STÉPHANE OUELLET

ÉTUDES DES MÉCANISMES DE RÉGULATION DE LA TRANSCRIPTION DES GÈNES HUMAINS *P21*$^{CIP1/WAF1}$, *GPC3* (*GLYPICAN 3*) ET *AβPP* (*AMYLOID β-PRECURSOR PROTEIN*).

Thèse présentée
à la Faculté des études supérieures de l'Université Laval
dans le cadre du programme de doctorat en biologie cellulaire et moléculaire
pour l'obtention du grade de Philosophiae Doctor (Ph.D.).

FACULTÉ DE MÉDECINE
UNIVERSITÉ LAVAL
QUÉBEC

2010

Résumé

La connaissance des mécanismes qui contrôlent la transcription des gènes, comme l'interaction de facteurs protéiques au niveau des promoteurs, est essentielle pour comprendre plusieurs fonctions biologiques et espérer traiter certaines pathologies. Nous nous sommes attardés à mieux comprendre les mécanismes qui régulent trois gènes humains dont les patrons d'expression sont différents et qui sont impliqués dans diverses pathologies en tentant de mettre en parallèle les différences structurales et fonctionnelles entre ceux-ci. Les gènes *AβPP* et *p21* sont exprimés chez l'adulte alors que *GPC3* est exprimé uniquement chez l'embryon de façon spécifique aux tissus. L'expression d'*AβPP* est aussi spécifique aux tissus et sa surexpression fait partie des mécanismes mis en cause chez les patients atteints de la maladie d'Alzheimer et du syndrome de Down. Le gène *p21* est quant à lui exprimé dans plusieurs types cellulaires et est fortement induit suite à des dommages à l'ADN. Enfin, nous avons montré que *GPC3* est exprimé de manière différentielle dans le neuroblastome (NB) et la tumeur de Wilms (WT), deux tumeurs embryonnaires.

p21 : La caractérisation du promoteur proximal de *p21* dans les fibroblastes humains normaux en prolifération nous a permis de localiser sept empreintes protéiques dont une au niveau de la séquence consensus pour NFI. Les études de retard sur gel, de transfection transitoire, d'immunoprécipitation de la chromatine et d'anti-ARN ont permis de confirmer la liaison de NFI et de le définir comme répresseur important de la transcription de *p21*.

AβPP : Nous avons montré que USF et Sp1 se lient au promoteur de *AβPP* et que leur liaison est essentielle pour générer une activité maximale du promoteur. La caractérisation *in cellulo* du promoteur dans les neurones et les astrocytes normaux a révélé huit sites d'interaction ADN-protéine, entre-autres au niveau des sites de liaison des facteurs de transcription CTCF, USF et Sp1.

GPC3 : La caractérisation du promoteur de *GPC3* nous a permis de montrer 1) une struture chromatinienne particulière tout le long du promoteur et 2) plusieurs empreintes protéiniques putatives dont certaines spécifiques à la lignée SJNB-7 qui exprime *GPC3*.

Parmi ces dernières, nous avons mis en évidence la liaison possible d'un facteur de transcription de type NF-Y.

Abstract

Gene transcription is the first step to the production of any given protein. Understanding of the molecular mechanisms regulating gene expression, such as the binding of transcription factors to genes promoters, is essential to the understanding of biological functions and to develop new powerful therapies against many clinically documented pathologies. We investigated the transcriptional regulatory mechanisms of three human genes very differently expressed and involved in diverse pathologies in an attempt to reveal structurals and functionals differencies between these mechanisms. *AβPP* and *p21* genes are both expressed in adult while *GPC3* is only transcribed in a tissus specific manner before birth. The expression of the *AβPP* gene is also specific to tissue and its over-expression may be involved in Alzheimer disease and Down syndrome. *P21* gene is expressed in many types of cells and is strongly induced by DNA damage. Finally, we demonstrated that GPC3 is differently expressed in neuroblastoma and Wilms' tumor.

P21 : The characterization of the proximal promoter from the *p21* gene in normal human proliferating fibroblasts revealed seven DNA-protein footprints of which one bears a perfect consensus sequence for the NFI family of transcription factors. EMSA, CHIP, anti-RNA and transient transfection of recombinant constructs analyses clearly demonstrated that NFI interact with the most proximal LMPCR footprint on the *p21* promoter and functions as a repressor. Upon serum starvation, a change in the electrophoretic mobility of the NFI DNA-protein complex was observed that may contribute to the activation mechanistic of the *p21* gene throughout cell senescence and differentiation.

AβPP : We demonstrated that Sp1, like USF, recognizes an element in the human *AβPP* gene that is necessary for full promoter activity. *In cellulo* footprinting analysis revealed at least eight DNA-protein interactions including CTCF, USF and many Sp1 target sites. These results were further supported by EMSA and transient transfection analysis.

GPC3 : The characterisation of the entire *GPC3* gene promoter revealed 1) a particular DNA structure in the promoter and 2) eight large protected regions. The use of competitor oligos in EMSA experiments and super-shift assays showed that an NFY-type transcription factor (TF) may explain the *GPC3* aberrant expression in SJNB-7.

Remerciements

Je voudrais tout d'abord souligner l'appui et la patience exceptionnelle de mon Directeur de thèse, le Dr Régen Drouin, qui, au cours de ces sept dernières années, a su développer en moi l'esprit scientifique et m'inculquer plusieurs qualités essentielles à l'art de la recherche. Ces années m'ont permis de connaître et de travailler avec plusieurs personnes qui ont tous collaboré à leur façon à ma formation personnelle et professionnelle. Un merci particulier à Isabelle, qui m'assiste depuis longtemps dans mes projets de recherches, d'escalade et de vélo. Merci à Martin qui fut en quelques sortes mon mentor, à Patrick à qui je dois quelques consommations pour son support informatique, et aux autres membres du laboratoire du Dr Drouin qui ont rendu mon séjour si agréable malgré des moments plus pénibles. Merci aussi à tous ceux qui ont, ou avec qui j'ai collaboré, aux cours des différents projets de recherches de mes études supérieures, dont les noms figurent au sein des articles de cette thèse. Plus particulièrement, merci au Dr Sylvain Guérin, mon co-Directeur, qui au cours de mes dernières années d'études doctorales a su renouveler mon appétit scientifique en plus de m'accueillir dans son laboratoire. Pour terminer, merci à tous les membres de ma famille et nombreux ami(e)s qui m'ont toujours supporté et encouragé dans ce projet de vie qu'est la poursuite d'études graduées.

Stéphane Ouellet

Contribution aux travaux

Dans les paragraphes qui suivent, je décris la teneur de ma contribution et celle des co-auteurs aux ouvrages scientifiques qui composent la thèse.

Chapitre 2

Ces travaux ont été publiés le 27 novembre 2006 dans périodique *Nucleic Acid Research.* Nous étions trois étudiants et un assistant de recherche à travailler sur ce projet. François Vigneault (étudiant au doctorat au moment de la réalisation des travaux) a essentiellement exécuté les manipulations d'immunoprécipitation de la chromatine (ChIP). Il a également participé à l'interprétation des résultats ainsi qu'à l'écriture et la révision de l'article. Maryse Lessard l'a accompagné dans les manipulations. En ce qui me concerne, j'ai contribué à l'élaboration du projet, participé aux manipulations pour environ 50%, produit plusieurs figures et contribué majoritairement à l'écriture de l'article et à l'interprétation des résultats. Quant à Steeve Leclerc, il fut le maître d'œuvre des manipulations *in vitro*, ce qui représente environ 40% des manipulations.

Chapitre 3

Les résultats de ce projet ont été publié dans la revue *Biochimica and Biophysica Acta* le 9 février 2007. Gino Boily et moi sommes les deux étudiants principaux de cet ouvrage, j'ai effectué une bonne partie des manipulations touchant la caractérisation intracellulaire du promoteur du gène à l'étude en compagnie de l'assistante de recherche Isabelle Paradis, en plus de rédiger une partie de l'article (10%) et de produire la figure 1. Gino Boily (à la maîtrise à ce moment) a développé le projet, participé aux manipulations, analysé les résultats et écrit l'article.

Chapitre 4

Cet article sera soumis prochainement à la revue *Journal of biological chemistry*. Nous étions deux étudiants à travailler sur ce projet. Martin Bourbonnière a essentiellement complété la réalisation des expériences *in vitro*, il a produit les figures s'y rapportant et participé à l'écriture de l'article. Avec l'assistante de recherche Isabelle Paradis, j'ai réalisé toutes les manipulations se rapportant à la caractérisation intracellulaire du promoteur du gène à l'étude. J'ai interprété les données qui résultèrent de ces manipulations et produit les figures qui décrivent ces résultats. J'ai participé à la rédaction des sections « matériel et méthodes », « résultats » et « discussion » du manuscrit.

Table des matières

Liste des tableaux et figures

CHAPITRE 1

CHAPITRE 2

ANNEXE

Chapitre 1 : Introduction générale

1.1 La mise en situation

Selon de récentes estimations, le génome humain serait composé d'environ 3×10^9 paires de bases (pb) et comprendrait environ 30,000 gènes répartis plus ou moins uniformément le long des 23 différents chromosomes de la femme ou des 24 de l'homme (1-3). D'autres équipes qui ont utilisé des méthodes différentes pour estimer le nombre de gènes avancent des chiffres différents et soutiennent que le génome humain compterait de 50,000 à 120,000 gènes (4, 5). Toutefois au-delà des chiffres, une question fondamentale demeure : quels sont les mécanismes qui orchestrent dans le temps et l'espace l'expression de tous ces gènes? La compréhension de ces mécanismes est l'objectif de nombreuses équipes de recherche (6, 7). Le développement d'un organisme vivant depuis sa fécondation jusqu'à sa maturité sexuelle et sa survie à l'âge adulte dépendent grandement de l'harmonisation de l'expression de ses gènes en réponse à des stimuli internes et externes. C'est l'expression organisée d'une suite encore mal définie de gènes qui permet le développement harmonieux d'un organisme. Au cours du développement de l'embryon humain, les cellules, contenant pourtant la même information génétique, se spécialisent et se multiplient pour former divers organes et tissus. Comme certains gènes ont une fonction très spécifique, il est impératif qu'ils soient exprimés aux moments opportuns et dans les tissus appropriés.

Sans régulation de la transcription, l'expression anarchique des gènes ruinerait le maintien du métabolisme cellulaire normal et conséquemment nuirait au développement et à la survie de l'organisme lui-même. La régulation de l'expression des gènes apparaît donc comme un phénomène indispensable à la vie. Chaque cellule de l'organisme est ainsi pourvue de divers mécanismes complexes qui assurent l'expression ordonnée de chacun de ses gènes. Le dérèglement de l'un de ces mécanismes peut avoir des conséquences dramatiques non seulement pour la cellule mais également pour l'organisme entier. Les médecins et les biologistes du développement fondent beaucoup d'espoir sur la contribution que pourrait apporter la compréhension de l'expression différentielle des gènes à la lutte contre certaines maladies congénitales, au vieillissement et au cancer (8).

L'exemple du cancer représente parfaitement les conséquences de l'expression inadéquate de certains gènes du cycle cellulaire. L'expression de certains oncogènes peut être totalement modifiée par une translocation chromosomique. Par exemple, l'expression de l'oncogène de la myélocytomatose cellulaire (*cellular myelocytomatosis*, *c-myc*) peut être augmentée suite à la translocation 8;14 (9). En effet, suite à cette translocation, l'expression de *c-myc* se retrouve sous le contrôle du promoteur du gène des chaînes lourdes des immunoglobulines. L'expression de *c-myc,* un oncogène qui stimule la division cellulaire, devient donc constitutive dans certains types cellulaires et contribue au développement tumoral.

Un éventuel contrôle médical de l'expression de certains gènes pourrait permettre le développement de nouvelles thérapies révolutionnaires ou du moins d'excellents moyens pour prévenir certaines maladies (7). Il existe déjà des molécules synthétiques qui peuvent interagir et séquestrer d'autres protéines, comme des facteurs de transcription ou des cofacteurs, ou se lier de façon spécifique à des séquences d'acide désoxyribonucléique (ADN) qui tous deux peuvent être essentiels à l'activation ou la répression de la transcription d'un gène cible comme *c-myc*.

1.2 Les principes de la régulation de l'expression des gènes

1.2.1 Les premiers modèles

En 2001, on annonça le séquençage complet du génome humain (1, 3). L'alignement de ces innombrables fragments de séquence ne cessant de progresser, l'objectif consiste maintenant à comprendre le fonctionnement des gènes, c'est-à-dire de déterminer la fonction et le patron d'expression de chacun d'eux. Toutefois, les scientifiques avaient

entrepris d'étudier les mécanismes de régulation de l'expression des gènes bien avant que la carte du génome humain ne soit disponible. Dès le milieu du vingtième siècle, les connaissances acquises en biologie cellulaire avaient permis de développer des modèles cohérents de régulation génique. Les premiers concepts en ce domaine ont été formulés par François Jacob et Jacques Monod, récipiendaires du Nobel de physiologie/médecine en 1965.
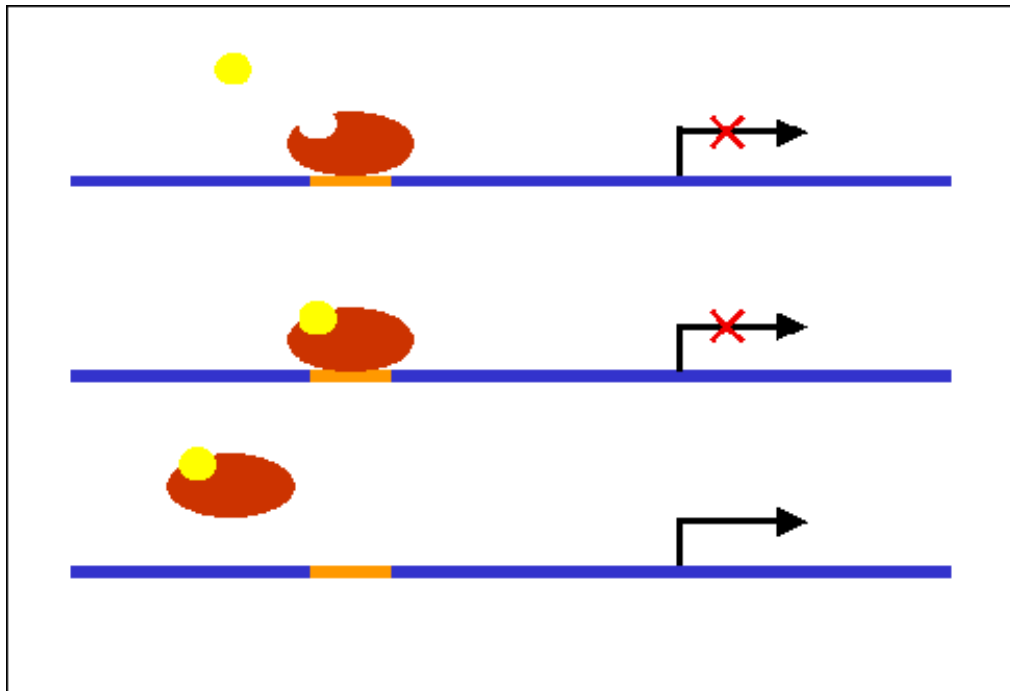
Figure 1 Représentation du modèle de Jacob et Monod. Le répresseur (rouge) est lié à l'opérateur (orange) et bloque l'expression du gène (bleu). Un agent (jaune) vient se lier au répresseur qui désormais ne peut plus se lier à l'opérateur, permettant ainsi l'expression du gène.

Ils furent les premiers à proposer l'existence d'un « messager » servant d'intermédiaire entre les gènes et leurs effecteurs biochimiques, les protéines. Comme le messager est tributaire de l'expression d'un gène, cela implique qu'il devait exister un mécanisme qui contrôlait la production du messager. Ils présentèrent alors le concept de l'« opérateur » qui est

aujourd'hui défini comme étant le « promoteur ». Dans leur modèle (Figure 1), l'opérateur contrôlait l'expression d'un gène à la manière d'un interrupteur; des « répresseurs » étaient liés sur l'opérateur et bloquaient l'expression de ce gène, donc la production du messager. C'est l'inhibition de l'action de ces répresseurs qui permettait l'expression du gène. Lorsqu'on s'attarde à l'histoire de ce domaine de recherche, on s'aperçoit que les modèles ont souvent précédé les explications moléculaires. Grâce au développement de techniques plus puissantes et spécialisées, aujourd'hui nous comprenons mieux la structure des gènes et les différents éléments biochimiques qui interagissent avec eux pour les réguler. Les connaissances actuelles ont permis de confirmer, de renforcer et même de raffiner les premiers modèles.

1.2.2 Une régulation implique une organisation

Dans le noyau d'une cellule humaine, l'ADN est fractionné en 23 paires de chromosomes et il est condensé sous forme de chromatine (10). Ainsi, une molécule d'ADN linéaire longue de dix centimètres sera condensée en un chromosome 10,000 fois plus petit mesurant à peine dix μm de long (10). La condensation de l'ADN en chromatine s'organise de manière séquentielle et ordonnée. En premier lieu, 146 paires de bases (pb) d'ADN s'enroulent ainsi autour d'un octamère d'histones pour former un nucléosome condensant l'ADN par un facteur de sept (11). Jusqu'à 95 % de l'ADN génomique serait ainsi associé à ces nucléosomes, constituant le premier niveau d'organisation de la chromatine (11). Dans un second niveau d'organisation, les nucléosomes se compactent et forment une hélice au rythme de six nucléosomes par tour produisant une fibre de 30 nm d'épaisseur (11). Cette fibre est finalement réorganisée de façon complexe en euchromatine (condensation légère) ou en hétérochromatine (condensation forte) constituant le troisième et dernier niveau d'organisation (11).

Or, seuls les gènes localisés dans l'euchromatine peuvent être potentiellement transcrits (11-18). Cette organisation structurale de l'ADN dans le noyau constituerait donc en elle-même un mécanisme de répression de la transcription des gènes. L'acétylation des histones et la méthylation des CpGs sont parmi les processus participant à la restructuration de la chromatine les plus étudiés (12-16). Les facteurs de transcription joueraient un rôle déterminant dans ces processus (11, 17, 18).

Les facteurs de transcription sont des protéines qui se lient généralement à une séquence bien précise de l'ADN et qui induisent, facilitent, nuisent ou inactivent la transcription des gènes de maintes façons (11-18). Ils peuvent notamment agir au niveau de la condensation de la chromatine, à l'initiation de la transcription et à l'élongation du transcrit. Un modèle hypothétique de régulation de la transcription d'un gène est illustré à la Figure 2. Si certains facteurs de transcription comme Sp1 et AP-1 (*clathrin-associated proteins type 1*; protéines de type 1 associées à la clathrine) sont associés à l'activation de plusieurs gènes, d'autres comme SREBP-2 (*sterol regulatory element binding protein-2*; protéine qui se lie à l'élément régulateur des stérols) limitent leur action à quelques gènes particuliers (19-21). Outre sa nature, la quantité de facteurs de transcription fonctionnels disponibles au locus d'un gène dans le noyau est déterminante pour initier avec succès sa transcription (28-32). Il a de plus été montré que chez la levure, plusieurs gènes situés très près les uns des autres sur un même chromosome étaient souvent co-exprimés, donnant lieu à une co-expression régionale de ces gènes (22). Il existerait aussi une certaine organisation qui favoriserait la formation de domaines de transcription au sein du noyau. Selon cette théorie, l'emplacement des chromosomes dans le noyau ne serait pas aléatoire. Ainsi, deux gènes localisés sur des chromosomes différents mais participant à un même phénotype ou faisant partie du même sentier métabolique pourraient être confinés dans un même domaine de transcription, soit très près l'un de l'autre dans le noyau interphasique (23-25). Ces deux phénomènes faciliteraient donc l'expression simultanée de gènes d'une même famille ou associés à un phénotype commun.
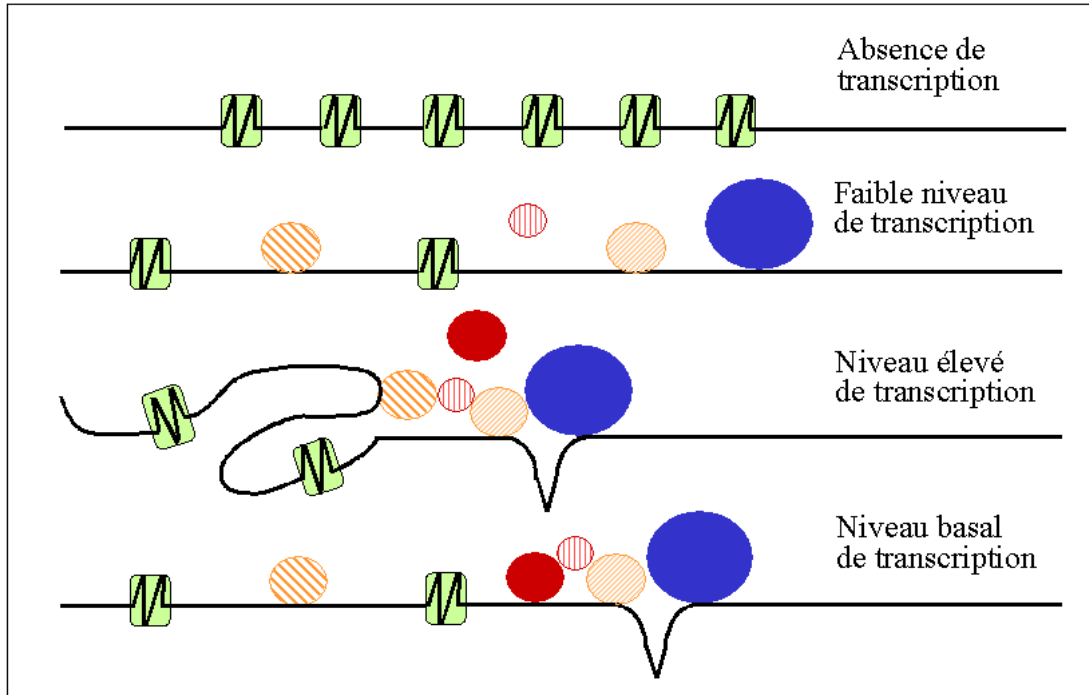
Figure 2 Illustration d'un des nombreux modèles théoriques qui peuvent expliquer la régulation de la transcription des gènes. Suite à l'ouverture de la chromatine (vert), la machinerie basale de transcription (bleu) a accès au promoteur et le gène produit des messagers. De manière coopérative, les facteurs de transcription activateurs (bariolés) forment un complexe qui facilite l'initiation de la transcription jusqu'à ce qu'un répresseur (rouge) vienne déstabiliser le complexe.

1.3 La structure et l'activation de la transcription d'un gène

1.3.1 La structure d'un gène

La plupart des gènes de classe II, c'est-à-dire ceux qui codent pour tous les acides ribonucléiques messagers (ARN_m), sont constitués de deux séquences d'ADN distinctes: la séquence codante et la séquence régulatrice. La séquence codante est celle qui est réellement transcrite en ARN_m. Elle est composée d'exons et d'introns et sert de gabarit à la production d'une chaîne polypeptidique. La séquence régulatrice, également appelée

séquence promotrice ou promoteur, est généralement située en amont de la séquence codante et contrôle la production du messager. Toutefois, il a été montré que certaines séquences introniques de gènes comme *BRCA1* (*breast cancer gene 1*; gène 1 du cancer du sein, type 1) et *CFTR* (*cystic fibrosis transmembrane conductance regulator*; régulateur de la conductivité transmembranaire de la fibrose kystique) participaient à la régulation de la transcription (157, 158). Le promoteur, qui peut parfois s'étendre sur des milliers de pb d'ADN, est divisé en trois parties distinctes (Figure 3): 1) la région promotrice proximale, 2) la région régulatrice et 3) la région stimulatrice/modératrice (*enhancer*/*silencer*).

1) La région promotrice proximale s'étend approximativement des nucléotides -30 à +30 par rapport au site d'initiation de la transcription (SIT). Elle est essentielle à la liaison et au positionnement de l'holoenzyme, constituée de l'ARN polymérase II (ARN pol II) et de ses cofacteurs, au SIT (Figure 4). Pour les nombreux gènes dont l'expression est limitée à certains tissus, le motif conservé thymine-adénine-thymine-adénine-adénine-adénine (TATAAA; *TATA box*; boîte TATA) y est souvent présent. La protéine TBP (*TATA box binding protein*; protéine qui se lie à la boîte TATA) se lie à ce motif et recrute par la suite l'ARN pol II et ses diverses composantes au SIT. Pour les autres gènes tissus-spécifiques qui n'ont pas de boîte TATA, le positionnement de l'ARN pol II au SIT est assuré par la séquence initiatrice (*initiator*) positionnée entre les nucléotides -4 et +3 et/ou par celle de l'élément qui suit la région promotrice proximale (*downstream core promoter element*; DPE), positionnées entre les nucléotides +30 à +36 (26, 27). Généralement, les gènes dont l'expression est ubiquitaire, les gènes d'entretien, ne contiennent ni le motif TATAAA ni la séquence initiatrice (28). En conséquence, ces gènes possèdent de multiples SIT. Leurs promoteurs sont généralement riches en nucléotides GC, ce qui génère souvent de nombreux motifs reconnus par le facteur de transcription Sp1 (19). Pour les gènes d'entretien, c'est le facteur de transcription Sp1 qui positionnerait l'holoenzyme au SIT. Alors que l'holoenzyme est capable d'initier à elle seule la transcription en éprouvette, des cofacteurs additionnels lui sont nécessaires pour l'initier dans le noyau d'une cellule vivante (28).

2) Dans le noyau d'une cellule vivante, la seconde région du promoteur est essentielle à l'initiation efficace de la transcription. C'est dans cette région, située juste en amont de la région promotrice proximale, que se trouvent la majorité des séquences d'ADN reconnues par des facteurs de transcription (29).
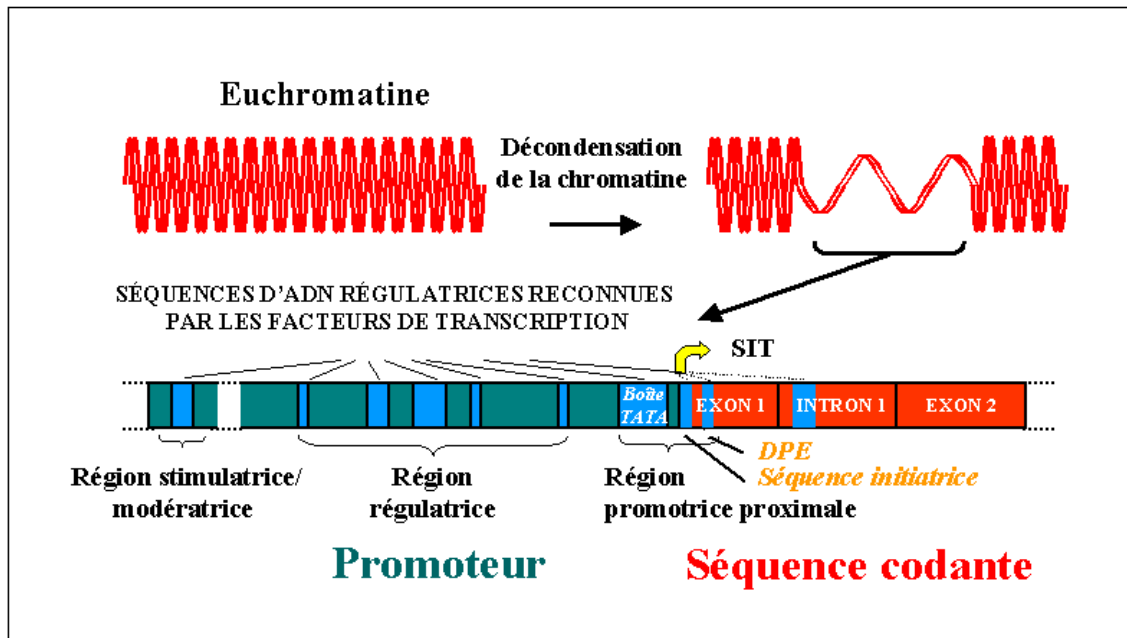
Figure 3 Illustration de la structure classique d'un gène. La décondensation de la chromatine rend les différentes séquences d'ADN régulatrices (boîte TATA, SIT, DPE, etc.) accessibles aux facteurs de transcription. Ces éléments régulateurs sont surtout situés dans le promoteur et sont composés de trois régions, mais ils peuvent également être situés dans la partie codante du gène, composée d'introns et d'exons.

3) La troisième région, la région stimulatrice/modératrice, est généralement située en amont du SIT, souvent de plusieurs milliers de pb d'ADN. Cette région participe également à la régulation de la transcription du gène mais surtout de manière ponctuelle et spécifique à un nombre restreint de tissus (30, 31). Si l'importance de ces séquences ne fait aucun doute, les mécanismes d'action de séquences régulatrices aussi éloignées du SIT demeurent encore vagues (32).

1.3.2 L'activation de la transcription d'un gène

Pour activer un gène donné, différents stimuli internes et/ou externes entraînent la liaison de quelques facteurs de transcription particuliers à certaines séquences d'ADN régulatrices situées sur le promoteur (13). Ces facteurs provoquent l'ouverture localisée de la chromatine permettant ainsi à d'autres facteurs de transcription d'accéder à leurs séquences d'ADN régulatrices sur le promoteur (18, 33). Ces dernières protéines vont recruter l'holoenzyme et stabiliser sa liaison au SIT du promoteur permettant une initiation efficace de la transcription (11, 17). Les facteurs de transcription qui se lient à la région stimulatrice stabiliseraient encore davantage la liaison de l'holoenzyme au SIT, augmentant ainsi le nombre d'initiations efficaces de la transcription.

Certains gènes comme celui de la β-globine, nécessiteraient une restructuration plus généralisée de la chromatine pour être activés (34). Les séquences LCR (*locus control region*; région de contrôle du locus), situées aux extrémités de certains gènes, favoriseraient la restructuration de la chromatine sur de grandes distances (35).

Certains facteurs, les activateurs, favorisent la transcription alors que d'autres, les répresseurs, l'inhibent. La liaison de répresseurs au promoteur d'un gène, tout comme la perte d'activateurs, peut mener à la condensation de la chromatine ou à la déstabilisation de la liaison de l'holoenzyme au SIT. Ces événements empêchent toute nouvelle initiation de la transcription. Les mécanismes d'actions des facteurs de transcription, qu'ils soient activateurs ou répresseurs, sont variés. Certains facteurs ont besoin d'un cofacteur ou d'une modification post-traductionnelle pour être actifs comme c'est le cas pour ATF-2 (*activating transcription factor 2*; facteur de transcription activateur 2) qui doit être phosphorylé pour se lier à l'ADN (36). D'autres facteurs, incluant les récepteurs des glucocorticoïdes, doivent être complexés avec leur ligand pour leur permettre de transactiver (37). Certains facteurs concentrent même leur action à un type cellulaire particulier comme c'est le cas pour Pitx3 (*paired-like homeodomain transcription factor 3*) qui semble exprimé uniquement dans certains neurones dopaminergiques à partir du jour 11 chez les embryons de souris (38).
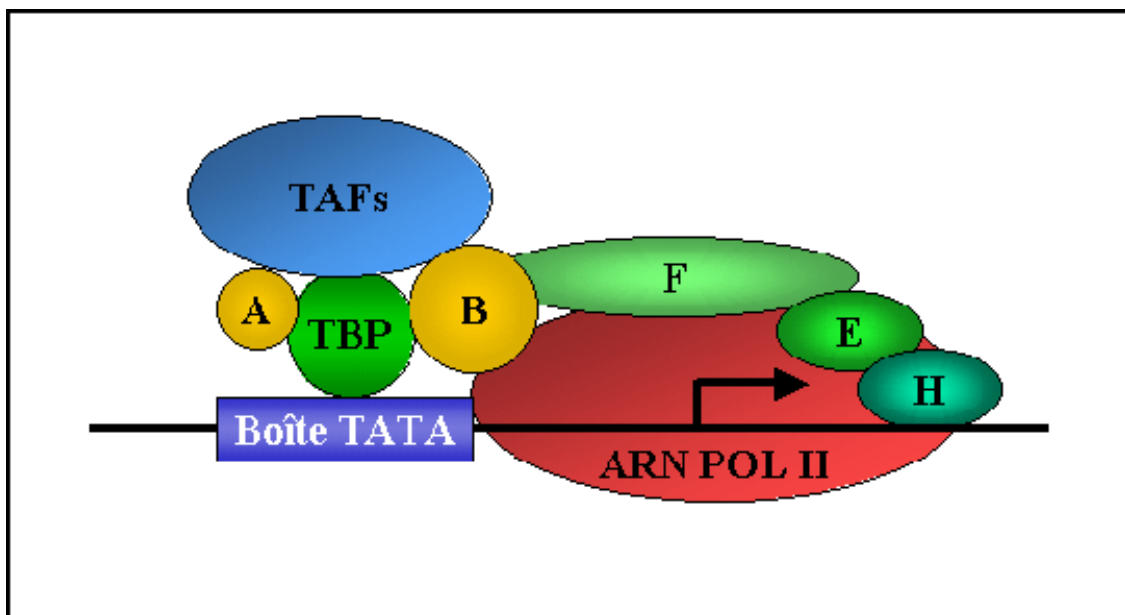
Figure 4 Représentation des principales composantes de la machinerie basale de transcription.

Si les facteurs de transcription jouent un rôle majeur dans la régulation d'une réponse à un stimulus, ils n'en sont pas les seuls acteurs puisque la régulation peut avoir lieu à différents niveaux (Figure 5): 1- Le signal provenant de l'environnement pourra ainsi être amplifié ou amoindri lors de sa transmission au noyau, selon les besoins de la cellule (39, 40). 2- La quantité disponible de facteurs de transcription au SIT est également critique pour initier la transcription comme c'est le cas pour les gènes qui nécessitent une translocation nucléaire de NF-κB (*nuclear factor-kappaB*; facteur nucléaire kappaB (159)). 3- Le transcrit du gène, l'ARN$_m$, peut lui-même être la cible d'une régulation lors de son épissage (maturation) et de son exportation extranucléaire. Il pourra également être dégradé plus ou moins rapidement ou même emmagasiné pour une utilisation ultérieure. Par exemple, les ARE (*AU-rich elements*; éléments riches en adénines et en uraciles) situés en 3' dans la région non-traduite (*untranslated region*, UTR) de plusieurs ARN$_m$, sont parmi les éléments responsables de l'instabilité de ces ARN$_m$ et donc de leur courte demi-vie (41). 4- La traduction de l'ARN$_m$ en protéine est également soumise à une régulation et est de plus en plus considéré comme un centre de régulation d'importance (42, 43). 5- Les modifications post-traductionnelles de diverses natures comme la glycosylation et la phosphorylation gèrent le maintien de la structure et de l'activité de la protéine plus ou moins longtemps dans la cellule (44). 6- Finalement, les diverses voies de dégradation protéique contrôlent aussi la quantité de protéines.

La régulation de l'expression d'un gène est donc très complexe et ne se limite pas au contrôle de sa transcription (Figure 5). Même si les connaissances en ce domaine ont beaucoup progressé, elles demeurent encore fragmentaires. Toutefois, les séquences d'ADN régulatrices situées sur le gène, les éléments *cis* et les facteurs de transcription qui s'y lient, les facteurs *trans*, ont été largement étudiés. Des décennies de travaux ont ainsi permis de produire plusieurs banques de données des séquences qui lient les facteurs de transcription (ex: TRANSFAC) (45, 46). De plus, de nombreux facteurs de transcription ont été purifiés et classés selon leurs propriétés régulatrices ou leur domaine de liaison à l'ADN.
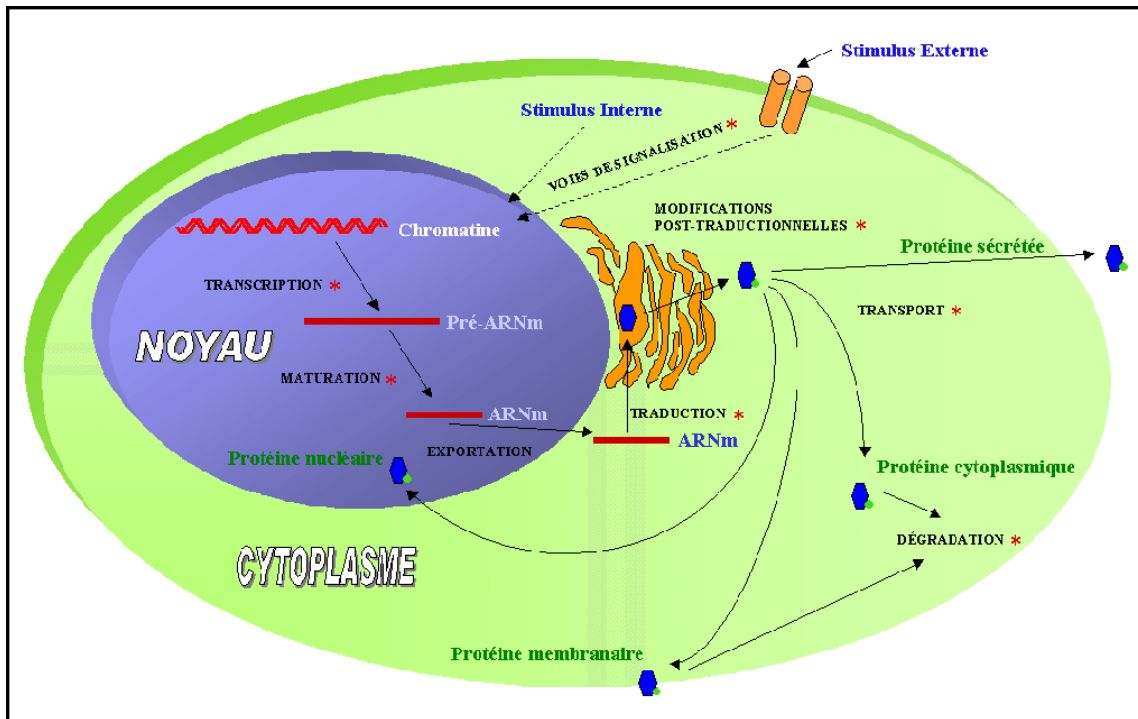
Figure 5 Illustration des principales étapes menant à la production d'une protéine. Chaque étape marquée d'un astérisque (*) correspond à un point de contrôle possible de l'expression d'un gène.

La famille des facteurs de transcription ATF/CREB (*activating transcription factor,* facteur de transcription activateur/*cyclic adénine monophosphate response element-binding protein*; protéine qui se lie à l'élément de réponse de l'adénine monophosphate cyclique) qui lient la séquence d'ADN consensus TGACGTCA est bien définie (47). Ainsi, on connaît certains acteurs qui participent au contrôle de l'initiation de la transcription des gènes et parfois comment ces acteurs interagissent entre eux et avec l'ADN dans la cellule (48). Cependant les réponses à d'autres questions restent évasives. Par exemple, pourquoi un facteur de transcription particulier peut-il se lier au promoteur d'un gène et participer à la régulation de sa transcription alors qu'il ne se lie pas au promoteur d'un autre gène bien que la séquence d'ADN consensus pouvant lier ce facteur est présente dans le promoteur des deux gènes? Malgré le manque de preuves, l'hypothèse retenue jusqu'ici fait référence

à l'intervention de cofacteurs conférant au facteur de transcription en question une spécificité autre que sa séquence d'ADN de liaison connue.

1.4 Les stratégies utilisées pour étudier la régulation génique à partir de molécules d'ADN purifiées ou synthétisées

L'analyse des interactions ADN-protéine, de la structure de la chromatine ainsi que l'identification des facteurs de transcription fournissent diverses informations essentielles à la compréhension des mécanismes qui affectent la transcription des gènes. Il y a quelques années, le scientifique était techniquement limité à l'utilisation de matériel soustrait, isolé de son environnement cellulaire normal pour effectuer ce type d'études. Les techniques qui utilisent ces molécules purifiées pour analyser les interactions ADN-protéine apportent néanmoins diverses informations très utiles. Elles permettent entre autres: 1) d'identifier la protéine qui se lie à une séquence spécifique d'ADN, 2) d'identifier le domaine de liaison à l'ADN de cette protéine ainsi que les nucléotides de la séquence d'ADN impliqués dans cette liaison, 3) de mesurer l'effet de la liaison de la protéine à sa séquence d'ADN consensus sur le niveau de transcription d'un gène rapporteur, 4) de délimiter les régions du promoteur qui stimulent ou inhibent la transcription et 5) de vérifier la présence de nucléosomes.

L'une des techniques les plus employées est le gel à retardement (*electrophoretic mobility shift assays*, EMSA) dont les principes ont été énoncés au début des années 80 (49, 50). Soumis à un courant électrique, les molécules d'ADN migrent à travers un gel de polyacrylamide à une vitesse proportionnelle à leur poids moléculaire. Or, en conditions non-dénaturantes, la liaison d'un ou de plusieurs facteurs protéiniques à un fragment d'ADN préalablement marqué (généralement à l'aide d'un radio-isotope), que l'on nomme sonde, augmente sa taille et freine sa migration à l'intérieur du gel (51) (chapitre 2). Les protéines de tailles, de poids moléculaires et de charges différentes ayant une mobilité

électrophorétique distincte, leurs liaisons engendreront des patrons de bande(s) retardée(s) qui leur sont spécifiques. La majorité des études utilisent les extraits nucléaires totaux des cellules spécifiques à leur recherche comme source de facteurs protéiques. Ainsi, la présence d'un ou de plusieurs facteurs ayant la faculté de lier la sonde génèrera sur un gel à retardement l'apparition d'une ou de plusieurs bandes distinctes de celles qui correspondent à la sonde d'ADN libre (dépourvue de liaison protéique). Cette technique est très versatile et selon les besoins, il est possible d'utiliser des compétiteurs non-marqués, des anticorps ou des fragments d'ADN modifiés comme sonde pour confirmer l'identité d'une protéine se liant à un fragment d'ADN (52) dans le but de déterminer le motif d'ADN reconnu par un facteur de transcription (53), pour mesurer l'affinité de la protéine pour une séquence d'ADN (54) ou pour localiser le site d'attachement à l'ADN sur les protéines (55) (chapitre 3).

Une autre approche couramment utilisée pour délimiter les séquences d'ADN qui sont liées par un facteur de transcription est l'analyse à l'endonucléase de l'ADN 1 (*deoxyribonucleotide endonuclease 1*, DNase I). La DNase I est une enzyme qui se positionne au niveau du sillon mineur de l'ADN, indépendamment de la séquence, et qui rompt de façon aléatoire le lien phosphate entre deux nucléotides adjacents (56). Par encombrement stérique, une protéine interagissant avec un fragment d'ADN protégera la partie de ce fragment avec laquelle il est en contact contre l'activité endonucléase de la DNase I. Ceci permet de bien délimiter la région d'interaction entre la protéine et le fragment d'ADN. L'effet protecteur du facteur de transcription pourra être indirectement observé en comparant l'activité de la DNase I sur de l'ADN purifié à celle sur ce même ADN traité en présence de protéines. La migration de ces fragments sur gel de séquence permettra de visualiser les nucléotides protégés dans la séquence d'ADN de l'échantillon traité en présence de protéines par l'absence ou la faible intensité des bandes correspondantes par rapport à la séquence de l'échantillon d'ADN traité seul. On observe aussi fréquemment des sites qui deviennent hypersensibles au clivage de la DNase I aux frontières de la séquence d'ADN liée par un facteur de transcription. La DNase I est cependant beaucoup moins sensible que la technique EMSA car elle requiert de grandes

quantités de matériel protéique. Néanmoins, elle a le mérite de pouvoir détecter des interactions ADN-protéine qui ne peuvent être détectées par les autres méthodes (57). Outre la DNase I, d'autres agents peuvent être utilisés pour détecter *in vitro* les régions d'ADN protégées par la liaison de facteurs de transcription. On peut, entre autres, utiliser des agents méthylants comme le sulfate de diméthyle (*dimethyl sulfate*, DMS) qui méthyle les guanines à la position de l'azote-7 (58), des agents éthylants (59), les rayons ultraviolets (UV) (60) et des exonucléases (61).

La liaison de facteurs protéiques à l'ADN entraîne et parfois même nécessite une réorganisation de sa structure ce qui comprend des torsions, des repliements et des ouvertures de sa double-hélice. Or, il existe des molécules qui permettent de détecter ces modifications. Le permanganate de potassium ($KMnO_4$) par exemple, est une molécule très réactive qui oxyde préférentiellement les thymines d'ADN monocaténaire (62, 63). Il peut ainsi détecter des portions de séquences d'ADN qui sont monocaténaires. Le $KMnO_4$ a été utilisé pour caractériser des distorsions dépendantes de la séquence (64), des structures d'ADN en épingle à cheveux (65) ou l'ADN de type B et l'ADN à hélice triple (66). De plus, l'utilisation d'autres agents et de certaines variations de cette technique peuvent servir à l'étude d'autres phénomènes ayant un impact sur la régulation de la transcription comme la méthylation des CpG et l'acétylation des histones (14, 67).

Dans un autre ordre d'idée, l'analyse d'un promoteur par transfections transitoires de cellules vivantes permet de délimiter les régions activatrices et inhibitrices du promoteur (68) et de mesurer l'activité des facteurs de transcription (69). À mi-chemin entre l'*in cellulo* et l'*in vitro*, cette technique consiste à intégrer dans le noyau de cellules vivantes des plasmides contenant diverses versions modifiées du promoteur à l'étude, que l'on nomme constructions, juxtaposées à un gène, que l'on qualifie de rapporteur, dont le produit de transcription est mesurable. Une construction contient généralement un promoteur dont une ou quelques régions particulières ont été enlevées ou dont le(s) site(s) de liaison pour un (des) facteur(s) de trancription a (ont) été muté(s). Les gènes rapporteurs

les plus communs sont les gènes de la *luciférase* et de la *β-galactosidase*. Les techniques de transfection sont de plus en plus variées, mais les plus courantes utilisent le calcium ($CaCl_2$), la lipofectamin ou un courant électrique (electroporation) pour permettre l'entrée des plasmides à l'intérieur des cellules. En comparant la quantité de produit du gène rapporteur généré par la transfection des différentes constructions du promoteur à l'étude, on peut évaluer l'importance de certaines régions promotrices ou mesurer l'effet de la liaison d'un facteur de transcription sur l'activité du promoteur. La particularité de cette technique est qu'elle utilise les mécanismes cellulaires endogènes de transcription et que l'ensemble des acteurs nucléaires sont présents. Il faut toutefois préciser que même si la transcription mesurée s'est produite à l'intérieur des noyaux de cellules vivantes, cette méthode est le plus souvent qualifiée d'*in vitro* puisque les gabarits d'ADN utilisés, les plasmides, ne se retrouvent pas sous forme de chromatine et ne comprennent le plus souvent qu'une partie du promoteur étudié.

1.5 Les principes de l'analyse intracellulaire de l'ADN

Dans les années 80, de nouvelles techniques ont été développées afin d'analyser l'ADN en tenant compte de l'environnement intracellulaire (70). En raison de cet avancement technologique, nous allons devoir en tenir compte dans nos définitions. Ainsi, nous allons définir une molécule d'ADN purifiée (donc extraite du noyau et dépouillée de toutes ses protéines) comme ADN *in vitro* et une molécule d'ADN sous forme de chromatine dans le noyau d'une cellule vivante comme ADN *in cellulo*. Ces qualificatifs (*in vitro* et *in cellulo* dans les articles nous avons conservé la dénomination de *in vivo*, plus employée il y a quelques années, mais *in vivo* réfère plus à des études chez l'animal) peuvent aussi se référer aux traitements ou aux techniques utilisant ces types d'ADN.

L'une des méthodologies à la base des techniques *in cellulo* d'investigation de l'ADN est l'utilisation d'agents de caractérisation qui causent des dommages à l'ADN intracellulaire.

Théoriquement, chaque nucléotide d'une molécule ADN purifiée possède la même probabilité d'être endommagé par un agent de caractérisation. Dans une cellule vivante cependant, l'environnement nucléaire peut modifier l'accessibilité de l'agent à certains nucléotides d'une séquence d'ADN (71-74). Ainsi, un facteur de transcription en contact avec certains nucléotides d'ADN *in cellulo* protégera ces nucléotides contre l'action de l'agent de caractérisation. Inversement, d'autres nucléotides seraient plus sensibles *in cellulo* à l'action de l'agent qu'une fois purifié et linéarisé (*in vitro*). En effet, dans les noyaux des cellules, certains nucléotides localisés au sein d'une structure particulière de la chromatine ou aux abords d'une interaction ADN-protéine ont, lorsqu'ils sont exposés à un agent, une réactivité plus grande qu'une fois que l'ADN a été purifié. Ainsi, ces nucléotides seront plus fréquemment endommagés par l'agent de caractérisation lors du traitement des cellules vivantes (*in cellulo*) que lors de celui de l'ADN purifié (*in vitro*).

L'analyse intracellulaire de l'ADN vise à comparer le degré d'accessibilité d'un agent de caractérisation pour chaque nucléotide entre l'ADN purifié et l'ADN d'une cellule vivante. Ceci permet de visualiser indirectement les interactions ADN-protéines et les structures secondaires spéciales à l'intérieur d'une cellule vivante.

1.5.1 In cellulo versus in vitro

Les analyses *in cellulo* et *in vitro* de l'ADN reposent sur les mêmes principes fondamentaux. La différence majeure réside dans le fait que dans les analyses *in vitro*, l'environnement intracellulaire est artificiellement recréé de façon plus ou moins fidèle, alors que dans les analyses *in cellulo*, les molécules d'ADN sont étudiées à l'intérieur même du noyau d'une cellule vivante. Même si elles sont toujours très utiles et largement répandues aujourd'hui, il est impossible de reproduire fidèlement la réalité intracellulaire avec les techniques *in vitro*. Dans une cellule vivante, la molécule d'ADN est sous forme de chromatine. Elle peut être méthylée ou adopter des structures secondaires particulières. De

plus, l'ADN interagit avec différentes protéines qui interviennent dans la régulation de la transcription des gènes, la réparation et la réplication de l'ADN, l'attachement de l'ADN à la matrice nucléaire, etc. Or ces éléments importants sont perdus suite à la purification ou au clonage de l'ADN (75). Ainsi, il s'est souvent avéré qu'une séquence promotrice participant à la régulation de l'expression d'un gène puisse lier un facteur de transcription particulier lors d'expérience *in vitro* et ne révéler aucune empreinte protéique *in cellulo*. L'étude de Chen et al (1995) du promoteur de l'antigène carcinoembryonaire de carcinomes du colon est un bel exemple de disparité qui peuvent surgir entre les études *in cellulo* et *in vitro* (76). Il est très difficile pour les études *in vitro* de reproduire l'influence du contexte nucléaire normal sur plusieurs phénomènes comme la liaison de facteurs protéiques à l'ADN. À la lumière de ces faits, il devient de plus en plus souhaitable de montrer ou de vérifier les mécanismes de régulation de la transcription des gènes avec des techniques qui tiennent compte de l'environnement intracellulaire en plus des techniques conventionnelles.

1.5.2 Les agents de caractérisation de l'ADN

Un agent de caractérisation doit posséder des caractéristiques bien précises afin d'être utilisé pour l'analyse intracellulaire de l'ADN. D'abord, il doit être sensible aux interactions ADN-protéines de l'ADN *in cellulo*. En d'autres termes, l'agent doit être suffisamment actif pour modifier l'ADN tout en ne déstabilisant pas la liaison des facteurs protéiques sur l'ADN. Idéalement, l'agent doit produire des cassures monocaténaires à l'ADN. Un agent qui ne produit pas directement de cassures à l'ADN peut toutefois être utilisé si les dommages qu'il cause à l'ADN sont convertibles en cassures monocaténaires. Incidemment, la conversion des dommages en cassures doit être efficace à près de 100% pour que la fréquence de cassures soit représentative des dommages initiaux au niveau de chaque nucléotide. Enfin, il est préférable que l'agent soit à même de pénétrer dans la cellule et d'agir rapidement sur l'ADN afin de réduire au minimum les perturbations qui pourraient modifier la nature des interactions ADN-protéines. Lorsque cela n'est pas

possible, certains agents qui perméabilisent la membrane plasmique permettant l'entrée de l'agent peuvent être utilisés.

La plupart des agents de caractérisation employés *in cellulo* sont les mêmes que ceux qui ont été utilisés avec les techniques *in vitro*. Trois d'entre eux sont généralement utilisés pour analyser les interactions ADN-protéines *in cellulo*. Il s'agit du DMS (un agent chimique), des irradiations UV de type B ou C (UVB, UVC; des agents physiques) et de la DNase I (un agent biochimique) (Tableau 1). D'autres agents, utilisés pour cartographier certaines structures particulières de l'ADN *in vitro*, peuvent également être utilisés *in cellulo*. Parmi ceux-ci, le $KMnO_4$ et la nucléase S1 qui sont sensibles à la présence d'ADN monocaténaire souvent associé à l'ADN transcrit (77). Certains groupes ont même utilisé le tétroxyde d'osmium (OsO_4) pour détecter des structures particulières de l'ADN comme l'ADN cruciforme. Nous pouvons également mentionner le bisulfite de sodium et l'hydrazine, deux molécules employées pour cartographier les cytosines méthylées de l'ADN génomique. Notez que la cartographie des cytosines méthylées peut être faits sur de l'ADN *in vitro* car la méthylation des cytosines n'est pas perdue lors de la purification de l'ADN.

Mis à part la DNase I et la nucléase S1, tous les autres agents de caractérisation produisent des dommages à l'ADN qui doivent être convertis en cassures monocaténaires. Il est préférable que le procédé utilisé pour convertir les dommages en cassures monocaténaires soit spécifique aux dommages induits par l'agent de caractérisation. Le traitement de l'ADN à la pipéridine à température élevée est une méthode de conversion très efficace mais pas spécifique à un dommage particulier (78). La forte alcalinité de la pipéridine déstabilise la base endommagée et ensuite la molécule de pipéridine provoque la rupture du lien glycosylé de la base endommagée à température élevée (78). Le site abasique ainsi créé est converti en cassure monocaténaire par les conditions fortement alcalines et la température élevée selon un processus de beta-élimination. Ce procédé laisse un phosphate à l'extrémité 5' et un site abasique avec un groupement OH (hydroxyle) à l'extrémité 3'.

Tableau 1 Les agents de caractérisation et leurs caractéristiques.

Approches	Applications	Avantages	Limites
SULFATE DE DIMÉTHYLE (DMS) Action : Méthylation préférentielle de l'azote en position 7 des guanines *Conversion en cassures : Par un traitement à la pipéridine (80 °C)*	1. Localise *in cellulo* les interactions ADN-protéines 2. Détecte certaines structures spéciales de l'ADN	Techniquement simple; le DMS est une petite molécule très réactive qui n'exige pas de perméabilisation pour pénétrer à l'intérieur des cellules	1. Requiert des guanines dans la séquence 2. Est incapable de détecter toutes les interactions ADN-protéines
IRRADIATION UV (UVB OU UVC) Action : Produit 2 types majeurs de dommages impliquant 2 pyrimidines adjacentes 1) dimères cyclobutyliques de pyrimidines (DCP) 2) photoproduits 6-4 (6-4pp) *Conversion en cassures :* *1) DCP : T₄ endoV suivie de la photolyase* *2) 6-4pp : photolyase suivie de la pipéridine (80 °C)*	1. Localise *in cellulo* les interactions ADN-protéines 2. Détecte certaines structures spéciales de l'ADN 3. Détermine parfois le positionnement des nucléosomes	1. Techniquement simple; les rayons UV traversent la membrane cytoplasmique sans la déstabiliser 2. Fixe les interactions ADN-protéines sans perturber le milieu intracellulaire	1. Requiert deux pyrimidines adjacentes dans la séquence 2. Il est parfois difficile de discerner une interaction ADN-protéine d'une structure spéciale de l'ADN
DNASE I Action : Produit des cassures à l'ADN sans égard à la séquence des quatre bases *Conversion en cassures :* *Pas nécessaire*	1. Localise *in cellulo* les interactions ADN-protéine 2. Cartographie *in cellulo* les sites hypersensibles à la DNase I 3. Détermine le positionnement des nucléosomes	1. Aucune restriction de séquence 2. N'exige pas de conversion 3. Capable de détecter la plupart des interactions ADN-protéines 4. Très sensible aux structures spéciales de l'ADN	1. Techniquement plus complexe et plus difficilement reproductible 2. La DNase I est une grosse protéine qui nécessite une perméabilisation membranaire pour pénétrer dans la cellule

La pipéridine est surtout utilisée pour convertir en cassures monocaténaires les guanines méthylées par le DMS. Elle peut également être utilisée avec les rayons UVB et UVC mais les dimères cyclobutyliques de pyrimidines (*cyclobutane pyrimidine dimers*, CPD) qu'ils produisent (79, 80) peuvent être spécifiquement convertis en cassures monocaténaires par un traitement avec des enzymes de réparation de l'ADN (75). Contrairement aux études de réparation de l'ADN par exemple, la conversion des dommages induit par les rayons UV en cassure par une méthode enzymatique n'est pas obligatoire pour détecter les interactions ADN-protéines *in cellulo* mais elle est recommandée (81).

1.5.3 Le séquençage génomique

Le séquençage chimique ou génomique fut mis au point par Church et Gilbert en 1984 (71). Brièvement, quatre parties aliquotes d'ADN génomique partiellement digérées par une enzyme de restriction sont respectivement traitées par un agent qui endommagera l'ADN de façon spécifique pour chacune des quatre bases de l'ADN. Après avoir converti les dommages en cassures monocaténaires par la pipéridine chaude, les quatre échantillons sont déposés sur un gel de polyacrylamide pour y être séparés selon leur poids moléculaire. Les molécules d'ADN sont ensuite transférées, fixées sur une membrane de nylon et détectées grâce à l'hybridation d'une sonde d'ADN monocaténaire spécifique à la région étudiée. Cette technique est suffisamment fiable et sensible pour détecter des différences d'intensité entre les bandes qui représentent les molécules d'ADN. Comme l'intensité d'une bande correspond à la fréquence de cassures monocaténaires induites par un agent de caractérisation au niveau d'un nucléotide particulier, le séquençage génomique peut être utilisé pour l'analyse intracellulaire de l'ADN.

Selon les mêmes principes que pour les analyses faites *in vitro*, si lors d'un traitement *in cellulo*, un nucléotide particulier est protégé de l'action d'un agent de caractérisation par la présence d'une protéine dans une forte proportion des cellules d'une population, les

fragments d'ADN isolés de ces cellules dont la cassure monocaténaire correspond à ce nucléotide seront moins nombreux que les fragments obtenus par le même traitement de l'ADN purifié. Dans ce cas, on observe une empreinte négative (*negative footprint*, Figure 6). Inversement, une bande plus intense *in cellulo* qu'avec l'ADN purifié sera appelée empreinte positive (*positive footprint*; Figure 6).
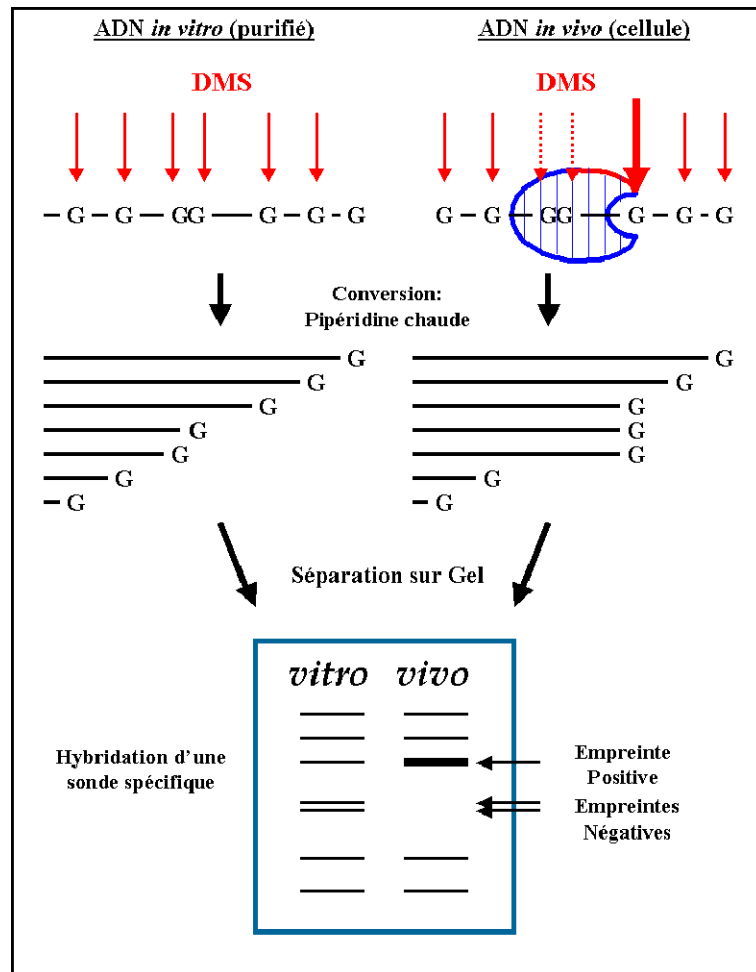
Figure 6 Principes de l'analyse intracellulaire de l'ADN par le DMS.

La technique de Church et Gilbert convient très bien à l'étude de génomes relativement petits comme celui de la levure. Cependant, le niveau de sensibilité de cette technique ne

permet pas d'étudier facilement des génomes plus volumineux comme ceux des mammifères (70). En effet, d'importantes quantités d'ADN doivent être utilisées pour obtenir un signal suffisamment visible pour être analysable.

À la fin des années 80, plusieurs alternatives dérivant de la technique de séquençage chimique de Church et Gilbert ont été développées afin de pallier à ce manque de sensibilité (70). Encore aujourd'hui, la seule alternative suffisamment sensible pour être appliquée aux grands génomes comme celui de l'humain, est la technologie de réaction de polymérisation en chaîne permise par un adaptateur (*ligation-mediated polymerase chain reaction*, LMPCR).

1.6 La technologie LMPCR

1.6.1 La description de la technologie LMPCR

À la base, la technique LMPCR est une technique de séquençage génomique très performante. Cette efficacité est dûe à l'utilisation de la réaction de polymérisation en chaîne (*polymerase chain reaction*, PCR) pour amplifier exponentiellement les fragments d'ADN qui présentent une cassure à une extrémité. Comme toutes les techniques de séquençage génomique, la technique LMPCR possède un degré de résolution qui se situe au niveau du nucléotide (75). De plus, l'utilisation d'amorces d'ADN permet à la technique LMPCR d'être très spécifique et plusieurs milliers de fois plus sensible que les autres techniques de séquençage génomique qui n'utilisent pas la puissance de l'amplification par PCR. Cette sensibilité accrue en permet l'utilisation sur les séquences uniques des génomes de mammifères tout en réduisant d'au moins 30 fois la quantité d'ADN requise (70).

La procédure LMPCR peut être divisée en plusieurs étapes (voir figure 1 du chapitre 2). Brièvement, un premier oligonucléotide spécifique à la séquence que l'on veut étudier se

fixe au gabarit complémentaire d'ADN suite à la dénaturation de cet ADN à haute température. Cet oligonucléotide sert d'amorce à la polymérisation qui est interrompue lorsque l'ADN polymérase atteint une cassure, produisant ainsi une molécule d'ADN bicaténaire avec une extrémité franche. Comme les sites de cassures monocaténaires sont aléatoires d'une molécule d'ADN à l'autre, les fragments d'ADN polymérisés seront de tailles variables et les séquences terminales du côté de l'extrémité franche (côté 5') ne seront pas les mêmes d'une molécule à l'autre. L'étape clé de la technique LMPCR est l'attachement d'un oligonucléotide bicaténaire asymétrique, un « adaptateur », à l'extrémité franche de toutes les molécules bicaténaires d'ADN. Cette procédure permet à toutes les molécules d'ADN polymérisées lors de l'étape précédente de posséder des séquences communes d'ADN aux deux extrémités, malgré le fait que les molécules soient de tailles différentes. La simple amplification par PCR de ces fragments est alors possible en utilisant une seconde amorce spécifique à la séquence étudiée et la plus longue amorce de l'adaptateur. Les molécules ainsi amplifiées correspondent fidèlement à la distribution et à la fréquence des cassures monocaténaires de l'ADN utilisé au départ. Les produits d'amplification par PCR sont séparés selon leur taille sur gel de polyacrylamide dénaturant et transférés sur membrane de nylon par électro-transfert. La membrane est alors hybridée avec une sonde d'ADN spécifique à la région d'intérêt et exposée sur film. Généralement, quatre échantillons d'ADN traités selon les réactions chimiques de Maxam et Gilbert (82) (pour les nucléotides G, T+C, C) et de Iverson (83) (pour les nucléotides A) sont simultanément amplifiés par la technique LMPCR de la même manière que les échantillons de notre étude. Sur gel polyacrylamide, ces quatre échantillons produiront une séquence, laquelle permettra de localiser sur la séquence analysée les cassures présentes au niveau des autres échantillons.

1.6.2 L'utilisation de la technologie LMPCR

En bref, la technologie LMPCR cartographie les cassures rares et en maintient la fréquence relative à travers toute la procédure. Combinée avec des agents de caractérisation de

l'ADN, cette technologie peut être utilisée pour cartographier *in cellulo* les interactions ADN-protéines (84), les structures spéciales de l'ADN (77), les segments d'ADN monocaténaire (85), le positionnement des nucléosomes (77) et la distribution et la fréquence de dommages causés par un mutagène (86), pour étudier *in cellulo* la réparation de l'ADN au niveau de chaque nucléotide (87) et pour cartographier *in vitro* les cytosines méthylées (88).

1.6.3 La contribution de la technologie LMPCR

Depuis sa mise au point en 1989 (84, 89), nous avons recensé plus de 550 publications où la technologie LMPCR a été employée. Pour la grande majorité de ces publications, les auteurs ont utilisé la technologie LMPCR pour cartographier *in cellulo* les interactions ADN-protéines au niveau des promoteurs de plusieurs gènes. L'analyse intracellulaire de l'ADN a apporté une contribution majeure sinon essentielle dans la plupart de ces travaux. Les quelques exemples qui suivent sont tirés de la littérature et reflètent l'importance de l'information que l'on ne pourrait obtenir autrement qu'avec ce type d'analyse.

1.6.3.1 Le gène PGK-1

Le gène humain de la PhosphoGlycérate Kinase 1 (*pgk-1*) est localisé sur le chromosome X et il est impliqué dans la glycolyse anaérobique. En utilisant la technologie LMPCR, les auteurs ont identifié *in cellulo* des modifications dans le patron d'interactions ADN-protéines entre le promoteur du gène *pgk-1* situé sur le chromosome X actif et le chromosome X inactif (77, 90). Les auteurs ont clairement montré que le promoteur actif de *pgk-1* liait plusieurs facteurs de transcription alors que le promoteur du gène inactif ne montrait aucune évidence de liaisons ADN-protéines. Suite à l'analyse intracellulaire à la DNase I, les auteurs ont également observé des empreintes disposées selon une périodicité

de 10 pb tout le long du promoteur inactif, caractéristiques d'une région d'ADN organisée avec des nucléosomes (77).

1.6.3.2 Les gènes NOS, TAT et cdc2

Dans certains cas, la transcription peut être induite ou son niveau être augmenté suite à une ou plusieurs nouvelles interactions ADN-protéines *in cellulo*. Les gènes *nos* (*nitric oxide synthase*; l'oxyde nitrique synthase) de souris (91) et *tat* (tyrosine aminotransférase) de rat (92) en sont de bons exemples. Les auteurs ont déterminé que ces gènes étaient inductibles et que leur activation dépendait de la liaison d'activateurs au niveau du promoteur. Alternativement, la perte d'une interaction ADN-protéine peut également activer la transcription *in cellulo*, comme pour le gène *cdc2* (*cell division cycle 2*; cycle 2 de division cellulaire) humain (93). Dans ce dernier exemple, les auteurs ont montré que le complexe p130/E2F-4 était lié au promoteur de cdc2 et qu'il agissait comme répresseur tout au long du cycle cellulaire. La perte de la liaison du complexe à l'entrée de la phase S permettait la transcription du gène. Dans ces systèmes, l'analyse intracellulaire s'effectue à la fois sur des cellules stimulées par un agent inducteur (les cytokines, par exemple) et sur des cellules non-stimulées. L'action directe ou indirecte de l'agent sur le promoteur peut être détectée par un changement dans le patron des empreintes ADN-protéines quand les promoteurs des cellules stimulées et non-stimulées sont comparés.

