oct-1 (*in vivo*), Oct-2 et un complexe spécifique aux cellules T (*in vitro*)
- Site D (Nt 4816 à 4851) : lie un facteur ubiquiste non identifié dont la liaison nécessite la présence de zinc.

L'impact des mutations individuelles ou simultanées de ces différents sites sur l'infectivité du virus a permis de montrer la pertinence fonctionnelle de ces sites *in vivo*. En effet, il semble y avoir une corrélation entre la liaison de ces facteurs à leurs sites respectifs, l'activité enhancer et la réplication du virus. Cette chute de la capacité de production de particules virales provoquée par la mutation des sites de liaison de la région SH7 peut être due à une diminution de la transcription étant donné que ni l'ARN ni les protéines encapsidés des virions ne semblent défectueux (Goffin, Demonte et al. 2005). Diverses hypothèses ont été émises quant au rôle de cette région. La présence d'un enhancer intragénique pourrait permettre de renforcer la transcription ayant lieu au niveau du LTR-5' ou apporter davantage de spécificité cellulaire. Une autre possibilité est la présence d'un promoteur interne au génome du virus HIV-1 comme cela a déjà été décrit pour le virus HTLV-1 (Ariumi, Shimotohno et al. 1998). Cependant, aucune des protéines identifiées à ce jour ne permet

a The NF- κ B family



b The I κ B family



c The IKK family

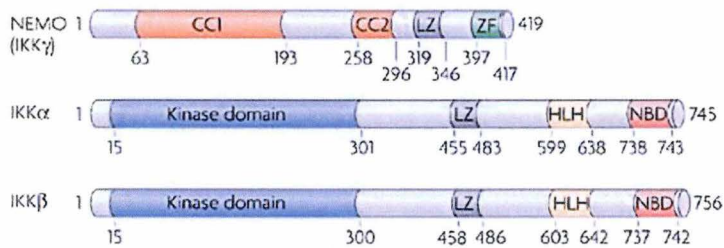


Figure 1.17 : Représentation de la structure des différents membres de la famille NF- κ B (A), I κ B (B) et IKK (C)

RHD : Rel-homology domain, TAD : domaine de trans-activation de la transcription, CC : Coiled-coil, DD: région homologue au domaine de mort, HLH :helix-loop-helix, LZ : domaine de trans-activation contenant un motif putatif homologue aux glissières leucine, NBD : domaine de liaison de NEMO, PEST : domaine riche en proline (P) ; glutamate (E) ; sérine (S) ; thréonine (T), ZF : domaine en doigt de zinc, ANK : motif de répétitions ankyrines (Perkins 2007).

d'expliquer l'inductibilité de cette région lorsque les cellules sont stimulées au TNF α (tumor necrosis factor) ou au PMA (phorbol 12-myristate-13-acetate, homologue du diacylglycérol (DAG)) + ionomycine (ionophore calcique), comme le montrent des expériences d'extension d'amorces dans les lignées Jurkat et U937 (C. Van Lint, communication personnelle).

Finalement, il est intéressant de remarquer que le virus HTLV-1 possède une protéine Tax semblable à la protéine Tat du virus HIV-1. Elle est également impliquée dans la régulation de l'efficacité de la transcription du provirus par l'ARNP II. En effet, un promoteur interne du HTLV-1, situé en amont du gène codant pour Tax, permet l'expression de cette protéine clé. Un mécanisme similaire pourrait donc être présent chez le HIV-1 afin de former les premières molécules de Tat qui vont permettre par la suite d'amplifier considérablement la transcription du provirus à partir du LTR-5'. Divers éléments sont en faveur de cette hypothèse, comme des expériences d'extension d'amorces qui ont permis de montrer une activité transcriptionnelle dans cette zone intragénique du HIV-1. Une meilleure caractérisation de cette zone et la compréhension de sa régulation au sein du génome hôte devrait permettre de donner davantage d'informations sur cette zone d'intérêt et sur le mécanisme global de réplication du provirus au niveau transcriptionnel.

I.4. Description des facteurs de transcription étudiés

Suite à la mise en évidence de l'inductibilité de cette région intragénique par le TNF α mais également par le PMA (capable d'activer la PKC suite à son homologie avec le DAG) + ionomycine (ionophore calcique) par extension d'amorces, trois facteurs de transcription pouvant expliquer ce résultat ont été étudiés : NF- κ B, AP-1 et NFAT.

I.4.1. NF- κ B

D'abord décrit comme un facteur de transcription spécifique aux lymphocytes B liant l'enhancer de la chaîne légère des immunoglobulines κ , NF- κ B (Nuclear Factor κ B) est maintenant reconnu comme régulateur ubiquiste contrôlant de nombreux processus comme la réponse immunitaire, l'apoptose, la prolifération et la migration cellulaire. Ce facteur appartient à une famille composée de 5 membres (Fig. 1.17) chez les mammifères : RelA (=p65), RelB, c-Rel, p50 et p52. Les deux derniers membres sont issus des gènes NF κ B1 et NF κ B2 codant pour les précurseurs p105 et p100, respectivement. Les différents membres de cette famille agissent sous la forme d'homo- ou d'hétérodimères, le plus classique étant le dimère RelA-p50, et sont dans la grande majorité des cas des activateurs transcriptionnels même si NF- κ B peut aussi réprimer l'expression de certains gènes (Perkins 2007).

Chaque membre de la famille possède à son extrémité N-terminale une séquence conservée d'environ 300 acides aminés formant le RHD (rel homology domain, Fig. 1.17). Ce domaine permet non seulement la liaison de NF- κ B à l'ADN, mais aussi la dimérisation avec les autres membres de la famille et l'interaction avec la famille d'inhibiteurs de NF- κ B : les I κ Bs. Cependant, seul trois membres, RelA, RelB et c-Rel, possèdent un domaine de transactivation dans leur extrémité C-terminale (Perkins 2007).

En condition basale et dans les autres types cellulaires que les lymphocytes B, NF- κ B est majoritairement localisé dans le cytoplasme. En effet, la séquence NLS présente dans leur RHD est totalement ou partiellement masquée par la liaison des I κ B (inhibitors of κ B), permettant la séquestration du facteur dans le cytoplasme. La famille des I κ Bs contient 4 membres (Fig. 1.17), I κ B α , I κ B β , I κ B γ , et Bcl3 qui possèdent tous des répétitions ankyrines,

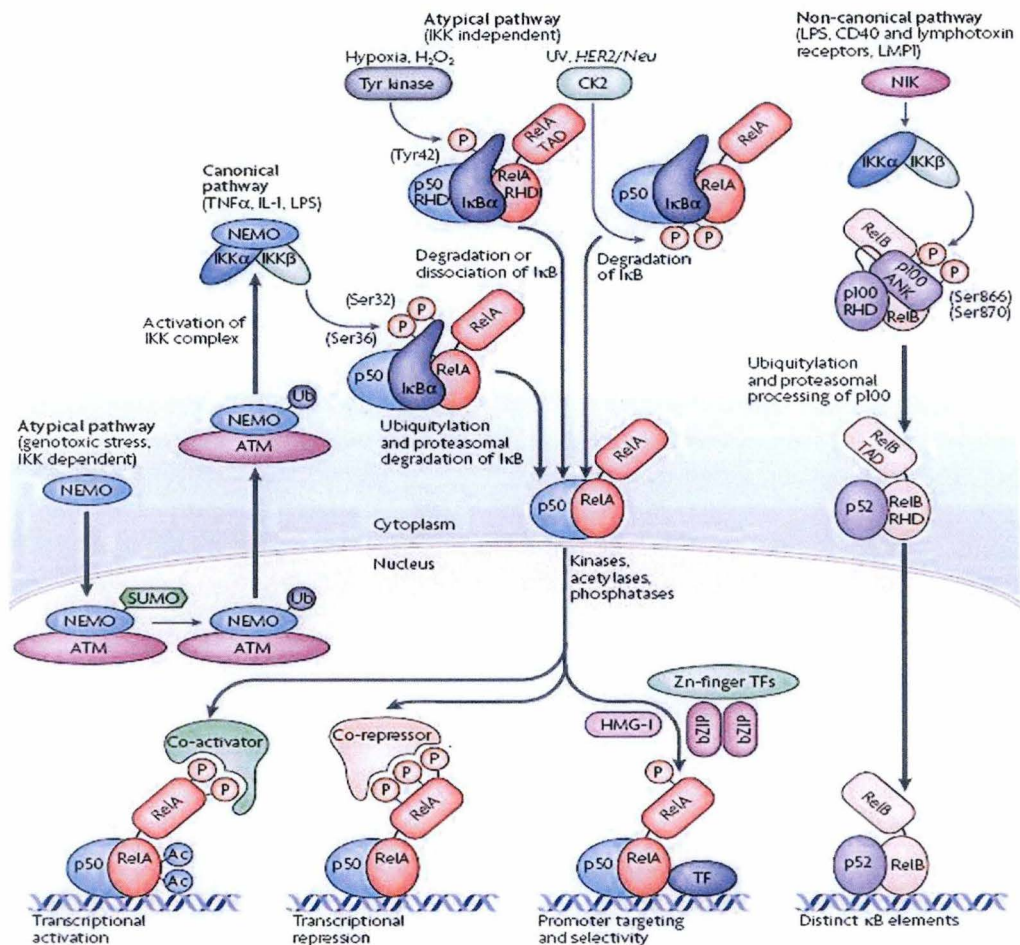


Figure 1.18 : Représentation des différentes voies d'activation de NF-κB. La voie d'activation classique est dépendante de l'activation de IKKβ (l'une des deux sous-unités catalytiques du complexe « signalosome » avec IKKα), phosphorylant IκBα qui va être ubiquitinylé et dégradé par le protéasome, permettant la migration de NF-κB dans le noyau. Différents types de récepteurs peuvent initier la cascade de transduction conduisant à la phosphorylation des IKKs comme le récepteur au TNF, à l'IL-1 ou spécifique aux cellules T. Dans ce dernier cas, la PKC fait partie des intermédiaires conduisant à l'activation des IKKs, expliquant l'inductibilité de NF-κB par le PMA. Dans la voie alternative, un homodimère IKKα va phosphoryler le précurseur p100 qui va être ubiquitinylé et partiellement dégradé par le protéasome afin de former p52 qui va s'associer à RelB. Finalement, deux voies atypiques, l'une dépendante et l'autre indépendante des IKKs, existent. La première est régulée par NEMO (NF-κB essential modifier, =IKKγ, sous-unité régulatrice du « signalosome ») qui va être sumoylé et ubiquitinylé en réponse à des stress cellulaires alors que la seconde repose sur la phosphorylation d'IκBα par la CK2 (Casein Kinase II) ou par des tyrosine kinases en réponse à divers stimuli (Perkins 2007).

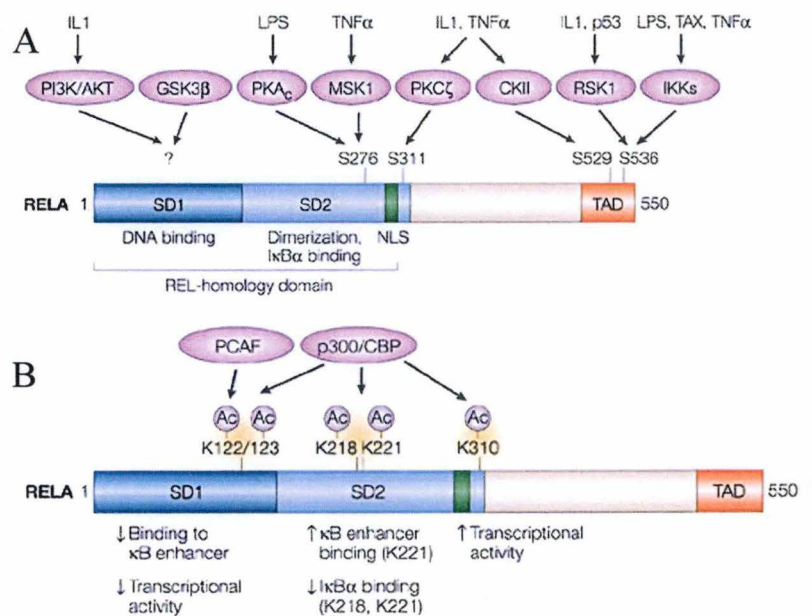


Figure 1.19 : Les stimuli et acteurs responsables des modifications post-traductionnelles sont indiqués pour les différents résidus phosphorylés (A) ou acétylés de RelA (Chen and Greene 2004).

domaine qui permet l'interaction entre protéines et capable de lier le RHD présent sur le facteur NF- κ B. Les précurseurs p100 et p105, possédant également ces domaines, agissent comme les I κ Bs en séquestrant les dimères qui leur sont associés dans le cytoplasme. Lorsque la cellule est stimulée, la phosphorylation des membres de cette famille, suivie par leur ubiquitinylation, va favoriser la dégradation des I κ Bs par le protéasome 26S, découvrant la séquence NLS présente dans le RHD des facteurs séquestrés. Le dimère NF- κ B peut alors migrer dans le noyau où il va reconnaître des éléments κ (consensus : 5'-GGGRNYYYCC-3', dont la correspondance des nucléotides dégénérés est décrite dans la table 1.1) présents dans le promoteur de différents gènes auxquels il va se lier. Une large gamme de stimuli peuvent activer NF- κ B, comme les cytokines proinflammatoires, ce qui explique l'importance des deux sites de liaison présents dans le LTR-5' du virus. Les différentes voies d'activation de NF- κ B sont représentées à la figure 1.18. (Bonizzi and Karin 2004; Perkins 2007)

De nombreux stimuli (cytokines pro-inflammatoires, facteurs de croissances, agents causant des dommages à l'ADN, virus,...) peuvent induire la translocation nucléaire de NF- κ B, qui est également responsable en retour de la régulation d'une large gamme de gènes codant pour des cytokines, chimiokines, protéines anti-apoptotiques, molécules d'adhésion, enzymes, et même I κ B α . La présence de I κ B α parmi les gènes cibles de NF- κ B permet une boucle de rétroaction négative afin de contrôler finement la durée d'activation de l'expression des gènes cibles du facteur de transcription. En effet, I κ B α possède une séquence NLS lui permettant de migrer dans le noyau où il lie NF- κ B et l'exporte dans le cytoplasme grâce à sa séquence NES. En condition basale, la séquence NES semble prendre le dessus sur la séquence NLS, ce qui a pour résultat la présence préférentielle de NF- κ B dans le cytoplasme tant que les I κ Bs ne sont pas phosphorylés et dégradés. Ce système de régulation par localisation subcellulaire de NF- κ B permet un contrôle précis et rapide de la transcription des gènes cibles du facteur (Bonizzi and Karin 2004; Chen and Greene 2004; Perkins 2007).

Outre le contrôle de la localisation subcellulaire de NF- κ B, de nombreuses modifications post-traductionnelles permettent de contrôler soit l'affinité du facteur pour l'ADN, soit sa capacité transactivatrice. La nature de ces modifications est variée : phosphorylation, acétylation, sumoylation, ubiquitinylation, etc. ... Parmi celles-ci, une partie des sites de phosphorylation et d'acétylation de NF- κ B sont repris à la figure 1.19. Cette fois encore, ces modifications post-traductionnelles dépendent d'une large gamme de stimuli, dont la cytokine pro-inflammatoire TNF α ou le PMA induisant par exemple l'acétylation de RelA par p300 (Chen and Greene 2004).

I.4.2. AP-1

Le facteur AP1 (adaptative protein 1) a d'abord été identifié comme une protéine liant le promoteur du virus SV40 (Simian Virus 40), après quoi il a été décrit pour être impliqué dans de nombreux autres processus biologiques comme la prolifération cellulaire ou l'apoptose. Cette famille de facteurs de transcription est formée d'homodimères et d'hétérodimères (Table 1.2) de protéines appartenant à la famille Jun (c-Jun, JunB, JunD), Fos (c-Fos, FosB, Fra1 et Fra2) et ATF (activating transcription factor). D'autres familles de protéines sont connues pour former exclusivement des dimères avec Jun (JDP1, JDP2, Jun dimérisation partner) ou Fos et Jun comme le facteur Maf. Seuls les membres de la famille Jun sont capables d'homodimériser alors que les protéines Fos doivent former obligatoirement des hétérodimères avec les membres des autres familles. Le dimère le plus classique est composé de c-Jun et c-Fos formant un complexe stable, caractérisé par un haut potentiel de transactivation (Shaulian and Karin 2001).

	c-Jun	JunB	JunD	c-Fos	Fos B	Fra1	Fra1
c-Jun	+	+	+	+	+	+	+
JunB	+	+	+	+	+	+	+
JunD	+	+	+	+	+	+	+
c-Fos	+	+	+	-	-	-	-
Fos B	+	+	+	-	-	-	-
Fra1	+	+	+	-	-	-	-
Fra1	+	+	+	-	-	-	-

Table 1.2 : Représentation des dimères possibles entre les membres de la famille Jun et Fos *in vitro* (Castellazzi, M et Sergeant, A., 1993)

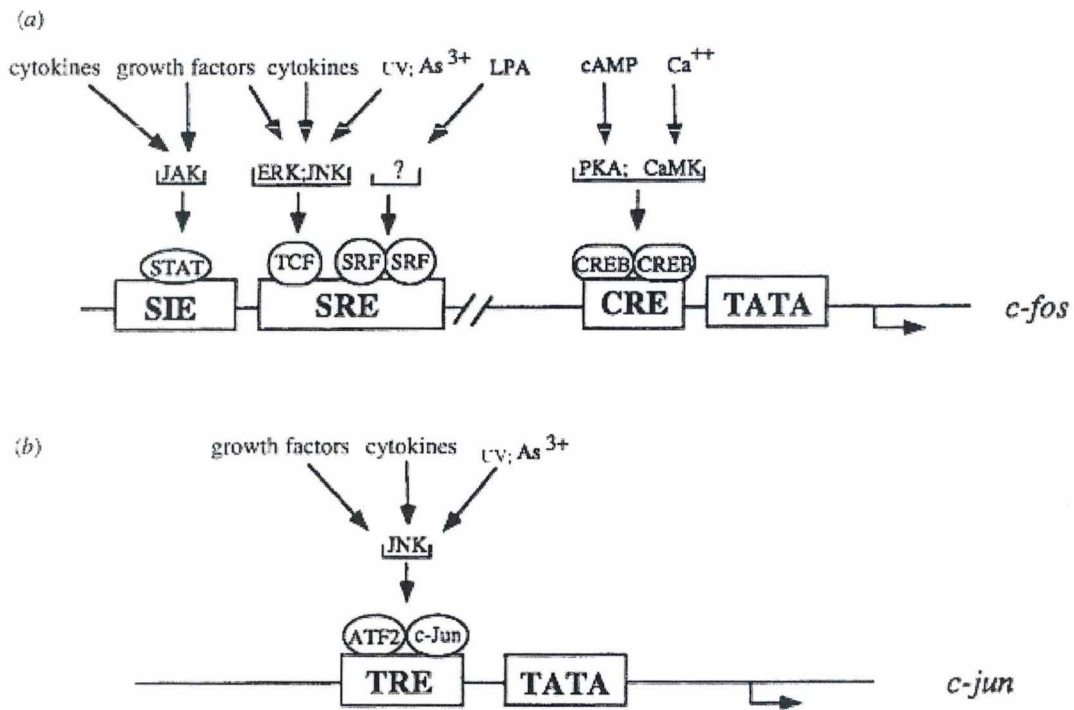


Figure 1.20 : Représentation des éléments *cis* régulant l'expression des gènes *c-fos* et *c-jun* (Karin 1996).

Ces différents membres possèdent tous un domaine bZIP qui permet la liaison à l'ADN mais également l'interaction physique entre les partenaires lors de la formation des dimères. Classiquement, AP-1 reconnaît et lie l'ADN par le biais d'un motif TRE (pour phorbol 12-O-tetradecanoate-13-acetate (TPA) responsive element, 5'-TGAG/CTCA-3'). Cependant, un autre motif, CRE (c-AMP responsive element, 5'-TGACGTCA-3'), peut également être reconnu par les homodimères Jun et les hétérodimères Jun/ATF (Whitmarsh and Davis 1996).

L'activité de AP-1 est régulée au niveau de son expression, tant quantitativement (abondance) que qualitativement (protéines composant le dimère), mais également par le biais de modifications post-traductionnelles (phosphorylation). L'expression des membres de la famille Fos et Jun est régulée par les MAPK (Mitogen-activated protein kinase) sous l'influence d'une large gamme de stimuli comme les facteurs de croissance, les cytokines proinflammatoires, le PMA, les UVs et par l'intermédiaire de Ras ou des rho GTPases.

Dans le cas qui nous intéresse, le $TNF\alpha$ va activer les MAPK JNK (Jun N-terminal Kinase) et p38, tandis que le PMA (homologue du diacylglycérol (DAG) qui est un activateur de la PKC) et l'ionomycine (permettant l'entrée de calcium) vont activer la MAPK ERK (extracellular signal-regulated protein kinase) par une voie Ras-dépendante. Ces kinases activées vont phosphoryler et activer le facteur de transcription Elk-1, membre de la famille de protéines TCF (ternary complex factor), qui contrôle positivement l'expression de c-fos (Fig 1.20). C-Fos pourra alors former des hétérodimères avec des protéines c-Jun présentes à l'état basal, augmentant l'expression de c-jun en liant des éléments *cis* présents dans le promoteur de ce gène. La régulation de l'expression de *c-jun* dépend également de deux éléments TRE liant préférentiellement le dimère c-Jun/ATF2. Cet hétérodimère est lié en condition basale au TRE et c'est suite à la phosphorylation du domaine trans-activateur de ces membres par la JNK que la transcription de *c-jun* va être augmentée (Fig. 1.20) (Karin 1996).

Cette description simplifiée de la régulation d'AP-1 met en évidence que les promoteurs de *c-fos* et de *c-jun* sont constitutivement liés *in vivo* par des facteurs de transcription et que les mécanismes de régulation de l'activité de ces promoteurs reposent sur des modifications post-traductionnelles. Ce système permet, comme pour NF- κ B, une réponse très rapide de la cellule, afin de synthétiser les divers membres de la famille AP-1 qui pourront ensuite réguler la transcription de gènes cibles en réponse au stimulus de départ.

Mis à part le contrôle de l'expression d'AP-1, l'activité de ce facteur est également contrôlée par phosphorylation. Par exemple, la phosphorylation de c-Jun (sérine 63 et 73) par la JNK, permet d'augmenter sa capacité transactivatrice suite au recrutement du coactivateur CBP (CREB binding protein) dans la partie N-terminale de son domaine de transactivation. Des facteurs de croissance, comme l'EGF (epidermal growth factor), activent la kinase FRK (Fyn-Related Kinase) qui va phosphoryler le domaine C-terminal de c-Fos (Karin 1996). ATF2 peut quant à lui être phosphorylé aussi bien par la MAPK JNK que p38. Les conséquences de ces modifications dépendent de la nature du dimère formé et des résidus ciblés. En effet, alors que la phosphorylation de résidus proches du domaine bZIP de c-Jun inhibe la liaison à l'ADN de l'homodimère, elle ne modifiera pas la capacité du dimère c-Fos/c-Jun à lier l'ADN. Par contre, la phosphorylation des Sérine 63 et 73 augmente aussi bien l'activité trans-activatrice de l'homodimère que de l'hétéro-dimère.

I.4.3. NFAT

NFAT (nuclear factor of activated T cells) a été identifié dans les cellules T en tant que facteur liant l'élément distal ARRE (antigen receptor response element) du gène de l'IL-

2. Il est maintenant reconnu comme un facteur inductible dans de nombreux autres types cellulaires. Cette famille de facteurs de transcription comprend 5 membres : NFAT1 (NFATp, NFATc2), NFAT2 (NFATc, NFATc1), NFAT3 (NFATc4), NFAT4 (NFATc3) et NFAT5. Ils appartiennent à la famille des facteurs nucléaires et possèdent des régions conservées communes avec le facteur NF- κ B. En effet, les différents membres de la famille NFAT présentent des régions d'homologie au niveau de leur domaine de liaison à l'ADN, de type RHD comme NF- κ B, et du domaine de régulation. Le RHD possède deux domaines fonctionnels : celui situé du côté N-terminal (RHD-N) permet la liaison à l'ADN alors que le côté C-terminal (RHD-C) est impliqué dans les événements de dimérisation. Alors que le RHD-N semble bien conservé, le RHD-C est beaucoup plus variable. NFAT peut se lier à l'ADN sous forme de monomère, de dimère à des sites dits « κ B-like » et sous forme d'hétérocomplexe avec d'autres facteurs de transcription, le plus connu étant AP-1 (Rao, Luo et al. 1997; Hogan, Chen et al. 2003).

NFAT est constitutivement séquestré dans le cytoplasme et sa migration dans le noyau est majoritairement régulée par le calcium. En effet, en condition basale, le domaine régulateur de NFAT est hyperphosphorylé au niveau de 4 sites conservés riches en sérines et la séquence NLS présente dans le domaine RHD est masquée. La phosphatase calcineurine (Cn) dépendante du complexe calcium/calmoduline permet la déphosphorylation de trois des quatre régions, induisant une modification de la conformation de NFAT, la découverte de la séquence NLS, l'importation dans le noyau et l'augmentation de l'affinité du facteur pour l'ADN. (Fig. 1.21 A). L'activité de la calcineurine nécessite une augmentation soutenue de la concentration intracellulaire en calcium qui a lieu en deux temps (Figure 1.21 B). Tout d'abord, la phospholipase C (PLC) est activée suite à la liaison de récepteur cellulaire et va hydrolyser le PIP2 (phosphatidylinositol biphosphate) en PI3 (phosphatidylinositol 3 phosphate) qui va lier son récepteur situé en surface du réticulum endoplasmique (RE). La concentration intracellulaire en calcium ($[Ca^{2+}]_i$) va augmenter suite au relargage des stocks calciques contenus dans le RE. Cette diminution des réserves calciques va induire l'ouverture des canaux calciques (CRAC) qui vont permettre de garder une $[Ca^{2+}]_i$ élevée. Suite à cette abondance soutenue de calcium, la calcineurine calmoduline-dépendante va être activée et va déphosphoryler les régions riches en sérine du domaine régulateur de NFAT. La calcineurine possède trois isoformes de la sous-unité A catalytique, CnA α , CnA β , CnA χ (dont la contribution respective est dépendante du type cellulaire) et de deux types de sous-unités régulatrices, CnB1 et CnB2 (Rao, Luo et al. 1997; Hogan, Chen et al. 2003).

De nombreuses kinases telles que la CK1 (Casein Kinase 1), GSK3 (Glycogen synthase kinase 3), JNK1 et p38 rephosphorylent NFAT, renvoyant directement dans le cytoplasme. C'est la raison pour laquelle un niveau intracellulaire élevé en calcium doit être soutenu afin que la calcineurine reste activée et déphosphoryle NFAT, provoquant continuellement son importation dans le noyau (Rao, Luo et al. 1997; Hogan, Chen et al. 2003).

La description de ces mécanismes de régulation de NFAT permettent de comprendre pourquoi ce facteur de transcription est activé dans des cellules T incubées en présence d'un ionophore calcique tel que l'ionomycine, qui provoque une élévation soutenue de la concentration calcique intracellulaire. La transcription NFAT-dépendante sera encore accrue en présence de PMA puisque celui-ci induit l'expression du gène codant NFAT, probablement en modifiant post-traditionnellement d'autres facteurs de transcription comme AP-1 permettant d'augmenter la transcription de NFAT (Negulescu, Shastri et al. 1994).

Mise à part sa liaison monomérique ou homodimérique à l'ADN, NFAT possède également la capacité de former des hétérocomplexes avec d'autres facteurs de transcription. Le plus décrit d'entre eux est le dimère NFAT :AP-1 au niveau de sites « composites », où les séquences nucléiques reconnues par les facteurs sont souvent dégénérées et qui peuvent être retrouvées dans les promoteurs de nombreux gènes comme l'IL-2, l'IL-4, le GM-CSF, ... Le point de contact entre les deux facteurs a lieu entre le domaine RHR situé dans la partie N-terminale de NFAT et le domaine bZIP d'AP-1 par des interactions polaires. Ces contacts étant plutôt faibles, le complexe ternaire ne peut être stabilisé en absence d'ADN. La liaison de l'hétérodimère NFAT :AP-1 à ces sites composites est caractérisée par une affinité plus importante que leur liaison individuelle. L'association de ces deux facteurs permet d'intégrer les stimuli contrôlant la voie des MAPKs à celle du calcium pour une régulation plus spécifique des gènes cibles (Macian, Lopez-Rodriguez et al. 2001; Hogan, Chen et al. 2003).

1.5. Objectifs

Comme nous venons de le voir, l'étude des interactions entre acides nucléiques et protéines représente un passage nécessaire vers la compréhension des mécanismes de régulation de l'expression génique. Alors que les techniques classiques comme l'EMSA, le footprinting ou encore le ChIP sont des outils couramment utilisés, ils représentent des approches avec *a priori* qui nécessitent la connaissance d'au moins un des deux partenaires étudiés (facteur de transcription ou séquence consensus).

Lors de ce mémoire, et dans le cadre de la thèse d'Aurélien Tacheny, une technique sans *a priori* d'identification par spectrométrie de masse des protéines interagissant avec une séquence oligonucléotidique d'intérêt est mise au point. Le principe général de la méthode a déjà été établi (figure 1.9) par A. Tacheny utilisant un fragment du LTR-5' du HIV-1 comme oligonucléotide afin de capturer les protéines interagissant avec cette séquence. Cette séquence a été choisie dans le cadre d'une collaboration avec l'équipe de C. Van Lint (ULB), parce qu'elle est très bien décrite et contient de nombreux éléments *cis* connus. Cette méthode a déjà permis d'identifier par spectrométrie de masse le facteur NF- κ B capturé par cette séquence dans un extrait nucléaire de cellules HeLa stimulés par l'IL-1 β .

