

# **Résumé des travaux effectués à l'IGBMC Strasbourg [Sept 1991 - Dec 1996]**

## **INTRODUCTION**

Dans le noyau des cellules eucaryotes, la plupart des RNA pré-messagers subissent des modifications post-transcriptionnelles qui aboutissent à la formation de RNA messagers matures. La modification la plus importante est représentée par le processus d'épissage qui conduit à l'excision des introns non codants. Ce processus a lieu dans une structure dynamique appelée complexe d'épissage ou "spliceosome" (1, 2). La formation de ce complexe survient en plusieurs étapes, et requiert à la fois la présence de petites particules ribonucléoprotéiques (snRNP U1,U2,U4,U5 et U6) et de facteurs protéiques (1, 2). A ce jour, peu de facteurs de type non-snRNP ont été identifiés chez les eucaryotes supérieurs, puisque seulement les facteurs U2AF pour "U2 auxilliary factor" (3), PSF pour "pPTB-associated splicing factor" (4), ainsi que la famille des facteurs riches en sérine/arginine (SR) (5-11) ont été clonés et caractérisés. L'existence d'une famille de facteurs SR a été révélée à l'aide d'un anticorps reconnaissant des protéines de 75, 55, 40, 30 et 20 kDa (5). Ces protéines sont très conservées d'une espèce à l'autre et ont été appelées respectivement SRp75, SRp55, SRp40, SRp30a, SRp30b et SRp20. Les facteurs SRp30a et 30b étaient déjà bien caractérisés et connus sous le nom de ASF/SF2 (8, 9) et SC35/PR264 (10, 11) respectivement.

Les protéines constituant la famille SR présentent des caractéristiques structurales et fonctionnelles très voisines. En effet, elles sont toutes constituées d'un ou plusieurs domaines de liaison au RNA (RBD pour RNA Binding Domain) dans la partie N terminale et d'une région riche en un dipeptide SR dans la partie C terminale (5-11). Par ailleurs, toutes ces protéines sont capables, prises individuellement, de compléter une fraction cytoplasmique (S100) pour l'épissage constitutif (5, 8, 9) et certaines modulent un épissage alternatif en favorisant l'utilisation de sites donneurs proximaux (5, 6, 8, 9). La modulation de l'épissage alternatif est importante car de nombreux gènes sont régulés par ce type de réaction. Ceci a été particulièrement bien démontré pour de nombreux facteurs de transcription et des proto-oncogènes pour lesquels les exons définissent des domaines structuraux et fonctionnels bien établis (12, 13). Il a été montré que des facteurs de type SR sont capables de moduler l'épissage alternatif de nombreux pré-messagers (14, 15), mais leur fonction réelle au sein de la cellule reste à démontrer.

## **RESUME DES RESULTATS OBTENUS AU COURS DE MA THESE**

Au cours de mes quatre années d'études dans ce laboratoire, j'ai largement participé à l'identification, au clonage et à la caractérisation d'un nouveau membre, de 35 kDa, de la famille des facteurs SR que nous avons appelé 9G8.

## **1. Caractérisation du gène 9G8**

### **1.1 Identification et clonage du cDNA codant pour le facteur d'épissage 9G8**

Celui-ci, comme les autres membres de la famille, possède un domaine N terminal de liaison au RNA (RBD), et une région C terminale riche en sérines et arginines. Cependant, ce facteur d'épissage possède des caractéristiques qui lui sont spécifiques comme un doigt à zinc dans la partie médiane et des répétitions RRSRSXSX dans le domaine SR. Nous avons montré que le doigt à zinc du facteur 9G8 est fonctionnel car capable de fixer le zinc.

### **1.2 Caractérisation du gène 9G8 et de ses produits d'expression**

Par ailleurs, nous avons caractérisé le gène codant pour ce facteur et nous avons pu observer qu'il contient huit exons et sept introns, la séquence codant pour ce facteur étant distribuée sur les huit exons. De plus, par des analyses en northern blot, nous avons montré que l'expression de notre gène varie dans les différents tissus humains. De plus, nous avons mis en évidence plusieurs isoformes de mRNA correspondant à notre gène, qui sont elles-mêmes modulées les unes par rapport aux autres dans ces tissus. Parmi ces isoformes, nous avons montré qu'il existe des mRNA conservant le troisième intron du gène. Cette rétention d'intron conduit à la traduction d'une protéine déléetée du domaine SR, par l'utilisation prématurée d'un codon STOP. Il a été précédemment décrit par les groupes d'A. Krainer (16) et de J. Manley (17) que ce domaine SR est indispensable pour la régulation de l'épissage alternatif, mais qu'il n'intervient pas dans l'épissage constitutif. Nous avons également montré que l'expression du gène 9G8 est soumise à une régulation par plusieurs facteurs. Ainsi, les facteurs CREB et CREM $\tau$ , l'oncogène myb ainsi que la protéine E1A d'adénovirus sont capables d'activer la transcription du gène 9G8.