1.6.3.3 Le gène c-jun

La cartographie des interactions ADN-protéines et des structures particulières de la chromatine ne permet pas toujours d'identifier les mécanismes d'activation d'un gène. En effet, d'autres mécanismes que l'action directe d'activateurs ou de répresseurs peuvent activer ou réprimer la transcription d'un gène. Dans ces cas, les interactions ADN-protéines

localisées au niveau des promoteurs des gènes actifs et inactifs sont les mêmes, malgré l'augmentation de la transcription engendrée par la stimulation. Par exemple, l'expression de l'oncogène cellulaire jun (*c-jun*), impliqué dans la réponse aux dommages à l'ADN, est rapidement et fortement induite suite à une irradiation aux UV. Cependant, l'analyse intracellulaire de 1 kilopaires de bases (kb) de séquence du promoteur de *c-jun* avec deux agents de caractérisation différents (DMS et DNase I) n'a révélé aucun changement dans le patron d'interactions ADN-protéines. Néanmoins, les auteurs ont proposé un modèle cohérent d'activation du gène compatible avec ces résultats. Dans ce modèle, les facteurs de transcription sont liés au promoteur de façon constitutive, mais ils sont peu actifs, ne permettant qu'un niveau basal de transcription (94, 95). L'irradiation des cellules entraînerait l'action d'un ou de plusieurs cofacteurs agissant en *cis* avec les facteurs de transcription déjà présents sur le promoteur. Les facteurs de transcription, maintenant « activés », assurent l'augmentation rapide du niveau de transcription du gène.

1.6.4 Les limites de la technologie LMPCR

L'avantage majeur de l'analyse *in cellulo* est de pouvoir sonder l'ADN de gènes actifs à l'intérieur d'une cellule vivante. Ce que ces 550 études ont montré de façon répétée, c'est que par rapport au nombre de séquences consensus possibles, très peu de ces séquences étaient effectivement liées par leur facteur de transcription dans les cellules vivantes. Il devient donc très imprudent de conclure que certaines séquences d'ADN sont impliquées dans la transcription des gènes dans une cellule vivante en étudiant le promoteur uniquement avec des techniques *in vitro*. Voilà pourquoi, malgré qu'elles soient fastidieuses, les techniques d'analyse intracellulaires doivent être de plus en plus utilisées par les chercheurs pour comprendre les mécanismes de régulation de l'expression des gènes.

Malgré ses nombreux avantages, la technique LMPCR demeure peu populaire. Elle demande un investissement important en équipement et en temps et elle nécessite un personnel qualifié. De plus, comme toutes les techniques qui utilisent l'amplification par PCR, l'analyse de certaines séquences génomiques, surtout celles riches en Guanine et Cytosine, présentent parfois des difficultés. L'interprétation des résultats peut également s'avérer délicate, même pour les initiés. Il faut aussi garder en tête que la technologie LMPCR ne permet pas l'identification des facteurs protéiques dont elle révèle les empreintes, ni leur influence (positive ou négative) sur la transcription du gène cible (les techniques *in vitro* sont essentielles pour y arriver). En définitive, même si la technique LMPCR fournie des résultats d'une validité peu commune, elle exige des efforts importants pour des résultats qui ne sont que partiels.

1.7 Immuno-précipitation de la chromatine (ChIP)

La technique ChIP est l'une des plus puissantes techniques servant à l'identification de protéines associées à une séquence d'ADN génomique *in cellulo*. Brièvement, la méthode consiste à fixer les interactions ADN-protéines au sein des cellules vivantes à l'aide de la formaldéhyde (96). La formaldéhyde, qui est un agent bipolaire soluble aux membranes cellulaires, fixe non seulement les interactions ADN-protéines mais aussi celles protéine-protéines et ARN-protéines qui sont présentes au moment du traitement. Ceci prévient la dissociation et la redistribution des protéines au cours des manipulations de la chromatine subséquentes, permettant ainsi l'analyse de protéines n'ayant pas la capacité de lier elles-mêmes l'ADN mais faisant partie de complexes d'interaction avec l'ADN étant recrutées par d'autres protéines liant l'ADN. L'une des particularités importantes de la formaldéhyde est que les fixations qu'elle engendre sont totalement réversibles lorsque exposées à une température élevée (67°C), permettant ainsi l'analyse subséquente des protéines ou de l'ADN. Donc, suite à la fixation, la chromatine est isolée et brisée en petits segments par sonication. Après incubation avec un anticorps spécifique à la protéine d'intérêt (normalement un facteur de transcription), nous procédons à l'immuno-précipitation des

ADN liés (directement ou indirectement) à cette protéine. La fixation est ensuite renversée et l'ADN ainsi relâché est purifié pour une amplification PCR standard à l'aide d'amorces spécifiques à la région à l'étude. Il est aussi possible de procéder à l'identification des protéines co-précipitées par analyse Western. Enfin, la méthode ChIP peut être combinée avec la spectrométrie de masse pour l'étude à grande échelle des interactions protéines-protéines (97).

1.8 L'utilization d'anti-ARN

L'utilisation d'anti-ARN permet d'analyser les effets de l'absence ou de l'inhibition de l'expression d'un gène sans avoir à créer de lignées transformées. Il s'agit d'introduire dans la cellule de petites séquences d'ARN antisens spécifiques au gène d'intérêt. La liaison de ces ARN antisens à l'ARNm du gène inhibe de manière spécifique et mesurable sa traduction. En éliminant ou en diminuant ainsi la présence d'un facteur de transcription dans une population cellulaire, il est possible d'évaluer son influence sur la transcription d'un gène en mesurant l'ARNm ou la protéine. La transfection de ces anti-ARN peut être combinée à celle de plasmides contenant un gène rapporteur pour vérifier l'influence du FT sur l'expression de gènes rapporteurs.

1.9 La problématique et les objectifs

La transcription d'un gène est le premier niveau de régulation de la production d'une protéine. La connaissance des mécanismes qui contrôlent l'expression génique est essentielle pour comprendre plusieurs fonctions biologiques et pour traiter certaines pathologies. Parmi les éléments qui régulent la transcription, l'interaction de facteurs protéiques à leurs sites de liaison au niveau des promoteurs est l'un des plus déterminants.

La cartographie de ces séquences régulatrices et l'identification de leurs ligands sont essentielles à notre compréhension des mécanismes de régulation de la transcription des gènes. Afin d'y parvenir, de nombreuses approches peuvent être employées dont certaines permettent l'analyse de l'ADN génomique à l'intérieur d'une cellule vivante, alors qu'un gène donné est actif d'un point de vue de sa transcription. Les modèles issus de telles études sont à la fois plus crédibles et valides. Dans notre étude, nous avons utilisé la technologie LMPCR, permettant la cartographie des interactions ADN-protéines *in cellulo*, en combinaison avec d'autres techniques *in vitro* afin de mieux définir les mécanismes de régulation de gènes humains de classe II dont les patrons d'expression sont tout à fait uniques et distincts. Ceci nous a non seulement permis d'approfondir notre compréhension de la régulation de l'expression de chaque gène étudié, mais aussi de mieux comprendre l'utilisation de certains mécanismes de régulation plutôt que d'autres en rapport avec le patron d'expression des gènes.

Les chapitres 2 à 4 de cette thèse représentent autant de travaux dont la publication est acceptée ou qui seront soumis dans des périodiques scientifiques reconnus. Les résultats de ces articles concernent l'étude des mécanismes de régulation de l'expression des gènes humains $p21^{cip1/waf1}$, *GPC3* (*Glypican 3*) et *AβPP* (*Amyloïde β-Precursor Protein*).

1.9.1 Les objectifs spécifiques de l'article #1 (Chapitre 2)

1.9.1.1 La mise en situation

P21 fait partie de la famille des Cip/Kip, des inhibiteurs des kinases Cycline-dépendantes (CDK). Le gène *p21* est exprimé dans plusieurs types cellulaires et module une variété de fonctions biologiques comme la croissance cellulaire, la différentiation et l'apoptose. Via l'interaction avec les CDK, les cyclines et le PCNA, p21 est le médiateur du blocage du cycle cellulaire aux transitions G1/S ou G2/M, nécessaire à la réparation de l'ADN endommagé par divers agents. Il a été montré que p53 activait directement l'expression de *p21* par interaction directe au niveau du promoteur suite à l'exposition à des agents qui

endommagent l'ADN comme les radiations ionisantes ou ultraviolettes (UV). La forte activation de *p21* induite par l'interaction de p53 à son site de liaison 2,2 Kb en amont du SIT se produirait via les facteurs de transcription sp1 au niveau du promoteur proximal du gène. D'autre part, *p21* est exprimé à un niveau basal de façon constitutive et dépendante du cycle cellulaire; cette activité du gène est indépendante de p53. Il semble que ce niveau de base de p21 soit favorable à la poursuite du cycle cellulaire en condition de prolifération normale. En effet, dans les cellules en croissance, la majorité des protéines p21 sont retrouvées au sein des complexes cycline/CDK actifs. Il a été montré qu'à faible concentration, p21 favorise l'interaction entre la cycline D et CDK4 et que l'expression de p21 est induite suite à l'exposition de cellules à des signaux mitogènes. P21 semble donc avoir un effet antagoniste sur l'activité des complexes cycline/CDK et donc sur le cycle cellulaire selon son niveau d'expression.

1.9.1.2 Les objectifs

Or, si le mécanisme d'induction de *p21* par p53 est connu, la régulation de la transcription du gène indépendante du facteur de transcription p53 est mal définie. La région contenant les sites putatifs d'interaction pour le facteur de transcription E2F1 est essentielle à l'activité du promoteur (134, 135). Nous avons donc tenté de mieux comprendre les mécanismes de régulation de la transcription de *p21* en condition de prolifération normale. Les objectifs de cet article sont : 1) cartographier *in cellulo* les sites d'interaction ADN-protéines au niveau du promoteur proximal de *p21* dans des fibroblastes mâles normaux en culture primaire à l'aide de la technique LMPCR combinée au DMS, aux UV de type C (UVC) et de la DNase I, 2) identifier les facteurs de transcription en cause dans ces interactions par EMSA (en combinaison avec des compétiteurs et des anticorps) et par immuno-précipitation de la chromatine (CHIP), et finalement 3) définir le rôle de ces facteurs sur l'activité du promoteur de p21 en transfections transitoires dans plusieurs types cellulaires et par la transfection d'ARN antisens spécifiques.

1.9.2 Les objectifs spécifiques de l'article #2 (Chapitres 3)

1.9.2.1 La mise en situation

Le gène *Glypican 3* (*GPC3*) est exprimé dans toutes les tumeurs de Wilms (WT) et dans certains neuroblastomes (NB), deux tumeurs embryonnaires. Il est exprimé selon une spécificité tissulaire avec une intensité qui culmine lors de l'embryogenèse. Après la naissance, *GPC3* est réprimé dans tous les tissus investigués. Il est situé sur le chromosome Xq26.1 et s'étend sur plus de 500 kb. Le produit du gène est un HSP (heparan sulfate proteoglycan) localisé à la surface cellulaire et attaché à la membrane cellulaire par une ancre glycosyl-phosphatidyl inositol (GPI; (98)). Son rôle n'est pas encore connu. Selon certaines données, GPC3 serait un régulateur négatif de la croissance cellulaire (98-101). Une mutation germinale du gène cause le syndrome de surcroissance de Simpson-Golabi-Behmel (98) et la souris knock-out pour le gène *Gpc3* reproduit partiellement ce syndrome (99). À l'inverse, d'autres indices suggèrent un rôle opposé. En effet, *GPC3* est surexprimé dans les carcinomes hépatocellulaires et est associé aux stades avancés et potentiellement invasif de ce cancer (102-105). De plus, les métastases hépatiques associées aux carcinomes colorectaux expriment *GPC3* de façon significativement plus importante que les tumeurs primaires (106). Il semble donc que, dépendamment du contexte cellulaire, *GPC3* régule de façon divergente les facteurs de croissance et de survie cellulaire (107).

La compréhension des mécanismes qui régulent la transcription anarchique du gène *GPC3* dans les cellules cancéreuses est d'un intérêt particulier. Certains auteurs ont récemment montré que quelques lignées cellulaires et des échantillons primaires femelles de NB présentaient une certaine perte de méthylation dans la région promotrice du gène. Ceci suggère que la déméthylation de l'allèle de *GPC3* du X inactif pourrait permettre l'expression du gène dans ces NB (108). Cette hypothèse est supportée par le fait que la surexpression de *GPC3* est plus fréquente dans les carcinomes hépatocellulaires femelles (102). Toutefois, la méthylation n'est pas nécessaire à la répression de la transcription de *GPC3* dans les lignées de NB n'exprimant pas le gène, ce qui indique que les conditions

d'expression permissives du gène sont dues à un autre mécanisme de régulation comme l'interaction de facteurs de transcription au niveau du promoteur. Il a été montré que Sp1 est nécessaire à l'activation de la transcription de *GPC3* (109). Toutefois, étant exprimé dans environ toutes les cellules, ce facteur ne peut contrôler à lui seul la spécificité tissulaire de l'expression de *GPC3*.

1.9.2.2 Les objectifs

Afin d'élucider les mécanismes de régulation de l'expression de *GPC3*, nous avons cartographié le promoteur du gène dans deux lignées cellulaires mâles issues de NB : SFNB-7 qui exprime *GPC3*, et SK-N-FI, qui ne l'exprime pas. Les objectifs précis de cet ouvrage sont : 1) cartographier *in cellulo* les sites d'interaction ADN-protéines et les structures secondaires sur le promoteur proximal entier de *GPC3* dans les lignées cellulaires ci-haut mentionnées, 2) identifier les disparités entre les 2 lignées cellulaires quand aux sites d'empreinte protéique afin de proposer un ou des régulateurs, positifs ou négatifs, pouvant expliquer l'expression différentielle de *GPC3* dans ces cellules et 3) identifier ces facteurs de transcription. Nous nous sommes attardés à caractériser le promoteur du gène humain *GPC3* à l'aide de la technique LMPCR combinée au DMS, aux UV de type C (UVC) et de la DNase I. Par la suite, certaines séquences ayant révélées des empreintes protéiques ont été analysées par EMSA en combinaison avec des compétiteurs et des anticorps afin d'identifier les facteurs en cause dans ces interactions.

1.9.3 Les objectifs spécifiques de l'article #3 (Chapitres 4)

1.9.3.1 La mise en situation

Il y a présentement 20 millions de personnes atteintes de la maladie d'Alzheimer, une maladie neuro-dégénérative très invalidante, et selon les pronostics, ce nombre doublera d'ici 2030. Il n'y a aujourd'hui encore aucune thérapie pour cette maladie dont les causes sont toujours mal définies. L'une des caractéristiques majeures de cette pathologie est l'accumulation sous forme de dépôts insolubles du peptide-Aβ42 dans le tissu cérébral des patients. Dans la maladie d'Alzheimer, le rôle pathogène du peptide bêta-amyloïde est bien établi. En revanche, ce n'est qu'en 2005 que des chercheurs allemands ont mis le doigt sur un possible rôle physiologique de ce peptide et de sa protéine précurseure : la régulation du métabolisme du cholestérol et des sphingomyélines. Le peptide-Aβ42 active les sphingomyélinases, diminuant alors les taux de sphingomyéline tandis que le peptide-Aβ40, qui contient 40 acides aminés, diminue la synthèse *de novo* de cholestérol par inhibition d'un enzyme de synthèse. Par ailleurs, il existe un système de régulation maintenant un équilibre entre les lipides, la production de la protéine amyloïde précurseur et le peptide-Aβ. Une hypercholestérolémie, possible facteur de risque de la maladie d'Alzheimer, dérégulerait ce cycle et provoquerait un accroissement de la production du peptide-Aβ42.

Ce peptide est neurotoxique et joue très certainement un rôle central dans l'initiation et la progression de la maladie de l'Alzheimer. Le peptide-Aβ42 est en fait le résidu de deux clivages spécifiques de l'AβPP (amyloide precursor protein; protéine précurseure de l'amyloide) par les sécrétases α et γ. L'AβPP est codé par un seul gène, *AβPP*, situé sur le chromosome 21. Il est exprimé de façon spécifique dans certains types cellulaires au cours du développement et d'une manière ubiquitaire dans la plupart des tissus à l'âge adulte. Toutefois, son niveau de transcrit est significativement accru dans certaines régions cervicales des patients atteints de la maladie d'Alzheimer. De plus, les mêmes symptômes ont été observés mais de façon plus précoce, soit environ 40 ans plus tôt, chez les individus trisomiques pour le chromosome 21. Pourtant, malgré la présence de trois copies du gène *AβPP* dans les cellules de ces derniers au lieu de deux, la quantité de transcrits du gène est non pas 0,5 fois mais 5 fois plus élevée que les individus normaux. Toutes ces observations ont conduit à l'hypothèse que la perturbation de la régulation de l'expression du gène

pourrait être la (ou l'une des) cause(s) de la formation des dépôts insolubles du peptide-Aβ42.

1.9.3.2 Les objectifs

Le contrôle de l'expression du gène humain *AβPP* est encore aujourd'hui mal défini. Son promoteur s'étend sur 3300 paires de bases en amont du site d'initiation de la transcription. Il contient bon nombre de sites putatifs d'interactions ADN-protéines pouvant être impliqués dans la régulation de l'expression du gène. Toutefois, plusieurs études démontrent que la région −96 à +100 par rapport au site d'initiation de la transcription serait la seule déterminante de la régulation de l'expression du gène (110-112). Dans cette région, deux zones comprenant trois sites d'interactions putatifs pour des facteurs de transcription (CTCF, USF et Sp1) seraient essentiels à l'activité maximale du promoteur. Les objectifs de ce papier sont : 1) cartographier dans les cellules vivantes normales toutes les interactions ADN-protéines au niveau de la séquence promotrice -695 à +45 à l'aide du DMS et des UVC, 2) faire la lumière sur l'importance relative des facteurs CTCF, USF mais surtout Sp1 pour la régulation du gène, 3) mettre en relief les disparités possibles entre les neurones et les astrocytes.

Chapitre 2 :   **Transcriptional regulation of the cyclin-dependent kinase inhibitor 1A (p21) gene by NFI in proliferating human cells**

2.1 Le résumé (français)

L'inhibiteur des kinases cycline-dépendantes 1A (CDKN1A), aussi appelé p21 (WAF1*I*CIP1) est connu pour moduler le cycle cellulaire, l'apoptose, la sénescence et la différenciation via l'interaction spécifique avec les cyclines, les kinases qui dépendent de la cycline et plusieurs autres. La régulation de l'expression de p21 s'opère principalement au niveau de la transcription. Il est une cible majeure du gène suppresseur de tumeur p53 mais possède aussi un mode d'induction de la transcription indépendant. L'analyse du promoteur du gène à l'aide des techniques LMPCR et ChIP a révélé une empreinte protéique entre les positions −161 et −149 par rapport au site d'initiation de la transcription de p21, contenant une séquence putative de haute affinité pour les facteurs de transcription membres de la famille des facteurs nucléaires I (NFI). Les analyses de retard de migration sur gel combinées avec des anticorps ont confirmé NFI comme étant le facteur nucléaire liant cette région. L'utilisation de constructions du promoteur de p21 contenant des mutations au site de liaison pour NFI ont permis de définir ce dernier comme répresseur important de l'expression de *p21* dans plusieurs types cellulaires. L'inhibition de NFI endogène avec les ARN antisens dans les fibroblastes cutanés humains normaux entraîna une augmentation considérable de l'activité du promoteur de p21, démontrant ainsi que NFI est un répresseur clé de la transcription de p21 *in cellulo*. La sur-expression de chacune des 4 isoformes de NFI dans la lignée HCT116 a montré que chacune d'elles contribuait à une extension variable de la répression du gène p21. Plus précisément, la sur-expression de NFI-B dans les HCT116 en arrêt de croissance suite au traitement avec de la doxorubicine, augmentait la proportion de cellules en phase S du cycle cellulaire alors que NFI-A et NFI-X la réduissait, établissant ainsi le rôle de NFI dans la régulation qui dépend du cycle cellulaire de *p21*.

2.2 le manuscrit (anglais)

TRANSCRIPTIONAL REGULATION OF THE *CYCLIN-DEPENDENT KINASE INHIBITOR 1A* (*P21*) GENE BY NFI IN PROLIFERATING HUMAN CELLS

Stéphane Ouellet[1], François Vigneault[2], Maryse Lessard[2], Steeve Leclerc[2], Régen Drouin[1], Sylvain L. Guérin[2,3]

[1]Service of Genetics, Dept of Pediatrics, Université de Sherbrooke, Sherbrooke, Québec, Canada; [2]Oncology and Molecular Endocrinology Research Center and [3]Unit of ophthalmology, CHUL, Centre Hospitalier Universitaire de Québec and Laval University, Québec, Québec, Canada

Running title: Transcriptional repression of p21 gene expression by NFI.

Address Correspondence to:
    Dr Sylvain L. Guérin
    Oncology and Molecular Endocrinology Research Center, CHUL
    2705 Laurier Blvd, Ste-Foy (Québec)
    G1V 4G2, Canada
    Tel.: (418) 654-2296
    Fax: (418) 654-2761
    E-Mail: Sylvain.guerin@crchul.ulaval.ca

2.2.1 Abstract

The cyclin dependent kinase inhibitor 1A (CDKN1A), also known as p21 (WAF1/CIP1) modulates cell cycle, apoptosis, senescence and differentiation via specific protein-protein interactions with the cyclins, cyclin-dependent kinase (Cdk), and many others. Expression of the p21 gene is mainly regulated at the transcriptional level. By conducting both ligation-mediated polymerase chain reaction (LMPCR) and chromatin immunoprecipitation (ChIP) *in vivo*, we identified a functional target site for the transcription factor nuclear factor I (NFI) in the basal promoter from the p21 gene. Transfection of recombinant constructs bearing mutations in the p21 NFI site demonstrated that NFI acts as a repressor of p21 gene expression in various types of cultured cells. Inhibition of NFI in human skin fibroblasts through RNAi considerably increased p21 promoter activity suggesting that NFI is a key repressor of p21 transcription. Overexpression of each of the four NFI isoforms in HCT116 cells established that each of them contribute to various extend to the repression of the p21 gene. Most of all, overexpression of NFI-B in doxorubicin, growth-arrested HCT116 increased the proportion of cells in the S-phase of the cell cycle whereas NFI-A and NFI-X reduced it, thereby establishing a role for NFI in the cell-cycle dependent expression of p21.

2.2.2 Introduction

p21 (CDKN1A) belongs to a family of cell cycle dependent kinase inhibitors. p21 modulates various processes such as cell growth (1), differentiation (2) and apoptosis (3). DNA-damaging agents such as UV-irradiations or carcinogens induce the transcriptional activation of p21 via a p53-dependent pathway. Phosphorylated p53 binds to regulatory elements present in distal regions of the p21 promoter and transactivates p21 transcription via physical and functional interactions with the ubiquitous transcription factor Sp1 bound to the proximal region (4,5). Strong p21 accumulation lead to suppression of cdk activity, allowing the accumulation of hypophosphorylated Rb, inhibition of E2F-dependent transcriptional processes, and cell cycle arrest in G1 (6-8). By interacting directly with PCNA, an auxiliary factor for DNA polymerases δ and ε, p21 also prevents DNA synthesis and regulates DNA methylation without alteration of the PCNA-dependent nucleotide-excision repair (9,10). Some reports also suggest a p53 independent activation of p21 following UV-irradiation (11,12) but the mechanisms of induction remained poorly understood. Inducers of differentiation such as steroid hormones (13), transforming growth factor-β (TGF-β, phorbol ester or phosphatase inhibitors (14), interferon γ (15), progesterone (16), nerve growth factor (17) and platelet-derived growth factor (18) enhance p21 gene expression in many different cell systems. As for p53, the majority of these modulators affect p21 gene transcription via Sp1 and Sp3 interactions at the proximal promoter. Novel alternate p21 transcripts that are upregulated as a result of DNA damage-induced p53 activation and are dependent on p53 for their basal or induced expression were recently discovered (19,20).

In proliferating cells, p21 is expressed at a basal level in a constitutive and cell cycle dependant way (21). Under these conditions, most of the p21 proteins are components of the cyclin/cdk active complex (22). Therefore, it is likely that p21, when expressed at a moderate level, can act as an anchor protein as well as an assembly factor for cyclin D-cdk4/cdk6 thereby promoting their mutual interactions and consequently cell cycle

progression, which is in complete contrast to its function as a cdk inhibitor (23). However, there seems to be a lack of understanding on the regulatory mechanisms involved in this basal expression of the p21 gene. In this study, we provided evidence that NFI is a major contributor of p21 gene expression as it could repress its transcription by interacting with the p21 proximal promoter *in vivo*.

2.2.3 Materials and methods

2.2.3.1 Cell culture

Human skin fibroblasts (HSFs) were isolated from skin biopsies of healthy male donors and primary cultured in Dulbecco's modified Eagle's medium (DMEM, Gibco BRL, Burlington, ON, Canada) supplemented with 10% fetal bovine serum (FBS, Gibco BRL). Human epitheloid carcinoma HeLa cells (ATCC CCL 2) were grown in DMEM supplemented with 10% FBS. Rat pituitary GH4C1 cells were grown in Ham's F10 medium (Sigma-Aldrich, Oakville, ON, Canada) supplemented with 10% FBS. The human colorectal cancer cell line HCT116, kindly donated by Dr. Chantal Guillemette (Oncology and Molecular Endocrinology Research Center, CHUL, Québec, Canada), was grown in McCoy's medium (Invitrogen cat# 16600) supplemented with 10% FBS. All cells were grown under 5% CO_2 at 37°C and gentamycin was added to all media at a final concentration of 15 µg/ml.

2.2.3.2 Cell cycle analysis

HCT116 cells (2.5×10^5 cells/well) were plated into 6-wells tissue culture plates (Sarstedt, Montréal, QC, Canada) in complete McCoy's medium and then growth-arrested by the

addition of 500 nM doxorubicin (Sigma-Aldrich). Approximately 48 hours later, cells were collected, fixed with 70 % cold ethanol, and stored at –20 °C overnight. Fixed cells were washed with PBS and incubated with PBS containing RNase (100 µg/ml) and PI (10 µg/ml) at room temperature in the dark for 30 minutes. The DNA content of the cells was analyzed using flow cytometry and acquisition of data for 10^5 events was performed using an Epics XL flow cytometer (Beckman Coulter, Miami, FL). The distribution of HCT116 cells within each phase of the cell cycle was analyzed from dual parameters histograms using an Epics® Elite flow cytometry workstation version 4.5 software.

2.2.3.3 In vivo footprinting

HSFs were seeded into 150mm^2 culture plates at a density of 50% cells/plate and grown as above for 3 days. Living cells and purified DNA (referred as *in vivo* and *in vitro*, respectively) were treated with one of the following probing agents: 0,02% dimethylsulphate (DMS; Sigma–Aldrich Canada), ultraviolet C (UVC; 1500 J/m^2) irradiation (G15T8 germicidal lamp; Philips, Montreal, QC, Canada) and DNaseI (8,75 µg/µl; Worthington Biochemical, Lakewood, NJ, U.S.A.) as previously described (24,25). Strand break frequencies were estimated on an alkaline agarose gel. LMPCR was carried out as previously described (24,25) using the 2 primer sets (1A: 5'-TCTCTCACCTCCTCTGA-3', position +97 to +81; 1B: 5'CTGAGTGCCTCGGTGCC TCGGCG-3', position +85 to +62; and 2A: 5'-CCAGA TTTGTGGCTCAC-3', position -280 to –264; 2B: 5'CTCACTTCGGGGGAAATGTGTCCAGCG-3'; position -268 to –241), allowing the analysis of the proximal promoter of p21 from nt –220 to +40. Briefly, gene specific primers were annealed to the genomic fragments of varying sizes and then extended using cloned Pfu exo- DNA polymerase (Stratagene, LaJolla, CA) to produce double-strand blunt ends. An asymmetric double-strand linker (L25 : 5'GCGGTGACCCGGGAGATCTG-AATTC-3' and oligo L11: 5'GAATTCAGATC-3') was then ligated to the phosphorylated terminal end of each fragment, providing a common sequence on the 5'-end of all fragments. Using Pfu exo- DNA polymerase, a linker specific

primer was used for a single round of linear amplification, followed by PCR amplification using the appropriate primer sets in combination with the linker primer. All primer extensions and PCR amplifications were carried out on T gradient thermocycler from Biometra (Montreal Biotech inc. Kirkland, Canada) as described (24,25). The PCR-amplified fragments were phenol/chloroform extracted, ethanol precipitated and subjected to electrophoresis on 8% polyacrylamide, 7 M urea gels alongside a Maxam and Gilbert sequencing ladder, followed by electrotransfer to nylon membranes (Roche Diagnostics Corp., Laval, Canada), hybridization to a ^{32}P-labeled gene-specific probe and visualization by autoradiography on Kodak films (Amersham Biosciences, Baie d'Urfé, Canada).

2.2.3.4 Plasmid Constructs

The p21 promoter fragment spanning region −192 to +36 (p21-192) relative to the mRNA start site was produced by KpnI/BglII digestion of the plasmid p21 (-2300)-Luc containing the entire p21 promoter (kindly provided by Dr. Claude Labrie, Oncology and Molecular Endocrinology Research Center, CHUL Research Center, Québec, Canada). Synthetic oligomers were used to produce two distinct linkers allowing the ligation of the fragment upstream of the chloramphenicol acetyltransferase (CAT) reporter gene in the HindIII/XbaI-linearized vector pCATbasic (Promega, Madison, WI). The plasmid lacking the region -192 to −125 (p21-124) was obtained through double-digestion of the p21-192 construct with the KpnI and BstX1 restriction enzymes. The p21-192 mNFI construct that bear a mutated NFI binding site was produced by the polymerase chain reaction (PCR), using p21-192 as a template and the synthetic oligomers p21-NFImA/B (5'-GGACCGGCTGGCCTGCTAAAACTCGATTAGGCTCAGCTG-GCTCC-3')/ (5'-GGAGCCAGCTGAGCCTAATCGAGTTTTAGCAGGCCACCGGTCC-3'). PCR amplifications were performed using the QuickChange® Site-Directed Mutagenesis Kit (Stratagene) according to the manufacturer's specifications. The DNA insert from each recombinant plasmid was sequenced by chain-termination dideoxy sequencing (26) to confirm the mutations. The pCH-NFI-A1.1, pCH-NFI-B, pCH-NFI-C and pCH-NFI-X

expression plasmids that encode high levels of the NFI-A, -B, -C and –X NFI isoforms, as well as the empty vector pCH that was used to derive the NFI expressing plasmids have been kindly provided by Dr. Richard M. Gronostajski (Dept. of Biochemistry, SUNY at Buffalo, 140 Farber Hall, 3435 Main St., Buffalo, NY, 14214, USA).

2.2.3.5 Protein Extracts and Electrophoretic Mobility Shift Assays

Nuclear extracts from primary cultured HSFs deprived or not of serum for 48 to 72 hrs, or from rat pituitary GH4C1 and HeLa cells were prepared as described previously (27,28) and kept frozen at –80°C until use. The NFI-L-enriched rat liver carboxymethyl (CM)-Sepharose fraction has been described previously (29). The protein concentration was determined for each nuclear extract by the Bradford procedure and further validated by Coomassie blue staining of SDS-polyacrylamide fractionated nuclear protein. Double-stranded synthetic oligonucleotides corresponding to: i) the p21 promoter region located from -167 to -127 and bearing the target region for NFI (p21.1: 5'-GCCTGCTGGAACTCGGCCAGGCTCAGCTGGCTCCGCGCGG-3'), ii) the high affinity binding site for CTF/NFI (5'-GATCTTATTTTGGATTGAAGCCAAT-ATGAG-3'), or iii) the 80-bp HindIII-SmaI fragment from p21-192 were 5'-^{32}P-end-labeled as described (30) and used as probes in the EMSAs. Approximately $3x10^4$ cpm labeled DNA was incubated for 5 min at room temperature with nuclear extracts in the presence of poly(dI:dC) (Amersham) and 50 mM KCl in buffer D (10 mM HEPES, 10% v/v glycerol, 0.1 mM EDTA, 0.25 mM phenylmethylsulfonyl fluoride). DNA-protein complexes were separated by electrophoresis on native 4% or 6% polyacrylamide gels run in Tris/glycine buffer at 4°C, as described previously (30).

Competition experiments in EMSA were conducted as described above except that up to 500-fold molar excesses of the following unlabeled double-stranded oligonucleotides were added to the reaction mixture as unlabeled competitors: p21.1, or derivatives of p21.1 that

bear mutations (bold, small letters) into either the 5' (in p21.1m5': 5'-GCCTGCTaaAACTCGGCCAGGCTCAGCTGGCTCCGCGCTGG-3') or the 3' NFI half-site (in p21.1m3': 5'-GCCTGCTGGAACTCGattAGGCTCAGCTGGCTCCGCGCTGG-3'.), or mutated into both half-sites (in p21.1m5'3': 5'-GCCTGCTaaAACTCGattAGGC-TCAGCTGGCTCCGCGCTGG-3'), p21.2 (5'-GCCTGCTGGAACTCGGCCAGGCTC-3') and p21.3 (5'-CAGCTGGCTCCGC-GCTGG-3') which correspond to each half of the p21.1 sequence, CTF/NFI, AP1 (5'-GATCCCCGCGTTGAGTCATTCGCCTC-3') and Sp1 (5'-GATCATATCTGCGGGGCGGG-GCAGACACAG-3')). EMSA supershift experiments were conducted as above except that 3 µl of either a pre-immune serum, or a polyclonal antibody that can recognize all isoforms of NFI (sc-5567, Santa Cruz Biotechnology, Santa Cruz, CA, USA) was incubated with the proteins for 5 min prior to the addition of the probe.

2.2.3.6 Methylation Interference

The 83 bp SmaI-HindIII fragment from the p21-192 plasmid that covers p21 promoter sequences from −110 to −192 was 5' end-labeled at the -110 position (top strand) and partially methylated with DMS as described (31). Approximately 3×10^5 cpm methylated, labeled probe was incubated with 100 µg crude nuclear proteins from HeLa cells in the presence of 1 µg poly(dI:dC) in buffer D and DNA-protein complexes were separated by electrophoresis through a 6% native polyacrylamide gel in Tris-glycine buffer. Following electrophoresis, the DNA-protein complexes were visualized by autoradiography and isolated by electroelution as described (32). The isolated labeled DNA was then treated with piperidine (60) and further analyzed on a 6% sequencing gel.

2.2.3.7 DNaseI in vitro Footprinting

The 228 bp XbaI/KpnI DNA fragment spanning the p21 promoter sequence from position –192 to +36 was 5' end-labeled and used as a probe in DNaseI footprinting. DNaseI digestion was performed in DNaseI buffer (33) by incubating $4x10^4$ cpm labeled probe with 75 µg crude nuclear proteins from HSF cells. Further analysis of the digested products onto polyacrylamide sequencing gels was done as described previously (62).

2.2.3.8 SDS-PAGE and Western blot

One volume of sample buffer (6 M urea, 63 mM Tris [pH 6.8], 10% [vol/vol] glycerol, 1% SDS, 0.00125% [wt/vol] bromophenol blue, and 300 mM β-mercaptoethanol) was added to 15 µg proteins before they were size fractionated on a 10 or 12% SDS-polyacrylamide minigel and electro-transferred onto a nitrocellulose membrane o/n at 4°C. The blot was then washed once (5 min) with TS buffer (150 mM NaCl and 10 mM Tris-HCl [pH 7.4]) and once (30 min) in TSM buffer (TS buffer plus 5% [wt/vol] fat-free powdered milk and 0.1% Tween 20). Then, a 1:5000 dilution of the NFI antibody in TSM buffer was added to the membrane and incubation proceeded for at least 90 minutes at 22°C. The blot was then washed 4 times in TSM buffer and incubated an additional 90 minutes at 22°C in a 1:5000 dilution of a peroxidase-conjugated goat anti-rabbit immunoglobulin G (Jackson Immuno-research, Biocan Scientific, Mississauga, ON, Canada). The membrane was successively washed in TSM (twice, 5 min each) and TS (twice, 5 min each) before immunoreactive complexes were revealed using western blot chemiluminescence reagents (Renaissance, NEN Dupont, Boston, MA) and autoradiographed.

2.2.3.9 Transient Transfections and CAT Assay

Primary cultured HSFs and the established tissue culture cell lines HeLa and GH4C1 were all transiently transfected with the p21-CAT recombinant constructs (p21-2300, p21-192

and its mutated derivatives p21-192 mNFI, p21-124) by the calcium phosphate precipitation method as described (34), using 15 μg of the test plasmid and 5 μg of the human growth hormone (hGH) gene-encoding plasmid pXGH5 (35). CAT activities were measured as described previously (36) and normalized to hGH secreted into the culture medium, which was assayed using a radioimmunoassay kit (Medicorp, Montréal, Québec, Canada). Each single value is expressed as 100 x (% CAT in 4h)/100 μg protein/ng hGH. To be considered significant, each individual value needed to be at least 3-fold the value of the background level caused by the reaction buffer used (usually corresponding to 0.15% chloramphenicol conversion). To compare the groups, Student's t-test was performed. Differences were considered to be statistically significant at $P < 0.05$. All data are expressed as mean ± SD.

2.2.3.10 Chromatin Immunoprecipitation Assays

The ChIP analyses were conducted as previously described (37). Briefly, HSFs were grown on 150-mm tissue culture dishes to ~75% confluence in complete DMEM, or in FBS-depleted DMEM for periods of time ranging from 48 to 72 hrs. Cells were then cross-linked with 1% formaldehyde for 10 min, harvested, and chromatin immunoprecipitated with 1 μg polyclonal antibodies directed against the Sp1 (sc-59; Santa Cruz Biotechnology), Sp3 (sc-644; Santa Cruz Biotechnology), NFI and E2F1 (sc-193x; Santa Cruz Biotechnology) transcription factors (Santa Cruz Biotechnology, Inc. Santa Cruz, CA). The resulting DNA was analyzed by PCR using a pair of primers (p21-279U: 5'CAGATTTGTGGCTC-ACTTCGTGG-3' and p21+223L: 5'AGAAACA-CCTGTGAACGCAGCAC-3') spanning the entire p21 gene promoter area from nt –279 to +223. As a negative control, each ChIP sample was also subjected to PCR using primers (p21-2587U: 5'AATTCCTCTGAAAGCTGACTGCC3' and p21-2132L: 5'AGGTTTACCTGGGGTCTTTA-GA-3') specific to a region located ~2 Kbp upstream from the p21 promoter.

2.2.3.11 RNAi assays

Silencer™ negative control (id #4611, #4613, #4615), and Silencer™ pre-designed siRNA duplexes against NFIA (id #115686, #115687, #144076), NFIB (id #115688, #115689, #115690), NFIC (id #215174, #215173) and NFIX (id #115298, #115297, #3296) were purchased from Ambion Inc. (Austin, TX, U.S.A.) and used according to manufacturer's specifications. Briefly, 2.5µg of all three siRNAs directed against NFIA mRNA were combined and transfected in triplicate into HSFs cultured to sub-confluence onto 39-mm tissue culture dishes ($1x10^6$ cells per dish at start) by the calcium phosphate precipitation method along with 10 µg of the appropriate plasmid construct (p21-124, p21-192, p21-192-mNFI, p21-2300 and p21-2300-mNFI). Cells were harvested 48 h following the addition of fresh medium and processed as mentioned above for the CAT assay.

2.2.4 Results

2.2.4.1 In vivo footprinting of the p21 gene promoter by LMPCR

In an attempt to identify the transcription factors that regulate p21 gene transcription, we performed in vivo footprinting analyses of the proximal promoter region from the p21 gene in normal, proliferating primary cultures of human skin fibroblasts (HSFs) by exploiting the LMPCR procedure (24). Footprint analyses were performed on the p21 promoter region covering both strands from position −220 to +40 relative to the p21 mRNA start site. Many protected and hyper-reactive sites were detected along the area from the p21 promoter that bear a putative NFI target site, on both the top and bottom strands. As shown on Figure 1A to 1C (and also summarized in Fig. 1D), the comparison of the band intensity from the in

vivo and in vitro treatments revealed many protected and hypersensitive nucleotides. Consistent with studies that reported binding of TBP to the TATA-box of other active gene promoters in different human cells (38,39), we also observed protected residues along the TATA-box of the p21 gene (from nt −28 to −14) as shown on Figure 1B (and summarized on Fig. 1D). The DNaseI protection extended many bp beyond the 3' end of the TATA-box that is typical of the association of other members of the holoenzyme with TBP. In addition, four of the six putative Sp1 binding sites reported to be required for p21 gene expression under basal and inducible conditions ((4,40); see discussion for details) were indeed protected from methylation by DMS. Two additional p21 promoter regions that revealed DNA-protein interactions were also identified by LMPCR. The first corresponds to the DNA target site for the transcription factor E2F1 (−84 to −80) that has already been shown to up-regulate transcription of the p21 gene ((41,42); see discussion for details). The second protected area extends from position −161 to −138 and encompasses a potential binding site for NFI (−161 to −149). This large DNA-protein interaction interface also extended 10 bp beyond the 3' end of the NFI site, from −148 to −138. However, search using computer databases for the identification of transcription factors target sites could not suggest any known putative transcription factor that could interact with this specific DNA segment from the p21 promoter.

2.2.4.2 Proteins from the NFI family bind the −161/−149 sequence of the p21 promoter

Both competition and supershift experiments in EMSA were conducted to demonstrate whether NFI possesses the ability to bind the −161/-149 region from the p21 gene, (Fig. 2). Incubation of a labeled probe (p21.1) bearing the entire footprinted region from positions −167 to −127 (Fig. 2A), with nuclear proteins from primary cultured HSFs (Fig. 2B) yielded a diffuse pattern of DNA-protein complexes (lane 2) identical to that reported previously for NFI (43-45). The formation of these DNA-protein complexes was almost completely abolished by the addition of a 150-fold molar excess of an unlabeled competitor oligonucleotide bearing the consensus sequence for CTF/NFI (lane 8). Formation of these

complexes was also prevented when either the p21.1 (lane 3) or the p21.2 double-stranded oligonucleotides (lane 4) which both bear the NFI binding site from the p21 promoter (Fig. 2A), were used as competitors (Fig. 2B). However, the p21.3 oligonucleotide, which is deleted from the NFI binding site (Fig. 2A), was unable to compete for the formation of the p21.1/protein complexes (data not shown). This result suggest that the -127 to -144 p21 sequence, which bear the protected residues identified immediately 3' from the NFI site in the p21 promoter, is not required for NFI binding and is likely recognized by a yet unknown nuclear protein. Derivatives of the p21.1 oligonucleotide that bear mutations either in the 5' (in p21.1m5') or the 3' (in p21.1m3') half-site of NFI could still compete, although not as efficiently as the wild-type p21.1 oligomer, for NFI binding to the labeled probe (28% reduction with p21.1m5' (lane 5) and 43% with p21.1m3' (lane 6) as evaluated using Biorad Quantity One software) indicating that NFI possesses the ability to bind to each single half-site from the p21 NFI site *in vitro*. The oligomer bearing both mutated sites (p21.1m5'3') was unable to compete for the formation of this complex in EMSA (lane 7). Competitors bearing the target sequences for the unrelated transcription factors AP1 (lane 9) and Sp1 (lane 10) were unable to compete for the formation of the shifted complexes yielded by the HSFs nuclear extract, a further evidence that the complexes detected indeed result from the recognition of the p21.1 labeled probe by NFI.

To validate the identity of NFI as the transcription factor binding the p21.1 labeled probe, supershift experiments were conducted using a polyclonal Ab directed against an N-terminal epitope shared by all members of the NFI family. The NFI Ab recognized at least one of the proteins binding the p21.1 probe and led to the formation of a larger supershifted complex (SC in Fig. 2B, lane 12). This new complex did not result from the non-specific recognition of the labeled probe by the NFI Ab or by an unrelated serum protein as neither condition (Fig. 2B, lanes 13 and 14) could yield the supershifted complex seen when the NFI Ab is incubated with the p21.1 probe in the presence of nuclear proteins (Fig. 2B, lane 12). Altogether, these results suggest that NFI binds the p21 promoter through its consensus sequence located from positions −161 to −149 in primary cultured HSFs.

The NFI target site from the p21 promoter differs from the prototypical sequence by 3 nucleotides over the entire 15 bp that constitutes this site (Fig. 2A). Recently, Roulet et al. (46) conducted a very elegant study in which they functionally validated a binding site predictor that could predict the affinity of a degenerated NFI site relative to that of the prototypical NFI sequence. According to their mathematical model, NFI is expected to bind the p21 NFI target site with an affinity of over 91% that normally yielded when it binds the high affinity NFI prototypical (consensus) sequence (which, in this case, is considered to correspond to an affinity of 100%). We therefore assessed the affinity of NFI toward both the p21 and the prototypical NFI target sites by incubating crude nuclear proteins from HSFs with the p21.1 labeled probe in the presence of varying concentrations (5- to 500-fold molar excesses) of either unlabeled p21.1 or the NFI consensus oligomer. As shown on Figure 2C, the unlabeled p21.1 oligomer was as efficient as the prototypical NFI-bearing oligonucleotide in competing for the formation of the NFI complex in EMSA. This result provided evidence that NFI binds to the p21 degenerated NFI site with an affinity undistinguishable from that obtained with the prototypical NFI site.

Recognition of DNA target sites by transcription factors might on occasion differ depending on whether the labeled probe used in the EMSA is synthetic (as the double-stranded oligonucleotide p21.1 used above) or derived from the gene's regulatory sequence. To further validate the results from the EMSAs conducted using the p21.1 oligonucleotide, EMSA experiments were repeated by using an 83 bp genomic fragment obtained from the SmaI-HindIII digestion of the p21-192 plasmid and covering the p21 promoter sequence from positions −110 to −192 as labeled probe. For this specific purpose, and to increase the sensitivity of the assay, we used a carboxymethyl (CM)-Sepharose enriched preparation of rat liver NFI-L (29) as the source of NFI protein. As shown on Figure 3A (lane 2), NFI-L indeed bound to the −110/-192 p21 labeled probe very efficiently. Specificity for the formation of the NFI complex was further demonstrated by both competition experiments in EMSA (only the NFI but not the unrelated Sp1 competitor oligonucleotide could compete for the formation of the shifted complex; Fig. 3A, lanes 3 and 4, respectively) and supershift experiments (Fig. 3A; addition of the NFI Ab (lane 6) supershifted the NFI

complex from its normal position on the gel (NFI) to that corresponding to a complex with low electrophoretic mobility in gel (SC)).

We next performed both dimethylsulfate (DMS) methylation interference assays and DNaseI footprinting *in vitro* in order to more precisely define the DNA target site bound by NFI along the −110 to −192 p21 promoter fragment. As shown in figure 3B, methylated G residues that interfere with binding of NFI to its target site in the p21 promoter *in vitro* also perfectly mapped within the site protected by NFI and defined through the use of the *in vivo* footprinting procedures (see Fig. 1). *In vitro* DNaseI footprinting identified only a short stretch of protected sequence (approximately 3 bp) that is contained within the NFI site identified *in vitro* by DMS footprinting and *in vivo* by LMPCR. The lack of DNaseI break points in the binding site of NFI, as revealed by the absence of bands on the autoradiogram, did not allow the determination of the entire sequence protected by the protein. This *in vitro* result contrasts with the *in vivo* footprinting data for this particular p21 promoter region as a larger protection could be observed (extending over approximately 14 bp; see Fig. 1). We therefore conclude that NFI indeed possesses the ability to bind efficiently to the p21 basal promoter both *in vivo* and *in vitro* by interacting with a target site located between positions −161 to −149.

2.2.4.3 NFI functions as a repressor of p21 promoter activity and triggers G2-arrested cells into the S phase of the cell cycle

In order to assess the regulatory influence exerted by the p21 NFI site, DNA fragments spanning the p21 basal promoter and extending from 5' position -192 up to either 3' positions +36 (which include the NFI site) or −124 to +36 (in which the NFI site has been deleted) was inserted upstream the CAT reporter gene (Fig. 4A) and transiently transfected into either primary cultured HSFs or tissue culture cells (HeLa and GH4C1). Deletion of the entire NFI site identified by *in vivo* footprinting (in p21-124), yielded a significant

increase in the activity of the CAT reporter gene upon transfection of the three different types of cells used in this study (3.1-, 2.9-, and 2.3-fold increase in CAT activity relative to the level directed by the p21-192 plasmid in HSFs, HeLa, and GH_4C_1 cells, respectively; Fig. 4B). In order to prevent undesirable regulatory effects that might have resulted from the deletion of sequences nearby the p21 NFI site in the p21-124 construct, the NFI site from the parental plasmid p21-192 was also altered through site-directed mutagenesis. The five residues identified through both DMS methylation interference and LM-PCR footprinting as the most important for NFI binding out of the 15 bp that constitute the p21 NFI site were changed for As and Ts (the wild-type NFI site 5'-CTGGAACTCGGCCAG-3' was changed to 5'-CTaaAACTCGattAG-3') to create the NFI mutated construct p21-192 mNFI. Consistent with the data from the deletion analysis, mutation of the p21 NFI site in p21-192 mNFI also yielded a substantial increase in CAT activity relative to the unmutated, wild-type NFI bearing construct p21-192 upon transfection of all cell types (8,5-, 2.3-, and 3.1-fold increase in HSFs, HeLa, and GH4C1 cells, respectively), further validating the results from the deletional analysis (Fig. 4B).

To further demonstrate the negative regulatory influence of NFI on the p21 promoter, p21-192 was co-transfected along with expression plasmids encoding high level of expression of each of the NFI isoforms into HCT116 cells. This cell line proved to be particularly interesting as it expresses very little NFI protein in Western blot analyses, unlike Sp1, which is abundantly expressed in these cells (Fig. 5A). Because HCT116 cells can easily be growth-arrested in the G2 phase of the cell cycle by the addition of doxorubucin, p21-192 was therefore transfected either alone or in the presence of each of the NFI expression plasmid (NFI-A, -B, -C and –X) in growth-arrested HCT116 cells. Part of the transfected HCT116 cells were used for the measurement of the CAT activity while the remaining of the cells were used to monitor cell-cycle progression of the doxo-growth arrested cells by FACS analyses. Transfection of HCT116 cells with the NFI expression plasmids revealed that all four NFI isoforms could reduce the activity of the p21-192 construct (up to 3.5-fold repression with NFI-X; Fig. 5B). Furthermore, co-transfection of p21-192 along with the combination of NFI-B, -C, and –X brought down p21 promoter function further to a 5-fold

repression. Culturing HCT116 cells in the presence of doxorubicin indeed arrested the cells primarily into the G2-phase of the cell cycle, with approximately 3% of the cells in the S-phase (compare 1st and 2nd columns on Fig. 5C) but did not affect the negative regulatory influence of the NFI isoforms on the activity of the p21 promoter (results not presented). Interestingly, while over-expressing NFI-A, -C and -X substantially reduced the proportion of the cells committed into growth-arrest (down to 1% with NFI-A), over-expression of NFI-B considerably increased those in the S-phase (up to approximately 9%). Transfection of doxorubicin-treated HCT116 cells with the combination of the NFI-B, -C and –X dramatically reduced the cells remaining in S-phase from 2.7±2.0 to 0.4±0.7.

2.2.4.4 Serum starvation alters the NFI-mediated repression of p21 gene transcription

P21 being cell cycle regulated, we investigated the NFI influence on p21 gene transcription upon transfection of both the wild-type p21-192 construct or its NFI-mutated derivative p21-192 mNFI into primary HSFs grown in serum-free medium, a condition under which p21 is known to be highly expressed (47-49). As expected, basal p21 promoter activity increased considerably upon serum starvation (about 3-fold for p21-192 and 4-fold for p21-192 mNFI; Fig. 6A). Under the experimental conditions used to conduct this transfection, which differ slightly from those used above in Figure 4 (see Experimental Procedures), mutation of the NFI site yielded an 11-fold increase in CAT activity when HSFs were maintained in complete medium. However, p21 promoter activity increased further up to 17-fold in serum-starved HSFs. This increase in p21 promoter activity when cells are maintained under serum-free condition correlated with a significant decrease in NFI binding in EMSA (Fig. 6B) when fibroblasts are deprived of serum for various periods of time (24, 48, and 72 h), despite that no alteration in the absolute amount of the NFI protein is observed in Western blot (Fig. 6C). Indeed, while extracts from HSFs grown in the presence of serum yielded a very clear and intense NFI-labeled probe DNA-protein complex in the EMSA (Fig.6B, lane 2), culturing them under serum starvation for various periods of time (48, 60, 66 and 72 hrs) resulted in a more diffused, strongly reduced NFI

signal (Fig. 6B, compare lanes 4, 6, 8 and 10 with lane 2). This reduction in NFI binding could not be accounted for from the degradation of NFI as no significant quantitative nor qualitative changes are observed in the total amount of NFI as revealed by Western blot (Fig. 6C). Besides, maintaining HSFs under serum free condition up till 72 hours did not committed the cells into apoptosis as poly(ADP-ribose) polymerase-1 (hPARP-1), a 113-kDa DNA repair enzyme whose cleavage by caspase-3 into degradation products of 89- and 22-kDa (of which the larger is efficiently recognized by the C-2-10 mAb) is recognized as an early marker of apoptosis (50,51), remained entirely intact during serum starvation (Fig. 6D). This is further validated by the fact that proteins in the high molecular range showed no evidence of degradation upon Coomassie blue staining on gel (Fig. 6E). In addition, HSFs exhibited a normal cell morphology under phase contrast microscopy with no sign of apoptosis, even when the cells were maintained for 72h without serum (Fig. 6F).

2.2.4.5 NFI binds the p21 promoter in vivo in dividing but not serum starved HSFs

The binding of NFI to the p21 gene promoter area was further examined *in vivo* in HSFs by ChIP assays. As a control, ChIP was also conducted on other transcription factors, such as Sp1, Sp3 and E2F1, also reported to bind the p21 basal promoter. Primers that spanned the entire p21 promoter from -279 down to +223 relative to the p21 mRNA start site were selected for the assay. Antibodies against Sp1, Sp3, and NFI all enriched the p21 promoter sequences in HSFs grown in complete, serum-containing medium, indicating that this genomic area is bound *in vivo* by these transcription factors (Fig. 7A). In contrast, only the Sp3 and E2F1 antibodies enriched the p21 promoter in HSFs maintained under serum starvation for 72h, relative to the "no antibody" control sample. These results suggest that Sp1, Sp3 and NFI are bound to cis-acting elements present on the p21 promoter in the *in vivo* chromatin configuration of HSF cells during the dividing phase of the cell cycle (basal level of p21). Upon serum starvation (where high levels of p21 expression are observed), binding of both Sp1 and NFI were completely abrogated. Interestingly, while the level of Sp3 binding to the p21 promoter remained unaffected by serum starvation, that of E2F1,

initially undetectable in actively growing cells, was found to bind weakly in serum-deprived HSFs. The significant occupancy of NFI in relation to the 'input' (Fig. 7B), suggest a critical role for this transcription factor on the p21 gene regulation. These results demonstrate that NFI binds the p21 promoter area *in vivo* in actively growing, but not serum-starved, growth-arrested HSFs.

2.2.4.6 Suppression of NFI expression by RNAi alleviates p21 repression

Evidence that NFI functionally represses p21 promoter activity was further examined through the suppression of the endogenous NFI transcript by RNAi (Fig. 7C). Because HeLa cells are much easier to culture and are transfected with a higher efficiency, they were selected over HSFs for conducting the RNAi assays. Recombinant constructs bearing various lengths from the p21 promoter (up to 5' positions -124, -192 and -2300) were co-transfected with and without a combined pool of siRNAs directed against NFIA, NFIB, NFIC and NFIX. As expected, the P21-124 construct lacking the NFI binding site yielded a promoter activity higher than the NFI-bearing constructs p21-192 and p21-2300, which are both under the negative regulatory influence of NFI. Co-transfection of these constructs along with the NFI siRNAs substantially released repression by NFI for both p21-192 (3-fold increase) and p21-2300 (near 7-fold increase). Again, mutating the NFI binding site in p21-192 (p21-192 mNFI) strongly increased CAT activity in HeLa cells (by approximately 8-fold). However, while inhibition of endogenous NFI through RNAi considerably increased the activity of the unmutated, wild-type p21-192 construct (3-fold increase) it had no such influence when the NFI site is mutated in p21-192 mNFI (which yielded a weak 27% reduction in p21 promoter activity).

2.2.5 Discussion

Cyclin-dependent kinase inhibitor p21 plays a key function in cell cycle arrest at the G1/S checkpoint in response to DNA damage, and is involved in the assembly of active cyclin-kinase complexes especially cyclin D-Cdk4/6. Transcription of the p21 gene relies on the control of numerous regulators of which many still have to be identified. In the present study, we precisely mapped, through both *in vivo* (LMPCR) and *in vitro* (DMS and DNaseI footprinting) approaches, a target site for the transcription factor NFI in the basal promoter of the human p21 gene that proved to function as a powerful repressor of p21 gene transcription in both non-damaged and proliferating primary cultures of human fibroblasts and established tissue culture cells.

To our knowledge, no evidence regarding a regulatory function exerted by NFI on the transcription of the p21 gene has ever been published until now. We demonstrated that members of the NFI family could bind a target site on the p21 basal promoter located between positions–161 and –149, and considerably repress its transcription in a variety of cells. Although each of the four NFI isoforms were found to contribute to this repression, NFI-X turned out to be the most effective. Intriguingly, each member from this family exerted very specific influences on the growth properties of the cells by either restricting or promoting progression of the cells into the S-phase from the cell cycle. Indeed, over-expression of NFI-A, -C and –X reduced the number of cells remaining in the S-phase in HTC116 cells that have been growth-arrested with doxorubicin, whereas over-expressing the NFI-B isoform triggered more cells to enter the S-phase, thereby establishing the biological significance for the regulatory influence of NFI on cell-cycle progression. Consistent with these results, Luciakova et al. have recently demonstrated that mutation of the NFI sites present in the promoter of the adenine nucleotide translocase-2 gene (ANT2) totally annihilated the growth-arrest properties of NFI in NIH3T3 cells (52). Most interestingly, they attributed these growth-arrest properties of NFI to the NFI-A, -C and –X isoforms but not NFI-B as the former were all found to repress expression of an ANT2-

driven reporter gene construct, whereas NFI-B activated it. Members from the NFI family can either repress or activate their target genes depending on the cell type or promoter context. This means that although all NFI isoforms were found to repress to varying degrees the transcription directed by the p21 promoter in HCT116 cells, yet they may activate transcription of other genes in the same cell that may or may not be related to cell growth. For instance, NFI has been shown to activate expression of p53 (53) and *gadd153* (54), both of which play key roles in growth arrest of damaged or environmentally stressed cells. Paradoxically, NF1-X1 was identified as one of three genes (together with c-*myc* and *MDM2*) that prevent TGFβ-induced growth arrest (55). Therefore, while all NFI isoforms clearly repress p21 gene expression in HCT116 cells, NFI-B might as well regulate expression of additional, cell-growth related genes that are not under the influence of the other NFI isoforms. This differential property of NFI-B may rely on its ability to recruit either co-repressors or co-activators that are distinct from those recruited by the other NFI proteins. For instance, of all the NFI isoforms, only NFI-B has been found to activate whey acidic protein (WAP) gene expression in cooperation with STAT5A and the glucocorticoid receptor (GR) in JEG-3 cells (56). Most interestingly, NFI-B deficient mice have been shown to die early after birth and display severe lung hypoplasia (57), lung development being arrested at the late pseudoglandular stage. These results suggest that NFI-B is most likely required for normal lung cell proliferation in order to ensure proper lung development. It is noteworthy that the *in vivo* footprint identified in the present study extends 10 bp beyond the 3' end of the NFI consensus site (from −148 to −138). A search conducted using databases for the identification of transcription factor target sites failed to identify any known nuclear protein that may interact with the −148/−138 p21 sequence. The presence of additional co-factors that can bind DNA *in vivo* but not *in vitro*, as well as structural events, such as DNA folding, could explain the detection of the protected residues located downstream from the −161/−149 NFI site. Indeed, the RNAi experiments conducted in this study (Fig. 7C) suggested that other DNA binding proteins might contribute to the appropriate transcription of the p21 gene by interacting somewhere along the -124/-148 segment from the p21 promoter as the p21-124 construct did not mimic levels achieved with the p21-192 that bears mutations in the -161/-149 NFI site. Besides, reducing the endogenous level of NFI through RNAi is also likely to considerably alter the

transcription of many genes in the transfected cells. Some of them may as well encode transcription factors other than NFI that also participate to the cascade of regulatory events required to ensure appropriate transcription of the p21 gene. Further works will be required in order to determine the precise identity of such cis-acting transcriptional regulators and whether their regulatory influences are mediated through their interaction with the -148/-138 area from the p21 promoter.

The NFI family of transcription factors is made-up of four protein subtypes (NFI-A, -B, -C et –X) (58), each encoded by a different gene, that can form either homo- or heterodimers. The large diversity of this family of proteins, for which 19 isoforms have been described to date (reviewed in ref. (59)), rely on the fact that each individual NFI mRNA transcript is subjected to alternative splicing (60,61). Members from the NFI family have been shown to function either as repressors (44,54,62) or activators (63,64) of gene transcription. Through interactions with weak binding sites, NFI may also regulate gene expression by its ability to either cooperate or compete with many transcription factors (62,65). Considering the negative influence exerted by NFI on p21 gene transcription, we propose that NFI may act in cooperation with other nearby co-activators or transcription factors to restrict the expression of the p21 gene throughout the cell cycle progression while DNA is undamaged. NFI was reported to interact with and antagonize Sp1, which result in the down-regulation of the platelet-derived growth factor (PDGF)-A gene expression (65). Interestingly, *in vivo* footprinting analysis of the p21 promoter also revealed DNA-protein interactions at 4 of the 6 putative consensus binding sites reported for Sp1 (4). Besides NFI and Sp1, these *in vivo* analyses also identified protections at target sites for nuclear proteins, such as E2F1 (Fig. 1) that have already been reported to bind the p21 proximal promoter. Members from the E2F family of transcription factors play a crucial role in the transactivation of G1/S transition specific genes (66). Previous studies already demonstrated the involvement of E2F1 in the p53-independent activation of p21 (41,42).

Results from the ChIP analyses revealed variations in the pattern of p21 promoter occupancy by the transcription factors Sp1, NFI and E2F1 between actively growing and serum-starved, quiescent HSFs. Indeed, whereas both Sp1 and NFI but not E2F1 substantially bind the p21 promoter in proliferating cells, HSFs grown under serum deprivation completely lost p21 promoter recognition by both Sp1 and NFI but gained binding of E2F1. The relative occupancy of the p21 promoter by NFI *in vivo* clearly exceeded that of both Sp1 and Sp3 in proliferating HSFs, thereby favoring repression rather than activation of p21 transcription. On the contrary, the lack of any NFI binding in serum-starved cells combined to the recognized positive influence of both Sp3 and E2F1 as both factors were shown to bind the p21 promoter under this condition, is consistent with the increased expression of p21 reported when cells progress toward growth arrest (42,67).

Interestingly the presence of Sp3, whose binding to the p21 promoter does not change with alterations in the proliferative state of the cells, raised the interesting possibility that most of the Sp1 target sites identified in the p21 promoter might became occupied by Sp3 under serum deprivation as both transcription factors have been reported to bind the same GC-rich target sites (67). It is noteworthy that Sp1/Sp3 share extensive structural and sequence homology (68), and often cooperate in activating gene transcription (69-71). The presence of Sp3 in HSFs grown with or without serum suggests that p21 is constitutively activated by this transcription factor and that fine tuning of p21 transcription is ensured by subtle variations in the ability of nuclear repressors such as NFI to interact with its corresponding target site in the p21 promoter. Our results are consistent with those published by other groups who demonstrated that while basal transcription of p21 is regulated by both Sp1 and Sp3 in cultured primary mouse keratinocytes, only Sp3 contributes to the calcium-induced p21 promoter activity (67).

The identification of each of the nuclear proteins yielding the new protected sites identified in this study as well as the precise regulatory function they play will be required to assess the precise mechanisms that modulate p21-dependent regulation of cell cycle progression.

We must now look upon NFI as a new and central player in the molecular mechanisms regulating p21 expression during cell cycle progression. The relationship between NFI and other transcription factors such as p53, Sp1, Sp3 and E2F1 that also bind the p21 promoter will be of a particular interest and should prove to be particularly fascinating regarding p21 gene regulation.

2.2.6 Acknowlegements

2.2.7 References

1.   Gartel, A.L., Serfas, M.S. and Tyner, A.L. (1996) p21--negative regulator of the cell cycle. *Proc Soc Exp Biol Med*, **213**, 138-149.
2.   Wu, H., Wade, M., Krall, L., Grisham, J., Xiong, Y. and Van Dyke, T. (1996) Targeted in vivo expression of the cyclin-dependent kinase inhibitor p21 halts hepatocyte cell-cycle progression, postnatal liver development and regeneration. *Genes Dev*, **10**, 245-260.
3.   Ostrovsky, O. and Bengal, E. (2003) The mitogen-activated protein kinase cascade promotes myoblast cell survival by stabilizing the cyclin-dependent kinase inhibitor, p21WAF1 protein. *J Biol Chem*, **278**, 21221-21231.
4.   Koutsodontis, G., Moustakas, A. and Kardassis, D. (2002) The role of Sp1 family members, the proximal GC-rich motifs, and the upstream enhancer region in the regulation of the

human cell cycle inhibitor p21WAF-1/Cip1 gene promoter. *Biochemistry*, **41**, 12771-12784.

5.   Pardali, K., Kurisaki, A., Moren, A., ten Dijke, P., Kardassis, D. and Moustakas, A. (2000) Role of Smad proteins and transcription factor Sp1 in p21(Waf1/Cip1) regulation by transforming growth factor-beta. *J Biol Chem*, **275**, 29244-29256.

6.   Afshari, C.A., Nichols, M.A., Xiong, Y. and Mudryj, M. (1996) A role for a p21-E2F interaction during senescence arrest of normal human fibroblasts. *Cell Growth Differ*, **7**, 979-988.

7.   Delavaine, L. and La Thangue, N.B. (1999) Control of E2F activity by p21Waf1/Cip1. *Oncogene*, **18**, 5381-5392.

8.   Linke, S.P., Harris, M.P., Neugebauer, S.E., Clarkin, K.C., Shepard, H.M., Maneval, D.C. and Wahl, G.M. (1997) p53-mediated accumulation of hypophosphorylated pRb after the G1 restriction point fails to halt cell cycle progression. *Oncogene*, **15**, 337-345.

9.   Ando, T., Kawabe, T., Ohara, H., Ducommun, B., Itoh, M. and Okamoto, T. (2001) Involvement of the interaction between p21 and proliferating cell nuclear antigen for the maintenance of G2/M arrest after DNA damage. *J Biol Chem*, **276**, 42971-42977.

10.   Cayrol, C., Knibiehler, M. and Ducommun, B. (1998) p21 binding to PCNA causes G1 and G2 cell cycle arrest in p53-deficient cells. *Oncogene*, **16**, 311-320.

11.   Haapajarvi, T., Kivinen, L., Heiskanen, A., des Bordes, C., Datto, M.B., Wang, X.F. and Laiho, M. (1999) UV radiation is a transcriptional inducer of p21(Cip1/Waf1) cyclin-kinase inhibitor in a p53-independent manner. *Exp. Cell Res*, **248**, 272-279.

12.   Loignon, M., Fetni, R., Gordon, A.J. and Drobetsky, E.A. (1997) A p53-independent pathway for induction of p21waf1cip1 and concomitant G1 arrest in UV-irradiated human skin fibroblasts. *Cancer Res*, **57**, 3390-3394.

13.   Liu, M., Iavarone, A. and Freedman, L.P. (1996) Transcriptional activation of the human p21(WAF1/CIP1) gene by retinoic acid receptor. Correlation with retinoid induction of U937 cell differentiation. *J Biol Chem*, **271**, 31723-31728.

14.   Biggs, J.R., Kudlow, J.E. and Kraft, A.S. (1996) The role of the transcription factor Sp1 in regulating the expression of the WAF1/CIP1 gene in U937 leukemic cells. *J Biol Chem*, **271**, 901-906.

15.   Chin, Y.E., Kitagawa, M., Su, W.C., You, Z.H., Iwamoto, Y. and Fu, X.Y. (1996) Cell growth arrest and induction of cyclin-dependent kinase inhibitor p21 WAF1/CIP1 mediated by STAT1. *Science*, **272**, 719-722.

16.   Owen, G.I., Richer, J.K., Tung, L., Takimoto, G. and Horwitz, K.B. (1998) Progesterone regulates transcription of the p21(WAF1) cyclin- dependent kinase inhibitor gene through Sp1 and CBP/p300. *J Biol Chem*, **273**, 10696-10701.

17.   Yan, G.Z. and Ziff, E.B. (1997) Nerve growth factor induces transcription of the p21 WAF1/CIP1 and cyclin D1 genes in PC12 cells by activating the Sp1 transcription factor. *J Neurosci*, **17**, 6122-6132.

18.   Michieli, P., Chedid, M., Lin, D., Pierce, J.H., Mercer, W.E. and Givol, D. (1994) Induction of WAF1/CIP1 by a p53-independent pathway. *Cancer Res*, **54**, 3391-3395.

19.   Gartel, A.L., Radhakrishnan, S.K., Serfas, M.S., Kwon, Y.H. and Tyner, A.L. (2004) A novel p21WAF1/CIP1 transcript is highly dependent on p53 for its basal expression in mouse tissues. *Oncogene*, **23**, 8154-8157.

20.   Radhakrishnan, S.K., Gierut, J. and Gartel, A.L. (2006) Multiple alternate p21 transcripts are regulated by p53 in human cells. *Oncogene*, **25**, 1812-1815.

21.   Li, Y., Jenkins, C.W., Nichols, M.A. and Xiong, Y. (1994) Cell cycle expression and p53 regulation of the cyclin-dependent kinase inhibitor p21. *Oncogene*, **9**, 2261-2268.