Bien que certaines mises au point doivent encore être réalisées, notamment afin d'identifier d'autres facteurs de transcription, nous avons voulu, toujours dans le cadre d'une collaboration avec C. Van Lint, appliquer cette méthode à une séquence de grand intérêt encore peu décrite. En effet, la zone d'hypersensibilité identifiée précédemment au sein du gène *pol* du HIV-1 pourrait être impliquée dans les phases précoces de la transcription du provirus. Alors que la liaison de différents facteurs de transcription a déjà été montrée par des techniques classiques, la nature de ces protéines ne permet pas d'expliquer l'inductibilité de cette région à des stimulations par le TNF α ou par le PMA + ionomycine.

L'identification des protéines interagissantes avec cette région pourrait permettre de comprendre l'inductibilité de cette région, et la régulation d'un promoteur interne potentiel susceptible d'expliquer la synthèse des premières molécules de Tat, régulateur crucial de la transcription du provirus.

Les objectifs de ce mémoire sont donc au nombre de deux. Premièrement participer à la mise au point de la méthode développée au laboratoire en étudiant une séquence intragénique d'intérêt du virus HIV-1 encore peu décrite à ce jour, à la recherche de nouveaux candidats potentiels interagissants avec cette séquence. Et deuxièmement, de réaliser en parallèle, une étude, cette fois avec *a priori*, de la liaison potentielle de trois facteurs de transcription (NF- κ B, AP-1 et NFAT) à cette même séquence d'intérêt, dans le but de pouvoir expliquer l'inductibilité de cette séquence au TNF α et PMA + ionomycine.

II. Matériels et Méthodes

II.1. Extractions des protéines nucléaires

II.1.1. Description des types cellulaires utilisés

Trois types cellulaires ont été utilisés lors de ce mémoire :

- Des cellules Hela : lignée de cellules épithéliales provenant d'un adénocarcinome glandulaire cervical (ATCC : CCL-2).
- Des cellules Jurkat : lignée lymphocytaire T provenant de lymphocytes T du sang périphérique d'un individu atteint de leucémie.
- Des cellules U937 : lignée de cellules issue d'un lymphome histiocytique diffus et utilisées comme modèle monocyttaire et macrophagique.

La lignée Hela est utilisée comme contrôle et pour les mises au point de la technique de purification sur bille. Les deux premiers types cellulaires représentent un modèle de choix étant donné qu'ils sont tous deux ciblés par le HIV-1 durant le cycle répliatif du virus.

II.1.2. Culture cellulaire

II.1.2.1. Matériel

- Milieu de culture DHG (Dulbecco Modified Eagle's Medium-High Glucose : 4,5 g/l ;Gibco, UK)
- Milieu de culture RPMI (Roswell Park Memorial Institute ; Gibco, UK)
- Sérum de veau foetal (SFV; Gibco, UK)
- PBS (phosphate buffer saline, Whittacker Belgium)
- Trypsine (0,5 g/l) (Gibco, 25050-014) + EDTA (0,2 g/l) (Gibco, 25300-014)
- Boîte de culture de 75 cm² (T75, Costar, Corning, USA)

Les milieux utilisés pour la culture cellulaire sont stériles et préchauffés à 37°C avant utilisation.

II.1.2.2. Méthode

Culture des cellules adhérentes (Hela) :

Les cellules sont passées tous les deux ou trois jours pour entretenir la culture. Le milieu des boîtes de culture est décanté et le tapis cellulaire est rincé avec 10 ml de PBS stérile. Les cellules sont détachées par une incubation d'une minute à 37°C en présence de 1 ml de trypsine + EDTA. L'activité de la trypsine est arrêtée avec 9 ml de DHG + 10% SFV. La suspension cellulaire est transférée dans un tube pour être centrifugée 7 minutes à 1000 RPM. Alors que le surnageant est décanté, le culot est resuspendu avec du milieu DHG + SFV dans un volume permettant la dilution adéquate (1/4 la semaine et 1/6 le week-end) et la suspension cellulaire est transférée dans une nouvelle boîte de culture. Du milieu DHG + 10% SFV est ajouté pour atteindre un volume final de 15ml par T75.

Culture des cellules en suspension (Jurkat, U937) :

Les cellules sont passées tous les deux ou trois jours. La suspension cellulaire est centrifugée 7 minutes à 1000 RPM. Le surnageant est décanté et le culot est resuspendu dans le volume adéquat de milieu RPMI + 10% SFV, après quoi la suspension cellulaire est répartie dans de

nouvelles boîtes de culture selon la dilution appropriée (1/5 la semaine et 1/7 le week-end). Finalement, le volume est porté à 15 ml avec du milieu RPMI + 10% SFV.

Les cellules sont gardées dans une étuve à 37°C dans une atmosphère humide contenant 5% de CO₂.

II.1.3. Stimulation cellulaire

II.1.3.1. Matériel

- IL-1 β (Interleukine 1 β) dilué à une concentration stock de 1 μ g/ μ l et stocké à -70°C.
- PMA (phorbol myristate acetate, Sigma P8139) dilué à une concentration de 0.5 μ g/ μ l dans du DMSO et stocké à -70°C
- Ionomycine (Sigma IO634) dilué à une concentration de 1mM dans du DMSO et stockés à -70°C.
- TNF α (tumor necrosis factor α , R&D 210-TA) dilué dans du PBS + BSA 0.1% à une concentration de 10 μ g/ μ l et stocké à -70°C).

II.1.3.2. Méthode

Cellules adhérentes :

Le milieu des boîtes est décanté et directement remplacé par du nouveau milieu sans sérum contenant de l'IL-1 β à une concentration de 5ng/ml. Les cellules sont stimulées pendant 45 minutes à 37°C avant d'extraire les protéines nucléaires.

Cellules en suspension :

La suspension cellulaire est centrifugée 7 minutes à 1000 RPM. Le surnageant est décanté et le culot est resuspendu dans la solution de stimulation adéquate : PMA 32 nM + ionomycine 500 nM dilué dans du RPMI pour les cellules Jurkat ou TNF α 10 ng/ml également dilué dans du RPMI pour les cellules U937. Les cellules sont stimulées pendant 5h (PMA + ionomycine) ou 1h (TNF α) dans l'étuve sèche à 37°C dans une atmosphère humide à 5% de CO₂.

II.1.4. Extraction nucléaire

II.1.4.1. Principe

L'extraction protéique nucléaire repose sur l'équilibre osmotique de part et d'autre des membranes cellulaires. Un tampon hypotonique est utilisé afin de fragiliser les membranes cellulaires externes qui sont ensuite claquées. Seuls les noyaux sont gardés et incubés en présence d'un tampon salin permettant de décrocher les facteurs de transcription liés à l'ADN et de faire sortir les protéines à l'extérieur du noyau.

II.1.4.2. Matériel

- PBS (NaCl 150 mM, 10 mM KH₂PO₄ / K₂HPO₄ ; pH 7,4)
- Tampon hypotonique 1X (Hepes 20mM, NaF 5mM, Na₂MoO₄ 10nM, EDTA 100nM)
- Tampon de lyse NP40 0,5% dilué dans du tampon hypotonique
- Tampon de resuspension (RE : Tampon hypotonique, Glycérol 17%)
- Tampon salin (SA : Tampon hypotonique , Glycérol 17%, NaCl 800 mM)
- PBS + NaF 5mM Na₂MoO₄ 1mM
- Tubes LoBind

RE et SA sont fraîchement complétés avec des cocktails d'inhibiteurs de protéases (Complete, Roche) et de phosphatases (Phosphostop, Roche).

II.1.4.3. Méthode

L'ensemble de l'extraction est réalisé sur glace afin de limiter l'activité des protéases et des phosphatases. Le maximum de précautions (gants, tablier) est pris afin d'éviter toutes contaminations en kératine en vue d'une analyse en spectrométrie de masse.

II.1.4.3.1. Cellules adhérentes :

Les boîtes de culture sont décantées et rincées avec du PBS froid, décantées à nouveau et rincées une seconde fois avec du PBS + NaF 5mM et Na₂MoO₄ 1mM. Après ces lavages, les cellules sont incubées 6 minutes dans du tampon hypotonique. Les boîtes sont ensuite décantées et le tapis cellulaire est raclé dans 1,2 ml de tampon de lyse. Les cellules dont la membrane plasmique est lysée sont récoltées dans un microtube pour être centrifugées 30 secondes à 13000 RPM. Le surnageant contenant le matériel cytosolique est éliminé et le culot contenant les noyaux est resuspendu dans un volume de RE variable en fonction de la taille du culot. Un volume de SA, équivalent au volume du culot et du RE additionné, est ajouté. Les noyaux sont incubés au minimum une heure sur roue à 4°C après quoi ils sont centrifugés 10 minutes à 13000 RPM à 4°C. Les surnageants sont regroupés, aliquotés dans des tubes Lobind. La concentration en protéines des extraits nucléaires est estimée par la méthode Bradford (voir II.1.5) avant d'être stockés à -70°C.

II.1.4.3.2. Cellules en suspension :

- Jurkat

Après stimulation, les cellules adhèrent légèrement à la boîte. Elles sont donc préalablement raclées délicatement après quoi elles sont transférées dans des tubes pour être centrifugées pendant 5 minutes à 1000 RPM à 4°C. Le surnageant est décanté, les cellules sont resuspendues dans du PBS, puis centrifugées 5 minutes à 1000 RPM à 4°C. Les tubes sont décantés et le culot est resuspendu dans 5ml de tampon hypotonique, incubé 3 minutes sur glace avant d'être centrifugé 4 minutes à 1000 RPM à 4°C. Le culot de cellules est resuspendu dans 1,2 ml de tampon de lyse. Une incubation de 5 minutes permet de perméabiliser les membranes cellulaires fragilisées lors de l'étape précédente. Une centrifugation d'une minute à 13000 RPM permet de sédimenter les noyaux. Ces derniers sont ensuite resuspendus dans le tampon RE auquel on rajoute ensuite un volume de tampon SA équivalent des noyaux et du RE. Les extraits sont incubés au minimum une heure sur roue à 4°C, après quoi ils sont centrifugés pendant 10 minutes à 13000 RPM et à 4°C. Les surnageants sont finalement regroupés et aliquotés dans des tubes LoBind pour être stockés à -70°C. Leur contenu protéique est évalué par la méthode Bradford (voir II.1.5)

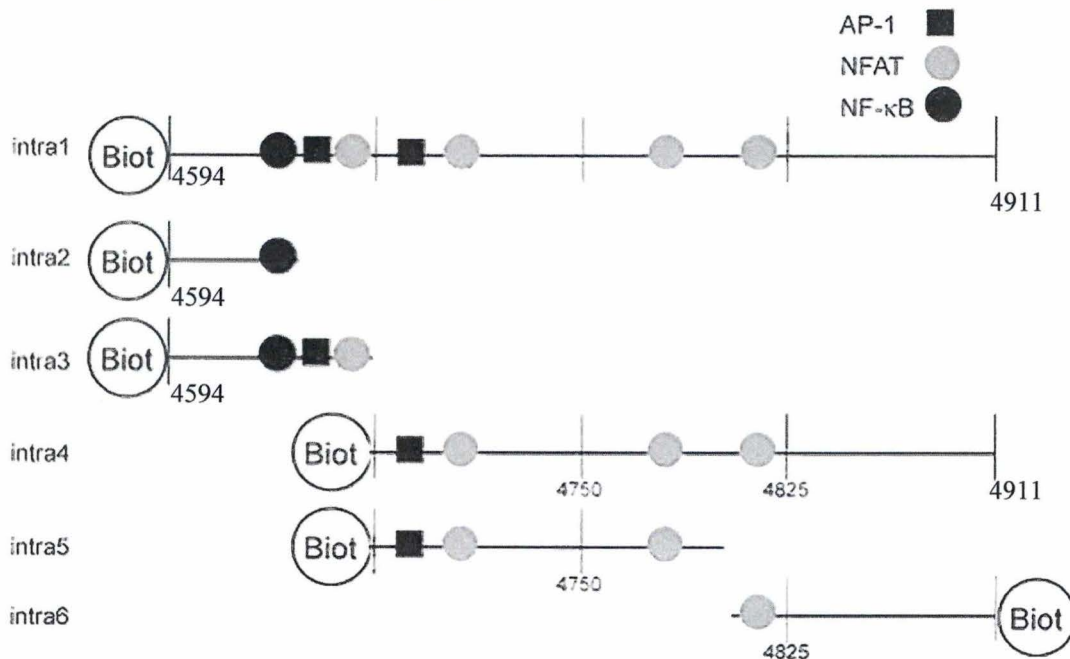
- U397 (méthode Dignam, (Dignam, Lebovitz et al. 1983))

- Tampon A : Hépès pH 4.9 10mM, MgCl₂ 1.5mM, HCl 10 mM, DTT 0.5 mM, inhibiteur de phosphatases 0.5 mM (phosphostop) et inhibiteurs de protéases (Complete, Roche) 1 pastille / 50 ml.
- Tampon C : Hépès pH 4.9 10mM, Glycérol 25%, NaCl 0.42 M, MgCl₂ 1.5mM, EDTA 0.2 mM, DTT 0.5 mM, inhibiteurs de phosphatases 0.5 mM (phosphostop) et inhibiteur de protéases (Complete, Roche) 1 pastille / 50 ml

Table 2.1 : Représentation des différentes sonde utilisées lors des expériences d'ELISA-like.

La numérotation des nucléotides est réalisée en considérant que le premier nucléotide est le site +1 de l'U3 du LTR-5'. Les amorces sens (excepté l'intra6 pour lequel c'est l'anti-sens) sont modifiées en 5' par ajout d'une biotine (pour les expériences d'ELISA-like) ou d'une desthiobiotine (pour la purification sur billes magnétiques).

Séquence	Nucléotide correspondant	Amorces (Sens/Anti-sens)
Intra1	4594-4911	S : 5'-CCAGTACTACAGTTA-3' AS : 5'-TCGGGTTTATTACA-3'
Intra2	4594-4667	S : 5'-CCAGTACTACAGTTA-3' AS : 5'-CCTACAATCCCCAA-3'
Intra3	4594-4674	S : 5'-CCAGTACTACAGTTA-3' AS : 5'-GGAGTAATAGAATCT-3'
Intra4	4674-4911	S : 5'-GGAGTAATAGAATCT-3' AS : 5'-TCGGGTTTATTACA-3'
Intra5	4674-4800	S : 5'-CCAGTACTACAGTTA-3' AS : 5'-ACTGTACCCCCCAAT-3'
Intra6	4800-4911	S : 5'-ATTGGGGGTACAGTG-3' AS : 5'-TCGGGTTTATTACA-3'
I3_NFAT*	4594-4674	S : 5'-CCAGTACTACAGTTA-3' AS : 5'-AGATTCTATTAGGCCTTGACT-3'
I3_AP1*	4594-4674	S : 5'-CCAGTACTACAGTTA-3' AS : 5'-AGATTCTATTACTCCTTAACC-3'
I3_AP1*NFAT*	4594-4674	S : 5'-CCAGTACTACAGTTA-3' AS : 5'-AGATTCTATTAGGCCTTAACC-3'

**Figure 2.1: Représentation des différents oligonucléotides utilisés pour les expériences de capture des facteur de transcription avec les différents sites putatifs prédits *in silico*.**

Les différents sites putatifs ont été détectés à l'aide du module MATCH de Transfac (<http://biobase-international.com>) et du logiciel Matinspector de Genomatix (<http://www.genomatix.de>)

- Tampon D : Hépès pH 4.9 10mM, Glycérol 20%, KCl 0.1 M, EDTA 0.2 mM, DTT 0.5 mM, inhibiteurs de phosphatases 0.5 mM (phosphostop) et inhibiteur de protéases (Complete, Roche) 1 pastille / 50 ml

Environ 10^9 cellules à une densité de 0.5 à 1 million par ml sont nécessaire pour ce mode d'extraction.

Les cellules sont tout d'abord centrifugées 10 minutes à 1200 RPM et à 4°C afin d'éliminer le milieu. Les cellules sont ensuite lavées dans 90 ml PBS et centrifugées 10 minutes à 1200 RPM. Cette étape est réalisée une seconde fois avec 10 ml de PBS. Les cellules sont resuspendues dans un volume de tampon A correspondant à 5 fois le volume estimé du culot cellulaire, incubées pendant 10 minutes sur glace et centrifugées 10 minutes à 1200 RPM à 4°C. Les cellules sont resuspendues dans un volume de tampon A d'environ deux fois le volume du culot et homogénéisées grâce à un Dounce (environ 25 aller-retour avec un piston de type B). La lyse cellulaire est estimée grâce à une coloration au bleu de trypan. Les cellules sont ensuite centrifugées 15 minutes à 4°C à 1500 RPM, le surnageant est décanté et le culot et resuspendu dans le tampon C pour être transféré dans un Dounce dans lequel une vingtaine d'aller retour du piston sont effectués. Les extraits sont alors mélangé doucement pendant 30 minutes à 4°C avant d'être centrifugés 30 minutes à 15000 g à 4°C. Les surnageants sont récupérés, assemblés et dialysé durant tout la nuit à 4°C dans le tampon D. Finalement, les extraits cellulaires sont centrifugés 30 minutes à 15000 g à 4°C afin d'éliminer les débris cellulaires. Les surnageants sont aliquotés, stockés à -70°C et le contenu protéique est estimé par la méthode Bradford.

II.1.5. Dosage par la méthode de Bradford (Bradford 1976)

Les protéines sont dosées selon la méthode Bradford. Pour ce faire, 2 μ l d'échantillon sont ajoutés à 1 ml de solution Bradford (Bio-rad Protein Assay 500-0006) diluée 5 fois dans de l'eau. L'absorbance est lue au spectrophotomètre après 5 minutes d'incubation à une longueur d'onde de 595 nm. La quantité de protéine est estimée grâce à un étalon (BSA) dont la concentration est connue. A partir de l'absorbance de cet étalon il est possible de déduire la concentration de l'échantillon par la formule suivant :

$$\frac{(\text{Moy éch} - \text{Moy blc}) / (\text{Moy Et} - \text{Moy blc Et}) \times 5}{\text{Volume de l'échantillon}}$$

Avec Et = étalon, éch = échantillon. Le blanc échantillon (blc) correspond à la solution dans laquelle il se trouve alors que le blanc étalon (blc Et) correspond à l'eau.

II.2. Amplification PCR

II.2.1. Description des séquences utilisées

Les séquences utilisées ainsi que les primers nécessaires à leur amplification sont décrits dans la table 1. Les sondes sont amplifiées à partir d'un plasmide (pCV10) donné par l'équipe de C. Van Lint.

Toutes les amorces sens sont modifiées en 5' par ajout d'une biotine (sondes pour l'ELISA-like, Eurogentec) ou d'une desthiobiotine (sonde oligonucléotidique pour la purification sur billes, IBA).

Toutes les amorces anti-sens possèdent une cyanine 3 à leur extrémité 3' qui permet, lors des mises au point, d'estimer par fluorescence la liaison des séquences oligonucléotidiques aux billes magnétiques.

Deux autres séquences ont été utilisées lors des expériences de compétition par ELISA-like, toutes deux provenant d'une région du génome du HIV-1 et pour lesquelles aucun site de liaison pour les facteurs de transcription étudiés n'a pu être mis en évidence. La première (CN1, contrôle négatif 1) possède le même nombre de nucléotide que la séquence intra 1 alors que la seconde (CN2) a le même pourcentage en GC que cette dernière.

II.2.2. Méthode

La préparation des mélanges PCR est réalisée dans deux pièces séparées. Tout d'abord, en pièce PCR1, l'eau distillée, le tampon 1X (Promega), les amorces (0,5 μ M, sens + anti-sens), les dNTPs (200 μ M, Eurogentec) et la polymérase GoTaq (2,5U, Promega #M2031) sont dispensés dans les différents tubes PCR et portés à un volume final de 90 μ l avec de l'eau. Seul l'ADN (10 ng de plasmide dans un volume de 10 μ l) est ajouté en pièce PCR2. Deux blancs sont également réalisés, contenant l'ensemble du mélange PCR, excepté l'ADN qui est remplacé par l'eau utilisée en pièce PCR 1 ou PCR2. L'amplification a lieu sur base du programme suivant :

- 5 minutes à 94°C
- 40 fois le cycle suivant : 30 secondes à 94°C, 30 secondes à la température d'hybridation correspondant aux amorces utilisés; 30 secondes à 72°C
- 10 minutes à 72°C
- En attente à 4°C

Après amplification, 2,5 μ l de produits PCR sont utilisés pour vérifier la bonne amplification de l'amplicon par migration sur gel d'agarose 1% en présence de bromure d'éthidium pour visualiser l'ADN.

II.3. Purification des produits PCR

Le but de cette étape est de ne garder que l'amplicon désiré, de concentrer le produit PCR et d'éliminer tout composant indésirable comme les dimères d'amorces qui, s'ils possèdent une biotine ou une desthiobiotine, pourraient également se lier à la streptavidine et fausser le test. Pour ce faire, le kit "Clean-Up with Wizard®SV for Gel and PCR (Promega)" a été utilisé et se base sur la capacité de l'ADN (40 μ g maximum par colonne d'une taille de 100 bp à 10kb) à lier la membrane de silice présente au fond des microcolonnes en présence de sels chaotropiques.

II.3.1. Méthode

400 μ l de solution de liaison à la membrane, contenant des sels chaotropiques, sont ajoutés à 400 μ l d'amplicon. 400 μ l de cette solution sont déposés sur une microcolonne, incubés pendant une minute puis centrifugés à 8000 RPM pendant 1 minute. Cette première étape est répétée une seconde fois avec les 400 μ l restants. Les microcolonnes sont ensuite rincées avec 700 μ l de solution de lavage et centrifugées 1 minute à 8000 RPM. Cette étape est répétée avec 500 μ l de solution de lavage et une centrifugation de 5 minutes à 8000 RPM, suivie d'une deuxième centrifugation de 1 minute à 8000 RPM afin d'éliminer au maximum la solution de rinçage. L'ADN est ensuite incubé 1 minute avec 25 μ l d'eau sans nucléases et récupéré par centrifugation (1 minute à 8000 RPM). Cette étape est répétée une seconde fois.

Table 2.2 : Composition des tampons de liaisons et de lyses pour les différents facteurs de transcription étudiés.

	Tampon de liaison (Tp B)	Tampon de lyse (Tp L)	Rapports utilisés pour les dosages d'activités de liaison à l'ADN en multipuits
NF- κ B	<ul style="list-style-type: none"> - Hepes 4 mM, - KCl 120 mM - Glycerol 8% - BSA 1% - ADN de sperme de saumon (0.01 mg/ml) - DTT 2 mM 	<ul style="list-style-type: none"> - Hepes 20 mM - NaCl 350 mM - MgCl₂ 1 mM - EDTA 0.5 mM - Glycérol 20%, NP-40 1% - DTT 5 mM - Cocktail d'inhibiteurs de protéases (Complete, Roche) 	30/20
NFAT	<ul style="list-style-type: none"> - Tris HCl 10 mM - MgCl₂ 1mM - EDTA 0.5 mM - NaCl 50 mM - glycerol 4% 	<ul style="list-style-type: none"> - Glycerol 17 %, - Hepes 40 mM - NaF 10 mM - Na₂MoO₄ 0.02 mM - EDTA 0.2 mM - NaCl 400 mM - DTT 1 mM - Cocktail d'inhibiteurs de protéases (Complete, Roche) - cocktail d'inhibiteurs de phosphatases (Phosphostop, Roche). 	40/10
AP-1	<ul style="list-style-type: none"> - Hepes 3 mM - EDTA 0.3 mM - NaCl 8 mM - Glycérol 12% - DTT 1 mM - Gloria 1%. 	<ul style="list-style-type: none"> - Glycerol 17 % - Hepes 40 mM - NaF 10 mM - Na₂MoO₄ 0.02 mM - EDTA 0.2 mM - NaCl 400 mM - DTT 1 mM - Cocktail d'inhibiteurs de protéases (Complete, Roche), - Cocktail d'inhibiteurs de phosphatases (Phosphostop, Roche). 	30/20

Après purification, la présence de l'amplicon purifié est vérifiée une seconde fois par migration sur gel d'agarose 1%, sa concentration est estimée en mesurant l'absorbance de l'échantillon grâce à un spectrophotomètre à une longueur d'onde de 260 nm. Les produits PCR sont conservés à 4°C pour les sondes desthiobiotinylées et -20°C pour les sondes biotinylées.

II.4. Dosage de la liaison à l'ADN par Elisa-like (TransAM)

II.4.1. Principe

Le test TransAM permet de mettre en évidence la liaison d'un facteur de transcription particulier à une séquence nucléique d'intérêt, immobilisée dans des puits. La détection de la liaison s'effectue suivant le principe ELISA grâce à un anticorps primaire et secondaire couplé à la peroxydase (voir figure 1.8) (Renard, Ernest et al. 2001)

II.4.2. Matériel

- Plaque 96 puits recouvert de streptavidine (Roche)
- PBS 150 (NaCl 150 mM, 10 mM KH₂PO₄ / K₂HPO₄ ; pH 7,4)
- PBS 50 (NaCl 50 mM, 10 mM KH₂PO₄ / K₂HPO₄ ; pH 7,4)
- Tétraméthylbenzidine (TMB, Tebu-bio Laboratories, #TMB100-0100)
- HCl 1N
- Trappeurs biotinylés, tampon et anticorps utilisés voir tableau 2.1, 2.2 et 2.3, respectivement.

II.4.3. Méthode

Tout d'abord, les sondes biotinylées sont immobilisées dans des plaques 96 puits. Pour ce faire, 5 picomoles d'ADN dont la séquence contient le site de liaison spécifique au facteur de transcription étudié sont déposées dans les puits tapissés de streptavidine et incubés une heure à 37°C. Les plaques sont ensuite rincées avec du PBS 150 + Tween 0,1 % puis avec de l'eau avant d'être séchées et stockées à 4°C.

Les extraits nucléaires (10 ou 20 µg dilués dans le tampon de lyse adéquat) sont incubés dans les puits, sous faible agitation, à température ambiante et pendant une heure. Cette étape a lieu en présence d'un rapport tampon de liaison/ tampon de lyse spécifique au facteur de transcription étudié (table 2.2) dans un volume final de 50 µl. Les puits sont ensuite lavés 3 fois avec du PBS 50 + tween 0,1% et incubés une heure à température ambiante avec l'anticorps primaire adéquat (table 2.3). Après trois nouveaux lavages au PBS 50 + tween 0,1%, un anticorps secondaire correspondant couplé à la peroxydase (tableau 2.3) est incubé une heure à température ambiante. Après 4 lavages au PBS 50 + 0,1% tween, l'activité de la peroxydase est évaluée par colorimétrie, en l'incubant en présence du substrat TMB pendant 10 min dans le noir à température ambiante, après quoi la réaction est arrêtée avec de l'HCl 1N. Finalement, l'activité est estimée en mesurant l'absorbance 450 nm.

II.5. Purification sur bille

II.5.1. Principe

Voir introduction point I.2.2.2 et figure 1.9

Table 2.3 : Anticorps utilisés pour détecter les différents facteurs de transcription lors des expériences d'ELISA-like.

	Anticorps primaire	Anticorps secondaire
NF- κ B	<ul style="list-style-type: none"> - IgG de lapin anti-p65 (Santa Cruz: SC372) - IgG de lapin anti-p50 (Santa Cruz : SC-7118) - IgG de lapin anti-p52 - IgG de lapin anti-c-rel - IgG de lapin anti-rel-B <p>Dilution : 1000 fois dans du PBS 50 + 1% gloria</p>	<p>IgG de chèvre anti- IgG de lapin, couplé à la peroxydase</p> <p>Dilution 1000 fois dans du PBS 50 + 1% gloria</p>
NFAT	<p>IgG de souris anti-NFATc1 (Santa Cruz SC-7294)</p> <p>Dilution : 1000 fois dans du PBS 50 + 1% BSA</p>	<p>IgG de chèvre anti- IgG de souris, couplé à la peroxydase</p> <p>Dilution 1000 fois dans du PBS 50 + 1% gloria</p>
AP-1	<p>IgG de lapin anti- :</p> <ul style="list-style-type: none"> - c-Fos (Santa Cruz: SC-7205) - Fos-B (Active motif #100529) - c-Jun (Santa Cruz: SC-1694) - Jun-B (Active motif #100677) - Jun-D (Active motif #100676) - Fra1 (Active motif #100969) - Fra2 (Active motif #100529) <p>IgG de souris anti- :</p> <ul style="list-style-type: none"> - c-jun phosphorylé (Santa-cruz: SC-822) <p>Dilution : 1000 fois dans du PBS 50 + 1% BSA</p>	<p>IgG de chèvre anti- IgG de (couplé à la peroxydase) :</p> <ul style="list-style-type: none"> - Lapin - Souris <p>Dilution 1000 fois dans du PBS 50 + 1% BSA</p>

II.5.2. Matériel

Le matériel utilisé lors de la purification des facteurs de transcription ne doit pas interférer avec l'analyse en spectrométrie de mass. Cela signifie par exemple qu'il faut éviter des sels.