## **2. Le facteur 9G8 et la modulation de l'épissage alternatif**

### **2.1 Exemple de l'épissage alternatif des transcrits E1A au cours de l'infection virale**

Nous avons également mis en évidence que le facteur 9G8 est essentiel pour la réaction constitutive et qu'il est capable de moduler une réaction d'épissage alternative. En utilisant comme modèle l'unité E1A d'adénovirus, nous avons montré que la modulation de l'épissage alternatif, observée sur cette unité au cours de l'infection virale, est due à une titration des facteurs de type SR, par des transcrits viraux synthétisés en grandes quantités dans le noyau en période tardive d'infection. Par ailleurs, nous avons démontré que le facteur 9G8 est capable de lier directement le RNA.

### **2.2 Etude des interactions entre la protéine 9G8 et le RNA**

Afin de mieux comprendre les mécanismes moléculaires impliqués dans la modulation de l'épissage alternatif, nous avons décidé de centrer nos recherches sur les interactions RNA/facteur d'épissage SR. En effet, bien que tous ces facteurs de type SR soient redondants pour l'épissage constitutif *in vitro*, des études récentes ont montré que la plupart des facteurs SR présente une spécificité de substrat plus ou moins marquée. Ainsi, X.D. Fu a démontré que SC35 était capable d'initier la formation d'un complexe précoce d'épissage d'un transcrit de  $\beta$  Globine alors que ASF/SF2

est dépourvu de cette activité (18). De plus, de nombreuses séquences exoniques riches en purines ont été identifiées comme des éléments *cis* activateurs de la réaction d'épissage en favorisant la reconnaissance de l'intron situé juste en amont. Pour l'hormone de croissance bovine (bGH), il a été clairement prouvé que ASF/SF2 fixe cette séquence et stimule l'épissage de l'intron situé en amont (19). Afin de déterminer les séquences cibles de la protéine 9G8, et de les comparer avec celles de deux autres facteurs SR (ASF/SF2 et SC35), nous avons choisi une approche *in vitro* basée sur la sélection et l'amplification de séquences ribonucléotidiques aléatoires (SELEX) (20). Pour cela, j'ai créé différents mutants des protéines 9G8, ASF/SF2 et SC35 tous dépourvus du domaine SR ( $\Delta$ SR), qui est susceptible de diminuer la spécificité d'interaction avec le RNA. De plus, afin d'analyser le rôle putatif du motif en doigt à zinc présent dans le facteur 9G8, celui-ci a été détruit par mutation ponctuelle. Ces mutants, exprimés dans *E. Coli* en fusion à la glutathion S transférase (GST) sont utilisés pour sélectionner les RNA contenant une séquence aléatoire de 20 nucléotides.

### 2.2.1 Analyse par SELEX

J'ai réalisé différents cycles de sélection avec les facteurs 9G8 $\Delta$ SR, SC35 $\Delta$ SR et 9G8Zn<sup>m</sup> $\Delta$ SR et de nombreux clones ont été séquencés et analysés. Parmi les 56 séquences obtenues avec la protéine 9G8 $\Delta$ SR, 51 ont été classées en deux catégories. Toutes les séquences sélectionnées sont riches en purines les consensus correspondants ont été identifiés (AGAC<sup>G/T</sup>ACGAC<sup>C/T</sup> et ACGAGAGAC<sup>C/T</sup>). Lorsque la sélection a été effectuée avec la protéine GST-9G8Zn<sup>m</sup> $\Delta$ RS dont les deux premières cystéines du doigt à zinc ont été mutées, le motif sélectionné est sensiblement différent. En effet, dans ce cas, la protéine GST-9G8Zn<sup>m</sup> $\Delta$ RS ne fixe plus de séquences riches en purines mais des séquences enrichies en C et en A. Ainsi, la destruction du doigt à zinc de la protéine 9G8 s'accompagne d'une variation dans la sélection des RNA. Ceci indique donc un effet coopératif des deux motifs: RBD et doigt à zinc, pour la définition de la spécificité de fixation. Parallèlement, des expériences de SELEX ont été réalisées avec la protéine SC35 $\Delta$ RS fusionnée à la GST (GST-SC35 $\Delta$ RS), cette protéine servant de référence. En effet, l'équipe de J. Manley a aussi utilisé cette protéine, mais fusionnée à une queue d'histidines dans le même type d'expériences. Ainsi, la comparaison de nos résultats avec ceux obtenus par Tacke et Manley nous permettent de vérifier deux points: (i) que deux facteurs différents, en l'occurrence 9G8 et SC35, reconnaissent des séquences ribonucléotidiques réellement différentes; (ii) que la fusion GST n'entraîne pas de variation dans les résultats. Dans cette analyse, 67 clones parmi les 68 séquencés, ont pu être classés en plusieurs grandes catégories en fonction de la séquence consensus identifiée. Si 27 des clones obtenus présentent des motifs riches en purines, les séquences des 40 autres s'écartent sensiblement des motifs classiquement reconnus par les facteurs 9G8 (voir ci-dessus) et ASF/SF2 (21). Notons que les consensus des séquences cibles de SC35 déterminés ici sont similaires à ceux retrouvés par l'équipe de Manley. Cependant, nous avons affiné cette étude puisque nous définissons, en plus des deux motifs déjà obtenus par le groupe de J. Manley, trois autres motifs consensus voisins.