22.   Besson, A. and Yong, V.W. (2000) Involvement of p21(Waf1/Cip1) in protein kinase C alpha-induced cell cycle progression. *Mol. Cell Biol*, **20**, 4580-4590.

23. LaBaer, J., Garrett, M.D., Stevenson, L.F., Slingerland, J.M., Sandhu, C., Chou, H.S., Fattaey, A. and Harlow, E. (1997) New functional activities for the p21 family of CDK inhibitors. *Genes Dev*, **11**, 847-862.

24. Drouin, R., Therrien, J.-P., Angers, M. and Ouellet, S. (2001) In Moss, T. (ed.), *DNA-Protein Interactions: Principles and Protocols*. second edition ed. Humana Press, Totowa, NJ., pp. 175-219.

25. Rouget, R., Vigneault, F., Codio, C., Rochette, C., Paradis, I., Drouin, R. and Simard, L.R. (2005) Characterization of the survival motor neuron (SMN) promoter provides evidence for complex combinatorial regulation in undifferentiated and differentiated P19 cells. *Biochem. J*, **385**, 433-443.

26. Sanger, F., Nicklen, S. and Coulson, A.R. (1977) DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. USA*, **74**, 5463-5467.

27. Bergeron, M.J., Leclerc, S., Laniel, M.A., Poirier, G.G. and Guerin, S.L. (1997) Transcriptional regulation of the rat poly(ADP-ribose) polymerase gene by Sp1. *Eur. J. Biochem*, **250**, 342-353.

28. Roy, R.J., Gosselin, P. and Guerin, S.L. (1991) A short protocol for micro-purification of nuclear proteins from whole animal tissue. *Biotechniques*, **11**, 770-777.

29. Roy, R.J. and Guerin, S.L. (1994) The 30-kDa rat liver transcription factor nuclear factor 1 binds the rat growth-hormone proximal silencer. *Eur. J. Biochem*, **219**, 799-806.

30. Laniel, M.A., Poirier, G.G. and Guerin, S.L. (2001) Nuclear factor 1 interferes with Sp1 binding through a composite element on the rat poly(ADP-ribose) polymerase promoter to modulate its activity in vitro. *J. Biol. Chem*, **276**, 20766-20773.

31. Baldwin, A.S.J. (1989) In Struhl, K. (ed.), *Current Protocols in Molecular Biology*. Eds Wiley, New York, pp. 12.18-12.10.

32. Harvey, M., Brisson, I. and Guerin, S.L. (1993) A simple apparatus for fast and inexpensive recovery of DNA from polyacrylamide gels. *Biotechniques*, **14**, 942-948.

33. Robidoux, S., Gosselin, P., Harvey, M., Leclerc, S. and Guerin, S.L. (1992) Transcription of the mouse secretory protease inhibitor p12 gene is activated by the developmentally regulated positive transcription factor Sp1. *Mol. Cell Biol*, **12**, 3796-3806.

34. Graham, F.L. and van der Eb, A.J. (1973) A new technique for the assay of infectivity of human adenovirus 5 DNA. *Virology*, **52**, 456-467.

35. Selden, R.F., Howie, K.B., Rowe, M.E., Goodman, H.M. and Moore, D.D. (1986) Human growth hormone as a reporter gene in regulation studies employing transient gene expression. *Mol. Cell Biol*, **6**, 3173-3179.

36. Pothier, F., Ouellet, M., Julien, J.P. and Guerin, S.L. (1992) An improved CAT assay for promoter analysis in either transgenic mice or tissue culture cells. *DNA Cell Biol*, **11**, 83-90.

37. Oberley, M.J. and Farnham, P.J. (2003) Probing chromatin immunoprecipitates with CpG-island microarrays to identify genomic sites occupied by DNA-binding proteins. *Methods Enzymol*, **371**, 577-596.

38. Temple, M.D., Cairns, M.J., Kim, A. and Murray, V. (1999) Protein-DNA footprinting of the human epsilon-globin promoter in human intact cells using nitrogen mustard analogues and other DNA-damaging agents. *Biochim Biophys Acta*, **1445**, 245-256.

39. Zhao, X.T., You, Z.S., Cheng, P. and Wang, Y. (1999) Investigation of Protein-DNA Interactions in Enhancer II and Core Promoter of HBV by in vivo Footprinting. *Sheng Wu Hua Xue Yu Sheng Wu Wu Li Xue Bao (Shanghai)*, **31**, 489-493.

40. Koutsodontis, G., Tentes, I., Papakosta, P., Moustakas, A. and Kardassis, D. (2001) Sp1 plays a critical role in the transcriptional activation of the human cyclin-dependent kinase inhibitor p21(WAF1/Cip1) gene by the p53 tumor suppressor protein. *J Biol Chem*, **276**, 29116-29125.

41. Gartel, A.L., Najmabadi, F., Goufman, E. and Tyner, A.L. (2000) A role for E2F1 in Ras activation of p21(WAF1/CIP1) transcription. *Oncogene*, **19**, 961-964.

42.   Hiyama, H., Iavarone, A. and Reeves, S.A. (1998) Regulation of the cdk inhibitor p21 gene during cell cycle progression is under the control of the transcription factor E2F. *Oncogene*, **16**, 1513-1523.

43.   Eskild, W., Simard, J., Hansson, V. and Guerin, S.L. (1994) Binding of a member of the NF1 family of transcription factors to two distinct cis-acting elements in the promoter and 5'-flanking region of the human cellular retinol binding protein 1 gene. *Mol Endocrinol*, **8**, 732-745.

44.   Steffensen, K.R., Holter, E., Tobin, K.A., Leclerc, S., Gustafsson, J.A., Guerin, S.L. and Eskild, W. (2001) Members of the nuclear factor 1 family reduce the transcriptional potential of the nuclear receptor LXRalpha promoter. *Biochem Biophys Res Commun*, **289**, 1262-1267.

45.   Wickenheisser, J.K., Nelson-DeGrave, V.L., Quinn, P.G. and McAllister, J.M. (2004) Increased cytochrome P450 17alpha-hydroxylase promoter function in theca cells isolated from patients with polycystic ovary syndrome involves nuclear factor-1. *Mol Endocrinol*, **18**, 588-605.

46.   Roulet, E., Bucher, P., Schneider, R., Wingender, E., Dusserre, Y., Werner, T. and Mermod, N. (2000) Experimental analysis and computer prediction of CTF/NFI transcription factor DNA binding sites. *J. Mol. Biol*, **297**, 833-848.

47.   Chang, B.D., Watanabe, K., Broude, E.V., Fang, J., Poole, J.C., Kalinichenko, T.V. and Roninson, I.B. (2000) Effects of p21Waf1/Cip1/Sdi1 on cellular gene expression: implications for carcinogenesis, senescence, and age-related diseases. *Proc. Natl. Acad. Sci. USA*, **97**, 4291-4296.

48.   Funato, N., Ohtani, K., Ohyama, K., Kuroda, T. and Nakamura, M. (2001) Common regulation of growth arrest and differentiation of osteoblasts by helix-loop-helix factors. *Mol. Cell Biol*, **21**, 7416-7428.

49.   Huang, Z.Y., Wu, Y., Hedrick, N. and Gutmann, D.H. (2003) T-cadherin-mediated cell growth regulation involves G2 phase arrest and requires p21(CIP1/WAF1) expression. *Mol. Cell Biol*, **23**, 566-578.

50.   Kaufmann, S.H., Desnoyers, S., Ottaviano, Y., Davidson, N.E. and Poirier, G.G. (1993) Specific proteolytic cleavage of poly(ADP-ribose) polymerase: an early marker of chemotherapy-induced apoptosis. *Cancer Res*, **53**, 3976-3985.

51.   Lazebnik, Y.A., Kaufmann, S.H., Desnoyers, S., Poirier, G.G. and Earnshaw, W.C. (1994) Cleavage of poly(ADP-ribose) polymerase by a proteinase with properties like ICE. *Nature*, **371**, 346-347.

52.   Luciakova, K., Barath, P., Poliakova, D., Persson, A. and Nelson, B.D. (2003) Repression of the human adenine nucleotide translocase-2 gene in growth-arrested human diploid cells: the role of nuclear factor-1. *J Biol Chem*, **278**, 30624-30633.

53.   Furlong, E.E., Keon, N.K., Thornton, F.D., Rein, T. and Martin, F. (1996) Expression of a 74-kDa nuclear factor 1 (NF1) protein is induced in mouse mammary gland involution. Involution-enhanced occupation of a twin NF1 binding element in the testosterone-repressed prostate message-2/clusterin promoter. *J Biol Chem*, **271**, 29688-29697.

54.   Nakamura, M., Okura, T., Kitami, Y. and Hiwada, K. (2001) Nuclear factor 1 is a negative regulator of gadd153 gene expression in vascular smooth muscle cells. *Hypertension*, **37**, 419-424.

55.   Sun, P., Dong, P., Dai, K., Hannon, G.J. and Beach, D. (1998) p53-independent role of MDM2 in TGF-beta1 resistance. *Science*, **282**, 2270-2272.

56.   Mukhopadhyay, S.S., Wyszomierski, S.L., Gronostajski, R.M. and Rosen, J.M. (2001) Differential interactions of specific nuclear factor I isoforms with the glucocorticoid receptor and STAT5 in the cooperative regulation of WAP gene transcription. *Mol Cell Biol*, **21**, 6859-6869.

57.   Grunder, A., Ebel, T.T., Mallo, M., Schwarzkopf, G., Shimizu, T., Sippel, A.E. and Schrewe, H. (2002) Nuclear factor I-B (Nfib) deficient mice have severe lung hypoplasia. *Mech Dev*, **112**, 69-77.

58.   Gronostajski, R.M. (2000) Roles of the NFI/CTF gene family in transcription and development. *Gene*, **249**, 31-45.

59.   Kane, R., Murtagh, J., Finlay, D., Marti, A., Jaggi, R., Blatchford, D., Wilde, C. and Martin, F. (2002) Transcription factor NFIC undergoes N-glycosylation during early mammary gland involution. *J. Biol. Chem*, **277**, 25893-25903.

60.   Marin, M., Karis, A., Visser, P., Grosveld, F. and Philipsen, S. (1997) Transcription factor Sp1 is essential for early embryonic development but dispensable for cell growth and differentiation. *Cell*, **89**, 619-628.

61.   Osada, S., Matsubara, T., Daimon, S., Terazu, Y., Xu, M., Nishihara, T. and Imagawa, M. (1999) Expression, DNA-binding specificity and transcriptional regulation of nuclear factor 1 family proteins from rat. *Biochem. J*, **342 (Pt 1)**, 189-198.

62.   Laniel, M.A., Poirier, G.G. and Guerin, S.L. (2001) Nuclear factor 1 interferes with Sp1 binding through a composite element on the rat poly(ADP-ribose) polymerase promoter to modulate its activity in vitro. *J Biol Chem*, **276**, 20766-20773.

63.   Gao, B., Jiang, L. and Kunos, G. (1996) Transcriptional regulation of alpha(1b) adrenergic receptors (alpha(1b)AR) by nuclear factor 1 (NF1): a decline in the concentration of NF1 correlates with the downregulation of alpha(1b)AR gene expression in regenerating liver. *Mol. Cell Biol*, **16**, 5997-6008.

64.   Clark, R.E., Jr., Miskimins, W.K. and Miskimins, R. (2002) Cyclic AMP inducibility of the myelin basic protein gene promoter requires the NF1 site. *Int. J. Dev. Neurosci*, **20**, 103-111.

65.   Rafty, L.A., Santiago, F.S. and Khachigian, L.M. (2002) NF1/X represses PDGF A-chain transcription by interacting with Sp1 and antagonizing Sp1 occupancy of the promoter. *Embo J*, **21**, 334-343.

66.   Araki, K., Nakajima, Y., Eto, K. and Ikeda, M.A. (2003) Distinct recruitment of E2F family members to specific E2F-binding sites mediates activation and repression of the E2F1 promoter. *Oncogene*, **22**, 7632-7641.

67.   Prowse, D.M., Bolgan, L., Molnar, A. and Dotto, G.P. (1997) Involvement of the Sp3 transcription factor in induction of p21Cip1/WAF1 in keratinocyte differentiation. *J Biol Chem*, **272**, 1308-1314.

68.   Kennett, S.B., Udvadia, A.J. and Horowitz, J.M. (1997) Sp3 encodes multiple proteins that differ in their capacity to stimulate or repress transcription. *Nucleic Acids Res*, **25**, 3110-3117.

69.   Lu, N., Heuchel, R., Barczyk, M., Zhang, W.M. and Gullberg, D. (2006) Tandem Sp1/Sp3 sites together with an Ets-1 site cooperate to mediate alpha11 integrin chain expression in mesenchymal cells. *Matrix Biol*, **25**, 118-129.

70.   Tsuda, M., Watanabe, T., Seki, T., Kimura, T., Sawa, H., Minami, A., Akagi, T., Isobe, K., Nagashima, K. and Tanaka, S. (2005) Induction of p21(WAF1/CIP1) by human synovial sarcoma-associated chimeric oncoprotein SYT-SSX1. *Oncogene*, **24**, 7984-7990.

71.   Li, L., He, S., Sun, J.M. and Davie, J.R. (2004) Gene regulation by Sp1 and Sp3. *Biochem Cell Biol*, **82**, 460-471.

2.2.8 Figure Legends

Figure 1. Genomic footprinting of the human p21 gene promoter.

The region shown was analyzed with primer set 1, to reveal the bottom strand sequence from nt -167 to −85 (A), and the primer set 2, to reveal the upper strand sequence from nt -114 to −12 (B) and nt −167 to −137 (C) relative to the transcription initiation site. Lane 1, 8 and 10: LMPCR of naked DNA purified from primary cultures of human skin fibroblasts (HSF) treated *in vitro* (*t*) with DMS (lane 1), UVC (lane 8) or DNaseI (lane 10). Lanes 2, 7 and 9: LMPCR of DNA purified from HSFs treated *in vivo* (*v*) with DMS (lane 2), UVC (lane 7) or DNaseI (lane 9) prior to DNA purification. Lanes 3-6: Maxam-Gilbert sequencing. DMS protected and hypersensitive guanines are indicated by opened and closed circles, respectively on each side of the autoradiograms, whereas UVC protected and hypersensitive sites are indicated by opened and closed squares. The DNaseI protected and hypersensitive sites are indicated by – and +, respectively. (D) Summary of the *in vivo* DMS, UVC and DnaseI footprints identified along the -187 to -136 human p21 gene promoter. The position of the consensus sequence for the specified transcription factors is also indicated above the sequence.

Figure 2. EMSA analysis of nuclear proteins from primary cultured HSFs interacting with the p21 NFI target site.

(A) DNA sequence of the double-stranded oligonucleotides used as probes or competitors in the EMSA experiments. The consensus sequence for NFI is also indicated for comparison purpose (NFIcon). (B) The 5' end-labeled p21.1 oligonucleotide was incubated with nuclear proteins (5 µg) from HSFs either alone (C; lane 2) or with a 150-fold molar excess of various unlabeled competitor oligonucleotides (p21.1, p21.1m5', p21.1m3', p21.1m5'3', p21.2, NFI, AP1, and Sp1; lanes 3 to 10). Formation of DNA–protein complexes was then monitored by EMSA on a 6% native polyacrylamide gel. The position of multiple DNA–protein complexes corresponding to the recognition of the labeled probe

by human NFI proteins is indicated (NFI). The p21.1 probe was also incubated with 5 μg HSFs nuclear proteins either alone (C; lane 11) or in the presence of either 2 μl of a polyclonal antibody directed against NFI (lane 12) or 2μl of a mouse non-immune serum (NIS ; lane 13). Formation of DNA-protein complexes was then monitored by EMSA as above. As an additional negative control, the probe was also incubated with the NFI Ab in the absence of nuclear proteins (lane 14). SC; super-shifted complex resulting from the recognition of the NFI-p21 labeled probe complex by the NFI Ab. P: labeled probe with no added protein (lane 1). U: unbound fraction of the probe. (C) The p21.1 labeled probe used in B was incubated with 5 μg nuclear proteins from HSFs either alone (lanes 1 and 8), or in the presence of increasing concentrations (5- to 500-fold molar excesses) of unlabeled, double-stranded competitors bearing the sequence of either the p21 NFI site (p21.1) or that of the high affinity, NFI prototypical target site (NFI). Formation of DNA-protein complexes was then monitored by EMSA as above. The position of the NFI complex is indicated along with that of the free probe (U).

Figure 3. *In vitro* DMS and DNaseI footprinting of NFI binding to the p21 promoter.

(A) The 5' end-labeled 83 bp SmaI-HindIII fragment from the p21-192 plasmid that covers p21 promoter sequences from –110 to –192 was incubated with 2 μg nuclear proteins from a CM-sepharose-enriched preparation of rat liver NFI in the presence of either no (C; lanes 2 and 5) or a 150-fold molar excess of unlabeled competitor oligomers (NFI or Sp1; lanes 3 and 4, respectively) Formation of DNA–protein complexes was then monitored by EMSA on a 4% native polyacrylamide gel. The position of the NFI/p21-192 DNA-protein complex is indicated (NFI). The –110/–192 labeled probe was also incubated with 2 μl of the NFI Ab to monitor the formation of the NFI/NFIAb/p21 protein-protein-DNA supershifted complex (SC in lane 6). P: labeled probe with no added protein (lane 1). U: unbound fraction of the labeled probe. (B) The labeled probe used in A was methylated with DMS and incubated with CM-Sepharose-enriched NFI before separation of the DNA/protein complex by EMSA. Both the labeled DNA from the NFI complex (NFI in panel A) and the unbound fraction of the probe (U in panel A) were isolated and further treated with piperidine before being analyzed on a 8% polyacrylamide sequencing gel. The DNA

sequence from the p21 promoter that includes the protected G residues (full and half circles correspond to fully and partly protected G residues, respectively) is indicated along with its positioning relative to the p21 mRNA start site. The p21 NFI target site is also indicated (box). (C) A 228 bp DNA fragment spanning p21 promoter sequences from position −192 to +36 was 5' end-labeled and incubated with 75µg CM-Sepharose-enriched NFI before being digested with DNaseI. The position of the NFI site protected from digestion by DNaseI is shown (shaded box) along with that of the p21 NFI site identified by *in vivo* footprinting (full line box). G: Maxam and Gilbert 'G' sequencing ladder; C: labeled DNA digested by DNaseI in the absence of proteins.

Figure 4. Transient transfection analysis.

(A) Schematic representation of the plasmids used. The NFI target site from the p21 promoter is shown. Numbers indicate positions relative to the p21 mRNA start-site. (B) The plasmids shown in A were transfected into HSFs, as well as in the tissue culture cell lines GH4C1 and HeLa. Cells were harvested and CAT activities determined and normalized to secreted hGH. *CAT activities that are statistically different from those obtained with the p21-192 construct ($P<0.05$; paired samples, *t*-test).

Figure 5. Influence of NFI over-expression on p21 promoter activity in growth-arrested HCT116 cells.

(A) Crude nuclear extracts were prepared from both HCT116 and HeLa cells (used as a control) and examined for expression of Sp1 and NFI in Western blot analyses. (B and C) The p21-192 recombinant construct was co-transfected with the empty vector pCH (+EV), or with expression plasmids encoding each of the NFI isoforms, either individually (+NFI-A, -B, -C, and –X) or in combination (+NFI-BCX), in doxorubicin, growth-arrested HCT116 cells. Cells were collected 48 h later and used either for the measurement of the CAT activity (panel B) or to determine the proportion of the cells engaged in the S phase of the cell cycle by FACS analyses (panel C).

Figure 6. Influence of serum starvation on NFI binding and p21 promoter function *in vitro*.

(A) The p21-192 construct and its NFI-mutated derivative p21-192 mNFI were transfected into HSF grown either in complete (+FBS) or serum-free DMEM (-FBS) for 72 hrs. Cells were then harvested and CAT activities determined and normalized to secreted hGH. * indicates CAT activities that are statistically different from those obtained with the p21-192 construct ($P<0.05$; paired samples, *t*-test). (B) A double-stranded oligonucleotide bearing the high affinity binding site for human CTF/NFI was incubated with 5 µg proteins from HSFs grown for various periods of time post-transfection (48, 60, 66, and 72 hours) in complete (+FBS) or serum-deprived DMEM (-FBS). Formation of DNA–protein complexes was then monitored by EMSA on a 6% native polyacrylamide gel. The position of multiple DNA–protein complexes corresponding to the recognition of the labeled probe by human NFI is indicated (A and B). Extracts from each condition were also incubated either alone (-; lanes 2, 4, 6, 8, and 10) or with 2 µl of the NFI Ab (+; lanes 3, 5, 7, 9, and 11). P: labeled probe with no added protein (lane 1); U: unbound fraction of the labeled probe. (C) Western blot analysis of NFI in nuclear extracts prepared from HSFs cultured in complete DMEM (+FBS) or FBS-free DMEM (- FBS) for 48 to 72 hours. (D) Western blot analysis of PARP-1 expression in HSFs cultured in complete DMEM (+FBS) or FBS-free DMEM (- FBS) for 48 to 72 hours. Nuclear extracts from HL60 cells cultured in the absence and presence of the apoptosis inducer VP16 were also loaded in lanes 6 and 7 as negative and positive controls, respectively. (E) 15 µg nuclear proteins from each of the extracts used above were loaded on a 10% SDS-polyacrylamide gel and stained with Coomassie blue for comparative purpose. (F) Phase-contrast images of HSFs cultured in complete DMEM (+FBS) or FBS-free DMEM (- FBS) for 48 to 72 hours. Magnification, x200.

Figure 7. Chromatin immunoprecipitation of NFI and RNAi assays in human skin fibroblasts. (A) ChIP assays were performed on HSFs in an exponential state of growth or after 72 hours of serum starvation. Chromatin was isolated and immunoprecipited with

antibodies directed against the transcription factors Sp1, Sp3, NFI, and E2F1. PCR of the p21 gene promoter was then carried out on the ChIP samples along with a "no antibody" control (No Ab) that contains chromatin but no antibody, an "input" sample corresponding to 0.2% of the total input chromatin, and a "mock" sample that does not contain chromatin. PCR amplification of a gene segment located ~2000 bp upstream from the p21 promoter was also conducted on the same sample as a negative control for all immunoprecipitates. (B) Graphical representation of the amount of specific PCR products expressed as the percentage of antibody binding versus the amount of PCR product obtained using a standardized aliquot of input chromatin. The signal in the no-antibody lane corresponds to the nonspecific binding background and was subtracted from each sample. (C) RNAi was performed using a combination of siRNA complementary to the NFIA, NFIB, NFIC and NFIX transcripts. The p21 promoter-bearing recombinant constructs p21-124, p21-192 and p21-2300, and the derivative from p21-192 that comprise a mutated NFI site (p21-192 mNFI) were transfected together with either the siRNA silencer negative control or with the combination of NFI siRNA duplexes (siNFI) into subconfluent HeLa cells. Cells were harvested and CAT activities determined. *CAT activities that are statistically different from those obtained with the p21 promoter constructs transfected solely with the siRNA silencer negative control ($P<0.05$; paired samples, t-test).

2.2.9 Figures
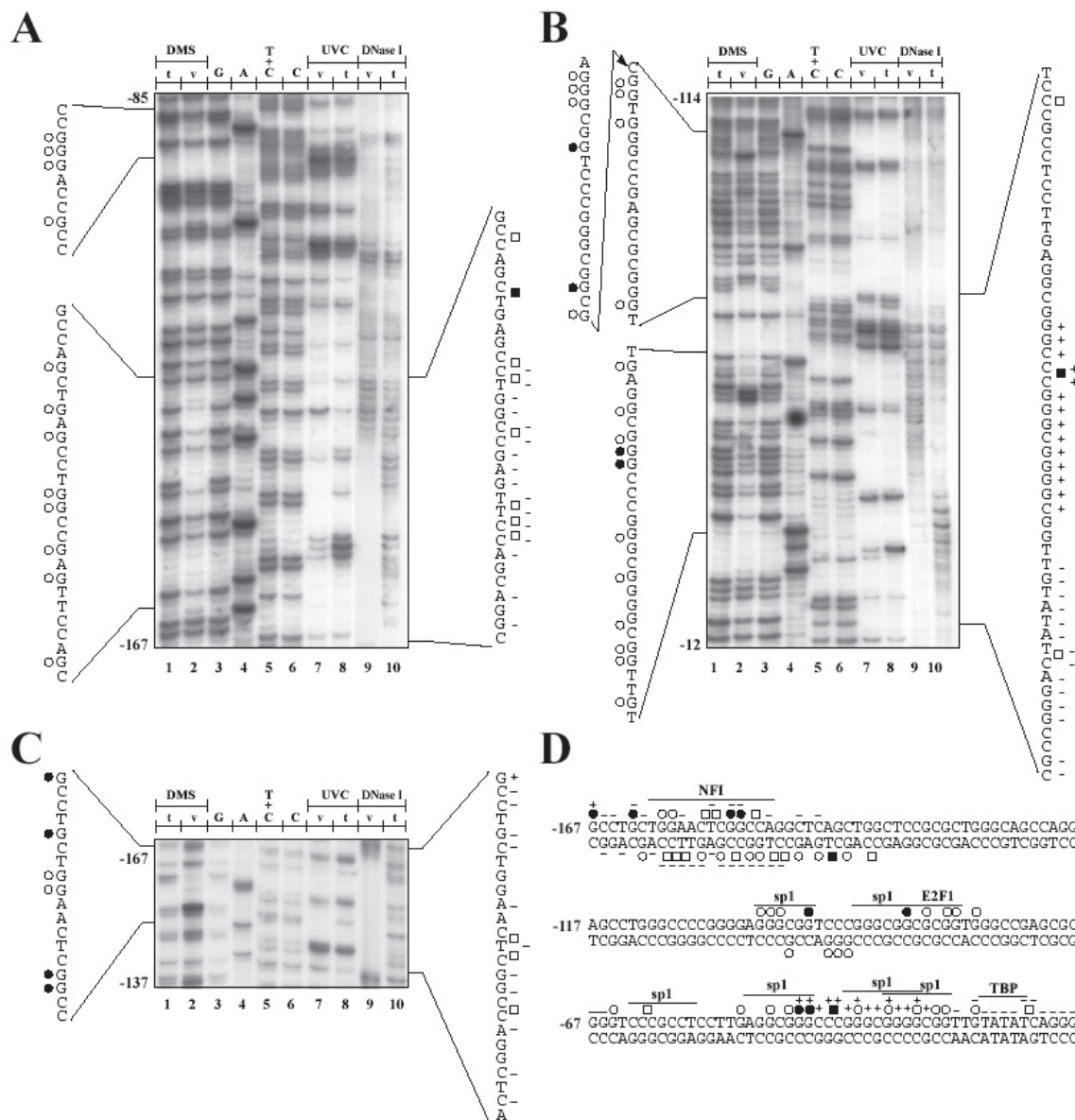
Figure 1 Genomic footprinting of the human p21 gene promoter.
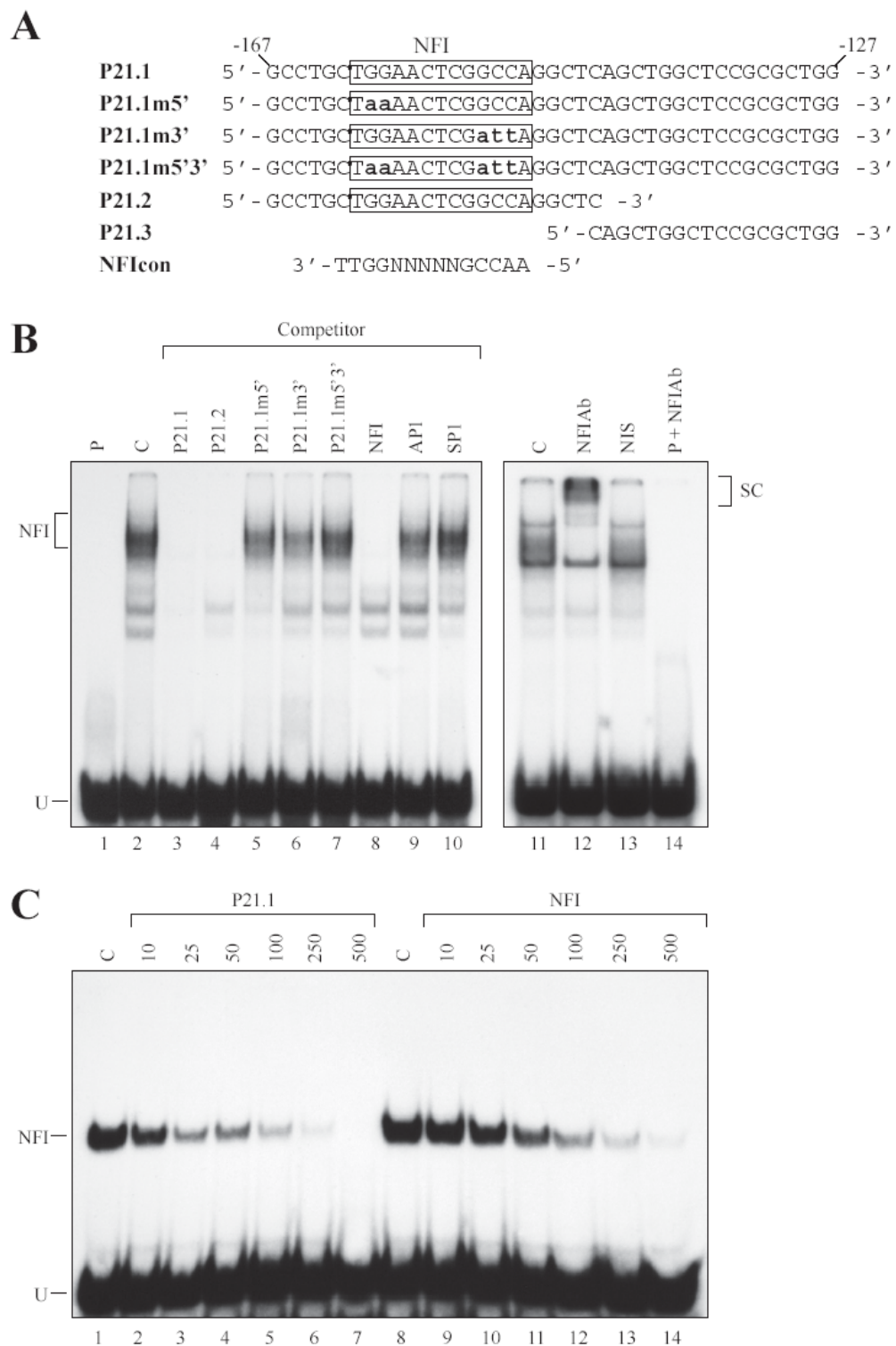
A

```
              -167              NFI                        -127
P21.1      5'-GCCTGC TGGAACTCGGCCA GGCTCAGCTGGCTCCGCGCTGG -3'
P21.1m5'   5'-GCCTGC TaaAACTCGGCCA GGCTCAGCTGGCTCCGCGCTGG -3'
P21.1m3'   5'-GCCTGC TGGAACTCGattA GGCTCAGCTGGCTCCGCGCTGG -3'
P21.1m5'3' 5'-GCCTGC TaaAACTCGattA GGCTCAGCTGGCTCCGCGCTGG -3'
P21.2      5'-GCCTGC TGGAACTCGGCCA GGCTC -3'
P21.3                             5'-CAGCTGGCTCCGCGCTGG -3'
NFIcon              3'-TTGGNNNNNGCCAA -5'
```

Figure 2 EMSA analysis of nuclear proteins from primary cultured HSFs interacting with the p21 NFI target site.
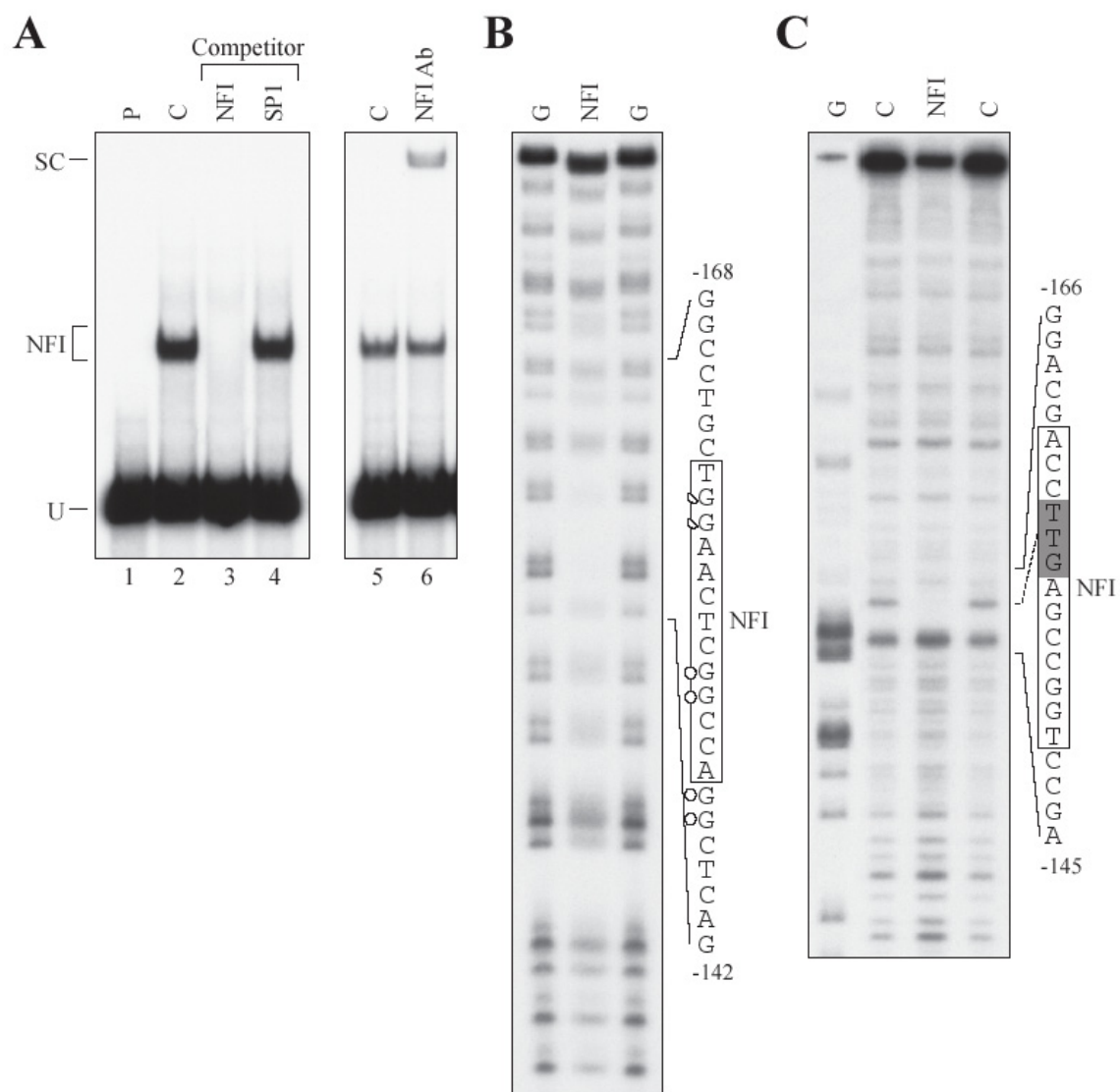
Figure 3 *In vitro* DMS and DNaseI footprinting of NFI binding to the p21 promoter.
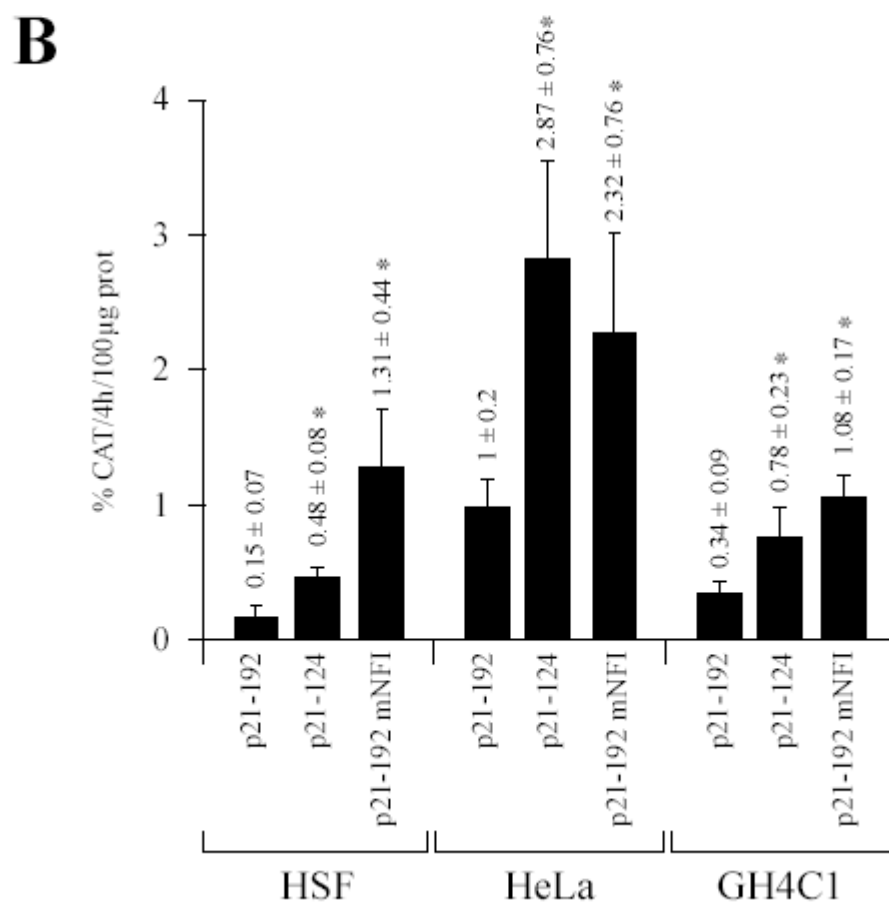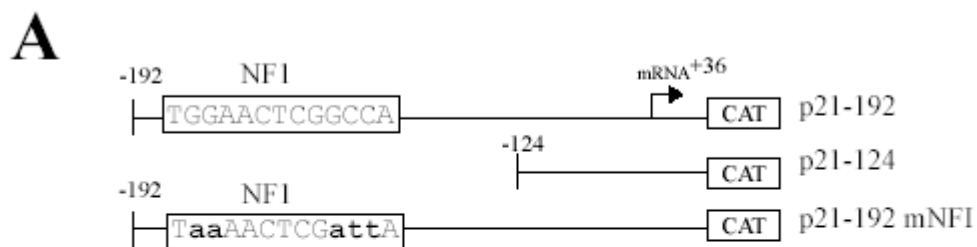
Figure 4 Transient transfection analysis.
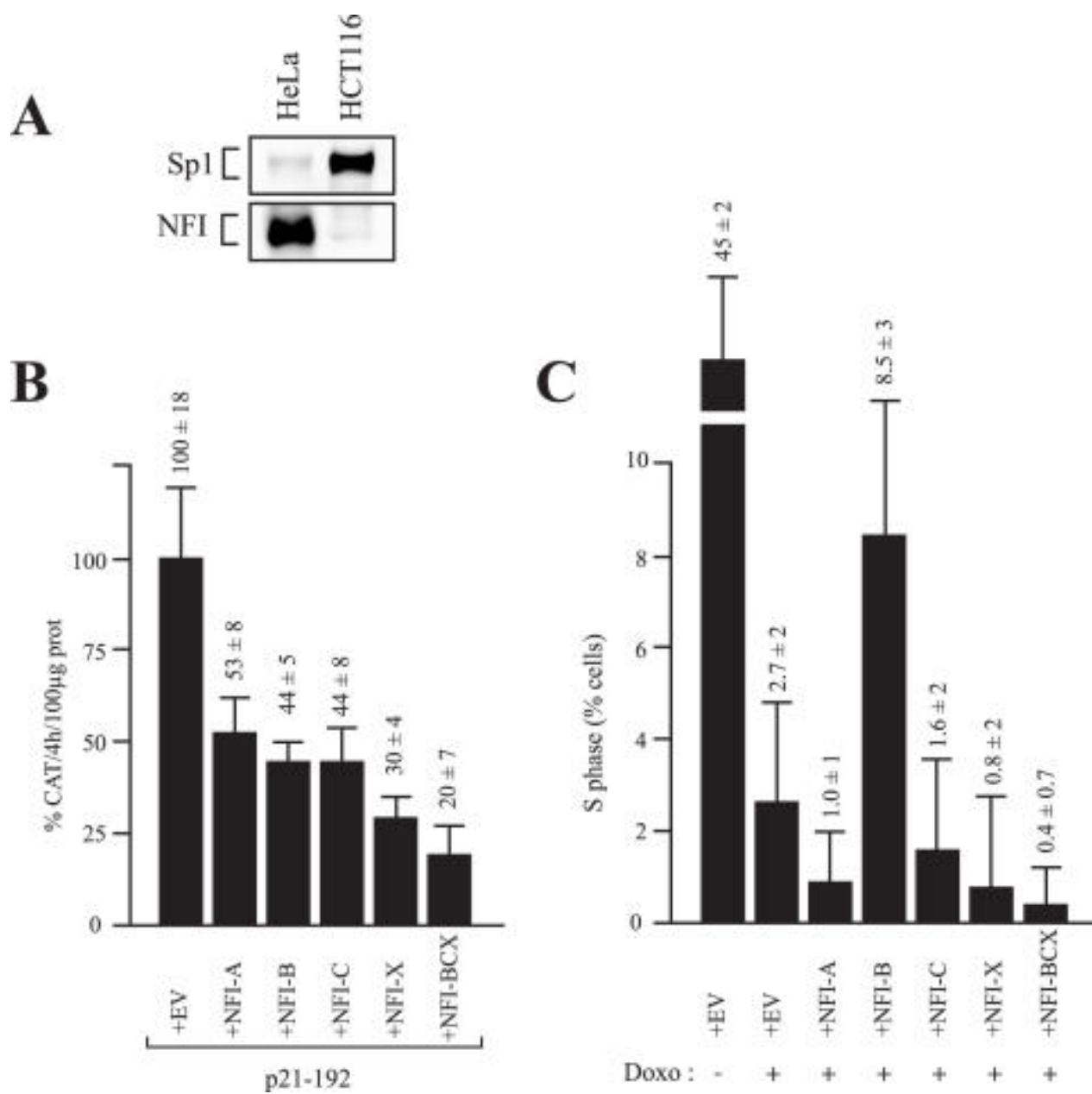
Figure 5 Influence of NFI over-expression on p21 promoter activity in growth-arrested HCT116 cells.
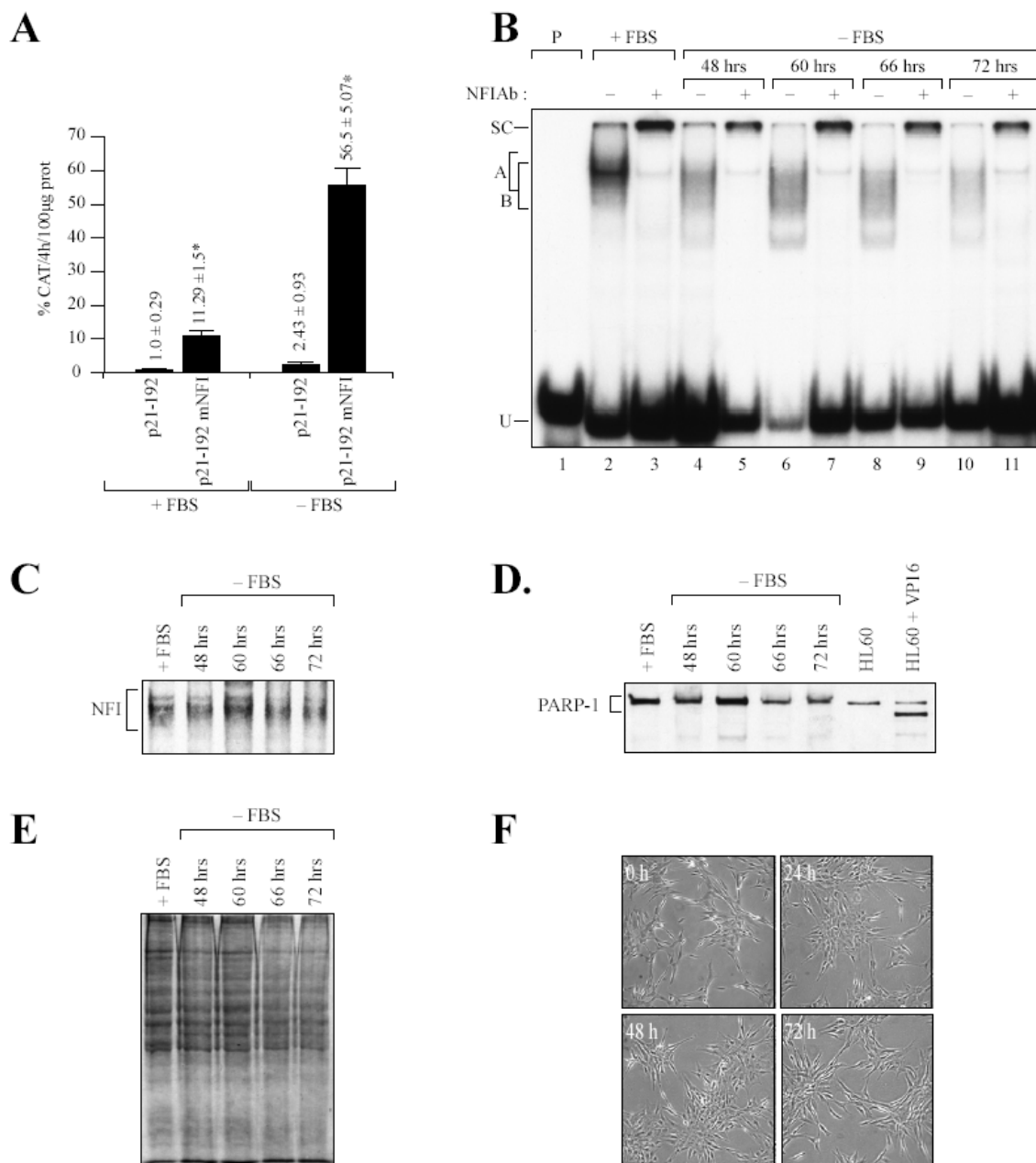
Figure 6 Influence of serum starvation on NFI binding and p21 promoter function *in vitro*.
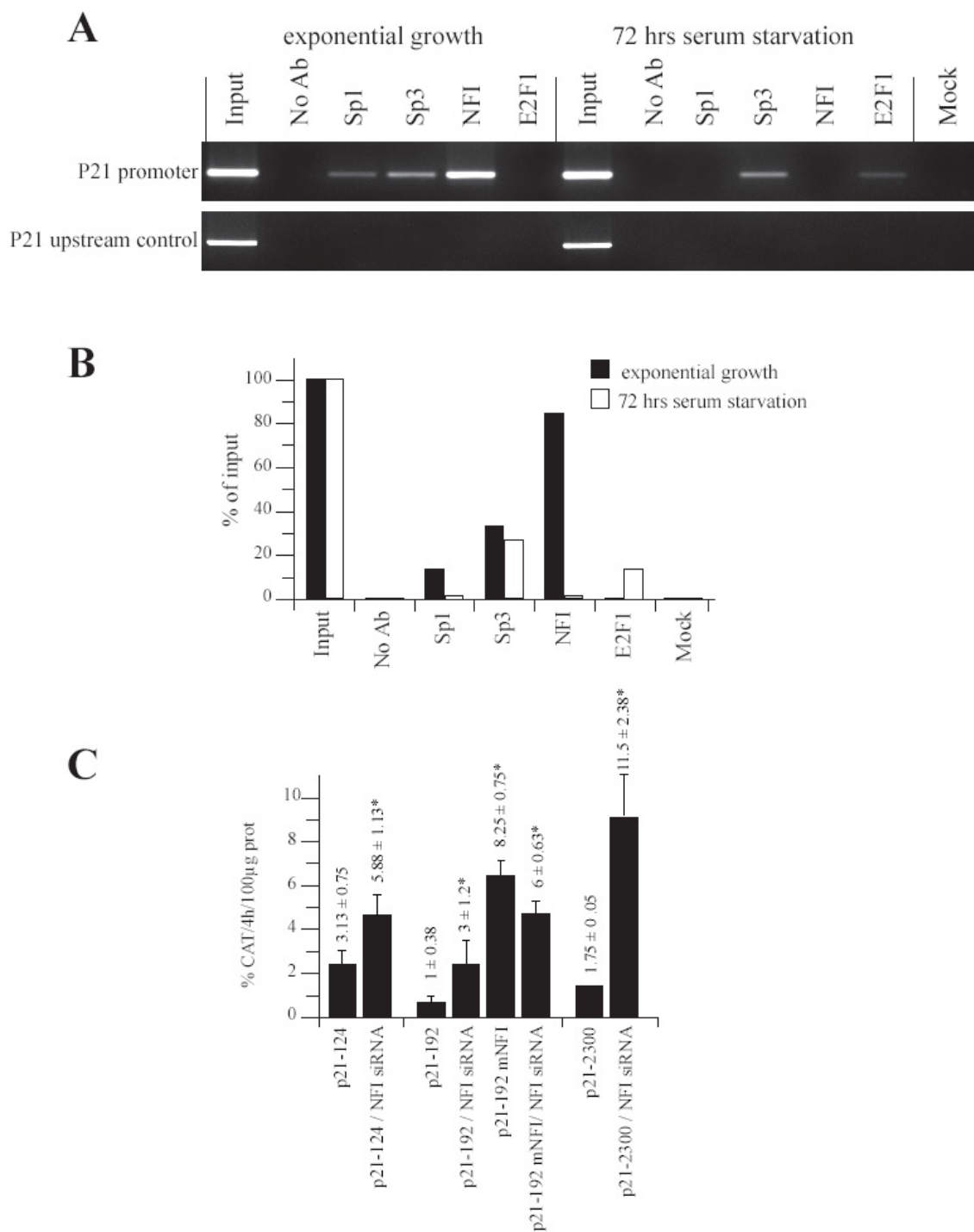
Figure 7 Chromatin immunoprecipitation of NFI and RNAi assays in human skin fibroblasts.

Chapitre 3 : *In vivo* footprinting analysis of the *Glypican 3 (GPC3)* promoter region in neuroblastoma cells

[publié dans *Biochimica Biophysica Acta* le 9 février 2007]

3.1 Résumé

Le gène *Glypican 3* (*GPC3*) est exprimé dans toutes les tumeurs de Wilms (WT) et dans certains neuroblastomes (NB), deux tumeurs embryonnaires. Il est exprimé selon une spécificité tissulaire avec une intensité qui culmine lors de l'embryogenèse. Après la naissance, *GPC3* est réprimé dans tous les tissus investigués. Il est situé sur le chromosome Xq26.1 et s'étend sur plus de 500 kb. Le produit du gène est un HSP (heparan sulfate proteoglycan) localisé à la surface cellulaire et attaché à la membrane cellulaire par une ancre glycosyl-phosphatidyl inositol (GPI). Son rôle n'est pas encore bien connu. Il semble que, dépendamment du contexte cellulaire, GPC3 régule de façon divergente les facteurs de croissance et de survie cellulaire. Les mécanismes qui régulent la transcription de *GPC3* sont d'un intérêt particulier pour comprendre l'expression anarchique du gène dans les cellules cancéreuses. Toutefois, la méthylation n'est pas nécessaire à la répression de la transcription de *GPC3* dans les NB, ce qui indique que les conditions d'expression permissives du gène sont dues à un autre mécanisme de régulation comme l'interaction de facteurs de transcription au niveau du promoteur. Il a été montré que Sp1 est nécessaire à l'activation de la transcription de *GPC3*. Toutefois, étant exprimé dans environ toutes les cellules, ce facteur ne peut contrôler à lui seul la spécificité tissulaire de l'expression de *GPC3*. Afin d'améliorer notre compréhension des mécanismes contrôlant la transcription de GPC3, nous avons cartographié *in vivo* les sites d'interactions ADN-protéines au promoteur de GPC3 dans deux lignés cellulaires de NB : SJNB-7 (exprimant GPC3) et SK-N-FI (n'exprimant pas GPC3). Nous avons ainsi mis en évidence une structure chromatinienne révélatrice de la présence de nucléosomes tout le long de la région promotrice étudiée. Nous avons de plus localisé huit régions protégées incluant des empreintes communes mais aussi spécifiques à la lignée SJNB-7, suggérant la présence d'interactions ADN-protéines dans les deux lignés mais certaines plus stables au sein des cellules exprimant le gène. Les analyses de compétition en retard sur gel combinées à des analyses de super rétention à l'aide d'anticorps suggèrent qu'un facteur de transcription de type NF-Y pourrait être en partie responsable de l'expression ectopique du gène dans la ligné SJNB-7, malgré la présence de nucléosomes.

3.2 Le manuscrit

IN VIVO FOOTPRINTING ANALYSIS OF THE GLYPICAN 3 (GPC3) PROMOTER REGION IN NEUROBLASTOMA CELLS.

Gino Boily,[1*] Stéphane Ouellet,[2,3*] Régen Drouin,[2,3] and Daniel Sinnett[1,4§]

[1]Division of Hemato-Oncology, Charles-Bruneau Cancer Center, Research Center, Sainte-Justine Hospital; 3175 Côte Sainte-Catherine, Montréal (Québec) Canada, H3T 1C5; [2]Unité de Recherche en Génétique Humaine et Moléculaire, Research Center, Hôpital Saint-François d'Assise, CHUQ, and Department of Medical Biology, Faculty of Medicine, Laval University, Québec (Québec), Canada; [3]Present address: Department of Pediatrics, Faculty of Medicine, Université de Sherbrooke, and Medical Genetics, Centre Hospitalier Universitaire de Sherbrooke, 3001, 12 Ave North, Sherbrooke (Québec), Canada, J1H 5N4 [4]Department of Pediatrics, Faculty of Medicine, University of Montreal, Montreal

*These authors contributed equally to this work
§Corresponding author

Email addresses:

      GB: gino_boily@hotmail.com
      SO: stephane.ouellet.1@ulaval.ca
      RD: Regen.Drouin@USherbrooke.ca
      DS: daniel.sinnett@umontreal.ca

3.2.1 Abstract

Glypican 3 (GPC3) is an X-linked gene that has its peak expression during development and is down-regulated in all studied tissues after birth. We have shown that GPC3 was expressed in neuroblastoma and Wilms' tumor. To understand the mechanisms regulating the transcription of this gene in neuroblastoma cells, we have focused our study on the identification of putative transcription factors binding the promoter. In this report we performed *in vivo* dimethylsulfate, UV type C radiation and DNaseI footprinting analyses coupled with ligation-mediated PCR on nearly 1000 bp of promoter in two neuroblastoma cell lines, SJNB-7 (expressing GPC3) and SK-N-FI (not expressing GPC3). Nucleosome signature footprints were observed in the most distal part of the studied region in both cell lines. We detected eight large differentially protected regions, suggesting the presence of binding proteins in both cell lines but more DNA-protein interactions in GPC3-expressing cells. Sp1 was previously shown to be able to bind some of these regions. Here by combining electromobility shift assays and chromatin immunoprecipitations we showed that the transcription factor NFY was part of the DNA-protein complex found in footprinted regions upstream of the described minimal promoter. These studies performed on chromatin *in situ* suggest that NFY and yet unknown cell type-specific factors may play an important role in the regulation of GPC3.

3.2.1 Introduction

Glypican 3 (GPC3) has been shown to be expressed in embryonal tumors including neuroblastoma (NB), Wilms' tumor and hepatoblastoma [1, 2]. This gene is expressed in a tissue-specific manner, has its peak expression during development and is down-regulated in adult tissues [3-5]. The product of this gene, located at chromosome Xq26 [6, 7] is a heparan sulfate proteoglycan located on the cell surface and attached to the cellular membrane by a glycosyl-phosphatidyl inositol anchor [7]. The role of this protein has not yet been defined. Many studies suggested that GPC3 is a negative cellular growth regulator [7-12]. A germline mutation of the human GPC3 gene is associated with the Simpson-Golabi-Behmel overgrowth syndrome [7] and GPC3 knock-out mice partly recapitulate the syndrome [8]. On the other hand, GPC3 has been shown to be over-expressed in hepatocellular carcinomas [4, 13, 14] and to be associated with advanced stages as well as with invasive potential of this cancer [4]. Moreover, colorectal carcinoma-associated liver metastases express GPC3 significantly more than primary tumors [15]. These data suggest that, depending on the cellular context, GPC3 is regulating different growth and survival factors [16].

The mechanisms regulating the transcription of GPC3 are of particular interest to understand the altered expression of GPC3 in cancer cells. We recently showed that female-derived NB cells presented a loss of methylation in the promoter region, suggesting that the demethylation of the inactive X-linked GPC3 allele may allow the expression of the gene at least in NB cells [17]. This hypothesis is supported by the fact that GPC3 is over-expressed preferentially in female as compared to male hepatocellular carcinomas [4]. However, methylation is not necessary for gene silencing in NB, indicating that permissive conditions for GPC3 expression must prevail at another regulatory level [17]. Transcription factors (TFs) are key players in the transcriptional regulation of genes. The factor Sp1 has been shown to be involved in the activation of the GPC3 transcription [18], but being expressed in virtually every cell, is not likely to regulate alone the tissue-specific expression of GPC3.

In this report, we carried out a comprehensive *in vivo* footprinting analysis using dimethylsulfate (DMS), 254-nm ultraviolet type C (UVC) light and deoxyribonuclease I (DNase I) treatments coupled with the ligation-mediated PCR approach to identify other regulatory elements in the GPC3 promoter in two NB cell lines, SJNB-7 and SK-N-FI, associated with opposite GPC3 expression pattern [1]. This method allows the probing of DNA-protein interactions in the native nuclear context, which is believed to have a great importance in the regulation of genes. We found several *in vivo* footprints specific to the GPC3 expressing cells. Two of these promoter regions have been further investigated *in vitro* by electrophoresis mobility shift assays and chromatin immunoprecipitations. Our results provide evidence that NFY binds the promoter region in its *in vivo* context but that other unknown cell-specific factors could be involved in the regulation of GPC3 expression.

3.2.2 Materials and methods

3.2.2.1 DMS, UVC and DNase I treatments of cells.

SK-N-FI (obtained from ATCC, Manassas, Virginia) and SJNB-7 (obtained from T. Look), both male-derived NB cell lines, were grown in DMEM supplemented with 10% of FBS. These cell lines were treated *in vivo* with DMS (0.02%) or irradiated with 254-nm UVC light (1500 J/m^2) as previously described [19]. The *in vivo* DNase I treatment was performed carrying out the permeabilization (using 0.05% of Lysolecithin) and the enzymatic digestion (using 8.75 μg/mL of DNaseI, Worthington biochemical corporation, Lakewood, NJ, USA) simultaneously in 4 mL of digestion buffer (150 mM sucrose, 80 mM KCl, 35 mM HEPES pH 7.4, 5 mM MgCl$_2$, and 2 mM CaCl$_2$) for 20 min at room temperature. At the end of the DNase I reaction, the cells were pelleted, and the supernatant was removed. The reaction was stopped and the cells were lysed by the addition of 500 μL of lysis buffer (150 mM NaCl, 5 mM EDTA pH 7.8), 500 μL of buffer C (20 mM Tris-HCl

pH 8, 20 mM NaCl, 20 mM EDTA, 1% sodium dodecyl sulfate (SDS)), and proteinase K to a final concentration of 300 μg/mL. This mixture was incubated at 37°C for 3 h. RNA was removed by digestion with RNase A (100 μg/mL) at 37°C for 1 h. Following phenol, phenol/chloroform and chloroform extractions, DNA was ethanol precipitated and dissolved in water. For footprinting controls with naked DNA, genomic DNA isolated from each cell line was treated in vitro with DMS, UVC or DNase I as previously described [19]. To generate a sequence marker (see Ligation-mediated PCR), genomic DNA from human peripheral blood leucocytes was chemically damaged according to Maxam and Gilbert's protocol [20]. Cyclobutane pyrimidine dimers (CPD) from UVC irradiation were converted into strand breaks by successive T4 endonuclease V and photolyase digestions [19]. Methylated guanines from DMS exposures and chemically damaged DNA were converted into strand breaks by hot piperidine treatment. Strand break frequencies were estimated on an alkaline agarose gel [21].

3.2.2.2 Ligation-mediated PCR.

LMPCR was carried out in duplicate as previously described [19] using the primer set shown in table 1, allowing the analysis of approximately 1000bp (between positions -970 to +175) in the GPC3 promoter. DNase I digestion of DNA leaves ligatable 5'-phosphorylated breaks, but the 3'-ends are free hydroxyl groups. To avoid nonspecific priming of these 3'-OH ends [22], we performed DNA denaturation, primer 1 hybridization and primer 1 extension at a higher temperatures using the thermostable Pfu exo⁻ DNA polymerase (Stratagene, La Jolla, CA, USA). One μg of treated DNA was used for each sample. The chemically cleaved G, A, T+C and C samples done on purified genomic DNA from human peripheral blood leucocytes were included along with the other samples in the LMPCR assays as sequence markers. For primer 1 extension, the samples were denatured at 98°C for 3 min, annealed at a temperature around the Tm of the first primer for 5 min and extended at 75°C for 15 min on a Biometra thermocycler using the Pfu exo⁻ DNA polymerase (1.5 U). The double-strand linker was then ligated to blunted ends with T$_4$

DNA ligase (3.25 U/reaction, Roche Diagnostics, Laval, QC, Canada), at 18°C for about 15 h. After the precipitation, the DNA fragments were amplified by conventional PCR using Taq DNA polymerase (0.03 U/µL) or Pfu exo⁻ DNA polymerase (0.07 U/µL) with the 25-mer of the linker and the second primer (Table 1). The resulting PCR fragments were phenol/chloroform extracted, ethanol precipitated and separated on an 8% polyacrylamide gel (65 cm long) containing 7 M urea and electroblotted onto a nylon membrane and fixed by UVC crosslinking [19]. Primer extension was used to make single stranded hybridization probes by utilizing the second primer. Hybridization of the membrane was done overnight at 65°C.

3.2.2.3 DNA sequence analysis.

To identify putative TFs, the sequence of the promoter region was searched for the presence of consensus sequences for TF binding sites using Matinspector and the TRANSFAC matrices at "optimized" settings ("thresholds (selection) that minimize false positives for each individual matrix"; see details on the Genomatix website: http://www.genomatix.de/software_services/software/MatInspector/MatInspector_stb.html)

3.2.2.4 Electromobility shift and super shift assays (EMSA).

Nuclear extracts were prepared as described in Dignam *et al* [23]. The labeling reaction mix contained 50 ng (350nM) of double-stranded oligo probe (sense strands: -318/-271: GCGGACGGCTGCTGGGAAGCCAATCAGCGCGCTCGAGCCTGCAGCCCC, -400/-341: GGGAAAAGCCCTCCAGGCTGTAGGCCAATGAGCGGCGGGAAGGAGGAGT GAGGCTGGGGA, NFY: ATTTTTCTGATTGGTTAAAAGT, NFY-mut: ATTTTTCTG ATTaaTTAAAAGT, OCT: CCTCTTGGATTTGCATATGGGCTG), 25 µCi of [γ-³²P]ATP (6000 Ci/mM), 10 U of T4 polynucleotide kinase (Invitrogen, Burlington, ON,

Canada), 1X kinase forward exchange buffer (Invitrogen, Burlington, ON, Canada), in a final volume of 10 µl. The mix was incubated at 37°C for 10 min. The reaction was stopped by the addition of 1 µl of 500 mM EDTA solution. The volume was completed to 100 µl with TE. The labeled probes were purified on a G-25 column (Amersham Biosciences, Piscataway, NJ, USA). For the binding reaction, the mixtures contained nuclear protein extract from SK-N-FI (25 µg), SJNB-7 (25µg) cells, 12% glycerol, 20 mM HEPES (pH 7.9), 50 mM KCl, 4 mM MgCl$_2$, 62.5 ng/µl of Poly dI-dC, 1 mM DTT, 1 mM EDTA, in a total volume of 16 µl. The extract premix was incubated at 4°C for 20 min. The probe premix contained 0.5 ng (4.4 nM) of oligo probe, 12% glycerol, 20 mM HEPES (pH 7.9), 50 mM KCl, 4 mM MgCl$_2$, 1 mM DTT, 1 mM EDTA in a final volume of 8 µl. The probe premix was incubated at 4°C for 20 min. 16 µl of the extract premix and 8 µl of the probe premix were mixed together and incubated at 4°C for 2 h. For competition assays, we added a 100-fold molar excess of unlabeled double-stranded oligonucleotide to the extract premix. EMSA supershift assays were performed with antibodies to NFYB (sc-10779X, Santa Cruz Biotechnology, California, USA) as described above. The DNA-protein complexes and unbound probes were resolved by electrophoresis on a 5% polyacrylamide (38:1) gel in 0.25X TBE buffer (210V, for 2 h at 4°C); the gels were dried and detected by autoradiography.

3.2.2.5 Chromatin immunoprecipitation assay.

Chromatin immunoprecipitation was done according to the ChIP-IT^{TM} kit protocol (Active Motif, California, USA). Briefly, SKN-FI and SJNB-7 cells were growth at 70-80% confluence, cross-linked with 1% formaldehyde for 10 min at room temperature, stopped with the addition of glycine, rinsed with 1x PBS and harvested. The resultant cell pellet was lysed and enzymaticaly digested to generate fragments ranging from 200 to 1500 bp. Protein–DNA complexes were enriched by immunoprecipitation using antibody for NF-YA (sc-7711X) or NF-YB (sc-10779X) (Santa Cruz Biotechnology, California, USA) or preimmune rabbit serum (negative control). Protein G agarose beads were added and

washed. DNA-protein complexes were eluted, reverse cross-linked, treated with proteinase K. Following DNA purification, PCR was performed using primers comprising the -271 to -318 footprint region (forward: GCGCTCTGGCATAACTACTG; reverse: AGTGAGGCTGGGGAACTTCT) as well as the -400 to -341 footprint region (forward: CTCTGACTGGCTCTGGGAGA; reverse, CCCAGGAAGGATGAAAAGGA).

3.2.3 Results

To identify cis-acting regulatory elements involved in the transcriptional regulation of GPC3, we applied a strategy combining *in vivo* footprinting analysis and ligation-mediated PCR to detect putative DNA-protein interactions at the gene promoter level in living cultured cells. This analysis is based on the differential sensitivity of discrete DNA sites that can be altered due to interactions of the DNA with intra-nuclear proteins or by the presence of particular DNA structures. In other words, *in vivo* footprinting assesses the local reactivity of modifying agents on the DNA of living cells as compared to that on purified DNA [19]. Consequently, when such a target site is treated with a given DNA damaging agent, it can be either protected (negative footprint) or over-exposed (positive footprint). In this study, we used three damaging agents, DMS, UVC light and DNaseI. DMS is a chemical compound that preferentially methylates guanines whereas UVC light leads to the formation of cyclobutane pyrimidine dimers (CPD) as well as pyrimidine (6-4) pyrimidone photoproducts. It is noteworthy that DMS or UVC alone allows a limited resolution because they target DNA at specific sites. Indeed, no DNA-protein interaction will be detected in the absence of guanine residues. UVC light has the potential to reveal all DNA-protein interactions provided there are dipyrimidine sites on either DNA strand within a putative binding sequence. DNaseI has virtually no preference on the target site it cleaves and then is expected to give more detailed information. Compared to DMS and UVC, DNase I is less base selective, is more efficient at detecting minor groove DNA-protein contacts, provides more information on chromatin structure, displays larger and clearer footprints, and better delimits the boundaries of DNA-protein interactions. On the

other hand, when using DNaseI, the cells need to be permeabilized and this could alter the native structure of the nucleus. Moreover, every one of these agents will affect the DNA integrity in a specific manner and the knowledge of their respective properties can be used to infer the promoter structure (see Discussion). When used together, these methods give complementary information on a promoter.

Footprint analysis was performed on two male NB cell lines to avoid any confounding results that could have been generated by the presence of the inactive X chromosome in female cells. The promoter region covering positions −970 to +175 relative to the transcriptional initiation site has been investigated by LMPCR. To correlate footprints with the expression status, we used two NB cell lines expressing GPC3 differentially: SJNB-7, expressing GPC3, and SK-N-FI, not expressing GPC3 [1]. This way changes in band intensity between samples reflect protein binding or any other changes in DNA structure that alter reactivity with the DNA damaging agent. Representative data are shown in Fig. 1 (and in supplementary materials). Along the investigated region, many protected and hyper-reactive sites have been detected (Figs. 1-2 and supplementary materials). These footprints may be classified into three broad categories: i) footprints common to both cell lines; ii) footprints specific to the GPC3-expressing cell line SJNB-7 and iii) footprints specific to the GPC3-non-expressing cell line SK-N-FI. An uncommonly high density of UVC- and DNaseI-generated footprints was detected, covering most parts of the investigated region (Figs. 1-2 and supplementary materials). Most of these footprints are very narrow and may not reflect the presence of DNA-binding proteins but rather that of particular intra-nuclear structures. A well-known conformation is the nucleosome structure which is characterized by a 10 bp-periodicity of DNaseI and UVC footprints [24, 25]. A DNaseI positive-footprint periodicity, which is common to both cell lines, was clearly observed in the −960 to −770 region although the corresponding UVC footprints were less obvious (Figs 1C-2), possibly because the distribution of pyrimidine dinucleotides (UVC DNA target sites) could not always allow such a pattern.