- Tube assurant une faible adsorption des protéines sur ses parois (Tube Lobind 1,5 ou 2,0 ml, #Z666505, Eppendorf)
- PBS 50 (NaCl 50mM, 10 mM KH₂PO₄ / K₂HPO₄ ; pH 7,4)
- Billes magnétique tapissées de streptavidine (Dynal M280-streptavidin, #112-05)
- Tampon de liaison (voir table 2.2)
- Tampon de lyse (voir table 2.2)
- H₂O (Biosolve)
- Tween 20
- Biotine
- NH₄HCO₃ 50 mM (Sigma A-6141)
- Silent Superfactant (PPS) (Protein-discovery, 21011)
- DTT 0,5M (Sigma D-0632)
- Iodoacétamide (Sigma I-1149)
- CaCl₂ 1mM
- Trypsine Gold (Promega V511A)
- HCl fumant 12 N

II.5.3. Méthode

Tout est réalisé en salle blanche afin de minimiser les contaminations en kératine et d'homogénéiser les conditions de température et d'humidité dans lesquelles la manipulation est réalisée.

Un mg de billes magnétiques, à une concentration de 10 µg/µl, est rincé 6 fois avec 100 µl de PBS 50. Pour ce faire, un aimant permet de cocentrer les billes magnétiques dans le fond du microtube. Alors que le surnageant est décanté, les billes sont resuspendues dans la solution de lavage.

20 pm de sonde oligonucléotidique (table 2.2), dilués dans du PBS 50 à volume final de 100 µl, sont ensuite incubés en présence des billes pendant une heure à température ambiante et sous légère agitation (roue). Après avoir décanté la solution d'oligonucléotides qui ne se sont pas liés aux billes, trois lavages sont réalisés avec respectivement 350, 200 et 200 µl de PBS 50.

Les extraits nucléaires sont préincubés pendant 15 minutes sur glace dans une solution de tampon de liaison/tampon de lyse. Cette étape permet aux protéines de se lier éventuellement aux ADN compétiteur présents dans le tampon de liaison. Si des protéines recombinantes sont incubées, aucune préincubation n'est effectuée et ces dernières sont directement incubées en présence de la sonde nucléique liée aux billes magnétiques. Les extraits nucléaires sont ensuite incubés 1h sur roue à température ambiante avec les billes magnétiques auxquelles l'oligonucléotide s'est lié à l'étape précédente. Les différentes protéines vont lier les sites de liaison qui leur sont spécifiques s'ils sont présents sur la sonde.

Après incubation, le surnageant est décanté, et les complexes billes-trappeurs-oligonucléotidiques sont rincés une première fois pendant 2 minutes sur roue avec 500 µl de tampon de liaison. Ce dernier est suivi par 3 lavages de 2 minutes sur roue avec 1ml de PBS 50 + tween 0,1%. Le surnageant est décanté et les billes sont resuspendues dans 200 µl de

NH_4HCO_3 50 mM et transférées dans un nouveau microtube. Les billes sont rincées une dernière fois avec 100 μl de NH_4HCO_3 , avant d'être incubées 2h sur roue et à température ambiante avec 30 μl de biotine 5mM diluée dans du NH_4HCO_3 . Lors de cette étape, les séquences nucléiques auxquelles sont liées les protéines vont être déplacées des billes magnétiques grâce à l'apport d'un excès de biotine, possédant une plus grande affinité pour la streptavidine que la desthiobiotine. Le surnageant, contenant les protéines liées à la sonde nucléique déplacée des billes, est transféré dans un microtube. Les billes sont rincées avec 10 μl de solution de biotine 5mM dont le surnageant est ajouté aux 30 μl précédemment transféré dans le microtube.

Ces 40 μl de surnageant sont gardés afin d'être préparés avant la digestion trypsique. Les protéines sont d'abord dénaturées grâce à un détergent compatible avec la spectrométrie de masse, le Silent Superfactant 0,75%, et chauffées pendant 5 min à 100°C. Les ponts disulfures sont réduits avec 0,5 μl de DTT 0,5M pendant 30 min à 50°C et sous une agitation de 300 RPM. Afin que les protéines restent sous forme réduite, elles sont carbamidométhylées 30 min dans le noir à température ambiante avec 4,7 μl d'iodoacétamide 750 mM dilué dans du NH_4HCO_3 . Finalement, les protéines sont digérées au niveau des arginines et des lysines avec une quantité de trypsine adaptée aux conditions expérimentales : 300 ng en cas de protéine recombinante seule ou 900 ng pour les extraits nucléaires et en présence 0,5 μl de CaCl_2 , afin de favoriser l'activité de la protéase. Le lendemain, 1 μl d'HCl fumant 12 N est ajouté à chaque échantillon pendant 45 min à 37°C. L'acidification provoque la dégradation du détergent qui va alors précipiter. Il est ensuite éliminé ainsi que les éventuelles billes magnétiques résiduelles, grâce à une centrifugation de 10 minutes à 13000 rpm à 4°C.

II.6. Séparation des échantillons par HPLC-SCX

Afin de décomplexifier les échantillons pour l'analyse en spectrométrie de masse, les peptides sont séparés par HPLC-SCX (High Performance Liquid Chromatography Strong Cation Exchange). Pour ce faire, les échantillons réalisés en double sont rassemblés après élimination du détergent, utilisé pour la dénaturation des protéines, par centrifugation. Etant donné qu'un volume d'environ 200 μl est nécessaire pour l'HPLC, approximativement 120 μl d'acétonitrile 30% (pH 2.6) est ajouté aux 80 μl d'échantillon. Le gradient salin croissant permettant de décrocher progressivement les peptides est réalisé à partir d'un mélange d'une solution A d'acétonitrile 30% pH 2.6, permettant également d'équilibrer la colonne avant la séparation, et d'une solution B composée d'acétonitrile 30% et de formate d'ammonium 540mM pH 4.5. Avec un débit de 200 $\mu\text{l}/\text{min}$, les peptides progressivement désorbés sont collectés dans un nouveau puits tout les 90 secondes permettant de générer une quarantaine de fractions au final. Le gradient croissant en concentration de formate d'ammonium, permettant un décrochage progressif des peptides, suit une courbe exponentielle durant 50 minutes en démarrant avec un mélange contenant 100% de solution A pour finir avec 100 % de B. Durant les 10 dernières minutes, la colonne est rincée avec 100% de B.

Etant donné qu'analyser les 40 fractions par spectrométrie de masse demanderait énormément de temps, ces dernières sont rassemblées : les 5 premières et les 5 dernières ensembles et toutes les autres par deux. Avant l'analyse en spectrométrie de masse, il faut encore éliminer le formate d'ammonium qui était présent dans le tampon B, utilisé pour le gradient de l'HPLC-SCX. Celui à la propriété unique d'être volatil et peut être évaporé par « SpeedVac ». Après évaporation, le volume restant des fractions est d'environ 40 μl , dont le pH est vérifié et, s'il est supérieur à 2 (pH acide nécessaire pour l'ionisation), est acidifié à l'aide d'acide formique 5%.

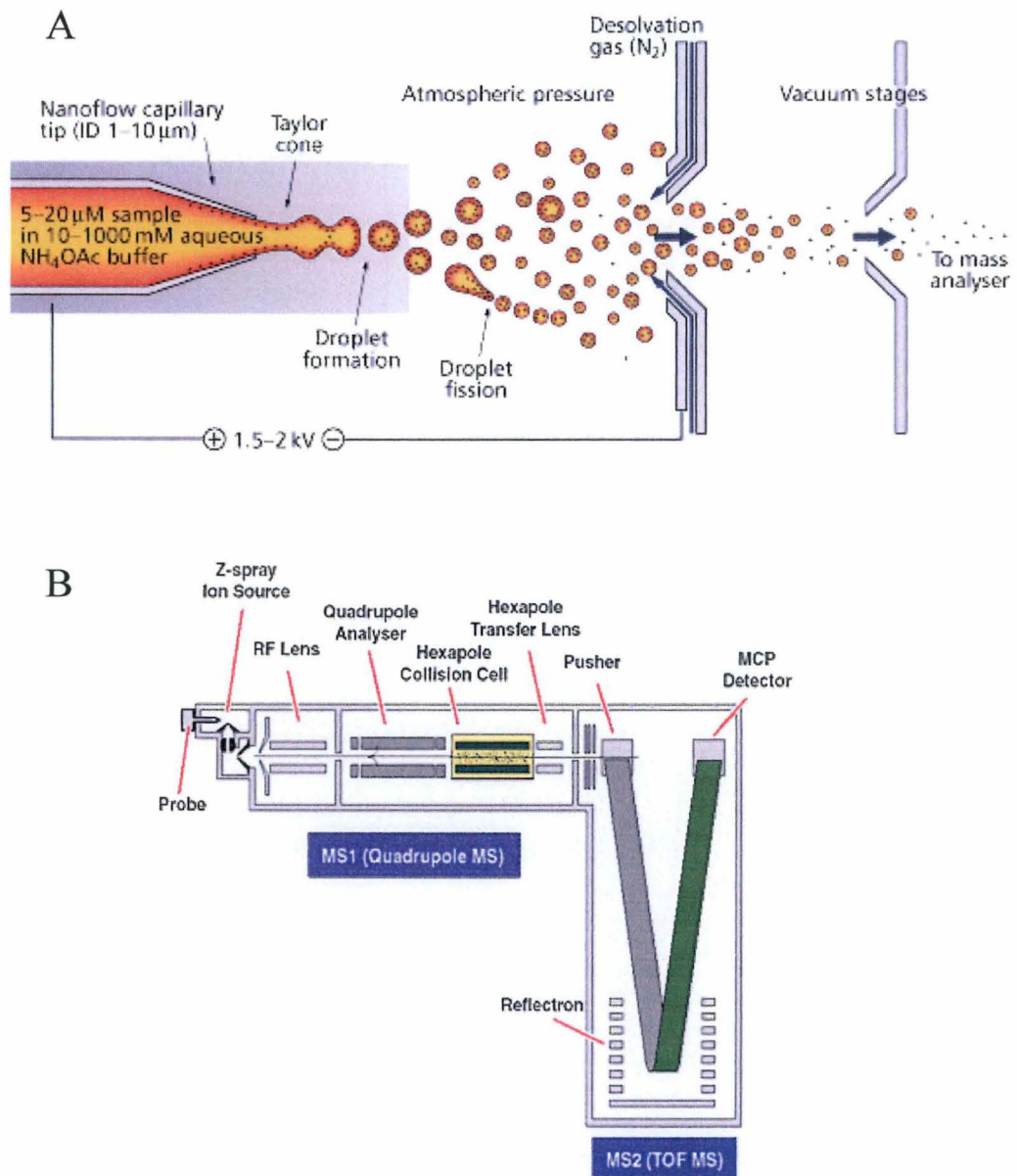


Figure 2.2 : Représentation des éléments d'un ESI-Q-TOF.

(A) Principe de l'ionisation par electrospray (<http://www.rsc.org/>). (B) Schéma de la structure générale d'une spectromètre de masse en tandem (<http://www.whri.qmul.ac.uk/>).

II.7. Analyse en spectrométrie de masse en tandem

II.7.1. Principe de la spectrométrie de masse

Le spectromètre de masse peut être divisé en trois parties : une source, un analyseur (séparateur) et un détecteur. La source permet d'ioniser l'analyte et de le mettre en phase gazeuse, deux caractéristiques indispensables pour pouvoir utiliser les analyseurs. Si l'ionisation des macromolécules biologiques est une chose aisée (par exemple par variation de pH), leur mise en phase gazeuse est beaucoup plus compliquée et représente l'étape qui a longtemps limité l'utilisation de la spectrométrie de masse dans le domaine de la biologie. C'est grâce à l'arrivée du MALDI (Matrix Assisted Laser Desorption Ionisation) et de l'ESI (Electrospray Ionisation), deux types de sources majoritairement représentées à l'heure actuelle, que la spectrométrie de masse est devenue un outil incontournable pour l'identification de composés biologiques. La seconde structure clé est l'analyseur, servant à séparer les peptides ionisés selon leur rapport masse sur charge (m/z) qui sera mesuré par le détecteur. Divers types d'appareillages, basés sur des principes de fonctionnement différents, existent pour chacune de ces trois structures. Nous focaliserons notre description sur celles utilisées lors de ce mémoire : une source de type ESI et un séparateur de type Q-TOF (quadripôle -temps de vol). En amont du spectromètre de masse proprement dit se trouve une micro-colonne de chromatographie en phase inverse possédant des chaînes lipidiques à 18 carbones. Elle permet une séparation des peptides en fonction de leur hydrophobicité, ceux-ci étant désorbés progressivement grâce à un gradient croissant d'acétonitrile.

Avec l'ESI, le passage en phase gazeuse est réalisé par évaporation d'un analyte en solution (Fig. 2.2A). En bref, l'échantillon est envoyé vers un fin capillaire porté à haut potentiel électrique, permettant la formation d'un cône de Taylor qui va s'étirer pour finalement former un spray de gouttelettes chargées positivement. Un flux d'azote favorise l'évaporation et la diminution de la taille des gouttelettes formées. Alors que le diamètre des gouttelettes va diminuer progressivement le nombre de charges par gouttelette va rester constant. Étant donné que la charge maximale portée par une goutte dépend directement de son diamètre, un seuil va être atteint pour lequel la répulsion entre les charges devient trop importante et provoque une explosion Coulombienne formant plusieurs gouttelettes filles de diamètre inférieur. Cette réaction se produit plusieurs fois, pour finalement donner un ion généralement multichargé qui est emmené vers l'analyseur grâce à une interface constituée de deux gradients : de pression et de potentiel. (Mémoire de thèse de J. VINH)

Le Q-TOF est un système hybride couplant deux types d'analyseurs : un quadripôle et un « temps de vol » (Fig. 2.2 B). Le principe de mesure de la masse par le TOF repose sur la relation entre le temps de vol, la vitesse (énergie cinétique) et le rapport m/z d'un ion dans des conditions de haut vide. Le temps de vol d'ions, préalablement accélérés, dépendant de leur m/z est mesuré dans un tube sous vide et en absence de champs magnétique. Outre le mode linéaire où la masse des ions est directement mesurée, la résolution, en mode réflectron, est augmentée en focalisant en un même point les ions de m/z identique mais d'énergie cinétique différente.

En amont du TOF se trouve un quadripôle formé de 4 barres d'aciers parallèles soumises à des différences de potentiel 2 à 2, de signes opposés entre chaque paire. Le fonctionnement du quadripôle peut être décrit en deux temps. Tout d'abord, en mode MS, tout ion, quel que soit son rapport m/z , pourra traverser au niveau de cet analyseur pour être séparé par le TOF, ce qui donnera le spectre de l'abondance des différentes m/z mesurées. Une fois en mode MS/MS, le quadripôle agit comme un filtre ne laissant passer que les ions de rapport m/z

déterminés. En fonction de la fréquence et des variations de tension appliquées au niveau des barres d'acier du quadripôle, seuls les ions des rapports m/z déterminé auront une trajectoire stable permettant de passer le quadripôle, alors que les autres seront déviés de la trajectoire ou percuteront les barres d'aciers. Chaque peptide filtré par le quadripôle est fragmenté suite à la cassure de leurs liaisons covalentes par un gaz inerte et grâce à l'énergie apportée au niveau de la cellule de collision située entre le quadripôle et le TOF. Dans notre cas, il s'agit d'une fragmentation de type CID (collision induced dissociation): l'énergie apportée permet de transférer des charges vers le lien peptidique qui va être clivé. Les peptides fragmentés, de plus petites tailles, sont séparés au niveau du TOF, et, la mesure de leur rapport m/z permet de connaître la séquence primaire du peptide précurseur. (Mémoire de thèse de J. VINH)

II.7.2. Matériel

LC (NaNoEase, Atlantiss dC18, 75 μ m by 150mm, Cat. 186002197, Waters)

ESI-Q-TOF II (Micromass, UK)

Solvant A : 95% H₂O (Biosolve) + 5% Acétonitrile (Biosolve)

Solvant B : 100% ACN + 0,1% acide formique

Solvant C : 99% H₂O + 0,1% acide formique

II.7.3. Méthode

Durant les 10 premières minutes d'injection, l'échantillon est transféré vers une précolonne, par un flux de solution C avec un débit de 4.5 μ l/minute. Cette première étape permet de concentrer les peptides de l'échantillon et d'éliminer les sels non tolérés par l'analyse en spectrométrie de masse en tandem. L'échantillon est ensuite emmené vers la colonne chromatographique, avec une chute du débit à 200 nl/min, composée de résine à laquelle sont greffées des chaînes lipidiques à 18 carbones. Les différents peptides vont s'y adsorber et seront élués progressivement en fonction de leur hydrophobicité grâce à un gradient croissant d'acétonitrile. Durant ces 30 minutes, les peptides sont désorbés grâce à un gradient d'hydrophobicité croissante de solution A et B (début par 95% de solution A pour 5% ACN +0,1% Ac. Formique (solution B) et fini par 60% de solution A pour 40% de B). Finalement, lors des 10 dernières minutes suivantes, 100% de la solution B permet de décrocher les peptides les plus hydrophobes et de nettoyer la colonne, avant de reconditionner cette dernière avec une solution 95% de solvant + 5% de solvant B.

II.7.4. Analyse des résultats

L'analyse bioinformatique des données acquises lors de l'analyse en spectrométrie de masse est effectuée avec Protein Lynx Global Server ou Mascot Daemon en utilisant les paramètres suivant :

- Base de donnée : NCBIInr0810 (7044224 sequences; 2433775060 residues)
- Enzyme : trypsine
- Modifications : carbamidométhylation (fixe), oxydation (variable)
- Valeur de masse : monoisotopique
- Tolérance de la masse peptidique : +/-0,2 Da
- Tolérance de la masse du fragment : +/- 0,15 Da

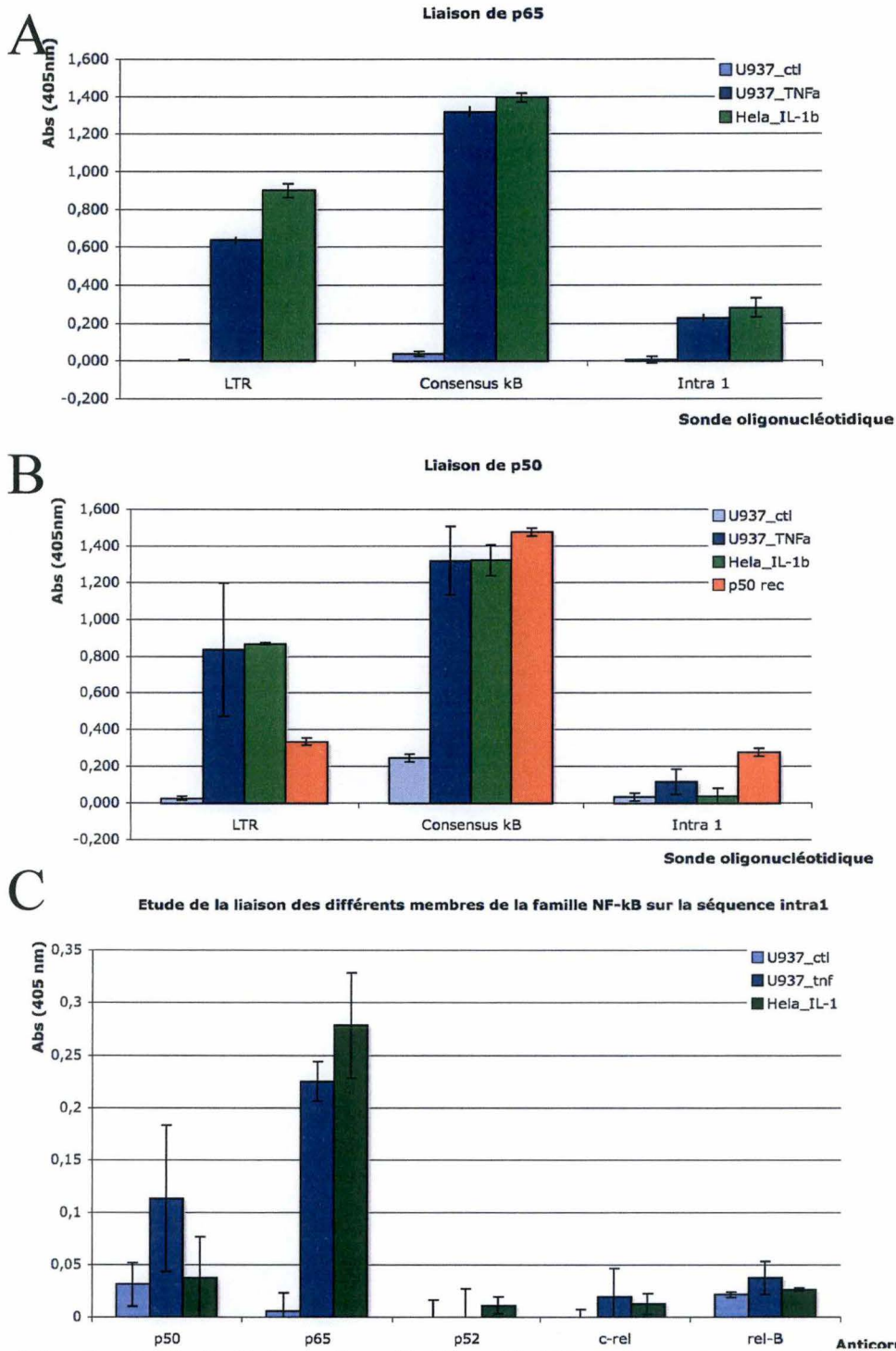


Figure 3.1 : Etude de la liaison de p65 (A), p50 (B) et des autres membres de la famille NF-κB (C) à la séquence intra1.

10 µg d'extraits nucléaires de cellules U937 non stimulées ou stimulées 1 heure au TNFα (10 ng/ml) ont été incubés en présence de la sonde intra1 immobilisée dans des puits (A, B, C). Deux séquences contenant des sites κB reconnus (le LTR-5' du HIV-1 et le site κB consensus du promoteur de la chaîne légère de l'immunoglobuline κ) ont également été utilisées en tant que contrôles positifs (en terme de séquence) pour A et B. Comme contrôle positif en terme d'extrait protéique pour la liaison de NF-κB, 10 µg d'extraits nucléaires de cellules HeLa stimulées 45 minutes à l'IL-1β (A, B, C) ou 2.5 pmoles de p50 recombinant (B), ont également été incubés en présence des différentes sondes.

III. Résultats et discussion

III.1. Etude de la liaison de facteurs de transcription susceptibles d'expliquer l'inductibilité de la région intragénique du HIV-1 : NF- κ B, NFAT et AP-1

III.1.1. NF- κ B

Dans la recherche de facteurs de transcription pouvant expliquer l'inductibilité de la région intragénique à une stimulation par le TNF α ou par le PMA + ionomycine, nous nous sommes tout d'abord intéressés à NF- κ B. En effet, ce facteur est bien connu pour être induit par le TNF α et par le PMA (voir Fig. 1.18). De plus, NF- κ B représente un acteur majeur de la régulation de la transcription du HIV-1 au niveau du LTR-5' (voir introduction point I.3.2).

Premièrement, la capacité de liaison des membres p50 et p65 (aussi appelé RelA), formant le dimère le plus largement représenté de la famille NF- κ B, sur la séquence intral a été analysée par ELISA-like (Fig 3.1 A et B). Pour rappel, cette technique permet d'évaluer la liaison d'un facteur de transcription particulier à une sonde oligonucléotidique immobilisée dans des puits, par l'intermédiaire d'anticorps primaires et secondaires (voir matériel et méthode point II.4). Pour ce faire, trois séquences ont été utilisées : la séquence d'intérêt correspondant à l'hypothétique enhancer intragénique du gène *pol* du HIV-1 (intra1), un oligonucléotide contenant le site consensus de NF- κ B (correspondant au site κ B du promoteur du gène codant les chaînes légères κ des immunoglobulines (Singh, Sen et al. 1986)) et une portion du LTR-5' (nt 228-461) du HIV-1. Deux sites NF- κ B (nt 350-359 and 364-373) ont été décrits dans cette dernière séquence et sont impliqués dans l'activation de la transcription du provirus. Comme attendu, la séquence consensus et le LTR capturent bien p65 (Fig. 3.1A) et p50 (Fig. 3.1B) dans les extraits nucléaires de cellules HeLa stimulées 45 minutes à l'IL-1 β (5ng/ml). La présence de la forme active de NF- κ B dans ce modèle cellulaire est bien connue, nous permettant d'utiliser ces extraits comme contrôle positif pour la liaison de p65 et p50 (Renard, Ernest et al. 2001). De plus, sur ces mêmes séquences, on peut également constater une liaison de p65 et de p50 dans les extraits de cellules U937 stimulées 1 heure par le TNF α (10ng/ml), validant ces extraits pour la présence du facteur activé.

De manière intéressante, cette liaison de p65 est détectable également sur la séquence intra1 (Fig. 3.1A) mais dans une moindre mesure que celle observée sur les deux autres séquences. En effet, la liaison de p65 dans les extraits nucléaires U937 stimulés par le TNF α est 6 et 4.5 fois moins forte sur la séquence intra1 que sur la séquence consensus et le LTR, respectivement (Fig 3.1A). Par contre, il est étonnant de constater que la liaison de p50 sur la séquence intra1 est très faible, quels que soient les extraits nucléaires considérés (Fig. 3.1B), alors que la séquence intra1 est capable de capturer des homodimères p50 comme le montrent les résultats obtenus avec la protéine p50 recombinante, mais cette fois encore dans une moindre mesure (7 fois moins) en comparaison avec la séquence consensus. Ces premiers résultats suggèrent la liaison de NF- κ B à la séquence intra1 sous une forme dimérique contenant le membre Rel-A, mais peut-être pas p50. Cependant, l'affinité de NF- κ B pour le site hypothétique présent sur la séquence d'intérêt est probablement faible comme le suggère la différence de l'intensité de liaison par rapport au consensus.

Etant donné la très faible liaison de p50 sur la séquence intra1, nous nous sommes intéressés à la liaison des autres membres de la famille NF- κ B afin d'étudier l'éventuelle formation de dimères non canoniques. Pour ce faire, la liaison des 3 membres de la famille

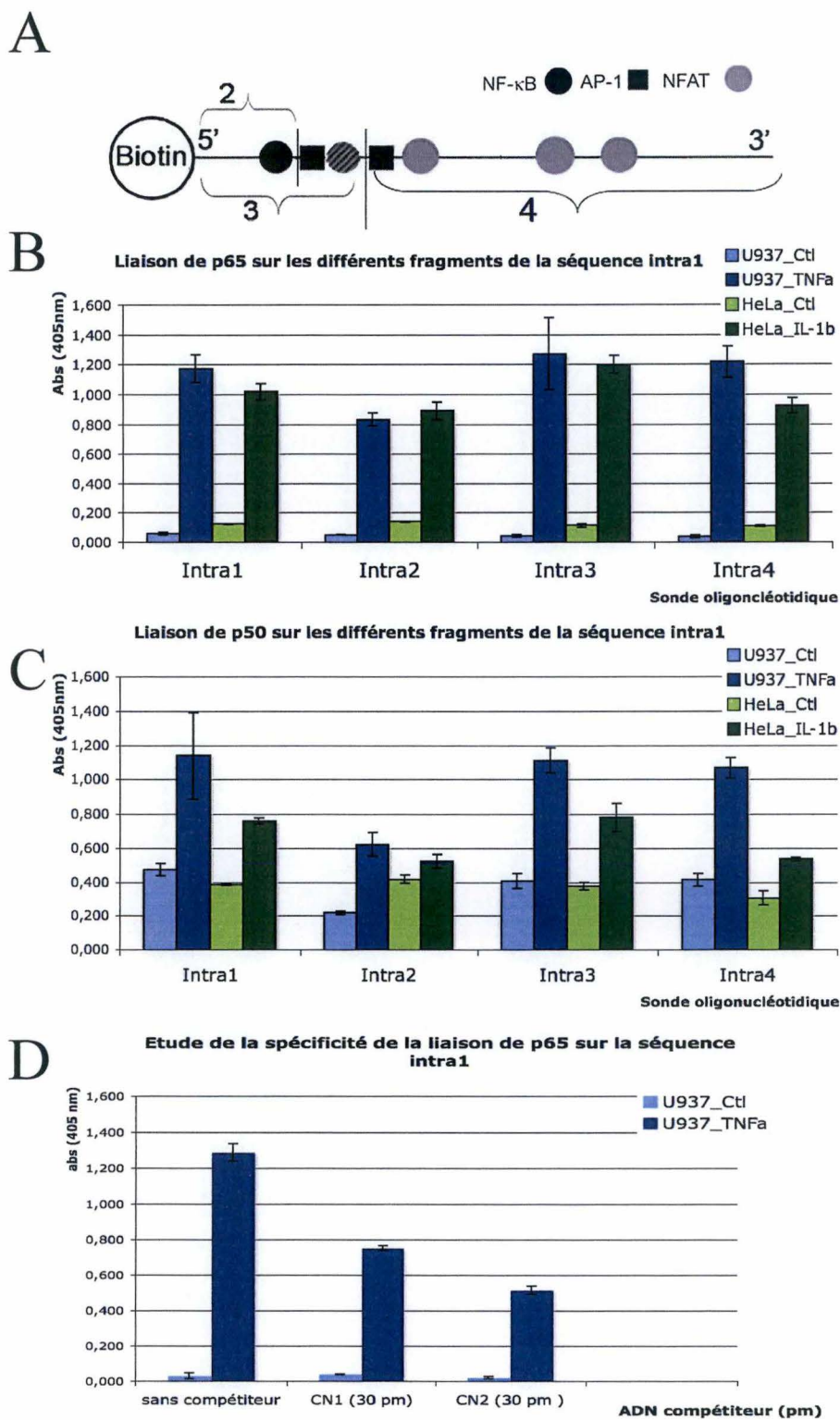


Figure 3.2 : Etude de la localisation du (des) site(s) NF- κ B putatifs sur la séquence intra1.