### 2.2.2 Etude des interactions directes RNA/protéine

Afin de vérifier de manière directe les résultats obtenus par la méthode du SELEX, différents clones individuels ont été analysés pour leur aptitude à être reconnu par les quatre protéines

chimériques étudiées. Cette étude a été menée par des essais de retard de migration sur gel non dénaturant. Il ressort que (i) chaque molécule de RNA est fixée de la manière la plus efficace par la protéine qui a servi à sa sélection, indiquant clairement que chaque protéine possède une spécificité de liaison au RNA; (ii) le doigt à zinc de 9G8 est bien impliqué dans la définition de la spécificité de liaison au RNA de cette protéine; (iii) chaque protéine fixe ses molécules cibles de RNA avec un Kd apparent de l'ordre de 1 nM, cette valeur étant du même ordre de grandeur que celle observée avec d'autres facteurs fixant des séquences spécifiques de RNA tels que les hnRNP A1 et C.

Ce travail s'est poursuivi par une étude des interactions dans des conditions plus physiologiques. Ainsi, nous avons analysé les interactions de divers RNA représentatifs de chacun des trois facteurs de 30 kDa, avec une population de facteurs SR contenant, en plus des précédents, les facteurs SRp20, 40, 55 et 75 (poids moléculaire de 20 à 75 kDa). Par compétition, analyse de “*supershift*” en présence d'anticorps dirigés contre les facteurs de 30 kDa et pontage covalent aux rayons ultra-violet, nous avons montré que les interactions facteurs SR/RNA s'effectuent avec une bonne spécificité. Un résultat de même type est également obtenu quand les RNA cibles sont mis en présence d'un extrait nucléaire contenant les facteurs SR endogènes. Dans ce cas, les facteurs SR interagissant avec les RNA sont essentiellement révélés par pontage aux U.V. du RNA radioactif aux facteurs impliqués, suivi d'une immunosélection des facteurs SR. Grâce à cette technique, que nous avons dû introduire dans le groupe, nous démontrons que chacun des facteurs SR étudié est capable de reconnaître préférentiellement une séquence ribonucléique, indiquant une possible spécificité dans leur action vis à vis de la réaction d'épissage alternative.

### *2.2.3 Implication de la liaison protéine SR/séquence cible de RNA sur l'épissage in vitro.*

L'intérêt d'une telle approche est également de pouvoir définir le rôle du facteur 9G8 dans la modulation et la régulation de l'épissage alternatif pour un substrat donné. Nous avons actuellement exprimé les trois facteurs SR de 30 kDa (9G8, ASF/SF2 et SC35) dans des cellules d'insecte infectées par le baculovirus. Ils sont utilisés dans des systèmes d'épissages alternatifs modèles afin de déterminer leur rôle respectif dans des modulations d'épissage alternatif. Il est important de noter que les séquences reconnues par les facteurs 9G8, ASF/SF2 et SC35 (dans le cas du motif riche en purines) sont très voisins de ceux décrits comme des activateurs exoniques d'épissage (ESE). Ces motifs, présents dans divers exons reconnus de manière alternative sont capables de compenser la faiblesse d'un site 3' d'épissage les bordant en amont (19, 22, 23). Ainsi, deux modèles ont été créés:

(i) le première consiste en une fusion entre le premier exon de l'unité majeure tardive d'adénovirus et de séquences introniques en aval d'une part, avec l'exon EDA de la fibronectine précédé de ses séquences introniques d'autre part. Dans ce modèle l'ESE naturel est remplacé par des séquences sélectionnées par les divers facteurs SR par la technique du SELEX.

(ii) l'exon K-SAM du récepteur du facteur de croissance des fibroblastes-2 (FGFR2), précédé de l'exon situé juste en aval. L'exon K-SAM possède un élément inhibiteur pour la réaction d'épissage, et celui-ci a été délété et remplacé par des séquences exoniques de taille similaire. Les séquences SELEX sont insérées au sein de cet exon.

Nous avons suivi l'activation de l'épissage de l'intron de chacun de ces modèles par la présence

des séquences SELEX, afin de vérifier si celles-ci peuvent agir comme des activateurs exoniques d'épissage. Ainsi, nous avons démontré que les séquences sélectionnées par 9G8 et ASF/SF2 sont effectivement capables d'activer efficacement l'épissage d'un intron bordé par un site 3' faible. Cependant, les séquences sélectionnées par le facteur SC35 ne sont pas dotées de cette aptitude. Notons que des résultats similaires ont été obtenus avec un autre modèle d'épissage alternatif pour les facteurs ASF/SF2 et SC35, indiquant que vraisemblablement, SC35 agit sur l'épissage alternatif selon un mode sensiblement différent des autres facteurs SR de 35 kDa (ASF/SF2 et 9G8). De plus, l'activation de la réaction d'épissage observée par des séquences cibles de 9G8 et ASF/SF2 peut être augmentée par l'addition de ces protéines recombinantes, démontrant clairement une stimulation directe, et que ces facteurs constituent des déterminants de l'activation de l'épissage alternatif via des séquences cibles présentes dans les exons faiblement reconnus.