Being mainly interested to delineate the regulatory mechanisms responsible for the differential mRNA expression, we focused on footprints that were specific to one cell line or the other. Region −770 to −400 presents mostly common footprints and a few narrow

SJNB-7- or SK-N-FI-specific ones (Fig. 2 and supplementary materials). Nucleosome DNaseI signature is also observed in some stretches of this region. Eight larger SJNB-7-specific footprints were found downstream of -400 (identified 1 to 8 in Fig. 2; see also Fig. 1). Note that footprints 1 and 3-8 contain DMS-generated footprints. Since DMS is transparent to nucleosomes, to most DNA secondary structure, and even to some unstable DNA-protein interactions [26, 27], these footprints more likely reflect actual DNA-protein interactions. Interestingly, most of these SJNB-7 footprints are partly constituted of both SJNB-7-specific and common protected and/or hyper-reactive sites. This might indicate the presence of DNA-binding proteins at these sites in both cell lines but that the interaction with DNA is somewhat extended or tighter in the GPC3-expressing cell line SJNB-7.

The sequence of the whole investigated region was submitted to MatInspector [28] to match footprints with putative TF binding sites. Fig. 2 shows all the predicted TF binding sites overlapping the in vivo footprints. However, consensus binding sites are highly degenerated and most predictions are not biologically relevant. In order to gain more confidence in predictions, "philogenetic footprinting" [reviewed in 29] approach has been used. A scan for consensus elements was carried out with MatInspector on mouse and rat sequences corresponding to the human studied region (Genbank entries: human: AF003529; mouse: AL671426; rat: U62019). The predicted sites are listed in Table 2 for each of the 8 footprints described above. If one assumes that the conservation of sequences between species is often indicative of functional importance, this comparative analysis highlights some consensus predictions that might more likely be true.

We wanted to determine whether some of these footprinted regions were indeed able to bind nuclear proteins and eventually to identify these proteins. In this regard, regions corresponding to footprints 6 and 7 (Fig. 2) have previously been shown to Sp1 factor [18]. Interestingly, footprints 1, 3 and 4 look very similar and overlap predicted CCAAT boxes (Figs 1B-2), including two that are conserved across all three species tested (Table 2). We decided to focus on this promoter region to perform EMSAs. Footprints 1, 2 and 4 were investigated with two probes: -400/–341 (nucleotides -400 to -341; covering footprints 1 and 2), and –318/–271 (nucleotides -318 to -271; covering footprint 4). We found that both probes bound specifically proteins of nuclear extracts from both cell lines as showed by the

addition of 100X excess of competitor cold probes (Fig. 3, lanes 1 and 2). In order to further characterize the proteins binding these probes, competition assays with 100X excess of cold NFY (both wild type and mutant) were performed. We chose these probes because NFY was our best candidate to bind the CCAAT boxes present in footprints 1 and 4 (see above). In both cell lines, only the wild type, but not mutant, NFY cold probes could abolish efficiently the specific upper complexes (Fig. 3, lanes 1, 4 and 5). When used to control for non-specificity, OCT2 cold probes failed to efficiently displace the complexes in both cell lines (Fig. 3, lane 3). Supershift experiments performed with NFYB antibody were able to displace the upper specific complexes formed with both probes and extracts from both cell lines (Fig. 3, lanes 1 and 6). Similar experiments performed with OCT2 antibody failed to efficiently displace the complexes (data not shown). Chromatin immunoprecipitation assays revealed an enrichment of NFY at the GPC3 promoter levels in both cell lines tested (Fig. 4). This enrichment was only observed when NFY-B antibody was used (Fig. 4). These results suggest that NFY sits in the regions covered by the two probes.

3.2.4 Discussion

In a first step towards the identification of the cis-acting elements involved in the regulation of GPC3 in NB cells, we applied an *in vivo* footprinting strategy to probe the GPC3 promoter region for the presence of putative DNA-protein binding sites in two NB cell lines expressing GPC3 differentially. For an adequate interpretation of our results, it is important to take into considerations that DNA damaging agents have their own sensitivity in detecting each of these contexts [reviewed in 19]. DMS, because of its small size, may not reveal particular DNA structures, nucleosomes or weak DNA-protein interactions [26, 27]. Accordingly, footprints detected by DMS are more likely to reflect real DNA-protein specific interactions. UVC is more sensitive than DMS to detect weaker DNA-protein interactions as well as particular DNA foldings and then, the results obtained by this approach may be more difficult to interpret. DNaseI is a very large molecule as compared

to DMS and then have a higher steric hindrance, which limits the access to DNA when proteins are bound. DNaseI usually gives a more detailed footprint map, although the resulting footprints are often more difficult to reproduce. These properties reveal complementary information and consequently give a more complete picture than any of the three approaches alone.

When the data of the chromatin structure were examined as a whole, interesting features emerge. We found an uncommonly high density of DNaseI- and UVC-generated footprints in the GPC3 promoter. Such a footprint density may reflect the presence of particular DNA structures. Region -960 to -770 presents a ~10-bp periodicity of DNaseI hyper-reactive sites which is the signature of nucleosome conformation [25]. Nucleosomes can also be revealed by UVC footprints in a ~10 bp-periodicity [24] but in our study such a pattern was only observed in some stretches of the same region. This may be due to the fact that pairs of pyrimidines were not distributed to allow such a pattern. A strikingly high amount of narrow footprints are also present in the region -770 to -400 as well but without any apparent particular pattern except for a few DNaseI nucleosome signature stretches. Most of these footprints may simply reflect other types of DNA foldings than nucleosomes as it seems unlikely to us that an equally high number of TFs might bind specifically the GPC3 promoter.

Further downstream, in the region -400 to +20, we found eight wider footprints of particular interest because they were specific to GPC3-expressing cell line SJNB-7 (Fig 2). Seven of these footprints involve DMS footprints suggesting the presence of true DNA-protein specific interactions. Earlier *in vitro* footprinting analysis of the minimal promoter sequence (-218 to +12) performed in Caco-2 cell extracts allowed the detection of a large footprint overlapping position -43 to -7 and a smaller one covering region -103 to -86 [18]. The *in vitro* footprint -43 to -7 overlaps with the *in vivo* footprints 6 and 7 detected in our study. Of note, *in vivo* footprints 5 and 8 have not been detected by the *in vitro* analysis [18]. This difference may be due to the difference in cell type and/or to the difference between native and *in vitro* promoter contexts. Huber *et al* [18] have shown by EMSAs that this region binds to transcription factor Sp1 and that this TF can activate the transcription of the minimal GPC3 promoter in Drosophila cells. Deletion mapping of the GPC3 promoter

also suggested that a cis-element(s) located between -218 and -82 is (are) important for transcription activation [18]. Some common protection footprints were detected in this region and may be functionally important for transcription regulation. Our *in vivo* footprinting analysis shows the presence of four SJNB-7-specific footprints upstream of the minimal promoter (footprints 1, 2, 3 and 4; Fig. 2), emphasizing the importance to study transcription regulation native nuclear contexts. Interestingly, footprints 1, 3 and 4 show a very similar pattern of footprinting, DMS footprints all centered on a CCAAT box. The only common TF predicted to bind these sites is NFY.

CCAAT boxes are extremely conserved within orthologous genes across species, in terms of positioning as well as orientation and are well-known consensus sequences for the binding of the NFY transcription factor [30]. Whereas NFY binding to the CCAAT box is often essential for gene transcription, particularly for TATA-less genes, NF-Y by itself is largely unable to activate transcription. Rather, its transcriptional role relates to its capacity to enhance transactivation from nearby activating elements and/or to facilitate the positioning of TFs at the transcriptional start site. In this regard it has been suggested that cooperation between Sp1 and NF-Y are responsible for the basal transcription of TATA-less genes [30, 31], which is particularly relevant to GPC3 that is characterized by a high density of CpG dinucleotides [17] and the absence of a TATA box [18]. We have shown by EMSA that footprinted regions -400 to -341 and -318 to -271 specifically bind proteins from nuclear extracts of both *GPC3*-expressing and non-expressing cell lines. The combination of EMSAs, supershift experiments and chromatin immunoprecipitation provided evidence that NFY is the transcription factor involved in the DNA-protein complex observed in both cell lines. These regions being differentially footprinted, it is intriguing that the difference of band pattern is not more obvious. In this regard, it is interesting to note that footprinted regions shown to bind NFY contain a combination of discrete specific but also common footprints (see Fig. 2). This observation may reflect that NFY binds to its elements in both cell lines but with a higher affinity or a more extended contact region in the expressing cell line. This could explain why we see a larger protection in the expressing cell line but also why NFY could still be able to be crosslinked its element in the non-expressing cell line when performing the ChIP experiment.

In conclusion, this *in vivo* footprinting study has confirmed the presence of two previously identified *in vitro* protected regions [18] as well as six other regions showing differential footprinting patterns associated to expression status and thus potentially important for the activation of GPC3. So far, the only two TFs identified as being able to bind this promoter are Sp1 [18] and NFY (this study TF). Since those TFs are ubiquitously expressed, it seems unlikely that they are on their own responsible for the differential expression of GPC3. However, because the regions they bind are differentially footprinted *in vivo*, in relation to the cells' expression status, we suggest that factors related to the native nuclear context determine the capacity of these TFs, and potentially others, to bind their respective binding regions and then activate the transcription or not. Such factors still need to be defined but could involve DNA structure, yet unknown cell line-specific TFs and/or cell-specific co-factors. Future work will need to address these questions to better understand the differential expression of GPC3.

3.2.5 Acknowledgements

3.2.6 References

[1]   Z. Saikali and D. Sinnett, Expression of glypican 3 (GPC3) in embryonal tumors, Int J Cancer 89 (2000) 418-22.

[2]   J.A. Toretsky, N.L. Zitomersky, A.E. Eskenazi, R.W. Voigt, E.D. Strauch, C.C. Sun, R. Huber, S.J. Meltzer and D. Schlessinger, Glypican-3 expression in Wilms tumor and hepatoblastoma, J Pediatr Hematol Oncol 23 (2001) 496-9.

[3]   J. Filmus, J.G. Church and R.N. Buick, Isolation of a cDNA corresponding to a developmentally regulated transcript in rat intestine, Mol Cell Biol 8 (1988) 4243-9.

[4]   H.C. Hsu, W. Cheng and P.L. Lai, Cloning and expression of a developmentally regulated transcript MXR7 in hepatocellular carcinoma: biological significance and temporospatial distribution, Cancer Res 57 (1997) 5179-84.

[5]   M. Pellegrini, G. Pilia, S. Pantano, F. Lucchini, M. Uda, M. Fumi, A. Cao, D. Schlessinger and A. Forabosco, Gpc3 expression correlates with the phenotype of the Simpson-Golabi- Behmel syndrome, Dev Dyn 213 (1998) 431-9.

[6]   R. Huber, L. Crisponi, R. Mazzarella, C.N. Chen, Y. Su, H. Shizuya, E.Y. Chen, A. Cao and G. Pilia, Analysis of exon/intron structure and 400 kb of genomic sequence surrounding the 5'-promoter and 3'-terminal ends of the human glypican 3 (GPC3) gene, Genomics 45 (1997) 48-58.

[7]   G. Pilia, R.M. Hughes-Benzie, A. MacKenzie, P. Baybayan, E.Y. Chen, R. Huber, G. Neri, A. Cao, A. Forabosco and D. Schlessinger, Mutations in GPC3, a glypican gene, cause the Simpson-Golabi-Behmel overgrowth syndrome, Nat Genet 12 (1996) 241-7.

[8]   D.F. Cano-Gauci, H.H. Song, H. Yang, C. McKerlie, B. Choo, W. Shi, R. Pullano, T.D. Piscione, S. Grisaru, S. Soon, L. Sedlackova, A.K. Tanswell, T.W. Mak, H. Yeger, G.A. Lockwood, N.D. Rosenblum and J. Filmus, Glypican-3-deficient mice exhibit developmental overgrowth and some of the abnormalities typical of Simpson-Golabi-Behmel syndrome, J Cell Biol 146 (1999) 255-64.

[9]   H. Lin, R. Huber, D. Schlessinger and P.J. Morin, Frequent silencing of the GPC3 gene in ovarian cancer cell lines, Cancer Res 59 (1999) 807-10.

[10]   S.S. Murthy, T. Shen, A. De Rienzo, W.C. Lee, P.C. Ferriola, S.C. Jhanwar, B.T. Mossman, J. Filmus and J.R. Testa, Expression of GPC3, an X-linked recessive overgrowth gene, is silenced in malignant mesothelioma, Oncogene 19 (2000) 410-6.

[11]   R. Weksberg, J.A. Squire and D.M. Templeton, Glypicans: a growing trend, Nat Genet 12 (1996) 225-7.

[12]   Y.Y. Xiang, V. Ladeda and J. Filmus, Glypican-3 expression is silenced in human breast cancer, Oncogene 20 (2001) 7408-12.

[13]   Y. Midorikawa, S. Ishikawa, H. Iwanari, T. Imamura, H. Sakamoto, K. Miyazono, T. Kodama, M. Makuuchi and H. Aburatani, Glypican-3, overexpressed in hepatocellular carcinoma, modulates FGF2 and BMP-7 signaling, Int J Cancer 103 (2003) 455-65.

[14]   Z.W. Zhu, H. Friess, L. Wang, M. Abou-Shady, A. Zimmermann, A.D. Lander, M. Korc, J. Kleeff and M.W. Buchler, Enhanced glypican-3 expression differentiates the majority of hepatocellular carcinomas from benign hepatic disorders, Gut 48 (2001) 558-64.

[15]   H. Lage, M. Dietel, G. Froschle and A. Reymann, Expression of the novel mitoxantrone resistance associated gene MXR7 in colorectal malignancies, Int J Clin Pharmacol Ther 36 (1998) 58-60.

[16]   J. Filmus, Glypicans in growth control and cancer, Glycobiology 11 (2001) 19R-23R.

[17]   G. Boily, Z. Saikali and D. Sinnett, Methylation analysis of the glypican 3 gene in embryonal tumours, Br J Cancer 90 (2004) 1606-11.

[18]   R. Huber, D. Schlessinger and G. Pilia, Multiple Sp1 sites efficiently drive transcription of the TATA-less promoter of the human glypican 3 (GPC3) gene, Gene 214 (1998) 35-44.

[19]   R. Drouin, J.P. Therrien, M. Angers and S. Ouellet, In vivo DNA analysis, Methods Mol Biol 148 (2001) 175-219.

[20]   A.M. Maxam and W. Gilbert, Sequencing end-labeled DNA with base-specific chemical cleavages, Methods Enzymol 65 (1980) 499-560.

[21]   R. Drouin, S. Gao and G.P. Holmquist, Agarose gel Electrophoresis for DNA damage analysis, in: Pfeifer, G.P. (Ed.) Technologies for detection of DNA damage and mutations, Plenum Press, New York, 1996, pp. 37-43.

[22]   G.P. Pfeifer and A.D. Riggs, Genomic sequencing, in: Griffin, G.H. and Griffin, A.M. Eds.), Methods in molecular biology, Humana Press Inc., Toyowa, NJ, 1993, pp. 169-181.

[23]   J.D. Dignam, R.M. Lebovitz and R.G. Roeder, Accurate transcription initiation by RNA polymerase II in a soluble extract from isolated mammalian nuclei, Nucleic Acids Res 11 (1983) 1475-89.

[24]   J.M. Gale, K.A. Nissen and M.J. Smerdon, UV-induced formation of pyrimidine dimers in nucleosome core DNA is strongly modulated with a period of 10.3 bases, Proc Natl Acad Sci U S A 84 (1987) 6644-8.

[25]   S. Tornaletti, S. Bates and G.P. Pfeifer, A high-resolution analysis of chromatin structure along p53 sequences, Mol Carcinog 17 (1996) 192-201.

[26]   P.L. Chin, J. Momand and G.P. Pfeifer, In vivo evidence for binding of p53 to consensus binding sites in the p21 and GADD45 genes in response to ionizing radiation, Oncogene 15 (1997) 87-99.

[27]   J.D. McGhee and G. Felsenfeld, Reaction of nucleosome DNA with dimethyl sulfate, Proc Natl Acad Sci U S A 76 (1979) 2133-7.

[28]   K. Quandt, K. Frech, H. Karas, E. Wingender and T. Werner, MatInd and MatInspector: new fast and versatile tools for detection of consensus matches in nucleotide sequence data, Nucleic Acids Res 23 (1995) 4878-84.

[29]   Z. Zhang and M. Gerstein, Of mice and men: phylogenetic footprinting aids the discovery of regulatory elements, J Biol 2 (2003) 11.

[30]   R. Mantovani, A survey of 178 NF-Y binding CCAAT boxes, Nucleic Acids Res 26 (1998) 1135-43.

[31]   B.F. Pugh and R. Tjian, Transcription from a TATA-less promoter requires a multisubunit TFIID complex, Genes Dev 5 (1991) 1935-45.

3.2.7 Figures legends

Figure 1: Representative *in vivo* footprints found in the *GPC3* promoter region of NB cell lines SJNB-7 and SK-N-FI. A) The region shown was analyzed with primer set GPC3-B to reveal upper strand sequences from nt + 16 to -61 relative to the transcription initiation site (TIS). B) The region shown was analyzed with primer set GPC3-K to reveal lower strand sequences from nt -284 to -389. C) The region shown was analyzed with primer set GPC3-G to reveal upper strand sequences from nt -962 to −798. A and B) Lanes 1, 3: LMPCR of naked DNA purified from SJNB-7 cells (1) or SK-N-FI cells (3) that was treated *in vitro* (*t*) with DMS. Lanes 2, 4: LMPCR of DNA purified from SK-N-B7 cells (2) or SK-N-FI cells (4) that were treated *in vivo* (*v*) with DMS prior to DNA purification. Lanes 5-8: LMPCR of DNA treated with standard Maxam-Gilbert cleavage reactions. Lanes 9, 11: LMPCR of DNA purified from SJNB-7 cells (9) or SK-N-B7 cells (11) that were irradiated *in vivo* (*v*) with UVC prior to DNA purification. Lanes 10, 12: LMPCR of DNA purified from SJNB-7 cells (10) or SK-N-FI cells (12) that were irradiated *in vitro* (*t*) with UVC. Lanes 13, 15: LMPCR of DNA purified from SJNB-7 cells (13) or SK-N-F1 cells (15) that were treated *in vivo* (*v*) with DNase I prior to DNA purification. Lanes 14, 16: LMPCR of DNA purified from SJNB-7 cells (14) or SK-N-FI cells (16) that were treated *in vitro* (*t*) with DNase I. DMS protections, all specific for SJNB-7 cells, are indicated with arrows on the left side of the autoradiogram. UVC protections and hyper-reactive sites are indicated by arrows on the right side of the autoradiogram. The DNase I protected and hypersensitive sites are indicated by − and + respectively, just on right of the UVC footprint arrows. All indicated footprints are mapped on right along the DNA sequence. DMS protected sites are indicated by empty circles. UVC protected and hypersensitive sites are indicated by empty and filled squares, respectively. DNaseI protected and hypersensitive sites are indicated by − and +, respectively. Footprints common to both cell lines are indicated in black and SJNB-7-specific ones are indicated in green. **C**) Lanes 1-4: LMPCR of DNA treated with standard Maxam-Gilbert cleavage reactions. Lanes 5, 7: LMPCR of DNA purified from SJNB-7 cells (5) or SK-N-B7 cells (7) that were treated *in vivo* (*v*) with DNase I prior to DNA purification. Lanes 6, 8: LMPCR of DNA purified from SJNB-7 cells (6) or SK-N-FI cells (8) that were treated *in vitro* (*t*) with DNase I. DNase I protected and hypersensitive sites

are indicated by – and + respectively on the right of the autoradiogram. Footprints common to both cell lines are indicated in black and SJNB-7-specific ones are indicated in green.

Figure 2: Detailed map of *in vivo* footprints identified in the *GPC3* promoter of NB cell lines SJNB-7 and SK-N-FI and predicted transcription factor binding sites. Footprints: DMS protected and hypersensitive sites are indicated by empty and filled circles, respectively. UVC protected and hypersensitive sites are indicated by empty and filled squares, respectively. DNaseI protected and hypersensitive sites are indicated by – and +, respectively. Footprints common to both cell lines are indicated in black, SJNB-7-specific ones are in green and SK-N-FI-specific ones are in red. Predicted TF binding sites: Only the predicted TF binding sites corresponding to observed footprints are shown. Arrows mark the position and direction of predicted transcription binding sites corresponding to the TF matrix identifier above it, as obtained with MatInspector.

Figure 3: Electromobility shift assay (EMSA) of regions –400 to –341 and –318 to –271 of the GPC3 promoter in NB cell lines SJNB-7 and SK-N-FI. A. EMSA with SJNB-7 nuclear extract and –318/-271 probe. B. EMSA with SK-N-FI nuclear extract and –318/-271 probe. C. EMSA with SJNB-7 nuclear extract and –400/-341 probe. D. EMSA with SK-N-FI nuclear extract and –400/-341 probe. EMSAs were performed with nuclear extracts from cell lines SJNB-7 (GPC3+) or SK-N-FI (GPC3-), and labeled double-stranded oligos (probes) made from the *GPC3* promoter sequence. –318/-271: oligo from region –318 to –271; –400/-341: oligo from region –400 to –341. Competition assays were performed in the presence of 100X unlabeled (cold) oligos –318/-271, -400/-341, NFY binding site consensus, NFY mutated binding site, or OCT binding site consensus. Supershift assays were performed in the presence of antibodies to NFYB.

Figure 4. ChIP assay with NF-YA and NF-YB antibody. PCR amplification of the -271/-318 and the -400/-341 footprint regions of the GPC3 promoter on immunoprecipitated chromatin obtained from SK-N-FI and SJ-N-B7 cells. Negative control corresponds to the immunoprecipitation done with the preimmune rabbit serum and the input (10-fold diluted) represents the positive control.

3.2.8 Tables

Table 1 Characteristics of oligonucleotide primers used for LMPCR analysis of the human *GPC3* gene promoter

Primer	Sequence (5'-3')	Position[a]	$T_m(°C)$[b]
Non-transcribed (upper) strand			
A1	CCAAGCTGAGCAGCAT	+251 to +236	54.1
A2	GCCACCACCAAGCACGCGGTGCGC	+234 to +211	72.2
B1	GGAGCAAGAGACGTGC	+106 to +91	56.6
B2	GCTACCCAGCCGCTGCAAAAGTTTCCT	+89 to +63	66.5
C1	GGGCGAGGCGAGCCG	-36 to -50	64.5
C2	GGGGGCGGTGGCGGTTGTGCGGCC	-59 to -82	75.6
D1	GCAGCCGAGCCCAGCT	-205 to -220	61.7
D2	CTGCCCGCTCGCAGCCGCTCTACACA	-219 to -244	71.4
E1	CTTCCCGCCGCTCATT	-359 to -374	56.6
E2	CAGCCTGGAGGGCTTTTCCCTTTAGGA	-381 to -407	66.5
F1	AGGTGTTCCTTCACACA	-531 to -547	52.1
F2	CAAAATCCACCTCCTCATTCTACTCTCTGAG	-546 to -576	64.6
G1	CTGCCTCACCTTTGAATA	-685 to -702	52.7
G2	TATCTTAAAAGTCTGAAAGCCTCTGTGGC	-706 to -734	62.0
Transcribed (bottom) strand			
H1	GAACTGGATTTCTAATGG	-1034 to -1017	50.4
H2	TGAAATCTAAGCAGGAGAGGTGGGATTT	-1026 to -999	62.0
I1	CTTTAAGCCCCTGGTTA	-884 to -868	52.1
I2	TATTTCGTTGTAACCAGTTAGGCTTTGCC	-869 to -841	62.0
J1	CAGACTTTTAAGATACAATA	-720 to -701	47.5
J2	ATTCAAAGGTGAGGCAGGCTGTGAAAAG	-701 to -674	63.4
K1	CACCTTGTGTGGCTCTG	-535 to -519	57.0
K2	CAGGAAAGAGCTGGCACAAGCTGAAAGAAGG	-513 to -483	67.3
L1	GGAAGGAGGAGTGAGGCT	-363 to -346	59.5
L2	GGAACTTCTCCCAGAGCCAGTCAGAGCG	-343 to -316	69.3
M1	GGCTCCGCACCCTCCT	-188 to -173	61.7
M2	CTCTCGCACTGCCTTCGCCCGGTCCC	-174 to -149	73.0

[a]Primer positions are given relative to the transcription initiation site
[b]T_m determined by the GeneJockey software.

Table 2 Inter-species conservation of predicted transcription factor binding sites overlapping the 8 *in vivo* footprinted regions in the *GPC3* promoter.

Footprinted region[a]	Predictions[b]		
	Human	Rat	Mouse
1	IRF2	PAX6	IRF3
	CREL		CREL
	MYOGNF1		MYOGNF1
	NF1		NF1
	NFY		NFAT
	CAAT		PPARA
	ELF2		HNF4
2	RFX1	RFX1	RFX1
	NFKAPPAB	PAX6	NFKAPPAB
	ELF2		MYOGNF1
	ZNF202		HNF4
3	RFX1	RFX1	RFX1
	CAAT	CAAT	CAAT
	NFY	NFY	NFY
	NFKAPPAB	MZF1	NFKAPPAB
			MZF1
4	NFY	NFY	NFY
	CAAT	CAAT	CAAT
	CLTR_CAAT	CLTR_CAAT	CLTR_CAAT
	CDE	PTX1	PTX1
		COMP1	
		IK2	
5	CMYB	EGR2	MUSCLE_INI
	ZNF202	MUSCLE_INI	
	EGR1		
	ZBP89		
	SP1		
	AP2		
	MYCMAX		
6	MAZR	MAZR	MAZR
	SP1	SP1	SP1
	ZF9	ZF9	ZF9
	MAZ	PAX3	MAZ
	MYCMAX		
7	MAZ	MAZ	MAZ
	PAX3	PAX3	PAX3
	MAZR	STAF	STAF
	ZF9	MOK2	MOK2
	SP1		
	BKLF		
	GC		
	PAX5		
	RREB1		
8	ZF9	ZF9	ZF9
	PAX3	PAX3	PAX3
	GC	GC	GC
	PAX5	PAX5	STAF
	BKLF	STAF	MOK2
	SP1	MOK2	AP2
	MAZR	AP2	NRSE
	RREB1		
	HES1		

[a] : Footprinted regions as identified in figure 2.
[b]: MatInspector predictions at optimal settings (see Genomatix website for details: http://www.genomatix.de).   Black boxes: predictions conserved in the three species. Grey boxes: predictions conserved in two species, human and rat or mouse.

3.2.9 Figures

Figure 1 Representative *in vivo* footprints found in the *GPC3* promoter region of NB cell lines SJNB-7 and SK-N-FI.

A

```
         --                  ++            ++  -        ++   -  ++              --  -- --    +   ■  -     ++
AAATAGCCTTCTACTCCATCTATGAAATAGGCTAGCTTTGGCTCAGTAAATTTGCTGTGTAATGATTTTCTAATGAGTTAGGCTGGCTTTAAGCCCCTGG  -871
         +++     +-+  ■  □                    -  --- +++++    ■ + □□ ++ □    +           -   ++
```

BTEB3.01   E2F.02   ISL1.01   MEIS1 HOXA9.01
                    HMGIY.01
                    HMGIY.01
                        PDX1.01
                        HNF1.02
                        GFI1B.01

```
  □  ++++  -  -  -    +  □       -         +           -  ++  -  +
TTATTTCGTTGTAACCAGTTAGGCTTTGCCTCTTGAAGGGCCACCTGGGACTGTCGTGCAGTAGATTTTCTTTTAACGCCCCAGAATCAGGTGCTTTCTC  -771
      ■       -- ++      ++    ++ □□□      -- □  □□        ■ ++- +     ■      +
  □                                                            ++          ++
```

HAND2 E12.01   E2F.02   MYOD.01
                                VMYB.04

```
       --    -       -- ++      -- ++  ■      --  ■         -  +  -    -- ++ -- -  +     -- - +
TGACTTTGTGTGGCTCTACTGAATCAAATCTAGCAAGCCACAGAGGCTTTCAGACTTTTAAGATACAATATTCAAAGGTGAGGCAGGCTGTGAAAAGCCC  -671
  □          ■     ++ ■  ■     -- ■      ++ □       +++ + ■ ++       ■  □□        □□
```

NFE2.01   CDX2.01   BARBIE.01   RBPJK.01
    AP1.01   AML3.01   EVI1.05
        CP2.02   OCT1P.01
                            CDP.01
                            CDPCR3.01
                            LEF1.02
                            CHR.01

```
-        -       --- --      +++  -
AGCGGTCCCTGGCTGTCCCTGAACGCGACTATTTGCAGGTTGGCTTTGAGAACCCGGTCAGAGCTGCGTTAGGAAAACGGTTCCCGGGAAGCTCCTCAGA  -571
           □□       □□□     ■   ■  ■       □   ■     --  +  □□□   □□ +  ■ □  + □    + [
```

PPARA.01   NFAT.01   NFKAPPAB.03
    VMYB.04   STAT6.01
    CEBPB.01
        VMYB.05
        RFX1.01
        STAT6.01

B

```
GAGTAGAATGAGGAGGTGGATTTTGTGTGAAGGAACACCTTGTGTGGCTCTGGTGGCCAGGAAAGAGCTGGCACAAGCTGAAAGAAGGCCTGTGGCGAAG  -471
```

OC2.01   SRF.02   ETS2.01   PAX3.01

```
CGGAGGGGGACCTAAGTCAGGGACCCCCACCTGCCCCCAGGAAGGATGAAAAGGAGACAAAAATCCTAAAGGGAAAAGCCCTCCAGGCTGTAGGCCAATG  -371
```

1

NGFIC.01   FXRE.01   ZNF202.01   ETS1.01   OCT.01   IRF2.01   NFY.01
CREL.01   CAAT.01
LMO2COM.01   MYOGNF1.01
MYOD.02   NF1.02
NEUROD1.01   NF1.01
RREB1.01

```
AGCGGCGGGAAGGAGGAGTGAGGCTGGGGAACTTCTCCCAGAGCCAGTCAGAGCGGACGGCTGCTGGGAAGCCAATCAGCGCGCTCGAGCCTGCAGCCCC  -271
```

2   3   4

ELF2.01   ZNF202.01   NFKAPPAB.03   NFY.03   NFY.03
RFX1.01   CAAT.01   CAAT.01
CLTR CAAT.01
CDE.01

```
TCTGCAGTAGTTATGCCAGAGCGCCCTGTGTAGAGCGGCTGCGAGCGGGCAGCTGGGCTCGGCTGCCGGGAGCCACCGCGCGGGCTCCGCACCCTCCTCT  -171
```

DBP.01   AP4.01   CDE.01
E47.01   CDE.01
NEUROD1.01   ZF9.01

C

```
            □  ■                        □ ■                ■■                  ++
- ++-- -- --    ----         --------------- -          □□         ++
CGCACTGCCTTCGCCCGGTCCCCGCGCCGCGGTGCCCCAGTGGCCCCCGCCGCGCTCCACGCCGCGCCCCCGCACCCCGCCGGCTACCGGCCGCACAACC  -71
+-   --      -  +            +--  -- --+                    +--        +++++ +   +    -
```

PAX9.01   MUSCLE INI.02   EKLF.01   CMYB.01
ZF9.01   PAX5.02   AP2.01
EBVR.01   ZF9.01
CDE.01   CDE.01
MZF1.01   WT1.01
NRSE.01   ZF5.01
PAX5.01   WT1.01
CDE.01   ZF9.01
    5   6   7   8

```
            ++        □□    ○○□□□○    ■ □□  □○○□□○■   -      --○- -- ------
- ++        □□    ○ ○○○○         ■ □□  □○○□□○■
GCCACCGCCCCCTGGCCGCGCGGCTCGCCTCGCCCCGCCCCGTCCCTCCTCGCCCCGCCCCACCCCAGTCAGCCCCGCCCTGCCCCGCGCCGCCAAGCGG  30
+   ●  ○○○○ -  +++  +  +-- +--  ○ ○○○○   --  +   ●○  ○○    +--    ○○○○    --  ●● ++   ■ ■
++   +                                                                    ++ ++
```

MYCMAX.03   ZF9.01   BKLF.01   MAZR.01
ZNF202.01   MAZ.01   MAZ.01   PAX5.01
EGR1.01   MYCMAX.03   MAZR.01   ZF9.01   GC.01
ZBP89.01   SP1.01   MAZR.01   RREB1.01   HES1.01
SP1.01   PAX3.01   GC.01
AP2.01   SP1.01
                 MAZR.01
                 GC.01

```
        □                  +   -   +      + +   -+ +  +    --  +  -      -   + +++ -- +++
TTCCCGCCCTCGCCCAGCGCCCAGGTAGCTGCGAGGAAACTTTTGCAGCGGCTGGGTAGCAGCACGTCTCTTGCTCCTCAGGGCCACTGCCAGGCTTGCC  130
        ●++            ++
```

SP1.01   ZF5.01   RFX1.01   ATF6.02   PAX3.01
NGFIC.01   MYT1.02

```
- --   +      -   -+ +-- -+   --   +++  -
GAGTCCTGGGACTGCTCTCGCTCCGGCTGCCACTCTCCCGCGCTC                                    175
```

ATF6.01

Figure 2 Detailed map of *in vivo* footprints identified in the *GPC3* promoter of NB cell lines SJNB-7 and SK-N-FI.

A

	1	2	3	4	5	6
-318/-271 probe	+	+	+	+	+	+
NFY mutated cold oligo					+	
NFY cold oligo				+		
OCT cold oligo			+			
-318/-271 cold oligo		+				
NFYB antibody						+

B

	1	2	3	4	5	6
-318/-271 probe	+	+	+	+	+	+
NFY mutated cold oligo					+	
NFY cold oligo				+		
OCT cold oligo			+			
-318/-271 cold oligo		+				
NFYB antibody						+

C

	1	2	3	4	5	6
-400/-341 probe	+	+	+	+	+	+
NFY mutated cold oligo					+	
NFY cold oligo				+		
OCT cold oligo			+			
-400/-341 cold oligo		+				
NFYB antibody						+

D

	1	2	3	4	5	6
-400/-341 probe	+	+	+	+	+	+
NFY mutated cold oligo					+	
NFY cold oligo				+		
OCT cold oligo			+			
-400/-341 cold oligo		+				
NFYB antibody						+

Figure 3: Electromobility shift assays (EMSA) of regions -400 to -341 and -318 to -271 of the *GPC3* promoter in NB cell lines SJNB-7 and SK-N-FI.

Figure 4: ChIP assay with NF-YA and NF-YB antibody.

Chapitre 4 :   Both Sp1 and USF elements of the human amyloid β-precursor protein gene are necessary for full promoter activity

[en préparation pour *Journal of Biological chemistry*]

4.1 Résumé

Il y a présentement 20 millions de personnes atteintes de la maladie d'Alzheimer, une maladie neuro-dégénérative très invalidante, et selon les pronostics, ce nombre doublera d'ici 2030. L'une des caractéristiques majeures de cette pathologie est l'accumulation sous forme de dépôts insolubles du peptide-Aβ42 dans le tissu cérébral des patients. Le peptide-Aβ42 est en fait le résidu de deux clivages spécifiques de l'AβPP (amyloide precursor protein; protéine précurseure de l'amyloïde) par les sécrétases α et γ. L'AβPP est codé par un seul gène, *AβPP*, situé sur le chromosome 21. La perturbation de la régulation de l'expression du gène pourrait être la ou l'une des causes de la formation des dépôts insolubles du peptide-Aβ42. Le contrôle de l'expression du gène humain *AβPP* est encore aujourd'hui mal défini. Son promoteur s'étend sur 3300 paires de bases en amont du site d'initiation de la transcription. Il contient bon nombre de sites putatifs d'interactions ADN-protéines pouvant être impliqués dans la régulation de l'expression du gène. Toutefois, plusieurs études montrent que la région –96 à +100 par rapport au site d'initiation de la transcription serait la seule déterminante de la régulation de l'expression du gène. Nous avons montré que USF et Sp1 se lient au promoteur de *AβPP* et que leur liaison est essentielle pour générer une activité maximale du promoteur. La caractérisation *in cellulo* du promoteur dans les neurones et les astrocytes normaux a révélé huit sites d'interactions ADN-protéines, entre autre au niveau des sites de liaison des facteurs de transcription CTCF, USF et Sp1.

4.2 Le manuscrit

BOTH SP1 AND USF ELEMENTS OF THE HUMAN *AMYLOID B-PRECURSOR PROTEIN* GENE ARE NECESSARY FOR FULL PROMOTER ACTIVITY

Stéphane Ouellet[1], Martin Bourbonnière[2], Régen Drouin[1,3], Joséphine Nalbantoglu[2]

[1]Division of Pathology, Department of Medical Biology, Faculty of Medicine, Laval University and Unité de recherche en génétique humaine et moléculaire, Research Centre, Hôpital St-François d'Assise, Centre Hospitalier Universitaire de Québec, 10 de l'Espinay Street, Quebec City, QC, Canada G1L 3L5; [2]Department of Neurology and Neurosurgery, and McGill Centre for Studies in Aging, McGill University, Montreal, Canada; [3]Present address: Service of Genetics, Department of Paediatrics, Faculty of Medicine and Health Sciences, Université de Sherbrooke, 3001, 12th Avenue North, Sherbrooke, Qc, Canada J1H 5N4

Correspondence: Dr Joséphine Nalbantoglu,
                    McGill Centre for Studies in Aging,
                    McGill University,
                    3801 University Street,
                    Montreal, (Québec), H3A 2B4, Canada
                    Phone number: (514) 398-5920;
                    Fax number: (514) 398-7371
                    E-mail: Josephine.Nalbantoglu@mcgill.ca

Running title: Transcriptionnal regulation of the APP gene

4.2.1 Summary

Post-transcriptionnal modifications of the amyloid β-precursor protein (APP) mRNA lead to a family of proteins produced by specific cleavage of the precursor. One of these, the Aβ42 (39 to 42 amino acids), accumulates as insoluble deposits in the cerebral tissue of Alzheimer and Down syndrome patients. The *APP* gene, located on chromosome 21, is expressed in specific tissues during development and is almost ubiquitous in the adult. Interestingly, in patients, the mRNA levels are significantly higher in certain regions. Perturbations of *APP* gene expression may thus be partly responsible for the subsequent protein accumulation. Binding of transcription factors (TFs) to their consensus DNA sequences on the gene promoter is one of the most important molecular mechanisms regulating gene transcription. The promoter of the *APP* gene extends upto 3.3 Kb proximal to the transcription start site and contains several putative binding sites for transcription factors. However, transfection studies have shown that only the region from −96 to +100 is relevant to *APP* gene regulation. The factors CTCF, Sp1 and USF have been postulated to play a role in *APP* gene regulation. Our objective was to better define the importance of each of the DNA-protein interactions at the APP proximal promoter. We show that Sp1 and USF recognize elements in the human *APP* gene that are both necessary for full promoter activity. *In vivo* footprinting analysis of the −695 to +45 region of the promoter in human normal primary neuronal cells and astrocytes revealed at least 8 consensus sequences showing DNA-protein interactions including at the CTCF, USF and many Sp1 binding elements, confirming the *in vitro* EMSA and transient transfection analyses.

4.2.2 Introduction

One of the major neuropathological hallmarks of Alzheimer's disease (AD) and Down syndrome (DS) is the presence of plaques in vulnerable regions of central nervous system (CNS). These plaques are formed of aggregated β-amyloid peptide of 39 to 42 amino acids (Aβ42) released by the cleavage of the amyloid β-precursor protein (APP). The proteolytic cleavage of APP by specific secretases is a key event in intracellular and extracellular accumulation of Aβ42 (1). The *APP* gene, located on chromosome 21, is present in 3 copies in DS, implicating upregulation of *APP* transcription as a possible mechanism for the initiation of plaque formation. The interplay between transcription and processing is not yet clear. In AD patients, APP mRNA levels are significantly higher in certain CNS regions (2). This supports the hypothesis that perturbations of *APP* gene expression might be responsible for subsequent protein accumulation. Here, we investigated the regulation of expression of the human *APP* gene.

Although there is both developmental and cell-type specific regulation of the *APP* gene, its expression in the adult is almost ubiquitous. Accordingly, the upstream region of the *APP* gene, extending over 3.3 Kb, resembles that of housekeeping genes with no CAAT or TATA boxes but with several GC boxes, multiple transcriptional start sites (+1 and –4) and a prominent initiator element (INR) associated with the primary transcriptional start site at position +1 (3). Progressive deletions of the *APP* promoter have revealed that the sequences located between –94 and +100 are sufficient to confer full activity to the promoter in transient transfection assays in a wide range of cell lines (4-6). Two *in vitro* DNase I footprinted regions were identified to be responsible for the full activity of the promoter. The first, designated APBα (position –53 to –42) contains an E-box (APP-E1) and is adjacent to an Sp1 site (–63 to –56). It was demonstrated that the primary cellular factor binding APBα is a heterodimer of USF43 and USF44 (7-10) . The deletion (6;11) or mutagenesis (12) of the second footprinted region at –92 to –87 (4;5;12), labelled APBβ,

results in a 60-70% decreases in transcriptional activity of the human and rodent genes. This region is bound by the transcription factor CTCF (13).

Previous mutational analyses of the human *APP* promoter yielded contradictory roles for the Sp1 and USF elements. In one case, mutations introduced in the Sp1 element had no effect on *APP* promoter activity whereas mutation in the USF element reduced the activity by only 20% (14). An *APP* construct carrying both mutations lost 90% of activity compared to the wild-type promoter suggesting a role for both elements. Similar results were also observed with the rat *APP* promoter which is conserved up to 85% with human *APP* promoter (7). However, Quitschke et al. (15) observed that mutations in both the Sp1 and USF elements (i.e. double mutant) had no effect on activity of the *APP* promoter in this system.

Our objective was to better define the importance of the DNA-protein interactions at the *APP* proximal promoter. We show that Sp1, like USF, recognizes an element in the human *APP* gene and that both are necessary for full promoter activity. Also, *in vivo* footprinting analysis of the –695 to +45 region of the promoter in human normal primary neuronal cells and astrocytes revealed at least 8 DNA-protein interaction sites including the CTCF (APBβ), the Sp1 and the USF (APBα) binding elements, confirming the EMSA and transient transfection analysis and reinstating the importance of the promoter region from –170 to –94 for *APP* gene regulation.

4.2.3 Experiment proceedures

4.2.3.1 Cell culture

The cell lines NG108-15 and HepG2 were obtained from the American Tissue Cell Collection and were maintained in Dulbecco's modified Eagle's medium (DMEM) supplemented with 10% fetal calf serum (FCS), 2 mM glutamine, 50 U/ml penicillin and 5 ug/ml streptomycin. *Drosophila* S2 cells (a gift from Dr. S. Stifani, McGill University, Montreal) were cultured in Shields and Sang M3 insect medium (Sigma-Aldrich) supplemented with 10% FCS and grown at room temperature in regular atmospheric conditions. Normal human fetal neurons and astrocytes were prepared as described previously (16). Tissue procurement was approved by the Institutional Review Board.

4.2.3.2 Plasmids

The plasmid pKCXB contains an *APP* promoter/reporter chloramphenicol acetyl transferase (CAT) construct generated by introduction of an Xma III / BamH I DNA fragment from the *APP* promoter (-203 to +104) into the Sma I / BamH I sites of pKK232-8 (17), a promoterless CAT vector distributed by Pharmacia Biotech Inc. To further characterize the *APP* promoter, a 3.8 kilobase (kb) BamH I fragment of genomic DNA, taken out from the plasmid pUC18BamHI3.8, containing a large portion of *APP* upstream sequences (-3699 to +100) (3), was introduced into the BamH I site of pBLCAt6, a high copy number, promoterless, CAT based reporter plasmid (18), creating pAPPCAT-3699. Internal deletions in the *APP* promoter were made from pAPPCAT-3699 by single or double digestions with the following restriction enzymes: Sal I and Nsi I, pAPPCAT-3343; Sal I and EcoR V, pAPPCAT-2991; HinD III, pAPPCAT-488; Sal I and Stu I, pAPPCAT-364; Xba I, pAPPCAT-303; Sal I and Eag I, pAPPCAT-204; Sal I and Nar I, pAPPCAT-96. pAPPCAT-77 and pAPPCAT-45 were created by digesting pAPPCAT-488 with Sal I

followed by a partial digestion with Bsa I and Pvu II respectively. After complete digestion of the plasmids, the ends were filled with the Klenow fragment of DNA polymerase I when necessary and ligated with T4 DNA ligase. The plasmid pBLCAT5 contains the promoter (-105 to +51) of the herpes simplex virus (HSV) thymidine kinase (TK) gene introduced in the BamHI/Bgl II sites of pBLCAT6, upstream of the CAT reporter. The plasmids pBCAT, pA2BCAT, pRSVAP2, pPac0, and pPacSp1 were gracious gifts from Dr. Robert Tjian (University of California at Berkeley). pBCAT contains the TATA box from the E1B gene of adenovirus introduced upstream of the CAT gene (19). pPac0 is a vector designed for expression of cloned genes in *Drosophila* cell lines. Whereas pPac0 does not carry any eukaryotic open reading frames, pPacSp1 contains the cDNA sequences coding for the human transcription factor Sp1 introduced downstream of the *Drosophila melanogaster* actin 5C promoter. The cDNA clone encoding the 43 kDa human USF-1 introduced into pBKS1, pΔ i2, was a generous gift from Dr. Michèle Sawadogo of the M.D. Anderson Cancer Centre, Houston, TX (20). This clone was used to produce recombinant USF-1 in a coupled transcription/translation assay in combination with T7 RNA polymerase. The eukaryotic expression plasmid pCXUSF, containing a cDNA coding for USF-1 driven by the cytomegalovirus (CMV) promoter, was kindly provided by Dr. Robert G. Roeder of the Rockefeller University, New York, NY (21). In another series of experiments, the cDNA coding for USF-1 was removed from pΔi2 by Xho I / EcoR I restriction digest and introduced in the Xho I / EcoR I sites of the expression vector pcDNAI/amp (Invitrogen) creating the expression vector pCMVUSF-1. Plasmid p2XAdMLTF/TKCAT, kindly supplied by Dr. Howard C. Towle from University of Minnesoata, MN, contains 2 recognition elements for USF from the adenovirus major late promoter (-69 to -49) introduced upstream of the minimal promoter from the TK gene of HSV (22). The vector pUT535, expressing the β-galactosidase gene under the control of a CMV enhancer/promoter, was a generous gift from Dr. Ken Hastings (McGill University, Montreal).

4.2.3.3 Oligonucleotides

Single stranded oligonucleotides were synthesized by either Dalton Chemical Laboratories Inc. (North York, ON, Canada), General Synthesis and Diagnostics (Toronto, ON, Canada) or the Sheldon Biotechnology Center (Montreal, QC, Canada). The double-stranded oligonucleotide Sp1 was purchased from Promega. Only the sequences from the coding strand are indicated for oligonucleotides used in EMSA experiments. The underlined regions mark the core binding site of the various transcription factors. Positions relative to the major transcriptional start site of the APP gene are indicated in parentheses. Mutations introduced in the *APP* promoter elements are in ***bold italics***.

SP1: 5'-ATTCGATC<u>GGGGCGGGG</u>CGAGC-3';

Ad-USF: 5'-TGTAGGC<u>CACGTG</u>ACCGGGT-3';

APP-GC1 (-97 to -78): 5'-CGGCG<u>CCGCTAGGGG</u>TCTCT-3';

APP-E1 (-56 to -37): 5'- GCCGGATC<u>AGCTG</u>ACTCGCC-3';

APPSP1 (-71 to -50): 5'-TGCCGAGC<u>GGGGTGGGC</u>CGGAT-3';

SP1-USF (-71 to -37): 5'-TGCCGAGC<u>GGGGTGGGC</u>CGGAT<u>CAGCTG</u>ACTCGCC-3';

APPUSF-M1-1 (-59 to -36): 5'-TGGGCCGGATC*GAT*TGACTCGCCT-3';

APPUSF-M1-2 (-36 to -59): 5'-AGGCGAGTCA***ATC***GATCCGGCCCA-3';

APPUSF-M2-1 (-59 to -36): 5'-***CA***GGCCGGATC***GAT***TGACTCGCCT-3';

APPUSF-M2-2 (-36 to -59): 5'-AGGCGAGTCA***ATC***GATCCGGCC***TG***-3';

APPSP1-M1-1 (-71 to -48): 5'-TGATCCGGCC***AACA***GGCCGGATCA-3';

APPSP1-M1-2 (-48 to -71): 5'-TGATCCGGCC***TGTT***CCGCTCGGCA-3';

APP-395FPst1 (-403 to -385): 5'-CACCCTAGC***T***GCAGTCCTT-3';

APPRBamH1 (+93 to +111): 5'AGATCTGGATCCGCCGCGT-3';

ASAMY (+71 to +100): 5'-CGCCGCGTCCTTGCTCTGCCCGCGCCGCCA-3'

4.2.3.4 Recombinant PCR mutagenesis

PCR buffer composition (final concentration): 50 mM KC1, 10 mM TRIS.HC1 (pH 8.3 at room temperature), 1.5 mM $MgC1_2$, 0.2 mM dNTP, 0.01% (w/v) gelatin, 10% formamide and 50 pmol of each primer. PCR mixtures were denatured for 10 minutes at 94ºC before the thermostable Taq polymerase was added. Typically, each PCR reaction contained 4 units/100 μl of reaction volume. The reactions were amplified for 24 cycles in the following condition: 94ºC 1 minute, 55ºC 1 minute and 72ºC 1 minute. Recombinant PCR is a two-step process. In the first step, two independent PCR reactions are performed using the same template. The reverse primer of reaction number one is complementary to the forward primer of the second reaction, creating an overlap. In the second step, the overlap between the two PCR products allows one to combine them in presence of the forward primer of reaction number one and the reverse primer of reaction number two in order to generate a longer PCR product. Introduction of mismatches in the middle of the overlapping oligonucleotides generates point mutations in the sequence of choice. The introduction of mutations in the Sp1 element is used in the following as an example for all the mutations introduced in the *APP* promoter. 500 ng of EcoR I linearized pAPPCAT-3699, which contains the wild type *APP* promoter, was used as the template. The first reaction contained the forward primer APP-3951Pst1 (this oligonucleotide contains one mismatch base pair in order to introduce a Pst I restriction site to facilitate the cloning of the final PCR products) and the reverse primer APPSP1-M1-2 generating of product of 355 bp. The second PCR reaction contained APPSP1-M1-1 as the forward primer along with APPRBamH1 as the reverse primer giving a fragment of 182 bp. The reaction products were ran on a 1.5% agarose gel and purified using the Qiaex gel extraction kit from Qiagen. In the second step of recombinant PCR, 100 ng of each PCR products were combined with forward primer APP-395FPst1 and reverse primer APPRBamH1 to generate a fragment of 514 pb. The final PCR product was gel purified and digested with Pst I and BamH I and introduced in the Pst I and BamH I sites of vector pBLCAT6 creating pAPPCAT-

395mSp1. Mutations in the USF element were generated in a similar fashion using instead the overlapping primers APPUSF-M1-1 and APUSF-M1-2 creating pAPPCAT-395mUSF. In order to create the double mutant pAPPCAT-395mSU, pAPPCAT-395mSP1 was linearized with the restriction enzyme EcoR I and used as template with the overlapping oligonucleotides APPUSF-M2-1 and APPUSF-M2-2. All mutations introduced in the *APP* promoter were verified by sequencing. The plasmids pAPPCAT-96mSp1, pAPPCAT-96mUSF, pAPPCAT-96mSU, pAPPCAT-77mSp1, pAPPCAT-77mUSF and pAPPCAT-77mSU were generated from pAPPCAT-395 derivatives as described previously for the wild type promoter.

4.2.3.5 Calcium phosphate transfection

Derivatives of the plasmid pBLCAT6 were transfected using a modification of the standard calcium phosphate protocol (22). Typical transfections used 10 ug of plasmid DNA for NG108-15 cells and 15 ug for HepG2 cells. Plasmid DNA was purified using the Qiagen Plasmid Maxi Kit following the manufacturer's specifications (Qiagen, Mississauga, ON, Canada). Cells were incubated for 16 hours at 35ºC in 3% CO_2 incubator, rinsed with 5 ml of phosphate-buffered saline (PBS) and fresh medium was added. Cells were then transferred to a 5% CO_2 incubator for 24 to 48 hours. Transfection conditions were optimized for each cell line. Refer to the figure legends for the specific transfection conditions of each experiment. Each transfection included 1 ug of the reporter plasmid pUT535 which contains the β-galactosidase gene under the control of the CMV enhancer/promoter region to standardize for transfection efficiencies.

4.2.3.6 Transfection of *Drosophila* S2 cells

Drosophila S2 cells were transfected using lipofecton™. 1 X 10^6 cells per 2 ml were plated in a 6-well tissue culture plate (Falcon). Cells were allowed to attach for 3-4 hours. For each well to be transfected, 2 polystyrene tubes were prepared containing 500 ul of serum free Shields and Sang M3 insect medium (Sigma-Aldrich, St.-Louis, MO, USA). One tube contained 10 μg of the DNA to be transfected whereas the other tube contained 10 μg of lipofecton™ (for detailed information on each transfection, see Figure legends). The DNA and lipofecton™ were gently combined and allowed to form a complex for 15 to 30 minutes at room temperature. The medium was removed from the wells and the cells were carefully rinsed with 2 ml of PBS, before the DNA- lipofecton™ mixture was gently added to the cells. Cells were incubated at room temperature on the bench for 6 hours before the DNA- lipofecton™ mixture was aspirated and replaced with 2 ml of Shields and Sang M3 medium supplemented with 10% FCS.

4.2.3.7 Reporter gene assays

The CAT reporter system was used to measure the activity of various constructs containing sequences from the APP promoter. Transfected cells were washed with ice cold PBS, removed from the culture dish mechanically, collected by centrifugation (1 000 RPM) and resuspended in 100 μl of TRIS-HC1 0.25 M pH 7.5. Cellular extracts from transfected cells were generated by three cycles of freezing and thawing. The cellular debris were removed by centrifugation at 13 000 RPM for 5 minutes. CAT and β-galactosidase assays were performed as described in Sambrook et al., 1989. Briefly, 50 μl of cellular extracts were incubated with 50 μl of TRIS 1 M pH 8.0, 20 μl of acetyl co-enzyme A, 60 mM, and 1.0 μl of ^{14}C-chloramphenicol, 0.1mCi/ml, for 30 minutes at 37ºC. Acetylated and unacetylated ^{14}C-chloramphenicol was extracted with 1.0 ml of ethyl acetate. 900 μl of ethyl acetate (upper phase) was collected and evaporated under vacuum for 30 minutes. Acetylated and unacetylated ^{14}C-chloramphenicol was resuspended in 20 μl of ethyl acetate. Acetylated ^{14}C-chloramphenicol was separated from unacetylated ^{14}C-

chloramphenicol by thin layer chromatography in 95:5 chloroform:methanol solvent. The 20 μl of ethyl acetate was spotted 3 μl at a time on a thin layer chromatography plate. The solvent front was allowed to migrate up the plate in a chromatography chamber for 20 to 30 minutes. Thin layer chromatography plates were exposed to a PhosphorImager screen overnight. The percentage of acetylation was determined using the ImageQuant software from Molecular Dynamics (Sunnyvale, CA, USA). For β-galactosidase assays, 30 μl of cellular extracts were incubated with 3 μl of 100X Mg solution (0.1 M $MgC1_2$, 4.5 M β-mercaptoethanol), 66 μl of o-nitrophenyl-β-D-galactopyranoside (ONPG), and 201 μl of 0.1 M sodium phosphate. The reactions were incubated for 10 to 30 minutes at 37ºC. The reactions were stopped by adding 500 μl of 1.0 M Na_2CO_3. Reactions were quantitated by measuring the optical densities at 420 nM. The percentages of acetylation from the CAT assays were standardized to β-galactosidase activity.

4.2.3.8 Preparation of nuclear extracts

Crude nuclear extracts were prepared using a modification of Dignam et al. (23). Cells were washed twice with cold PBS before they were harvested from the tissue culture dishes. Cells were then centrifuged at 1 000 RPM for 10 minutes, the pellet was resuspended in 5-packed cell pellet volume of buffer A (10 mM HEPES pH 7.9 at 4ºC, 1.5 mM $MgC1_2$, 10 mM KC1 and 0.5 mM DTT) and incubated for 5-10 minutes on ice before NP-40 was added at a final concentration of 0.5% (v/v). Cells were lysed by pipeting 4-5 times in a 10 ml tissue culture pipette. Lysis was verified under the microscope. The lysate was centrifuged for 2 to 3 minutes at 3 000 rpm at 4ºC, the supernatant discarded and the nuclei pellet rinsed with 5-packed volumes of buffer A without NP-40. The nuclei were centrifuged for 2-3 minutes at 3 000 RPM and resuspended in 1-packed volume of buffer B (20 mM HEPES pH 7.9, 25% glycerol, 0.42 M NaC1, 1.5 mM $MgC1_2$, 0.5 mM phenylmethylsulfonyl fluoride (PMSF) and 0.5 mM DTT). Nuclei were lysed by passing 10-15 times in a Kontes all glass Dounce homogenizer with a B type pestle. Extent of the lysis was verified under the microscope. The nuclei was gently agitated on ice for 30

minutes. The lysate was then centrifuged at 25 000 G for 20 minutes at 4ºC. The supernatant was carefully removed and dialysed twice (2 hours and overnight) against 50 volumes of buffer C (20 mM HEPES pH 7.9, 20% glycerol, 0.1 M KC1, 0.2 mM EDTA, 0.5 mM PMSF and 0.5 mM DTT). The dialysate was centrifuged at 25 000 G for 20 minutes and the supernatant was quickly aliquoted and frozen at -80ºC. All manipulations were performed on ice or at 4ºC. DTT and PMSF were added fresh. Protein concentrations were determined using the Pierce BCA Protein Assay Reagent kit (Pierce). HeLa nuclear extracts were purchased from Promega.

4.2.3.9 Electrophoretic mobility shift analysis

For annealing, equimolar amounts of complementary single stranded oligonucleotides were heated at 95ºC for 10 minutes and allowed to cool down at room temperature over a period of 2 hours in 10 mM TRIS-HC1 pH 7.8, 1 mM EDTA and 100 mM NaC1. Double stranded oligonucleotides were end-labeled using γ-^{32}P-ATP and T4 polynucleotide kinase. When end-labeled restriction fragments were used, they were labeled according to Sambrook et al. (1989) using the Klenow fragment of DNA polymerase I, α-^{32}P-dCTP and/or α-^{32}P-dGTP and purified on a polyacrylamide gel (10%). Binding reactions were performed at room temperature in a final volume of 15 μl (4% glycerol, 1.0 mM MgC1$_2$, 0.5 mM dithiothreitol, 50 mM NaC1 and 10 mM TRIS-HC1 pH 7.5) in presence of 0.5-1.0 μg of poly (dI-dC) – poly (dI-dC) and 6 to 8 μg of nuclear proteins. Probes (0.1 ng; 3.5-5.0 x 10^5 cpm/ng) were added 10 minutes after the nuclear extracts and the binding reactions were allowed to proceed for 20-30 minutes at room temperature. The concentration of DTT was increased from 0.5 mM to 4.0 mM for the studies on the USF transcription factor. For binding reactions involving the AP-2 transcription factor, the following binding buffer was used: 25 mM TRIS-HC1 pH 8.0, 6.25 mM MgC1$_2$, 0.5 mM EDTA, 0.5 DTT and 10% glycerol. When present, specific cold oligonucleotide competitor was added 5 minutes before the probe. For "supershifts", polyclonal antibodies (Santa Cruz Biotechnology Inc.) were added to the binding reaction after 20 minutes and incubated at room temperature for

an additional 30 minutes. Samples were analyzed on a 5% polyacrylamide gel containing 2.5% glycerol. Electrophoresis was performed at 10 mA for 1.5-2.0 hours in 0.5X TRIS-borate-EDTA (TBE) (Sambrook et al., 1989) with circulating cold water. The gels were dried and exposed to Kodak XAR-5 film for 16 hours at -70ºC or exposed to phosphor screens and analyzed on a Molecular Dynamic PhosphorImager using the Imae Quant software.

4.2.3.10 In vitro DNase I footprinting anaylsis

DNase I footprinting analysis of the *APP* proximal promoter was performed on an end-labeled Xma III / BamH I (-203 to +100) fragment. In order to specifically label one strand, the plasmid pUC18BamHI3.8 was either digested with the restriction enzymes Xma III (for non-coding strand labeling) or BamH I (for coding strand labeling). The restriction fragments were end-labeled with α-^{32}P-dCTP and α-^{32}P-dGTP using Klenow polymerase (Sambrook et al., 1989). The end-labeled fragments were further digested in order to generate the 303 bp Xma III / BamH I fragment. The restriction fragments were gel purified before they were subjected to DNase I footprinting analysis. The binding of nuclear proteins to the end-labeled Xma III / BamH I fragments was performed under the same buffer conditions as those of the electrophoretic mobility shift analysis. The volume of the reaction was increased to 50 µl with 20 000 cpm of probe (3-10 X 10^4 cpm/ng) in presence of 5 µg of poly (dI-dC) – poly (dI-dC) and 60 µg of crude nuclear extract proteins. When purified Sp1 was used, the reaction included 0.05 µg poly (dI-dC) – poly (dI-dC) and 160 ng of Sp1. After 20 minutes of binding, 50 µl of solution A (5 mM CaC1$_2$, 10 mM MgC1$_2$) was added and DNase I digestion was performed at room temperature for 60 seconds and stopped with 100 µl of solution B (20 mM EDTA, 15 SDS, 100 mg/ml tRNA). The products were phenol-chloroform extracted, ethanol precipitated and then resolved on a 6% denaturing polyacrylamide gel. The gels were dried and exposed to Kodak XAR-5 film for 16 hours at -70ºC.

4.2.3.11 In Vitro Transcription and Translation-TNT^{TM}

Coupled Reticulocyte Lysate System from Promega was used as recommended by the manufacturer to synthesize *in vitro* the 43 kDa USF-1 protein from 2 μg of plasmid pΔi2. Efficiency of USF-1 synthesis was verified by radioactivity labeling an aliquot with ^{35}S-methionine and the molecular weight of the translation product was confirmed by SDS-PAGE.

4.2.3.12 In Vitro Transcription

Run-off *in vitro* transcription assays were performed using the HeLa Nuclear extract *in vitro* transcription system (Promega). Template pAPPCAT-96 was linearized by restriction enzyme digest with Msc I, resulting in a transcript of 638 nucleotide fragment after run-off transcription assays. The final reaction contained 1 μg of linearized templates incubated in 8 mM Hepes, 40 mM KC1, 0.08 mM EDTA, 1.25 mM DTT, 3 mM MgC1$_2$, 8% glycerol, 0.4 mM rATP, 0.4 mM rUTP, 0.016 mM rGTP and 1 μl of α-^{32}P-GTP (3000 Ci/mmol, 10 mCi/ml) in a volume of 25 μl. The amount of protein varied from 20 μg to 80 μg per 25 μl. The amount of protein varied from 20 μg to 80 μg per 25 μl reactions (refer to specific figure legends). The transcription reactions were incubated at 30ºC for 60 minutes and stopped by adding 175 μl of 0.3 M TRIS-HC1 (pH 7.4), 0.3 M sodium acetate, 0.5% SDS (w/v), 2 mM EDTA and 3 μg/ml tRNA. The resulting 200 μl were extracted with phenol:chloroform:isoamyl alcohol (25:24:1) and the aqueous phase was ethanol precipitated by adding 500 μl of ice cold ethanol. After 30 minutes at -80ºC, the precipate was collected by centrifuging at 12 000 g for 20 minutes at 4ºC. The supernatant was carefully decanted and the pellet was vacuum dried for 5-10 minutes. The pellet was resuspended in 10-20 μl of RNase free water. One volume of loading dye (98%

formamide, 10 mM EDTA, 0.1% xylene cyanol and 0.1% bromophenol blue) was added to the samples before they were incubated at 95ºC for 5 minutes and loaded onto a denaturing polyacrylamide gel [6% acrylamide (39:1 acrylamide:BIS), 7.0 M urea and IX TBE (0.09 M TRIS-borate, 0.002 M EDTA)]. The gel was run in IX TBE at 250 volts for 2.5 hours, dried and exposed to XAR-5 film or to phosphor screens and analyzed on a Molecular Dynamic PhosphorImager using the ImageQuant software.

For run-on *in vitro* transcription assays, 2 μg of supercoiled pAPPCAT-96, pAPPCAT-96mSP1, pAPPCAT-96mUSF and pAPPCAT-96mSU were used as templates, radioactive nucleotides were omitted, consequently the final concentration of cold rGTP was increased to 0.4 mM, the amount of proteins as 160 μg in a final volume of 50 μl. After ethanol precipitation, the products of run-on *in vitro* transcription were subjected to primer extension assays in order to determine the transcriptional start sites of the *APP* promoter mutants.

4.2.3.13 Primer extension assays

Primer extension assays were performed according to Sambrook et al., 1989. 1×10^5 cpm of the end-labeled primer ASAMY (Anti-Sens AMYloid), complementary to APP and mRNA (+71 to +100), were added either to the products of run-on *in vitro* transcription assays or to 50 μg of total mRNA and ethanol precipitated by adding 0.1 volume of 3.0 M sodium acetate (pH 5.2) and 2.5 volume of ethanol. The mixture was stored at -20ºC for 30 minutes. The RNA-primer precipitate was collected by centrifuging at 12 000 g for 20 minutes at 4ºC. The pellet was rinsed with ice cold 70% ethanol, air dried and dissolved in 30 μl of hybridization buffer (40 mM PIPES (pH 6.4), 1 mM EDTA (pH 8.0), 0.4 NaCl and 80% formamide), heated at 85ºC for 10 minutes and then quickly transferred to 30ºC for 8 to 12 hours. After the hybridization was complete, 170 μl of RNase free water and 400 μl of ethanol were added and the mixture was stored at -20ºC for one hour. The precipitate was collected as previously described and dissolved in 20 μl of reverse transcription buffer

(50 mM TRIS-HCl (pH 7.6), 60 mM KC1, 10 mM MgCl$_2$, 1 mM of each dNTP, 1 mM DTT, 1 unit/μl of placental RNase inhibitor (Gibco-BRL) and 50 μg/ml of actinomycin D). Murine reverse transcriptase, (50 units per reaction) was added and polymerization was allowed to proceed for 2 hours at 37ºC. To terminate the reaction, 1 μl of 0.5 M EDTA was added, followed by 1 μl of pancreatic DNase-free RNase (final concentration 5 μg/ml). The reaction was incubated for 30 minutes at 37ºC before 150 μl of TNE (10 mM TRIS-HCl (pH 7.6), 0.1 M NaCl and 1 Mm EDTA). The resulting 200 μl was extracted with phenol:chloroform (1:1) (200 μl). The aqueous phase was ethanol precipitated as previously described with 500 μl of ice cold ethanol. The pellet was dissolved in 4 μl of TE (pH 7.4). After completely resuspending the pellet, 6 μl of formamide loading buffer (80% formamide, 10 mM EDTA (pH 8.0), 1 mg/ml xylene cyanol and 1 mg/ml bromophenol blue) were added and the samples were heated at 95ºC before being loaded onto at a denaturing polyacrylamide gel [6% acrylamide (39:1 acrylamide:BIS), 7.0 M urea and 1X TBE]. Sequencing reactions (performed as recommended by the manufacturer, NEB) was present on the gels and were used to determine the size of the primer extension assay products.

4.2.3.14 In vivo footprinting analysis

In vivo footprinting analysis using DMS (0,02%) and UVC (1500 J/m^2) of the APP gene core promoter from −695 to +45 was conducted on human primary neuronal cells and astrocytes. The *APP* gene is expressed in these two types of cells. They were grown as a monolayer to about 80% confluence in 150-mm Petri dish at 37°C in supplemented Dulbecco's modified Eagles medium (DMEM). Cells were then treated *in vivo* with DMS (0,02%) and irradiated with ultraviolet C radiation (UVC; 1500 J/m^2) as previously described (14). LMPCR procedures were carried out as previously described (14) using Pfu exo- DNA polymerase on both first extension and PCR amplification. The amplification of GC-rich regions (primer sets 1, 2, 7 and 8; Table I) was made using betain (Sigma-Aldrich)

at 1.5 M concentration. The addition of aqueous betain to a PCR mixture has been reported to reduce the base pair composition dependence on DNA strand melting. The chemically cleaved G, A, T+C and C samples carried out on purified genomic DNA from human peripheral blood lymphocytes were included along with the other samples in the LMPCR assays as sequence markers.