(A) Découpage de la séquence intra1 et sites putatifs prédits *in silico*. Etude de la liaison de p65 (B) et p50 (C) : 20 μ g d'extraits nucléaires de cellules U937 non stimulées ou stimulées 1 heure au TNF α (10 ng/ml) ont été incubés en présence des différents fragments de l'intra1. Les extraits nucléaires de cellules HeLa sont toujours utilisés comme contrôle positif. (D) Etude de la spécificité de la liaison de p65 sur la séquence intra1 pour les extraits nucléaires (20 μ g) de cellules U937 en absence et en présence d'un excès 12 fois d'ADN compétiteur (CN1 et CN2).

NF- κ B (p52, c-Rel, RelB) a été recherchée sur la séquence *intra1* dans des extraits nucléaires de cellules U937 non stimulées ou stimulées 1 heure en présence de TNF α (10 ng/ml) et pour des extraits nucléaires de cellules HeLa stimulées 45 minutes par l'IL-1 β (Fig. 3.1C). Aucun des trois facteurs étudiés ne semblent se lier à la séquence d'intérêt, quelle que soit la condition étudiée. Cependant, on ne peut exclure définitivement l'implication de ces membres dans la régulation de la séquence intragénique du HIV-1, car le temps de stimulation nécessaire à l'activation de certains membres peut être plus long que celui de RelA. En effet, à l'opposé de RelA, présent à l'état basal dans la cellule et rapidement délocalisable dans le noyau, RelB n'est pas exprimé constitutivement et doit être synthétisé, ce qui peut demander des temps d'activation plus longs (Bren, Solan et al. 2001). D'autre part, la liaison de p50 sur l'*intra1* étant faible mais pas nulle, cette expérience pourrait être refaite avec davantage d'extraits. Malgré tout, ces résultats suggèrent un homodimère RelA-RelA, forme de NF- κ B qui n'est pas classique mais qui a déjà été décrite (Ganchi, Sun et al. 1993).

Afin de localiser plus précisément la position de site de liaison de NF- κ B sur la séquence *intra1* une analyse bioinformatique a été réalisée, pointant différents sites putatifs. Cette analyse a été réalisée à l'aide du module MATCH de Transfac (<http://biobase-international.com>) et du logiciel MatInspector de Genomatix (<http://www.genomatix.de>). Parmi les sites putatifs identifiés, un site NF- κ B hypothétique (5'-ATTTGGCATTCCC-3' Nt 4643) a été identifié en aval du site B (voir Fig. 1.16) et à 13 nucléotides d'un site AP-1 putatif (5'-AGTCA-3' Nt 4668). Afin de tester une éventuelle coopération entre ces deux sites, la séquence *intra1* a été fragmentée entre ceux-ci, générant une première sonde appelée « *intra2* » (Nt 4594-4667) et ne contenant que le site NF- κ B putatif alors que l'oligonucléotide « *intra3* » (Nt 4594-4688) correspond à l'*intra2* allongé d'une vingtaine de nucléotides parmi lesquels se trouve le site putatif AP-1 (Fig. 3.2A). Finalement la séquence dite « *intra4* » (Nt 4674-4911) correspond au restant de la séquence *intra1* situé en aval de l'*intra3* et qui ne contient aucun site putatif pour NF- κ B (Fig. 3.2A).

Etant donné les faibles valeurs de liaison indiquées sur la séquence *intra1* (Fig. 3.1), à la limite du seuil de détection du spectrophotomètre, nous avons décidé de doubler la quantité d'extraits nucléaires (passant de 10 μ g à 20 μ g) incubés avec les sondes afin de diminuer la variabilité et d'augmenter les niveaux de liaison détectés. Les résultats de l'étude de la liaison de RelA et p50 aux différents fragments correspondant à la séquence *intra1* sont illustrés aux figures 3.2B et 3.2C, respectivement. Cette fois encore, les extraits nucléaires de cellules HeLa sont utilisés comme contrôles positifs pour lesquels une importante liaison de RelA (Fig. 3.2B) peut être observée dans les extraits stimulés 45 minutes à l'IL-1 β (5ng/ml) comparativement aux extraits non stimulés et ce pour chaque sonde utilisée. Cette activation peut également être observée pour p50 (Fig. 3.2C) mais dans une moindre mesure (au maximum 2 fois). Le profil de liaison de RelA et de p50 pour les extraits nucléaires de cellules U937 non stimulées ou stimulées 1 heure par le TNF α (10 ng/ml) est similaire à celui observé pour les cellules HeLa. La moindre liaison de p50 en réponse à la stimulation au TNF α ou à l'IL-1 β peut être due à la capacité des homodimères p50 à entrer dans le noyau et à lier l'ADN lorsque les cellules ne sont pas stimulées. En effet, sous la forme homodimérique p50 n'est pas ciblé par I κ B comme c'est le cas pour les hétérodimères p50-RelA. De la sorte, p50 agit en temps que répresseur en occupant des sites de liaison κ B au niveau de l'ADN alors que l'homodimère ne possède aucune capacité trans-activatrice. Il peut être déplacé du site de liaison par un hétérodimère, tel que p50-RelA, ayant une plus grande affinité pour la séquence nucléique que l'homodimère p50-p50 (Ghosh and Karin 2002).

La présence de RelA et p50 sur les fragments intra2 et intra3 serait un argument en faveur du site NF- κ B putatif suggéré par l'analyse bioinformatique (Fig. 3.2 A). Cependant, une liaison de NF- κ B est également mise en évidence au niveau de l'intra4, région pour laquelle aucun site de liaison pour ce facteur n'a pu être prédit. Ces résultats pourraient s'expliquer par la présence de faux négatifs lors de l'analyse bioinformatique de la séquence intra4 ou par une liaison non spécifique du facteur au niveau de cette même séquence.

C'est afin de tester cette hypothèse, renforcée par l'observation d'une liaison homogène de RelA (et de p50 dans une moindre mesure) quel que soit le fragment étudié (intra 1 à 4) que des expériences de compétitions ont été réalisées. Pour ce faire, deux séquences oligonucléotidiques ont été utilisées, toutes deux provenant d'autres régions du génome du HIV-1 ne correspondant pas à des sites d'hypersensibilité à la DNase. La séquence CN1 (contrôle négatif 1) possède une longueur comparable à celle de l'intra1 alors que la séquence CN2 contient le même pourcentage en GC que la séquence intra1. Etant donné que la quantité de sondes immobilisées dans les puits est estimée à 2.5 pmoles, un excès 12 fois de sondes compétitrices (soit 30 pmoles) est ajouté lors de l'incubation des extraits nucléaires en présence des sondes d'intérêts. Le résultat de cette expérience de compétition de la liaison de RelA sur la séquence intra1 pour des extraits nucléaires de cellules U937 non stimulées ou stimulées 1 heure au TNF α est représenté à la figure 3.2D. En l'absence de compétiteur, l'activation de RelA est similaire au résultat obtenu précédemment. Cependant, une chute de cette activation est observée en présence d'un excès 12 fois de compétiteur, diminuant la liaison de RelA de plus de 50% dans le cas du compétiteur CN2.

En résumé, la liaison de RelA (et dans une moindre mesure, p50) a pu être montrée dans un premier temps sur la séquence intra1 puis, par après, sur les fragments intra2 et 3, en accord avec la présence du site putatif prédit *in silico*. Cependant, le fait que la liaison de p65 ait aussi été détectée sur l'intra4, alors qu'aucun site putatif n'ait été prédit dans cette séquence, pose question et remet en cause la spécificité de la liaison sur les fragments intra1, 2 et 3. En effet, l'expérience de compétition réalisée semble être en faveur d'une liaison non spécifique de RelA sur la séquence intra1.

Des expériences complémentaires pourraient être réalisées en vue de confirmer ce dernier résultat, notamment la mutation du site putatif identifié sur le fragment intra2, afin de voir si la liaison de RelA chute dans ces conditions. Une autre expérience de compétition similaire pourrait aussi être réalisée en utilisant cette fois des sondes compétitrices beaucoup plus courtes contenant soit le consensus NF- κ B ou soit son homologue muté.

Finalement, nous n'avons pas continué dans cette voie car, suite à nos premiers résultats, l'équipe de C. Van Lint a recherché par gel retard la liaison de NF- κ B au niveau du site putatif, en vain, ce qui renforce l'hypothèse d'une liaison non spécifique.

III.1.2. NFAT

Etant donné que la région intragénique du HIV-1 est également inductible au PMA + ionomycine, nous avons étudié le facteur de transcription NFAT qui est classiquement décrit pour être activé par ce traitement (voir introduction point I.4.3) dans des lignées cellulaires T lymphoïdes. La démarche appliquée est similaire à celle utilisée pour NF- κ B.

Premièrement, la liaison de NFAT a été testée sur la séquence intra1. Cette fois encore, une séquence nucléotidique contenant le site consensus de NFAT (5'-T/AGGAAA-3', (Luo, Burgeon et al. 1996)) a également été utilisée afin de contrôler la présence de la forme activée de NFAT dans les extraits stimulés. L'étude de la liaison de NFAT à la séquence intra1 pour des extraits nucléaires de cellules Jurkat non stimulées ou stimulées 5 heures en

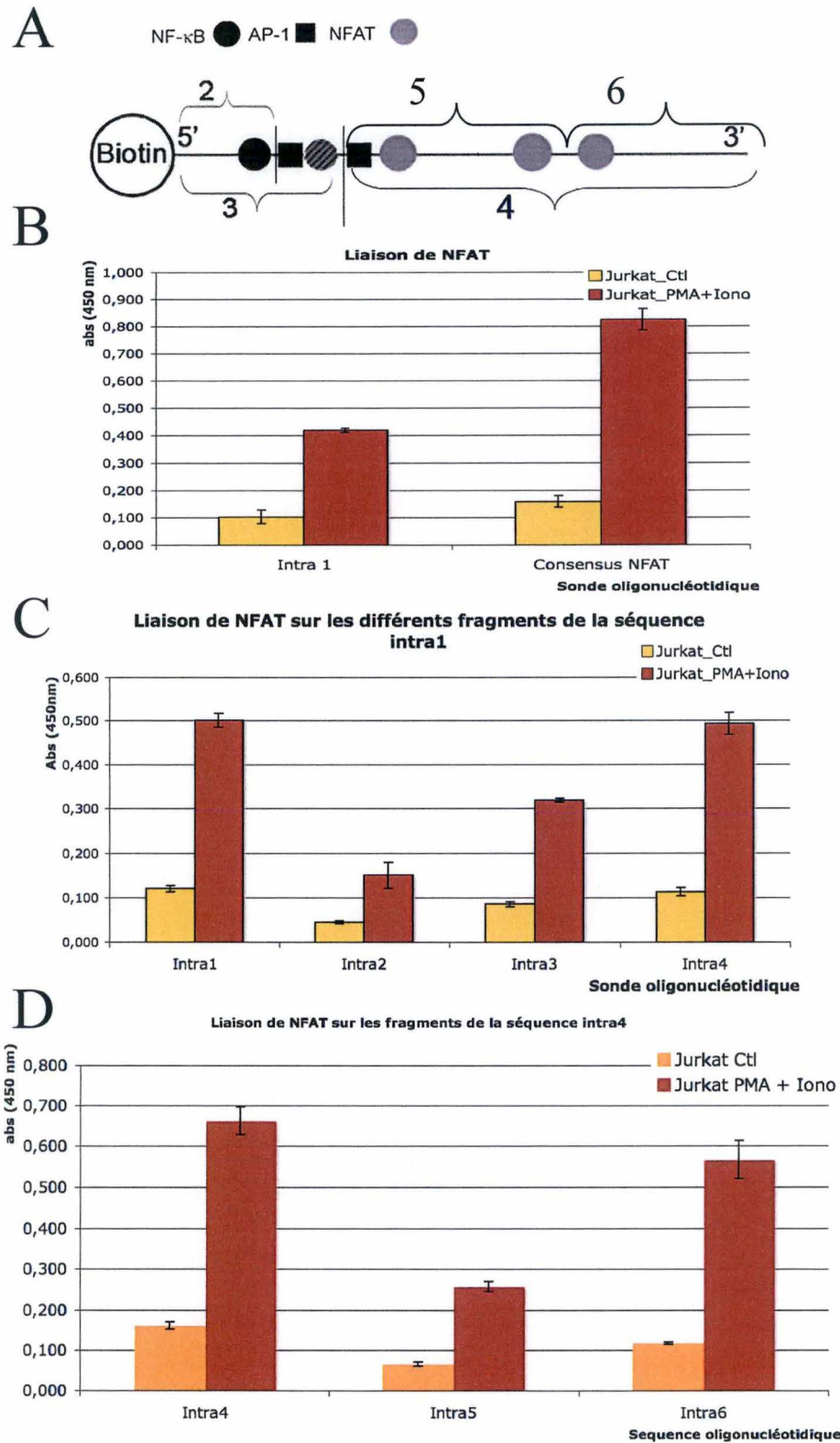


Figure 3.3 : Etude de la liaison de NFAT à la séquence intra1

(A) Visualisation des sites de liaison putatifs pour NF-κB, NFAT et AP-1, ainsi que du « découpage » de la séquence intra1 en 6 fragments. (B) 10 μg d'extraits nucléaires de cellules Jurkat non stimulées ou stimulées 5 heures au PMA (32 nM) + ionomycine (500 nM) ont été incubés en présence de 2,5 pmoles de la séquence intra1 ou d'un oligonucléotide contenant le consensus NFAT. (C), (D) 20 μg d'extraits nucléaires de cellules Jurkat stimulées ou non en présence de PMA + ionomycine ont été incubés en présence des différents fragments issus du découpage de la séquence intra1 (C) ou intra4 (D).

présence de PMA (32nM) + ionomycine (500 nM) peut être observée à la figure 3.3A. La liaison de NFAT sur la séquence consensus dans les extraits de cellules stimulées par le PMA + ionomycine confirme que ce traitement a provoqué l'activation de ce facteur de transcription. Cette augmentation de la liaison de NFAT consécutive au traitement par le PMA + ionomycine peut être observée sur la séquence *intra1*, mais, comme pour NF- κ B, dans une moindre mesure en comparaison avec la séquence consensus.

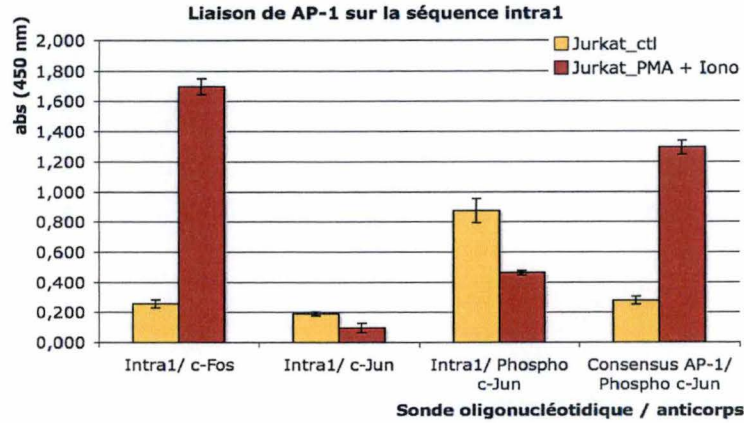
Afin de localiser plus précisément la position de ce ou ces sites de liaison, la présence de sites putatifs de liaison pour NFAT a été également recherchée par analyse bioinformatique et a permis de mettre en évidence 4 sites putatifs aux nucléotides 4674 (5'-GGAG-3'), 4705 (5'-AGAAA-3'), 4785 (5'-AGAAA-3'), 4817 (5'-GGAAA-3') (Figure 3.3 B). L'étude de la liaison de NFAT a également été réalisée sur les différents fragments de la séquence *intra1* (Figure 3.3C) pour les extraits nucléaires de cellules Jurkat stimulées ou non. Cette fois le profil de liaison varie entre les différents fragments de l'*intra1*. En effet, alors que la liaison est très faible au niveau de la sonde *intra2*, les sondes *intra3* et surtout *intra4* capturent ce facteur de transcription. NFAT ne semble pas capturé par l'*intra2*, contenant le site NF- κ B putatif auquel il aurait pu se lier. En effet, bien que la séquence consensus de NFAT soit 5'-T/AGGAAA-3', le facteur NFAT est également capable de lier certains sites NF- κ B, comme, par exemple, les deux sites présents au niveau du LTR-5' du HIV-1 (Bates, Barthel et al. 2008). L'augmentation (environ 2 fois) de la liaison de NFAT entre la sonde *intra2* et *intra3* laisse suspecter la présence d'un site de liaison dans les 20 pb différenciant les deux séquences. Un examen minutieux de ces 20 pb indique que seules quelques bases d'un site putatif fort dégénéré (5'-GGAG-3') pourraient expliquer la liaison de NFAT à cet endroit.

Trois autres sites putatifs de liaison pour NFAT (Nt 4705, 4785, 4817) ont été localisés sur la séquence *intra1* dans la région correspondant à l'*intra4*, ce qui permettrait d'expliquer le signal mesuré pour ce fragment. Afin de déterminer si ces trois sites contribuent à la liaison observée, la séquence *intra4* a été fragmentée en *intra5*, d'environ 125 pb et contenant deux des sites putatifs, et *intra6* de 110 pb où est localisé le site putatif restant (Fig. 3.3 B). Comme on peut le voir à la figure 3.3D, l'activation de NFAT est majoritairement observée au niveau de la séquence *intra6*. Même si une faible activation est mise en évidence au niveau du fragment *intra5*, le niveau de liaison pour les extraits nucléaires de cellules Jurkat stimulées 5 heures au PMA + ionomycine reste très faible. Il semblerait donc que NFAT lie davantage les sites putatifs localisés sur le fragment *intra6*.

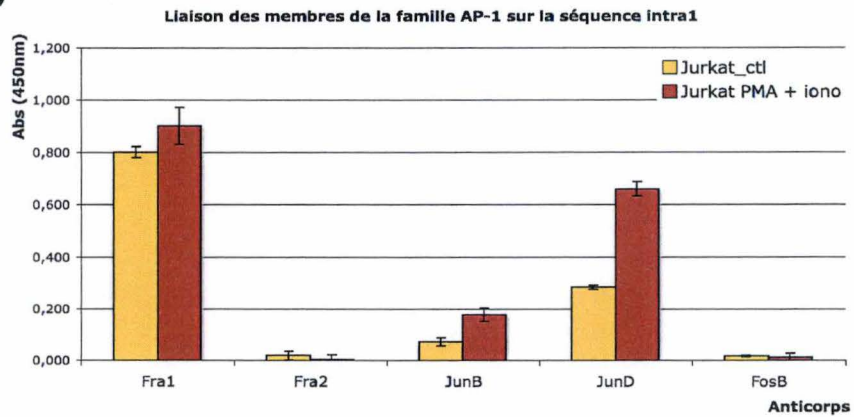
III.1.3. AP-1

Pour finir, le troisième candidat choisi susceptible d'induire la transcription de la région intragénique du HIV-1 est le facteur de transcription AP-1, étant donné qu'il peut être induit par un traitement par le PMA + ionomycine ou par le TNF α (voir introduction I.4.2). Comme pour NF- κ B, nous nous sommes intéressés en premier lieu aux membres de la famille AP-1 formant le dimère le plus répandu : c-Fos et c-Jun. Comme attendu, une différence de liaison de c-Jun phosphorylé peut être observée entre les extraits nucléaires de cellules Jurkat non stimulées et stimulées 5 heures en présence de PMA (32nM) + ionomycine (500 nM) sur la séquence consensus (Fig. 3.4A). Néanmoins, aucune liaison similaire ne peut être observée sur la séquence *intra1*, que l'on détecte c-Jun ou sa forme active phosphorylée. Par contre, une forte augmentation de la liaison de c-Fos est observée dans les extraits nucléaires de cellules Jurkat activées en comparaison à celle mesurée dans les conditions contrôles. Ces résultats suggèrent une liaison de AP-1 sous forme d'un dimère contenant c-Fos mais pas c-Jun. Cependant, c-Fos ne peut former d'homodimères, c'est pourquoi ce membre de la famille AP-1 est obligé de former des hétérodimères. C'est dans cette optique que la liaison des autres membres de la famille AP-1 (Fra1, Fra2, JunB, JunD, FosB) a également été recherchée sur la

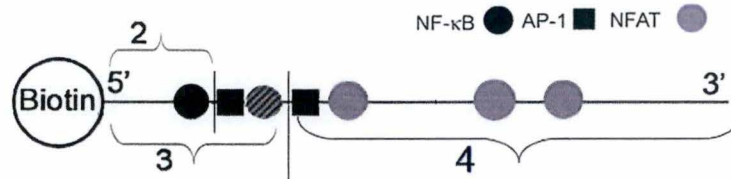
A



B



C



D

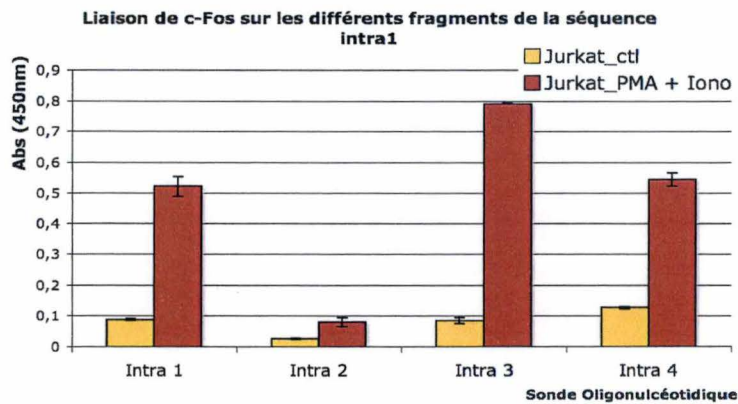


Figure 3.4 : Etude de la liaison de AP-1 sur la séquence intra1.

(A) Etude des membres les plus représentés de la famille AP-1 : c-Fos et c-Jun. 10 µg d'extraits nucléaires de cellules Jurkat non stimulées ou stimulées 5 heures au PMA (32 nM) + ionomycine (500nM) ont été incubés en présence de la séquence intra1, ou d'une séquence consensus de liaison pour AP-1. La liaison éventuelle d'AP-1 a été détectée à l'aide d'un anticorps dirigé contre c-Fos, c-Jun ou phospho-cjun. (B) Etude des autres membres de la famille AP-1. 20 µg d'extraits nucléaires de cellules Jurkat non stimulées ou stimulées 5 heures au PMA (32 nM) + ionomycine (500nM) ont été incubés en présence de la séquence intra1. La liaison éventuelle d'AP-1 a été détectée à l'aide d'un anticorps dirigés contre Fra1, Fra2, JunB, JunD et FosB (C) Représentation du découpage de la séquence intra1 et des différents sites putatifs prédits pour ces différents fragments. (D) 20 µg d'extraits nucléaires de cellules Jurkat stimulées ou non en présence de PMA + ionomycine ont été incubés en présence des différents fragments issus du découpage de la séquence intra1. La liaison de c-Fos a été détectée à l'aide d'un anticorps spécifique.

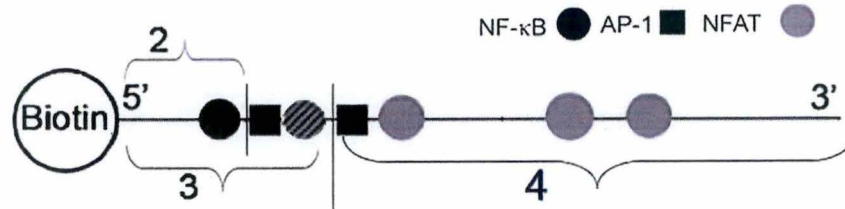
séquence *intra1*, toujours pour des extraits nucléaires de cellules Jurkat non stimulées ou stimulées par le PMA + ionomycine (Fig. 3.4B). *Fra2* et *FosB* ne semblent pas se lier à la séquence *intra1*, même pour les extraits stimulés. *Fra1* présente une importante liaison sur la séquence *intra1*, et ce, aussi bien dans les extraits de cellules stimulées que non stimulées, ce qui signifie qu'il ne peut être le facteur de transcription responsable de l'inductibilité de cette séquence. Une activation peut être observée pour les facteurs *JunB* et *JunD* avec une liaison du facteur qui double entre les extraits contrôles et stimulés. Malgré tout, la liaison de *JunB* reste très faible et il est plus probable que le dimère liant la séquence *intra1* soit de type c-Fos-*JunD*.

Comme pour les deux autres facteurs, les différents fragments de la séquence *intra1* ont été utilisés dans le but de localiser plus précisément la position du ou des sites de liaison suggérés par les résultats précédents. Comme on peut le voir figure 3.4D, la liaison de c-Fos varie fortement en fonction du fragment de l'*intra1* utilisé, pour des extraits nucléaires de cellules Jurkat non stimulées ou stimulées par le PMA + ionomycine. En effet, alors que pratiquement aucune liaison de c-Fos n'est détectable sur la séquence *intra2*, la plus forte différence de liaison entre les deux conditions est observée sur la séquence *intra3*. Une augmentation de la liaison de c-Fos est également mise en évidence lorsque l'on compare les extraits contrôles et stimulés sur la sonde *intra4*, mais dans une moindre mesure que celle mesurée pour l'*intra3*. La liaison de c-Fos observée sur l'*intra3* dans les extraits de cellules stimulées par le PMA + ionomycine semble appuyer la présence du site putatif (Nt 4668) identifié *in silico* et situé parmi les 20 nucléotides différenciant l'*intra2* et l'*intra3* (voir Fig. 3.4.C). Un second site putatif prédit *in silico* est situé dans la région correspondant à la séquence *intra4* et pourrait expliquer la liaison de c-Fos sur ce fragment.

Afin de déterminer si un dimère c-Fos/*JunD*, ou éventuellement c-Fos/*JunB*, peut se lier sur cette séquence de 20 nucléotides différenciant l'*intra3* de l'*intra2*, la liaison des deux membres de la famille *Jun* a été recherchée sur les séquences *intra2* et *intra3*, pour lesquelles une différence de liaison a pu être montrée pour c-Fos (Fig. 3.4D). La liaison de c-Fos, *JunB* et *JunD* aux séquences *intra2* et *intra3* pour des extraits de cellules Jurkat non stimulées ou stimulées par le PMA + ionomycine peut être observée à la figure 3.5B. Aucune liaison n'est mesurée pour *JunB* au niveau de l'*intra2*, quel que soit l'extrait nucléaire utilisé. Même si la séquence *intra3* semble capturer ce facteur dans les extraits de cellules stimulées en comparaison avec les contrôles, cette liaison reste très faible et à la limite du seuil de détection. Le profil observé est similaire pour *JunD* : la liaison est comparable dans les deux types d'extraits sur la séquence *intra2* alors qu'une différence de liaison (environ 2 fois) de *JunD* sur le fragment *intra3* est mesurée entre les deux types d'extraits nucléaires. Le probable dimère AP-1 se liant à la séquence *intra1* est donc de type c-Fos/*JunD*, bien qu'on ne puisse pas exclure c-Fos/*JunB*. Il faut également mentionner que, bien que la différence de liaison de c-Fos entre les extraits nucléaires de cellules contrôles et stimulées soit toujours observable, le niveau de liaison de ce facteur est environ 3 fois plus faible sur le fragment *intra3* que ce qui a pu être observé précédemment (voir ci-après).

Comme cela a été fait pour NF- κ B, nous avons également voulu tester la spécificité de la liaison de c-Fos sur la séquence *intra1*. Les deux séquences compétitrices sont les mêmes que celles utilisées précédemment (CN1, CN2, Fig. 3.2 D). Cette fois, l'expérience de compétition est également effectuée par rapport à la séquence contenant le site consensus de AP-1 précédemment utilisée, afin de s'assurer que les sondes compétitrices ne possèdent pas de site de liaison pour c-Fos. Les résultats pour les sondes consensus (Fig. 3.6A) montrent une liaison du facteur en réponse à la stimulation cellulaire lorsque que l'on compare les

A



B

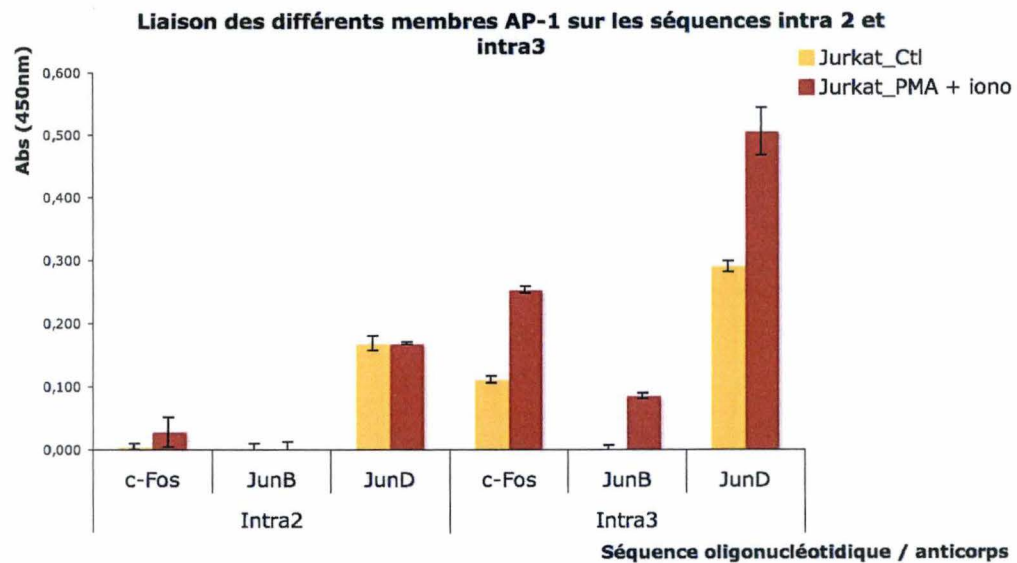


Figure 3.5 : Etude de la liaison de c-Fos, JunB et JunD sur les séquences intra 2 et 3

(A) Représentation du découpage de la séquence intra1 et des différents sites putatifs prédits pour ces différents fragments. (B) Liaison de c-Fos, JunB et JunD sur les séquences intra2 et intra3. 20 µg d'extraits nucléaires de cellules Jurkat non stimulées ou stimulées 5 heures au PMA (32 nM) + ionomycine (500nM) ont été incubés en présence de 2.5 pmoles de la séquence intra2 ou intra3.