## **RESUME DES RESULTATS OBTENUS AU COURS DE L'ANNÉE** **SUIVANT MA THESE**

Au cours de l'étude des produits d'expression du gène 9G8 dans divers tissus adultes ou embryonnaires humains, nous avons montré que ce gène est soumis à un épissage alternatif de l'intron 3 (voir paragraphe 1.2). Cette rétention alternative de l'intron 3 semble physiologiquement importante à plus d'un titre:

- dans certains tissus les mRNA ayant retenu l'intron 3 peuvent représenter jusqu'à 60-70 % des messagers totaux et ce taux varie considérablement selon les tissus.

- la nature des signaux d'épissage bordant cet intron indique que ce phénomène n'est pas fortuit.

- la persistance de l'intron 3 dans le messager conduit à la synthèse d'une protéine dépourvue de son domaine SR. Parmi les autres membres de la famille des facteurs SR, il a été montré que le facteur ASF/SF2, débarassé de son domaine SR, conserve ses capacités à moduler l'utilisation de sites d'épissage alternatifs, mais perd son activité pour la réaction d'épissage constitutive.

Sur la base de ces résultats et observations, il était donc possible de postuler que l'épissage alternatif du gène 9G8, permettant à la cellule de moduler le taux de protéine "sauvage" par rapport à celui de la protéine "tronquée", représente un évènement biologiquement important.

Ceci nous a conduit à aborder cette thématique par deux approches qui visent à:

- (i) démontrer que la modulation de l'épissage alternatif au sein de l'intron 3 est basée sur la reconnaissance d'éléments en *cis* présents sur le pré-mRNA par des éléments agissant en *trans*. Certains motifs exoniques impliqués dans la modulation de cet épissage ont déjà été identifiés et les facteurs agissant en *trans* sont en cours de caractérisation.

- (ii) démontrer que cet évènement est également présent dans d'autres espèces. En effet la conservation d'un mécanisme alternatif au cours de l'évolution tend à confirmer l'importance de son rôle.

Ainsi, nous avons effectué une analyse comparative du gène 9G8 et de ses produits d'expression

chez la souris et chez l'homme. Par la suite, ce gène a été caractérisé dans d'autres espèces telles que le poulet, le poisson zèbre, le xénope, la drosophile et même chez *Arabidopsis thaliana*. La première étape de cette étude consiste en la caractérisation de l'ADNc codant pour la protéine 9G8 murine. Pour cela, une approche par PCR sur différentes banques d'ADNc a été choisie et m'a permis d'isoler le cDNA codant pour ce facteur. De la même manière, une amplification d'ADN génomique de souris par PCR m'a permis d'isoler trois introns, dont l'intron 3, chez cette espèce. La seconde étape de mon travail a consisté en l'étude comparative des produits d'expression du gène 9G8 dans les espèces considérées. Cette analyse sera effectuée sur différents tissus murins à la fois de manière qualitative, par RT-PCR, que semi quantitative, par la technique de Northern blot. Le choix de l'espèce murine est intéressant puisque nous disposons d'un panel de tissus (adultes ou embryonnaires), et de banques de cDNA beaucoup plus larges que pour l'homme et il permet de rendre ce type d'étude plus exhaustif. De plus, la bonne connaissance du gène murin et de son expression nous permet d'envisager la destruction sélective du gène 9G8 chez la souris dans un avenir proche.

Nous avons ainsi montré que les séquences composant l'intron 3 du gène 9G8 sont très conservées au cours de l'évolution, indiquant fortement leur importance physiologique. En effet, les séquences des deux autres introns sont nettement moins conservées. Nous avons ainsi pu délimiter des zones qui semblent importantes pour la régulation de l'activité du gène 9G8. Par ailleurs, diverses autres isoformes d'épissage alternatif ont pu être caractérisées à la fois chez l'homme que chez la souris. Un rôle putatif pour ces isoformes n'a pourtant pas pu être démontré.