4.2.4 Results

4.2.4.1 Transient transfection of progressive deletions of the APP promoter

As a first step, we performed transient transfection experiments using progressive deletions of the *APP* promoter in order to confirm previous experiments from Pollwein (14) and Quitschke (11) and to validate the promoter activity in the cells used in this report. As illustrated in Figure 1, results of transient transfection assays in both NG108-15 neuroblastoma cells and HepG2 hepatoma cells show that sequences present upstream of the first −96 bp do not significantly affect the activity of the promoter. Deletion of the sequences present between −96 and −77 (pAPPCAT-77) reduced the activity of the *APP* promoter to approximately 10% in both cell lines. However, deletion of the region from −364 to −96 did not lead to a significant change in the average CAT activity (pAPPCAT-96), but obviously reduced the variability of promoter activity as can be seen by the smaller standard errors.

4.2.4.2 Sp1 recognizes the consensus Sp1 element upstream of the USF binding site but fails to transactivate the APP promoter in Drosophila S2 cells

We investigated whether the transcription factor Sp1 could bind the consensus Sp1 sequence present in the *APP* minimal promoter. In order to determine the sequences recognized by the Sp1 protein in the *APP* promoter, we performed DNaseI footprinting analysis in the presence of recombinant SP1 protein using the end-labeled genomic DNA fragment-spanning the region from -203 to $+100$ relative to the major transcriptional initiation site (TIS). Three regions of the *APP* promoter were protected from DNase I digestion (Fig. 2A). Region -67 to -53 corresponds to the consensus SP1 element upstream of the USF binding site in the minimal promoter. The second region, -137 to -122, also corresponds to a consensus Sp1 binding site as revealed by analysis with the TRANSFAC data-base. The third region could not be precisely determined due to its relative position in the gel. In the present study, we focused our attention on the regions between -96 and the TIS in order to evaluate their role in the APP basal expression. EMSA was used to confirm that Sp1 could bind its consensus element in the proximal *APP* promoter (-67 to -53). Incubation of a labeled probe bearing the entire Sp1 element (APPSP1, -71 to -50), with HeLa nuclear extracts, which contain Sp1 activity (24), yielded a specific DNA-protein complex (Fig. 2B, lane 2) that was gradually abolished by the addition of a 20, 50 and 100-fold molar excess of an unlabeled competitor oligonucleotide bearing the consensus sequence for Sp1 (data not showed). Furthermore, incubation of the EMSA reaction with antibody directed against Sp1 further slowed down the migration of the APPSP1-Sp1 complex (supershift, data not showed). Minor complexes could be observed that were not competed with SP1. These complexes were often observed with the APPSP1 oligonucleotide and could not be competed with unlabeled APPSP1 either, suggesting that they were non-specific in origin (data not shown and Fig. 4A).

We co-transfected the minimal (pAPPCAT-96) or the CTCF-deleted (pAPPCAT-77) *APP* promoter fragments along with an expression vector for Sp1 (pPacSp1) in *Drosophila* Melanogaster S2 cells that are devoid of Sp1 activities (25). As shown in Figure 2C, Sp1 failed to transactivate both *APP* promoters whereas similar conditions increased more than two fold the thymidine kinase promoter (pTKCAT) from the herpes simplex virus (HSV), which contains two Sp1 binding sites (26). These results suggest that Sp1 can recognize the

element located between −67 and −53, but that it might have roles other than direct transactivation of *APP*.

4.2.4.3 Binding of USF and Sp1 does not occur simultaneously on the same molecule in EMSA

In an attempt to investigate the dynamics of the binding of Sp1 and USF to their own but closely located binding sites on the *APP* promoter, we used recombinant Sp1 protein and *in vitro* synthesized USF-1 in EMSA with the SP1-USF oligonucleotide as probe. Both Sp1 and *in vitro* synthesized USF-1 recognized the SP1-USF oligonucleotide and formed distinct complexes (Fig. 3A, lanes 2 and 6). When both proteins were added at the same time to the binding reaction, the complexes were the sum of the individual complexes formed with the SP1-USF probe (Lane 5). The order in which Sp1 or *in vitro* synthesized USF-1 were added did not affect the intensity of the complexes formed (Lane 3 and 4). In the conditions used, we did not detect a slower complex containing both Sp1 and USF bound to the same DNA molecule.

To confirm the results with recombinant proteins, we used a similar approach using HeLa nuclear extracts which contain both Sp1 and USF activity (26;27). Three major complexes were observed (Fig. 3B, lane 1). We used competition EMSA to visualize if removal of either Sp1 or USF had an effect on the recognition of the adjacent element (Fig. 3B). Competition with the unlabeled Sp1 oligonucleotide, which contains the Sp1 biding site from the SV40 promoter, abolished the slower migrating complex without significantly affecting the other complexes (Fig. 3B, lanes 2-4). Furthermore, the complex which was competed by the SP1 oligonucleotide could be supershifted by an antibody directed against Sp1 (lane 8) confirming the presence of Sp1 in the slower migrating complex. In the converse experiment, competition with an oligonucleotide (Ad-USF) containing the adenovirus major late promoter USF binding site gradually abolished the two faster

migrating complexes without affecting the Sp1 complex (lane 5-7). Taken together, these results suggest that Sp1 and USF recognize in an independent manner, and maybe compete with each other to bind to their respective adjacent elements in the *APP* promoter.

4.2.4.4 Introduction of mutations in Sp1 and USF elements

In our efforts to characterize the individual function of the elements present in the minimal *APP* promoter, we introduced mutations into the *APP* promoter to study their effects in the context of various promoter sequences. As a first step, we performed EMSA to confirm that the mutations to be introduced in the Sp1 and USF elements would not be recognized by their respective transcriptional activator. As observed in Figure 4A and B, both mutant oligonucleotides failed to compete with their unmutated probes whereas, in similar conditions, the oligonucleotide without mutation abolished the formation of the complexes. We concluded from these competition studies that these mutations, when introduced into the *APP* promoter, would prevent the binding of Sp1 and USF to their respective elements.

4.2.4.5 Mutation in Sp1 and USF elements do not affect the selection of multiple transcriptional start sites

We verified that the mutations introduced in the Sp1 and USF elements did not affect the utilization of the major transcriptional start sites of the *APP* promoter. In order to detect the transcriptional start sites of the various constructs of the minimal *APP* promoter, we performed run-on in vitro transcription assays, using HeLa nuclear extracts, followed by primer extension assays. As seen in Figure 4C, mutations in the Sp1 and USF elements (lanes 2-4) of the *APP* promoter did not affect the utilization of transcriptional start site when compared to the wild type promoter (lane 1). The transcriptional start site used in vitro by the *APP* promoter with nuclear extracts from HeLa cells are identical to the ones

used in vivo by HepG2 cells (lane 5). When in vitro transcription templates were used for primer extension assays, we observed 2 close bands corresponding to the major transcriptional start site (+1) and another single one corresponding to the alternative transcriptional start site (–4). Determination of the ratio (average of three independent experiments) between the two transcriptional start sites (+1/–4) revealed a slight reduction in the constructs carrying a mutation in the USF element. Furthermore, we constantly observed a decrease in the efficiency of in vitro transcription run-on and run-off assays with constructs carrying mutations (Fig. 4C and data not shown) reflecting an effect on transcriptional activity. These results suggest that Sp1 and USF are not involved in start site selection but rather in transcriptional activity of the *APP* promoter.

4.2.4.6 The Sp1 and USF are necessary for full promoter activity

In order to investigate the effects of the interaction of Sp1 and USF with their elements on the *APP* promoter, we performed transient transfection experiments in NG108-15 and HepG2 cells using both deletion and mutant constructs. First, we show that USF-1 can transactivate the minimal *APP* promoter in a concentration dependent manner in HepG2 cells and that this activation is dependant on the presence of a functional USF element (Fig. 5A). In co-transfection assays, increasing amounts of the pCMVSF-1 expression vector activated pAPPCAT-96 up to two-fold (Fig. 5A, 5μg of pCMVSF-1). Mutations introduced in the USF element (pAPPCAT-96mUSF) reduced the activity of the promoter to 40% of the wild type promoter and abolished the two-fold activation generated by the co-tranfection with 5μg of pCMVSF-1.

Then we analyzed the effect of mutating the Sp1 and USF elements separately or in combination in context of full or deleted *APP* promoter (Fig. 5B and 5D). As shown in Figure 5B, pAPPCAT-96mSp1 activity was reduced to 67% whereas pAPPCAT-96mUSF had a more important effect reducing activity to 42% when compared to pAPPCAT-96 in NG108-15 cells. In HepG2 cells, similar results were obtained: pAPPCAT-96mSp1 had a

reduced activity of 59% and pAPPCAT-96mUSF, 40%. The presence of both mutations in the double mutant pAPPCAT-96mSU further reduced the relative promoter activity to 22.3% in NG108-15 cells and to 14.3% in HepG2 cells. pAPPCAT-488 had a level of activity similar to pAPPCAT-96. The relative activity of pAPPCAT-395mSU in NG108-15 was 34 % and in HepG2 32%. These differences between pAPPCAT-96mSU and pAPPCAT-395mSU reinforce the assumption that the sequences located upstream of −96 could participate in *APP* gene regulation and that CTCF proteins binding to the proximal GC-element cannot be held responsible for the entire activity of the *APP* promoter as previously proposed (15).

To further characterize the role of Sp1 and USF elements, we studied the functional activity of the mutants in the absence of the GC-element of the *APP* promoter (that bind CTCF) in NG108-15 (Fig. 5C and 5D). The wild type promoter present in pAPPCAT-77 had a relative CAT activity of 7.4% when compared to pAPPCAT-96. Mutations in the Sp1 or USF elements lead to an increase to 10.7% and a decrease to 4.8% in the relative CAT activity respectively. The activity of the double-mutant pAPPCAT-77mSU was similar to pAPPCAT-77mUSF (4.8%), suggesting that the proteins recognizing the Sp1 element need the presence of the GC-element binding proteins to properly activate the *APP* promoter. These results illustrate well that both the Sp1 and USF elements are essential and that they cooperate with CTCF binding the GC-element, located between −96 and −77, to confer full activity to the *APP* promoter.

4.2.4.7 In vivo footprinting of the APP gene promoter by LMPCR

In order to validate the previous results obtained by us and others, we used an *in vivo* footprinting approach to detect DNA-protein interactions at the *APP* gene promoter level in living cells (28). DNA has an intrinsic susceptibility to damaging agents that is dependent on its chemical composition. The *in vivo* footprinting analysis is based on the fact that the

susceptibility of one nucleotide to be affected by a damaging agent can be altered due to interactions of the DNA with intra-nuclear proteins or by DNA secondary structures. Consequently, when nuclear DNA is treated with a particular damaging agent, it can be either protected (negative footprint; protection) or over-exposed (positive footprint; hyper-reactivity or hypersensitivity). By comparing the damage frequency induced in a native nuclear context with that of naked DNA, DNA-protein interactions as well as DNA secondary structures can be detected. In this study, we used two damaging agents, DMS and UVC radiations to study the *cis*-acting regulatory sequences from the *APP* gene promoter (from −695 to +45) that are recognized by nuclear proteins in primary neurons and astrocytes. These cells were chosen because they are at the origin of the alterations in the *APP* gene regulation in the Alzheimer disease.

As observed in figure 6 A and B (summarized in C), the comparison of the band intensity from the *in vivo* and *in vitro* treatments revealed many protected and hypersensitive nucleotides which are identical in both human neuronal cells and astrocytes. This is in agreement with the almost equivalent expression of the *APP* gene in these cells. First, we clearly demonstrate that the binding domains for CTCF (GC-element: −93 to −82), Sp1 (−63 to −56) and USF (−53 to −42) show protein protection from damaging agents, confirming the importance of these factors and their binding elements for the *APP* gene regulation in normal living cells. However, many other DNA-protein interactions are also detected on the proximal promoter of APP gene from −170 to −20. The protection visualized using both DMS and UVC radiation at position −130 to −123 on the bottom strand underline the possible interaction of Sp1 at its consensus binding site located between −133 and −124 as observed in Figure 2A. The sequence analysis revealed the possible involvement of another Sp1 binding site (−113 to −106), of a NF-kB (−110 to −101), and a CAC-binding protein (−41 to −30). We also found a clear footprint on the bottom strand at position −168 to − 160 where the sequence does not match with the binding sequence of known transcription factors.

4.2.5 Discussion

The *APP* gene is expressed ubiquitously. The sequence surrounding the major transcriptional start site has a high percentage of GC content (85%) and does not contain a TATA-box. Sequences responsible for the transcriptional activity of the promoter have been narrowed down to the sequence downstream –96. We first confirm these previous studies in the context of our cell types (Fig. 1). We demonstrate that two elements of the *APP* gene promoter were recognized by the Sp1 (–63 to –55) and USF (–50 to –43) transcription factors. Of the two, only USF was capable of transactivating the *APP* promoter in transient co-transfection assays but both were necessary for full activity in the two different cells lines.

The E box found in the proximal promoter region of the *APP* gene has previously been shown to be important for expression of the human and rodent genes (4-6;11;12;14). EMSA studies (5;11;12;14), DNase I footprinting (4-6;12), and methylation interference (5) had revealed that a sequence of 12 base pairs (–53 to –42) was essential for nuclear factor binding. This motif, which is conserved 100% between human, mouse and rat genes, is also footprinted by nuclear proteins in rodents. Besides activating the adenovirus major late promoter by binding to an E box (29), USF has been shown to regulate a variety of cellular genes through interactions with upstream E boxes (20;30-33). The stimulatory effect of USF may be mediated through direct protein-protein interaction between USF and the transcriptional apparatus (34). It has also been proposed that binding of USF to its recognition sequences may lead to formation of a nucleosome-free region over the promoter (35). Further support for the involvement of USF in chromatin dynamics is provided by the report that USF binds with high affinity to an E box in the DNase I-hypersensitive site HS2 in the β-globin locus control region (32). This might also be the functional mechanism in regulating basal expression of the *APP* gene since a DNase I hypersensitive site has been demonstrated in the proximity of the USF binding site (36).

Our results using recombinant Sp1 in DNase I footprinting analysis and EMSA with oligonucleotide APPSP1 (–71 to –50) are in complete agreement with what was proposed by Pollwein (14) and by Hoffman and Chernak (7) for the rat *APP* promoter. We failed to observe a transactivation of the *APP* promoter in *Drosophila melanogaster* S2 cells, a system from which Sp1 activity has been shown to be lacking (25). One of the reasons for this lack of activation could be that the affinity of Sp1 for the *APP* Sp1 element. Although we did not perform affinity studies, two observations suggest the Sp1 does not have the same affinity for the *APP* Sp1 element when compared with the SV40 Sp1 element. First, intensity of the complex formed between recombinant Sp1 and the SP1 oligonucleotide is stronger than the one observed between recombinant Sp1 and the oligonucleotide APPSP1 (Fig. 2B). Second, in competition EMSA with HeLa nuclear extracts and oligonucleotide APPSP1, the SP1 oligonucleotide totally removed the complex containing Sp1 activity with a 20 fold molar excess whereas for oligonucleotide APPSP1 the complex disappeared with a 50 molar excess (Fig 4A, compare lane 5 to lane 8 and 9). Another factor that could have an influence in our transactivation study is the number of Sp1 binding sites present in the *APP* promoter. Our positive control (pTKCAT), which was activated more than two-fold, contained two Sp1 binding sites (26). The activation of pTKCAT by Sp1 in our transactivation assays is well within what was previously reported under similar conditions (2.8 fold) (25). The presence of only one low affinity binding site for Sp1 in the *APP* promoter could explain why we did not detect any increase in activity during co-transfection studies in *Drosophila* cells. It remains to be seen if transactivation would have been observed within the context of a longer portion of the *APP* promoter, in which there is an additional Sp1-binding site in close proximity (–137 to –122) of what was footprinted *in vivo* (Fig. 6A, B and C).

The activity of both pAPPCAT-77 and pAPPCAT-96 is very similar when transfected in *Drosophila* S2 cells (Fig. 5A) whereas in mammalian cell lines we observed a decrease of 90% in activity (Fig. 1). This suggests that the sequences located between –96 and –77 (GC-element) are not active in *Drosophila* cells. Although the transcriptional machinery is mostly conserved in eukaryotic cells, some components are species specific. One example

is the human TBP associated factor (hTAF) 55 (55 kDa) of TFIID which is absent from *Drosophila* cells and shown to physically interact with Sp1 (37;38). Whether this particularity affected the potential activation of the *APP* promoter by Sp1 is not known.

The role of Sp1 in *APP* regulation might not only be to directly activate transcription. Binding sites for Sp1 in promoters of housekeeping genes, usually located in CpG islands in the chromatin, are protected from de novo methylation which turns off gene transcription (39). Another possible function of Sp1 would be to create an environment free of nucleosomes and facilitate the assembly of a transcriptionally active RNA polymerase II complex (40). Sp1 has been shown to directly interact with two subunits of TFIID, hTAF 135 and hTAF 55 and possibly help in the recruitment of TFIID to TATA-less promoters (38).

The proximity of an Sp1 binding site in the TATA-less human Ha-ras promoter influences the selection of the transcriptional start site (41). In the *APP* promoter, neither the Sp1 nor the USF binding sites affected this choice although we observed a decrease in overall activity of the mutants reminiscent of the results obtained in transient transfection assays. We determined in these assays that both elements were necessary for full activity of the *APP* promoter (Fig. 5B and C). Our approach is the first one that looks at the effects of mutations in the Sp1 and USF elements, separately or in combination, in presence or absence of the GC-element which binds CTCF.

The roles of the Sp1 and USF elements in regulation of the *APP* promoter have been difficult to define. Pollwein (14) observed that mutations introduced into the Sp1 element (which only differs from our mutation by one nucleotide) had no effect on the activity of the *APP* promoter in HeLa cells. Mutations in the USF element (not known at the time to be recognized by USF) reduced the activity to 78% of the wild type promoter. When both mutations were present, the activity of the *APP* promoter fell to 13%, a level similar to

what we observed in human HepG2 cells (14.3%). Pollwein proposed that the close proximity of the Sp1 and USF elements led to competition of the factors in their recognition and regulation of the *APP* promoter and that the factors recognizing these elements could compensate for the absence of each other. When both factors were prevented from interacting with their respective element, the activity of the promoter was compromised. In the rat, opposite results were observed (7). Mutation in the Sp1 element reduced the activity to 70% whereas mutation in the USF element had no effect. They demonstrated the role of USF by mutating the *APP* USF element into a previously described consensus sequence for USF (29) and obtaining higher levels of activation. This study did not look at the consequences on the rat *APP* promoter activity of the effect of mutations in both elements. Our results show that mutations introduced into Sp1 and USF element, alone or in combination, affect the *APP* promoter activity. These differences could be due to the specific changes introduced in the *APP* promoter. Pollwein (14) reported only partial competition in EMSA with an oligonucleotide containing the mutated Sp1 element (GTG to ACA, one zinc finger) whereas in our case the APPSP1m did not interfere with formation of the complex between APPSP1 and Sp1 under similar conditions. Since mutations in both elements have an effect, our results do not support the compensatory mechanism between Sp1 and USF proposed by Pollwein (14), but rather point to an additive mechanism where both the Sp1 and USF elements contribute to *APP* promoter activity. However, our data do not rule out the possible competition between Sp1 and USF for the regulation of *APP* promoter. Whether Sp1 and USF are simultaneously present on the same DNA molecule when they participate in *APP* transcription cannot be determined by transient tranfection assays or EMSA studies which use an excess of DNA molecules. Moreover, *in vitro* studies so far reported, including ours, do not take into consideration DNA topology, the presence of histones and the nearby RNA polymerase II holoenzyme. The LMPCR *in vivo* technology as well cannot investigate the potential presence of two factors on a single DNA molecule because this method analyzes the protein-DNA interactions in a population of cells at the same time.

In contrast to the studies described above, in experiments performed by Quitschke and colleagues (15), mutations introduced into the Sp1 and USF elements had no effect on *APP* promoter activity in the presence of 488 bp of upstream regulatory sequences. In a previous report by the same author(11), mutations introduced into the USF element affected the *APP* promoter activity only in absence of the GC-element (77 bp of upstream regulatory sequences), but not in its presence (488 bp of upstream regulatory sequences), both in transient transfection and *in vitro* transcription assays. These results led to the proposal that the Sp1 and USF elements played no role in *APP* transcriptional regulation in presence of the proteins recognizing the GC-element, CTCF (15). These results contrast with our own and those reported by Pollwein (14). We observed a strong effect on *APP* promoter when mutating both elements in presence of both 96 and 395 bp of upstream regulatory sequences. The high level of conservation of the *APP* promoter between species and the observation of a similar region protected from DNase I digestion argue against the fact that Sp1 and USF elements are not essential. Again, the specific mutations introduced in the Sp1 and USF elements could in part explain these discrepancies. However, in their studies of the *APP* promoter, Quitschke and co-workers used a reporter plasmid carrying two reporter genes under the control of two different promoters located on the same DNA strand. The CAT reporter gene was under the control of the *APP* promoter whereas a β-galactosidase gene was controlled by the chicken β-actin gene to standardize for transfection efficiency. Previously, presence of the SV40 promoter/enhancer sequences downstream of two separate promoters had been shown to activate transcription in a non-specific manner (42). When only the TATA-box of these promoters was present, the SV40 promoter/enhancer had no effect suggesting that the binding sites for proteins located upstream of the TATA-box are essential for this sort of activation. Similarly, presence of β-actin promoter in *cis* could affect the *APP* promoter in an unidentified manner and mask the effect of the mutations. In absence of the GC-element, however, mutations in the USF element affected *APP* promoter activity suggesting that this *cis*-activation needed the presence of CTCF recognizing the GC-element (11).

Our results have revealed that in the presence of the GC-element, the Sp1 and USF elements regulate the *APP* promoter in additive manner (Fig. 5B, 5D). However, in the

absence of the GC-element, there seems to be competition between the proteins recognizing the Sp1 and USF elements since mutating the Sp1 element increased the activity of the *APP* promoter (Fig.5C, 5D and Table 2), although the converse was not true in that mutating the USF element actually reduced the activity of the promoter to levels similar to the double mutant. These results suggest that Sp1 needs the presence of CTCF binding the GC-element to properly activate the *APP* promoter whereas USF does not. Possible roles for Sp1 and USF in *APP* regulation could be in the recruitment of TFIID to the *APP* promoter since both Sp1 and USF have been reported to physically contact hTAF 55, a subunit of TFIID. Recruitment of TFIID to the promoter regions of genes has been shown to be a limiting event in transcriptional initiation. TFIID recognizes the TATA box presents in promoters through its TBP subunit. Although the *APP* promoter does not possess a consensus TATA box, TFIID has been shown to be present in TATA-less promoters in a region located approximately 30 bp upstream of the transcriptional start site and its recruitment seems to involve proteins located upstream of the −30 region which could implicate Sp1 and USF (43).

The investigation of the *APP* promoter using *in vivo* footprinting method revealed many consensus sequences where DNA-protein interactions are present in normal human neurons and astrocytes. We first demonstrated protein interactions occur in the regions which are also DNase I footprinted *in vitro* (Figure 2). The *in vivo* data showing occupancy of these regions in live neurons and astrocytes completely support the implication of USF and Sp1 in the regulation of the *APP* gene. This method also revealed the presence of many other DNA-protein interactions on the *APP* gene promoter. These occurred at consensus sequences for Sp1, NF-kB, CAC-binding protein and in the DNase I-protected domain UE. The identification of each of these factors bound on the promoter and their involvement in *APP* gene expression will have to be determined in future experiments. Our *in vivo* data also point to a potential importance of the region from −170 to −96 where inducers or repressors may act to influence the *APP* expression. Further investigations of this region including mutational analysis may reveal the relative importance of these sequences for the promoter activity and confirm the identities and the roles of the factors.

The DNA-protein interaction patterns are exactly the same in neurons and astrocytes. In all likelihood, these DNA-protein interactions are essential for the basal expression of APP in normal cells. In perspective, it would be interesting to compare the protein interaction patterns on the *APP* gene promoter of these normal cells to Alzheimer affected neurons or astrocytes.

Differences in tissue expression of the *APP* gene could be caused by the varying proportions of the factors as well as the combination of their family members. So far there is no indication that the proximal promoter of the *APP* gene and the proteins that recognize it could play a role in increasing *APP* expression. However, other sequences outside of the proximal region of the *APP* promoter could be involved not only in tissue specific expression but also cause an alteration of *APP* expression under certain conditions (44).

The advent of high throughput screening methods of an almost limitless number of bio-molecules could allow the identification of compounds able to recognize specific DNA sequences. In conditions where overexpression of the *APP* gene could lead to symptoms related to AD, for example as what occurs in Down syndrome, interfering with the basal activity of the *APP* promoter, alone or in combination with other therapies, could slow down the development of the disease, even if the source of the signal leading to altered expression is unknown.

4.2.6 References

1. Selkoe, D. J. (2002) *J.Clin.Invest* **110,** 1375-1381

2. Schmechel, D. E., Goldgaber, D., Burkhart, D. S., Gilbert, J. R., Gajdusek, D. C., and Roses, A. D. (1988) *Alzheimer Dis.Assoc.Disord.* **2,** 96-111

3. Salbaum, J. M., Weidemann, A., Lemaire, H. G., Masters, C. L., and Beyreuther, K. (1988) *EMBO J.* **7,** 2807-2813

4. Pollwein, P., Masters, C. L., and Beyreuther, K. (1992) *Nucleic Acids Res.* **20,** 63-68

5. Quitschke, W. W. and Goldgaber, D. (1992) *J.Biol.Chem.* **267,** 17362-17368

6. Izumi, R., Yamada, T., Yoshikai, S., Sasaki, H., Hattori, M., and Sakaki, Y. (1992) *Gene* **112,** 189-195

7. Hoffman, P. W. and Chernak, J. M. (1995) *Nucleic Acids Res.* **23,** 2229-2235

8. Bourbonniere, M. and Nalbantoglu, J. (1996) *Brain Res.Mol.Brain Res.* **35,** 304-308

9. Kovacs, D. M., Wasco, W., Witherby, J., Felsenstein, K. M., Brunel, F., Roeder, R. G., and Tanzi, R. E. (1995) *Hum.Mol.Genet.* **4,** 1527-1533

10. Vostrov, A. A., Quitschke, W. W., Vidal, F., Schwarzman, A. L., and Goldgaber, D. (1995) *Nucleic Acids Res.* **23,** 2734-2741

11. Quitschke, W. W. (1994) *J.Biol.Chem.* **269,** 21229-21233

12. Hoffman, P. W. and Chernak, J. M. (1994) *Biochem.Biophys.Res.Commun.* **201,** 610-617

13. Vostrov, A. A. and Quitschke, W. W. (1997) *J.Biol.Chem.* **272,** 33353-33359

14. Pollwein, P. (1993) *Biochem.Biophys.Res.Commun.* **190,** 637-647

15. Quitschke, W. W., Matthews, J. P., Kraus, R. J., and Vostrov, A. A. (1996) *J.Biol.Chem.* **271,** 22231-22239

16. D'Souza, S., Alinauskas, K., McCrea, E., Goodyer, C., and Antel, J. P. (1995) *J.Neurosci.* **15,** 7293-7300

17. Bourbonniere, M. and Nalbantoglu, J. (1993) *Brain Res.Mol.Brain Res.* **19,** 246-250

18. Boshart, M., Kluppel, M., Schmidt, A., Schutz, G., and Luckow, B. (1992) *Gene* **110,** 129-130

19. Lillie, J. W. and Green, M. R. (1989) *Nature* **338,** 39-44

20. Gregor, P. D., Sawadogo, M., and Roeder, R. G. (1990) *Genes Dev.* **4,** 1730-1740

21. Du, H., Roy, A. L., and Roeder, R. G. (1993) *EMBO J.* **12,** 501-511

22. Shih, H. and Towle, H. C. (1994) *J.Biol.Chem.* **269,** 9380-9387

23. Dignam, J. D., Martin, P. L., Shastry, B. S., and Roeder, R. G. (1983) *Methods Enzymol.* **101,** 582-598

24. Kadonaga, J. T., Carner, K. R., Masiarz, F. R., and Tjian, R. (1987) *Cell* **51,** 1079-1090

25. Courey, A. J. and Tjian, R. (1988) *Cell* **55,** 887-898

26. Jones, K. A., Yamamoto, K. R., and Tjian, R. (1985) *Cell* **42,** 559-572

27. Sawadogo, M., Van Dyke, M. W., Gregor, P. D., and Roeder, R. G. (1988) *J.Biol.Chem.* **263,** 11985-11993

28. Drouin, R., Therrien, J. P., Angers, M., and Ouellet, S. (2001) *Methods Mol.Biol.* **148,** 175-219

29. Sawadogo, M. and Roeder, R. G. (1985) *Cell* **43,** 165-175

30. Chodosh, L. A., Carthew, R. W., Morgan, J. G., Crabtree, G. R., and Sharp, P. A. (1987) *Science* **238,** 684-688

31. Sato, M., Fukushi, Y., Ishizawa, S., Okinaga, S., Muller, R. M., and Shibahara, S. (1989) *J.Biol.Chem.* **264,** 10251-10260

32. Bresnick, E. H. and Felsenfeld, G. (1993) *J.Biol.Chem.* **268,** 18824-18834

33. Reisman, D. and Rotter, V. (1993) *Nucleic Acids Res.* **21,** 345-350

34. Roy, A. L., Meisterernst, M., Pognonec, P., and Roeder, R. G. (1991) *Nature* **354,** 245-248

35. Workman, J. L., Roeder, R. G., and Kingston, R. E. (1990) *EMBO J.* **9,** 1299-1308

36. Lukiw, W. J., Rogaev, E. I., Wong, L., Vaula, G., McLachlan, D. R., and St George, H. P. (1994) *Brain Res.Mol.Brain Res.* **22,** 121-131

37. Chiang, C. M. and Roeder, R. G. (1995) *Science* **267,** 531-536

38. Burley, S. K. and Roeder, R. G. (1996) *Annu.Rev.Biochem.* **65,** 769-799

39. Brandeis, M., Frank, D., Keshet, I., Siegfried, Z., Mendelsohn, M., Nemes, A., Temper, V., Razin, A., and Cedar, H. (1994) *Nature* **371,** 435-438

40. Li, B., Adams, C. C., and Workman, J. L. (1994) *J.Biol.Chem.* **269,** 7756-7763

41. Lu, J., Lee, W., Jiang, C., and Keller, E. B. (1994) *J.Biol.Chem.* **269,** 5391-5402

42. Schatt, M. D., Rusconi, S., and Schaffner, W. (1990) *EMBO J.* **9,** 481-487

43. Wiley, S. R., Kraus, R. J., and Mertz, J. E. (1992) *Proc.Natl.Acad.Sci.U.S.A* **89,** 5814-5818

44. Lahiri, D. K., Ge, Y. W., Maloney, B., Wavrant-De Vrieze, F., and Hardy, J. (2005) *Neurobiol.Aging* **26,** 1329-1341

4.2.7 Legend to figures

Figure 1. *APP* **promoter activity evaluated by transient transfection in NG108-15 and HepG2 cells.** NG108-15 and HepG2 cell lines were transiently transfected using a modification of the calcium phosphate protocol. For NG108-15, each transfection was performed with a total of 10 µg of DNA composed of 5 µg of pAPPCAT reporter plasmids, 4 µl of carrier DNA (pBLCAT6) and 1 µg of pUT535, a β-galactosidase reporter plasmid. HegG2 cells were transfected with a total of 15 µg of pAPPCAT reporter plasmids and 1 µg of pUT535. The CAT activity of each construct is expressed as a percentage of the minimal promoter contained in pAPPCAT-96. The error bars represent the standard error of the mean (NG108-15, n=10-15; HepG2, n=5-9). Results were obtained with at least two independent preparations of plasmid DNA for each construct. Transfections were performed in triplicate.

Figure 2. Characterization of the Sp1 binding to the *APP* gene promoter. A) DNase I footprinting analysis of the *APP* promoter with recombinant Sp1 protein. A Xma iii / BamH I (-203 to +100) restriction fragment end-labeled on the coding strand was incubated with either no protein (lane 1) or 160 ng of recombinant Sp1 (Promega) (land 2) and subjected to DNase I digestion. The brackets illustrate the protected regions compared to control. The numbers represent the positions relative to the major transcriptional start site. **B**) Sp1 binds the consensus Sp1 element in the *APP* promoter. EMSA was performed with an oligonucleotide APPSP1, -71 to -50, containing the protected region observed by DNase

I footprinting analysis. Recombinant Sp1 was incubated with oligonucleotide SP1 (lane 1) which contains a consensus Sp1 sequence from the SV40 promoter region and oligonucleotide APPSP1 (lane 2). The arrow indicates the complexes formed. F: free probe. **C**) Co-transfection assays with Sp1 and the *APP* minimal promoter in *Drosophila* S2 cells. The constructs pAPPCAT-77 and pAPPCAT-96 (1 μg) were co-transfected either with the *Drosophila* expression vector pPac0 or its derivative pPacSp1 which carries the cDNA for human Sp1 (pPacSp1). The amount of expression vector was kept constant at 0.2 μg: 0 μg of pPacSp1 = 0.2 μg of pPac0; 0.05 μg of pPacSp1 = 0.5 μg of pPacSp1 + 0.15 μg of pPac0). The activity pTKCAT, which contains two Sp1 binding sites, was increased when 0.05 μg and 0.2 μg of pPacSp1 were present.

Figure 3. Characteristics of Sp1 and USF biniding to the *APP* gene promoter. A) Sp1 and USF-1 bind to an oligonucleotide from the *APP* promoter containing both elements. Recombinant Sp1 and *in vitro* synthesized USF-1 were used in EMSA with oligonucleotide SP1-USF spanning region -71 to -37 of the *APP* promoter. Each binding reaction contained 1 μl of an *in vitro* coupled transcription/translation reaction. 100 ng of recombinant Sp1 was added (lanes 3-6) to *in vitro* coupled transcription/translation reactions where USF-1 was either present (lanes 3-5) or absent (lane 6). Lane 1 represents the probe SP1-USF incubated with 1 μl of *in vitro* coupled transcription/translation reaction where the T7 polymerase was omitted. The number *1* and *2* in overlay describe the order in which Sp1 and USF-1 were added did to the reaction (lanes 3 and 4). **B**) Oligonucleotide APPSP1-USF was end-labeled and incubated with HeLa nuclear extracts Competition EMSA with oligonucleotide APPSP1 and Ad-USF were performed (20, 50 and 100-fold molar excess respectively). Lane 8: Supershift experiment using the Ab-SP1 antibody. Comp., competition; F, free probe; arrowhead: supershift complex.

Figure 4. Effects of mutations in the Sp1 and USF elements. A) EMSA of the mutations introduced into the Sp1 element. Oligonucleotide APPSP1 was end-labeled and incubated with HeLa nuclear extracts (lane 1). Competition EMSA with oligonucleotide APPSP1m:

lane 2, 20-fold molar excess; lane 3, 50-fold molar excess; lane 4, 100-fold molar excess. Competition with the oligonucleotide SP1 (lane 5-7, 20, 50 and 100-fold molar excess respectively). Competition under similar conditions with APSP1 (lane 8-10, 20, 50 and 100-fold molar excess respectively). Comp., competition; F, free probe. **B**) EMSA of the mutations introduced into the USF element. HeLa nuclear extracts were used in competition EMSA with end-labeled oligonucleotide APP-E1 which contains the USF element from the *APP* promoter. APP-E1 formed two major complexes with proteins from HeLa nuclear extract (large arrow). Competition EMSA with increasing amount of unlabeled competitor APP-USFm (lane 2, 20-fold; lane 3, 50 excess; lane 4, 100-fold molar excess), and APP-USF (lane 5, 20-fold; lane 6, 50-fold; lane 7, 100-fold molar excess). Comp.: competition; F: free probe. **C)** Analysis of the transcriptional start sites of mutants of the *APP* promoter**.** Run-on *in vitro* transcription assays were performed with 80 μg of HeLa nuclear extracts on 2 μg of supercoiled plasmids pAPPCAT-96, pAPPCAT-96mSP1, pAPPCAT-96mUSF and pAPPCAT-96mSU. Transcription start sites were determined by primer extension assays with the antisense oligonucleotide ASAMY (+71 to +100) on the products of run-on *in vitro* transcription assays (lane 1-4) or using 50 μg of total RNA from HepG2 cells (lane 5). The products were separated on a 6% denaturing polyacrymide gel and their size was assessed by comparing with a sequencing reaction used as a ladder. The major transcriptional start site is indicated by +1. The ratio +1/-4 is indicated below.

Figure 5. Promoter activity of USF-1 and Sp1 mutants. A) Co-transfection assays of USF-1 and the *APP* promoter in HepG2 cells. HepG2 cells were co-transfected with a combination pcDNAI/amp (5, 4, 3 or 0 μg) or pCMVUSF-1 (0, 1, 2 or 5 μg), keeping the amount of plasmids bearing the CMV promoter constant at 5 μg; 5 μg of the reporter plasmid pAPPCAT-96; 4 μg of pBLCAT6 (carrier) and 1 μg of pUT535 (β-galactosidase) for transfection efficiency, for a total of 15 μg. Fold increases are expressed in function of untreated pAPPCAT-96. The error bars indicate the standard error of the mean (SEM), n=3. **B**) Transient transfections of mutants of the *APP* promoter in NG108-15 and HepG2 cells. NG108-15 cells were transfected with 5 μg of pAPPCAT derivatives, 5 μg of pBLCAT6 (carrier) and 1 μg of pUT535 (β-galactosidase) to standardize for transfection

efficiency. Transient transfection assays with HepG2 cells used 15 μg of pAPPCAT derivatives and 1 μg of pUT535. Transfections were performed in triplicate with 2 independent preparation of plasmid DNA for each construct (n=4). The presence or absence of the GC-element, Sp1 and USF elements are indicated by boxes in the schematic representation of the *APP* promoter. Relative CAT activity is expressed as a function of pAPPCAT-96. **C)** Transient transfections of truncated mutants of the *APP* promoter in NG108-15. NG108-15 cells were transfected with 10 μg of pAPPCAT-77 derivatives and 1 μg of pUT535 for β-galactosidase standardization of transfection efficiency. Presence or absence of the Sp1 and USF elements is indicated by the boxes in the schematic drawing of the truncated *APP* promoter. The relative CAT activity is expressed as a function of pAPPCAT-96 (not shown). Transfections were performed in triplicate with at least two independent preparations of plasmid DNA. The error bars indicated the standard error of the mean (SEM), n=3. **D)** Representative CAT assays of mutants of the *APP* promoter. NG108-15 cells were transfected with 10 μg of pAPPCAT derivatives and 1 μg of pUT535. For the purpose of this Figure, cellular exracts were standardized to β-galactosidase activity before CAT assays were performed. Mock transfection contained no DNA.

Figure 6. Genomic footprinting of the human *APP* gene promoter. The region shown was analyzed **A)** with primer set 1 to reveal upper strand sequences from nt −136 to −16, **B)** with primer set 7 to reveal bottom strand sequences from nt −175 to −19 relative to the transcription initiation site (TIS). Lane 1 and 10: Ligation-mediated polymerase chain reaction (LMPCR) of naked DNA purified from human primary cultured fibroblasts (F) that was treated *in vitro* (*t*) with DMS (1) or UVC (10). Lanes 2, 7 and 9: LMPCR of DNA purified from primary Astrocytes (AS) or neuronal cells (NE) that was treated *in vivo* (*v*) with DMS (2 and 3) or UVC (8 and 9) prior to DNA purification. Lanes 4-7 : Maxam-Gilbert sequence samples. All the footprints are illustrated on the corresponding sequence. DMS protected and hypersensitive guanines are indicated by opened and closed circles respectively on the left side of the autoradiograms. UVC protections and hypersensitives sites are indicated by opened and closed squares on the right side of the autoradiograms. **C)** Summary of in vivo dimethylsulfate (DMS) and ultraviolet C (UVC) footprints identified

along the −189 to +19 human *p21* gene promoter. The consensus binding sites for transcription factors are shown with black lines above the sequences.

3.2.8 Table

TABLE I. Synthetic Oligonucleotide Primers for LMPCR Analysis
of the Human APP Gene Promoter

| Primer	Sequence (5'-3')	Position*	Tm(°C)**
Non-transcribed (upper) strand			
1.1	GCGTCCTTGCTCTGCC	+96 to +81	59.2
1.2	CGCCGCCACCGCCGCCGTCTCC	+78 to +57	74.9
2.1	GAGACCCCTAGCGGCG	-79 to -94	61.7
2.2	GGGAACTGCGCCCGCTCGCG	-101 to -120	70.0
3.1	GTGGGAGGGAGAGTCTG	-231 to -247	59.4
3.2	GCCAGGAGAGGGACGGTGCAGGATCAG	-250 to -276	71.1
4.1	TCCGCATTTCGTTTTTTCTC	-414 to -433	53.7
4.2	GGGGCAGGCGTTTCTGGAAGAGAATGAG	-449 to -476	67.8
Transcribed (bottom) strand			
5.1	ACATCCCTGCTTAACAACAA	-791 to -772	53.7
5.2	CCCCCGCCCCGCAAAATCCACACTG	-763 to -739	70.1
6.1	TTGGTTCGTTCTAAAGATAG	-602 to -583	51.6
6.2	GCTCGTGCCTGCTTTTGACGTTGG	-570 to -547	65.3
7.1	CTCACCTTTCCCTGATC	-287 to -271	54.5
7.2	GCACCGTCCCTCTCCTGGCCCCAGACTC	-268 to -241	73.7
8.1	GCGTGGGGTGCAGGC	-185 to -171	61.7
8.2	GCCAAGGGCGCTGCACCTGTGGG	-167 to -145	70.8

*Primer positions are given relative to the main transcription initiation site.
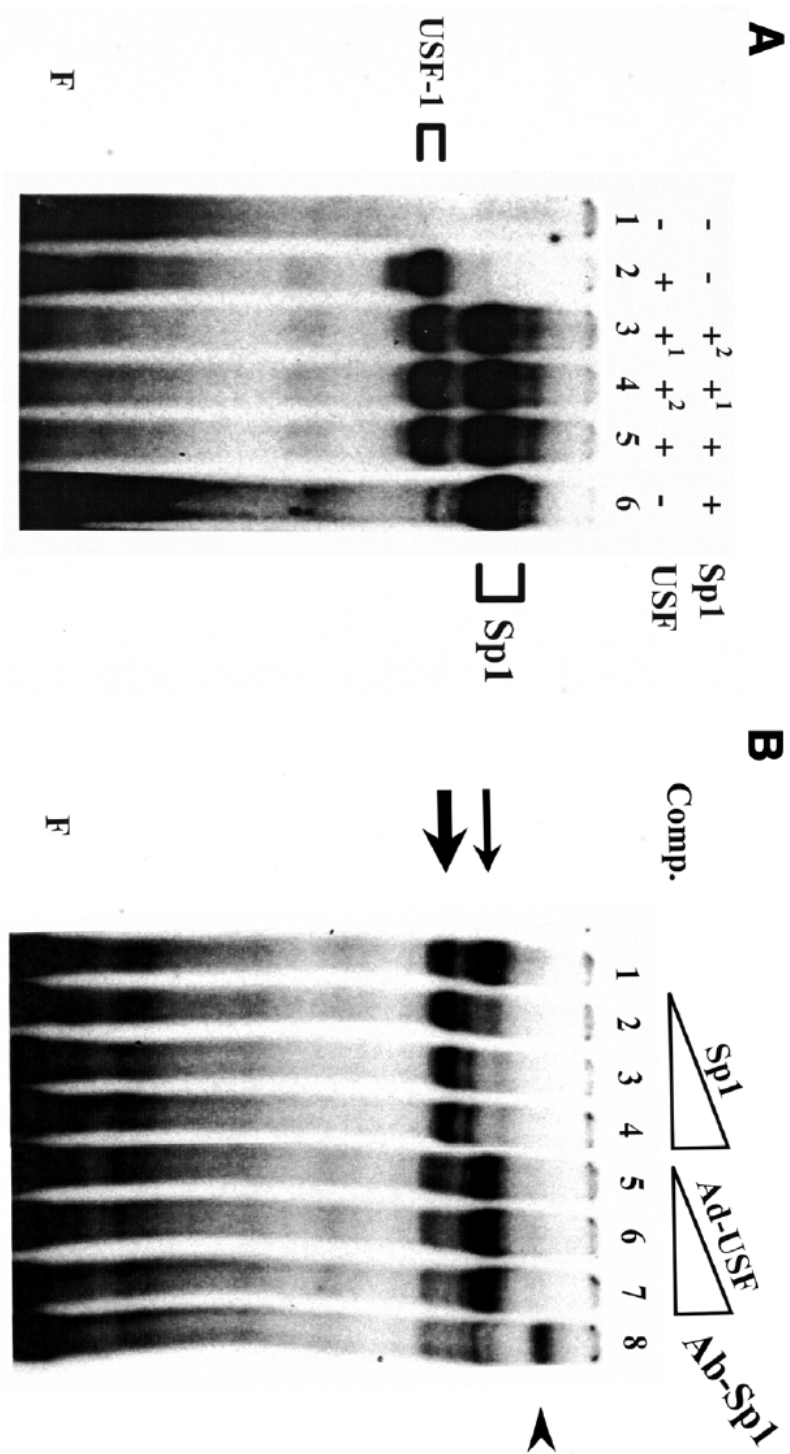**Tm determined by the GeneJockey software program.
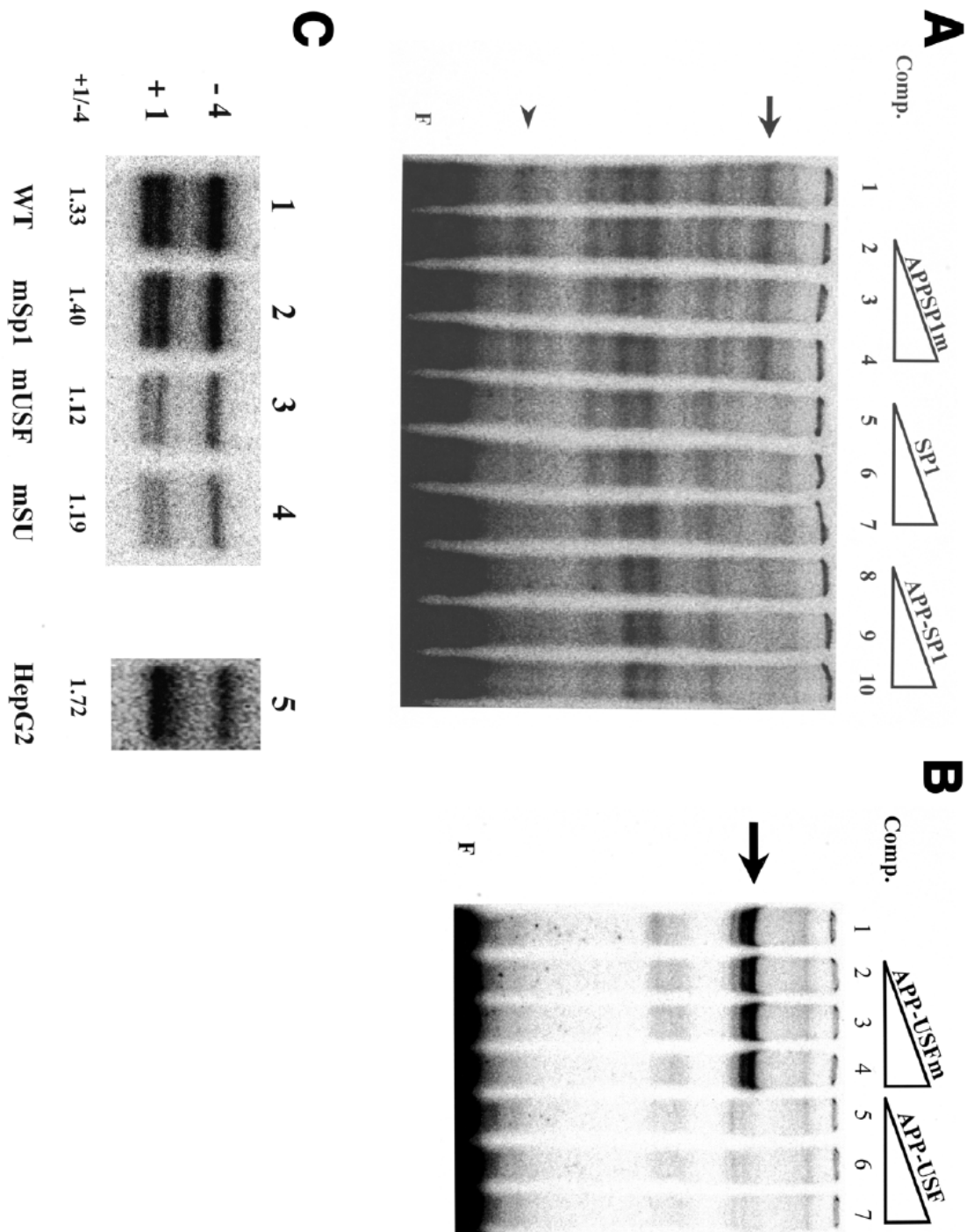
3.2.9 Figures
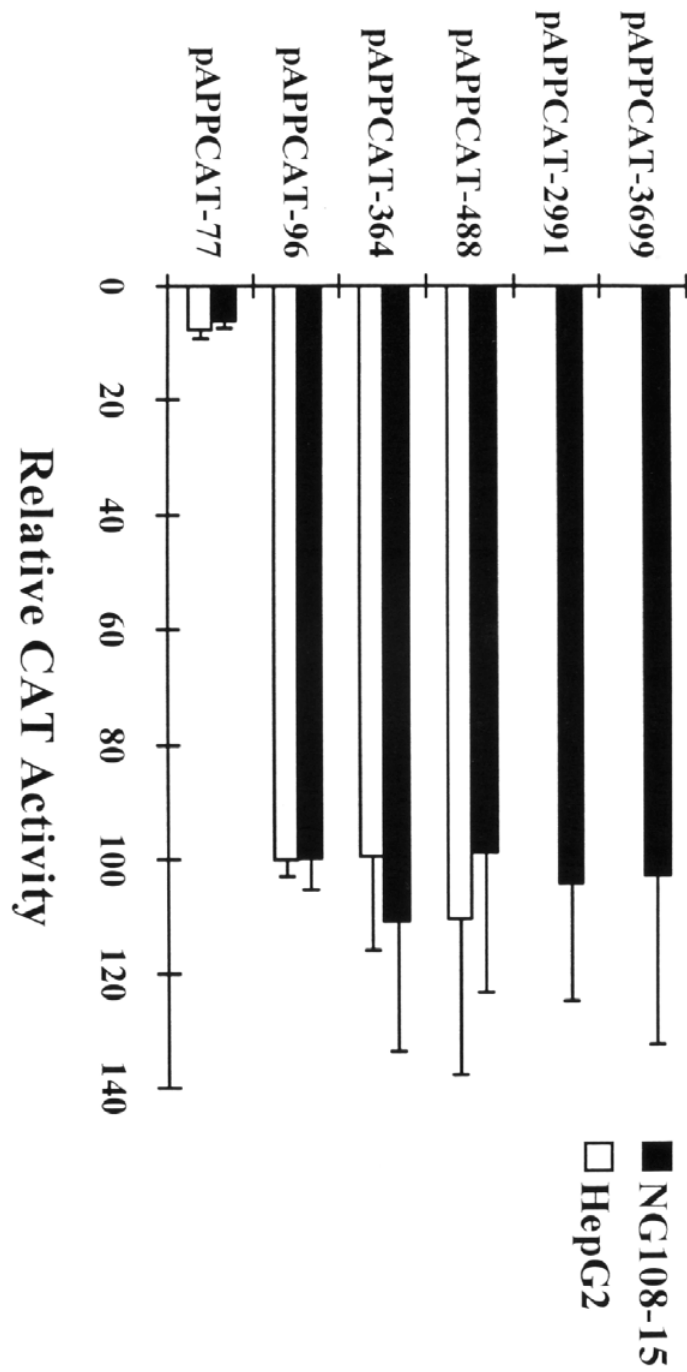
Figure 1. *APP* promoter activity evaluated by transient transfection in NG108-15 and HepG2 cells.
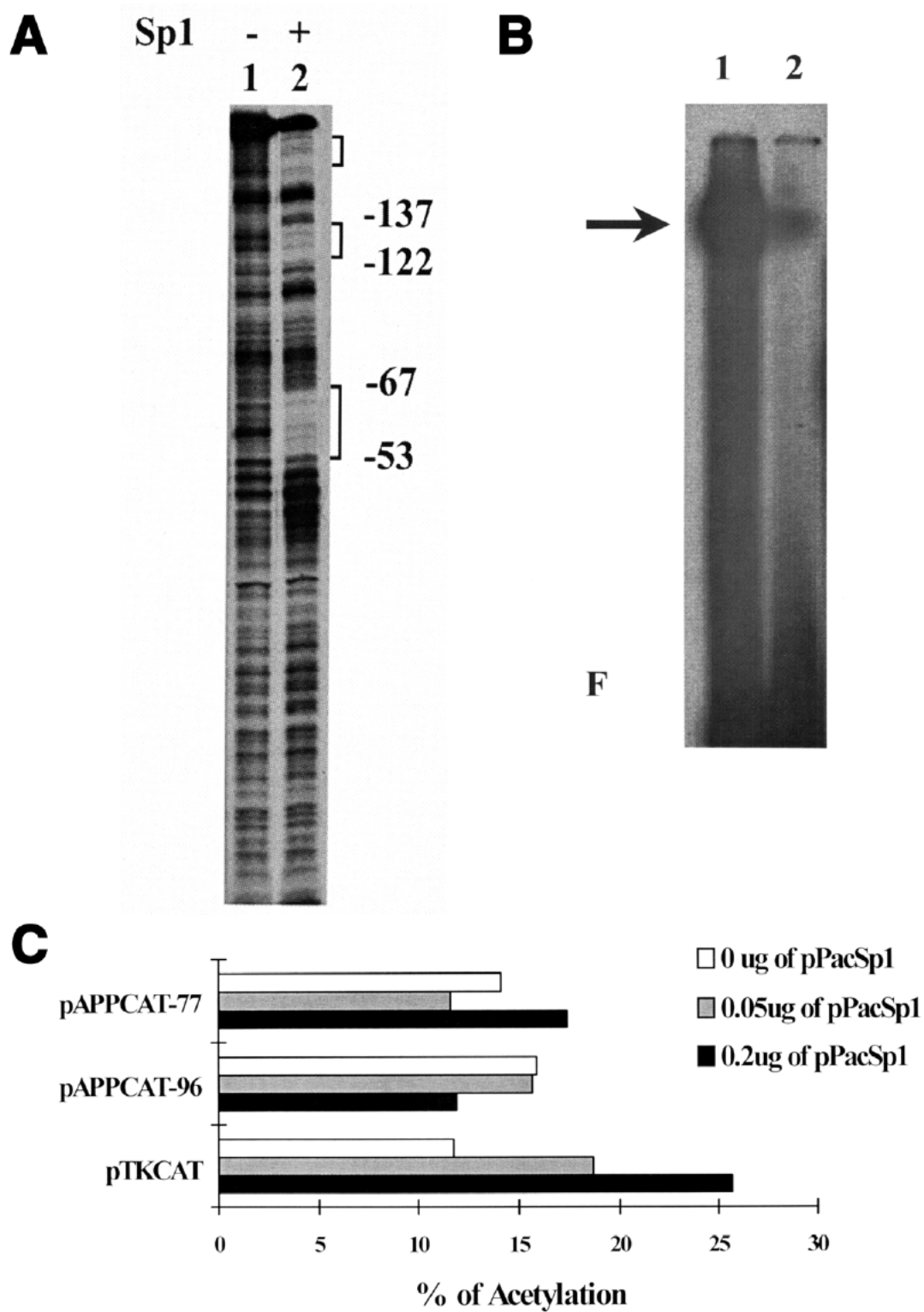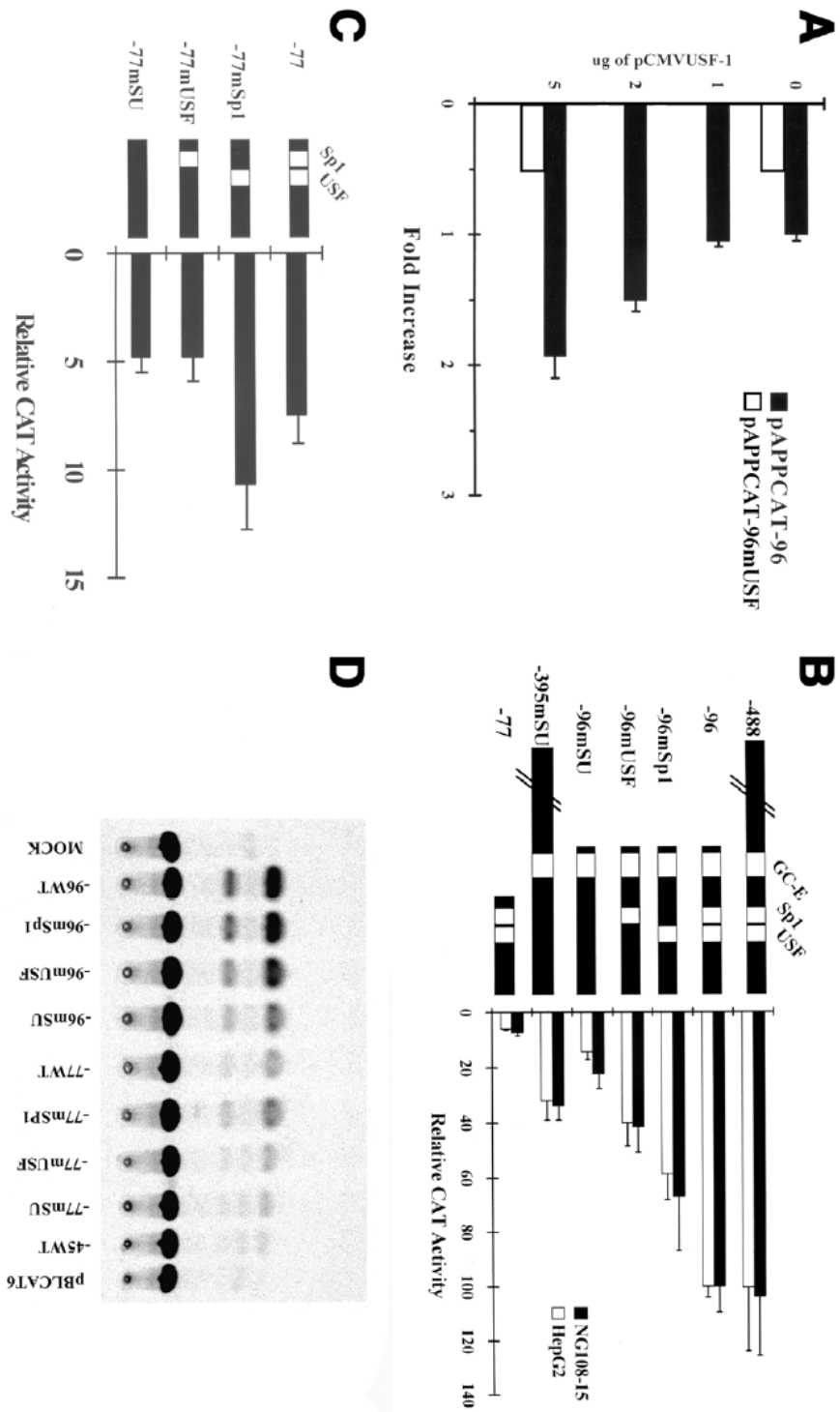
Figure 2. Characterization of the Sp1 binding to the *APP* gene promoter.

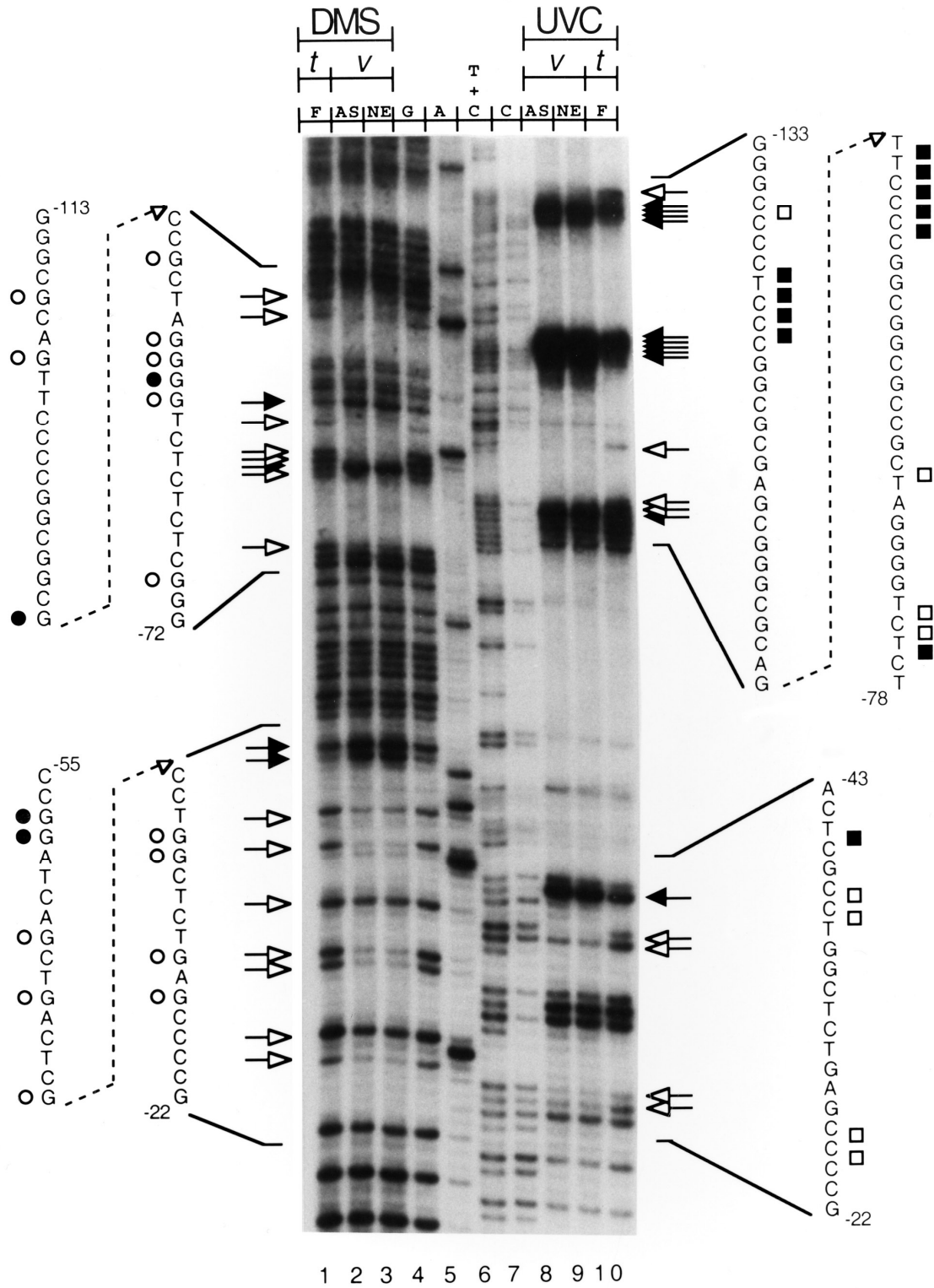**Figure 3.** Characteristics of Sp1 and USF biniding to the *APP* gene promoter.

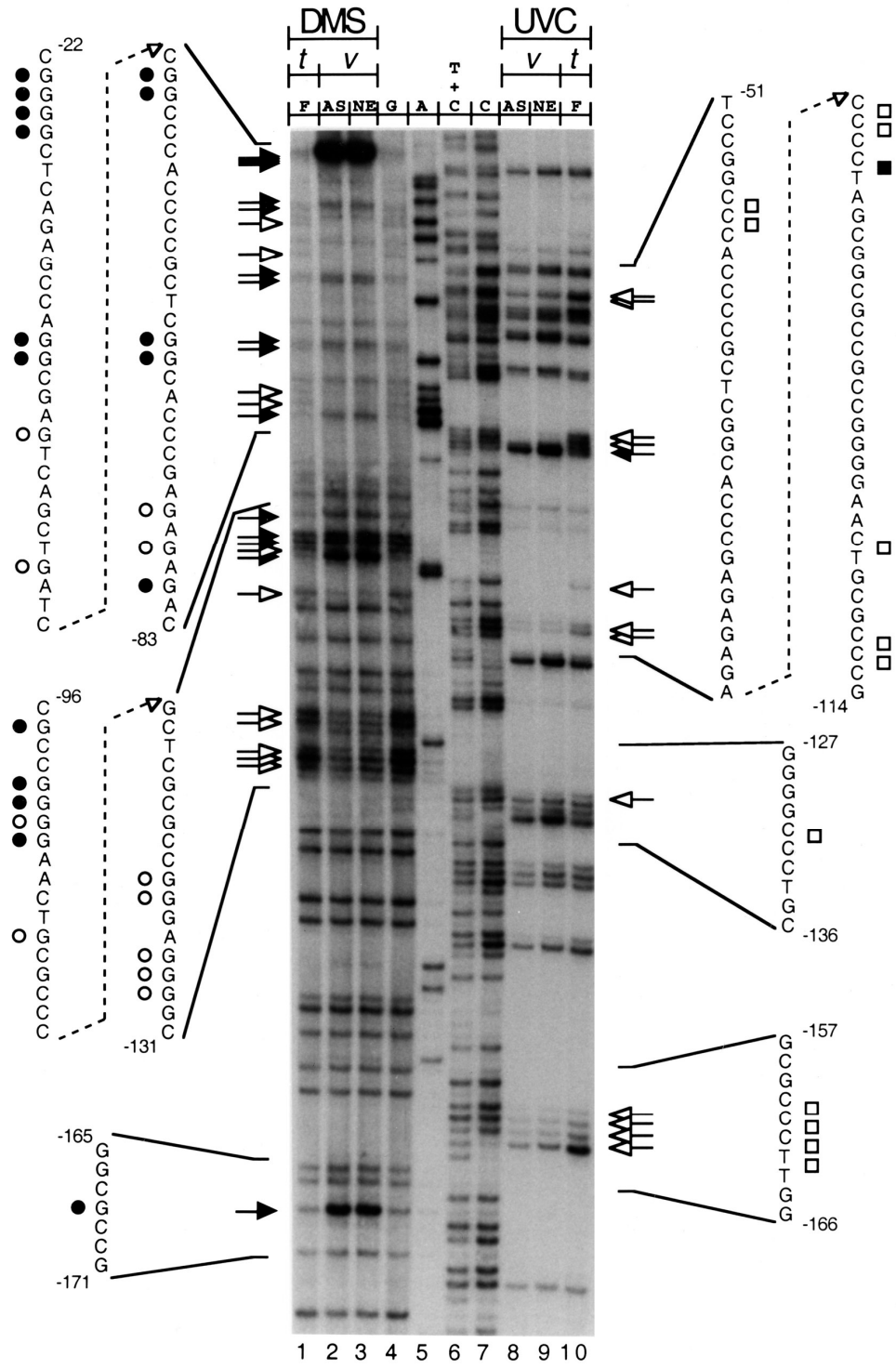**Figure 4.** Effects of mutations in the Sp1 and USF elements.

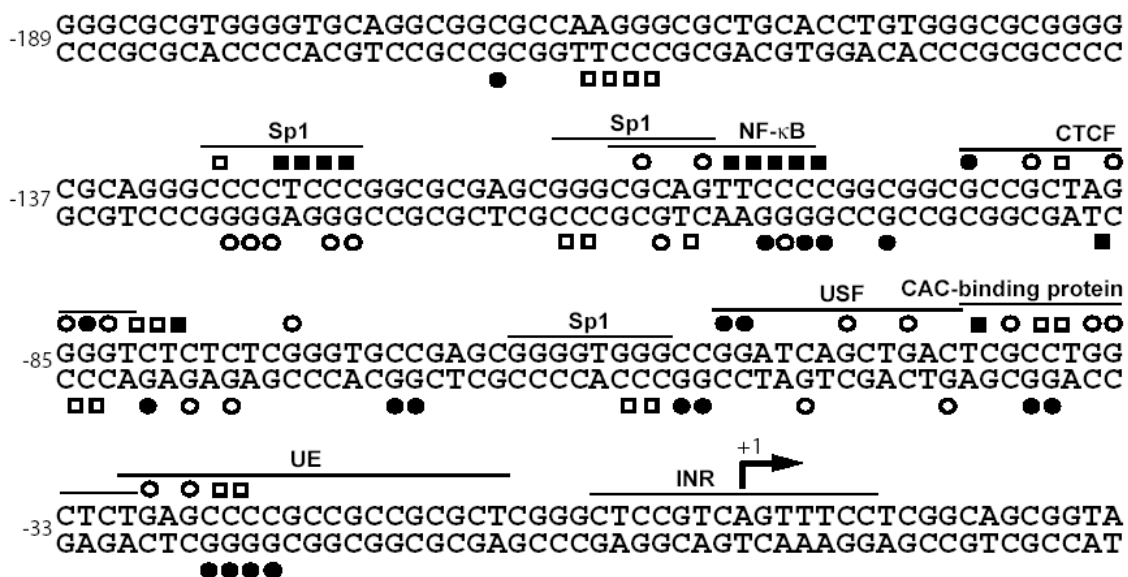**Figure 5.** Promoter activity of USF-1 and Sp1 mutants.