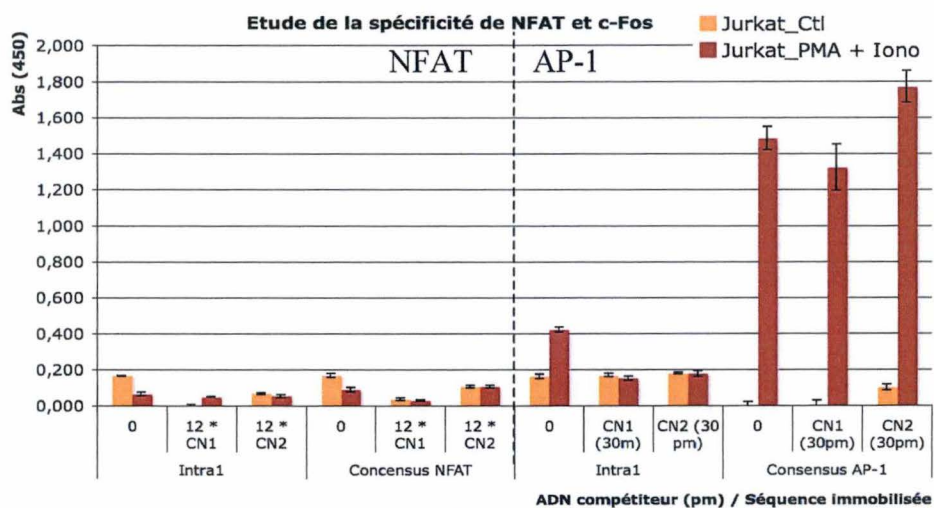


Figure 3.6 : Etude de la spécificité de NFAT et d'AP-1 sur la séquence intra1

Spécificité de la liaison de NFAT (à gauche) et de c-Fos (à droite) sur la séquence intra1 pour les extraits nucléaires (20 µg) de cellules Jurkat en absence et en présence d'un excès 12 fois d'ADN compétiteur (CN1 et CN2).

extraits contrôles et stimulés avec du PMA + ionomycine. Par contre, l'ajout de compétiteurs (excès 12 fois) ne diminue pas la liaison dans les extraits nucléaires de cellules Jurkat stimulées, quelle que soit la séquence compétitrice utilisée (CN1 ou CN2). On peut observer une liaison de c-Fos sur la séquence intra1 et en l'absence de compétiteurs lorsque l'on compare les extraits contrôles et stimulés. Cependant, en présence de l'excès 12 fois de sondes nucléiques compétitrices (CN1 et CN2), le signal mesuré pour les extraits stimulés avec du PMA + ionomycine chute à un niveau comparable à celui des extraits contrôles. Ce résultat est surprenant étant donné qu'il indique une liaison non spécifique de c-Fos alors que la différence de liaison de ce facteur dans les extraits stimulés entre la séquence intra2 et intra3 (Fig. 3.4D) laisse penser le contraire.

Une expérience de compétition a été réalisée en parallèle avec les mêmes extraits nucléaires, mais cette fois en détectant la liaison de NFAT et plus de c-Fos. Cependant, aucune liaison de NFAT n'a pu être observée quel que soit l'extrait testé et même en l'absence de compétiteurs (Fig. 3.6 A). Étant donné que la liaison de NFAT à la séquence intragénique avait été montrée précédemment (Fig. 3.3C et 3.3D), ce résultat est plutôt surprenant. De plus, les deux dernières expériences réalisées pour détecter la liaison de c-Fos (Fig. 3.5 B et 3.6 A) montrent un niveau de liaison plus faible sur la séquence intra3 que celui observé précédemment (Fig. 3.4D). Devant ces résultats troublants et sachant que les dernières expériences ont été réalisées avec un nouveau batch d'extraits nucléaires de cellules, nous avons mis en doute la qualité de nos extraits. L'activation de NFAT dans les extraits nucléaires de cellules stimulées a tout d'abord été testée par ELISA-like en utilisant une sonde contenant le consensus NFAT. Nous n'avons pas pu détecter de liaison de NFAT à sa séquence consensus (données non montrées), expliquant les résultats obtenus avec NFAT pour l'expérience de compétition. De nouveaux extraits ont par la suite été produits, permettant de mettre à nouveau en évidence la liaison de NFAT à sa séquence consensus dans des extraits nucléaires de cellules Jurkat stimulées avec du PMA + ionomycine.

Suite à la confirmation de ces problèmes d'extraction protéique, les informations collectées lors des expériences présentées dans les figures 3.5B et 3.6A sont à considérer avec prudence, à titre informatif, et ne peuvent pas permettre de tirer une conclusion définitive. Les expériences de compétition pour NFAT et AP-1 sont à refaire, ainsi que l'étude de la liaison de JunB et JunD sur les différents fragments de la séquence intra1.

Néanmoins, nous nous sommes interrogés sur l'interprétation à donner à ces résultats de la liaison de c-Fos sur la séquence intra-1, obtenus avec ces extraits nucléaires dans lesquels AP-1 est activé et pas NFAT. En effet, la liaison de c-Fos sur sa séquence consensus est toujours optimale (figure 3.6 A), alors qu'elle est clairement réduite sur la séquence intra 1, en comparaison avec des extraits dans lesquels AP-1 et NFAT sont activés (comparer les figures 3.6 et 3.4.A). Cette constatation, ajoutée au fait que le profil de liaison de c-Fos, (et JunD) sur les séquences intra2 et intra3 est similaire à celui de NFAT (voir figures 3.4D, 3.5B et 3.3C, respectivement) nous a fait penser à une éventuelle coopération entre ces facteurs de transcription.

III.1.4. Etude de la coopération entre AP-1 et NFAT

La coopération de AP-1 avec NFAT a été mise en évidence dans le contexte de plusieurs promoteurs (voir introduction I.4.3). En particulier, la vingtaine de nucléotides qui constitue l'extrémité 3' de l'intra3 (séquence qui n'est pas présente dans l'intra2 et pour laquelle nos résultats suggèrent une liaison de NFAT et d'AP-1) présente une certaine similitude avec un fragment du promoteur du gène codant le GM-CSF, fragment dans lequel

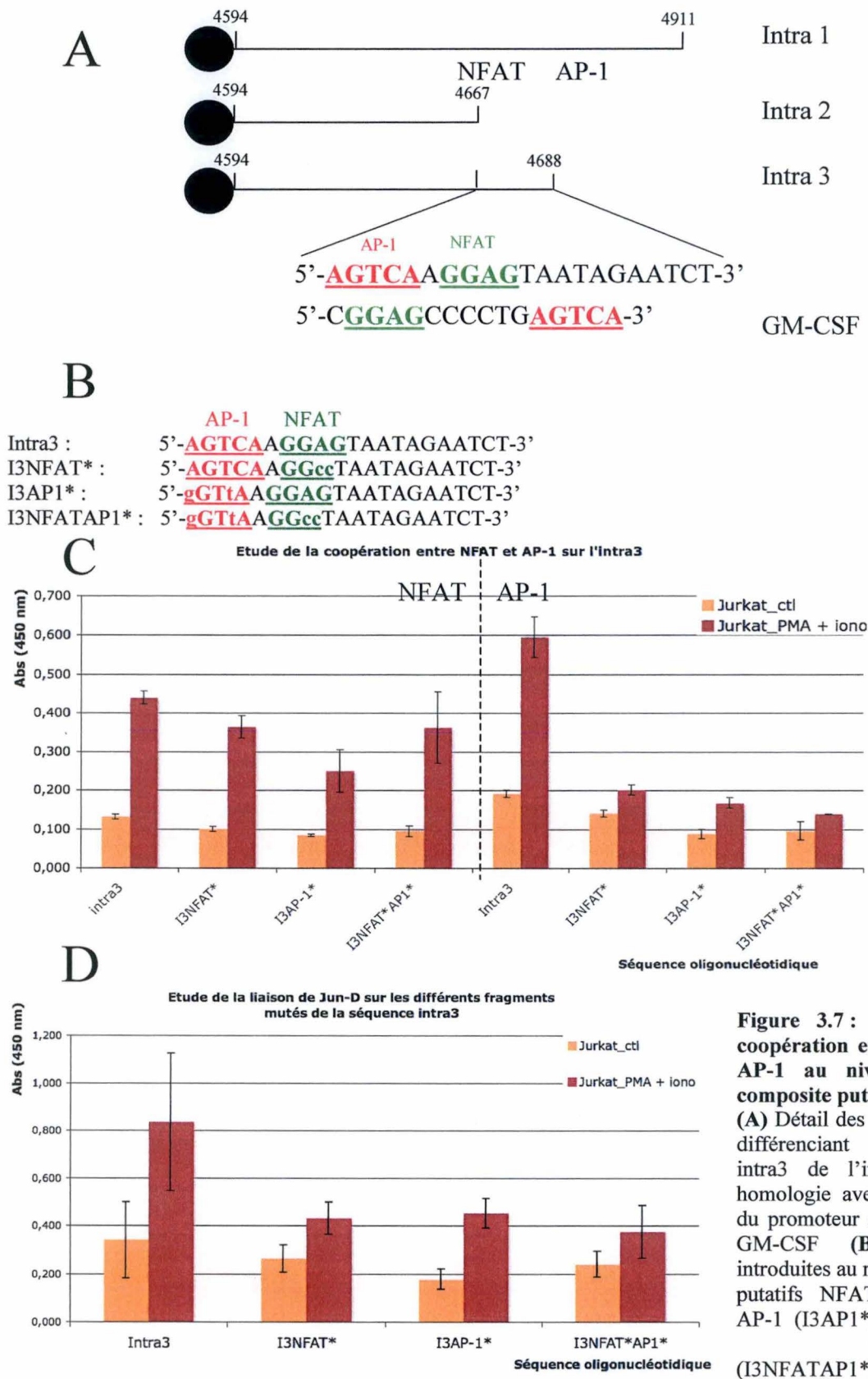


Figure 3.7: Etude de la coop ration entre NFAT et AP-1 au niveau du site composite putatif.

(A) D tail des 21 nucl otides diff renciant la s quence intra3 de l'intra2 et leur homologie avec la s quence du promoteur codant pour le GM-CSF (B) Mutations introduites au niveau des sites putatifs NFAT (I3NFAT*), AP-1 (I3AP1*) ou les deux simultan ment (I3NFATAP1*). (C) et (D) :

r sultats obtenus pour l' tude de la liaison de NFAT (  gauche), c-Fos (  droite) (C) et JunD (D) sur les sondes mut es pour les sites putatifs et incub es en pr sence de 20 µg d'extraits nucl aires de cellules Jurkat non stimul es ou stimul es 5 heures au PMA (32 nM) + ionomycine (500 nM).

un site composite NFA- AP-1 a été mis en évidence (Macian, Lopez-Rodriguez et al. 2001), comme l'illustre la figure 3.7.A La liaison de NFAT et AP-1 à ces sites composites est caractérisée par une plus grande affinité et stabilité lorsqu'ils interagissent physiquement que lorsqu'ils se lient individuellement aux acides nucléiques. Dans l'extrémité 3' de la séquence intra3, nous avons détecté des sites putatifs dégénérés pour AP-1 (Nt 4668) et NFAT (Nt 4674) dont la disposition est inversée dans la séquence intragénique par rapport au promoteur du GM-CSF, mais la nécessité d'une orientation particulière n'a pas été décrite dans la littérature.

Afin de déterminer si les sites AP-1 et NFAT dégénérés et situés à l'extrémité 3' de la sonde intra3 peuvent fonctionner comme un site composite, des sondes homologues à ce fragment intra3 ont été synthétisées (Fig. 3.7B) mais pour lesquelles le site putatif AP-1 (Nt 4668), ou NFAT (Nt 4674) et les deux simultanément ont été mutés. Comme on peut le voir sur la partie gauche du graphique présenté à la figure 3.7C, il y a une liaison de NFAT plus importante dans les extraits de cellules Jurkat stimulées en présence de PMA + ionomycine que dans les extraits contrôles sur la séquence intra3. Cette liaison est légèrement plus faible pour les extraits stimulés sur la sonde où les sites NFAT et AP-1 putatifs ont été mutés. Il est étonnant de constater que la liaison sur la sonde possédant simultanément les deux mutations présente encore une liaison de NFAT environ 3.5 fois supérieure pour des extraits de cellules stimulées par rapport aux cellules non traitées. A droite du graphique (Fig. 3.7C) sont présentés les résultats pour c-Fos. Alors qu'une liaison du facteur activé est observable sur la sonde sauvage, comme montré précédemment, le niveau de liaison dans les extraits stimulés chute d'un facteur 3, atteignant presque le niveau basal des contrôles, pour les 3 sondes mutées. La même démarche a été effectuée pour JunD. Les résultats (Fig. 3.7D) sont similaires à ceux obtenus pour c-Fos.

Les résultats de ces expériences de liaison à la sonde intral pour le facteur de transcription AP-1 en présence de sondes mutées suggèrent une coopération avec NFAT. En effet, la chute de la liaison mesurée pour AP-1 sur les sondes mutées pour le site putatif de NFAT coïncide avec la description de la coopération physique entre ces facteurs de transcription : en absence d'un des deux facteurs, il en résulte une plus faible stabilité de la liaison du partenaire à l'ADN.

Ces deux expériences permettent aussi de répondre à la question de la spécificité de la liaison de AP-1 à la séquence intragénique, du moins pour la partie correspondant au fragment intra3. En effet, la chute de la liaison de c-Fos et JunD sur la sonde mutée pour le site AP-1 dans les extraits stimulés par le PMA + ionomycine laisse penser que la liaison au niveau du site AP-1 situé au nucléotide 4668 est spécifique. La différence de liaison de NFAT entre les séquences sauvage et mutée au niveau du site putatif pour les extraits nucléaires stimulés par le PMA + ionomycine étant moins évidente, une conclusion similaire à celle de AP-1 ne peut être également tirée.

Pour finir, étant donné que la région intragénique du HIV-1 est également inductible par un traitement au PMA + ionomycine dans les lignées monocytaires/macrophagiques, nous avons également voulu étudier la liaison de AP-1 et NFAT dans des extraits nucléaires de cellules U937 stimulées 1 heure en présence de PMA (810 μ M) + ionomycine (285 nM). Bien que NFAT soit classiquement exprimé dans les lignées cellulaires de type T, des données indiquent qu'il pourrait aussi être exprimé et activé dans des monocytes (Shin, Lee et al. 2007). Deux types d'oligonucléotides ont été utilisés : la séquence d'intérêt intral et un oligonucléotide contenant le consensus spécifique au facteur étudié afin de vérifier sa présence dans les extraits nucléaires de cellules U937 stimulées au PMA + ionomycine. Comme on peut le voir dans la partie gauche du graphique présenté à la figure 3.8A, aucune

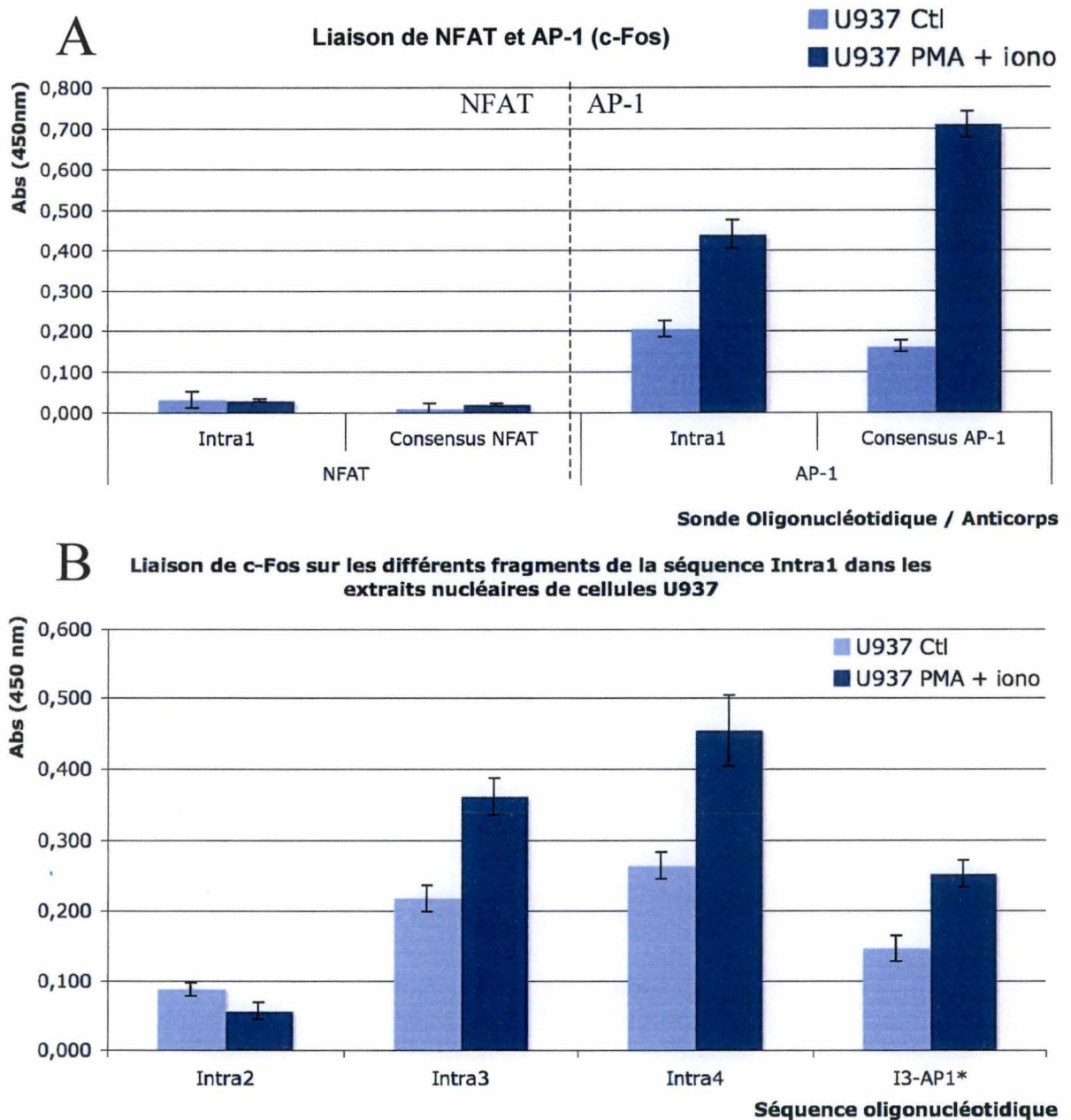


Figure 3.8 : Etude de la liaison de NFAT et c-Fos sur la séquence intra1 à partir d'extraits nucléaires de cellules U937

(A) 20 µg d'extraits nucléaires de cellules U937 non stimulées ou stimulées 1 heure au PMA (50nM) + ionomycine (285 nM) ont été incubés en présence de la séquence intra1 ou d'un oligonucléotide contenant la séquence consensus de NFAT (à gauche du graphique) ou d'AP-1 (à droite). (B) Liaison de c-Fos pour des extraits nucléaires de cellules U937 non stimulées ou stimulées 1 heure au PMA (50nM) + ionomycine (285 nM) après avoir été incubés en présence de 2,5 pmoles de la séquence intra2, intra3, intra4 et intra3 mutée au niveau du site putatif AP-1 (I3AP1*, voir Fig. 3.7 B).

liaison de NFAT n'est révélée quel que soit la sonde ou le type d'extraits nucléaires utilisés. Etant donné qu'aucune liaison de NFAT n'est observée sur la sonde contenant le site consensus de NFAT, il semblerait que la forme active de ce facteur ne soit pas présente dans ce type d'extrait nucléaire. Cependant, c-Fos semble quant à lui activé par le PMA + ionomycine dans les cellules U937, comme l'indique la liaison de AP-1 sur sa séquence consensus et que l'on retrouve également dans une moindre mesure sur la séquence intra1.

Etant donné que NFAT n'est pas activé dans les monocytes, nous avons voulu étudier la liaison de c-Fos sur les différents fragments de la séquence intra1 puisque théoriquement c-Fos ne peut pas se lier fortement au niveau du site composite en l'absence de NFAT. Aucune différence de liaison de c-Fos à la séquence intra2 n'est visible entre les extraits contrôles et stimulés alors que pour les trois autres oligonucléotides une liaison plus importante peut être observée dans les extraits provenant des cellules stimulées, par rapport aux cellules non traitées. Comme pour les extraits Jurkat, la différence de liaison entre les séquences intra2 et intra3 suggère la présence d'un site AP-1 dans les 20 nucléotides propres au fragment intra3. Cependant, le niveau de liaison de c-Fos à l'intra3 mesuré dans les extraits stimulés est faible. Il se pourrait donc que AP-1 puisse lier le site putatif présent au niveau du site composite mais avec une affinité plus faible, comme décrit précédemment pour les extraits Jurkat en absence de NFAT (Fig. 3.6 A). La liaison de c-Fos dans les extraits nucléaires U937 stimulés diminue encore pour la sonde I3-AP1* mutée pour le site AP-1 au nucléotide 4668 (Fig. 3.8B). Finalement, une activation de c-Fos est également observable pour le fragment intra4, comme dans les extraits nucléaires de cellules Jurkat stimulées par le PMA + ionomycine et en accord avec les prédictions *in silico*, suggérant la présence d'un site AP-1 dans la région intragénique correspondante.

En résumé pour cette première partie de recherche avec *a priori* de facteurs de transcription susceptibles d'expliquer l'inductibilité de la région intragénique du HIV-1 aux traitements par le TNF α et par le PMA + ionomycine, nous avons étudié trois facteurs de transcription : NF- κ B, AP-1 et NFAT. Les résultats obtenus pour NF- κ B suggèrent une liaison non spécifique à la séquence intra1, à l'opposé des facteurs AP-1 et NFAT.

Pour AP-1, le dimère liant probablement la séquence est de type c-Fos/JunD, bien que l'on ne puisse pas exclure un dimère c-Fos/JunB. Ces expériences demandent cependant à être répétées avec des extraits nucléaires de meilleure qualité. Les résultats obtenus pour c-Fos dans les extraits nucléaires de cellules Jurkat et U937 suggèrent une coopération entre NFAT et AP-1, comme le montre la chute de liaison de c-Fos et JunD lorsque le site putatif NFAT est muté (Fig. 3.7C et D). Cependant, les résultats obtenus pour NFAT sur les sondes mutées sont plus difficiles à interpréter. La coopération entre NFAT et AP-1 pour se lier à des sites composites est majoritairement décrite pour NFATc2. Or, les expériences d'ELISA-like ne permettent d'évaluer que la liaison de NFATc1. Il serait donc intéressant de pouvoir étudier la liaison de NFATc2 au niveau du site composite putatif par une autre méthode comme l'EMSA, étant donné que seul l'anticorps pour NFATc1 est disponible pour les ELISA-like. De plus, l'utilisation de l'EMSA est intéressante dans le cadre de l'étude de la coopération entre NFAT et AP-1 car elle permet de distinguer les facteurs liant l'ADN seule ou sous forme de complexe avec d'autres protéines augmentant le retard de migration.

Table 3.1 : Protéines identifiées par spectrométrie de masse dans des extraits nucléaires de cellules Jurkat (1mg) non stimulées après avoir été incubées en présence de 20 pmoles de la séquence intra1.

Nom de la protéine	Nombre de peptides identifiés
AHNAK nucléoprotéin isoform 1	15
poly (ADP-ribose) synthetase	15
Tubuline β	6
Structural maintenance of chromosome protein 2	1
Ef-1 δ (elongation factor-1-delta)	2
pancreatic tumor-related protein	1
eukaryotic translation elongation factor 1 beta	2
Frameshift mutation results in premature stop	1
Tubuline α	5
KIF4	4

Table 3.2 : Protéines identifiées par spectrométrie de masse dans des extraits nucléaires de cellules Jurkat (1mg) stimulées 5 heures au PMA (32 nM) + ionomycine (500nM) incubés en présence de 20 pmoles de la séquence intra1.

Nom de la protéine	Nombre de peptides identifiés
Thymopoietin α	5
RBP4 (retinoblastoma binding protein 4)	1
hnRNP (Heterogeneous nuclear ribonucleoproteins) A1	5
hnRNP (Heterogeneous nuclear ribonucleoproteins) D	3
hnRNP (Heterogeneous nuclear ribonucleoproteins) A2/B1	5
Keratin, type II cytoskeletal 1 (CK-1)	3
hnRNP (Heterogeneous nuclear ribonucleoproteins) AB	2
lamin B1	2
Ef-1 Δ (elongation factor-1-delta)	1

III.2. Etude sans a priori des protéines liant la région intragénique du HIV-1

III.2.1. Résultats préliminaires

Après une première partie visant à étudier la liaison de trois facteurs de transcription susceptibles d'expliquer l'inductibilité de la région intragénique du HIV-1 suite à un traitement par le TNF α ou par le PMA + ionomycine, le second objectif de ce travail était d'étudier cette même séquence, sans *a priori*, afin d'identifier de nouveaux régulateurs potentiels.

Un premier essai d'identification sans a priori des protéines liant la région intragénique du HIV-1 a été réalisé pour 1 mg d'extraits nucléaires de cellules Jurkat non stimulées ou stimulées 5 heures en présence de PMA (32 nM) + ionomycine (500 nM), 1 mg de billes magnétiques et 20 pm de trappeur desthiobiotinylés suivant la méthode décrite à la figure 1.9. Les différentes protéines identifiées et le nombre de peptides séquencés pour chacune d'entre elles sont référencées dans les tables 3.1 et 3.2.

Etant donné que ces résultats préliminaires sont le fruit d'un premier essai sur la séquence intral, nous ne nous sommes intéressés, à titre indicatif, qu'à la description des protéines retrouvées dans la condition d'intérêt, à savoir les cellules stimulées en présence de PMA + ionomycine. En effet, l'objectif principal de cette expérience est avant tout d'avoir un premier aperçu du nombre et du type de protéines que l'on peut identifier, afin de mettre en avant d'éventuelles améliorations de la technique.

Parmi les protéines identifiées dans les extraits nucléaires des cellules stimulées en présence de PMA + ionomycine, quatre facteurs identifiés appartiennent à la famille des hnRNPs (heterogeneous nuclear ribonucleoproteins). Ces ribonucléoprotéines, dont une vingtaine de protéines différentes ont été décrites chez l'humain, sont impliquées dans des processus comme la maturation, le maintien de la stabilité et le transport cytoplasmique des ARNms. Elles possèdent des domaines KH (K homologous) qui leur permettent d'interagir entre elles, mais aussi de lier l'ARN avec une forte affinité, et dans une moindre mesure l'ADN. Certaines ribonucléoprotéines comme l'hnRNP K sont aussi impliquées dans la régulation de la transcription, en interagissant avec des séquences régulatrices spécifiques et l'ARN polymérase II (Kim, Hahm et al. 2000). La protéine Ef-1 Δ (Eukaryotic translation elongation factor 1 delta) est une sous-unité du complexe facteur d'élongation 1 impliqué dans le processus de régulation de l'élongation de la traduction, en agissant comme un facteur d'échange de la guanine. Cette protéine est également connue pour interagir avec la protéine transactivatrice Tat du HIV-1, réprimant la traduction des gènes de l'hôte et favorisant celle des ARNm du virus (Xiao, Neuveut et al. 1998).