## BIBLIOGRAPHIE

- (1) Moore M.J., Query C.C et Sharp P.A. (1993) *Splicing of precursor to messenger RNAs by the spliceosome - In the RNA world*. Gesteland R.F. et Atkins J.F., eds (Cold Spring Harbor Laboratory Press) 303-358
- (2) Green M.R. (1991) *Biochemical mechanisms of constitutive and regulated pre-mRNA splicing*. Annu. Rev. Cell. Biol., **7**, 559-599
- (3) Zamore P.D., Patton J.G. et Green M.R. (1992) *Cloning and domain structure of the mammalian splicing factor U2AF*. Nature, **355**, 605-614
- (4) Patton J.G., Porro E.B., Galceran J. Tempst P. et Nadal-Ginard B. (1993) *Cloning and characterization of PSF, a novel pre-mRNA splicing factor*. Genes & Dev., **7**, 393-406
- (5) Zahler A.M., Lane W.S., Stolk J.A. et Roth M.B. (1992) *SR proteins: a conserved family of pre-mRNA splicing factors*. Genes Dev., **6**, 837-847
- (6) Zahler A.M., Neugebauer K.M., Stolk J.A. et Roth M.B. (1993a) *Human SR proteins and isolation of a cDNA encoding SRp75*. Mol. Cell. Biol., **13**, 4023-4028
- (7) Roth M.B., Zahler A.M. et Stolk J.A. (1991) *A conserved family of nuclear phosphoproteins localized to sites of polymerase II transcription*. J. Cell. Biol., **115**, 587-596
- (8) Krainer A.R., Mayeda A., Kozak D. et Binns G. (1991) *Functional expression of cloned human splicing factor SF2: homology to RNA binding proteins U1 70K and drosophila splicing regulators*. Cell, **66**, 383-394
- (9) Ge H., Zuo P. et Manley J.L. (1991) *Primary structure of the human splicing factor ASF reveals similarities with drosophila regulators*. Cell, **66**, 373-382
- (10) Fu X.D. et Maniatis T. (1992) *Isolation of a complementary DNA that encodes the mammalian splicing factor SC35*. Science, **256**, 535-538
- (11) Vellard M., Sureau A., Soret J., Martinerie C. et Perbal B. (1992) *A potential splicing factor is encoded by the opposite strand of the trans-spliced c-myc exon*. Proc. Natl. Acad. Sci. USA, **89**, 2511-2515
- (12) Foulkes N.S., Borelli E. et Sassone-Corsi P. (1991) *CREM gene: Use of alternative DNA binding domains generates multiple antagonists of cAMP-induced transcription*. Cell, **64**, 739-749
- (13) Foulkes N.S. et Sassone-Corsi P. (1992) *More is better: activators and repressors from the same gene*. Cell, **68**, 411-414
- (14) Zahler A.M., Neugebauer K.M., Lane W.S. et Roth M.B. (1993b) *Distinct functions of SR proteins in alternative pre-mRNA splicing*. Science, **260**, 219-222.

- (15) Cáceres J.F., Stamm S., Helfman D.M. et Krainer A.R. (1994) *Regulation of alternative splicing in vivo by overexpression of antagonistic splicing factors*. Nature, **265**, 1706-1709
- (16) Cáceres J.F. et Krainer A.R. (1993) *Functional analysis of pre-mRNA splicing factor SF2/ASF structural domains*. EMBO J., **12**, 4715-4726
- (17) Zuo, P. et Manley J.L. (1993) *Functional domains of the human splicing factor ASF/SF2*. EMBO J., **12**, 4727-4737
- (18) Fu X.D. (1993) *Specific commitment of different pre-mRNAs to splicing by single SR proteins*. Nature, **365**, 82-85
- (19) Sun Q., Mayeda A., Hampson R.K., Krainer A.R. et Rottman F.M. (1993) *General splicing factor SF2/ASF promotes alternative splicing by binding to an exonic splicing enhancer*. Genes Dev., **7**, 2598-2608
- (20) Tuerk C. et Gold L. (1990) *Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase*. Science, **249**, 505-510.
- (21) Tacke R. et Manley J.L. (1995) *The human splicing factors ASF/SF2 and SC35 possess different, functionally significant RNA binding specificities*. EMBO J., **14**, 3540-3551
- (22) Lavigne A., La Branche H., Kornblihtt A.R. et Chabot B. (1993) *A splicing enhancer in the human fibronectin alternate ED1 exon interacts with SR proteins and stimulates U2 snRNP binding*. Genes Dev., **7**, 2405-2417
- (23) Wang Z., Hoffmann H.M. et Grabowski P.J. (1995) *Intrinsic U2AF binding is modulated by exon enhancer signals in parallel with changes in splicing activity*. RNA, **1**, 21-35

# Résumé des travaux effectués à l'UPR 41 CNRS Rennes [depuis Fev 1997]

## INTRODUCTION

### 1. Besoins de nouvelles approches de cartographie

L'émergence de très nombreux projets de séquençage de génomes et notamment du génome humain nécessite inmanquablement le développement de nouvelles méthodes rapides et précises de cartographie afin d'effectuer l'assemblage des données brutes issues du séquençage systématique. De plus, si comme la tendance actuelle de procéder par des shotguns de génomes entiers se confirme, il sera nécessaire de disposer de cartes génomiques détaillées.