A

B

C

```
      GGGCGCGTGGGGTGCAGGCGGCGCCAAGGGCGCTGCACCTGTGGGCGCGGGG
-189  CCCGCGCACCCCACGTCCGCCGCGGTTCCCGCGACGTGGACACCCGCGCCCC
                              ●         □□□□

                  Sp1                 Sp1        NF-κB            CTCF
            □    ■■■■             ○    ○ ■■■■■■        ●    ○ □    ○
      CGCAGGGCCCCTCCCGGCGCGAGCGGGCGCAGTTCCCCGGCGGCGCCGCTAG
-137  GCGTCCCGGGGAGGGCCGCGCTCGCCCGCGTCAAGGGGCCGCCGCGGCGATC
            ○○○   ○○              □□    ○ □    ●○●●    ●              ■

                                                    USF    CAC-binding protein
      ○●○ □□■       ○              Sp1          ●●       ○    ○   ■ ○ □□ ○○
      GGGTCTCTCTCGGGTGCCGAGCGGGGTGGGCCGGATCAGCTGACTCGCCTGG
-85   CCCAGAGAGAGCCCACGGCTCGCCCCACCCGGCCTAGTCGACTGAGCGGACC
         □□  ●  ○  ○              ●●           □□ ●●    ○         ○    ●●

                  UE                        INR    +1
            ○  ○ □□                               ⌐►
      CTCTGAGCCCCGCCGCCGCGCTCGGGCTCCGTCAGTTTCCTCGGCAGCGGTA
-33   GAGACTCGGGGCGGCGGCGCGAGCCCGAGGCAGTCAAAGGAGCCGTCGCCAT
            ●●●●
```

Figure 6. Genomic footprinting of the human *APP* gene promoter.

Chapitre 5 : Discussion et Conclusion

5.1 Discussion

La régulation de l'expression génique qui est assurée d'une manière coordonnée par l'ensemble des facteurs de transcription, des co-régulateurs (ou co-facteurs) et de la machinerie basale de transcription, est essentielle à la spécificité et à l'homéostasie cellulaire. La transcription d'un gène en ARNm est l'étape première de son expression et de la production d'une protéine fonctionnelle. Du même fait, elle est le premier, et souvent le plus décisif des points de contrôle de cette expression. Alors que de plus en plus de gènes sont découverts et que de plus en plus de ceux-ci sont associés à des pathologies, la régulation de la transcription génique chez l'humain reste encore très peu comprise. Les modèles traditionnels se voulaient beaucoup trop statique étant donné la variété de mécanismes et d'acteurs qui agissent ensemble de façon dynamique dans la cellule pour régir l'expression des gènes. C'est pourquoi les études ayant pour but la démystification des mécanismes de régulation de la transcription des gènes et leur comparaison sont d'un si grand intérêt. Les travaux présentés dans cette thèse ont permis de mieux définir les mécanismes de régulation de la transcription de trois gènes associés à des pathologies, fournissant des informations précieuses et utiles au développement de futures thérapies.

5.1.1 Article # 1 (Chapitre 2)

L'activité fonctionnelle de p21 semble dépendre du stade cellulaire car elle est régulée par l'interaction directe et indirecte de divers modulateurs ou par des modifications post-transcriptionnelles (phosphorylation, protéolyse partielle). Dans notre étude, nous avons localisé dans le promoteur basal du gène humain *p21*, via des approches *in cellulo* (LMPCR) et *in vitro* (footprint avec le DMS et la DNaseI) un site de liaison (-161 à −149) pour le facteur de transcription (FT) NFI qui s'est révélé être un puissant répresseur de la transcription du gène au sein de fibroblastes humains normaux en prolifération et de cultures cellulaires établies (HeLa et GH4C1). Jusqu'ici, aucune publication n'avait fait état

d'une possible implication de NFI dans la régulation de l'expression de *p21*. L'empreinte cartographiée *in cellulo* excède d'environ 10 pb l'extrémité 3' du site consensus de liaison pour NFI (de −148 à −138). L'analyse de retard sur gel à clairement montrée que la liaison de NFI ne nécessitait pas cette séquence d'ADN, ce qui laisse croire qu'elle est le site de liaison d'un autre facteur. L'analyse à l'aide des bases de données n'a pas permis d'identifier un facteur nucléaire capable de reconnaître cette séquence du promoteur. La présence de facteurs ou de co-facteurs pouvant lier *in cellulo* mais non *in vitro* cette région, tout comme le repliement de l'ADN, peuvent être à l'origine de la protection des résidus juste en 3' du site lié par NFI. Toutefois, la délétion de la région s'étendant de −192 à −124 lors des expériences d'anti-ARN (Fig. 6C) suggère l'action d'un autre facteur liant cette région puisque le niveau d'activation obtenu avec la construction p21-124 ne correspond pas à celui obtenu avec p21-192 muté au site NFI. Néanmoins, il est important de mentionner que la réduction du niveau nucléaire de NFI avec les anti-ARN peut aussi entraîner des répercussions indésirables sur la transcription de multiples gènes qui peuvent coder pour des facteurs autres que NFI participant aussi à régulation de l'expression de *p21* de façon directe ou indirecte. D'autres travaux seront essentiels pour éclaircir cette situation.

La famille de facteurs de transcription NFI est divisée en quatre groupes (NFI-A, -B, -C, et -X) (113), chacun étant codé par un gène différent et pouvant former des homo- et des hétéro-dimères (114). La grande diversité de cette famille, dont 19 membres ont été décrits à ce jour (revue dans (115)), est due à l'épissage alternatif des ARNm (116, 117). Les membres de la famille NFI ont été décris comme répresseurs (118-121) ou activateurs (122, 123) de la transcription des gènes. Il y aussi quelques exemples ou NFI participait à la régulation d'un gène via la coopération ou la compétition avec d'autres facteurs (119, 124, 125). Considérant l'effet inhibiteur qu'exerce NFI sur la transcription du gène *p21*, nous proposons qu'il agit en coopération avec d'autres facteurs et/ou co-facteurs environnant pour restreindre l'expression du gène au cours de la progression du cycle cellulaire lorsque l'ADN n'est pas endommagé. Il a été rapporté que NFI interagissait et exerçait une action antagoniste avec le facteur Sp1 au niveau du promoteur du gène PDGF-A (*platelet-derived*

growth factor) pour en diminuer l'expression (124). Or, la cartographie *in cellulo* du promoter de *p21* a révélé des empreintes à 4 des 6 sites de liaison putatifs pour Sp1 rapportés (126) en absence théorique de dommages à l'ADN et de p53 actifs. Plusieurs groupes ont déjà établi l'importance critique de Sp1 pour l'expression de *p21* (127-129). Récemment, il a été montré que Sp1 facilitait l'activation de p21 via son interaction directe avec p53 (130). D'un autre côté, il a été montré que cette influence positive de Sp1 peut être renversée par la déacétylase des histones 1 (HDAC1; *histone deacetylase 1*) qui en modifiant le niveau de condensation de la chromatine prévient l'accès de Sp1 au promoteur de *p21* (129, 131, 132). Néanmoins, comme la liaison de Sp1 au promoteur de *p21* a été observée *in cellulo* en absence de dommages à l'ADN ou d'induction de p53, il semble que le recrutement d'autres facteurs ou co-facteurs soit responsable de l'activation du gène qui dépend de p53. Outre NFI et Sp1, l'analyse *in cellulo* a permis l'identification d'empreintes au niveau du site de liaison pour des facteurs nucléaires comme E2F1 (Fig. 1) dont la liaison au promoteur proximal de *p21* avait déjà été montrée (134, 135). Les membres de la famille de facteurs de transcription E2F jouent un rôle crucial dans l'activation de gènes spécifiques à la transition G1/S (133). D'autres études ont aussi montré l'implication de E2F1 dans l'activation de *p21* indépendamment de p53 (134, 135). La délétion du segment du promoteur de *p21* contenant le site de liaison de E2F1 diminue de façon significative l'expression du gène rapporteur de la luciférase lors de transfections transitoires de cellules U343 (135). La transactivation de *p21* via l'interaction de E2F1 avec son site de liaison est associée avec l'augmentation concomitante des niveaux endogènes de p21 et d'E2F1 à la frontière G1/S. Tous ces données en plus du fait que p21 a été identifiée au sein de complexes cyclines/cdk/p21 ayant une activité kinase (136), supporte l'idée que p21 ait une rôle promoteur sur le cycle cellulaire, agissant comme facteur d'assemblage pour les cyclines et les kinases dépendantes des cyclines (cdk).

Les résultats d'immunoprécipitation de la chromatinne (ChIP) ont révélé certaines variations au niveau du promoteur de *p21* quant au patron d'occupation des facteurs de transcriptions Sp1, NFI et E2F1 entre les HSF (fibroblastes cutanés humains normaux) en prolifération et quiescents (privés de sérum) (Fig. 6). Alors que Sp1 et NFI mais pas E2F1

sont substantiellement liés au promoteur de *p21* dans les fibroblastes en prolifération, la privation de sérum entraîne la perte de la reconnaissance des facteurs Sp1 et NFI mais l'addition de la liaison d'E2F1. Or, l'occupation relative du promoteur de *p21* par NFI *in cellulo* excède clairement celle de Sp1/Sp3 dans les HSF en prolifération, favorisant ainsi la répression de la transcription de *p21* plutôt que l'activation, et pouvant donc être à l'origine du niveau basal d'expression du gène. À l'opposé, l'absence totale de liaison de NFI dans les cellules privées de sérum combiné à l'influence positive de Sp3 et de E2F1, tous deux présents au niveau du promoteur dans ces conditions, est consistante avec l'augmentation déjà rapportée de l'expression de p21 lors de virage des cellules vers l'arrêt du cycle et de la croissance (135, 137). La présence de Sp3 au niveau du promoteur de *p21* n'est pas affectée par l'altération de l'état de prolifération cellulaire et d'expression du gène. Ceci supporte l'idée que la majeure partie des sites de liaison Sp1 du promoteur de *p21* sont occupés par Sp3 suite à la privation de sérum, puisque les deux facteurs interagissent avec les mêmes sites de reconnaissance riches en GC (137, 138). La présence de Sp3 dans les HSF proliférant ou en quiescence suggère que *p21* est activé de façon constitutive par ce facteur et que la régulation fine de l'expression du gène est assuré par la capacité de liaison des autres facteurs, comme NFI, à leurs sites de liaison respectifs. Nos résultats sont en parfait accord avec ceux déjà publiés par d'autres groupes qui stipulaient que même si l'expression basale de *p21* était sous le contrôle des deux facteurs Sp1 et Sp3 dans les kératinocytes primaires de souris, seul Sp3 contribuait à l'induction du promoteur de *p21* qui dépendait du calcium (137). En prennant en considération que l'accumulation importante de p21 mène à l'hypophosphorylation de Rb, l'inhibition de transcription qui dépend de E2F et de l'arrêt du cycle cellulaire en G1 (133, 139-141), nous proposons que E2F1 pourrait être l'un des nombreux constituants d'une boucle de rétro-action requise pour l'induction de l'arrêt du cycle cellulaire plutôt qu'être un activateur direct majeur de la transcription de *p21*. Nos résultats suggèrent que suite au relâchement de la répression exercée par NFI, Sp3 et les facteurs E2F1 disponibles, donc qui n'ont pas encore été séquestrés par l'hypophosphorylation éminente de Rb, sont recrutés au niveau du promoteur de *p21* et induisent un haut niveau de transcription du gène. L'accumulation du niveau de p21 mène à l'hypophosphorylation Rb et théoriquement à la séquestration des facteurs de type E2F libre. Or, même si le niveau de E2F1 libre diminue, la transcription

soutenue de p21 n'est plus requise à partir du moment où la cellule est entrée en processus de dormance ou de quiescence. Or, plus d'évidences sur la dynamique de recrutement d'E2F1 et de sa relation avec la transcription de *p21* seront requises afin de vérifier ce modèle.

L'identification de chacun des facteurs nucléaires interagissant aux nouveaux sites identifiés dans cette étude et la détermination de leur rôle précis quant à la transcription de *p21* sera requise afin de comprendre les mécanismes qui contrôlent l'expression de *p21* en fonction du cycle cellulaire. Nous devons maintenant considérer NFI comme un nouvel acteur central dans la régulation de l'expression de *p21*. L'analyse des relations entre NFI et les autres facteurs de transcription comme p53, Sp1, Sp3 et E2F1 sera fascinante et essentielle à la compréhension de la régulation complexe du gène *p21*.

5.1.2 Article # 2 (Chapitre 3)

Dans le cas de l'étude sur le gène *GPC3*, l'équipe du Dr Sinnett avait déjà déterminé que la méthylation n'était pas impliquée dans la répression transcriptionnelle du gène et que la méthylation partielle observée chez les filles est probablement associée au X inactif (108). Des profils de méthylation similaires avaient été relevés dans la plupart des échantillons tumoraux analysés. Ces résultats laissaient donc supposer que la perturbation de l'expression de *GPC3* dans les neuroblastomes et les tumeurs de Wilms impliquait un autre niveau de régulation. Nous nous sommes attardé à la caractérisation des interactions ADN-protéines au niveau du promoteur du gène *GPC3* dans deux lignées de neuroblastomes exprimant le gène de façon différentielle. Nous avons tout d'abord mis en évidence une densité élevée peu commune d'empreintes protéiques à l'aide de la DNase I et des UVC, révélatrice d'une structure particulière de la chromatine au niveau du promoteur de *GPC3*. La région s'étendant de −960 à −770 présente une périodicité d'environ 10 pb de sites

hyper-réactifs à la DNase I ce qui représente la signature typique de la présence de nucléosomes (chapitre en annexe). Ces derniers peuvent aussi être révélés par l'analyse avec les UVC avec une périodicité semblable (chapitre en annexe) ce qui dans notre étude n'est observé que de façon interrompue. Ceci est probablement dû à la distribution des paires de pyrimidines qui ne peuvent que rarement permettre la visualisation claire des nucléosomes à l'aide des UVC sur une longue région. Une très grande quantité d'empreintes protéiques est aussi présente dans la région s'étendant de −770 à −400. Nous y retrouvons toutefois pas la même périodicité à l'exception de quelques courtes régions. La plupart de ces empreintes seraient donc le reflet d'une structure particulière de l'ADN autre que l'assemblage avec les nucléosomes puisqu'il semble très improbable qu'un si grand nombre de facteurs protéiques se lient de façon spécifique au promoteur de *GPC3*.

Étant donné la haute fréquence d'empreintes protéiques révélées à l'aide de la DNase I et des UVC, nous avons centré notre attention sur les empreintes protéiques détectées à l'aide du DMS qui ne révèle normalement pas la présence de nucléosomes et très peu de structure particulière de l'ADN, afin de localiser les sites d'interactions entre les facteurs de transcription et leur cible spécifique au niveau du promoteur. Nous avons ainsi trouvé au niveau du promoteur proximal (-400 à +20) huit régions (numéroté de 1 à 8) présentant des empreintes protéiques intéressantes puisqu'elles sont spécifiques à la lignée exprimant le gène *GPC3*, SJNB-7 (Fig. 2). La caractérisation *in vitro* du promoteur minimal (-218 à +12) avec les extraits de cellules Caco-2 avait permis la détection d'une large région d'interactions s'étendant de −43 à −7 et d'une autre plus petite de −103 à −86 (109). Lorsque l'on compare avec les résultats que nous avons obtenus, l'empreinte protéique révélée *in vitro* de −43 à −7 correspond aux régions 6 et 7 identifiées *in cellulo*, alors que les régions 5 et 8 n'ont pas été détectées par l'analyse *in vitro*. Cette distinction peut être due à la différence de type cellulaire et/ou aux nombreuses différences entre l'ADN natif et le contexte *in vitro*. Huber *et al* (109) avait déjà montré par retard de migration sur gel (EMSA) que la région −43 à −7 liait le facteur de transcription Sp1 et que celui-ci pouvait activer la transcription via le promoteur minimal de *GPC3* dans les cellules de drosophile.

La transfection de promoteurs délétés de *GPC3* suggérait aussi que les éléments situés dans la région −218 à −82 étaient importants pour l'activation de la transcription (109).

Nous avons de plus détecté quatre empreintes spécifiques à SJNB-7 en amont du promoteur minimal (empreintes 1, 2, 3 et 4; Fig. 2), rappelant l'importance d'étudier la régulation de la transcription dans un contexte nucléaire natif. Or, de façon intéressante, les régions 1, 3 et 4 présentent des patrons d'empreintes vraiment similaires, soient des empreintes protéiques observées avec le DMS centrés au niveau d'une boite CCAAT et un facteur de transcription putatif commun, NFY.

Les boîtes CCAAT sont des séquences extrêmement bien conservées et connues pour lier le facteur de transcription NFY (voir référence 30 du chapitre 3). Même si la liaison de NFY à la boîte CCAAT est souvent essentielle à la transcription d'un gène, particulièrement pour ceux dépourvus d'une boîte TATA, NFY ne peut de lui même activer la transcription. Son rôle serait plus précisément de promouvoir la transcription via la liaison directe à ses éléments et/ou de faciliter le positionnement de facteurs de transcription au SIT. À cet égard, il a été proposé que la coopération entre Sp1 et NFY soit responsable de l'expression basale des gènes dépourvus d'une boîte TATA (voir référence 30 et 31 du chapitre 3), ce qui est particulièrement pertinent pour *GPC3* qui possède une haute densité de dinucléotides CpG (108) et aucune boîte TATA (109).

Nos expériences de retard de migration sur gel (EMSA) ont montré que les régions −400 à −341 et −318 à −271 liaient spécifiquement des protéines des extraits des deux lignées cellulaire étudiées une exprimant différentiellement *GPC3*. Ces régions présentent des empreintes spécifiques à la lignée exprimant le gène, il est surprenant de ne remarquer aucune différence notable quant au patron de bandes retardées obtenu. Ceci renforce l'idée que le contexte *in cellulo* est important pour la régulation de la transcription. Or, la formation des complexes retardés est annulée par l'incubation préalable des extraits avec

l'anticorps NFYB mais pas par la compétition avec une cible consensus pour NFY, ce qui suggère que le complexe contient un facteur de transcription de type NFY avec une spécificité pour une séquence consensus particulière.

À la lumière de ces résultats, comment peut-on expliquer l'expression spécifique au type cellulaire de *GPC3*. Les 2 seuls facteur identifiés comme pouvant interagir avec le promoteur sont Sp1 et potentiellement un facteur de transcription du type NFY. Sp1 est exprimé de façon ubiquitaire et ne peut donc probablement pas être responsable de cette spécificité cellulaire. Nous proposons donc un modèle dans lequel l'état de la chromatine serrait impliqué dans la régulation de *GPC3*. La très grande majorité des empreintes sont localisées au niveau des mêmes régions dans les 2 lignées cellulaires et les empreintes particulières à celles exprimant le gène (SJNB-7) sont souvent entourées d'autres communes au 2 lignées. Il est donc très probable que certains facteurs soient présent aux sites de liaison du promoteur dans les 2 lignées, mais que leurs interactions plus compact et stable soient à l'origine de l'expression du gène dans SJNB-7. Dans ce modèle, l'activation du gène dans la lignée SJNB-7 ne serait pas le résultat de l'apparition et de la liaison d'un nouveau facteur de transcription, mais à un changement dans la structure physique du promoteur ou à l'action de co-facteurs (nouveaux co-activateurs ou perte de co-suppresseurs), entraînant le renforcement des interactions ADN-protéines déjà existantes aux niveaux des régions 1 à 8. L'implication de la structure native du promoteur dans l'inhibition de l'expression du gène pourrait de plus expliquer le fait que 3,2 kb du promoteur est capable d'induire la transcription d'un gène rapporteur dans les cellules HeLa où *GPC3* endogène n'est pas exprimé (109), la construction n'ayant pas cette structure native inhibitrice. D'autres travaux seront nécessaires pour identifier les autres facteurs contrôlant l'expression de *GPC3* incluant l'implication des changements locaux de la chromatine.

5.1.3 Article # 3 (Chapitre 4)

Le gène *AβPP* est exprimé de façon ubiquitaire chez l'adulte. Son promoteur ne contient pas de boîte TATA et selon plusieurs études, seule la région de −96 à +100 serait essentielle à la régulation du gène. Or, jusqu'ici, aucune caractérisation *in cellulo* du promoteur n'avait été faites. Nous avons clairement démontré que deux éléments du promoteur d'*AβPP* sont reconnus par les facteurs de transcription Sp1 et USF. De ces derniers, seul USF est capable de transactiver le promoteur d'*AβPP* en co-transfection transitoire même s'ils sont tous deux essentiels à l'activité maximale du promoteur dans deux types de cellules humaines.

Le facteur de transcription USF (Upstream Stimulatory Factor) est connu pour activer une variété de gènes via l'interaction directe avec les éléments régulateurs (E-box) au niveau des promoteurs. L'effet activateur d'USF pourrait avoir lieu via l'interaction protéine-protéine avec la machinerie de transcription basale et/ou par la formation d'une région libre de nucléosomes au niveau du promoteur.

Nos résultats de retard de migration sur gel (EMSA) et d'empreintes protéiques *in vitro* en utilisant l'oligonucléotide correspondant à la séquence de liaison pour Sp1 de la région promotrice (−71 à −50; APPSP1) sont en parfait accord avec ce qui avait été proposé par Pollwein (110) et par Hoffman et Chernak (143) pour le gène *AβPP* chez le rat. Toutefois, si la liaison de Sp1 à cet élément du promoteur d'*AβPP* est presque indéniable, son rôle dans la régulation du gène est complexe. Nous n'avons pu observer de transactivation du promoteur d'*AβPP* en co-transfection transitoire chez la Drosophile. L'une des raisons possible réside en l'affinité de Sp1 pour son élément du promoteur *AβPP*. Même si nous n'avons fait aucun essai d'affinité, plusieurs indices laisse supposer que Sp1 possède moins d'affinité pour l'élément présent dans le promoteur d'*AβPP* que pour celui de SV40 qui a été utilisé comme contrôle. L'autre facteur ayant put influencer le potentiel d'activation de Sp1 est le nombre de ces éléments de liaison dans les promoteurs que nous avons

comparés. Notre contrôle positif (pTKCAT), qui a été activé par Sp1 par un facteur supérieur à 2, contient deux sites de liaison pour Sp1. Il se peut fort bien que la présence d'un seul site de faible affinité pour la liaison de Sp1 dans le promoteur d'*AβPP* soit la raison de la faible augmentation de l'activité observée lors des essais de co-transfection dans les cellules de Drosophile. Il sera important de vérifier la transactivation induite par Sp1 dans le contexte de promoteur entier d'*AβPP* qui contient en fait deux autres éléments putatifs Sp1 qui ont révélé des empreintes *in cellulo* (-130 à −123 et −111 à −104). Le nombre de sites de liaison pour un facteur de transcription, dont Sp1, s'est déjà révélé garant de l'activité d'un promoteur (144-146). D'autre part, le rôle de Sp1 pourrait ne pas se limiter à l'induction directe de la transcription. Il a été montré que les sites de liaison pour Sp1 au sein des promoteurs de gènes d'entretien étaient protégés de la métylation *de novo* responsable de l'inactivation permanente des gènes lors de l'embryogenèse (147). Sp1 pourrait donc participer à la régulation génique en créant un environnement libre de nucléosomes et du même coup favoriser l'assemblage du complexe actif de l'ARN polymérase II (148). Ceci pourrait survenir via l'interaction directe de Sp1 avec les sous-unités hTAF 135 et hTAF 55 de TFIID, ce qui favoriserait le recrutement de TFIID au niveau des promoteurs sans boîte TATA. À ce titre, il a aussi été montré qu'USF pouvait interagir avec hTAF 55 de TFIID. Or le recrutement de TFIID aux promoteurs de gènes à été défini comme étant une étape limitant l'initiation de la transcription. TFIID reconnaît généralement les boîtes TATA présentes dans les promoteurs via sa sous-unité TBP (TATA-binding protein). Même si le promoteur ne contient aucune boîte TATA, il a été montré que TFIID était présent au niveau des promoteurs sans boîte TATA dans une région localisée à environ 30 pb en amont du SIT et que son recrutement impliquerait des facteurs en amont de la région −30 ce qui pourrait, dans le cas d'*AβPP*, être Sp1 et USF (149).

Notre approche fut la première à étudier l'effet de mutations des sites de liaison pour Sp1 et USF séparément ou en combinaison en présence ou en absence de l'élément de liaison pour CTCF. Le rôle de ces facteurs dans la régulation d'*AβPP* reste difficile à définir. Nos résultats montrent que les sites de liaison pour Sp1, USF et CTCF agissent de façon additive sur l'activité du promoteur d'*AβPP*. Toutefois, en absence du site de liaison pour

CTCF, il semble y avoir une compétition entre la protéine reconnaissant le site Sp1 et l'élément USF puisqu'une mutation du site Sp1 entraîne une augmentation de l'activité du promoteur, alors que le la mutation de l'élément USF réduit l'activité du promoteur à un niveau similaire au double-mutant. Ces résultats indiquent que Sp1 requiert la présence de la protéine liant l'élément-GC pour activer le promoteur d'*AβPP* alors que ce n'est pas le cas pour USF.

La caractérisation du promoteur d'*AβPP* à l'aide de la technologie LMPCR a permis de révéler plusieurs séquences consensus où une interaction ADN-protéine était présente au sein des neurones et des astrocytes humains normaux. Nous avons tout d'abord confirmé la présence d'interactions au niveau des sites de liaison pour les facteurs CTCF, USF et Sp1 dans les cellules vivantes. Alors que la liaison de facteurs protéiques aux sites pour CTCF et USF est évidente, la liaison putative de Sp1 juste en amont d'USF dans les cellules vivantes semble soit moins fréquente au sein de la population cellulaire ou moins stable puisque nous avons seulement deux nucléotides protégés de façon significative de l'irradiation UVC et seulement sur un brin. Ces résultats n'affectent en rien l'implication de Sp1 dans la régulation d'*AβPP*, mais nous donne plutôt des indices sur la dynamique de liaison des facteurs au sein de la population cellulaire pour activer le gène. Combinés aux résultats des expériences de transfection et de compétition sur gel à retardement, ces résultats nous permettre d'émettre l'hypothèse que les rôles de CTCF et de USF dans l'activation d'*AβPP* nécessitent leur présence assidue à leurs sites respectifs de liaison alors que celle de Sp1 pourrait être épisodique.

Maintenant, plusieurs autres intercations ADN-protéines ont été localisées au niveau du promoteur d'*AβPP*. Parmi ces empreintes, certaines se situent à des sites dont on ignorait jusqu'ici l'implication dans la régulation de la transcription d'*AβPP*. L'analyse subséquente de la séquence révéla que ces empreintes se situaient au niveau des séquences consensus pour la liaison des facteurs NF-kB et de la protéine liant les boîtes CACC (CACC-binding protein), en l'occurrence, EKLF (erythroid kruppel-like factor). En plus de nouveaux

acteurs putatifs, l'analyse *in cellulo* du promoteur d'*AβPP* a mis en cause deux nouveaux sites de liaison pour Sp1 en plus de confirmer la présence d'une empreinte au site UE (Upstream Element) ayant déjà révélé une empreinte lors de l'analyse *in vitro*.

La présence d'empreintes en amont de la région prédéterminée comme étant seule responsable de l'activité complète du promoteur rétablit l'importance de la région −170 à −96 dans la régulation du gène. Comme nous le voyons à la figure 1 du chapitre 4, page 160, la délétion de la région de −364 à −96 n'entraîne pas de changements significatifs de l'activité CAT mesurée à un niveau moyen, mais diminue incontestablement la variabilité de celle-ci. Cette région pourrait donc être le centre d'une modulation fine de l'activité du promoteur. Nous pouvons aussi mesurer l'importance de cette région en absence des sites Sp1 et USF fonctionnels par l'augmentation de l'activité du promoteur −395mSU par-rapport au promoteur −96mSU lors de leur transfection au sein des deux types cellulaires. L'identification, l'implication et le rôle de tous ces facteurs dans la régulation de l'expression d'*AβPP* devront être déterminés par d'autres expériences.

Le patron d'interactions ADN-protéines révélé est exactement le même dans les neurones et les astrocytes. Selon toutes probabilités, ces régions et les facteurs qui s'y lient sont responsables de l'expression basale du gène *AβPP* dans les cellules normales. Il sera intéressant de comparer le patron d'interactions ADN-protéines au niveau du promoteur du gène *AβPP* que nous avons obtenu au sein de ces cellules normales à celui que l'on obtiendra peut-être un jour au sein de cellules neuronales ou d'astrocytes de patients qui souffrent de la maladie d'Alzheimer, et à celui d'autres types de cellules humaines normales exprimant à des niveaux divers le gène *AβPP*.

Les différences d'expression d'*AβPP* parmi les tissus humains peuvent être dues à la variation dans la proportion des facteurs aussi bien que de la combinaison des membres de

leurs familles, que de leurs modifications post-traductionnelles ou de l'action de co-facteurs particuliers. En ce qui concerne Sp1 et USF, aucun signe ne permet de supposer qu'ils subissent une régulation spécifique. Mais comme nous l'avons mentionné, d'autres acteurs et d'autres régions sont impliqués dans ce système et tous peuvent avoir un rôle majeur dans la régulation de l'expression d'*AβPP* ou être à l'origine de sa perturbation dans certaines conditions.

Dans les conditions où la sur-expression du gène *AβPP* mène à des symptômes associés à la maladie d'Alzheimer, par exemple chez les patients atteints du syndrome de Down, l'interférence avec les facteurs impliqués dans la régulation basale du gène, seule ou en combinaison avec d'autres thérapies, pourrait ralentir le développement de la maladie même si la source du signal menant à l'altération de l'expression du gène reste inconnue.

5.2 La conclusion

Dans ces travaux, nous avons tenté de mieux comprendre la régulation des gènes *p21*, *AβPP*, et *GPC3*, dont les patrons d'expression et la structure des promoteurs sont très différents. Hormis les informations importantes que ces travaux nous ont fournies sur les mécanismes qui régulent l'expression de ces gènes, ils nous permettent maintenant de dresser un portrait des stratégies cellulaires quant aux choix de ces mécanismes en rapport avec le patron d'expression des gènes. Il nous est aussi possible d'évaluer les approches utilisées dans chaque projet.

Le gène de la *protéine précurseure de l'amyloïde β (AβPP)* est exprimé de façon spécifique au niveau des tissus au cours du développement et est ubiquitaire dans la plupart des tissus adultes. Nos études sur des cellules primaires normales ont montré un patron clair d'empreintes protéiques au niveau du promoteur proximal du gène. Ces empreintes se situent au niveau de séquences régulatrices qui lient des facteurs de transcription déjà mis

en causes pour la régulation du gène *AβPP*, soit Sp1, CTCF et USF. De plus, nous avons localisé des empreintes au niveau de sites de liaison putatifs de facteurs de transcription jusqu'alors insoupçonnés de participer à la régulation de la transcription de ce gène (ex : NF-kB (–110 à –101), protéine qui lie les séquences CACC (–41 à –30)). Certaines de celles-ci se situent en amont de la région promotrice considérée jusqu'alors comme pleinement fonctionnelle (-96 à +100). Or, lors des analyses *in vitro* préalablement faites, ces régions n'avaient pas été incluses dans l'étude. Il est intéressant de comparer la méthodologie employée lors de l'étude du gène *AβPP* et de *p21*. Dans le premier cas, soucieux de confirmer les résultats obtenus par les méthodes *in vitro* traditionnelles, nous avons procédé à la cartographie des interactions ADN-protéines *in cellulo* à l'aide de la technique LMPCR. De cette manière, il ne fut pas possible d'intégrer l'analyse des régions insoupçonnées révélant des empreintes *in cellulo* dans les analyses *in vitro*. À l'inverse, la presque totalité de l'article portant sur la régulation du gène *p21* découle de la découverte de l'action d'un nouvel acteur, NFI, au niveau d'une région du promoteur jusqu'alors ignorée. Ceci montre que la cartographie intracellulaire des interactions ADN-protéines devrait être l'étape première de l'étude de la régulation de la transcription des gènes, ou tout au moins, suivre la transfection de promoteurs macro-délétés, ces dernières pouvant servir à établir les régions du promoteur à cibler lors de l'analyse *in cellulo*. De cette façon, nous évitons non seulement de négliger l'implication de certains facteurs, mais nous évitons de gaspiller temps et énergie à l'étude *in vitro* du rôle de facteurs putatifs alors qu'aucune donnée *in cellulo* valable n'implique ce facteur et son site putatif de liaison.

D'autre part, les empreintes mises en évidence sur le promoteur proximal d'*AβPP* et de *p21* restent très évidentes et concentrées au niveau de certains sites d'interaction strictes, c'est-à-dire qu'il n'y a pas d'ambiguïté sur l'atténuation ou l'augmentation de l'intensité des bandes et que les frontières des interactions sont généralement claires. Ceci n'est pas le cas des empreintes révélées au niveau du promoteur de *GPC3* dû à la présence probable d'une structure chromatinienne atypique du promoteur actif. *GPC3* est normalement régulé de façon spécifique au tissu lors de l'embryogenèse alors qu'il n'est exprimé dans aucun type cellulaire après la naissance. Même si le gène est situé sur le chromosome X, les lignés étant tous deux d'origine masculine, les empreintes ne peuvent donc pas être liées aux

phénomènes d'inactivation du chromosome X inactif (Xi). La structure chromatinienne mise en évidence est donc fort probablement liée à la répression généralisée du gène à la naissance. Cette régulation négative de *GPC3* étant, jusqu'à preuve du contraire, étendu à tout les types cellulaires chez l'adulte, il n'est pas surprenant qu'elle utilise un mécanisme plus généraliste que la perte de facteurs de transcription activateurs ou l'ajout de facteurs de transcription répresseurs. En effet, l'assemblage de l'ADN avec les nucléosomes et le glissement (ou le remodelage) de ceux-ci le long de l'ADN constituent des mécanismes de régulation de la transcription des gènes (150). Dans ce cas-ci, la liaison de facteurs de transcription n'est pas impossible mais l'initiation de la transcription du gène requiert une réorganisation préalable de la chromatine, qui est induite par la liaison de certains facteurs comme SWI/SNF (151) permettant l'accès des différents membres de la machinerie de transcription (Chapitre en annexe). Ce mécanisme est tellement important qu'il constitue un mécanisme de régulation génique essentiel à un grand nombre de fonctions cellulaires comme la différentiation, l'apoptose et le cycle cellulaire (152, 153). Néanmoins, dans les NB, la liaison de certains facteurs de transcription semble être impliqué dans la ré-activation anormale de *GPC3* sans que la structure soit déstabilisée à ce que l'on peut constater par l'absence de variations dans le patron d'empreintes entre les deux lignées, autour des empreintes spécifique à SJNB-7 (qui exprime *GPC3*).

Or, on peut se demander si la formation de cette structure est un mécanisme de régulation de la transcription utilisé de façon commune pour réprimer l'expression de gènes qui comme *GPC3*, ne doivent plus être exprimés après un certain stade de développement. Il sera donc essentiel, avant de conclure, d'analyser les promoteurs d'autres gènes ayant ce type d'expression et d'inclure dans ces études l'utilisation de types cellulaires non cancéreux. Il est intéressant de constater que le gène *GPC3* est fortement exprimé dans la lignée SJNB7 alors que cette structure particulière, supposée bloquer la transcription du gène, est toujours présente. Ce mécanisme d'inactivation génique n'est-il donc pas supposé être définitif? Si oui, la liaison d'activateur comme NF-Y est-il l'unique moyen que la cellule à l'origine de la lignée SJNB7 a trouvé pour contourner ce feu rouge?

Dans un autre ordre d'idée, les discordances entre les résultats de certaines études utilisant des techniques *in vitro* et les nôtres soulèvent quelques questions dont les réponses pourraient modifier notre approche de recherche actuelle. En effets, l'absence d'empreinte à des sites de régulation déjà mis en cause par d'autres équipes est devenue monnaie courante au cours de nos études. Le contraire, soit la découverte d'empreinte à des sites dont la séquence ne correspond pas à la séquence consensus d'aucun facteur de transcription connu ou dont l'implication dans la régulation du gène à l'étude était jusqu'alors inconnue (ex : NF1 pour *p21*), est aussi fréquent. Il est donc de plus en plus évident que la seule présence de séquences d'ADN pouvant liées un facteur de transcription n'est pas garant de leur implication dans la régulation de la transcription du gène dans la cellule vivante. En effet, les résultats des études sur *p21* et *AβPP* montrent que la réalité intranucléaire est bien plus complexe que la simple interaction linéaire entre un facteur de transcription unique et une séquence d'ADN isolée. À titre d'exemple, le positionnement des séquences de liaison pour les facteurs de transcription sur le promoteur relativement au SIT et entre elles mêmes est supposé être déterminant de leurs rôles comme en fait foi le concept des éléments composites (154-156). Selon ce concept, un premier facteur de transcription lié au promoteur pourrait servir d'ancre à un second qui, plutôt que de se lier à une séquence régulatrice de haute affinité ailleurs sur le promoteur, se liera préférentiellement à une séquence régulatrice de faible affinité aux abords du site de liaison du premier facteur. De plus, ces facteurs pourraient varier selon le type cellulaire pour la même séquence d'ADN. Pour toutes ces raisons, l'analyse intracellulaire des mécanismes de régulation est nécessaire à leur compréhension. Un jour prochain les techniques de génie génétique en médecine permettront d'utiliser les connaissances en matière de régulation génique pour développer et utiliser de nouvelles thérapies dont le potentiel s'annonce grandiose.

Bibliographie

1.	Lander, E.S. and Linton, L.M. and Birren, B. and Nusbaum, C. and Zody, M.C. and Baldwin, J. and Devon, K. and Dewar, K. and Doyle, M. and FitzHugh, W. *et al.* (2001) Initial sequencing and analysis of the human genome. *Nature,* **409,** 860-921.
2.	Roest Crollius, H., Jaillon, O., Bernot, A., Dasilva, C., Bouneau, L., Fischer, C., Fizames, C., Wincker, P., Brottier, P., Quetier, F. *et al.* (2000) Estimate of human gene number provided by genome-wide analysis using Tetraodon nigroviridis DNA sequence. *Nat Genet,* **25,** 235-8.
3.	Venter, J.C. and Adams, M.D. and Myers, E.W. and Li, P.W. and Mural, R.J. and Sutton, G.G. and Smith, H.O. and Yandell, M. and Evans, C.A. and Holt, R.A. *et al.* (2001) The sequence of the human genome. *Science,* **291,** 1304-51.
4.	Aparicio, S.A. (2000) How to count ... human genes. *Nat Genet,* **25,** 129-30.
5.	Liang, F., Holt, I., Pertea, G., Karamycheva, S., Salzberg, S.L. and Quackenbush, J. (2000) Gene index analysis of the human genome estimates approximately 120, 000 genes. *Nat Genet,* **25,** 239-40.
6.	Bentley, D.R. (2000) The Human Genome Project--an overview. *Med Res Rev,* **20,** 189-96.
7.	Broder, S. and Venter, J.C. (2000) Whole genomes: the foundation of new biology and medicine. *Curr Opin Biotechnol,* **11,** 581-5.
8.	Bentley, D.R. (2000) Decoding the human genome sequence. *Hum Mol Genet,* **9,** 2353-8.
9.	Rabbitts, T.H. (1994) Chromosomal translocations in human cancer. *Nature,* **372,** 143-9.
10.	Blum, Z. and Lidin, S. (1988) DNA packing in chromatine, a manifestation of the Bonnet transformation. *Acta Chem Scand B,* **42,** 417-22.
11.	Johnson, W. and Jameson, J.L. (1998) Transcriptional control of gene expression. In Jameson, J.L. (ed.), *Principles of molecular medecine.* Humana Press, Totowa, NJ., pp. 25-41.
12.	Davie, J.R. and Chadee, D.N. (1998) Regulation and regulatory parameters of histone modifications. *J Cell Biochem Suppl,* **31,** 203-13.
13.	Felsenfeld, G., Boyes, J., Chung, J., Clark, D. and Studitsky, V. (1996) Chromatin structure and gene expression. *Proc Natl Acad Sci U S A,* **93,** 9384-8.
14.	Kass, S.U., Landsberger, N. and Wolffe, A.P. (1997) DNA methylation directs a time-dependent repression of transcription initiation. *Curr Biol,* **7,** 157-65.
15.	Kingston, R.E., Bunker, C.A. and Imbalzano, A.N. (1996) Repression and activation by multiprotein complexes that alter chromatin structure. *Genes Dev,* **10,** 905-20.
16.	Sandaltzopoulos, R., Blank, T. and Becker, P.B. (1994) Transcriptional repression by nucleosomes but not H1 in reconstituted preblastoderm Drosophila chromatin. *Embo J,* **13,** 373-9.
17.	Beato, M. and Eisfeld, K. (1997) Transcription factor access to chromatin. *Nucleic Acids Res,* **25,** 3559-63.

18. Eberharter, A. and Becker, P.B. (2002) Histone acetylation: a switch between repressive and permissive chromatin. Second in review series on chromatin dynamics. *EMBO Rep,* **3,** 224-9.

19. Bouwman, P. and Philipsen, S. (2002) Regulation of the activity of Sp1-related transcription factors. *Mol Cell Endocrinol,* **195,** 27-38.

20. Shimano, H. (2001) Sterol regulatory element-binding proteins (SREBPs): transcriptional regulators of lipid synthetic genes. *Prog Lipid Res,* **40,** 439-52.

21. Wisdom, R. (1999) AP-1: one switch for many signals. *Exp Cell Res,* **253,** 180-5.

22. Cohen, B.A., Mitra, R.D., Hughes, J.D. and Church, G.M. (2000) A computational analysis of whole-genome expression data reveals chromosomal domains of gene expression. *Nat Genet,* **26,** 183-6.

23. Bex, F. and Gaynor, R.B. (1998) Regulation of gene expression by HTLV-I Tax protein. *Methods,* **16,** 83-94.

24. Stein, J.L., van Wijnen, A.J., Lian, J.B. and Stein, G.S. (1996) Control of cell cycle regulated histone genes during proliferation and differentiation. *Int J Obes Relat Metab Disord,* **20 Suppl 3,** S84-90.

25. Stein, D.C., Gunn, J.S. and Piekarowicz, A. (1998) Sequence similarities between the genes encoding the S.NgoI and HaeII restriction/modification systems. *Biol Chem,* **379,** 575-8.

26. Burke, T.W. and Kadonaga, J.T. (1996) Drosophila TFIID binds to a conserved downstream basal promoter element that is present in many TATA-box-deficient promoters. *Genes Dev,* **10,** 711-24.

27. Kaufmann, J. and Smale, S.T. (1994) Direct recognition of initiator elements by a component of the transcription factor IID complex. *Genes Dev,* **8,** 821-9.

28. Roeder, R.G. (1996) The role of general initiation factors in transcription by RNA polymerase II. *Trends Biochem Sci,* **21,** 327-35.

29. Roeder, R.G. (1991) The complexities of eukaryotic transcription initiation: regulation of preinitiation complex assembly. *Trends Biochem Sci,* **16,** 402-8.

30. Dorsett, D. (1999) Distant liaisons: long-range enhancer-promoter interactions in Drosophila. *Curr Opin Genet Dev,* **9,** 505-14.

31. Ogbourne, S. and Antalis, T.M. (1998) Transcriptional control and the role of silencers in transcriptional regulation in eukaryotes. *Biochem J,* **331,** 1-14.

32. Blackwood, E.M. and Kadonaga, J.T. (1998) Going the distance: a current view of enhancer action. *Science,* **281,** 61-3.

33. Grunstein, M. (1997) Histone acetylation in chromatin structure and transcription. *Nature,* **389,** 349-52.

34. Li, Q., Harju, S. and Peterson, K.R. (1999) Locus control regions: coming of age at a decade plus. *Trends Genet,* **15,** 403-8.

35. Milot, E., Strouboulis, J., Trimborn, T., Wijgerde, M., de Boer, E., Langeveld, A., Tan-Un, K., Vergeer, W., Yannoutsos, N., Grosveld, F. *et al.* (1996) Heterochromatin effects on the frequency and duration of LCR-mediated gene transcription. *Cell,* **87,** 105-14.

36. Curthoys, N.P. and Gstraunthaler, G. (2001) Mechanism of increased renal gene expression during metabolic acidosis. *Am J Physiol Renal Physiol,* **281,** F381-90.

37. Wessely, O., Deiner, E.M., Beug, H. and von Lindern, M. (1997) The glucocorticoid receptor is a key regulator of the decision between self-renewal and differentiation in erythroid progenitors. *Embo J,* **16,** 267-80.

38. Lebel, M., Gauthier, Y., Moreau, A. and Drouin, J. (2001) Pitx3 activates mouse tyrosine hydroxylase promoter via a high-affinity binding site. *J Neurochem,* **77,** 558-67.

39. Benes, C., Poitout, V., Marie, J.C., Martin-Perez, J., Roisin, M.P. and Fagard, R. (1999) Mode of regulation of the extracellular signal-regulated kinases in the pancreatic beta-cell line MIN6 and their implication in the regulation of insulin gene transcription. *Biochem J,* **340 (Pt 1),** 219-25.

40. Freeman, M. and Gurdon, J.B. (2002) Regulatory principles of developmental signaling. *Annu Rev Cell Dev Biol,* **18,** 515-39.

41. Chen, C.Y. and Shyu, A.B. (1994) Selective degradation of early-response-gene mRNAs: functional analyses of sequence features of the AU-rich elements. *Mol Cell Biol,* **14,** 8471-82.

42. Dever, T.E. (2002) Gene-specific regulation by general translation factors. *Cell,* **108,** 545-56.

43. Menotti, E., Henderson, B.R. and Kuhn, L.C. (1998) Translational regulation of mRNAs with distinct IRE sequences by iron regulatory proteins 1 and 2. *J Biol Chem,* **273,** 1821-4.

44. Gottifredi, V., Shieh, S., Taya, Y. and Prives, C. (2001) From the Cover: p53 accumulates but is functionally impaired when DNA synthesis is blocked. *Proc Natl Acad Sci U S A,* **98,** 1036-41.

45. Kolchanov, N.A., Ignatieva, E.V., Ananko, E.A., Podkolodnaya, O.A., Stepanenko, I.L., Merkulova, T.I., Pozdnyakov, M.A., Podkolodny, N.L., Naumochkin, A.N. and Romashchenko, A.G. (2002) Transcription Regulatory Regions Database (TRRD): its status in 2002. *Nucleic Acids Res,* **30,** 312-7.

46. Prestridge, D.S. (2000) Computer software for eukaryotic promoter analysis. *Methods Mol Biol,* **130,** 265-95.

47. Hai, T. and Hartman, M.G. (2001) The molecular biology and nomenclature of the activating transcription factor/cAMP responsive element binding family of transcription factors: activating transcription factor proteins and homeostasis. *Gene,* **273,** 1-11.

48. Poulin, G., Lebel, M., Chamberland, M., Paradis, F.W. and Drouin, J. (2000) Specific protein-protein interaction between basic helix-loop-helix transcription factors and homeoproteins of the Pitx family. *Mol Cell Biol,* **20,** 4826-37.

49. Fried, M. and Crothers, D.M. (1981) Equilibria and kinetics of lac repressor-operator interactions by polyacrylamide gel electrophoresis. *Nucleic Acids Res,* **9,** 6505-25.

50. Garner, M.M. and Revzin, A. (1981) A gel electrophoresis method for quantifying the binding of proteins to specific DNA regions: application to components of the Escherichia coli lactose operon regulatory system. *Nucleic Acids Res,* **9,** 3047-60.

51. Laniel, M.A., Beliveau, A. and Guerin, S.L. (2001) Electrophoretic mobility shift assays for the analysis of DNA-protein interactions. *Methods Mol Biol,* **148,** 13-30.

52. Gille, J., Swerlick, R.A. and Caughman, S.W. (1997) Transforming growth factor-alpha-induced transcriptional activation of the vascular permeability factor (VPF/VEGF) gene requires AP-2-dependent DNA binding and transactivation. *Embo J,* **16,** 750-9.

53. Zhang, W., Shields, J.M., Sogawa, K., Fujii-Kuriyama, Y. and Yang, V.W. (1998) The gut-enriched Kruppel-like factor suppresses the activity of the CYP1A1 promoter in an Sp1-dependent fashion. *J Biol Chem,* **273,** 17917-25.

54. Gerstle, J.T. and Fried, M.G. (1993) Measurement of binding kinetics using the gel electrophoresis mobility shift assay. *Electrophoresis,* **14,** 725-31.

55. Bergeron, M.J., Leclerc, S., Laniel, M.A., Poirier, G.G. and Guerin, S.L. (1997) Transcriptional regulation of the rat poly(ADP-ribose) polymerase gene by Sp1. *Eur J Biochem,* **250,** 342-53.

56. Suck, D., Lahm, A. and Oefner, C. (1988) Structure refined to 2A of a nicked DNA octanucleotide complex with DNase I. *Nature,* **332,** 464-8.

57. Leblanc, B. and Moss, T. (2001) DNase I footprinting. *Methods Mol Biol,* **148,** 31-8.

58. Shaw, P.E. and Stewart, A.F. (2001) Identification of protein-DNA contacts with dimethyl sulfate. Methylation protection and methylation interference. *Methods Mol Biol,* **148,** 221-7.

59. Manfield, I.W. and Stockley, P.G. (2001) Ethylation interference. *Methods Mol Biol,* **148,** 229-43.

60. Molnar, G., O'Leary, N., Pardee, A.B. and Bradley, D.W. (1995) Quantification of DNA-protein interaction by UV crosslinking. *Nucleic Acids Res,* **23,** 3318-26.

61. Metzger, W. and Heumann, H. (2001) Footprinting with exonuclease III. *Methods Mol Biol,* **148,** 39-47.

62. Kahl, B.F. and Paule, M.R. (2001) The use of diethyl pyrocarbonate and potassium permanganate as probes for strand separation and structural distortions in DNA. *Methods Mol Biol,* **148,** 63-75.

63. Rubin, C.M. and Schmid, C.W. (1980) Pyrimidine-specific chemical reactions useful for DNA sequencing. *Nucleic Acids Res,* **8,** 4613-9.

64. McCarthy, J.G. and Rich, A. (1991) Detection of an unusual distortion in A-tract DNA using KMnO4: effect of temperature and distamycin on the altered conformation. *Nucleic Acids Res,* **19,** 3421-9.

65. Balagurumoorthy, P. and Brahmachari, S.K. (1994) Structure and stability of human telomeric sequence. *J Biol Chem,* **269,** 21858-69.

66. Jiang, H., Zacharias, W. and Amirhaeri, S. (1991) Potassium permanganate as an in situ probe for B-Z and Z-Z junctions. *Nucleic Acids Res,* **19,** 6943-8.

67. Razin, A. (1998) CpG methylation, chromatin structure and gene silencing-a three-way connection. *Embo J,* **17,** 4905-8.

68. Nguyen, V.T., Morange, M. and Bensaude, O. (1988) Firefly luciferase luminescence assays using scintillation counters for quantitation in transfected mammalian cells. *Anal Biochem,* **171,** 404-8.

69. Rhodius, V., Savery, N., Kolb, A. and Busby, S. (2001) Assays for transcription factor activity. *Methods Mol Biol,* **148,** 451-64.

70. Cartwright, I.L. and Kelly, S.E. (1991) Probing the nature of chromosomal DNA-protein contacts by in vivo footprinting. *Biotechniques,* **11,** 188-90, 192-4, 196 passim.

71. Church, G.M. and Gilbert, W. (1984) Genomic sequencing. *Proc Natl Acad Sci U S A,* **81,** 1991-5.

72. Ephrussi, A., Church, G.M., Tonegawa, S. and Gilbert, W. (1985) B lineage--specific interactions of an immunoglobulin enhancer with cellular factors in vivo. *Science,* **227,** 134-40.

73. Giniger, E., Varnum, S.M. and Ptashne, M. (1985) Specific DNA binding of GAL4, a positive regulatory protein of yeast. *Cell,* **40,** 767-74.

74. Nick, H. and Gilbert, W. (1985) Detection in vivo of protein-DNA interactions within the lac operon of Escherichia coli. *Nature,* **313,** 795-8.

75. Drouin, R., Therrien, J.-P., Angers, M. and Ouellet, S. (2001) In vivo DNA analysis. In Moss, T. (ed.), *DNA-Protein Interactions: Principles and Protocols.* second edition ed. Humana Press, Totowa, NJ., pp. 175-219.

76. Chen, C.J., Li, L.J., Maruya, A. and Shively, J.E. (1995) In vitro and in vivo footprint analysis of the promoter of carcinoembryonic antigen in colon carcinoma cells: effects of interferon gamma treatment. *Cancer Res,* **55,** 3873-82.

77. Pfeifer, G.P. and Riggs, A.D. (1991) Chromatin differences between active and inactive X chromosomes revealed by genomic footprinting of permeabilized cells using DNase I and ligation-mediated PCR. *Genes Dev,* **5,** 1102-13.

78. Mattes, W.B., Hartley, J.A. and Kohn, K.W. (1986) Mechanism of DNA strand breakage by piperidine at sites of N7-alkylguanines. *Biochim Biophys Acta,* **868,** 71-6.

79. Brash, D.E. (1988) UV mutagenic photoproducts in Escherichia coli and human cells: a molecular genetics perspective on human skin cancer. *Photochem Photobiol,* **48,** 59-66.

80. Cadet, J., Anselmino, C., Douki, T. and Voituriez, L. (1992) Photochemistry of nucleic acids in cells. *J Photochem Photobiol B,* **15,** 277-98.

81. Tornaletti, S. and Pfeifer, G.P. (1995) UV light as a footprinting agent: modulation of UV-induced DNA damage by transcription factors bound at the promoters of three human genes. *J Mol Biol,* **249,** 714-28.

82. Maxam, A.M. and Gilbert, W. (1980) Sequencing end-labeled DNA with base-specific chemical cleavages. *Methods Enzymol,* **65,** 499-560.

83. Iverson, B.L. and Dervan, P.B. (1987) Adenine specific DNA chemical sequencing reaction. *Nucleic Acids Res,* **15,** 7823-30.

84. Mueller, P.R. and Wold, B. (1989) In vivo footprinting of a muscle specific enhancer by ligation mediated PCR. *Science,* **246,** 780-6.

85. Cloutier, J.-F., Castonguay, A., O'Connor, T.R. and Drouin, R. (2001) Alkylating agent and chromatin structure determine sequence context- dependent formation of alkylpurines. *J Mol Biol,* **306,** 169-88.

86. Denissenko, M.F., Pao, A., Tang, M. and Pfeifer, G.P. (1996) Preferential formation of benzo[a]pyrene adducts at lung cancer mutational hotspots in P53. *Science,* **274,** 430-2.

87. Therrien, J.-P., Loignon, M., Drouin, R. and Drobetsky, E.A. (2001) Ablation of p21waf1cip1 expression enhances the capacity of p53- deficient human tumor cells to repair UVB-induced DNA damage. *Cancer Res,* **61,** 3781-6.

88. Hornstra, I.K., Nelson, D.L., Warren, S.T. and Yang, T.P. (1993) High resolution methylation analysis of the FMR1 gene trinucleotide repeat region in fragile X syndrome. *Hum Mol Genet,* **2,** 1659-65.

89.   Pfeifer, G.P., Steigerwald, S.D., Mueller, P.R., Wold, B. and Riggs, A.D. (1989) Genomic sequencing and methylation analysis by ligation mediated PCR. *Science,* **246,** 810-3.

90.   Pfeifer, G.P., Tanguay, R.L., Steigerwald, S.D. and Riggs, A.D. (1990) In vivo footprint and methylation analysis by PCR-aided genomic sequencing: comparison of active and inactive X chromosomal DNA at the CpG island and promoter of human PGK-1. *Genes Dev,* **4,** 1277-87.

91.   Goldring, C.E., Reveneau, S., Algarte, M. and Jeannin, J.F. (1996) In vivo footprinting of the mouse inducible nitric oxide synthase gene: inducible protein occupation of numerous sites including Oct and NF-IL6. *Nucleic Acids Res,* **24,** 1682-7.

92.   Rigaud, G., Roux, J., Pictet, R. and Grange, T. (1991) In vivo footprinting of rat TAT gene: dynamic interplay between the glucocorticoid receptor and a liver-specific factor. *Cell,* **67,** 977-86.

93.   Tommasi, S. and Pfeifer, G.P. (1995) In vivo structure of the human cdc2 promoter: release of a p130-E2F-4 complex from sequences immediately upstream of the transcription initiation site coincides with induction of cdc2 expression. *Mol Cell Biol,* **15,** 6901-13.

94.   Rozek, D. and Pfeifer, G.P. (1993) In vivo protein-DNA interactions at the c-jun promoter: preformed complexes mediate the UV response. *Mol Cell Biol,* **13,** 5490-9.

95.   Rozek, D. and Pfeifer, G.P. (1995) In vivo protein-DNA interactions at the c-jun promoter in quiescent and serum-stimulated fibroblasts. *J Cell Biochem,* **57,** 479-87.

96.   Solomon, M.J. and Varshavsky, A. (1985) Formaldehyde-mediated DNA-protein crosslinking: a probe for in vivo chromatin structures. *Proc Natl Acad Sci U S A,* **82,** 6470-4.

97.   Wells, J. and Farnham, P.J. (2002) Characterizing transcription factor binding sites using formaldehyde crosslinking and immunoprecipitation. *Methods,* **26,** 48-56.

98.   Pilia, G., Hughes-Benzie, R.M., MacKenzie, A., Baybayan, P., Chen, E.Y., Huber, R., Neri, G., Cao, A., Forabosco, A. and Schlessinger, D. (1996) Mutations in GPC3, a glypican gene, cause the Simpson-Golabi-Behmel overgrowth syndrome. *Nat Genet,* **12,** 241-7.

99.   Cano-Gauci, D.F., Song, H.H., Yang, H., McKerlie, C., Choo, B., Shi, W., Pullano, R., Piscione, T.D., Grisaru, S., Soon, S. *et al.* (1999) Glypican-3-deficient mice exhibit developmental overgrowth and some of the abnormalities typical of Simpson-Golabi-Behmel syndrome. *J Cell Biol,* **146,** 255-64.

100.   Murthy, S.S., Shen, T., De Rienzo, A., Lee, W.C., Ferriola, P.C., Jhanwar, S.C., Mossman, B.T., Filmus, J. and Testa, J.R. (2000) Expression of GPC3, an X-linked recessive overgrowth gene, is silenced in malignant mesothelioma. *Oncogene,* **19,** 410-6.

101.   Weksberg, R., Squire, J.A. and Templeton, D.M. (1996) Glypicans: a growing trend. *Nat Genet,* **12,** 225-7.

102.   Hsu, H.C., Cheng, W. and Lai, P.L. (1997) Cloning and expression of a developmentally regulated transcript MXR7 in hepatocellular carcinoma: biological significance and temporospatial distribution. *Cancer Res,* **57,** 5179-84.

103.   Midorikawa, Y., Ishikawa, S., Iwanari, H., Imamura, T., Sakamoto, H., Miyazono, K., Kodama, T., Makuuchi, M. and Aburatani, H. (2003) Glypican-3, overexpressed

in hepatocellular carcinoma, modulates FGF2 and BMP-7 signaling. *Int J Cancer,* **103,** 455-65.

104. Toretsky, J.A., Zitomersky, N.L., Eskenazi, A.E., Voigt, R.W., Strauch, E.D., Sun, C.C., Huber, R., Meltzer, S.J. and Schlessinger, D. (2001) Glypican-3 expression in Wilms tumor and hepatoblastoma. *J Pediatr Hematol Oncol,* **23,** 496-9.

105. Zhu, Z.W., Friess, H., Wang, L., Abou-Shady, M., Zimmermann, A., Lander, A.D., Korc, M., Kleeff, J. and Buchler, M.W. (2001) Enhanced glypican-3 expression differentiates the majority of hepatocellular carcinomas from benign hepatic disorders. *Gut,* **48,** 558-64.

106. Lage, H., Dietel, M., Froschle, G. and Reymann, A. (1998) Expression of the novel mitoxantrone resistance associated gene MXR7 in colorectal malignancies. *Int J Clin Pharmacol Ther,* **36,** 58-60.

107. Filmus, J. (2001) Glypicans in growth control and cancer. *Glycobiology,* **11,** 19R-23R.

108. Boily, G., Saikali, Z. and Sinnett, D. (2004) Methylation analysis of the glypican 3 gene in embryonal tumours. *Br J Cancer,* **90,** 1606-11.

109. Huber, R., Schlessinger, D. and Pilia, G. (1998) Multiple Sp1 sites efficiently drive transcription of the TATA-less promoter of the human glypican 3 (GPC3) gene. *Gene,* **214,** 35-44.

110. Pollwein, P., Masters, C.L. and Beyreuther, K. (1992) The expression of the amyloid precursor protein (APP) is regulated by two GC-elements in the promoter. *Nucleic Acids Res,* **20,** 63-8.

111. Quitschke, W.W. and Goldgaber, D. (1992) The amyloid beta-protein precursor promoter. A region essential for transcriptional activity contains a nuclear factor binding domain. *J Biol Chem,* **267,** 17362-8.

112. Izumi, R., Yamada, T., Yoshikai, S., Sasaki, H., Hattori, M. and Sakaki, Y. (1992) Positive and negative regulatory elements for the expression of the Alzheimer's disease amyloid precursor-encoding gene in mouse. *Gene,* **112,** 189-95.

113. Gronostajski, R.M. (2000) Roles of the NFI/CTF gene family in transcription and development. *Gene,* **249,** 31-45.

114. Kruse, U. and Sippel, A.E. (1994) Transcription factor nuclear factor I proteins form stable homo- and heterodimers. *FEBS Lett,* **348,** 46-50.

115. Kane, R., Murtagh, J., Finlay, D., Marti, A., Jaggi, R., Blatchford, D., Wilde, C. and Martin, F. (2002) Transcription factor NFIC undergoes N-glycosylation during early mammary gland involution. *J Biol Chem,* **277,** 25893-903.

116. Marin, M., Karis, A., Visser, P., Grosveld, F. and Philipsen, S. (1997) Transcription factor Sp1 is essential for early embryonic development but dispensable for cell growth and differentiation. *Cell,* **89,** 619-28.

117. Osada, S., Matsubara, T., Daimon, S., Terazu, Y., Xu, M., Nishihara, T. and Imagawa, M. (1999) Expression, DNA-binding specificity and transcriptional regulation of nuclear factor 1 family proteins from rat. *Biochem J,* **342 (Pt 1),** 189-98.

118. Gao, B. and Kunos, G. (1998) Cell type-specific transcriptional activation and suppression of the alpha1B adrenergic receptor gene middle promoter by nuclear factor 1. *J Biol Chem,* **273,** 31784-7.

119. Laniel, M.A., Poirier, G.G. and Guerin, S.L. (2001) Nuclear factor 1 interferes with Sp1 binding through a composite element on the rat poly(ADP-ribose) polymerase promoter to modulate its activity in vitro. *J Biol Chem,* **276,** 20766-73.

120. Nakamura, M., Okura, T., Kitami, Y. and Hiwada, K. (2001) Nuclear factor 1 is a negative regulator of gadd153 gene expression in vascular smooth muscle cells. *Hypertension,* **37,** 419-24.

121. Steffensen, K.R., Holter, E., Tobin, K.A., Leclerc, S., Gustafsson, J.A., Guerin, S.L. and Eskild, W. (2001) Members of the nuclear factor 1 family reduce the transcriptional potential of the nuclear receptor LXRalpha promoter. *Biochem Biophys Res Commun,* **289,** 1262-7.

122. Clark, R.E., Jr., Miskimins, W.K. and Miskimins, R. (2002) Cyclic AMP inducibility of the myelin basic protein gene promoter requires the NF1 site. *Int J Dev Neurosci,* **20,** 103-11.

123. Gao, B., Jiang, L. and Kunos, G. (1996) Transcriptional regulation of alpha(1b) adrenergic receptors (alpha(1b)AR) by nuclear factor 1 (NF1): a decline in the concentration of NF1 correlates with the downregulation of alpha(1b)AR gene expression in regenerating liver. *Mol Cell Biol,* **16,** 5997-6008.

124. Rafty, L.A., Santiago, F.S. and Khachigian, L.M. (2002) NF1/X represses PDGF A-chain transcription by interacting with Sp1 and antagonizing Sp1 occupancy of the promoter. *Embo J,* **21,** 334-43.

125. Yamada, K., Tanaka, T. and Noguchi, T. (1997) Members of the nuclear factor 1 family and hepatocyte nuclear factor 4 bind to overlapping sequences of the L-II element on the rat pyruvate kinase L gene promoter and regulate its expression. *Biochem J,* **324 (Pt 3),** 917-25.

126. Koutsodontis, G., Moustakas, A. and Kardassis, D. (2002) The role of Sp1 family members, the proximal GC-rich motifs, and the upstream enhancer region in the regulation of the human cell cycle inhibitor p21WAF-1/Cip1 gene promoter. *Biochemistry,* **41,** 12771-84.

127. Kim, Y.K., Han, J.W., Woo, Y.N., Chun, J.K., Yoo, J.Y., Cho, E.J., Hong, S., Lee, H.Y., Lee, Y.W. and Lee, H.W. (2003) Expression of p21(WAF1/Cip1) through Sp1 sites by histone deacetylase inhibitor apicidin requires PI 3-kinase-PKC epsilon signaling pathway. *Oncogene,* **22,** 6023-31.

128. Koutsodontis, G., Tentes, I., Papakosta, P., Moustakas, A. and Kardassis, D. (2001) Sp1 plays a critical role in the transcriptional activation of the human cyclin-dependent kinase inhibitor p21(WAF1/Cip1) gene by the p53 tumor suppressor protein. *J Biol Chem,* **276,** 29116-25.

129. Han, J.W., Ahn, S.H., Kim, Y.K., Bae, G.U., Yoon, J.W., Hong, S., Lee, H.Y., Lee, Y.W. and Lee, H.W. (2001) Activation of p21(WAF1/Cip1) transcription through Sp1 sites by histone deacetylase inhibitor apicidin: involvement of protein kinase C. *J Biol Chem,* **276,** 42084-90.

130. Lagger, G., Doetzlhofer, A., Schuettengruber, B., Haidweger, E., Simboeck, E., Tischler, J., Chiocca, S., Suske, G., Rotheneder, H., Wintersberger, E. *et al.* (2003) The tumor suppressor p53 and histone deacetylase 1 are antagonistic regulators of the cyclin-dependent kinase inhibitor p21/WAF1/CIP1 gene. *Mol Cell Biol,* **23,** 2669-79.

131. Huang, L., Sowa, Y., Sakai, T. and Pardee, A.B. (2000) Activation of the p21WAF1/CIP1 promoter independent of p53 by the histone deacetylase inhibitor

suberoylanilide hydroxamic acid (SAHA) through the Sp1 sites. *Oncogene,* **19,** 5712-9.

132. Xiao, H., Hasegawa, T. and Isobe, K. (1999) Both Sp1 and Sp3 are responsible for p21waf1 promoter activity induced by histone deacetylase inhibitor in NIH3T3 cells. *J Cell Biochem,* **73,** 291-302.

133. Araki, K., Nakajima, Y., Eto, K. and Ikeda, M.A. (2003) Distinct recruitment of E2F family members to specific E2F-binding sites mediates activation and repression of the E2F1 promoter. *Oncogene,* **22,** 7632-41.

134. Gartel, A.L., Najmabadi, F., Goufman, E. and Tyner, A.L. (2000) A role for E2F1 in Ras activation of p21(WAF1/CIP1) transcription. *Oncogene,* **19,** 961-4.

135. Hiyama, H., Iavarone, A. and Reeves, S.A. (1998) Regulation of the cdk inhibitor p21 gene during cell cycle progression is under the control of the transcription factor E2F. *Oncogene,* **16,** 1513-23.

136. Besson, A. and Yong, V.W. (2000) Involvement of p21(Waf1/Cip1) in protein kinase C alpha-induced cell cycle progression. *Mol Cell Biol,* **20,** 4580-90.

137. Prowse, D.M., Bolgan, L., Molnar, A. and Dotto, G.P. (1997) Involvement of the Sp3 transcription factor in induction of p21Cip1/WAF1 in keratinocyte differentiation. *J Biol Chem,* **272,** 1308-14.

138. Kennett, S.B., Udvadia, A.J. and Horowitz, J.M. (1997) Sp3 encodes multiple proteins that differ in their capacity to stimulate or repress transcription. *Nucleic Acids Res,* **25,** 3110-7.

139. Afshari, C.A., Nichols, M.A., Xiong, Y. and Mudryj, M. (1996) A role for a p21-E2F interaction during senescence arrest of normal human fibroblasts. *Cell Growth Differ,* **7,** 979-88.

140. Delavaine, L. and La Thangue, N.B. (1999) Control of E2F activity by p21Waf1/Cip1. *Oncogene,* **18,** 5381-92.

141. Fotedar, R., Fitzgerald, P., Rousselle, T., Cannella, D., Doree, M., Messier, H. and Fotedar, A. (1996) p21 contains independent binding sites for cyclin and cdk2: both sites are required to inhibit cdk2 kinase activity. *Oncogene,* **12,** 2155-64.

142. Linke, S.P., Harris, M.P., Neugebauer, S.E., Clarkin, K.C., Shepard, H.M., Maneval, D.C. and Wahl, G.M. (1997) p53-mediated accumulation of hypophosphorylated pRb after the G1 restriction point fails to halt cell cycle progression. *Oncogene,* **15,** 337-45.