Après ce groupe de protéines liant préférentiellement l'ARN, trois protéines impliquées dans l'architecture nucléaire ont aussi été identifiées : la lamine B1, la thymopoiétine α et la protéine KIF4. La lamine B1 appartient à la famille des lamines de type B et est encodée par le gène *LMNB1*. Ce groupe de protéines est connu pour jouer un rôle majeur dans l'architecture nucléaire, l'organisation de la chromatine et la distribution des complexes formant les pores nucléaires. Cependant, elles peuvent également être impliquées dans des processus de transcription, réparation et réplication de l'ADN. Par exemple, il semblerait que l'activité de répression de Oct-1 soit corrélée avec son association avec la lamine B1 (Imai, Nishibayashi et al. 1997). Pour une revue complète sur les fonctions des lamines et des pathologies qui leurs sont associées voir (Parnaik 2008). La thymopoiétine

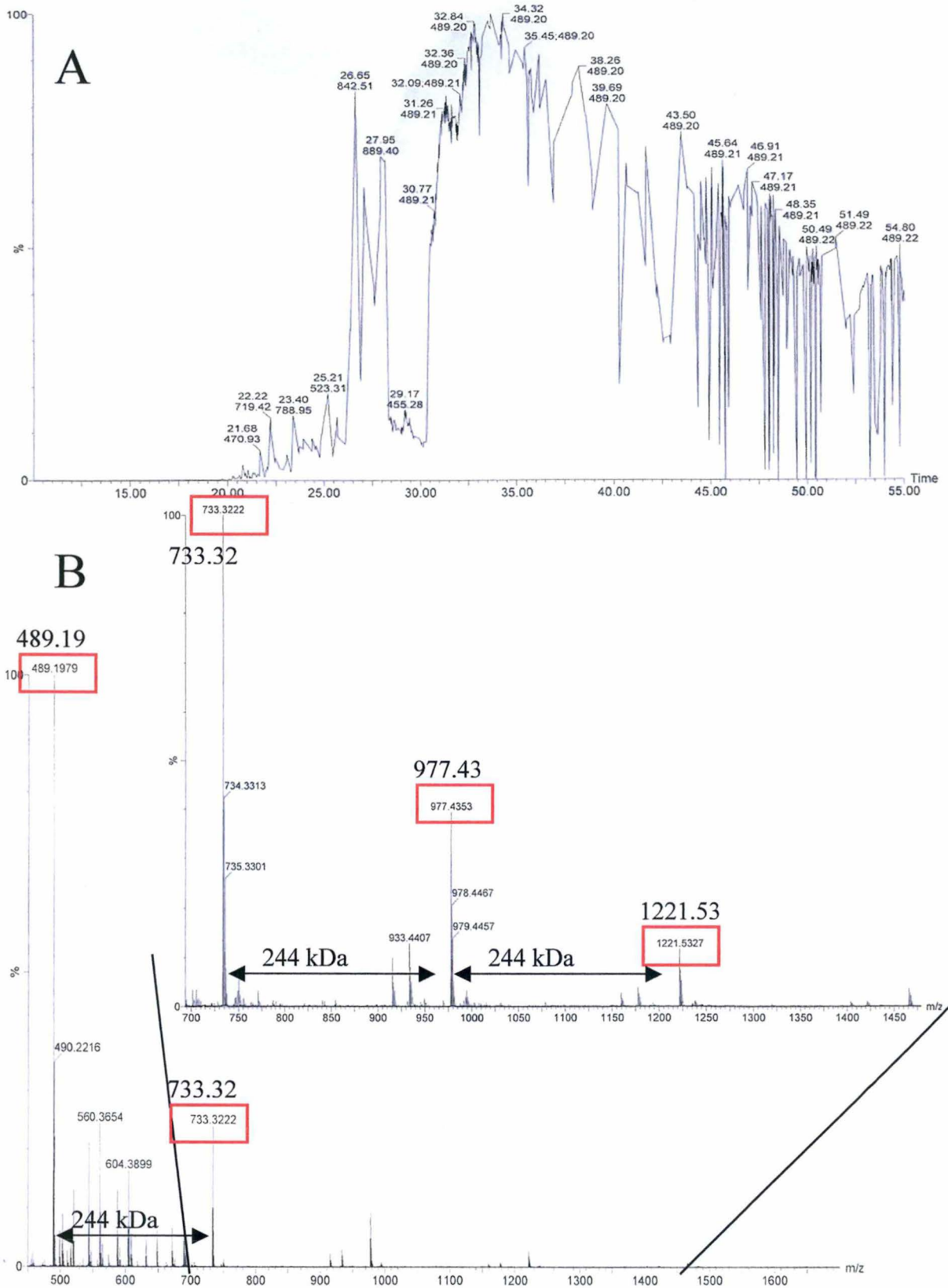


Figure 3.9 : Chromatogramme (A, représente l'abondance des peptides en fonction du temps d'éluion) et spectre MS (B, représente l'abondance des peptides en fonction des rapports m/z) obtenus lors de l'analyse en spectrométrie de masse des protéines capturées par la séquence intral en présence d'extraits nucléaires (1mg) de cellules Jurkat stimulées 5 heures au PMA + ionomycine.

α est issue de l'épissage alternatif du gène *tmpo* et est exprimée de manière diffuse dans le noyau, alors que les isoformes β et γ sont localisées au niveau de la membrane nucléaire. La protéine murine KIF4 (kinesin superfamily member 4) est impliquée dans le transport de vésicules cytoplasmiques, cependant, des expériences d'immunocytochimie ont également montré une localisation nucléaire de son homologue humain, associé avec la matrice nucléaire. Durant la mitose, KIF4 semble être associée aux chromosomes et pourrait avoir une fonction motrice dans le transport de l'ADN en association avec les microtubules (Lee, Lee et al. 2001).

Malgré les précautions employées (travail en salle blanche), des contaminants expérimentaux comme la kératine sont tout de même détectés. Finalement, le facteur liant la protéine RBBP4 (retinoblastoma binding protein 4) sera décrit par après (point III.2.3.3).

Ces premiers résultats préliminaires, pour lesquels la majeure partie des facteurs identifiés sont décrits comme localisés dans le noyaux et capables de lier l'ADN, ont permis de mettre en évidence la nécessité de diverses mises au point.

Premièrement, aucun pic d'autodigestion de la trypsine n'a pu être détecté (ou en faible quantité lors d'expériences précédentes) dans aucune des deux conditions. La digestion des protéines n'est peut-être pas optimale, suite à une quantité de trypsine (900 ng) trop faible.

Deuxièmement, le chromatogramme généré lors de la séparation des peptides par la nano-chromatographie en phase inverse (nano-LC) qui précède la phase d'ionisation du spectromètre de masse fournit également des informations qualitatives sur l'échantillon (Fig 3.9A). Ce chromatogramme, qui indique l'abondance des peptides en fonction du temps d'élution, nous a permis de poser l'hypothèse d'une trop grande complexité de nos échantillons, limitant la détection des peptides présents en plus faible quantité. En effet, comme on peut le voir majoritairement entre 25 et 40 minutes, l'abondance des peptides reste très élevée continuellement au cours du temps, et est responsable d'une saturation de détection des masses peptidiques et de leur séquençage. Suite à cette saturation du signal, seuls les peptides les plus abondants, correspondant comme nous venons de le voir à des protéines liant l'ADN de manière non spécifique, seront détectés et séquencés. Les peptides recherchés, comme par exemple ceux correspondant à des facteurs de transcription liant spécifiquement la séquence intral, sont quant à eux en plus faible quantité et leur probabilité de détection est moindre. Outre la saturation de détection, la complexité de l'échantillon a également pour conséquence que l'ensemble des sites de liaison de la nano-LC en amont du spectromètre de masse sont occupés. La séparation des peptides selon leur hydrophobicité n'est alors plus optimale. Ces différents éléments nous ont amenés à penser qu'il fallait décomplexifier les échantillons pour l'analyse en spectrométrie de masse afin de favoriser la détection des peptides en plus faible abondance.

Finalement, la figure 3.9B, correspondant au spectre MS pour ces mêmes échantillons, représente l'abondance des peptides en fonction de leur m/z . Ces spectres nous ont permis de mettre en évidence la présence d'un contaminant non identifié et présent en abondance, dont la m/z correspond toujours à un multiple de 244 kDa : 489.19, 733.32, 977.43 et 1221.53. La présence de ce contaminant pourrait en partie expliquer la complexité de l'échantillon mise en évidence précédemment et la faible détection des peptides peu abondants. Des expériences ultérieures isolant chacune des sources potentielles de ce contaminant (billes magnétiques, sondes oligonucléotidiques, extrait nucléaire, biotine et étape de digestion) ont permis d'identifier ce contaminant : il correspond à des polymères de biotine (données non montrées). En effet, la biotine, de poids moléculaire 244.13 kDa, est utilisée en excès afin de déplacer les complexes sonde nucléiques-protéines des billes.

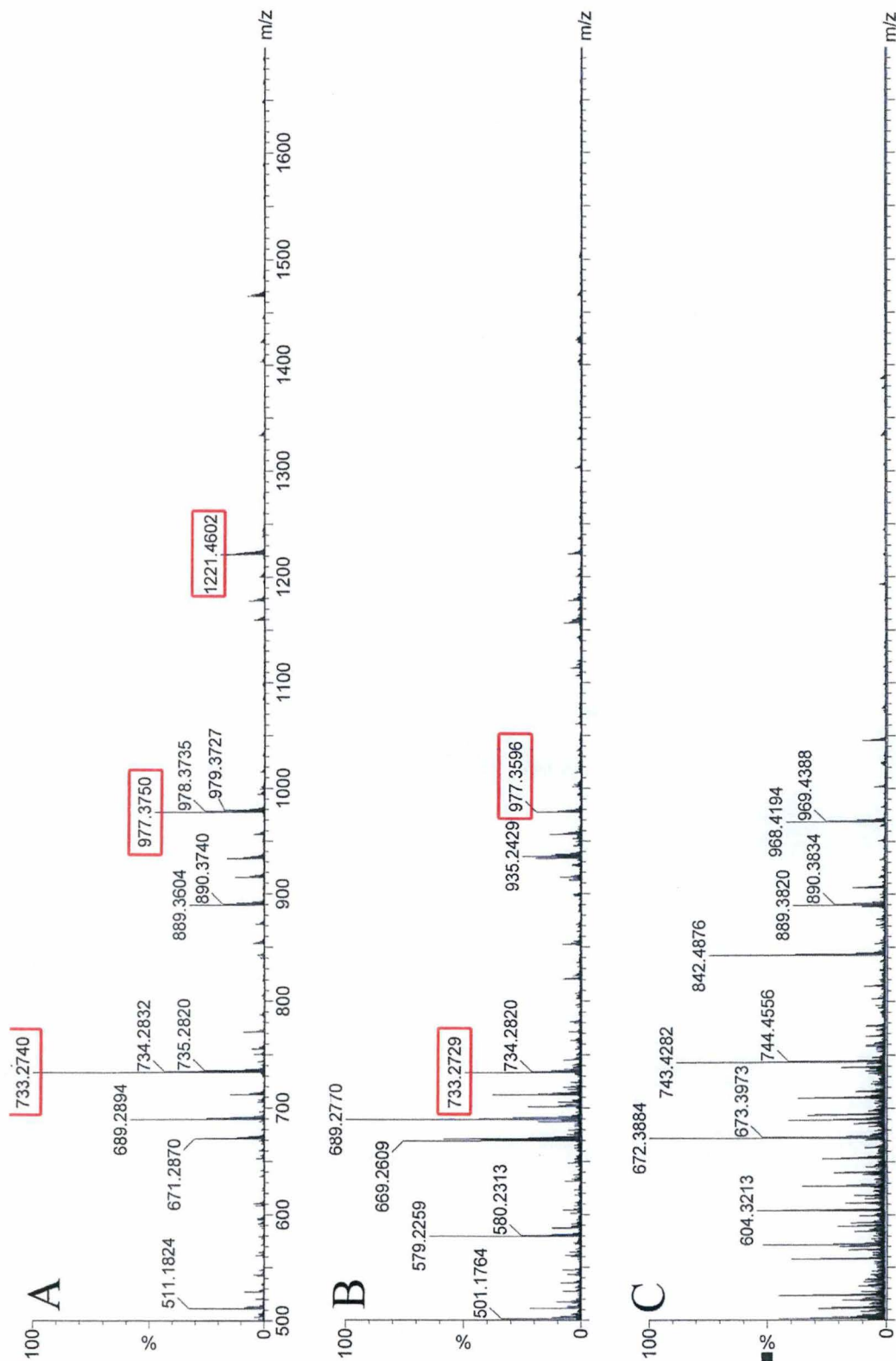


Figure 3.10 Spectres MS des protéines identifiées après une étape d'élimination de la biotine ou non.

Abondance des peptides en fonction des rapports m/z suite à l'incubation de 1 mg d'extraits nucléaires de cellules Jurkat avec la séquence intral : sans lavage de la biotine (A), lavage de la biotine avec 300 µg (B) et 500 µg (C) de billes magnétiques tapissées de streptavidine.

Suite à ces résultats préliminaires obtenus avec la séquence intral et les extraits nucléaires de cellules Jurkat stimulées 5 heures avec du PMA + ionomycine, trois paramètres doivent être améliorés : optimiser la digestion des protéines purifiées, éliminer l'excès de biotine et simplifier l'échantillon pour son analyse en spectrométrie de masse en le fractionnant. Pour améliorer ces paramètres, nous avons choisi d'utiliser la séquence oligonucléotidique LTR du HIV-1, précédemment utilisée par A. Tacheny pour la mise au point générale de la technique (voir point I.2.2.2 de l'introduction). En effet, nous savons que la capture des protéines présentes dans un extrait nucléaire de cellules HeLa stimulées 45 minutes à l'IL-1 β (5ng/ml) par la séquence du LTR du HIV-1 immobilisée sur billes permet la détection d'au moins un facteur de transcription : le facteur NF- κ B. L'abondance des peptides correspondant à NF- κ B sera donc le témoin de l'amélioration de ces 3 paramètres.

III.2.2. Mises au point

III.2.2.1. Amélioration de la digestion trypsique.

Tout d'abord, l'optimisation de la digestion des protéines par la trypsine a été réalisée. Pour ce faire, nous avons doublé la quantité de trypsine incubée avec les protéines purifiées grâce à la sonde oligonucléotidique, passant de 900 à 1800 ng de trypsine pour 1mg d'extraits nucléaires incubés en présence de 20 pmoles de sonde oligonucléotidique immobilisée. Les premiers résultats obtenus montrent que le nombre de peptides différents ayant pu être séquencés est plus élevé lorsque la quantité de trypsine est doublée (données non montrées). Nous avons donc utilisé 1800 ng de trypsine lors de nos analyses ultérieures.

III.2.2.2. Elimination de la biotine

Le second obstacle auquel il fallait faire face était la présence de polymères de biotine. Pour résoudre ce problème, nous avons incubé l'échantillon, directement après la digestion trypsique, 30 minutes en présence des billes magnétiques tapissées de streptavidine afin de capturer la biotine en excès. Cette nouvelle étape d'incubation ne pose pas de problèmes par rapport aux protéines purifiées car ces dernières ont été préalablement digérées en peptides et ne sont plus liées aux sondes oligonucléotidiques qui pourraient se lier à la streptavidine grâce à leur desthiobiotine. Seuls les surnageants contenant les peptides exempts de la vitamine non désirée sont gardés pour être analysés par spectrométrie de masse. Trois quantités de billes magnétiques ont été testées (300, 400 et 500 μ g), sur base de l'estimation d'excès de biotine présent après déplacement de la sonde et de la capacité de liaison des billes magnétiques. Le spectre MS des analyses réalisées peut être observé à la figure 10. Comme on peut le voir, lorsque la biotine n'est pas éliminée (Fig. 3.10 A) les polymères de m/z multiples de 244 kDa sont toujours présents. Après une incubation avec 300 μ g de billes recouvertes de streptavidine (Fig. 3.10 B), la présence de ces peptides est déjà fortement diminuée mais ne disparaît totalement que pour une incubation avec 500 μ g de billes (Fig. 3.10 C). Cet essai d'élimination n'a été réalisé qu'une seule fois et des expériences ultérieures nous ont montré qu'il pouvait encore y avoir des traces de biotine, mais en quantité beaucoup plus faible que précédemment, c'est pourquoi nous avons finalement choisi d'incuber le mélange peptidique avec 600 μ g de billes magnétiques.

III.2.2.3. Diminution de la complexité des échantillons

Si l'élimination de la biotine permet de désaturer les sites de liaison de la colonne C18 en amont du spectromètre de masse et d'augmenter la détection des peptides d'intérêt, il semble que la complexité de l'échantillon représente toujours un obstacle à l'identification des peptides les moins abondants. Il paraît nécessaire de fragmenter l'échantillon avant l'analyse en spectrométrie de masse afin d'obtenir des fractions plus simples pour lesquelles il n'y aura pas de saturation de détection.

Pour ce faire, le mélange peptidique obtenu après digestion et retrait de la biotine est séparé par chromatographie échangeuse de cations (HPLC-SCX, en collaboration avec Aurélien Boland, URBV). Lors de cette étape, les différents peptides vont être progressivement désorbés de la colonne, chargée négativement, en fonction des charges qu'ils portent, par le biais d'un gradient croissant de formate d'ammonium. Ce sel particulier possède la caractéristique d'être volatile, ce qui permet de l'éliminer avant l'analyse en spectrométrie de masse. Un volume de départ d'environ 200 μ l, approximativement 80 μ l d'échantillon mis à volume avec une solution d'acétonitrile 30%, a de la sorte été fractionné en 40 fractions. Les 5 premières et les 5 dernières fractions sont rassemblées car elles ne contiennent que peu de peptides et correspondent à ce qui ne s'est pas lié à la colonne ou à l'étape finale de rinçage. Les autres fractions sont réunies par deux afin de diminuer le nombre d'échantillons à analyser par spectrométrie de masse. Il est important de noter qu'il est maintenant nécessaire de travailler à partir de 2 mg d'extraits nucléaires afin d'avoir du matériel en suffisance vu le nombre important de fractions générées lors de l'HPLC-SCX. Pour ce faire, la purification sur billes magnétiques a lieu dans des conditions similaires aux précédentes, c'est-à-dire 1 mg d'extraits nucléaires pour 1 mg de billes magnétiques et 20 pmoles de sondes oligonucléotidiques. Cependant, chaque condition est réalisée en double, et les surnageants sont réunis après retrait de la biotine pour être fractionnés par l'HPLC-SCX. Finalement, toutes les données acquises pour les différentes fractions sont rassemblées pour les confronter aux bases de données et identifier les protéines capturées.

Le premier essai de fractionnement des échantillons a été réalisé sur un fragment du LTR du HIV-1 et pour des extraits nucléaires de cellules HeLa non stimulées ou stimulées 45 minutes à l'IL-1 β (voir point I.2.2.2), permettant l'identification de plus de 100 protéines différentes en comparaison à la vingtaine habituellement trouvée. Parmi celles-ci, nous avons pu identifier le facteur NF- κ B (p65) ainsi que les habituelles hnRNPs et d'autres protéines liant l'ARN. Il paraît évident que le nombre de protéines identifiées est énorme pour une séquence d'environ 250 pb et que bon nombre des candidats ont été capturés suite à une liaison non spécifique à cet oligonucléotide. Certaines protéines ont été identifiées dans les deux conditions, c'est pourquoi il est intéressant de faire une analyse différentielle des protéines ne se trouvant que dans les extraits contrôles ou stimulés. Cependant, des problèmes techniques liés à l'analyse en spectrométrie de masse ont conduit à la perte d'informations pour la condition contrôle pouvant biaiser cette comparaison.

Etant donné que d'une part l'identification des protéines capturées par le LTR du HIV-1 constitue davantage le sujet de recherche d'Aurélien Tacheny que le mien, et que d'autre part les données de cette expérience sont incomplètes, nous ne détaillerons pas ici la fonction de cette centaine de protéines identifiées dans ce modèle cellulaire.

Ces différentes mises au point, reprises à la figure 3.11, nous ont donc permis d'améliorer nettement le nombre de protéines identifiées par spectrométrie de masse lorsque celles-ci sont capturées par une séquence oligonucléotidique particulière. Dans la dernière partie des résultats, nous décrirons comment ces améliorations ont permis d'identifier des

Table 3.3 : Protéines identifiées par spectrométrie de masse dans les extraits nucléaires de cellules Jurkat (2mg) non stimulées et stimulées 5 heures au PMA + ionomycine après avoir été incubés en présence de 20 pmoles de la séquence intra1.

N°	Nom uniprot
Q14008	Cytoskeleton-associated protein 5
P22626	hnRNP A2/B1 (Heterogeneous nuclear ribonucleoproteins A2/B1)
P09651	hnRNP A1 (Heterogeneous nuclear ribonucleoprotein A1)
P52272	hnRNP M (Heterogeneous nuclear ribonucleoprotein M)
P07910	hnRNP C1/C2 (Heterogeneous nuclear ribonucleoproteins C1/C2)
Q12771	P37 AUF1
P26641	Ef-1 γ (Elongation factor 1-gamma)
Q92841	DDX17 (Probable ATP-dependent RNA helicase)
P43243	Matrin-3
P31943	hnRNP H (Heterogeneous nuclear ribonucleoprotein H)
P23246	SFPQ (Splicing factor, proline- and glutamine-rich)
Q15717	ELAV-like protein 1
Q07666	KH domain-containing, RNA-binding, signal transduction-associated protein 1
Q7Z780	U2 small nuclear RNA auxiliary factor 1
P61978	hnRNP K (Heterogeneous nuclear ribonucleoprotein K)
O75494	FUSE-interacting serine-arginine-rich protein 1
P26599	hnRNP I (Heterogeneous nuclear ribonucleoprotein I)
P68104	Ef-1 α (Elongation factor 1-alpha 1)
Q99729	hnRNP A/B (Heterogeneous nuclear ribonucleoprotein A/B)
Q92945	FUSE (Far upstream element) binding protein 2
Q53EW3	Regulatory factor X, 5 variant
Q16666	Ifi-16 (Gamma-interferon-inducible protein)
Q12906	Interleukin enhancer-binding factor 3
Q59EQ0	Special AT-rich sequence binding protein 1 variant
P78347	TFII-I (General transcription factor II-I)
Q15424	SAB-B1 (Scaffold attachment factor B1)
O00287	Regulatory factor X-associated protein
Q96AE4	Far upstream element-binding protein 1
Q14151	Scaffold attachment factor B2
Q12905	Interleukin enhancer-binding factor 2
Q86V81	THO complex subunit 4
P10244	Myb-related protein B
Q6MZP7	Protein lin-54 homolog
Q8CD94	Protein lin-52 homolog
Q5TKA1	Lin-9 homolog
O14593	DNA-binding protein RFXANK
O95239	Chromosome-associated kinesin KIF4A
Q9UQE7	Structural maintenance of chromosomes protein 3
Q14683	Structural maintenance of chromosomes protein 1A
A6NHR9	Structural maintenance of chromosomes flexible hinge domain-containing protein 1
Q9NTJ3	Structural maintenance of chromosomes protein 4
P13645	Keratin, type I cytoskeletal 10
P02768	Serum albumin
P18887	DNA-repair protein XRCC1
Q15233	NonO (Non-POU domain-containing octamer-binding protein)
O60216	Double-strand-break repair protein rad21 homolog

P11387	DNA topoisomerase 1
P49916	DNA ligase 3
P09874	PARP (Poly [ADP-ribose] polymerase 1)
Q9NTI5	Sister chromatid cohesion protein PDS5 homolog B
P42166	Lamina-associated polypeptide 2 isoform alpha
P20700	Lamin-B1
P11182	Lipoamide acyltransferase component of branched-chain alpha-keto acid dehydrogenase complex, mitochondrial
P00747	Plasminogen
Q00325	Phosphate carrier protein, mitochondrial
P04406	Glyceraldehyde-3-phosphate dehydrogenase
Q9P258	Protein RCC2
Q65ZQ3	FBRNP
Q92608	Dedicator of cytokinesis protein 2
Q9ULM3	YEATS domain-containing protein 2

« Abération »

Protéines majoritairement décrites pour lier l'ARN

Protéines associées à la maintenance des chromosomes

Contaminants expérimentaux

Protéines impliquées dans la régulation transcription

Protéines associées à des processus liés à l'ADN

protéines capturées par le enhancer intragénique du HIV-1 (séquence intra-1), mise en présence d'extraits nucléaires provenant de cellules Jurkat stimulées ou non.

III.2.2.4. Etude des protéines liant la région intragénique du HIV-1 après amélioration de la technique

La séquence intra1 immobilisée sur des billes magnétiques a été utilisée pour capturer des protéines d'extraits nucléaires de cellules Jurkat non stimulées ou stimulées 5 heures avec du PMA + ionomycine. Comme pour les extraits nucléaires de cellules HeLa, le fractionnement des échantillons a permis d'augmenter le nombre de protéines identifiées : 102 pour les extraits provenant de cellules non stimulées et 114 protéines dans les extraits de cellules stimulées par le PMA + ionomycine.

Etant donné le nombre important de protéines identifiées pour une séquence de 300 paires de bases, et afin de mettre en évidence les protéines spécifiques à chaque condition, les protéines identifiées ont été classées selon 3 catégories : celles présentes dans les deux conditions, uniquement dans les contrôles et exclusivement dans la condition PMA + ionomycine. Cette analyse différentielle a permis de mettre en évidence 43 protéines spécifiques à la condition PMA + ionomycine et 32 pour les extraits nucléaires contrôles.

III.2.2.5. Protéines identifiées dans les conditions stimulées et non stimulées

Dans la première catégorie, constituée des protéines présentes dans les deux conditions, une importante fraction d'entre elles sont décrites pour interagir et être impliquées dans des processus biologiques liés aux ARNms. Comme précédemment, bon nombre de ces protéines interagissent avec l'ARN par l'intermédiaire de leur domaine KH, comme par exemple les **hnRNP**. D'autres protéines décrites pour interagir avec les acides ribonucléiques ont aussi été identifiées, comme l'hélicase putative **DDX17**, des protéines du cytosquelette impliquées dans le transport (**CKAP5** : Cytoskeleton-associated protein 5), l'épissage (**FUSE-binding protein 2**) et la traduction (**Ef-1 γ**) des ARNms. La signification de la présence de ces protéines n'est pas évidente car bon nombre sont décrites majoritairement pour leur rôle régulateur au niveau des ARNms, mais certaines d'entre elles sont aussi connues pour agir au niveau de la transcription comme l'**hnRNP K** ou la protéine **NonO** (Non-POU domain-containing octamer-binding protein). En effet, cette protéine est impliquée dans la maturation et l'épissage des ARNm mais également dans la réplication, la réparation de l'ADN, ainsi que dans la régulation de la transcription. Dans ce dernier cas, il a été montré que NonO pouvait se lier à l'enhancer du LTR de l'IAP (intracisternal A-particle) murin et activer sa transcription. NonO pourrait également être un co-activateur du domaine transactivateur N-terminal de l'AR (androgen receptor) (Ishitani, Yoshida et al. 2003). Enfin, en plus de sa capacité à lier l'ADN, NonO peut aussi recruter certains facteurs de transcription ou lier directement certains d'entre eux comme PU.1 (Shav-Tal and Zipori 2002). Le **SFPQ** (Splicing factor, proline- and glutamine-rich, aussi appelé PSF) est un facteur essentiel pour l'épissage des ARNm et pour la formation du « spliceosome ». Il peut interagir avec NonO, également lors de processus d'épissage mais aussi pour dérouler l'ADN et réguler la transcription, notamment en réprimant les gènes par recrutement d'histone déacétylases (HDAC). Enfin, parmi d'autres exemples de régulation de la transcription, le complexe PSF-NonO a été décrit pour permettre le couplage de la transcription avec l'épissage des ARNm en présence du domaine C-terminal de l'ARN PII (Rosonina, Ip et al. 2005). Pour une revue

Table 3.4 : Protéines identifiées par spectrométrie de masse uniquement dans les extraits nucléaires de cellules Jurkat (2mg) non stimulées après avoir été incubés avec 20 pmoles de la séquence intra1.

N° Uniprot	Nom Uniprot	Nombre de ppt. Identifiés
Q9H7G6	CDNA: FLJ20897 fis, clone ADKA03573	4
Q05639	Elongation factor 1-alpha 2	2
Q15046	Lysyl-tRNA synthetase	2
P41252	Isoleucyl-tRNA synthetase, cytoplasmic	2
P07814	Bifunctional aminoacyl-tRNA synthetase	2
Q07955	Splicing factor, arginine/serine-rich 1	3
Q00839	hnRNP U (Heterogeneous nuclear ribonucleoprotein U)	2
Q2KHR2	RFX 7 (Regulatory factor X 7)	2
Q13547	HDAC 1 (Histone deacetylase 1)	1
P51610	HCF-1 (Host cell factor)	1
P78527	DNA-PK (DNA-dependent protein kinase catalytic subunit)	3
Q8N3U4	Cohesin subunit SA-2	1
Q86UE8	TLK2 (Serine/threonine-protein kinase tousled-like 2)	1
O95347	SMC2 (Structural maintenance of chromosomes protein 2)	6
Q03252	Lamin-B2	5
P31689	DnaJ homolog subfamily A member 1	1
P11498	Pyruvate carboxylase, mitochondrial	2
P02768	Serum albumin	2
Q9UFC0	Leucine-rich repeat and WD repeat-containing protein 1	2
Q8IWQ1	TGS2	2
O75410	Transforming acidic coiled-coil-containing protein 1	3
O43933	Peroxisome biogenesis factor 1	1
O15187	T-cluster binding protein (TIA1 cytotoxic granule-associated RNA binding protein-like 1, isoform CRA_a)	1

sur PSF et NonO voir (Shav-Tal and Zipori 2002). La facteur **Matrin-3** a été décrit pour participer à la rétention nucléaire des ARNm défectueux en association avec NonO-SFPQ.

Mises à part ces protéines liant préférentiellement l'ARN, des facteurs impliqués dans des processus biologiques associés à l'ADN sont également identifiés. C'est par exemple le cas de la **topoisomérase 1**, l'**ADN ligase 1**, de protéines associées au maintien des chromosomes (**hCAP** : Chromosome-associated polypeptide) ou à la réparation de l'ADN (**PARP** : Poly [ADP-ribose] polymerase 1).

Une dizaine de protéines identifiées jouent un rôle dans le contrôle de l'expression des gènes comme par exemple le facteur général **TFII-I** (revu par (Roy 2007)). TFII-I est capable de lier des séquences régulatrices de type Inr (voir introduction point I.1.2) et pourrait dès lors faire le lien entre des éléments *cis* et la machinerie basale de transcription. Différentes isoformes ont été décrites et semblent avoir des fonctions opposées. Alors que l'isoforme TFII- β est recrutée en conditions basales au niveau du promoteur du gène *c-fos* et est responsable de la répression de ce dernier, elle est exportée dans le cytoplasme en présence de facteurs de croissance. A l'opposé, TFII- Δ est localisée dans le cytoplasme à l'état basal mais est exportée dans le noyau et active l'expression de c-Fos sous l'influence de facteurs de croissance (Hakre, Tussie-Luna et al. 2006).

Les protéines **b-Myb** et les homologues **Lin9**, **Lin54** et **Lin52** sont décrits comme appartenant à un complexe DREAM impliqué dans la répression des gènes spécifiques au cycle cellulaire (Pilkinton, Sandoval et al. 2007).