Plus d'une douzaine de génomes microbiens ont d'ors et déjà été séquencés entièrement, suite à un shotgun direct de leur génome (<http://www.tigr.org/tdb/mdb/mdb.html>). Cependant, ceux-ci présentent des tailles allant de 580 kb à 4 Mb. L'utilisation de cet approche dans le but de séquencer de plus grands génomes tels que le génome humain (3 000 Mb), comme cela a été annoncé (1) suscite néanmoins certaines interrogations. En effet, la taille importante de ces génomes additionnée à la présence de nombreuses séquences répétées ajoute une complexité à l'assemblage des résultats de séquence. Le soutien d'une carte solide aiderait grandement au traitement de ces données.

L'établissement de cartes génomiques détaillées peut être également d'un grand secours pour des études phylogéniques. En effet, si l'on a longtemps étudié l'évolution des génomes par l'analyse comparée de certaines séquences géniques, il semble désormais clair que l'expression de nombreux gènes dépend de leur localisation dans un certain contexte génétique. Les développements récents de la génomique permettent maintenant d'étudier plus finement l'évolution des génomes en se fondant sur les changements de relation de synténie. Dès lors que des cartes détaillées peuvent être obtenues pour deux espèces quelconques, il devient possible d'avoir une vision plus claire des liaisons génétiques qui se sont maintenues entre ces deux espèces au cours de l'évolution.

Un autre domaine pour lequel l'apport de cartes génétiques détaillées serait d'un intérêt capital est la localisation de QTL (*Quantitative Trait Loci*). En effet, la plupart des variations au sein d'une population ou entre des races différentes, par exemple, est de nature quantitative. C'est à dire que les variations ne rentrent pas dans des classes bien définies suivant les proportions Mendéliennes, mais plutôt s'opèrent de manière continue selon un gradient d'une extrême à l'autre. Pour exemple, nous pouvons citer la taille, le poids des individus, ou encore la date de floraison chez les plantes ou bien la quantité de lait produite chez les mammifères *etc...* Dès lors que ces variations sont conservées et donc que la descendance tend à ressembler à ses parents, il y a une évidence génétique de transmission. Les loci impliqués dans la variation de traits phénotypiques quantitatifs sont appelés QTL. La détection de liaisons entre un QTL et des marqueurs génétiques apportent une méthode robuste pour l'identification de ces QTL. Il est possible de localiser un QTL par la méthode dite 'Interval mapping'

(2) entre deux marqueurs informatifs séparés de plus de 20 cM. Dès lors que cet espace est restreint, il est d'un grand secours de pouvoir effectuer un zoom sur la région d'intérêt, et donc une cartographie fine, afin de localiser précisément le gène recherché.

## **2. Analyse comparée des génomes de primates supérieurs**

Les relations phylogéniques entre les grandes espèces de primates (homme, chimpanzé, gorille et orang-outan) sont toujours l'objet de nombreux débats, d'autant plus que la manière dont ont pu se produire les spéciations successives n'est pas encore définitivement établie. L'origine de l'homme et des grands singes a été particulièrement étudiée par analyse caryotypique (3). Les hominidés présentent des caryotypes cytogénétiques très voisins et la comparaison du génome humain avec ceux des autres primates nous informe sur les modifications survenues au cours des spéciations. L'actuelle connaissance du génome des différents singes anthropoïdes montre une conservation des divers gènes analysés supérieure à 98%, comparativement au génome humain. Toutefois, ce pourcentage élevé reflète la conservation globale des séquences et notamment de certains gènes orthologues mais ne traduit pas la représentation du génome dans son intégralité. En effet, elle n'inclut pas la notion de synténie et de rupture de synténie et explique *a priori* difficilement les différences phénotypiques entre ces espèces.

Au cours de ces dix dernières années, le développement de la technique de l'hybridation chromosomique suppressive (CISS, pour *Chromosomal In Situ Suppression*; 4, 5) a été utilisé pour établir des homologies entre les chromosomes humains et ceux des autres primates (6-9). Les différences les plus notables entre les génomes des membres de la famille des Hominidés concernent des variations dans la répartition de l'hétérochromatine ainsi qu'un certain nombre d'inversions péricentriques (10). A ce jour, aucune translocation, autre que celle bien connue qui a donné naissance au chromosome 2 humain par fusion des chromosomes 12 et 13 trouvés chez les chimpanzés ou que la translocation 5;17 chez le gorille n'a été décelée chez les grands singes, par rapport à l'homme (6). Ces dernières années, plusieurs gènes ont été cartographiés chez les grands singes, notamment grâce à l'utilisation de l'hybridation *in situ* par fluorescence (FISH, pour *Fluorescent In Situ Hybridization*) (11-13) ajoutant des informations précieuses sur la synténie. Par ailleurs, si ces techniques ont permis de mettre en évidence des régions sur les chromosomes de chimpanzé, hybridant avec du DNA humain (régions convergentes), elles ont également montré des régions non hybridées (régions divergentes) qui ont été attribuées à un excès d'environ 10% de DNA génomique pour le chimpanzé par rapport à l'homme (7). Toutefois, toutes ces approches n'ont en général pas permis de détecter d'éventuels remaniements intrachromosomiques, dont on pense qu'ils ont joué un rôle important lors de l'évolution de l'homme et des grands singes.