143. Hoffman, P.W. and Chernak, J.M. (1995) DNA binding and regulatory effects of transcription factors SP1 and USF at the rat amyloid precursor protein gene promoter. *Nucleic Acids Res,* **23,** 2229-35.

144. Force, W.R., Tillman, J.B., Sprung, C.N. and Spindler, S.R. (1994) Homodimer and heterodimer DNA binding and transcriptional responsiveness to triiodothyronine (T3) and 9-cis-retinoic acid are determined by the number and order of high affinity half-sites in a T3 response element. *J Biol Chem,* **269,** 8863-71.

145. Miyahara, K., Kuge, H., Shizuta, Y. and Honke, K. (2004) Three repeats of CCCCTCC on the pyrimidine-rich sequence in the proximal 5' flanking region are required for efficient transcriptional activity of the human endothelial nitric oxide synthase gene. *Free Radic Res,* **38,** 87-95.

146. Zeiser, S., Liebscher, H.V., Tiedemann, H., Rubio-Aliaga, I., Przemeck, G.K., de Angelis, M.H. and Winkler, G. (2006) Number of active transcription factor binding sites is essential for the Hes7 oscillator. *Theor Biol Med Model,* **3,** 11.

147. Brandeis, M., Frank, D., Keshet, I., Siegfried, Z., Mendelsohn, M., Nemes, A., Temper, V., Razin, A. and Cedar, H. (1994) Sp1 elements protect a CpG island from de novo methylation. *Nature,* **371,** 435-8.

148. Li, B., Adams, C.C. and Workman, J.L. (1994) Nucleosome binding by the constitutive transcription factor Sp1. *J Biol Chem,* **269,** 7756-63.

149. Wiley, S.R., Kraus, R.J. and Mertz, J.E. (1992) Functional binding of the "TATA" box binding component of transcription factor TFIID to the -30 region of TATA-less promoters. *Proc Natl Acad Sci U S A,* **89,** 5814-8.

150. Lomvardas, S. and Thanos, D. (2001) Nucleosome sliding via TBP DNA binding in vivo. *Cell,* **106,** 685-96.

151. Lee, D., Kim, J.W., Seo, T., Hwang, S.G., Choi, E.J. and Choe, J. (2002) SWI/SNF complex interacts with tumor suppressor p53 and is necessary for the activation of p53-mediated transcription. *J Biol Chem,* **277,** 22330-7.

152. Gresh, L., Bourachot, B., Reimann, A., Guigas, B., Fiette, L., Garbay, S., Muchardt, C., Hue, L., Pontoglio, M., Yaniv, M. *et al.* (2005) The SWI/SNF chromatin-remodeling complex subunit SNF5 is essential for hepatocyte differentiation. *Embo J,* **24,** 3313-24.

153. Muchardt, C. and Yaniv, M. (2001) When the SWI/SNF complex remodels...the cell cycle. *Oncogene,* **20,** 3067-75.

154. Kel-Margoulis, O.V., Kel, A.E., Reuter, I., Deineko, I.V. and Wingender, E. (2002) TRANSCompel: a database on composite regulatory elements in eukaryotic genes. *Nucleic Acids Res,* **30,** 332-4.

155. Rao, A., Luo, C. and Hogan, P.G. (1997) Transcription factors of the NFAT family: regulation and function. *Annu Rev Immunol,* **15,** 707-47.

156. Ramirez-Carrozzi, V. and Kerppola, T. (2003) Asymmetric recognition of nonconsensus AP-1 sites by Fos-Jun and Jun-Jun influences transcriptional cooperativity with NFAT1. *Mol Cell Biol,* **23,** 1737-49.

157. Smith, TM., Lee, MK., Szabo, CL., Jerome, N., McEuen, M., Taylor, M., Hood, L. and King,MC. (1996) Complet genomic sequence and analysis of 117 kb of human DNA containing the gene BRCA1. *Genome Res,* **11,** 1029-49.

158. Suen, TC. and Goss, PE. (2001) Identification of a novel transcriptional repressor element located in the first intron of the human BRCA1 gene. *Oncogene,* **20(4),** 440-50.

159. Menon, SD., Qin, S., Guy, GR. and Tan, YH. (1993) Differential induction of nuclear NF-kappa B by protein phosphatase inhibitors in primary and transformed human cells. Requirement for both oxidation and phosphorylation in nuclear translocation. *J Biol Chem,* **268(35),** 26805-12.

Annexe

Chapitre de livre : In vivo DNA analysis (anglais)

In Vivo DNA Analysis

Régen Drouin, Jean-Philippe Therrien, Martin Angers and Stéphane Ouellet

Division of Pathology, Department of Medical Biology, Faculty of Medicine, Laval University and Research Unit in Human and Molecular Genetics, Research Centre, Hôpital Saint-François d'Assise, Centre Hospitalier Universitaire de Québec, Québec (Québec), Canada, G1L 3L5

<u>**Corresponding address**</u>:
 Régen Drouin
 Research Unit in Human and Molecular Genetics
 Research Centre
 Hôpital Saint-François d'Assise, CHUQ
 10 de l'Espinay street
 Québec (Québec)
 Canada   G1L 3L5
  Tel.:   (418) 525-4402
  Fax: (418) 525-4195
  e-mail: regen.drouin@crsfa.ulaval.ca

1 Introduction

The in vivo analysis of DNA-protein interactions and chromatin structure can provide several kinds of critical information regarding regulation of gene expression and gene function. For example, DNA sequences spanned by nuclease-hypersensitive sites or bound by transcription factors often correspond to genetic regulatory elements. Using the ligation-mediated polymerase chain reaction (LMPCR) technology it is possible to map such DNA sequences and to demonstrate the existence of unusual DNA structures directly in living cells. LMPCR analyses can thus be used as a primary investigative tool to identify the regulatory sequences involved in gene expression. Once specific promoter sequence sites shown to be bound by transcription factors in living cells, it is often possible to establish the identity of these factors simply by comparison with the consensus binding sites of known factors such as Sp1, AP-1, NF-1, and so forth. The identity of each factor can then be confirmed using in vitro gel shift (electrophoretic mobility shift assay [EMSA]) or footprinting assays.

Clearly, gene promoters are best studied in their natural state in the living cell and, thus, it is not surprising that in vivo DNA footprinting is one of the most accurate predictors of the state of transcriptional activity of genes *(1-3)*. The native state of a gene and most of the special DNA structures are unavoidably lost when DNA is cloned or purified *(1-4)*. Hence, the commonly used in vitro methods, such as in vitro footprinting and EMSAs, cannot demonstrate that a given DNA-protein interaction actually occurs within the cell of interest. With the advent of in vivo DNA footprinting, in vitro studies have been extended to the situation in living cells, revealing the cellular processes implicated in the regulation of gene expression. LMPCR is the method of choice for in vivo footprinting and DNA structure studies because it can be used to investigate complex animal genomes, including that of human. The quality and usefulness of the information obtained from any in vivo DNA analysis, however, depends on three parameters: (1) the integrity of the native chromatin substrate used in the experiment, (2) the structural specificity of the chromatin probe, and (3) the sensitivity of the assay. The ideal chromatin substrate is, of course, that found inside

intact cells. However, a near-ideal chromatin substrate is still to be found in permeabilized cells, allowing the application of a wider range of DNA cleavage agents, including DNase I.

In vivo footprinting assesses the local reactivity of modifying agents on the DNA of living cells as compared to that on purified DNA (*see* **Figs. 1-4**). Two steps characterize an in vivo footprinting analysis: (1) the treatment of purified DNA and of cells with a given modifying agent and (2) the visualization of nucleotide modifications on a DNA sequencing gel. The latter step requires that the modifying agent either directly induces DNA strand breaks or modifies DNA nucleotides such that strand breaks can subsequently be induced in vitro. A comparison is then made between the modification frequency on purified DNA and that on the DNA in living cells. For example, each guanine residue of purified DNA has a near-equivalent probability of being methylated by dimethylsulfate (DMS) and, thus, the cleavage pattern of in vitro modified DNA appears on a sequencing gel as a ladder of bands of roughly equal intensity. However, as a result of the presence of DNA-binding proteins, all guanine residues do not show the same accessibility to DMS in living cells (**Fig. 1**). Thus, differences between banding patterns obtained from in vitro and in vivo modified DNA can be used to infer the sites of protein binding in living cells. As will be seen, it is always advisable to validate such interpretations using more than one footprinting agent.

The step of visualizing in vivo footprints has historically been problematic because of the dilute nature of the sequences of interest and the complexity of the genomes of higher eukaryotes. The development of an extremely sensitive and specific technique, such as LMPCR, was thus necessary. The LMPCR technique quantitatively maps single-strand DNA breaks having phosphorylated 5' ends within single-copy DNA sequences. It was first developed by Mueller and Wold *(5)* for DMS footprinting, and, subsequently, Pfeifer and colleagues adapted it to DNA sequencing *(6)*, methylation analyses *(1,6,7)*, DNase I footprinting *(2)*, nucleosome positioning *(2)* and UV photofootprinting *(4,8)*. LMPCR can be combined with a variety of DNA-modifying agents used to probe the chromatin structure in vivo. It is our opinion that no single technique can provide as much information

on the DNA-protein interactions and DNA structures existing within living cells as can LMPCR.

1.1 General Overview of LMPCR

Genomic sequencing techniques such as that developed by Church and Gilbert *(9)* can be used to map strand breaks in mammalian genes at nucleotide resolution. However, by incorporating an exponential amplification step, LMPCR (outlined in **Fig. 5**) constitutes a genomic sequencing method orders of magnitude more sensitive than the direct technique of Church and Gilbert. It uses 20 times less DNA than this latter technique to obtain a nucleotide-resolution banding pattern and allows short autoradiographic exposure times. The unique aspect of LMPCR is the blunt-end ligation of an asymmetric double-stranded linker (5' overhanging to avoid self-ligation or ligation in the wrong direction) onto the 5' end of each cleaved blunt-ended DNA molecule *(5,6)*. The blunt end is created by the extension of a gene-specific primer (primer 1 in **Fig. 5**) until a footprint strand break is reached. Because the generated breaks will be randomly distributed along the genomic DNA and thus have 5' ends of unknown sequence, the asymmetric linker adds a common and known sequence to all 5' ends. This then allows exponential PCR amplification from an adjacent genomic sequence to that of the generated breaks using the longer oligonucleotide of the linker (linker-primer) and a second nested gene-specific primer (primer 2, *see* **Fig. 5**). After 20-22 cycles of PCR, the DNA fragments are size-fractionated on a sequencing gel. LMPCR preserves the quantitative representation of each fragment in the original population of cleaved molecules *(10-13)*, allowing quantification on a phosphorimager *(14-17)*. Thus, the band intensity pattern obtained by LMPCR directly reflects the frequency distribution of 5'-phosphoryl DNA breaks along a 200-bp sequence adjacent to the nested primer.

Two methods exist to reveal the sequence and footprinting ladders created by LMPCR. Pfeifer and colleagues *(6)* took advantage of electroblotting DNA onto a nylon membrane followed by hybridization with a gene-specific probe to reveal sequence ladders, otherwise known as "indirect end labeling". On the other hand, Mueller and Wold *(5)* used a nested third radiolabeled primer for the last one or two cycles of the PCR amplification step. We find Pfeifer's method much more sensitive than Mueller and Wold's (unpublished data). In this chapter, we will describe our LMPCR protocol as modified from the protocol of Pfeifer and colleagues.

In summary, LMPCR is the method of choice to study the in vivo structure of promoters with respect to the positions of DNA-protein interactions, of special DNA structures and chromatin structures such as nucleosomes. To perform in vivo DNA analysis, three probing agents are regularly combined with LMPCR: DMS, ultraviolet (UV) and DNase I (**Figs. 1-4**, **Table 1**). These probing agents provide complementary information and each has its associated advantages and drawbacks (**Table 2**). To best characterize DNA-protein interactions, it is often necessary to use two or even all three of these methods. Treatments with any probing agents must produce either strand breaks or modified nucleotides that can be converted to DNA strand breaks with a 5'-phosphate in vitro (**Figs. 1-4**, **Table 3**). In this chapter, we describe protocols routinely used in our laboratory for DMS, UV, and DNase I in vivo treatments as well as the associated LMPCR technology. These protocols may also be adapted to footprinting with other probing agents, such as $KMnO_4$ and OsO_4 (*see* Chapters 6 and 9), although a detailed description is beyond the scope of the present chapter.

1.2 In Vivo Dimethylsulfate (DMS) Footprint Analysis (Fig. 1)

Dimethylsulfate is a small, highly reactive molecule that easily diffuses through the outer cell membrane and into the nucleus. It preferentially methylates not only the N7 position of

guanine residues via the major groove but, to a lesser extent, also the N3 position of adenine residues via the minor groove. The most significant technical advantage of in vivo DMS footprinting is that DMS can be simply added to the cell culture medium, requiring no cell manipulation (*see* **Table 2** for advantages and drawbacks). Each guanine residue of purified DNA displays about the same probability of being methylated by DMS. Because DNA inside living cells forms chromatin and is often found associated with a number of proteins, it is expected that its reactivity toward DMS will differ from purified DNA. **Figures 6** and **7** show in vivo DMS treatment patterns compared to the treatment of purified genomic DNA. Proteins in contact with DNA either decrease accessibility of specific guanines to DMS (protection) or, as frequently observed at the edges of a footprint, increase reactivity (hyperreactivity) *(1)*. Hyperreactivity can also indicate a greater DMS accessibility of special in vivo DNA structure *(19)*. Hot piperidine cleaves the glycosylic bond of methylated guanines and adenines, leaving a ligatable 5'-phosphate *(20)*.

Genomic footprinting using DMS reveals DNA-protein contacts located in the major groove of the DNA double helix (**Table 1**). However, it should be noted that in vivo DNA studies using DMS alone may not detect some DNA-protein interactions *(21)*. First, no DNA-protein interaction will be detected in the absence of guanine residues. Second, some proteins do not affect DNA accessibility to DMS. Third, certain weak DNA-protein contacts could actually be disrupted because of the high reactivity of the DMS. Thus when using DMS, it is often important to also apply alternative footprinting approaches *(21,22)*.

1.3 Photofootprint Analysis (Figs. 2 and 3)

Ultraviolet light (UVC: 200-280 nm; UVB: 280-320 nm) can also be used as a modifying agent for in vivo footprinting *(4,8,23-25)*. When cells are subjected to UV light (UVC or UVB), two major classes of lesions may be introduced into the DNA at dipyrimidine sequences (CT, TT, TC, and CC): the cyclobutane pyrimidine dimer (CPD) and the

pyrimidine (6-4) pyrimidone photoproduct (6-4PP) *(26)*. CPDs are formed between the 5,6 bonds of any two adjacent pyrimidines, whereas a stable bond between positions 6 and 4 of two adjacent pyrimidines characterizes 6-4PPs. 6-4PP are formed at a rate 15-30% of that of CPDs *(27)* and are largely converted to their Dewar valence isomers by direct secondary photolysis (photoisomerization) *(27)*. In living cells, the photoproduct distribution is determined both by sequence context and chromatin structure *(28)*. In general, CPDs and 6-4PPs appear to form preferentially in longer pyrimidine runs. Because UVB and UVC radiation are primarily absorbed in the cell by the DNA, there are relatively few perturbations of other cellular processes, and secondary events that could modify the chromatin structure or release DNA-protein interactions. Furthermore, intact cells are exposed for a short period of time only to a high-intensity UV irradiation. Thus, UV irradiation is probably one of the least disruptive footprinting method and, hence, truly reflects the in vivo situation (**Table 2**). As for DMS, DNA-binding proteins influence the distribution of UV photoproducts in a significant way *(23)*. When the photoproduct spectrum of irradiated purified DNA is compared with that obtained after irradiation of living cells, some striking differences become apparent. These are referred to as "photofootprints" *(23)*. The photoproduct frequency within sequences bound by sequence-specific DNA-binding proteins (transcription factors) is suppressed or enhanced in comparison to purified DNA *(4,8,29)*. Effects of chromatin structure may be significant in regulatory gene regions that bind transcription factors (**Fig. 6**). Mapping of CPDs at the single-copy gene level can reveal positioned nucleosomes because CPDs are modulated in a 10-bp periodicity within nucleosome core DNA *(30,31)*. 6-4PPs form more frequently in linker DNA than in core DNA *(32)*.

Photofootprints reveal variations in DNA structure associated with the presence of transcription factors or other proteins bound to the DNA. UV light has the potential to reveal all DNA-protein interactions provided there is a dipyrimidine sequence on either DNA strand within a putative protein-binding sequence. Because photofootprints can be seen outside protein-binding sites, UV light should not be used as the only in vivo footprinting agent. The precise delimitations of the DNA-protein contact are difficult to determine with the simple in vivo UV probing method.

The distribution of UV-induced CPDs and 6-4PPs along genomic DNA can be mapped at the sequence level by LMPCR following conversion of these photoproducts into ligatable 5'-phosphorylated single-strand breaks. CPD are enzymatically converted by cleavage with T_4 endonuclease V followed by UVA (320-400 nm) photoreactivation of the overhanging pyrimidine using photolyase (**Fig. 2**) *(8)*. Because the 6-4PPs and their Dewar isomers are hot alkali-labile sites, they can be cleaved by hot piperidine (**Fig. 3**) *(29)*.

1.4 In Vivo DNase I Footprint Analysis (Fig. 4)

DNase I treatment of permeabilized cells gives clear footprints when the DNase I-induced breaks are mapped by LMPCR *(2)*. Both living cells (in vivo) and purified DNA (in vitro) are treated with DNase I. As with DMS and UV, footprint analyses are obtained by comparing in vivo DNase I digestion patterns to patterns obtained from the digestion of purified genomic DNA (**Fig. 7**). When compared to purified DNA, permeabilized cells show protected bands at DNA-protein interaction sequences and DNase I hypersensitive bands in regions of higher-order nucleoprotein structure *(2)*. Compared to DMS, DNase I is less base selective, is more efficient at detecting minor groove DNA-protein contacts, provides more information on chromatin structure, displays larger and clearer footprints, and better delimits the boundaries of DNA-protein interactions (**Fig. 7**). The nucleotides covered by a protein are almost completely protected on both strands from DNase I nicking, allowing a better delimitation of the boundaries of DNA-protein contacts. However, it should be underlined that the relatively bulky DNase I molecule cannot cleave the DNA in the immediate vicinity of a bound protein because of steric hindrance. Consequently, the regions protected from cutting can extend beyond the actual DNA-protein contact site. On the other hand, when DNA is wrapped around a nucleosome-size particle, DNase I cutting activity is increased at 10-bp intervals and no footprint is observed (**Tables 1** and **2**).

DNase I, a relatively large 31-kDa protein, cannot penetrate cells without previous cell-membrane permeabilization. Cells can be efficiently permeabilized by lysolecithin *(2)* or Nonidet P40 *(33)*. It has been shown that cells permeabilized by lysolecithin remain intact, replicate their DNA very efficiently, and show normal transcriptional activities *(34,35)*. There are numerous studies showing that lysolecithin-permeabilized cells maintain a normal nuclear structure to a greater extent than isolated nuclei, because the chromatin structure can be significantly altered during the nuclear isolation procedures *(2)*. Indeed, DNase I footprinting studies using isolated nuclei can be flawed because transcription factors are lost during the isolation of nuclei in polyamine containing buffers *(2)*. Even though other buffers may be less disruptive, factors can still be lost during the isolation procedure, leading to the loss of footprints or partial loss of footprints.

DNase I digestion of DNA leaves ligatable 5'-phosphorylated breaks, but the 3'-ends are free hydroxyl groups. Pfeifer and colleagues *(2,36)* observed that these genomic 3'-OH ends can be used as primers and extended by the DNA polymerases during the primer extension and/or PCR steps of LMPCR, thereby reducing significantly the overall efficiency of LMPCR and giving a background smear on sequencing gels. To avoid the nonspecific priming of these 3'-OH ends, three alternative solutions have been applied: (1) blocking these ends by the addition of a dideoxynucleotide *(2,36)*; (2) enrichment of fragments of interest by extension product capture using biotinylated gene-specific primers and magnetic streptavidin-coated beads *(18,37-39)*; and (3) performing primer 1 hybridization and primer 1 extension at a higher temperature (52-60°C vs 48°C, and 75°C vs 48°C, respectively) using a thermostable enzyme such as *Vent* exo⁻ DNA polymerase and cloned *Pfu* DNA polymerase *(3,40-42)*. Although effective, the first two alternatives involve additional manipulations that are time-consuming. Because of its simplicity, we select primer 1 with higher T_m (52-60°C) and use the cloned *Pfu* DNA polymerase for the primer 1 extension.

1.5 Choice of DNA Polymerases for LMPCR

Ligation-mediated PCR involves the PCR amplification of a mixture of genomic DNA fragments of different size. During the LMPCR procedure, DNA polymerases are required for two steps: primer extension (PE) and PCR amplification. For the PE step, the best DNA polymerase would be one that (1) is thermostable and very efficient, (2) has no terminal transferase activity, (3) is able to efficiently polymerize about 0.75 kb of DNA even when the DNA is very GC rich, and (4) is able to polymerize through any DNA secondary structures. For the PCR step, the best DNA polymerase would be (1) thermostable, (2) very efficient, (3) able to amplify indiscriminately a mixture of DNA fragments of different lengths (between 50 and 750 bp) and of varying GC-richness (from 5 to 95%), and (4) able to efficiently resolve DNA secondary structures. We find cloned *Pfu* DNA polymerase that corresponds to *Pfu* exo⁻ is the best enzyme for the PE and PCR steps of LMPCR *(42)*. In this chapter, LMPCR protocols using cloned *Pfu* DNA polymerase for PE and PCR steps will be described in detail. However, because the more frequently used combination of DNA polymerases is Sequenase™ 2.0 for the PE step and *Taq* DNA polymerase for the PCR step, a description of an alternative LMPCR protocol using Sequenase 2.0 and *Taq* DNA polymerase will also be included.

2 Materials

2.1 DNA Purification (for 10^7 to 10^8 cells)

1. Any types of cells (i.e., fibroblasts, lymphocytes, etc.).

2. Trypsin-EDTA (Gibco-BRL).

3. Hank's Balanced Salt Solution (HBSS) (Gibco-BRL).

4. Buffer A: 300 mM sucrose, 60 mM KCl, 15 mM NaCl, 60 mM Tris-HCl, pH 8.0, 0.5 mM spermidine, 0.15 mM spermine, and 2 mM EDTA. Store at -20˚C.

5. Buffer A + 1% Nonidet P40. Store at -20˚C.

6. Conical tissue culture tubes, 50 mL.

7. Buffer B: 150 mM NaCl and 5 mM EDTA, pH 7.8.

8. Buffer C: 20 mM Tris-HCl pH 8.0, 20 mM NaCl, 20 mM EDTA, and 1% sodium dodecyl sulfate (SDS).

9. Proteinase K from *Tritirachium album* (Roche Molecular Biochemicals).

10. RNase A from bovine pancreas (Roche Molecular Biochemicals).

11. Phenol, equilibrated with 0.1M Tris-HCl, pH 8.0 (Roche Molecular Biochemicals, cat. no. 108-95-2).

12. Chloroform.

13. 5 M NaCl.

14. Precooled absolute ethanol (-20°C).

15. Precooled 70% ethanol (-20°C).

16. N-2-Hydroxyethylpiperazine-N'-2-ethanesulfonic acid (HEPES).

17. 4'-6-Diamidino-2-phenylindole (DAPI).

18. Nanopure H_2O should be used in making any buffers, solutions, and dilutions, unless otherwise specified.

2.2 Chemical Cleavage for DNA Sequencing Products

1. Potassium tetrachloropalladate(II) (K_2PdCl_4, Aldrich).

2. K_2PdCl_4 solution: 10 mM K_2PdCl_4 and 100 mM HCl, pH 2.0 (adjusted with NaOH). Store at -20˚C.

3. K_2PdCl_4 stop: 1.5 M sodium acetate, pH 7.0 and 1 M ß-mercaptoethanol.

4. Dimethylsulfate (DMS, 99+%, Fluka). Considering its toxic and carcinogenic nature, DMS should be manipulated in a well-ventilated hood. DMS is stored under nitrogen at 4˚C and should be replaced every 12 mo. DMS waste is detoxified in 5 M NaOH.

5. DMS buffer: 50 mM sodium cacodylate and 1 mM EDTA, pH 8.0. Store at 4°C.

6. DMS stop: 1.5 M sodium acetate, pH 7.0 and 1 M ß-mercaptoethanol. Store at -20°C.

7. Hydrazine (Hz, anhydrous, Aldrich). Considering its toxic and carcinogenic potentials, Hz should be manipulated in a well-ventilated hood. Hz is stored under nitrogen at 4˚C in an explosion-proof refrigerator and the bottle should be replaced at least every 6 mo. Hz waste is detoxified in 3 M ferric chloride.

8. Hz stop: 300 mM sodium acetate, pH 7.0 and 0.1 mM EDTA. Store at 4°C.

9. 5 M NaCl.

10. 3 M Sodium acetate, pH 7.0.

11. Precooled absolute ethanol (-20°C).

12. Precooled 80% ethanol (-20°C).

13. Dry ice.

14. Piperidine (99+%, Fluka or Sigma): 10 M stock diluted to 2 M with H_2O just before use by adding 250 μL stock under 1 mL H_2O in a 1.5-mL microtube on ice. Cap

immediately to minimize evaporation. Considering its toxic and carcinogenic potentials, piperidine should be manipulated in a well-ventilated hood. Piperidine 10 *M* is stored at 4°C under nitrogen atmosphere.

15. Teflon tape.

16. Lock caps.

17. 3 *M* Sodium acetate, pH 5.2.

18. 20 μg/μL glycogen.

19. Vacuum concentrator (SpeedVac concentrator, Savant).

2.3 Treatment of Purified DNA and Living Cells with Modifying Agents

2.3.1 DMS Treatment

1. DMS (99+%, Fluka).

2. Trypsin-EDTA (Gibco-BRL).

3. Hank's Balanced Salt Solution (HBSS).

2.3.2 254-nm UV and UVB Irradiation

1. Germicidal lamp (254 nm) for UVC irradiation (Philips G15 T8, TUV 15W).

2. UVB light for UVB irradiation (Philips, FS20T12/UVB/BP).

3. UVX digital radiometer (Ultraviolet Products, Upland, CA).

4. 0.9% NaCl.

5. UV irradiation buffer: 150 mM KCl, 10 mM NaCl, 10 mM Tris-HCl, pH 8.0, and 1 mM EDTA.

6. Buffer A + 0.5% Nonidet P40. Store at -20˚C.

7. Scraper.

2.3.3 DNase I Treatment

1. Deoxyribonuclease I (DNase I, Worthington biochemical corporation; 45A134).

2. Trypsin-EDTA (Gibco-BRL).

3. Hank's Balanced Salt Solution (HBSS).

4. L-α-Lysophosphatidylcholine (L-α-Lysolecithin).

5. Nonidet P40.

6. Solution I: 150 mM sucrose, 80 mM KCl, 35 mM HEPES, pH 7.4, 5 mM MgCl$_2$, and 0.5 mM CaCl$_2$.

7. Solution II: 150 mM sucrose, 80 mM KCl, 35 mM HEPES, pH 7.4, 5 mM MgCl$_2$, and 2 mM CaCl$_2$.

8. Conical tubes, 15 and 50 mL.

9. Buffer B: 150 mM NaCl and 5 mM EDTA, pH 7.8.

10. Buffer C: 20 mM Tris-HCl, pH 8.0, 20 mM NaCl, 20 mM EDTA, and 1% SDS.

11. Proteinase K from *Tritirachium album* (Roche Molecular Biochemicals).

12. RNase A from bovine pancreas (Roche Molecular Biochemicals).

13. Phenol (*see* **Subheading 2.1., item 11**).

14. Chloroform.

15. 5 *M* NaCl.

16. Precooled absolute ethanol (-20°C).

17. Precooled 80% ethanol (-20°C).

2.4 Conversion of Modified Bases to DNA Single-Strand Breaks

2.4.1 DMS-Induced Base Modifications

1. Piperidine (99+%, *see* **Subheading 2.2., item 14**).

2.4.2 UV-Induced Base Modifications

1. 10X dual buffer: 500 m*M* Tris-HCl, pH 7.6, 500 m*M* NaCl, and 10 m*M* EDTA.

2. 1 *M* 1,4-Dithiothreitol (DTT, Roche Molecular Biochemicals).

3. 5 mg/mL nuclease-free bovine serum albumine (BSA, Roche Molecular Biochemicals).

4. T_4 endonuclease V enzyme (Epicentre Technologies). The saturating amount of T_4 endonuclease V enzyme can be estimated by digesting UV-irradiated genomic DNA

with various enzyme quantities and separating the cleavage products on alkaline agarose gel *(43)*. The saturating amount of the enzyme is the next to the minimum quantity that produces the maximum cleavage frequency as evaluated on the alkaline agarose gel.

5. *E. coli* photolyase enzyme (Pharmingen). The saturating amount of photolyase can be estimated by photoreactivating UV-irradiated genomic DNA with various enzyme quantities, digestion with T_4 endonuclease V, and separating the cleavage products on alkaline agarose gel *(43)*. The saturating amount of photolyase is the next to the minimum enzyme quantity which produces no cleavage following T_4 endonuclease V digestion as evaluated on the gel. Because photolyase is light sensitive, all steps involving photolyase should be carried out under yellow light.

6. UVA black light (UV F15T8BLB 360 nm, Philips, 15W).

7. Plastic film (plastic wrap).

8. 0.52% SDS solution.

9. Phenol (*see* **Subheading 2.1., item 11**).

10. Chloroform.

11. 5 *M* NaCl.

12. Precooled absolute ethanol (-20°C).

13. Precooled 80% ethanol (-20°C).

14. Piperidine (99+%, *see* **Subheading 2.2., item 14**).

2.5 Ligation-Mediated Polymerase Chain Reaction Technology

2.5.1 Primer Extension (Steps II and III, Fig. 5)

1. A gene-specific primer (primer 1) is used to initiate primer extension. The primer 1 used in the first-strand synthesis are 15- to 22-mer oligonucleotides and have a calculated melting temperature (T_m) of 50-60°C. They are selected using a computer program (Oligo 4.0 software, National Biosciences) *(44)* and, optimally, their T_m, as calculated by a computer program (GeneJockey software), should be about 10°C lower than that of subsequent primers (*see* **Note 1**) *(45)*. The first-strand synthesis reaction is designed to require very little primer 1 with a lower T_m so that this primer does not interfere with subsequent steps *(11-13,46)*. The primer 1 concentration is set at 50 μM in H_2O and then diluted 1:100 in H_2O to give 0.5 pmol/μL.

2. Siliconized microtubes (0.625 μL) (National Scientific Supply Co, Inc.).

3. Thermocycler (PTC^{TM}, MJ research, Inc.).

4. 10X cloned *Pfu* buffer: 200 mM Tris-HCl, pH 8.8, 20 mM MgSO$_4$, 100 mM NaCl, 100 mM (NH$_4$)$_2$SO$_4$, 1% (v/v) Triton X-100, and 1 mg/mL nuclease-free BSA (*see* **Note 2**).

5. Cloned *Pfu* mix: 1.5 mM of each dNTP and 1.5 U cloned *Pfu* DNA polymerase, also named *Pfu* exo⁻ (2.5 U/μL, Stratagene).

6. 5X Sequenase buffer: 200 mM Tris-HCl, pH 7.7, and 250 mM NaCl.

7. Mg-dNTPs mix: 20 mM MgCl$_2$, 20 mM DTT, and 0.375 mM of each dNTP.

8. T7 Sequenase V.2 (Amersham).

9. 310 mM Tris-HCl, pH 7.7.

2.5.2 Ligation (Step IV, Fig. 5)

1. The DNA molecules that have a 5'-phosphate group and a double-stranded blunt end are suitable for ligation. A DNA linker with a single blunt end is ligated directionally onto the double-stranded blunt end of the extension product using T_4 DNA ligase. This linker has no 5' phosphate and is staggered to avoid self-ligation and provide directionality. Also, the duplex between the 25-mer (5' GCGGTGACCCGGGAGATCTGAATTC) and 11-mer (5' GAATTCAGATC) is stable at the ligation temperature, but denatures easily during subsequent PCR reactions *(5,46)*. The linker was prepared in aliquots of 500 µL by annealing in 250 m*M* Tris-HCl, pH 7.7, 20 pmol/µL each of the 25-mer and 11-mer, heating at 95°C for 3 min, transferring quickly at 70°C, and cooling gradually to 4°C over a period of 3 h. Linkers are stored at -20°C and thawed on ice before use. Linker: L25 (60 pmol/µL, 5'-GCGGTGACCCGGGAGATCTGAATTC), L11 (60 pmol/µL, 5'-GAATTCAGATC), 2 *M* Tris-HCl, pH 7.7, and 1 *M* MgCl$_2$.

2. T_4 DNA ligase (1 U/µL, Roche Molecular Biochemicals).

3. Ligation mix: 30 m*M* DTT, 1 m*M* ATP, 83.3 µg/mL of BSA, 100 pmol of linker, and 3.25 U/microtube of T_4 DNA ligase. If cloned *Pfu* DNA polymerase was used for primer extension (step III, **Fig. 5**), the ligation mix is prepared by adding per microtube: 1.35 µL of 1 *M* DTT, 0.5 µL of 100 m*M* ATP, 0.15 µL of 5 µg/µL BSA, 1.1 µL of Tris-HCl, pH 7.4, 5.0 µL of 20 pmol/µL linker, 3.25 µL of 1 U/µL T_4 ligase, and 33.65 µL of H$_2$O. If Sequenase was used for primer extension (step III, **Fig. 5**), the ligation mix is prepared by adding per microtube: 1.35 µL of 1 *M* DTT, 0.5 µL of 100 m*M* ATP, 0.75 µL of 5 µg/µL BSA, 5.0 µL of 20 pmol/µL linker, 3.25 µL of 1 U/µL T_4 ligase, and 34.15 µL of H$_2$O.

4. 7.5 *M* Ammonium acetate.

5. 0.5 *M* EDTA, pH 8.0.

6. 20 µg/µL glycogen.

7. Precooled absolute ethanol (-20°C).

8. Precooled 80% ethanol (-20°C).

2.5.3 Polymerase Chain Reaction (Steps V and VI, Fig. 5)

1. At this step, gene-specific fragments can be exponentially amplified because primer sites are available at each end of the target fragments (i.e., primer 2 on one end and the longer oligonucleotide of the linker on the other end). Primer 2 may or may not overlap with primer 1. The overlap, if present, should not be more than seven to eight bases *(11-13,46)*. Primer 2 is diluted in H_2O to give 50 pmol/µL.

2. 10X cloned *Pfu* buffer: 200 mM Tris-HCl, pH 8.8, 20 mM MgSO$_4$, 100 mM NaCl, 100 mM (NH$_4$)$_2$SO$_4$, 1% (v/v) Triton X-100, and 1 mg/mL nuclease-free BSA (*see* **Note 2**).

3. Cloned *Pfu* DNA polymerase, also named *Pfu* exo⁻ (2.5 U/µL, Stratagene).

4. Cloned *Pfu* DNA polymerase mix per microtube: 2X cloned *Pfu* buffer, 0.5 mM of each dNTP, 10 pmol of LP25 (Linker Primer), 10 pmol of primer 2, and 3.5 U of cloned *Pfu* DNA polymerase.

5. Mineral oil.

6. Cloned *Pfu* DNA polymerase stop: 1.56 M sodium acetate, pH 5.2 and 20 mM EDTA.

7. Formamide loading dye: 94% formamide, 2 mM EDTA, pH 7.7, 0.05% xylene cyanole FF, and 0.05% bromophenol blue *(11-13)*. The formamide loading dye is freshly premixed by adding 1 part H_2O to 2 parts formamide loading dye.

8. 5X *Taq* buffer: 50 mM Tris-HCl, pH 8.9, 200 mM NaCl, and 0.05% [w/v] gelatin (*see* **Note 2**).

9. *Taq* DNA polymerase (5 U/μL, Roche Molecular Biochemicals).

10. *Taq* DNA polymerase mix per microtube: 2X *Taq* buffer, 4 mM MgCl$_2$, 0.5 mM of each dNTP, 10 pmol LP25 (Linker Primer), 10 pmol primer 2, and 3 U *Taq* DNA polymerase.

11. *Taq* DNA polymerase stop: 1.56 M sodium acetate, pH 5.2 and 60 mM EDTA.

12. Phenol (*see* **Subheading 2.1., item 11**) premixed with chloroform in a ratio of 92 μL of phenol for 158 μL of chloroform.

13. Precooled absolute ethanol (-20°C).

14. Precooled 80% ethanol (-20°C).

2.5.4 Gel Electrophoresis and Electroblotting (Step VII, Fig. 5)

1. 60-cm-long x 34.5-cm-wide sequencing gel apparatus (Owl Scientific).

2. Spacers (0.4-mm thick).

3. Plastic well-forming comb (0.4-mm thick, BioRad).

4. 5X (0.5 M) Tris-Borate-EDTA (TBE) buffer: 500 mM Tris, 830 mM boric acid, and 10 mM EDTA, pH 8.3. Use this stock to prepare 1X (100 mM) TBE buffer.

5. 8% Polyacrylamide, to prepare 1 L, add 77.3 g of acrylamide, 2.7 g of *bis*-acrylamide, 420.42 g of urea, and 200 mL of 0.5 M TBE dissolved in H$_2$O. Polyacrylamide solution should be kept at 4°C.

6. Gel preparation: Mix 100 mL of 8% polyacrylamide with 1 mL of 10% ammonium persulfate (APS) and 30 μL of *N,N,N′,N′*-tetra-methylethylenediamide (TEMED). This mix is prepared immediately before pouring the solution between the glass plates. Without delay, take the gel mix into a 50-mL syringe and inject the mix between the

plates, maintaining a steady flow. During pouring, the plates should be kept at a 30˚ angle and tilted to the side into which the mix is injected. Any air bubbles should be avoided and removed if they form. The gel should be left to polymerize for a minimum of 3 h before use. If the gel is to be left overnight, 45 min after pouring, place a moistened paper tissue over the comb, and cover the upper end of the assembly with a plastic film to prevent the gel from drying out.

7. Flat gel loading tips (National Scientific Supply Co).

8. Power supply (Bio-Rad PowerPac 3000).

9. Electroblotting apparatus (HEP3, Owl Scientific Inc.) used according to the manufacturer's instructions.

10. Whatman 3MM Chr paper (Fisher Scientific).

11. Plastic film (plastic wrap).

12. Whatman 17 Chr papers (Fisher Scientific).

13. Nylon membrane, positively charged (Roche Molecular Biochemicals, cat. no. 1 417 240).

14. Power supply (Bio-Rad, model 200/2.0).

15. UVC (254 nm) germicidal lamp.

2.5.5 Hybridization (Step VII, Fig. 5)

The hybridization is performed in a rolling 8-cm-diameter x 22 cm long borosilicate glass hybridization tubes in a hybridization oven (Hoefer). The nylon membrane is soaked in 100 mM TBE and, using a 25-mL pipet, placed in the tube so that the membrane sticks completely to the wall of the hybridization tube. Following hybridization and washing, the

membranes are placed in an autoradiography cassette FBAC 1417 (Fisher Scientific) and exposed to Kodak X-ray film (XAR-5, 35 x 43 cm, Kodak Scientific Imaging Film) with intensifying screens (35 x 43 cm, Fisher Scientific, cat. no. FB-IS-1417) at -70°C when a radiolabeled probe has been hybridized and without intensifying screens at room temperature when a digoxigenin-labeled probe has been hybridized.

2.5.5.1 Radiolabeled Probe

1. Hybridization buffer: 250 mM sodium phosphate, pH 7.2, 1 mM EDTA, 7% SDS, and 1% BSA.

2. Radiolabeled probe diluted in 6-7 mL of hybridization buffer.

3. Washing buffer I: 20 mM sodium phosphate, pH 7.2, 1 mM EDTA, 0.25% BSA, and 2.5% SDS.

4. Washing buffer II: 20 mM sodium phosphate, pH 7.2, 1 mM EDTA, and 1% SDS.

5. Plastic film (plastic wrap).

2.5.5.2 Digoxigenin-Labeled Probe

1. Pre-hybridization buffer: 5X SSC (750 mM NaCl and 75 mM sodium citrate, pH 7.0), 1% casein, 0.1% N-lauroylsarcosin, and 0.02% SDS.

2. Digoxigenin-labeled probe diluted in 15 mL of pre-hybridization buffer (use only 7.5 mL for hybridization).

3. 2X washing solution: 2X SSC and 0.1% SDS.

4. 0.5X washing solution: 0.5X SSC and 0.1% SDS.

5. Buffer 1: 150 mM NaCl and 100 mM maleic acid, pH 7.5.

6. Buffer 2: buffer 1 + 1% (w/v) casein.

7. Antidigoxigenin antibodies (Roche Molecular Biochemicals).

8. Buffer 1 + 0.3% Tween-20.

9. Buffer 3: 100 mM Tris-HCl, pH 9.5, 100 mM NaCl, and 50 mM MgCl$_2$.

10. CSPD$^®$ [Disodium 3-(4-methoxyspiro {1,2-dioxetane-3,2'-(5'-chloro)tricyclo[3.3.1.1^{3,7}]decan}-4-yl)phenyl phosphate] substrate (Roche Molecular Biochemicals, cat. no. 1 655 884).

11. Acetate sheets.

12. Doubleseal (Model 855, Decosonic).

2.6 Preparation of Single-Stranded Hybridization Probes (Step VIII, Fig. 5)

2.6.1 Template Preparation: PCR Products

2.6.1.1 PCR Amplification

1. 5X *Taq* buffer: 50 mM Tris-HCl, pH 8.9, 200 mM NaCl, and 0.05% (w/v) gelatin (*see* **Note 2**).

2. *Taq* DNA polymerase (5 U/µL, Roche Molecular Biochemicals).

3. One primer 2 (50 pmol/µL) for each strand of the DNA fragment to be amplified distant from 150 to 450 bp.

4. *Taq* DNA polymerase mix per microtube: 2X *Taq* buffer, 4 mM MgCl$_2$, 0.4 mM of each dNTP, 10 pmol of each primer 2, and 3 U *Taq* DNA polymerase.

5. Mineral oil.

6. *Taq* DNA polymerase stop: 1.56 M sodium acetate, pH 5.2, and 60 mM EDTA.

7. Phenol (*see* **Subheading 2.1., item 11**) premixed with chloroform in a ratio of 92 µL of phenol to 158 µL of chloroform.

8. Precooled absolute ethanol (-20°C).

9. Precooled 80% ethanol (-20°C).

10. 5X TAE loading buffer: 5X TAE (200 mM Tris base, 100 mM glacial acetic acid, and 5 mM EDTA, pH 8.0), 0.025% bromophenol blue, 30% Ficoll 400, and 2% SDS.

2.6.1.2 Purification and Quantification of PCR Products

1. Agarose.

2. 1X TAE buffer: 40 mM Tris base, 20 mM glacial acetic acid, and 1 mM EDTA, pH 8.0.

3. DNA size standards (øX 174 RF, Canadian life technologies, cat. no. 15611-015).

4. Ethidium bromide.

5. Siliconized microtubes (0.625 mL) and 1.5-mL microtubes.

6. Glass wool.

7. 3 M sodium acetate, pH 7.0.

8. Precooled absolute ethanol (-20°C).

9. Precooled 80% ethanol (-20°C).

10. Low DNA mass ladder (Gibco BRL, cat. no. 10068-013).

11. 5X universal neutral loading buffer: 0.25% bromophenol blue, 0.25% xylene cyanol FF, and 30% glycerol in H_2O. Store at 4˚C.

2.6.2 Labeling of Single-Strand Hybridization Probes

2.6.2.1 Isotopic Labeling

1. Siliconized microtubes (0.625 mL) and 1.5-mL microtubes.

2. 5X *Taq* buffer: 50 m*M* Tris-HCl, pH 8.9, 200 m*M* NaCl, and 0.05% (w/v) gelatin (*see* **Note 2**).

3. 100 m*M* $MgCl_2$.

4. DNA templates: PCR products (10 ng/µL) or DNA plasmids (20 ng/µL).

5. Primer 2 (50 pmol/µL).

6. dNTP (dATP, dGTP, dTTP) mix (200 µ*M* of each).

7. dNTP (dATP, dGTP, dTTP) mix diluted 1:10 in H_2O. This mix is changed every 2 wk.

8. *Taq* DNA polymerase (5 U/µL, Roche Molecular Biochemicals).

9. α-$[^{32}P]$dCTP (3000 Ci/mmol, New England Nuclear).

10. 7.5 *M* ammonium acetate.

11. 20 µg/µL glycogen

12. Precooled absolute ethanol (-20°C)

13. Geiger counter.

14. TE buffer pH 8.0: 10 mM Tris-HCl, pH 8.0 and 1 mM EDTA, pH 7.8.

15. Hybridization buffer: 250 mM sodium phosphate, pH 7.2, 1 mM EDTA, 7% SDS, and 1% BSA.

2.6.2.2 Digoxigenin (Nonisotopic) Labeling

1. Siliconized microtubes (0.625 mL) and 1.5-mL microtubes.

2. 5X *Taq* buffer: 50 mM Tris-HCl, pH 8.9, 200 mM NaCl, and 0.05% (w/v) gelatin (*see* **Note 2**).

3. 100 mM MgCl$_2$.

4. DNA templates: PCR products (10 ng/µL) or DNA plasmids (20 ng/µL).

5. Primer 2 (50 pmol/µL).

6. dNTP mix (A:G:C:T = 25 mM : 25 mM : 25 mM : 20 mM).

7. dNTP mix diluted 1:8.3 in H$_2$O. This mix is changed every 2 wk.

8. 1 mM digoxigenin-11-dUTP (Roche Molecular Biochemicals) diluted 1:2 in H$_2$O.

9. *Taq* DNA polymerase (5 U/µL, Roche Molecular Biochemicals).

10. 7.5 M ammonium acetate.

11. 20 µg/µL glycogen.

12. Precooled absolute ethanol (-20°C).

13. TE buffer pH 8.0: 10 mM Tris-HCl, pH 8.0 and 1 mM EDTA, pH 7.8.

14. Pre-hybridization buffer: 5X SSC (750 mM NaCl and 75 mM sodium citrate, pH 7.0), 1% casein, 0.1% N-lauroylsarcosine, and 0.02% SDS.

3 Methods

3.1 DNA Purification (for 10^7 to 10^8 Cells)

1. Detach cells using trypsin (if needed) and sediment the cell suspension by centrifugation in 50-mL conical tubes.

2. Resuspend the cells in 5-15 mL of buffer A.

3. Add 1 volume (5-15 mL) of buffer A containing 1% Nonidet P40.

4. Incubate on ice for 5 min.

5. Sediment nuclei by centrifugation at 4500g for 10 min at 4˚C.

6. Remove the supernatant. Resuspend nuclei in 1-10 mL of buffer A by gentle vortexing. Resediment nuclei at 4500g for 10 min at 4˚C.

7. Remove supernatant. It is recommended to leave a small volume (100-500 µL) of buffer A to facilitate resuspension of nuclei.

8. Dilute the nuclei in 1-2 mL of buffer B.

9. Add an equivalent volume of buffer C and proteinase K to a final concentration of 450 µg/mL.

10. Incubate at 37°C for 3 h, shake occasionally (*see* **Note 3**).

11. Add RNase A to a final concentration of 150 µg/mL.

12. Incubate at 37°C for 1 h.

13. Purify DNA by extraction with 1 volume phenol (one to two times as needed), 1 vol phenol:chloroform (one to two times as needed), and 1 vol chloroform. Phenol extraction and phenol-chloroform extraction should be repeated if the aqueous phase is not clear (*see* **Note 3**).

14. Precipitate DNA in 200 mM NaCl and 2 vol of precooled absolute ethanol. Ethanol should be added slowly and to facilitate DNA recovery, rock the tube very gently.

15. Recover DNA by spooling the floating DNA filament with a micropipet tip. If DNA is in small pieces or not clearly visible, recover DNA by centrifugation (5000g for 30 min at 4°C), but expect RNA contamination (*see* **Note 4**). RNA contamination does not cause any problems for LMPCR. RNase digestion can be repeated if needed.

16. Remove supernatant and wash DNA once with 10 mL of 70% ethanol.

17. Centrifuge the DNA (5000g for 30 min at 4°C).

18. Remove supernatant and air-dry DNA pellet.

19. Dissolve DNA in 10 mM HEPES, pH 7.4, and 1 mM EDTA (HE buffer) at an estimated concentration of 60-100 µg/mL. The quantity of DNA can be estimated based upon the number of cells that were initially used for DNA purification. About 6 µg of DNA should be purified from 1 x 10^6 cells.

20. Carefully measure DNA concentration by spectrophotometry at 260 nm. Alternatively, DNA can be measure by fluorometry after staining with DAPI. Only double-strand DNA concentration has to be measured, be careful if there is RNA contamination (*see* **Note 5**).

3.2 Chemical Cleavage for DNA Sequencing Products

In vivo DNA analysis using LMPCR requires complete DNA sequencing ladders from genomic DNA. Base-specific chemical modifications are performed according to Iverson and Dervan *(47)* for the A reaction and Maxam and Gilbert for the G, T+C and C reactions. DNA from each of these base modification reactions is processed by LMPCR concomitantly with the analyzed samples and loaded in adjacent lanes on the sequencing gel to allow the identification of the precise location and sequence context of footprinted regions. The chemical modifications induced by DMS, Hz, and K_2PdCl_4 and cleaved by piperidine destroy the target base. Therefore, one must bear in mind that when analyzing a chemical-sequencing ladder, each band corresponds to a DNA fragment ending at the base preceding the one read. In this section, we will describe the chemical sequencing of genomic DNA. The cleavage protocol below works optimally with 10-50 μg of genomic DNA per microtube. Before chemical sequencing, the required amount of DNA per microtube is ethanol precipitated and the pellet is air-dried. For each base-specific reaction, we usually carried out the treatment in three microtubes containing 50 μg of genomic DNA for three different incubation times with the modifying agent in order to obtain low, medium and high base-modification frequencies.

3.2.1 A Reaction

1. Add 160 μL of H_2O and 40 μL of K_2PdCl_4 solution to the DNA pellet and carefully mix on ice using a micropipet.

2. Incubate at 20°C for 5, 10, or 15 min.

3. Add 50 μL of K_2PdCl_4 stop.

4. Add 750 μL of precooled absolute ethanol.

3.2.2 G Reaction

1. Add 5 μL of H_2O, 200 μL of DMS buffer, and 1 μL of DMS to the DNA pellet and carefully mix on ice using a micropipet.

2. Incubate at 20˚C for 30, 45, or 60 s.

3. Add 50 μL of DMS stop.

4. Add 750 μL of precooled absolute ethanol.

3.2.3 T+C Reaction

1. Add 20 μL of H_2O and 30 μL of Hz to the DNA pellet and carefully mix on ice using a micropipet.

2. Incubate at 20˚C for 120, 210, or 300 s.

3. Add 200 μL of Hz stop.

4. Add 750 μL of precooled absolute ethanol.

3.2.4 C Reaction

1. Add 5 μL of H_2O, 15 μL of 5 *M* NaCl, and 30 μL of Hz to the DNA pellet and carefully mix on ice using a micropipet.

2. Incubate at 20˚C for 120, 210, or 300 s.

3. Add 200 μL of Hz stop.

4. Add 750 µL of precooled absolute ethanol.

All samples are processed as follows:

1. Mix samples well and place on dry ice for 15 min.

2. Centrifuge for 15 min at 15,000g at 4˚C.

3. Remove supernatant, then recentrifuge for 1 min and remove all the liquid using a micropipet.

4. Carefully dissolve pellet in 405 µL of H_2O.

5. Add 45 µL of 3 M sodium acetate, pH 7.0.

6. Add 1 mL of precooled absolute ethanol.

7. Leave on dry ice for 15 min.

8. Centrifuge for 10 min at 15,000g at 4˚C.

9. Take out supernatant and then respin.

10. Wash with 1 mL of precooled 80% ethanol; spin 5 min at 15,000g in a centrifuge at 4˚C.

11. Remove the supernatant, spin quickly, remove the liquid with a micropipet and air-dry pellet.

12. Dissolve pellet in 50 µL of H_2O, add 50 µL of freshly prepared 2 M piperidine, and mix well using a micropipet.

13. Secure caps with Teflon™ tapes and lock the caps with "lock caps".

14. Incubate at 82˚C for 30 min.

15. Pool all three microtubes of the same chemical reaction in a new 1.5-mL microtube.

16. Add H_2O until a volume of 405 µL is reached, then add 10 µL of 3 *M* sodium acetate, pH 5.2, 1 µL of glycogen, and 1 mL of precooled absolute ethanol.

17. Leave on dry ice for 15 min.

18. Spin 10 min at 15,000*g* at 4˚C.

19. Take out the supernatant and wash twice with 1 mL of precooled 80% ethanol, then respin for 1 min and remove all the liquid using a micropipet.

20. Add 200 µL of H_2O and remove traces of remaining piperidine by drying the sample overnight in a Speedvac concentrator.

21. Dissolve DNA in H_2O to a concentration of 0.5 µg/µL.

22. Determine the DNA strand break frequency by running the samples on a 1.5% alkaline agarose gel *(43)*. The size range of the fragments should span 100-500 bp.

3.3 Treatment of Purified DNA and Cells with Modifying Agents

3.3.1 DMS Treatment

1. If cells are grown to confluence as monolayer, replace the cell culture medium with a freshly prepared serum-free medium containing 0.2% DMS and incubate at room temperature for 6 min. If cells are grown in suspension, sediment the cells by centrifugation and remove the cell culture medium. The cells are diluted in a freshly prepared serum-free medium containing 0.2% DMS and are then incubated at room temperature for 6 min.

2. Remove the DMS-containing medium and quickly wash the cell monolayer with 10-20 mL of cold HBSS. Sediment cells by centrifugation if they are treated in suspension and remove the DMS-containing medium and wash the cells with 10 mL of cold HBSS.

3. Detach cells using trypsin for cells grown as monolayer.

4. Nuclei are isolated and DNA purified as described in **Subheading 3.1.**

5. Purified DNA obtained from the same cell type is treated as described in **Subheading 3.2.2.** Usually, a DMS treatment of 45 s should give a break frequency corresponding to that of the in vivo treatment described in this section. This DNA is the in vitro treated DNA used to compare with DNA DMS-modified in vivo (*see* **Notes 5** and **6**).

3.3.2 254-nm UV and UVB Irradiation

1. If cells are grown as monolayer in Petri dishes, replace cell culture medium with cold 0.9% NaCl. If cells are grown in suspension, sediment the cells by centrifugation and remove the cell culture medium. The cells are diluted in cold 0.9% NaCl at a concentration of 1 x 10^6 cells/mL (*see* **Note 7**) and, to avoid cellular shielding, a thin layer of the cell suspension is placed in 150-mm Petri dishes.

2. Expose the cells to 0.5-2 kJ/m^2 of UVC (254-nm UV) or 25-100 kJ/m^2 of UVB. The cells should be exposed on ice with uncovered Petri dishes. The UV intensity is measured using a UVX digital radiometer.

3. Remove the 0.9% NaCl by aspiration for cells grown as monolayer in Petri dishes or by sedimentation for cell suspensions.

4. If cells were irradiated in suspension; follow the procedure described in **Subheading 3.1.** to isolate nuclei and purify DNA. After DNA purification, DNA is dissolved in H$_2$O at a concentration of 0.2 μg/μL. For cells cultured in Petri dishes, add in each dish 8 mL of buffer A containing 0.5% Nonidet P40.

5. Incubate on ice for 5 min.

6. Scrape the cells and transfer them in a conical 50-mL tube. In the same conical 50-mL tube, pool cells from the Petri dishes that undergo the same procedure.

7. Wash the dishes twice with 8 mL of buffer A + 0.5% Nonidet P40 per each of three identical Petri dishes.

8. Continue from **step 5** of **Subheading 3.1.** After DNA purification, DNA is dissolved in H_2O at a concentration of 0.2 µg/µL.

9. Expose purified DNA to the same UVC or UVB dose as the cells. Purified DNA should be irradiated on ice and diluted in the UV irradiation buffer at a concentration of 60-75 µg/mL (*see* **Note 6**). Purified DNA should be obtained from the same type of cells as the type irradiated in vivo (*see* **Note 8**). This DNA is used as control DNA to compare with DNA UV modified in vivo (*see* **Notes 6** and **7**).

10. Following UV irradiation, DNA is ethanol precipitated and DNA is resuspended in H_2O at a concentration of 0.2 µg/µL.

3.3.3 DNase I Treatment

Genomic footprinting with DNase I requires cell permeabilization (*see* **Note 9**). Cells grown as a monolayer can be permeabilized while they are still attached to the Petri dish or in suspension following trypsinization. Here, we will describe cell permeabilization using lysolecithin applied to monolayer cell cultures (steps labeled **a**). For monolayer cultures, cells are grown to about 80% of confluency. Alternatively, we describe cell permeabilization using lysolecithin or Nonidet P40 applied to cells in suspension (steps labeled **b**). For cells in suspension, cells are diluted at a concentration of approximately 1 x 10^6 cells/mL. To permeabilize the vast majority of cells in suspension, they must not be clumped and not form aggregates during the permeabilization step and subsequent DNase I

treatment. To achieve this, we gently flick the microtubes during permeabilization and DNase I treatment and keep the cell concentration below 2 x $10^6/\mu L$.

1a. For cells in monolayers, permeabilize the cells by treating them with 4 mL of 0.05% lysolecithin in solution I (prewarmed) at 37˚C for 1-2 min *(48)*.

2a. Remove the lysolecithin and wash with 10 mL of solution I. Add 3 mL of DNase I (2-4 U/mL) to solution II and incubate at room temperature for 3-5 min. DNase I concentration and incubation times may have to be adjusted for different cell types. During this incubation, no more than 10% of the cells should be released from the dish.

3a. Stop the reaction and lyse the cells by removal of the DNase I solution and addition of 1.5 mL of buffer C containing 600 µg/mL of proteinase K. Add 1.5 mL of buffer B and mix gently by rocking the flask or the Petri dish. Transfer the lysis solution to a 15-mL tube (then continue to **step 4**).

Alternatively:

1b. Sediment the cell suspension by centrifugation. Wash the cells with HBSS. Resuspend the cells in solution II at a concentration of 20 x 10^6/mL and aliquot by transferring 100 µL of the cell suspension per 1.5-mL microtube. Add to each microtube 100 µL of solution II prewarmed at 37˚C containing 0.1% lysolecithin or 0.25% Nonidet P40. Mix gently by flicking. Incubate at room temperature for 3 min.

2b. Quickly spin to pellet the cells. Add 50 µL of 2000 U/mL DNase I and mix gently by flicking. Incubate at room temperature for 5 min.

3b. Quickly spin and remove supernatant, resuspend the cells in 1.5 mL of buffer B, and, using a pipet, rapidly transfer to a 15-mL tube in which there are already 1.5 mL of buffer C containing 600 µg/mL of proteinase K (then continue to **step 4**).

4. Incubate at 37°C for 3 h, shake occasionally.

5. Add RNase A to a final concentration of 200 µg/mL and incubate at 37°C for 1 h.

6. Purify DNA by phenol-chloroform extraction (*see* **Subheading 3.1., step 13**).

7. Precipitate DNA in 200 m*M* NaCl and 2 volumes of precooled absolute ethanol.

8. Leave on dry ice for 20 min.

9. Recover DNA by centrifugation (5000*g* for 30 min at 4°C), but expect RNA contamination. RNA contamination does not cause any problems for LMPCR. RNase A digestion can be repeated if needed.

10. Remove supernatant and wash DNA once with 10 mL of precooled 80% ethanol.

11. Centrifuge the DNA (5000*g* for 10 min at 4°C). Remove supernatant and air-dry DNA pellet.

12. Dissolve DNA in H$_2$O and carefully measure DNA concentration (*see* **Subheading 3.1., step 20**).

13. To obtain purified DNA controls (*see* **Notes 6** and **8**), digest 50 µg of purified DNA in solution II with 4-8 U/µL of DNase I at room temperature for 10 min. Stop the reaction by adding 400 µL of phenol. Extract once with phenol-chloroform and once with chloroform. Dissolve DNA in H$_2$O at a concentration of 0.5 µg/µL.

3.4 Conversion of Modified Bases to DNA Single-Strand Breaks

When purified DNA or living cells are treated with DMS and UV, DNA base modifications are induced (**Table 3**). These modifications must be converted to single-strand breaks before running LMPCR. Following UV exposure, CPDs and 6-4PPs are converted individually because they use different conversion procedures (**Table 3**). On the other hand, DNase I digestion generates DNA strand breaks suitable for LMPCR without any conversion procedures. Before running LMPCR, the DNA strand break frequency must be

determined by running the samples on a 1.5% alkaline agarose gel *(43)*. The size range of the fragments should span 200-2000 bp (*see* **Note 6**).

3.4.1 DMS-Induced Base Modifications (see Fig. 1)

1. Dissolve DNA (10-50 μg) in 50 μL H_2O, add 50 μL of 2 *M* piperidine and mix well using a micropipet.

2. Samples are processed as described in **Subheading 3.2., steps 13-20.**

3. Dissolve DNA in H_2O to a concentration of 0.2 μg/μL.

3.4.2 UV-Induced Base Modifications

3.4.2.1 CPD (*see* **Fig. 2**)

1. To specifically cleave CPDs, dissolve 10 μg of UV-irradiated DNA in 50 μL H_2O, add 50 μL of a solution containing 10 μL of 10X dual buffer, 0.1 μL of 1 *M* DTT, 2 μL of 5 mg/mL BSA, a saturating amount of T_4 endonuclease V, and complete with H_2O to a final volume of 50 μL. Mix well by flicking the microtube and quick spin.

2. Incubate at 37°C for 1 h.

3. To perform the photolyase digestion to remove the overhanging dimerized base that would otherwise prevent ligation *(8)*, add 10 μL of the following mix: 1 μL of 10X dual buffer, 1 μL of 1 *M* DTT, 0.2 μL of 5 mg/mL BSA, a saturating amount of photolyase, and complete with H_2O to a final volume of 10 μL. Mix well by flicking the microtube and quick spin.

4. Preincubate the microtubes at room temperature for 3-5 min under yellow light with their caps opened.

5. Leaving their caps opened, cover the microtubes with a plastic film to prevent UVB-induced damage and place open ends 2-3 cm from a UVA black light for 1 h.

6. Add 290 µL of 0.52% SDS, mix well, and extract DNA using 1 vol (400 µL) phenol, 1 vol phenol:chloroform, and 1 vol chloroform.

7. To precipitate DNA, add 18 µL of 5 M NaCl and 1 mL of precooled absolute ethanol.

8. Leave 15 min on dry ice, spin 20 min at 15,000g in a centrifuge at 4˚C.

9. Wash once with 1 mL of precooled 80% ethanol.

10. Spin 8 min at 15,000g in a centrifuge at 4˚C.

11. Air-dry the pellet and dissolve DNA in H_2O to a concentration of 0.2 µg/µL.

3.4.2.2 6-4PP (*see* Fig. 3)

1. Dissolve DNA (10-50 µg) in 50 µL of H_2O, add 50 µL of 2 M piperidine and mix well using a micropipet.

2. Samples are processed as described in **Subheading 3.2., steps 13-20**.

3. Dissolve DNA in H_2O to a concentration of 0.2 µg/µL.

3.5. Ligation-Mediated Polymerase Chain Reaction Technology

The LMPCR protocol using cloned *Pfu* DNA polymerase for primer extension and PCR steps is labeled with **a** in **Subheadings 3.5.1.** and **3.5.3.** An alternative LMPCR protocol using Sequenase for primer extension steps and *Taq* DNA polymerase for PCR steps is

labeled with **b** in **Subheadings 3.5.1.** and **3.5.3.** Aside from the ligation mix (*see* **Subheading 2.5.2.**), the ligation step (**Subheading 3.5.2.**) is identical with both enzyme combinations. The primer extension, ligation, and PCR steps are carried out in siliconized 0.625-mL microtubes and a thermocycler is used for all incubations.

3.5.1 Primer Extension (Steps II and III, Fig. 5)

1a. Mix 0.5-2 µg of genomic DNA, 3 µL of cloned *Pfu* buffer, and 1 pmol of primer 1 in a final volume of 25 µL.

2a. Denature DNA at 98°C for 3 min.

3a. Incubate for 20 min at 45°C to 55°C, depending of the T_m of the primer 1.

4a. Cool to 4°C.

5a. Add 5 µL of the cloned *Pfu* mix. Flick and quick spin.

6a. Incubate the samples at the annealing temperature for 30 s, then increase the temperature to 75°C at a rate of 0.3°C/s and incubate at 75°C for 10 min. Finally, the samples are cooled to 4°C.

Alternatively:

1b. Mix 0.5-1.6 µg of DNA in Sequenase buffer with 1 pmol of primer 1 in a final volume of 15-18 µL.

2b. Denature DNA at 98°C for 3 min.

3b. Incubate for 20 min at 45°C to 50°C, depending of the T_m of the primer 1.

4b. Cool to 4°C.

5b. Add 9 μL of the following mix: 7.5 μL of Mg-dNTP mix, 1.1 μL of H_2O, and 0.4 μL of T7 Sequenase V.2. Flick and quick spin.

6b. Incubate at 48°C for 5 min, 50°C for 1 min, 51°C for 1 min, 52°C for 1 min, 54°C for 1 min, 56°C for 1 min, 58°C for 1 min, and 60°C for 1 min. Then, the samples are cooled to 4°C.

7b. Add 6 μL of cold 310 mM Tris-HCl, pH 7.7.

8b. Incubate at 67°C for 15 min to inactivate the Sequenase, then cool to 4°C.

3.5.2 Ligation (Step IV, Fig. 5)

1. To the primer extension reaction, add 45 μL of the ligation mix and mix well with the pipet. Note that the composition of the ligation mix (*see* **Subheading 2.5.2.**) is different whether Sequenase or cloned *Pfu* DNA polymerase was used for the primer extension (Section 3.5.1 and Step III in **Fig. 5**).

2. Incubate at 18°C overnight.

3. On ice, precipitate DNA by adding 28.75 μL of 7.5 M ammonium acetate, 0.25 μL of 0.5 M EDTA, pH 8.0, 1 μL of 20 μg/μL glycogen, and 275 μL of precooled absolute ethanol.

4. Leave 15 min on dry ice, and spin 20 min at 15,000g in a centrifuge at 4˚C.

5. Wash once with 500 μL of precooled 80% ethanol.

6. Spin 8 min at 15,000g in a centrifuge at 4˚C.

7. Air-dry DNA pellets and dissolve DNA pellets in 50 μL of H_2O.

3.5.3 Polymerase Chain Reaction (Steps V and VI, Fig. 5)

1a. Add 50 µL of the cloned *Pfu* DNA polymerase mix and mix with the pipet. The reaction mix is overlaid with 50 µL of mineral oil.

2a. Cycle 22 times as described in **Table 4** for cloned *Pfu* DNA polymerase. The last extension should be done for 10 min to fully extend all DNA fragments.

3a. Add 25 µL of cloned *Pfu* DNA polymerase stop under the mineral oil layer. Then, continue to **step 4**.

Alternatively:

1b. Add 50 µL of the *Taq* DNA polymerase mix and mix with the pipet. The reaction mix is overlaid with 50 µL of mineral oil.

2b. Cycle 22 times as described in **Table 4** for *Taq* DNA polymerase. The last extension should be done for 10 min to fully extend all DNA fragments.

3b. Add 25 µL of *Taq* DNA polymerase stop mix under the mineral oil layer. Then, continue to **step 4**.

4. Extract with 250 µL of premixed phenol-chloroform (92 µL:158 µL) and transfer to 1.5-mL microtubes.

5. Add 400 µL of precooled absolute ethanol.

6. Leave 15 min on dry ice; spin 20 min at 15,000g in a centrifuge at 4˚C.

7. Wash once with 500 µL of precooled 80% ethanol.

8. Spin 8 min at 15,000g in a centrifuge at 4˚C.

9. Air-dry DNA pellets.

10. Dissolve DNA pellets in 7.5 µL of premixed formamide loading dye in preparation for sequencing gel electrophoresis. For the sequence samples G, A, T+C, and C, it is often advisable to dissolve DNA pellets in 15 µL of premixed formamide loading dye.

3.5.4 Gel Electrophoresis and Electroblotting (Step VII, Fig. 5)

The PCR-amplified fragments are separated by electrophoresis through a 8% polyacrylamide/7 M urea gel, 0.4 mm thick and 60-65 cm long, then transferred to a nylon membrane by electroblotting *(11-13)*.

1. Prerun the 8% polyacrylamide gel until the temperature of the gel reaches 50˚C. Running buffer is 100 mM TBE. Before loading the samples, wash the wells thoroughly using a syringe.

2. To denature DNA, heat the samples at 95°C for 2-3 min, then keep them on ice prior to loading.

3. Load an aliquot of 3-3.5 µL using flat tips.

4. Run the gel at the voltage or power necessary to maintain the temperature of the gel at 50°C. This will ensure that the DNA remains denatured.