Plusieurs protéines induites par l'IFN γ sont également identifiées telles que le facteur **Ifi16** (IFN γ inducible protein), protéine nucléaire possédant les caractéristiques des facteurs de transcription, à savoir un domaine de liaison à l'ADN et un domaine de régulation de la transcription. De plus, il peut exercer une activité de répresseur transcriptionnel en se liant directement à p53 (Johnstone, Kerry et al. 1998; Kwak, Ongusaha et al. 2003).

Enfin, le facteur de régulation X (**RFX**) est impliqué dans la régulation de la transcription des gènes MHCII (Major histocompatibility class II) (Garvie, Stagno et al. 2007).

Malgré tout, la présence de certaines protéines semble aberrante puisqu'elles n'ont aucun lien évident avec les acides nucléiques ou tout autre processus qui pourrait leur être associé, comme le plasminogène ou encore des composants de complexes mitochondriaux.

Enfin, des contaminants expérimentaux comme la kératine ou l'albumine, présentent dans le tampon de conservation des billes magnétiques, ont également été identifiés dans les deux conditions.

III.2.2.6. Protéines identifiées uniquement dans les conditions non stimulées

La deuxième catégorie comprend les protéines que l'on ne retrouve que dans les extraits nucléaires de cellules Jurkat non stimulées et qui pourraient contenir des facteurs de répression, dont la localisation subcellulaire ou l'activité a été modifiée suite à la stimulation.

A nouveau, une dizaine de protéines majoritairement décrites dans des processus liés aux ARNs ont été identifiées. Certains facteurs appartiennent à la même famille que des protéines de la catégorie précédente : **Ef-1 α** , les **hnRNPs U** ainsi que des facteurs d'épissage (Splicing factor, arginine/serine-rich 1).

Trois protéines sont décrites pour être impliquées dans la régulation de la transcription : l'**HDAC 1** (histone deacetylase 1), l'**HCF-1** (host cell factor) et le **RFX 7** (Regulatory factor X 7). L'HDAC est connue pour sa capacité à induire la répression des

Table 3.5 : Protéines identifiées par spectrométrie de masse uniquement dans les extraits nucléaires de cellules Jurkat (2mg) stimulées 5 heures au PMA (32 nM) + ionomycine (500 nM) après avoir été incubés avec 20 pmoles de la séquence intra1

N°	Uniprot	Nom uniprot	Nombre de ppt ≠ identifiés
Q9Y230	RuvB-like 2		6
Q9Y265	RuvB-like 1		5
Q13769	THO complex subunit 5 homolog		2
Q96FV9	THO complex subunit 1		1
Q8NI27	THO complex subunit 2		4
Q6I9Y2	THO complex subunit 7 homolog		1
O96019	Arp Baf53a (Actin-related protein Baf53a)		3
Q9H981	Arp8 (Actin-related protein 8)		1
Q6P4R8	NFRKB (Nuclear factor related to kappa-B-binding protein)		9
Q09028	RBBP4 (retinoblastoma binding protein 4)		2
Q9UKL0	coREST (REST corepressor 1)		1
Q01658	Dr1 (=NC-2β)		1
P25490	YY1		3
O60341	histone demethylase AOF1		2
Q9UHB7	MCEF (Major CDK9 elongation factor-associated protein)		1
P19338	Nucleolin		1
Q14839	Chromodomain-helicase-DNA-binding protein 4		2
Q8TET7	INO80 complex subunit E		1
Q9NX08	COMMD8 (COMM domain-containing 8)		1
Q6K0P9	PYHIN1 (Interferon-inducible protein X)		2
Q99986	VRK1 (Serine/threonine-protein kinase 1)		3
P60709	Beta-actin (Actin, cytoplasmic 1)		1
Q13422	Ikaros		5
Q8IVW6	ARID domain-containing protein 3B		1
P62306	Small nuclear ribonucleoprotein F		1
P29692	Ef-1Δ (Elongation factor 1-delta)		2
Q8NC51	Plasminogen activator inhibitor 1 RNA-binding protein		1
Q567R9	CIP29 protein (cytokine induced protein 29 kDa)		2
O75526	RNA-binding motif protein, X-linked-like-2		2
Q71U36	Tubulin alpha-1A chain		3
Q9H853	Putative tubulin-like protein alpha-4B		2
Q9NY65	Tubulin alpha-8 chain		2
P54652	Heat shock-related 70 kDa protein 2		2
P11142	Heat shock cognate 71 kDa protein		3
Q13416	ORC2L (Origin recognition complex subunit 2)		1
Q29RF7	SCC-112 (Sister chromatid cohesion protein PDS5 homolog A)		2
Q9Y3Y2	Uncharacterized protein C1orf77		2
O43805	SSNA1 (Sjoegren syndrome nuclear autoantigen 1)		1
O95785	Wiz (widely interspaced zinc finger motifs)		2
O75152	ZC3H11A(Zinc finger CCCH domain-containing protein 11A)		1

gènes suite à la déacétylation des lysines présentes dans le domaine amino-terminal des histones. Le facteur HCF-1 est un coactivateur transcriptionnel pour plusieurs facteurs de transcription (Luciano and Wilson 2003). Il participe également à la régulation du virus HSV-1. En effet, le facteur viral VP16 pourrait activer la phase lytique du virus en permettant l'expression des gènes précoces du HSV-1 suite à la formation d'un complexe avec Oct-1 et HCF-1 (Wysocka and Herr 2003).

Quatre autres protéines sont en relation avec des processus liés à l'ADN. L'**ADN-PK** (DNA-dependent protein kinase catalytic subunit) est une sérine thréonine kinase agissant comme senseur des dommages à l'ADN et possédant de nombreux substrats comme histone H1, PARP1, c-Jun, p53 et c-myc. La protéine du virus HIV-1 Tat est connue pour interagir avec cette ADN-PK. Il a été montré que Tat pouvait se lier à la kinase et augmenter la phosphorylation de Sp1 dépendante de l'ADN-PK lors de la transcription du provirus à partir du LTR-5' (Chun, Semmes et al. 1998).

La protéine « **Cohesin subunit SA-2** » est comme son nom l'indique une sous-unités du complexe cohésine impliqué dans la cohésion des chromatides sœurs après la réplication de l'ADN (source : uniprot). La protéine **TLK2** (Tousled-like kinase 2) est une sérine thréonine kinase nucléaire impliquée dans l'assemblage de la chromatine. Enfin, la protéine **SMC2** (Structural maintenance of chromosomes protein 2) appartient au complexe « condensin » impliqué dans la dynamique des chromosomes.

Finalement, quelques protéines dont la fonction ne semble pas en relation avec la régulation de la transcription, ont également été identifiées comme la co-chaperone mitochondriale homologue de DnaJ et la pyruvate carboxylase.

III.2.2.7. Protéines identifiées uniquement dans les extraits nucléaires provenant de cellules stimulées par le PMA+ionomycine

Finalement, la troisième catégorie comprend les protéines qui n'ont été identifiées que dans les extraits nucléaires de cellules Jurkat stimulées 5 heures en présence de PMA + ionomycine (table 3.5). A l'opposé des deux autres catégories, peu de protéines liant l'ARN sont identifiées, parmi lesquelles deux ribonucléoprotéines et un facteur d'élongation (**Ef-1δ**). Enfin, la protéine **CIP29** (cytokine-induced protein 29 kDa) a été décrite comme facteur associé au cycle cellulaire mais a également été identifiée pour interagir avec l'ARN Helicase DDX39 en liant l'ARN et favorisant l'activité de cette protéine (Sugiura, Sakurai et al. 2007).

Bon nombre de facteurs identifiés dans cette catégorie sont décrits pour être impliqués directement ou indirectement dans la régulation de la transcription. La protéine **MCEF** (Major CDK9 elongation factor-associated protein) membre de la famille AF4/FMR2, un composant du facteur P-TEFb (Positive Transcription Elongation Factor b) contenant la CDK9 (cycin dependent kinase 9) et la cycline T1, toutes deux nécessaires pour la phosphorylation du domaine carboxy-terminal (CTD) de l'ARN PII et le passage en phase d'élongation de la transcription (Estable, Naghavi et al. 2002).

La protéine **Dr1** (ou NC2-β, Negative co-factor 2) est quant à elle capable de réprimer l'initiation de la transcription en s'associant, sous forme d'hétérodimère avec DRAP1 (DR1-associated protein 1) pour former NC2, ou avec les TBPs pour empêcher la liaison des facteurs généraux de TFIIA et TFIIB (Masson, Leimgruber et al. 2008).

La protéine **VRK1** (Vaccinia-Related Kinase) est une sérine-thréonine kinase qui serait capable de phosphoryler l'histone H3 (Kang, Park et al. 2007), ainsi que certains

facteurs de transcription comme c-Jun (Sevilla, Santos et al. 2004), ATF2 (Sevilla, Santos et al. 2004) et p53. Dans le dernier cas, la phosphorylation d'une thréonine dans le domaine de liaison de p53 à MDM2, prévient la liaison de l'ubiquitine ligase et la dégradation de p53 (Low, Lopez et al. 2000). **PYHIN1**, une protéine inductible à l'interféron, est également décrite pour prévenir l'interaction de p53 avec la MDM2 en favorisant la dégradation de la ligase. Les protéines VRK1 et PYHIN1 sont donc susceptibles d'exercer un effet positif sur la transcription dépendante de p53.

La protéine **ORC2L** est une sous-unité du complexe ORC (origin recognition complex) permettant l'initiation de la réplication dans les cellules eucaryotes et pouvant même lier l'origine de réplication du virus Epstein-Barr (EBV). Cette protéine a également été décrite pour être impliquée dans la répression de l'expression génique (Takahara, Bong et al. 1996).

La protéine 3B contenant un domaine ARID (**ARID domain-containing protein 3B**) possède des régions d'homologie avec Rbp1, Rbp2 (liant la protéine de retinoblastome) et le facteur Bright spécifique des lymphocytes B. Cette protéine pourrait participer à la régulation de gènes tissus-spécifiques et la différenciation cellulaire (Numata, Claudio et al. 1999).

La protéine **COMMD8** appartient à une famille de protéines, homologues à la protéine MURR1, possédant un domaine conservé unique COMM (copper metabolism gene MURR1), permettant l'interaction entre protéines. Il semblerait que certaines de ces protéines puissent inhiber NF- κ B en agissant directement sur la capacité du facteur de transcription à interagir avec la chromatine (Burstein, Hoberg et al. 2005).

La protéine **Wiz** est peu décrite. La présence de cette protéine à doigt de zinc est essentielle pour la stabilisation d'un complexe entre les histones méthyltransférases G9a et GLP, dont l'association semble nécessaire, mais pas indispensable, pour l'activité de méthylation de la lysine 9 de l'histone 3 (H3K9). La présence de deux sites de liaison sur la protéine Wiz et dont l'association avec CtBP a pu être montrée, laisse penser que cette protéine à doigt de zinc pourrait faire le lien entre le complexe G9a/GLP et le co-répresseur CtBP (Ueda, Tachibana et al. 2006).

Quatre sous-unités appartenant au **complexe THO** ont également été mises en évidence pour la condition PMA + ionomycine. Ce complexe pourrait permettre de coupler la maturation des ARNm avec l'élongation de la transcription. Cependant, une sous-unité de ce complexe est également identifiée dans les extraits de cellules non traitées, ce qui laisse douter quant au rôle éventuel que THO pourrait jouer dans l'inductibilité de cette séquence suite à un traitement au PMA + ionomycine.

Parmi les protéines identifiées et connues pour participer de près ou de loin à la régulation de la transcription, quelques-unes d'entre elles jouent un rôle ou appartiennent à des complexes impliqués dans le remodelage de la chromatine. Par exemple, la nucléoline est un facteur capable de déstabiliser les octamères d'histones afin de favoriser l'activité des complexes de remodelage de la chromatine (Angelov, Bondarenko et al. 2006).

Le facteur **coREST** (REST corepressor 1) appartient au complexe BHC (BRAF-HDAC complex) décrit pour réprimer les gènes spécifiques des neurones dans les cellules non neuronales, suite au recrutement par REST (RE1-silencing transcription factor). Le BHC possède des activités enzymatiques de déacétylation et déméthylation des histones. Dans le second cas, il semblerait que le facteur coREST joue un rôle crucial pour l'activité de déméthylation des histones mais également des nucléosomes en favorisant l'interaction de BHC avec ces derniers (Lee, Wynder et al. 2005). De plus, la protéine responsable de cette activité enzymatique est l'histone deméthylase **AOF1**, également identifiée dans les extraits de la condition PMA + ionomycine (Lysine-specific histone demethylase 1).

Le facteur de transcription **Ikaros** joue un rôle essentiel dans la différenciation et la prolifération des lymphocytes. Il représente un candidat intéressant car des sites de liaison pour cette protéine ont pu être prédits *in silico* sur la séquence étudiée (Intra1). Ce facteur peut agir aussi bien en tant qu'activateur que répresseur de la transcription, selon le stade de différenciation des lymphocytes (Kim, Sif et al. 1999). En effet, lorsque les cellules sont immatures, Ikaros va recruter des complexes régulateurs négatifs comme SWI-SNF ou des HDAC, responsables de la répression de l'expression de CD4. Cependant, à un stade de différenciation plus avancé, Ikaros va pouvoir recruter de nouveaux éléments régulateurs tels que Mi-2 β , surpassant l'influence répressive des complexes initialement recrutés (SWI-SNF) et favorisant l'expression de CD4 (Naito, Gomez-Del Arco et al. 2007). Il est intéressant de constater que ce facteur **Mi-2 β** [connu également sous le nom de CHD4 (Chromodomain-helicase-DNA-binding protein 4)], fait également partie des protéines identifiées dans les extraits provenant de cellules incubées en présence de PMA + ionomycine. Cette protéine est connue pour appartenir à un complexe NuRD (Nucleosome remodeling and DNA methylation) capable de remodeler la chromatine de manière ATP-dépendante et de modifier les histones. Mi-2 β a été décrit aussi bien en tant qu'activateur de la transcription, suite à l'observation de son association avec des histones acétyltransférases et des histones méthyltransférases (Williams, Naito et al. 2004), que répresseur lorsqu'il est associé à des histones déacétylases (Zhang, LeRoy et al. 1998). Une troisième protéine appartenant au complexe NuRD, **RBBP4** (retinoblastoma binding protein 4), qui avait déjà été identifiée dans les résultats préliminaires, fait aussi partie des facteurs exclusivement identifiés dans les extraits PMA + ionomycine. Cette protéine est également connue pour appartenir à des complexes de déacétylation des histones (Nicolas, Morales et al. 2000).

Les protéines **RuvB-like 1 et 2**, **Arp8** (actin-related protein), **l'Arp Baf53a** (Arp4), **l'actine**, la protéine **NFR κ B** (nuclear factor related to kappaB binding protein) et la **sous-unité E** sont des sous-unités appartenant au complexe INO80. D'abord mis en évidence chez la levure, ce complexe également présent chez l'homme possède une activité de remodelage et de glissement des nucléosomes qui est ATP-dépendante. Il pourrait également être impliqué dans la régulation de la transcription ainsi que la réparation de l'ADN (Jin, Cai et al. 2005). L'étude du promoteur INO1 chez la levure a permis de montrer que ce complexe était nécessaire à mobilisation et à la réorganisation de la position des nucléosomes, réprimant l'expression du gène en conditions basales (Ford, Odeyale et al. 2007). Les protéines RuvB-like 1 et 2, partageant des régions d'homologies avec l'hélicase bactérienne RuvB, sont nécessaires à l'activité de remodelage de la chromatine du complexe INO80. Ces deux protéines sont des hélicases appartenant à la famille des ATPases de type AAA+ impliquées dans la dégradation, le dépliement et le désassemblage de complexes protéiques (Bao and Shen 2007). L'actine et les protéines associées (Arp) font également partie de ce complexe (Boyer and Peterson 2000). Elles ont été décrites pour participer à des processus de remodelage de la chromatine ATP-dépendant : le phénotype résultant de la mutation de Arp8 a permis de montrer l'importance de ce facteur dans l'activité de remodelage de ce complexe (Shen, Ranallo et al. 2003).

Le facteur de transcription **YY1**, identifié également dans les extraits PMA + ionomycine, est capable d'interagir et même de recruter le complexe co-activateur INO80 au niveau des gènes activés par YY1 (Cai, Jin et al. 2007). L'identification de cette protéine est importante car, en plus des divers sites putatifs qui ont pu être prédits *in silico*, la liaison de ce facteur a été mise en évidence par gel retard dans une partie de la région intragénique correspondant à l'intra1 (C. Van Lint, communication personnelle). Mise à part son interaction avec le complexe INO80, le facteur YY1 est connu pour pouvoir à la fois activer

ou réprimer la transcription, comme le suggère son appellation (Yin and Yang 1). Sa capacité à activer ou réprimer l'expression des gènes dépend des interactions que ce facteur de transcription établit d'autres protéines. Par exemple, en présence de la protéine virale E1A, YY1 agit en temps qu'activateur de la transcription alors qu'en son absence il devient répresseur (Gordon, Akopyan et al. 2006). Ce résultat est d'autant plus intéressant que lorsque YY1 interagit avec INO80, il exerce une activité positive sur la transcription du gène concerné (Cai, Jin et al. 2007). Rappelons enfin que YY1 était déjà connu pour réguler l'expression du HIV-1. En effet, un site de liaison est également présent en aval du site d'initiation de la transcription du LTR-5' du HIV-1, où il a été décrit pour réprimer l'expression du provirus en absence de la protéine trans-activatrice Tat (voir introduction point I.3.2) (Coull, Romerio et al. 2000).

Finalement, quelques protéines n'ayant pas de relation apparente avec les acides nucléiques ont été identifiées, comme des isoformes de tubuline ou des chaperones nucléaires et trois protéines (**C1orf77**, **SSNA1**, **ZC3H11A**) pour lesquelles aucune fonction n'est encore connue à l'heure actuelle.

Grâce aux améliorations techniques détaillées au point III.2.2, les résultats générés par la méthode de purification et d'identification des protéines interagissant avec une séquence oligonucléotidique sont encourageants. En effet, le nombre de protéines identifiées a largement augmenté par rapport aux résultats préliminaires et l'analyse différentielle entre les conditions testées permet de mettre en évidence des protéines présentant un certain intérêt dans la problématique abordée. C'est par exemple le cas du complexe INO80, étant donné qu'il est décrit pour activer l'expression des gènes en modifiant le motif nucléosomal réprimant la transcription en condition basale. Ce complexe est d'autant plus intéressant qu'il peut être recruté par le facteur YY1, également identifié dans les extraits PMA / ionomycine, et pour lequel une interaction physique avec la séquence intragénique a été montrée par gel retard. Par contre, plusieurs facteurs identifiés dans la condition PMA + ionomycine sont également associés à des activités de répression de l'expression génique alors que la région intragénique du HIV-1 étudiée présente une inductibilité sous ces mêmes conditions.

Cependant, il est avant tout indispensable de reproduire cette expérience, qui n'a pu être réalisée qu'une seule fois, afin de vérifier les résultats obtenus, d'autant plus que l'identification de certaines protéines ne repose que sur le séquençage d'un seul peptide. Néanmoins, le fait que certains complexes protéiques soient identifiés par l'intermédiaire de plusieurs de leurs sous-unités (7 sous-unités pour le complexe INO80), donne un certain poids à ces données.

Il serait intéressant de tester si davantage de protéines ne peuvent pas être identifiées en analysant chaque fraction séparément plutôt que de les regrouper, de manière à s'assurer que les peptides en faible quantité soient séquencés le plus souvent possible et non plus de manière aléatoire comme c'est le cas lorsqu'il y a saturation de détection. Ceci permettrait également d'assurer une meilleure reproductibilité de la méthode.

Le fait de retrouver le facteur YY1 parmi les protéines identifiées est rassurant puisque sa liaison avait également été montrée à cette région par gel shift. Même si ces résultats sont très encourageants, nous n'avons cependant pas réussi à identifier les facteurs de transcription Oct-1 ou Sp1 dont la liaison à la séquence intragénique a été démontrée par gel retard (Goffin, Demonte et al. 2005). Nous n'avons également pas pu montrer la liaison de AP-1, mise en évidence sur la séquence intral par ELISA-like (voir point III.1.3), à partir d'extraits nucléaires de cellules Jurkat stimulées en présence de PMA + ionomycine. Cette absence

d'AP-1 peut être expliquée par l'utilisation dans le tampon de liaison de PolyIdC, une petite molécule jouant le rôle d'ADN compétiteur, classiquement utilisée dans la technique de retard sur gel. Les concentrations utilisées (0,1 mg/ml) sont connues pour faire chuter la liaison de ce facteur de transcription particulier à l'ADN (Park, Ko et al. 2005). Nous avons également observé cet effet néfaste de trop haute concentration de ce compétiteur sur la liaison de AP-1 lors de nos analyses en spectrométrie de masse. En effet, la liaison de c-Fos recombinant ajouté à des extraits nucléaires de cellules Jurkat a été testée pour trois tampons différents : le tampon habituellement utilisé, ce même tampon sans polyIdC et le tampon de liaison spécifique à AP-1 employé pour les ELISA-like (voir table 2.1). c-Fos a pu être identifié dans le cas des deux derniers tampons ne contenant pas ou peu (0,05 mg/ml) de polyIdC, mais pas dans les conditions contenant (0.1 mg/ml) de polyIdC.

Ce résultat met en évidence l'importance du tampon de liaison entre les protéines et l'oligonucléotide et de la concentration en ADN compétiteur utilisée. En effet, malgré des concentrations en polyIdC affectant déjà la liaison spécifique de certains facteurs de transcription comme AP-1, des protéines comme les hnRNPs sont toujours capturées en abondance. Bien que l'analyse différentielle des conditions contrôles et stimulées permettent d'en exclure une majorité, le système idéal serait de travailler dans des conditions où elles ne puissent pas se lier à la sonde oligonucléotidique d'intérêt. En effet, cela éviterait la saturation de sites de liaison et d'éventuels encombrements stériques au niveau de la sonde nucléique, pouvant entraver la liaison spécifique de facteurs d'intérêt. De plus, l'identification de ces facteurs lors de l'analyse par spectrométrie de masse serait dès lors favorisée en absence des nombreux peptides provenant de protéines comme les hnRNPs.

IV. Conclusion et perspectives

A l'heure actuelle, l'analyse bioinformatique d'une séquence est la seule solution disponible pour le chercheur qui souhaite entreprendre l'étude d'une séquence nucléique encore non caractérisée, en recherchant les facteurs de transcription susceptibles de se lier à cette séquence. Toutefois, les résultats obtenus par ce type d'approche peuvent demander beaucoup de temps puisqu'ils nécessitent de vérifier la pertinence des nombreux candidats générés, et ces analyses génèrent de nombreux faux positifs et faux négatifs.

Dans le but de développer une stratégie permettant d'initier une telle recherche, une méthode sans *a priori* d'identification par spectrométrie de masse des protéines interagissant physiquement avec une séquence oligonucléotidique est mise au point au laboratoire. Lors de ce mémoire, alors que le principe général de la méthode avait déjà fait l'objet de nombreuses mises au point, nous avons eu l'occasion de confronter cette technique à une séquence d'intérêt encore peu décrite du HIV-1, qui pourrait permettre de mieux comprendre la phase précoce de transcription du provirus. En effet, lors du cycle répliatif du virus, le génome viral va s'intégrer dans le génome de l'hôte, adoptant un statut équivalent aux autres gènes cellulaires, afin d'être transcrit et répliqué. Cependant, en l'absence de la protéine transactivatrice virale Tat, la transcription du HIV-1, initiée au niveau du LTR-5', ne peut passer en phase d'élongation suite à l'action répressive d'un nucléosome nucl1 situé en aval du site d'initiation, du facteur YY1 et N-TEF. Etant donné que le virus n'importe pas de protéine Tat dans sa capsid, la question de l'origine des premières molécules Tat dans la cellule infectée se pose. Le modèle accepté actuellement postule que parmi les courts transcrits générés, de rares transcrits pourraient être complets et permettre la synthèse des premières protéines Tat. Entre-temps, des recherches initiées par l'équipe du professeur A. Burny ont identifié un promoteur intragénique au sein du gène *pol*. Il a dès lors été envisagé que ce promoteur puisse permettre la synthèse des premières protéines Tat, comme cela a pu être décrit pour le HTLV-1 et la protéine trans-activatrice Tax, dont la fonction est relativement similaire à Tat. Les recherches actuelles ont permis de mettre en évidence une inductibilité de cette séquence intragénique au TNF α , ainsi qu'au PMA + ionomycine, cependant, aucun des facteurs identifiés à ce jour pour s'y lier ne peut expliquer cette propriété. Lors de ce mémoire, nous avons eu recours à deux stratégies complémentaires afin d'identifier des candidats qui pourraient expliquer l'inductibilité de la région intragénique du HIV-1 et de mieux comprendre la régulation de cette région d'intérêt. Dans un premier temps, nous avons étudié de manière ciblée trois facteurs de transcription potentiels, connus pour être induits par le TNF α , ainsi que par le PMA + ionomycine, alors que la seconde approche est basée sur une identification sans *a priori* de protéines liant un fragment de la région intragénique par la technique actuellement développée au laboratoire.

IV.1. Etude de la liaison de NF- κ B, AP-1 et NFAT par ELISA-like

Lors de cette première partie, nous nous sommes intéressés à trois facteurs de transcription en particulier (NF- κ B, AP-1 et NFAT) car ils sont bien connus pour être inducibles par le TNF α mais aussi par le PMA + ionomycine. Ils sont donc susceptibles d'expliquer l'inductibilité de la région intragénique à ces deux traitements.

Lors de l'approche avec *a priori*, une démarche similaire a été appliquée pour les trois facteurs de transcription étudiés. Dans un premier temps, la liaison a d'abord été testée sur

une séquence d'environ 300 paires de bases de la région intragénique du gène *pol* du HIV-1 (intra1), pour ensuite essayer de localiser plus précisément les sites potentiels de liaison des facteurs de transcription. Les premiers résultats ont montré une liaison de chacun des trois candidats à la séquence intra1, qui semble cependant d'affinité beaucoup faible que celle observée pour des sites consensus. Par la suite, nous avons donc recherché plus précisément le(s) site(s) reconnu(s) par chacune des trois protéines grâce à l'utilisation de séquences plus petites et couvrant l'ensemble de la séquence initialement utilisée. Les résultats pour NF- κ B ont rapidement suggéré une liaison non spécifique à la séquence intra1 comme le révèlent les expériences de compétition réalisées. En contre partie, ceux pour NFAT et AP-1 sont en accord avec l'analyse bioinformatique réalisée et indiquent deux sites putatifs pour AP-1 (plus précisément le dimère c-Fos-JunD) et deux également pour NFAT aux nucléotides 4668, 4765 et 4674, 4785, respectivement.

Suite à l'homologie de séquence entre le promoteur du GM-CSF et la région intragénique pour les sites AP-1 et NFAT aux nucléotides 4668 et 4674, respectivement, nous avons envisagé une éventuelle coopération physique qu'il pourrait y avoir entre ces deux facteurs de transcription. Ces sites composites permettent une régulation fine des gènes qu'ils contrôlent, en intégrant simultanément les stimuli propres aux voies d'activation des deux facteurs. Etant donné qu'une liaison stable des facteurs de transcription à ces sites composites nécessite la présence des deux partenaires, nous avons testé la coopération de AP-1 et NFAT grâce à l'utilisation de sondes mutées individuellement ou simultanément pour les sites présents aux nucléotides 4668 et 4765. Les premiers résultats obtenus pour la liaison de AP-1 à ces sondes semblent en faveur de l'existence de ces sites composites. En effet, la liaison de c-Fos chute pour chacune des sondes mutées, nous informant sur la spécificité de la liaison de c-Fos et sa coopération avec NFAT. Alors que des expériences de compétition mettaient en doute la spécificité de la liaison de c-Fos à la séquence intra1, la chute du signal de liaison observée lorsque le site putatif AP-1 est muté sur la séquence intra3 est rassurante pour ce fragment. Une démarche similaire devrait également être appliquée pour le site AP-1 présent au nucléotide 4765 afin de s'assurer de la spécificité de la liaison observée pour c-Fos sur la sonde intra4. Ensuite, la chute de la liaison est également observable lorsque le site putatif de NFAT est muté, en accord avec le principe des sites composites selon lequel la liaison à l'ADN est instable en absence d'un des deux partenaires. Enfin, les résultats obtenus pour JunD sont similaires à ceux de c-Fos. Toutefois, la forte variabilité observée pour cette expérience nécessite qu'elle soit avant tout reproduite.

Les résultats sont plus difficilement interprétables pour NFAT, étant donné que la liaison pour la sonde doublement mutée chute à peine. Toutefois, le membre étudié (NFATc1) n'est pas celui qui a été le plus souvent décrit pour une coopération avec AP-1. Il serait donc intéressant de pouvoir tester la liaison de NFATc2 au niveau du site composite. Ceci permettrait également de répondre à la question de spécificité de la liaison de NFAT, comme cela a été fait pour c-Fos. Si l'étude de la liaison de NFATc2 est en accord avec les résultats obtenus pour AP-1, il serait alors intéressant d'approfondir les recherches en terme de fonctionnalité. Des expériences de gène rapporteur avec la séquence sauvage et mutée individuellement ou simultanément pour les deux sites pourraient être réalisées afin d'évaluer la capacité trans-activatrice du complexe ternaire. Par après, le test fonctionnel le plus intéressant serait d'étudier l'infectivité et l'expression de protéines virales, dont Tat, pour des lignées cellulaires T et monocytaires contenant des virus intégrés et mutés pour le site composite putatif. Bien que l'idée de coopération soit renforcée par le fait que les résultats obtenus pour les extraits nucléaires de cellules U937 soient similaires à ceux observés pour AP-1 lorsque le site putatif de NFAT est muté, ils ne coïncident pas avec le fait que la région intragénique soit également inductible dans les lignées macrophagiques. En effet, étant donné que NFAT n'est pas présent dans les U937, l'hypothétique coopération ne peut avoir lieu et

pourtant cette lignée cellulaire est tout de même inductible. Il pourrait y avoir une régulation différentielle en fonction du type cellulaire, expliquant la différence d'inductibilité existant entre les lignées T et monocytaires/macrophagiques, permettant une adaptation de la régulation de l'expression du provirus en fonction de la cellule hôte. Dans cette optique, il serait également intéressant de tester la liaison à la séquence intra1 des autres membres de la famille AP-1 pour des extraits nucléaires de cellules U937, étant donné que les membres présents et activés peuvent différer des cellules Jurkat.