Les progrès récents de la cartographie génétique des génomes permettent maintenant d'étudier plus finement l'évolution de ceux-ci en se fondant sur les changements de relation de synténie. Dès lors que des cartes détaillées peuvent être obtenues pour deux espèces quelconques, il devient possible d'avoir une vision plus claire des liaisons génétiques qui se sont maintenues entre ces deux espèces au cours de l'évolution. Jusqu'à présent, de tels travaux de cartographie comparée ont été conduits principalement entre l'homme, la souris, les bovidés ou le porc (14-19).

Dans le cas des singes anthropoïdes, il n'existe pas de carte génétique détaillée. Cependant, la constitution d'une telle carte serait d'un très grand intérêt puisque les grands singes constituent les plus proches parents de l'homme. Ainsi, les cartes comparées de ces espèces apporteraient une information unique sur les événements récents de l'évolution qui ont permis d'organiser le génome humain. Si, le très grand nombre de marqueurs humains actuellement disponible (20-21) constitue un atout important qui devrait permettre de dresser une telle carte, l'analyse de liaisons génétiques au moyen d'études familiales telles que celles qui ont fait leurs preuves chez l'homme est impraticable.

La constitution de cartes génomiques chez les primates supérieurs devrait permettre de déterminer les zones de rupture de synténie avec une précision suffisante pour envisager raisonnablement le clonage ultérieur de la région et l'analyse de son anatomie moléculaire. Dans tous les cas où les relations de synténie sont modifiées, l'analyse des points de rupture permettra de préciser les phénomènes qui s'y sont produits, notamment s'il y a eu gain, perte ou bien modification de l'information.

### ***3. La méthode du HAPPY mapping***

#### 3.1 Choix de l'approche

Une approche de cartographie envisageable est la cartographie par hybrides d'irradiation (RH, pour *Radiation Hybrid mapping*) (23-24). Elle consiste à irradier des lignées cellulaires, ce qui provoque des cassures chromosomiques. Les différents fragments de chromosomes générés sont alors intégrés dans le génome de cellules de rongeur. Ainsi, il est possible de déterminer la distance séparant deux marqueurs sachant que plus ils sont proches, plus ils auront de chance d'être incorporés au sein d'un même fragment et donc d'être détectés dans une même lignée. Néanmoins, cette approche présente certains désavantages en plus de sa lourdeur dans la mise en place d'un panel d'hybrides d'irradiation. En effet, (i) certains loci qui ne se seraient pas clonés ne peuvent être intégrés dans une carte du génome; (ii) l'interprétation des résultats peut être confuse dans le cas où les inserts se sont réarrangés ou ligués entre eux; (iii) la présence de DNA exogène, en l'occurrence celui de la cellule hôte, de hamster, oblige bien souvent à écarter un nombre de marqueurs donnant une réponse positive avec cet ADN exogène. La méthode HAPPY, plus souple, permet de contourner ces différents problèmes (25).

La méthode HAPPY permet de reproduire *in vitro* les deux événements analysés par les croisements de la génétique formelle, c'est à dire le *crossing-over* et la ségrégation. En pratique, le *crossing-over* est mimé par une cassure aléatoire du DNA en fragments dont la taille dépend de la cartographie recherchée. Les marqueurs sont alors ségrégués par distribution aléatoire de ces fragments en dépôts de moins d'un équivalent de génome haploïde par aliquot, puis détectés par PCR. Ceux qui sont génétiquement liés tendent à demeurer ensemble dans le même aliquot tandis que ceux qui ne sont pas liés se distribuent de façon aléatoire. Leur ordre et la distance qui les sépare peuvent être déduits de la fréquence de leur co-ségrégation par le calcul. Il est important de rappeler ici qu'un panel utilisable pour la cartographie HAPPY est simple à produire puisqu'il ne requiert que quelques jours voire quelques semaines. En outre, il est adaptable à n'importe quel niveau de résolution, selon la taille des fragments sélectionnés et peut même aboutir au clonage moléculaire des fragments d'intérêt en vue d'un séquençage.

### 3.2 Description de la méthode (Figure 1)

Des cellules, récoltées chez le sujet à étudier [1] sont encapsulées dans des blocs d'agarose, puis lysées afin de libérer le DNA intact. Une partie du DNA est ensuite cassée de manière aléatoire (par irradiation  $\gamma$ , digestion enzymatique, etc...) pour obtenir des fragments dont la taille dépend de la résolution choisie [2]. Le DNA est alors migré par électrophorèse en champs pulsés (PFGE, pour *Pulsed Field Gel Electrophoresis*) pour séparer des fragments de taille homogène [3] qui sont répartis dans les 96 puits d'une plaque de microtitration de façon à n'avoir qu'une quantité voisine d'un équivalent génome haploïde de DNA par puits [4]. De cette façon, chaque puits a une représentation incomplète du génome qui lui est propre. L'intégralité de l'information génétique contenue dans chaque puits est alors amplifiée par PCR.