5. Stop the gel when the green dye (xylene cyanole FF) reaches 1-2 cm from the bottom of the gel.

6. Separate the glass plates using a spatula, then remove one of the plates by lifting it carefully. The gel should stick to the less treated plate (*see* **Note 10**).

7. Cover the lower part of the gel (approx 40-42 cm) with a clean Whatman 3MM Chr paper, carefully remove the gel from the glass plate and cover it with a plastic film (*see* **Note 10**).

8. On the bottom plate of the electroblotter, individually layer three sheets of Whatman 17 CHR paper, 43 cm x 19 cm, presoaked in 100 mM TBE, and squeeze out the air bubbles between the paper layers by rolling with a bottle or pipet.

9. Add 150 mL of 100 mM TBE on the top layer and place the gel quickly on the Whatman 17 CHR papers before TBE is absorbed. Before removing the plastic film, remove all air bubbles under the gel by gently rolling a 10-mL pipet.

10. Remove the plastic film and cover the gel with a positively charged nylon membrane presoaked in 100 mM TBE, remove all air bubbles by gently rolling a 10-mL pipet, then cover with three layers of presoaked Whatman 17 CHR paper and squeeze out air bubbles with rolling bottle. Paper sheets can be reused several times except for those immediately under and above the gel.

11. Place the upper electrode onto the paper.

12. Electrotransfer for 45 min at 2 A. The voltage should settle at approximately 10-15 V.

13. UV-crosslink (1000 J/m^2 of UVC) the blotted DNA to the membrane, taking care to expose the DNA side of the membrane. If probe stripping and rehybridization are planned, keep the membrane damp.

3.5.5 Hybridization (Step VII, Fig. 5)

3.5.5.1 Radiolabeled Probe

1. Prehybridize with 15 mL of hybridization buffer at 60-68°C for 20 min. The prehybridization temperature is based on the T_m of the primer used to prepare the probe.

2. Decant the prehybridization buffer and add the labeled probe in 6-8 mL of hybridization buffer.

3. Hybridize at 60-68°C (2°C below the calculated T_m of the probe) overnight.

4. Wash the membrane with prewarmed washing buffers. The buffers should be kept in an incubator or water bath set at a temperature of 4°C higher than the hybridization temperature. The membrane is placed into a tray on an orbital shaker. Wash with buffer I for 10 min and with buffer II three times for about 10 min each time.

5. Wrap the membrane in plastic film (*see* **Note 10**). Do not let the membrane become dry if stripping and rehybridization are planned after exposure to the film.

6. Expose membrane to X-ray films with intensifying screens at -70°C. Although longer exposure might be necessary, an exposure of 0.5-8 h is usually enough to produce a sharp autoradiogram. Nylon membranes can be rehybridized if more than one set of primers have been included in the primer extension and amplification reactions *(11-13)*. Probes can be stripped by soaking the membranes in boiling 0.1% SDS solution twice for 5-10 min each time.

3.5.5.2 Digoxigenin-Labeled Probe

1. Prehybridize with 20 mL of prehybridization buffer at 60-68°C for at least 3 h.

2. Decant the prehybridization buffer and add 7.5 mL of digoxigenin-labeled probe in prehybridization buffer.

3. Hybridize at 60-68°C (2°C below the calculated T_m of the probe) overnight.

4. Wash the membrane twice with 20 mL of 2X washing solution for 5 min each at room temperature, followed by two washes with 20 mL of 0.1X washing solution for 15 min each at 65°C. The membrane is placed into a rolling 8-cm-diameter x 22-cm-long borosilicate glass hybridization tube in a hybridization oven. Manipulate the membrane exclusively with tweezers (*see* **Note 11**) and do not let it dry following the hybridization step.

5. Wash the membrane with 50 mL of buffer 1 for 1 min at room temperature.

6. Transfer the membrane to a new hybridization tube and incubate with 20 mL of buffer 2 for 1 h at room temperature.

7. Replace the buffer 2 with 20 mL of buffer 2 containing the antidigoxigenin antibody diluted 1:10,000 (prepared 5 min before use) and incubate for 30 min at room temperature.

8. Remove the antibody solution and wash the membrane with 20 mL of buffer 1.

9. Transfer the membrane to a new hybridization tube and incubate with 20 mL of buffer 1 containing 0.3% Tween-20, for 15 min at room temperature.

10. Replace the solution with 20 mL of buffer 3 and incubate for 5 min at room temperature.

11. Place the membrane between two cellulose acetate sheets and pour 0.5 mL:100 cm^2 of CSPD$^®$ diluted 1:100 in buffer 3 onto the membrane between the acetate sheet sandwich. Carefully remove the air bubbles and seal the acetate sheets using heat (Doubleseal). Incubate the membrane for 15 min at 37°C.

12. Expose membrane to X-ray films for 40 min at room temperature (*see* **Note 11**).

3.6 Preparation of Single-Stranded Hybridization Probes (Step VIII, Fig. 5)

The [^{32}P]-dCTP or digoxigenin-labeled single-stranded probe is prepared by 30 cycles of repeated linear primer extension using *Taq* DNA polymerase. Primer 2 (or primer 3, *see* **Note 12**) is extended on a double-stranded template which can be a plasmid or a PCR product. The latter is produced by using two opposing primers 2, separated by a distance of

150-450 bp. Alternatively, any pair of gene specific primers suitable for amplifying a DNA fragment containing a suitable probe sequence (*see* **Note 12**) can be employed.

3.6.1 Template Preparation: PCR Products

3.6.1.1 PCR Amplification

1. To 50 µL of purified genomic DNA (100 ng) in H_2O, add 50 µL of the *Taq* DNA polymerase mix and mix with the pipet. The reaction is overlaid with 50 µL of mineral oil.

2. Cycle 35 times at 95°C for 1 min (97°C for 3 min for the first cycle), 61-73°C (1-2°C below the calculated T_m of primer 2 with the lowest T_m) for 2 min, and 74°C for 3 min. The last extension should be done for 10 min.

3. Add 25 µL of *Taq* DNA polymerase stop under the mineral oil layer.

4. Extract with 250 µL of premixed phenol-chloroform (92 µL:158 µL) and transfer to 1.5-mL microtubes.

5. Add 400 µL of precooled absolute ethanol.

6. Leave 15 min on dry ice, spin 20 min at 15,000g in a centrifuge at 4˚C.

7. Wash once with 1 mL of precooled 80% ethanol.

8. Spin 8 min at 15,000g in a centrifuge at 4˚C.

9. Air-dry DNA pellets.

10. Resuspend DNA pellets in 25 µL of 1X TAE loading buffer.

3.6.1.2 Purification and Quantification of PCR Products

1. Load 25 µL of PCR products per well along with an appropriate DNA mass ladder.

2. Migrate the PCR products on a neutral 1.2-1.5% agarose gel.

3. Stain the gel with ethidium bromide and recover the band containing the DNA fragment of expected molecular weight on a UV transilluminator using a clean scalpel blade. Minimize the size of the slice by removing as much extraneous agarose as possible.

4. Crush the slice and put it in a 0.625-mL microtube pierced at the bottom, and containing a column of packed dry glass wool (*see* **Note 13**).

5. Insert the 0.625-mL microtube containing the column in a 1.5-mL microtube and spin 15 min at 7000g. Transfer the flowthrough to a new 1.5-mL microtube. If there is still some agarose remaining, repeat **step 5**.

6. Add 50 µL of H_2O to wash the column of any remaining DNA by spinning 8 min at 7000g. Pool all of the flowthrough contents in one 1.5-mL microtube.

7. Complete the volume to 405 µL with H_2O, add 45 µL of 3 M sodium acetate, pH 7.0, and 1 mL of precooled absolute ethanol to precipitate DNA. Leave 15 min on dry ice, spin 20 min at 15,000g in a centrifuge at 4°C.

8. Wash once with 1 mL of precooled 80% ethanol and spin 8 min at 15,000g in a centrifuge at 4°C.

9. Air-dry DNA pellet.

10. Dissolve DNA pellets in 103 µL of H_2O.

11. Load aliquots of 1 and 2 µL of the DNA template dissolved in 1X universal neutral loading buffer along with a quantitative DNA molecular-weight ladder and electrophorese on a neutral 1.5% agarose gel.

12. Stain the gel with ethidium bromide and photograph on a UV transilluminator. The DNA concentration of the aliquots is estimated by comparison with the DNA ladder band intensities and H_2O is added to obtain a final concentration of template DNA of 3 ng/µL. The DNA template is aliquoted and stored at -20˚C.

3.6.2 Labeling of Single-Strand Hybridization Probes

3.6.2.1 Isotopic Labeling

1. Prepare 150 µL of the following mix: 30 µL of 5X *Taq* buffer, 3 µL of 100 m*M* MgCl$_2$, 1 µL of dNTPs mix diluted 1:10 in H_2O, 20-40 ng of plasmid or 10-20 ng of PCR products, 1.5 µL of 50 pmol/µL primer 2, 2.5 U of *Taq* DNA polymerase, and 10 µL of α-[^{32}P]-dCTP (3000 ci/mmol).

2. Cycle 30 times at 95°C for 1 min (97°C for 3 min for the first cycle), 60-68°C for 2 min, and 74°C for 3 min.

3. Transfer the mixture to a conical 1.5-mL microtube with screw cap.

4. Precipitate the probe by adding 50 µL of 10 *M* ammonium acetate, 1 µL of glycogen, and 420 µL of precooled absolute ethanol.

5. Leave 5 min at room temperature and spin 5 min at 15,000*g* in a centrifuge at room temperature.

6. Transfer the supernatant to into a new 1.5-mL microtube. Using a Geiger counter, compare the counts per minute between the pellet (probe) and the supernatant, count from the probe should be more or equal to the supernatant for optimal results.

7. Dissolve the probe in 100 µL of TE buffer.

8. Add the probe to 6-8 mL of hybridization buffer and keep the probe at 65°C.

3.6.2.2 Digoxigenin (Nonisotopic) Labeling

1. Prepare 150 µL of the following mix: 30 µL of 5X *Taq* buffer, 3 µL of 100 m*M* MgCl$_2$, 1 µL of dNTP mix diluted 1:8.3, 20-40 ng of plasmid or 10-20 ng of PCR products, 1.5 µL of 50 pmol/µL primer 2, 2.5 U of *Taq* DNA polymerase, and 1.2 µL of 0.5 m*M* digoxigenin-11-dUTP.

2. Cycle 30 times at 95°C for 1 min (97°C for 3 min for the first cycle), 60-68°C for 2 min, and 74°C for 3 min.

3. Precipitate the probe by adding 50 µL of 10 *M* ammonium acetate, 1 µL of glycogen, and 420 µL of precooled absolute ethanol. Spin for 10 min at 15,000*g* in a centrifuge at room temperature.

4. Check the incorporation of the digoxigenin-labeled nucleotide by a dot blot (*see* **Note 14**).

5. Resuspend the probe in 100 µL of TE buffer.

6. Add the probe to 15 mL of prehybridization buffer.

4 Notes

1. Primers should be selected to have a higher T_m at the 5' end than in the 3' end. This higher annealing capacity of the 5' end lowers false priming, thus allowing a more specific extension and less background *(49)*. A guanine or a cytosine residue should also occur at the 3' end. This stabilizes the annealing and facilitates the initiation of the primer extension. It is important that the selected primer does have long runs of purines

or pyrimidines, does not form loops or secondary structure, and does not anneal with itself. If primer dimerization occurs, less primer will be available for annealing and polymerization will not be optimal. The purity of the primers is verified on a 20% polyacrylamide gel (to prepare 500-mL mix: dissolve 96.625 g of acrylamide, 3.375 g of *bis*-acrylamide, 210.21 g of urea corresponding to 7 *M*, in 100 m*M* TBE); if more than one band is found, the primer is reordered. The primers are also tested in a conventional PCR to prepare the template for the probe synthesis (*see* **Note 12**).

2. Originally, *Pfu* and *Taq* buffers were prepared using KCl, which was, however, shown to stabilize secondary DNA structures, thus preventing an optimal polymerization *(50)*. The use of NaCl prevents, to some extent, the ability of DNA to form secondary structures. This is particularly helpful when GC-rich regions of the genome are being investigated.

3. The genomic DNA used for LMPCR needs to be very clean and undegraded. Any shearing of the DNA during preparation and handling before the first primer extension must be avoided. After an incubation of 3 h, if clumps of nuclei are still visible, proteinase K at a final concentration of 300 µg/mL should be added and the sample reincubated at 37°C for another 3 h.

4. If no DNA can be seen, add glycogen (1-2 µg) to the DNA solution and put the DNA on dry ice for 20 min and centrifuge the DNA (5000*g* for 20 min at 4°C). This should help DNA recovery but increases the probability of RNA contamination.

5. Because in vivo DNA analysis is based on comparison of DNA samples modified in vivo with DNA control modified in vitro, given the quantitative characteristic and high sensitivity of LMPCR technology, the DNA concentrations should be as accurate as possible. Indeed, it is critical to start LMPCR with similar amounts of DNA in every sample to be analyzed. The method used to evaluate DNA concentration should measure only double-stranded nucleic acids. RNA contamination does not affect LMPCR, although it can interfere with the precise measurement of the DNA concentration.

6. The DNA frequency of DNA breakage is even more critical than the DNA concentration. For DMS and UV, the base-modification frequency determines the break frequency following conversion of the modified bases to single-strand breaks, whereas for DNase I, the frequency of cleavage is exactly the break frequency. The break frequency must be similar among the samples to be analyzed. It should not average more than one break per 150 bp for in vivo DNA analysis, the optimal break frequency varying from one break per 200 bp to one break per 2000 bp. When the break frequency is too high, we typically observe dark bands over the bottom half of the autoradiogram and very pale bands over the upper half, reflecting the low number of long DNA fragments. In summary, to make the comparison of the in vivo modified DNA sample with a purified DNA control easily interpretable and valid, the amount of DNA and the break frequency must be similar between the samples to be compared. On the other hand, it is not so critical that the break frequency of the sequence ladders (G, A, T+C, and C) be similar to that of the samples to be studied. However, to facilitate sequence reading, the break frequency should be similar between the sequence reactions. It is often necessary to load less DNA for the sequence ladders.

7. If the cell density is too high, multiple cell layers will be formed and the upper cell layer will obstruct the lower ones. This will result in an inhomogeneous DNA photoproduct frequency.

8. It is imperative that the purified DNA samples used as DNA control and the in vivo DNA samples come from the same cell type. For instance, differing cytosine methylation patterns of genomic DNA from different cell types affect photoproduct formation *(17,29)* and give altered DNase I cleavage patterns *(2)*.

9. A nearly ideal chromatin substrate can be maintained in permeabilized cells. Nonionic detergents such as lysolecithin *(48)* and Nonidet P40 *(32)* permeabilize the cell membrane sufficiently to allow the entry of DNase I. Conveniently, this assay can be performed with cells either in a suspension or in a monolayer. One concern is that permeabilized cells will lyse after a certain amount of time in a detergent, thus care must be taken to monitor cell integrity by microscopy during the course of the experiment. A further difficulty with the permeabilization technique concerns the

relatively narrow detergent concentration range over which the assay can be performed. Each cell type appears to require specific conditions for the detergent cell permeabilization. Furthermore, the DNase I concentration must be calibrated for each cell type to produce an appropriate cleavage frequency. Optimally, the in vivo DNase I protocol works better if the enzyme has cleaved the DNA backbone every 1.5-2 kb. Cutting frequencies greater than 1 kb are associated with higher LMPCR backgrounds because the number of 3'-OH ends is much higher, making the suppression of the extension of these ends more difficult.

10. To facilitate sequencing gel removal following migration, it is crucial to siliconize the inner face of both glass plates prior to pouring the gel. For security, cost effectiveness, efficiency, and time-saving, we recommend treating the glass plates with RAIN-AWAY® solution (Wynn's Canada, product no. 63020). We apply 0.75 mL on one plate and 1.5 mL on the other before each five utilizations as specified by the manufacturer. In this way, the gel is easier to pour and will tend to stick on the less siliconized plate. Whenever plastic film is needed, we recommend Saran Wrap® brand. This brand was found to be less permeable to liquids and more resistant to tears then other brands. This is particularly important when membranes are exposed on the phosphorimager plate in order to avoid the moistening of the plate and irreversibly damaging it.

11. We adapted the nonisotopic digoxigenin-based probe labeling method and chemiluminescent detection system (Roche Molecular Biochemicals) to reveal DNA sequence ladders after LMPCR amplification, sequencing gel electrophoresis, and electroblotting. Compared to the isotopic method, the nonisotopic method has a higher specificity, higher sensitivity, lower background, and lower cost, and is therefore a highly recommendable alternative. As shown in **Fig. 6B**, the sequence ladder revealed by non-isotopic labeling was clearer, sharper and presented lesser background as compared to isotopic labeling method (**Fig. 6A**). Unlike isotopic probes, digoxigenin-labeled probes are innocuous, can be easily disposed of, can be stored for long periods, and can even be reused. It is worth noting, however, that this nonisotopic detection method requires some minor precautions. First, the nylon membrane used for this type of detection, must bear a specific density of homogeneous distribution positive charge.

Among membranes we tested, the one sold by Roche Molecular Biochemicals, unquestionably gave the best results. Secondly, care should be taken with the manipulation of the membrane. The use of tweezers is strongly recommended in order to reduce the presence of nonspecific spots and background. As seen in **Fig. 6B**, in spite of taking every precaution, some small spots are still observed on the "chemiluminogram". These might be explained by the powder from gloves. An alternative explanation for these spots could be the presence of nondissolved crystals in the antibody solution (to minimize this problem, this solution can be spun for 15-30 s before use) or in the detection buffer. However, the use of an appropriate membrane and meticulous manipulations can produce very good results with the nonisotopic detection method.

12. To avoid long probes, (i.e., greater than 200 bp), plasmid DNA is cut with an appropriate restriction enzyme (e.g., *see* **ref. 16)**. If a third primer (primer 3) is used to make the probe, it should be selected from the same strand as the amplification primer (primer 2) just 5' to primer 2 sequence and with no or not more than seven to eight bases of overlap on this primer, and have a T_m of 60-68°C. As first reported by Hornstra and Yang *(41,51,52)*, we use the primer 2 employed in the amplification step and we produce the probe from PCR products. Such probes cost less (no primer 3), are more convenient (the preparation of the PCR products permits the testing of primers) and simplify the assay because no cloning requirement is needed as long as the sequence is known.

13. The bottom of a capless 0.625-mL siliconized microtube can be easily pierced with a heated needle. It is important to emphasize that the hole should be made as small as possible for the column to efficiently retain agarose. The pierced microtube is packed with wetted glass wool. Three successive centrifugation steps of 1 min each at 16,000g are necessary to compact and dry the glass wool. The water is recuperated in a capless 1.5-mL microtube. If glass wool is found with the effluent, the column should be discarded. A final 5-min centrifugation at 16,000g should be carried out to ensure the glass wool is fully compacted and dry. The glass wool column is stored at room temperature in a new capless 1.5-mL microtube and covered with a plastic film to

protect the column from dust. In this way, the column can be stored indefinitely until it is used.

14. To verify whether digoxigenin was incorporated in the probe, use an aliquot of 1 μL from the 100-μL probe preparation and pipet it onto a small piece of positively charged membrane (*see* **Subheading 2.5.4., item 13**). Expose the membrane to 1000 J/m^2 of 254-nm UV to crosslink the probe onto the membrane. Place the membrane in a Petri dish and add 15 mL of buffer 1 (*see* **Subheading 2.5.5.2., item 5**). Discard the buffer 1, add 20 mL of buffer 2 (*see* **Subheading 2.5.5.2., item 6**), and place the dish on a shaker for 10 min at room temperature. Discard the buffer 2, add 20 mL of digoxigenin-antibody coupled with a peroxidase (anti-Digoxigenin-AP, Roche Molecular Biochemicals, cat. no. 1 093 274) diluted 1:5000 in buffer 2. Incubate 10 min at room temperature. Add 20 mL of buffer 1 in a new Petri dish, transfer the membrane to this new dish and wash the membrane for 10 min at room temperature. Always manipulate the membrane with tweezers (*see* **Note 11**). Remove the buffer 1 and add 20 mL of buffer 3 (*see* **Subheading 2.5.5.2., item 9**). Wait 5 min to allow the membrane to reach the appropriate pH (pH 9.5) for the detection. During this time, prepare the detection solution by adding 90 μL of NBT (4-Nitroblue tetrazolium chloride, Roche Molecular Biochemicals, cat. no. 1 383 213) and 70 μL of BCIP (X-phosphate/5-bromo-4-chloro-3-indolyl-phosphate, Roche Molecular Biochemicals, cat. no. 1 383 221) to 20 ml of buffer 3. Discard buffer 3 and add the detection solution to the dish containing the membrane and place it in a dark room. Check occasionally and monitor the appearance of staining. If no staining appears after 1 hr, this means that the incorporation of DIG was not efficient. The detection solution is very toxic, manipulate it carefully and eliminate this solution as a toxic waste.

5 Acknowledgments

6 References

1. Pfeifer, G. P., Tanguay, R. L., Steigerwald, S. D., and Riggs, A. D. (1990) *In vivo* footprint and methylation analysis by PCR-aided genomic sequencing: comparison of active and inactive X chromosomal DNA at the CpG island and promoter of human PGK-1. *Genes Dev.* **4,** 1277-1287.

2. Pfeifer, G. P. and Riggs, A. D. (1991) Chromatin differences between active and inactive X chromosomes revealed by genomic footprinting of permealized cells using DNase I and ligation-mediated PCR. *Genes Dev.* **5,** 1102-1113.

3. Chen, C.-J., Li, L. J., Maruya, A., and Shively, J. E. (1995) *In vitro* and *in vivo* footprint analysis of the promoter of carcinoembryonic antigen in colon carcinoma cells: effects of interferon γ treatment. *Cancer Res.* **55,** 3873-3882.

4. Tornaletti, S. and Pfeifer, G. P. (1995) UV light as a footprinting agent: modulation of UV-induced DNA damage by transcription factors bound at the promoters of three human genes. *J. Mol. Biol.* **249,** 714-728.

5. Mueller P. R. and Wold, B. (1989) *In vivo* footprinting of a muscle specific enhancer by ligation mediated PCR. *Science* **246,** 780-786.

6. Pfeifer, G. P., Steigerwald, S. D., Mueller, P. R., Wold, B., and Riggs, A. D. (1989) Genomic sequencing and methylation analysis by ligation mediated PCR. *Science* **246,** 810-813.

7. Pfeifer, G. P., Steigerwald, S. D., Hansen, R. S., Gartler, S. M., and Riggs, A. D. (1990) Polymerase chain reaction-aided genomic sequencing of an X chromosome-linked CpG island: methylation patterns suggest clonal inheritance, CpG site autonomy, and an explanation of activity state stability. *Proc. Natl. Acad. Sci. USA* **87,** 8252-8256.

8. Pfeifer, G. P., Drouin, R., Riggs, A. D., and Holmquist, G. P. (1992) Binding of transcription factors creates hot spots for UV photoproducts *in vivo*. *Mol. Cell. Biol.* **12,** 1798-1804.

9. Church, G. M. and Gilbert, W. (1984) Genomic sequencing. *Proc. Natl. Acad. Sci. USA* **81,** 1991-1995.

10. Pfeifer, G. P. (1992) Analysis of chromatin structure by ligation-mediated PCR. *PCR Methods Appl.* **2,** 107-111.

11. Pfeifer, G. P. and Riggs, A. D. (1993) Genomic footprinting by ligation mediated polymerase chain reaction, in *PCR protocols: Current Methods and Applications* (White, B., ed.), Humana, Totowa, NJ, pp. 153-168.

12. Pfeifer, G. P. and Riggs, A. D. (1993) Genomic sequencing, in *DNA Sequencing Protocols* (Griffin, H. G. and Griffin, A. M., eds.). Humana, Totowa, NJ, pp. 169-181.

13. Pfeifer, G. P., Singer-Sam, J., and Riggs, A. D. (1993) Analysis of methylation and chromatin structure. *Methods Enzymol.* **225,** 567-583.

14. Gao, S., Drouin, R., and Holmquist, G. P. (1994) DNA repair rates mapped along the human PGK1 gene at nucleotide resolution. *Science* **263,** 1438-1440.

15. Tornaletti, S. and Pfeifer, G. P. (1994) Slow repair of pyrimidine dimers at p53 mutation hotspots in skin cancer. *Science* **263,** 1436-1438.

16. Rodriguez, H., Drouin, R., Holmquist, G. P., O'Connor, T. R., Boiteux, S., Laval, J., Doroshow, J. H., and Akman, S. A. (1995) Mapping of copper/hydrogen peroxide-induced DNA damage at nucleotide resolution in human genomic DNA by ligation-mediated polymerase chain reaction. *J. Biol. Chem.* **270,** 17,633-17,640.

17. Drouin, R. and Therrien, J.-P. (1997) UVB-induced cyclobutane pyrimidine dimer frequency correlates with skin cancer mutational hotspots in p53. *Photochem. Photobiol.* **66,** 719-726.

18. Rozek, D. and Pfeifer, G. P. (1993) *In vivo* protein-DNA interactions at the c-jun promoter: preformed complexes mediate the UV response. *Mol. Cell. Biol.* **13,** 5490-5499.

19. Cartwright, I. L. and Kelly, S. E. (1991) Probing the nature of chromosomal DNA-protein contacts by *in vivo* footprinting. *BioTechniques* **11,** 188-203.

20. Maxam, A. M. and Gilbert, W. (1980) Sequencing end-labeled DNA with base-specific chemical cleavages. *Methods Enzymol.* **65,** 499-560.

21. Chin P. L., Momand, J., and Pfeifer, G. P. (1997) *In vivo* evidence for binding of p53 to consensus binding sites in the *p21* and *GADD45* genes in response to ionizing radiation. *Oncogene* **15,** 87-99.

22. Angers, M., Drouin, R., Bachvarova, M., Paradis, I., Marceau, F., and Bachvarov, D. R. (2000) *In vivo* protein-DNA interactions at the kinin B$_1$ receptor gene promoter: no modification upon interleukin-1 beta or lipopolysaccharide induction. *J. Cell. Biochem.* **78,** 278-296.

23. Becker, M. M. and Wang, J. C. (1984) Use of light for footprinting DNA *in vivo*. *Nature* **309,** 682-687.

24. Pfeifer, G. P. and Tornaletti, S. (1997) Footprinting with UV irradiation and LMPCR. *Methods* **11,** 189-196.

25. Pfeifer, G. P., Chen, H. H., Komura, J., and Riggs, A.D. (1999) Chromatin structure analysis by ligation-mediated and terminal transferase-mediated polymerase chain reaction. *Methods Enzymol.* **304,** 548-571.

26. Cadet, J., Anselmino, C., Douki, T., and Voituriez, L. (1992) Photochemistry of nucleic acids in cells. *J. Photochem. Photobiol. B: Biol.* **15,** 277-298.

27. Mitchell D. L., Nairn R. S. (1989) The biology of the (6-4) photoproducts. *Photochem. Photobiol.* **49,** 805-819.

28. Holmquist, G. P. and Gao, S. (1997) Somatic mutation theory, DNA repair rates, and the molecular epidemiology of p53 mutations. *Mutat. Res.* **386,** 69-101.

29. Pfeifer, G. P., Drouin, R., Riggs, A. D., and Holmquist, G. P. (1991) *In vivo* mapping of a DNA adduct at nucleotide resolution: detection of pyrimidine (6-4) pyrimidone photoproducts by ligation-mediated polymerase chain reaction. *Proc. Natl. Acad. Sci. USA* **88,** 1374-1378.

30. Gale, J. M. and Smerdon, M. J. (1990) UV induced (6-4) photoproducts are distributed differently than cyclobutane dimers in nucleosomes. *Photochem. Photobiol.* **51,** 411-417.

31. Gale, J. M., Nissen, K. A., and Smerdon, M. J. (1987) UV-induced formation of pyrimidine dimers in nucleosome core DNA is strongly modulated with a period of 10.3 bases. *Proc. Natl. Acad. Sci. USA* **84,** 6644-6648.

32. Mitchell, D. L., Nguyen, T. D., and Cleaver, J. E. (1990) Nonrandom induction of pyrimidine-pyrimidone (6-4) photoproducts in ultraviolet-irradiated human chromatin. *J. Biol. Chem.* **265,** 5353-5356.

33. Rigaud, G., Roux ,J., Pictet, R., and Grange, T. (1991) *In vivo* footprinting of Rat TAT gene: dynamic interplay between the glucocorticoid receptor and a liver-specific factor. *Cell* **67,** 977-986.

34. Miller, M. R., Castellot, J. J., and Pardee, A. B. (1978) A permeable animal cell preparation for studying macromolecular synthesis. DNA synthesis and the role of deoxyribonucleotides in S phase initiation. *Biochemistry* **17,** 1073-1080.

35. Contreras, R. and Fiers, W. (1981) Initiation of transcription by RNA polymerase II in permeable SV40-infected CV-1 cells; evidence of multiple promoters for SV40 late transcription. *Nucleic Acids Res.* **9,** 215-236.

36. Tanguay, R. L., Pfeifer, G. P., and Riggs, A. D. (1990) PCR-aided DNase I footprinting of single-copy gene sequences in permeabilized cells. *Nucleic Acids Res.* **18,** 5902.

37. Törmänen, V. T., Swiderski, P. M., Kaplan, B. E., Pfeifer, G. P., and Riggs, A. D. (1992) Extension product capture improves genomic sequencing and DNase I footprinting by ligation-mediated PCR. *Nucleic Acids Res.* **20,** 5487-5488.

38. Tornaletti, S., Bates, S., and Pfeifer, G. P. (1996) A high-resolution analysis of chromatin structure along p53 sequences. *Mol. Carcinogen.* **17,** 192-201.

39. Szabo, P. E., Pfeifer, G. P., and Mann, J. R. (1998) Characterization of novel parent-specific epigenetic modifications upstream of the imprinted mouse *H19* gene. *Mol. Cell. Biol.* **18,** 6767-6776.

40. Garrity, P. A. and Wold, B. J. (1992) Effects of different DNA polymerases in ligation-mediated PCR: Enhanced genomic sequencing and *in vivo* footprinting. *Proc. Natl. Acad. Sci. USA* **89,** 1021-1025.

41. Hornstra, I. K. and Yang, T. P. (1994) High resolution methylation analysis of the human hypoxanthine phosphoribosyltransferase gene 5' region on the active and inactive X chromosomes: Correlation with binding sites for transcription factors. *Mol. Cell. Biol.* **14,** 1419-1430.

42. Angers, M., Cloutier, J.-F., Castonguay A, and Drouin, R. (2001) Optimal conditions to use *Pfu* exo⁻ DNA polymerase for highly efficient ligation-mediated polymerase chain reaction protocols. *Nucleic Acids Res.* **29,** e83-e92.

43. Drouin, R., Gao, S., and Holmquist, G. P. (1996) Agarose gel electrophoresis for DNA damage analysis, in *Technologies for Detection of DNA Damage and Mutations* (Pfeifer, G. P., ed.), Plenum, New York, pp. 37-43.

44. Rychlik, W. and Rhoads, R. E. (1989) A computer program for choosing optimal oligonucleotides for filter hybridization, sequencing and in vitro amplification of DNA. *Nucleic Acids Res.* **17,** 8543-8551.

45. Drouin, R., Rodriguez, H., Holmquist, G. P., and Akman, S. A. (1996) Ligation-mediated PCR for analysis of oxidative DNA damage, in *Technologies for Detection of DNA Damage and Mutations* (Pfeifer, G. P., ed.), Plenum Press, New York, pp. 211-225.

46. Mueller P.R. and Wold, B. (1991) Ligation-mediated PCR: applications to genomic footprinting. *Methods* **2,** 20-31.

47. Iverson, B. L. and Dervan, P. B. (1987) Adenine specific DNA chemical sequencing reaction. *Nucleic Acids Res.* **19,** 7823-7830.

48. Zhang, L. and Gralla, J. D. (1989) *In situ* nucleoprotein structure at the SV40 major late promoter: melted and wrapped DNA flank the start site. *Genes Dev.* **3,** 1814-1822.

49. Rychlik, W. (1993) Selection of primers for polymerase chain reaction, in *PCR protocols: Current Methods and Applications* (White, B., ed.), Humana, Totowa, NJ, pp. 31-40.

50. Fry, M. and Loeb, L. A. (1994) The fragile X syndrome d(CGG)n nucleotide repeats form a stable tetrahelical structure. *Proc. Natl. Acad. Sci. USA* **91,** 4950-4954.

51. Hornstra, I. K. and Yang, T. P. (1992) Multiple *in vivo* footprints are specific to the active allele of the X-linked human hypoxanthine phosphoribosyltransferase gene 5' region: implications for X chromosome inactivation. *Mol. Cell. Biol.* **12,** 5345-5354.

52. Hornstra, I. K. and Yang, T. P. (1993) *In vivo* footprinting and genomic sequencing by ligation-mediated PCR. *Anal. Biochem.* **213,** 179-193.

7 Tableaux

Table 1 Purposes of the Three Main In Vivo Footprinting Approaches

Approaches	Activities
Dimethylsulfate (DMS)	**i**.Localizes in vivo DNA-protein contacts located in the major groove of the DNA double helix **ii**.Can detect special DNA structures
UV irradiation (UVB or UVC)	**i**.Localizes in vivo DNA-protein interactions and shows how DNA structure is affected in the presence of transcription factors **ii**.Can detect special DNA structures **iii**.Can show evidence of positioned nucleosomes
DNase I	**i**.Localizes in vivo DNA-protein contacts **ii**.Precisely maps in vivo DNase I hypersensitive sites **iii**.Shows evidence of nucleosomes and their positions; can differentiate core DNA from linker DNA

Table 2 Advantages and Drawbacks of the Three Main In Vivo Footprinting Approaches

Approaches	Advantages	Drawbacks
DMS	Treatment is technically easy to carry out; the DMS is a small molecule that penetrates very easily into living cells with little disruption.	1.Requires guanines, therefore is sequence dependent. 2.Does not detect all DNA-protein interactions.
UV irradiation (UVB or UVC)	1.Treatment is technically easy to carry out; UV light penetrates through the outer membrane of living cells without disruption. 2.Detects many DNA-protein interactions. 3.Very sensitive to particular DNA structures.	1.Requires two adjacent pyrimidines, therefore is sequence dependent. 2.The interpretation of the results is sometime difficult; to differentiate between DNA-protein interactions and special DNA structures can be very difficult.
DNase I	1.Little sequence dependency. 2.No conversion of modified bases required. 3.Detects all DNA-protein contacts. 4.Very sensitive to particular DNA structures.	1.Technically difficult to carry out; reproducibility is often a problem. 2.DNase I is a protein that can penetrate in living cells only following membrane permeabilization, thus causing some cell disruption.

Table 3 Mapping Schemes Used with the Three Main In Vivo Footprinting Approaches

Approaches	Strand breaks	Modified bases	Conversion of modified bases to DNA single-strand breaks
DMS	Few	Guanine: methylated guanines at N7 position Adenine: to a much lesser extent, methylated adenines at N3 position	Hot piperidine
UV irradiation (UVB or UVC)	Very few	(i)Cyclobutane pyrimidine dimers (ii)6-4 Photoproducts	(i)T_4 endonuclease V followed by photolyase (ii)Photolyase followed by hot piperidine
DNase I	Yes	None	No conversion required

Table 4 Exponential Amplification Steps Using Cloned *Pfu* DNA Polymerase or *Taq* DNA Polymerase

Cycle	Denaturation (T in °C for D in s)		Annealing (T is the T_m of the oligonucleotide for D in s)	Polymerization (D in s) T is the same for all cycles: 75°C for *Pfu* and 74°C for *Taq*
	Pfu	*Taq*	*Pfu* or *Taq*	
0	-	93 for 120	-	-
1	98 for 300	98 for 150	T_m for 180	180
2	98 for 120	95 for 60	T_m -1°C for 150	180
3	98 for 60	95 for 60	T_m -2°C for 120	180
4	98 for 30	95 for 60	T_m -3°C for 120	180
5	98 for 20	95 for 60	T_m -4°C for 90	150
Repeat cycle 5, 13 more times (add 5 s per cycle for annealing and polymerization)				
19	98 for 20	95 for 60	T_m -3°C for 240	240
20	98 for 20	95 for 60	T_m -2°C for 240	240
21	98 for 20	95 for 60	T_m -1°C for 240	240
22	98 for 20	95 for 60	T_m for 240	600

Note: Temperature (T) and duration (D) of the denaturation, annealing and polymerization steps.

8 Légendes des figures

Figure 1. Overall scheme for in vivo DNA analysis using DMS. The methylation of guanine residues following DMS treatment of purified DNA (in vitro) and cells (in vivo) is shown by vertical arrows and methylated residues (Me). When purified DNA is treated with DMS, every guanine residue has a similar probability of being methylated. However, the guanine residue in intimate contact with a sequence-specific DNA-binding protein illustrated by the dotted oval is protected from DMS methylation, whereas the guanine residue localized close to the boundary of the DNA-protein contact that modifies DNA structure, allowing a better accessibility to DMS, is methylated more frequently. The methylated guanine residues are cleaved by hot piperidine leaving phosphorylated 5' ends. On the sequencing ladder following LMPCR, guanine residues that are protected from methylation appear as missing or less intense bands when compared with the sequencing ladder from the same DNA sequence obtained after DMS treatment of purified DNA. On the other hand, guanine residues that undergo enhanced DMS methylation appear as darker bands in the sequencing ladder relative to the purified DNA control.

Figure 2. Overall scheme for in vivo DNA analysis using UVC and CPD formation. The CPD formation following UVC exposure of purified DNA (in vitro) and cells (in vivo) is shown with curved arrows and brackets linking two adjacent pyrimidines (Y). When purified DNA is irradiated with UVC, the frequency of CPD formation at dipyrimidine sites is determined by the DNA sequence. However, the presence of a sequence-specific DNA-binding protein illustrated by the dotted oval as well as DNA structure can prevent (negative photofootprint) or enhance (positive photofootprint) CPD formation. The CPDs are cleaved by T_4 endonuclease V digestion and photolyase photoreactivation leaving phosphorylated 5' ends. On the sequencing ladder following LMPCR, the negative photofootprints appear as missing or less intense bands when compared with the sequencing ladder from the same DNA sequence obtained after UVC irradiation of purified DNA. On the other hand, positive photofootprints appear as darker bands in the sequencing ladder relative to the purified DNA control.

Figure 3. Overall scheme for in vivo DNA analysis using UVC and 6-4PP formation. The 6-4PP formation following UVC exposure of purified DNA (in vitro) and cells (in vivo) is shown with curved arrows and brackets linking two adjacent pyrimidines (Y). When purified DNA is irradiated with UVC, the frequency of 6-4PP formation at dipyrimidine sites is determined by the DNA sequence. However, the presence of a sequence-specific DNA-binding protein illustrated by the dotted oval as well as DNA structure can prevent (negative photofootprint) or enhance (positive photofootprint) 6-4PP formation. First, CPDs are photoreactivated by photolyase and then 6-4PPs are cleaved by hot piperidine treatment leaving phosphorylated 5' ends. On the sequencing ladder following LMPCR, the negative photofootprints appear as missing or less intense bands when compared with the sequencing ladder from the same DNA sequence obtained after UVC irradiation of purified DNA. On the other hand, positive photofootprints appear as darker bands in the sequencing ladder relative to the purified DNA control.

Figure 4. Overall scheme for in vivo DNA analysis using DNase I. The DNase I enzyme (the solid black) digestion of purified DNA (in vitro) and cells (in vivo) is shown. When purified DNA is digested with DNase I, the cleavage pattern shows that sites of the nucleotide sequence have similar probabilities of being cleaved. However, the presence of a sequence-specific DNA-binding protein illustrated by the dotted oval as well as DNA structure can prevent (protection) or enhance (hypersensitive) DNase I cleavage. The DNase I cleavage leaves phosphorylated 5' ends. On the sequencing ladder following LMPCR, DNA sequences that are protected from DNase I cleavage appear as missing or less intense bands when compared with the sequencing ladder from the same DNA sequence obtained after DNase I digestion of purified DNA. On the other hand, hypersensitive sites that undergo enhanced DNase I cleavage appear as darker bands in the sequencing ladder relative to the purified DNA control.

Figure 5. Outline of the LMPCR procedure. Step I: specific conversion of modified bases to phosphorylated 5' single-strand breaks; Step II: denaturation of genomic DNA; Step III: annealing and extension of primer 1 (although both strands can be studied, each LMPCR protocol only involves the analysis of either the nontranscribed strand or the transcribed

strand); Step IV: ligation of the linker; Step V: first cycle of PCR amplification, this cycle is a linear amplification because only the gene-specific primer 2 can anneal; Step VI: cycle 2 to 22 of exponential PCR amplification of gene-specific fragments with primer 2 and the linker primer (the longer oligonucleotide of the linker); Step VII: separation of the DNA fragments on a sequencing gel, transfer of the sequence ladder to a nylon membrane by electroblotting, and visualization of the sequence ladder by hybridization with a labeled single-stranded probe; Step VIII: preparation and isotopic or nonisotopic labeling of single-stranded probe.

Figure 6. LMPCR analysis of methylated guanines and CPD along the nontranscribed strand of the c-jun promoter following DMS treatment, and UVB and UVC irradiation, respectively. (**A**) The membrane was hybridized with an isotopic [^{32}P]-dCTP-labeled probe. The membrane was exposed on film between two intensifying screens for 25 min at -70°C. (**B**) The membrane was hybridized with a digoxigenin-labeled probe and exposed on film for 40 min at room temperature. For this experiment, one LMPCR protocol was carried out and only one gel was run on which all the samples (20 in total) were loaded symmetrically in duplicate. Each symmetrical well of each set of samples was loaded with exactly the same amount of DNA. Lanes 1-4: LMPCR of DNA-treated with chemical cleavage reactions. These lanes represent the sequence of the *c-jun* promoter analyzed with JD primer set *(18)*. Lanes 5-6: LMPCR of DMS-treated naked DNA (T: in vitro) and fibroblasts (V: in vivo) followed by hot piperidine treatment. Lanes 7-10: LMPCR of UVC- and UVB-irradiated naked DNA (T) and fibroblasts (V) followed by T$_4$ endonuclease V/photolyase digestions. On the right, the consensus sequences of transcription factor binding sites are delimited by brackets. The numbers indicate their positions relative to the major transcription initiation site.
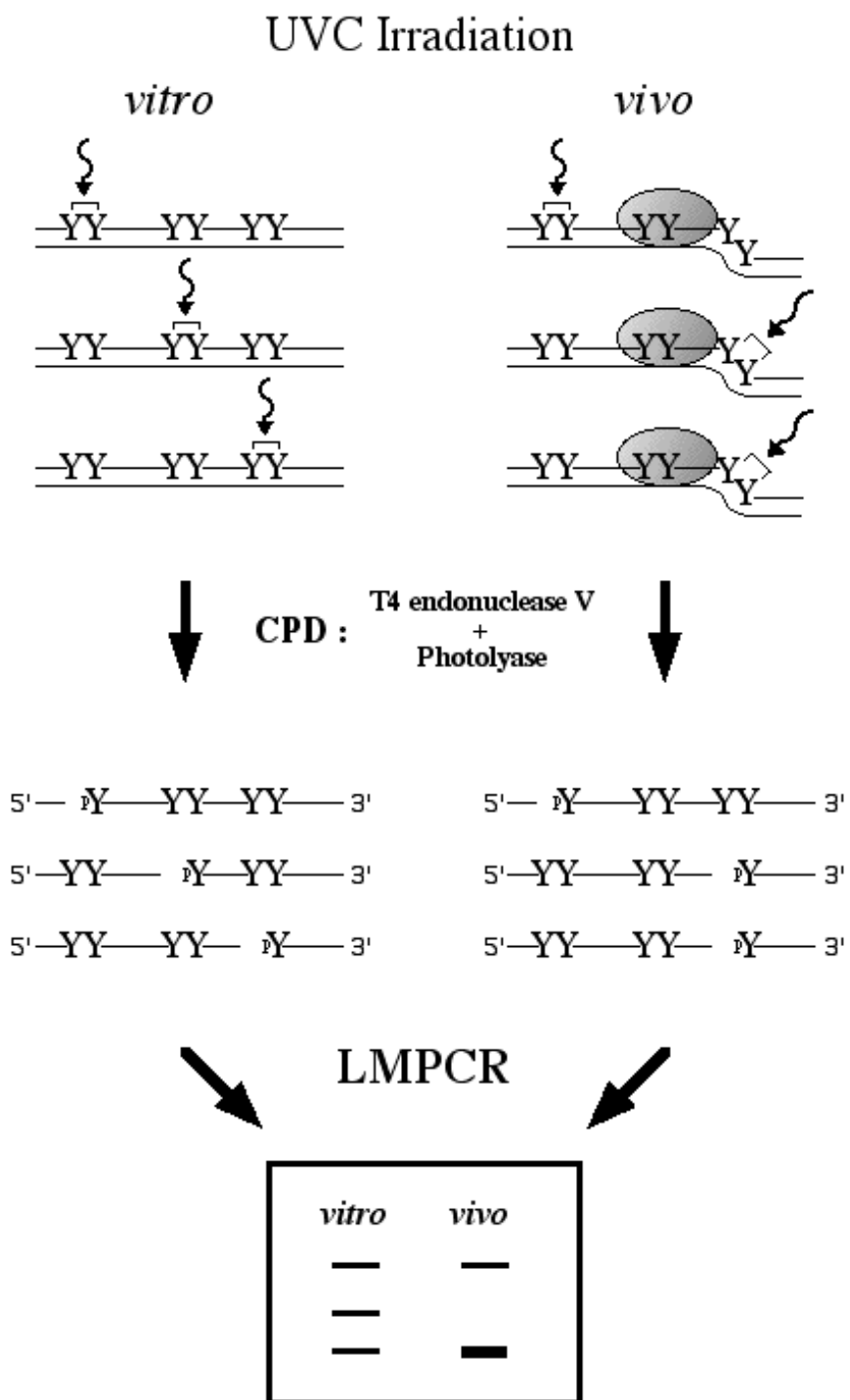
Figure 7. LMPCR analysis of methylated guanines and DNA strand breaks along the transcribed strand of the c-jun promoter following DMS treatment and DNase I digestion respectively. The membrane was hybridized with an isotopic [^{32}P]-dCTP-labeled probe. Lanes 1-2: LMPCR of DMS-treated purified DNA (t: in vitro) and fibroblasts (v: in vivo) followed by hot piperidine treatment. Lanes 3-6: LMPCR of DNA treated with chemical

cleavage reactions. These lanes represent the sequence of the *c-jun* promoter analyzed with JC primer set *(18)*. Lanes 7-8: LMPCR of DNase I-digested permeabilized fibroblasts (v) and purified DNA (t). As a reference, a small portion of the chemically derived sequence is shown on the right of the autoradiogram, the AP-1-like binding sequence is enclosed by a box, and the numbers indicate its position relative to the major transcription initiation site. Open circles represent guanines that are protected against DMS-induced methylation (negative DMS footprints) in vivo. The black bar shows the protected sequence against DNase I-induced cleavage in vivo. Thus, in vivo DNase I footprinting analysis delimits much better the DNA-protein interactions.
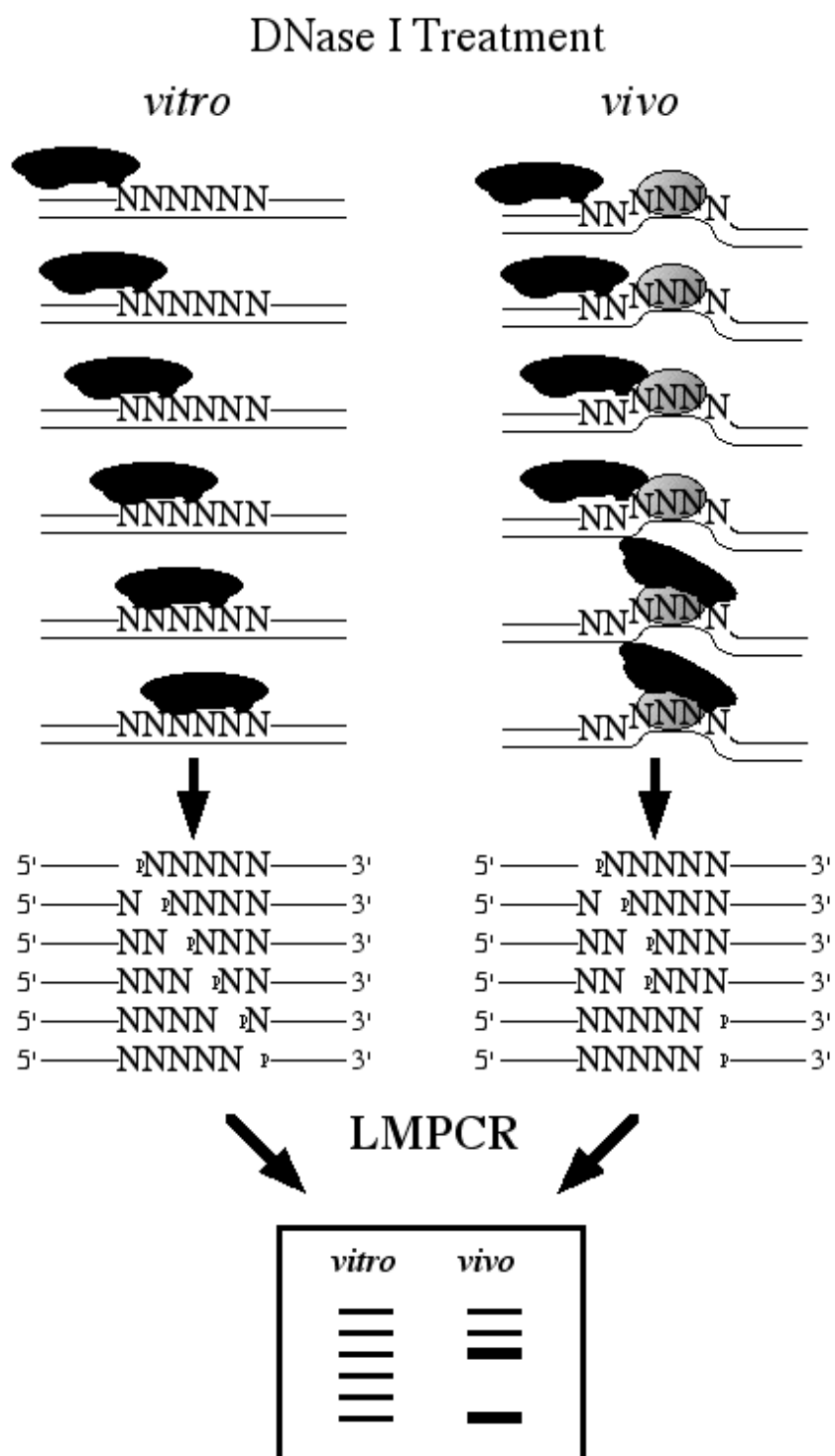
9 Figures
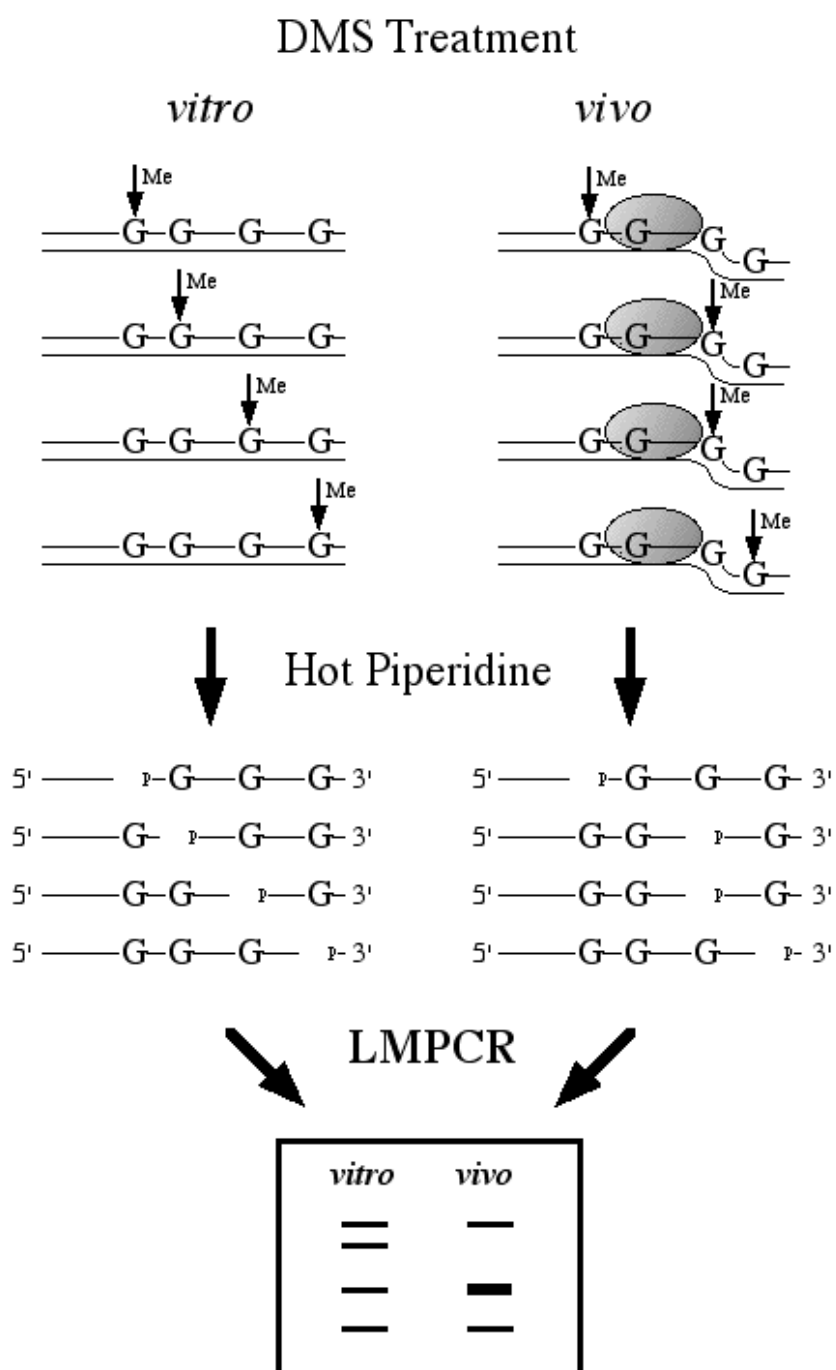
Figure 1 Overall scheme for in vivo DNA analysis using DMS.

Figure 2 Overall scheme for in vivo DNA analysis using UVC and CPD formation.
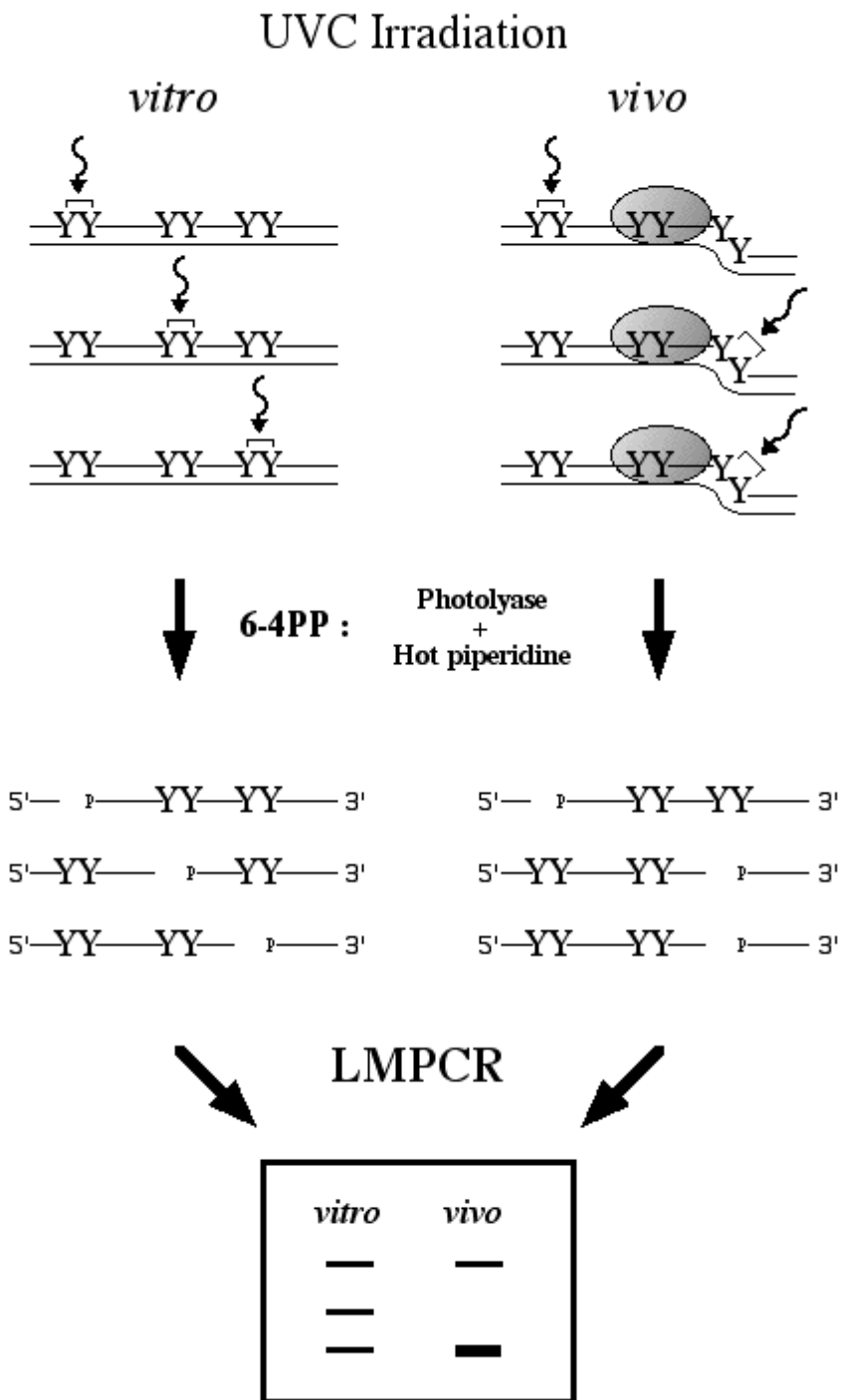
Figure 3 Overall scheme for in vivo DNA analysis using UVC and 6-4PP formation.

Figure 4 Overall scheme for in vivo DNA analysis using DNase I.
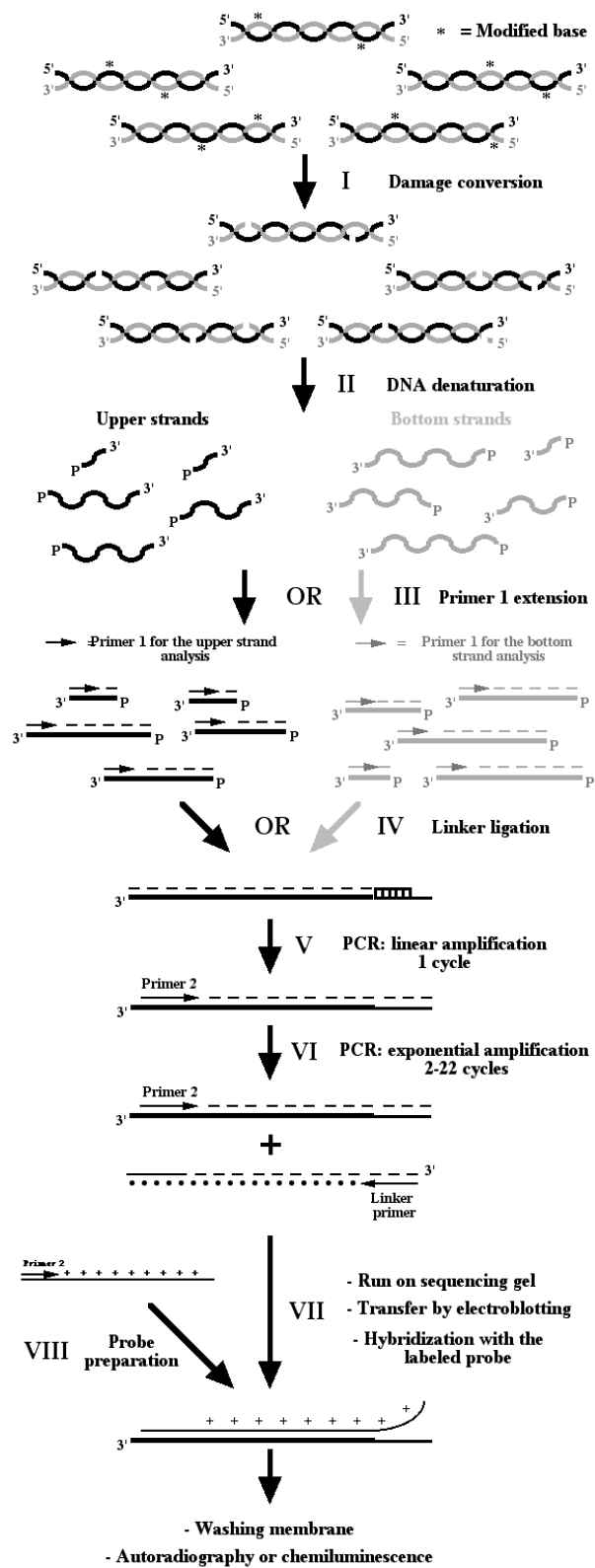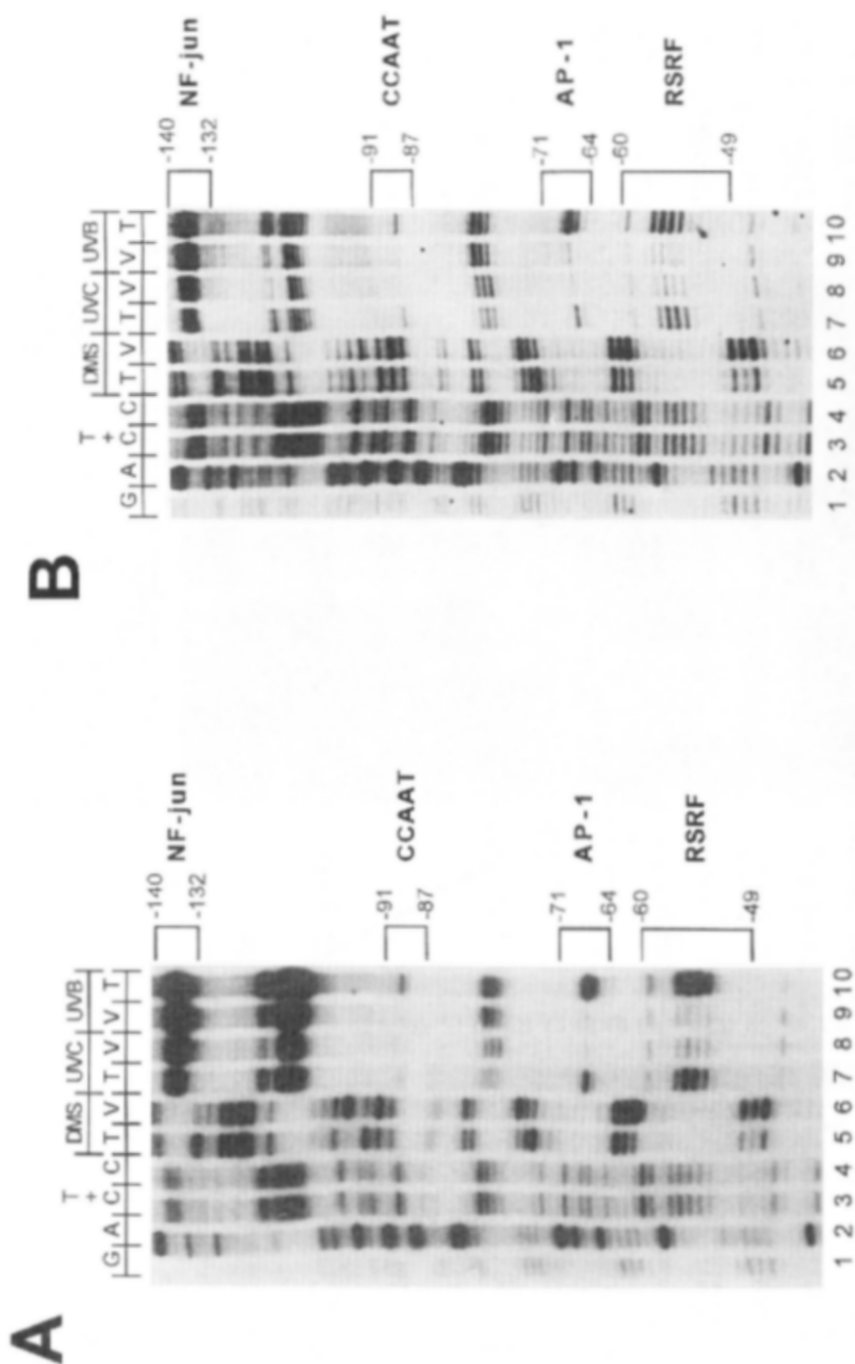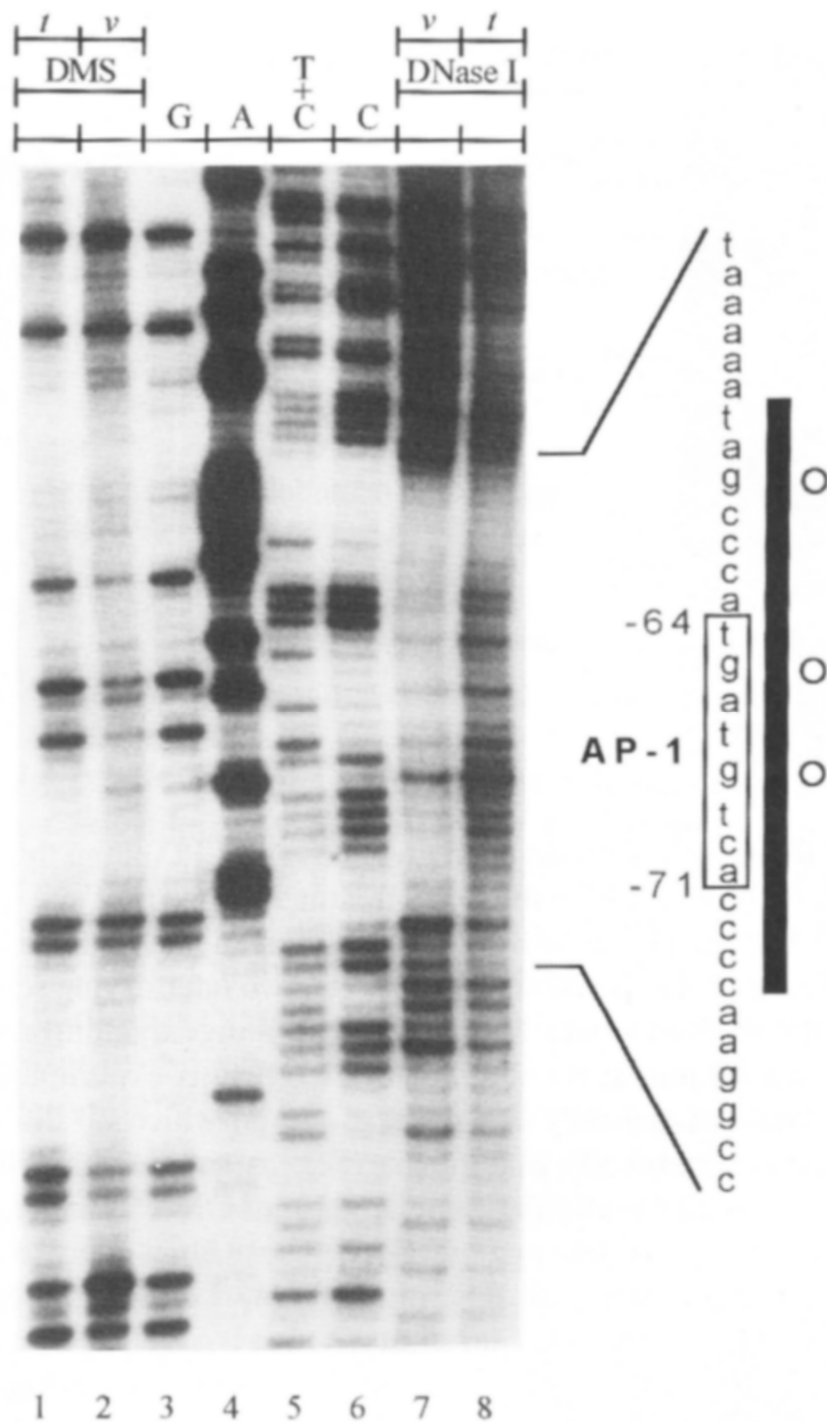
Figure 5 Outline of the LMPCR procedure.

Figure 6 LMPCR analysis of methylated guanines and CPD along the nontranscribed strand of the c-jun promoter following DMS treatment, and UVB and UVC irradiation, respectively. The membrane was hybridized with an (**A**) isotopic [^{32}P]-dCTP- or a (**B**) digoxigenin-labeled probe.

Figure 7 LMPCR analysis of methylated guanines and DNA strand breaks along the transcribed strand of the c-jun promoter following DMS treatment and DNase I digestion respectively.