Il est important de remarquer que la liaison de AP-1 et NFAT qui a pu être mise en évidence sur la séquence intragénique est de faible affinité en comparaison avec celle observable sur un site consensus. Cependant, cela peut provenir du fait que les sites présents au niveau de la séquence intragénique sont fortement dégénérés. Néanmoins, dans le cadre de la problématique étudiée, une forte affinité n'est pas nécessairement requise. En effet, si une transcription au niveau du gène *pol* a bien lieu, il n'est pas nécessaire qu'elle soit intense car une fois les premières protéines Tat synthétisées, la transcription pourra être initiée de manière beaucoup plus importante à partir du LTR-5'.

IV.2. Recherche sans a priori de facteurs interagissant avec la séquence intra1

Toujours dans l'optique d'identifier des facteurs de transcription interagissant avec la séquence intragénique du HIV-1 et pouvant expliquer son inductibilité au PMA + ionomycine, une première analyse par spectrométrie de masse des protéines capturées par la région d'intérêt a soulevé la nécessité de trois mises au point.

Premièrement, la quantité de trypsine a été augmentée, ce qui a permis de favoriser la digestion des protéines purifiées et d'augmenter le nombre de peptides différents identifiés par protéine.

Ensuite, la présence de polymères de biotine très abondants et entravant l'identification des peptides en faible quantité a été éliminée en incubant les peptides digérés en présence de billes magnétiques tapissées de streptavidine.

Enfin, afin de résoudre le problème de complexité qui représente le facteur limitant majeur de la capacité d'identification par spectrométrie de masse, les peptides sont séparés selon leurs charges et leur hydrophobicité avant d'être analysés. La première étape de séparation est réalisée par chromatographie échangeuse de cations (HPLC-SCX), générant une quarantaine de fractions beaucoup moins complexes et qui vont être séparées par une deuxième chromatographie, en phase inverse cette fois (nano LC) en amont du spectromètre de masse.

Suite à ces améliorations, une seconde analyse des protéines purifiées grâce au fragment de la séquence intragénique du HIV-1 (« intra 1 ») a été réalisée à partir de 2 mg d'extraits nucléaires de cellules Jurkat non stimulées ou stimulées 5 heures par le PMA + ionomycine. La différence majeure des résultats obtenus par rapport à l'analyse préliminaire est le nombre de protéines identifiées, passant d'une vingtaine à plus d'une centaine. Alors que la première analyse a généré une majorité de protéines liant préférentiellement l'ARN, l'analyse différentielle des protéines identifiées entre la condition contrôle et stimulée lors de ce second essai permet de mettre en avant divers candidats intéressants.

Parmi ces derniers, nous mettons majoritairement l'accent sur deux d'entre eux : Ikaros et YY1. Tout d'abord parce que des sites putatifs ont pu être prédits sur la séquence étudiée pour ces deux facteurs de transcription. Ensuite parce qu'ils peuvent tous les deux

recruter des complexes de remodelage de la chromatine afin de favoriser l'expression génique. Bien que chacun d'eux soit décrit aussi bien pour favoriser que pour réprimer la transcription, les protéines également identifiées lors de notre analyse sont connues pour s'associer avec YY1 et Ikaros lors de processus d'activation de l'expression génique. En effet, le recrutement du facteur de remodelage de la chromatine Mi-2 β par Ikaros permet de surpasser la répression due aux complexes présents à l'état basal sur le promoteur du gène *CD4* (Naito, Gomez-Del Arco et al. 2007). De la même manière, YY1 recrute le complexe INO80 au niveau de promoteurs où il active la transcription (Cai, Jin et al. 2007). De plus, le fait que 6 protéines différentes du complexe INO80 aient pu être identifiées dans les extraits stimulés conforte ce résultat, même si cette expérience n'a encore été réalisée qu'une seule fois. Etant donné que la région d'hypersensibilité n'a pu être montrée que dans des lignées monocytaires, une régulation différentielle de la région intragénique pourrait avoir lieu en fonction du type cellulaire. En effet, puisque cette région n'est pas dépourvue de nucléosomes dans les lignées T, des complexes de remodelage comme INO80 sont peut être nécessaires afin de permettre l'accès de régulateurs transcriptionnelles à l'ADN.

En effet, l'ensemble de ces résultats doivent être reproduits afin de s'assurer que l'on retrouve bien les protéines identifiées. Si c'est le cas, il faudrait dans un premier temps étudier la liaison de ces protéines à la séquence intragénique par des méthodes classiques, pour ensuite s'intéresser à une approche plus fonctionnelle, débouchant finalement sur des expériences *in vivo* dans des lignées cellulaires contenant le provirus.

Des expériences complémentaires à celles réalisées sur les cellules Jurkat pourraient également être réalisées. Tout d'abord, il faudrait réaliser l'analyse sans *a priori* pour des cellules U937 non stimulées ou stimulées 1 heure en présence de TNF α , afin de voir non seulement l'éventuelle présence de facteurs de transcription communs aux deux lignées et pouvant expliquer l'inductibilité au TNF α et au PMA + ionomycine, mais également afin de mettre en évidence certains régulateurs spécifiques aux différentes lignées cellulaires.

Enfin, des expériences d'extension d'amorces à partir de fragments de la région intragénique sauvage ou mutée ont permis de montrer que le site D (situé environ au milieu de la séquence intra4), liant un facteur encore non identifié, est capital pour l'inductibilité de cette séquence (C. Van Lint, communication personnelle). Il serait donc intéressant de réaliser une sonde plus courte que l'intra1 (qui est de 300 bp), afin de limiter les protéines capturées non spécifiquement mais également pour prévenir les encombrements stériques, et contenant le site D pour essayer d'identifier le ou les facteur(s) interagissant avec ce dernier.

En ce qui concerne la technique en elle-même, l'augmentation de la quantité de trypsine, l'élimination de la biotine et le fractionnement des échantillons ont permis d'augmenter le nombre de protéines identifiées et le nombre de peptides différents séquencés pour chacune d'entre elles. Le fait d'identifier diverses sous-unités appartenant à un même complexe montre que la technique permet de capturer les protéines interagissant ensemble, ce qui laisse penser qu'il est dès lors possible d'identifier les protéines interagissant directement et indirectement avec la séquence étudiée.

Il est encore nécessaire de s'assurer de la reproductibilité de ces résultats mais la diminution de la complexité des échantillons à analyser semble une étape indispensable. Les facteurs Oct-1 et Sp1 n'ayant pas pu être identifiés lors nos analyses, alors que la liaison de ce facteur a pu être démontrée par gel retard sur la séquence étudiée, suggère la nécessité de mises au point supplémentaires. Pour ce faire, il est possible d'agir principalement à deux niveaux : la capture des facteurs sur la sonde oligonucléotidique et la capacité d'identification par spectrométrie de masse des protéines purifiées. La seconde possibilité semble maintenant

déjà bien améliorée, notamment grâce à la digestion optimisée des protéines, le retrait de la biotine et surtout le fractionnement de l'échantillon par HPLC-SCX. En effet, ces mises au point ont permis de passer de la vingtaine de protéines habituellement identifiées à une centaine. Il semblerait donc que le fractionnement des échantillons permette de limiter l'impact, du moins dans une certaine mesure, de la présence des protéines « contaminantes » sur la détection des protéines recherchées.

Par contre, des progrès doivent encore être réalisés au niveau de l'étape de purification des protéines par la sonde oligonucléotidique, d'une part pour éviter d'identifier des protéines qui se lient aux acides (ribo)nucléiques en général mais pas spécifiquement à la séquence d'intérêt (protéines type hNRPs), et d'autre part pour augmenter les chances de capturer les facteurs de transcription d'intérêt attirés spécifiquement par cette séquence.

En effet, les protéines liant préférentiellement l'ARN sont toujours présentes, le fractionnement ne permettant d'éliminer que le problème de leur abondance pour l'analyse en spectrométrie de masse. Des encombrements stériques peuvent toujours avoir lieu lors de la purification sur billes magnétiques, empêchant la liaison de facteurs de transcription ou de protéines qui leur sont associés. L'idéal serait donc d'arriver à ne capturer que les protéines liant spécifiquement la séquence étudiée ou, de manière intermédiaire, faire chuter au maximum la liaison des protéines non désirables. Cependant, la démarche sans *a priori* limite considérablement les solutions envisageables et à l'heure actuelle les ADN compétiteurs restent la meilleure solution malgré leur limite. En effet, comme cela a pu être décrit dans la littérature (Park, Ko et al. 2005) et lors de nos expériences avec la protéine c-Fos recombinante, la concentration en compétiteur utilisée est un paramètre pouvant affecter la liaison spécifique de ce facteur de transcription à l'ADN. Leur utilisation n'en reste pas moins indispensable et il est nécessaire de trouver le meilleur compromis entre l'élimination des protéines liant l'ADN et la liaison spécifique des facteurs de transcription.

L'importance des contaminants peut également dépendre de la structure de la séquence elle-même. En effet, en fonction de sa taille et des éléments *cis* présents, une séquence capture davantage de protéines qu'une autre. Deux situations peuvent alors être envisagées. Dans la première, le nombre de protéines capturées par l'oligonucléotide n'est pas trop important et n'affecte pas la purification de ces dernières. A l'opposé, une séquence riche en sites de liaison reconnus par de nombreux facteurs de transcription, en plus des protéines liant l'ARN, pourraient provoquer la perte d'information suite à des problèmes d'encombrement. Dans ce cas, il serait peut-être intéressant d'incuber les extraits nucléaires en présence de faibles concentrations d'ARN, afin de lier une partie des protéines ayant une affinité préférentielle pour les acides ribonucléiques, lors d'une étape précédant la purification proprement dite. Il est évident que le risque encouru par l'ajout d'une telle étape est la perte de protéines d'intérêts qui pourraient également se lier à cette séquence.

Le second paramètre majeur nécessitant une future mise au point est le tampon de liaison qui sera utilisé lors de l'incubation des extraits nucléaires avec les sondes nucléiques. En effet, les conditions de liaison *in vitro* des facteurs de transcription sur un oligonucléotide peuvent être fort variables. Nous avons d'ailleurs utilisé 3 tampons différents pour étudier la capacité de liaison de NF- κ B, AP-1 et NFAT sur la séquence intra1. Il est donc essentiel de définir un tampon de liaison permettant de couvrir la plus large gamme possible de facteurs de transcription. Pour ce faire, il sera nécessaire de réaliser un criblage des différents tampons actuellement utilisés et d'identifier selon leur composition les protéines capables de lier l'ADN. Il serait d'ailleurs intéressant de voir s'il existe une corrélation entre le type de domaine de liaison à l'ADN présent sur la protéine et la composition des tampons de liaison ad hoc. On pourrait éventuellement envisager de ne pas utiliser un seul mais plusieurs

tampons de liaison pour des purifications réalisées en parallèles afin de couvrir les différentes familles de facteurs de transcription et de cibler un maximum de candidats potentiels. Malgré tout, il est bon de remarquer que le développement d'un tampon « universel » renforcera également la purification de protéines liant non spécifiquement la sonde nucléique, pointant une fois encore l'importance de trouver un moyen de les éliminer.

Enfin, la purification est réalisée à partir d'ADN nu ce qui fait que certaines protéines peuvent se lier aux acides nucléiques alors qu'elles n'y auraient pas accès *in vivo*. A plus long terme, lorsque tous les aspects de la technique seront optimisés, il serait intéressant de tester la possibilité d'utiliser des sondes oligonucléotidiques immobilisées sous forme de chromatine reconstituée, même si elles ne mimeront toujours pas parfaitement les conditions cellulaires, étant donné que des phénomènes dynamiques comme les modifications post-traductionnelles des histones représentent un élément de contrôle majeur. Pour conclure, il est d'ailleurs important de rappeler que le but premier de cette technique est d'identifier des candidats interagissant potentiellement avec une séquence oligonucléotidique et qui pourront, par la suite, être validés par le biais de techniques plus classiques, étudiant leur liaison *in vivo* (ChIP) mais également leur fonctionnalité (gènes rapporteurs).

- Angelov, D., V. A. Bondarenko, et al. (2006). "Nucleolin is a histone chaperone with FACT-like activity and assists remodeling of nucleosomes." *EMBO J* **25**(8): 1669-79.
- Ariumi, Y., K. Shimotohno, et al. (1998). "Characterization of the internal promoter of human T-cell leukemia virus type I." *FEBS Lett* **423**(1): 25-30.
- Atreya, I., R. Atreya, et al. (2008). "NF-kappaB in inflammatory bowel disease." *J Intern Med* **263**(6): 591-6.
- Bachler, M., R. Schroeder, et al. (1999). "StreptoTag: a novel method for the isolation of RNA-binding proteins." *RNA* **5**(11): 1509-16.
- Bao, Y. and X. Shen (2007). "INO80 subfamily of chromatin remodeling complexes." *Mutat Res* **618**(1-2): 18-29.
- Bates, D. L., K. K. Barthel, et al. (2008). "Crystal structure of NFAT bound to the HIV-1 LTR tandem kappaB enhancer element." *Structure* **16**(5): 684-94.
- Berger, E. A., P. M. Murphy, et al. (1999). "Chemokine receptors as HIV-1 coreceptors: roles in viral entry, tropism, and disease." *Annu Rev Immunol* **17**: 657-700.
- Bonizzi, G. and M. Karin (2004). "The two NF-kappaB activation pathways and their role in innate and adaptive immunity." *Trends Immunol* **25**(6): 280-8.
- Boyer, L. A. and C. L. Peterson (2000). "Actin-related proteins (Arps): conformational switches for chromatin-remodeling machines?" *Bioessays* **22**(7): 666-72.
- Bradford, M. M. (1976). "A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding." *Anal Biochem* **72**: 248-54.
- Cai, Y., J. Jin, et al. (2007). "YY1 functions with INO80 to activate transcription." *Nat Struct Mol Biol* **14**(9): 872-4.
- Chen, L. F. and W. C. Greene (2004). "Shaping the nuclear action of NF-kappaB." *Nat Rev Mol Cell Biol* **5**(5): 392-401.
- Chun, R. F., O. J. Semmes, et al. (1998). "Modulation of Sp1 phosphorylation by human immunodeficiency virus type 1 Tat." *J Virol* **72**(4): 2615-29.
- Cullen, B. R. (1987). "Use of eukaryotic expression technology in the functional analysis of cloned genes." *Methods Enzymol* **152**: 684-704.
- Deng, W. and S. G. Roberts (2007). "TFIIB and the regulation of transcription by RNA polymerase II." *Chromosoma* **116**(5): 417-29.
- Dignam, J. D., R. M. Lebovitz, et al. (1983). "Accurate transcription initiation by RNA polymerase II in a soluble extract from isolated mammalian nuclei." *Nucleic Acids Res* **11**(5): 1475-89.
- Estable, M. C., M. H. Naghavi, et al. (2002). "MCEF, the newest member of the AF4 family of transcription factors involved in leukemia, is a positive transcription elongation factor-b-associated protein." *J Biomed Sci* **9**(3): 234-45.
- Ford, J., O. Odeyale, et al. (2007). "A SWI/SNF- and INO80-dependent nucleosome movement at the INO1 promoter." *Biochem Biophys Res Commun* **361**(4): 974-9.
- Freed, E. O. (2001). "HIV-1 replication." *Somat Cell Mol Genet* **26**(1-6): 13-33.
- Fried, M. G. (1989). "Measurement of protein-DNA interaction parameters by electrophoresis mobility shift assay." *Electrophoresis* **10**(5-6): 366-76.
- Gabrielsen, O. S., E. Hornes, et al. (1989). "Magnetic DNA affinity purification of yeast transcription factor tau--a new purification principle for the ultrarapid isolation of near homogeneous factor." *Nucleic Acids Res* **17**(15): 6253-67.
- Galas, D. J. and A. Schmitz (1978). "DNase footprinting: a simple method for the detection of protein-DNA binding specificity." *Nucleic Acids Res* **5**(9): 3157-70.

- Ganchi, P. A., S. C. Sun, et al. (1993). "A novel NF-kappa B complex containing p65 homodimers: implications for transcriptional control at the level of subunit dimerization." *Mol Cell Biol* **13**(12): 7826-35.
- Garvie, C. W., J. R. Stagno, et al. (2007). "Characterization of the RFX complex and the RFX5(L66A) mutant: implications for the regulation of MHC class II gene expression." *Biochemistry* **46**(6): 1597-611.
- Ghosh, S. and M. Karin (2002). "Missing pieces in the NF-kappaB puzzle." *Cell* **109** Suppl: S81-96.
- Goffin, V., D. Demonte, et al. (2005). "Transcription factor binding sites in the pol gene intragenic regulatory region of HIV-1 are important for virus infectivity." *Nucleic Acids Res* **33**(13): 4285-310.
- Gordon, S., G. Akopyan, et al. (2006). "Transcription factor YY1: structure, function, and therapeutic implications in cancer biology." *Oncogene* **25**(8): 1125-42.
- Griffin, G. E., K. Leung, et al. (1989). "Activation of HIV gene expression during monocyte differentiation by induction of NF-kappa B." *Nature* **339**(6219): 70-3.
- Gygi, S. P., B. Rist, et al. (1999). "Quantitative analysis of complex protein mixtures using isotope-coded affinity tags." *Nat Biotechnol* **17**(10): 994-9.
- Hakre, S., M. I. Tussie-Luna, et al. (2006). "Opposing functions of TFII-I spliced isoforms in growth factor-induced gene expression." *Mol Cell* **24**(2): 301-8.
- Heintzman, N. D. and B. Ren (2007). "The gateway to transcription: identifying, characterizing and understanding promoters in the eukaryotic genome." *Cell Mol Life Sci* **64**(4): 386-400.
- Hogan, P. G., L. Chen, et al. (2003). "Transcriptional regulation by calcium, calcineurin, and NFAT." *Genes Dev* **17**(18): 2205-32.
- Hou, P., Z. Chen, et al. (2007). "Real-time PCR assay for ultrasensitive quantification of DNA-binding proteins." *Clin Chem* **53**(4): 581-6.
- Ishitani, K., T. Yoshida, et al. (2003). "p54nrb acts as a transcriptional coactivator for activation function 1 of the human androgen receptor." *Biochem Biophys Res Commun* **306**(3): 660-5.
- Johnstone, R. W., J. A. Kerry, et al. (1998). "The human interferon-inducible protein, IFI 16, is a repressor of transcription." *J Biol Chem* **273**(27): 17172-7.
- Juven-Gershon, T., J. Y. Hsu, et al. (2006). "Perspectives on the RNA polymerase II core promoter." *Biochem Soc Trans* **34**(Pt 6): 1047-50.
- Kadonaga, J. T. (2004). "Regulation of RNA polymerase II transcription by sequence-specific DNA binding factors." *Cell* **116**(2): 247-57.
- Karin, M. (1996). "The regulation of AP-1 activity by mitogen-activated protein kinases." *Philos Trans R Soc Lond B Biol Sci* **351**(1336): 127-34.
- Karin, M. (2006). "Nuclear factor-kappaB in cancer development and progression." *Nature* **441**(7092): 431-6.
- Karn, J. (1999). "Tackling Tat." *J Mol Biol* **293**(2): 235-54.
- Kwak, J. C., P. P. Ongusaha, et al. (2003). "IFI16 as a negative regulator in the regulation of p53 and p21(Waf1)." *J Biol Chem* **278**(42): 40899-904.
- Laniel, M. A., A. Beliveau, et al. (2001). "Electrophoretic mobility shift assays for the analysis of DNA-protein interactions." *Methods Mol Biol* **148**: 13-30.
- Lee, M. G., C. Wynder, et al. (2005). "An essential role for CoREST in nucleosomal histone 3 lysine 4 demethylation." *Nature* **437**(7057): 432-5.
- Lewis, B. A. and D. Reinberg (2003). "The mediator coactivator complex: functional and physical roles in transcriptional regulation." *J Cell Sci* **116**(Pt 18): 3667-75.
- Low, J. N., M. D. Lopez, et al. (2000). "N-(6-Amino-3,4-dihydro-3-methyl-5-nitroso-4-oxopyrimidin-2-yl) derivatives of glycine, valine, serine, threonine and methionine:

- interplay of molecular, molecular-electronic and supramolecular structures." *Acta Crystallogr B* **56 (Pt 5)**: 882-92.
- Luciano, R. L. and A. C. Wilson (2003). "HCF-1 functions as a coactivator for the zinc finger protein Krox20." *J Biol Chem* **278(51)**: 51116-24.
- Luo, C., E. Burgeon, et al. (1996). "Recombinant NFAT1 (NFATp) is regulated by calcineurin in T cells and mediates transcription of several cytokine genes." *Mol Cell Biol* **16(7)**: 3955-66.
- Macian, F., C. Lopez-Rodriguez, et al. (2001). "Partners in transcription: NFAT and AP-1." *Oncogene* **20(19)**: 2476-89.
- Marciniak, R. A. and P. A. Sharp (1991). "HIV-1 Tat protein promotes formation of more-processive elongation complexes." *EMBO J* **10(13)**: 4189-96.
- Masson, P., E. Leimgruber, et al. (2008). "The dual control of TFIIB recruitment by NC2 is gene specific." *Nucleic Acids Res* **36(2)**: 539-49.
- Matys, V., O. V. Kel-Margoulis, et al. (2006). "TRANSFAC and its module TRANSCompel: transcriptional gene regulation in eukaryotes." *Nucleic Acids Res* **34(Database issue)**: D108-10.
- McKenna, N. J. and B. W. O'Malley (2002). "Combinatorial control of gene expression by nuclear receptors and coregulators." *Cell* **108(4)**: 465-74.
- Naito, T., P. Gomez-Del Arco, et al. (2007). "Antagonistic interactions between Ikaros and the chromatin remodeler Mi-2beta determine silencer activity and Cd4 gene expression." *Immunity* **27(5)**: 723-34.
- Negulescu, P. A., N. Shastri, et al. (1994). "Intracellular calcium dependence of gene expression in single T lymphocytes." *Proc Natl Acad Sci U S A* **91(7)**: 2873-7.
- Nicolas, E., V. Morales, et al. (2000). "RbAp48 belongs to the histone deacetylase complex that associates with the retinoblastoma protein." *J Biol Chem* **275(13)**: 9797-804.
- Nikolov, D. B. and S. K. Burley (1997). "RNA polymerase II transcription initiation: a structural view." *Proc Natl Acad Sci U S A* **94(1)**: 15-22.
- Numata, S., P. P. Claudio, et al. (1999). "Bdp, a new member of a family of DNA-binding proteins, associates with the retinoblastoma gene product." *Cancer Res* **59(15)**: 3741-7.
- Orphanides, G. and D. Reinberg (2002). "A unified theory of gene expression." *Cell* **108(4)**: 439-51.
- Park, S. S., B. J. Ko, et al. (2005). "Mass spectrometric screening of transcriptional regulators using DNA affinity capture assay." *Anal Biochem* **344(1)**: 152-4.
- Parnaik, V. K. (2008). "Role of nuclear lamins in nuclear organization, cellular signaling, and inherited diseases." *Int Rev Cell Mol Biol* **266**: 157-206.
- Perkins, N. D. (2007). "Integrating cell-signalling pathways with NF-kappaB and IKK function." *Nat Rev Mol Cell Biol* **8(1)**: 49-62.
- Pilkinton, M., R. Sandoval, et al. (2007). "Mammalian Mip/LIN-9 interacts with either the p107, p130/E2F4 repressor complex or B-Myb in a cell cycle-phase-dependent context distinct from the Drosophila dREAM complex." *Oncogene* **26(54)**: 7535-43.
- Rao, A., C. Luo, et al. (1997). "Transcription factors of the NFAT family: regulation and function." *Annu Rev Immunol* **15**: 707-47.
- Renard, P., I. Ernest, et al. (2001). "Development of a sensitive multi-well colorimetric assay for active NFkappaB." *Nucleic Acids Res* **29(4)**: E21.
- Rieger, R. A., E. I. Zaika, et al. (2006). "Proteomic approach to identification of proteins reactive for abasic sites in DNA." *Mol Cell Proteomics* **5(5)**: 858-67.
- Roeder, R. G. and W. J. Rutter (1970). "Specific nucleolar and nucleoplasmic RNA polymerases." *Proc Natl Acad Sci U S A* **65(3)**: 675-82.

- Rosenau, C., D. Emery, et al. (2004). "Development of a high-throughput plate-based chemiluminescent transcription factor assay." *J Biomol Screen* **9**(4): 334-42.
- Rosonina, E., J. Y. Ip, et al. (2005). "Role for PSF in mediating transcriptional activator-dependent stimulation of pre-mRNA processing in vivo." *Mol Cell Biol* **25**(15): 6734-46.
- Roy, A. L. (2007). "Signal-induced functions of the transcription factor TFII-I." *Biochim Biophys Acta* **1769**(11-12): 613-21.
- Rusconi, F., F. Guillonéau, et al. (2002). "Contributions of mass spectrometry in the study of nucleic acid-binding proteins and of nucleic acid-protein interactions." *Mass Spectrom Rev* **21**(5): 305-48.
- Shaulian, E. and M. Karin (2001). "AP-1 in cell proliferation and survival." *Oncogene* **20**(19): 2390-400.
- Shav-Tal, Y. and D. Zipori (2002). "PSF and p54(nrb)/NonO--multi-functional nuclear proteins." *FEBS Lett* **531**(2): 109-14.
- Shen, X., R. Ranallo, et al. (2003). "Involvement of actin-related proteins in ATP-dependent chromatin remodeling." *Mol Cell* **12**(1): 147-55.
- Shimkus, M., J. Levy, et al. (1985). "A chemically cleavable biotinylated nucleotide: usefulness in the recovery of protein-DNA complexes from avidin affinity columns." *Proc Natl Acad Sci U S A* **82**(9): 2593-7.
- Singh, H., R. Sen, et al. (1986). "A nuclear factor that binds to a conserved sequence motif in transcriptional control elements of immunoglobulin genes." *Nature* **319**(6049): 154-8.
- Smale, S. T. and J. T. Kadonaga (2003). "The RNA polymerase II core promoter." *Annu Rev Biochem* **72**: 449-79.
- Sugiura, T., K. Sakurai, et al. (2007). "Intracellular characterization of DDX39, a novel growth-associated RNA helicase." *Exp Cell Res* **313**(4): 782-90.
- Taverner, N. V., J. C. Smith, et al. (2004). "Identifying transcriptional targets." *Genome Biol* **5**(3): 210.
- Thomas, M. C. and C. M. Chiang (2006). "The general transcription machinery and general cofactors." *Crit Rev Biochem Mol Biol* **41**(3): 105-78.
- Ueda, J., M. Tachibana, et al. (2006). "Zinc finger protein Wiz links G9a/GLP histone methyltransferases to the co-repressor molecule CtBP." *J Biol Chem* **281**(29): 20120-8.
- Van Lint, C., J. Ghysdael, et al. (1994). "A transcriptional regulatory element is associated with a nuclease-hypersensitive site in the pol gene of human immunodeficiency virus type 1." *J Virol* **68**(4): 2632-48.
- Verdin, E. (1991). "DNase I-hypersensitive sites are associated with both long terminal repeats and with the intragenic enhancer of integrated human immunodeficiency virus type 1." *J Virol* **65**(12): 6790-9.
- Verdin, E., N. Becker, et al. (1990). "Identification and characterization of an enhancer in the coding region of the genome of human immunodeficiency virus type 1." *Proc Natl Acad Sci U S A* **87**(12): 4874-8.
- Vigneault, F. and S. L. Guerin (2005). "Regulation of gene expression: probing DNA-protein interactions in vivo and in vitro." *Expert Rev Proteomics* **2**(5): 705-18.
- Wang, J., T. Li, et al. (2005). "Exonuclease III protection assay with FRET probe for detecting DNA-binding proteins." *Nucleic Acids Res* **33**(2): e23.
- Whitmarsh, A. J. and R. J. Davis (1996). "Transcription factor AP-1 regulation by mitogen-activated protein kinase signal transduction pathways." *J Mol Med* **74**(10): 589-607.
- Williams, C. J., T. Naito, et al. (2004). "The chromatin remodeler Mi-2beta is required for CD4 expression and T cell development." *Immunity* **20**(6): 719-33.

- Wysocka, J. and W. Herr (2003). "The herpes simplex virus VP16-induced complex: the makings of a regulatory switch." Trends Biochem Sci **28**(6): 294-304.
- Xiao, H., C. Neuveut, et al. (1998). "Interaction of the second coding exon of Tat with human EF-1 delta delineates a mechanism for HIV-1-mediated shut-off of host mRNA translation." Biochem Biophys Res Commun **244**(2): 384-9.
- Yang, V. W. (1998). "Eukaryotic transcription factors: identification, characterization and functions." J Nutr **128**(11): 2045-51.
- Zhang, Y., G. LeRoy, et al. (1998). "The dermatomyositis-specific autoantigen Mi2 is a component of a complex containing histone deacetylase and nucleosome remodeling activities." Cell **95**(2): 279-89.