Chaque puits contient maintenant une grande quantité de DNA ce qui permet de distribuer ce matériel dans de nombreuses plaques filles [5]. Ainsi chaque plaque fille sera utilisée pour déterminer la distribution d'un marqueur au moyen de la PCR en utilisant comme amorces des oligonucléotides spécifiques du marqueur en question [6]. Les produits de PCR sont ensuite analysés en électrophorèse classique sur gel d'agarose. Cette opération est répétée pour un nombre élevé de marqueurs.

### 3.3 Analyse des résultats

Dans le cas où un équivalent de génome est réparti par puits, chaque marqueur sera retrouvé, d'après la loi de Poisson dans 65% des puits. Ainsi, deux marqueurs non liés (séparés par une distance supérieure à la taille moyenne des fragments aliquotés) vont se distribuer de façon aléatoire et se retrouver ensemble dans 42% des puits ( $65\% \times 65\%$ ). A l'inverse, deux marqueurs séparés par une distance inférieure à cette taille moyenne ségrégeront d'autant plus fréquemment ensemble qu'ils seront proches l'un de l'autre, jusqu'à se retrouver ensemble dans 65% des aliquots dans le cas de marqueurs très étroitement liés. Chaque occurrence est enregistrée et les liaisons entre les marqueurs sont calculées (lod score) comme en génétique classique à l'aide de logiciels informatiques tels que RHmap (26) ou RHmapper (27) [7]. La fréquence d'association entre deux marqueurs, traduisant la distance physique qui les sépare sur le chromosome, permet alors de déduire une carte regroupant tous les marqueurs associés, deux à deux, au sein d'un même groupe de liaison [8].

Cette méthode de cartographie a déjà fait ses preuves pour des génomes aussi variés en taille que le chromosome 14 humain —~100 Mb— (28) ou celui d'un protozoaire parasite de l'épithélium intestinal de nombreux mammifères, *Cryptosporidium parvum* —10 Mb— (29). Toutefois, pour ces deux études, aucune méthode d'amplification du panel initial, permettant de cartographier un nombre illimité de marqueurs n'a été décrite. Dans Dear *et al.* (28), seulement une faible partie de l'ADN total, encadrée par des séquences répétées, a pu être cartographiée. Dans Piper *et al.* (29), le niveau d'amplification n'était pas suffisant pour permettre la détection directe par PCR des marqueurs. Ces auteurs ont du procéder par PCR nichée afin de visualiser les marqueurs à cartographier. De plus, la méthode d'amplification utilisée ne permet que de cartographier un nombre limité de marqueur, nécessitant de reconstruire un panel de cartographie pour localiser des marqueurs supplémentaires.

Afin d'être envisageable pour la cartographie HAPPY, la méthode d'amplification doit répondre à trois critères stricts (*i*) une quantité d'ADN voisine d'un équivalent de génome haploïde doit être

suffisante comme matrice; (ii) l'intégralité de l'information génétique doit être amplifiée; (iii) le panel constitué doit pouvoir être réamplifié à l'infini afin de permettre la cartographie d'un nombre illimité de marqueurs. Nous décrivons ici une approche d'amplification totale, homogène et à l'infini de l'ADN génomique. Nous montrons que cette technique, que nous avons appelé *FGM*, pour "*Full Genome Mapping*" peut être utilisée pour la cartographie HAPPY.

## **RÉSULTATS**

### **1. Constitution du panel de cartographie**

La détermination des conditions d'amplification par PCR du contenu génomique de chaque puits de la microplaque mère est prépondérante pour la cartographie HAPPY. Pour cette analyse, nous avons choisi aléatoirement un marqueur sur chaque chromosome humain. Si l'on part d'un équivalent génome haploïde, chaque marqueur doit être présent dans 65% des puits analysés, d'après la loi statistique de Poisson (25). Nous avons vérifié, par PCR nichée, que chaque marqueur est effectivement distribué de cette manière lors d'une répartition d'un équivalent génome haploïde dans les 96 puits d'une microplaque. Dans les conditions que nous avons déterminées, nous avons vérifié que tous ces marqueurs présentent la même distribution après la première phase d'amplification. La seconde phase d'amplification s'effectue de la même manière sans perte apparente d'information.

### **2. Création d'une carte d'une portion du chromosome 2 humain**

Dans un premier temps, 190 marqueurs humains ont été sélectionnés dans les banques de données sur une zone couvrant environ 60 Mégabases du chromosome 2 humain. Chacun de ces marqueurs a été testé par PCR sur les puits de la microplaque constituant notre panel de cartographie. La présence/absence de ces marqueurs est visualisée par PCR suivie d'une électrophorèse et coloration au BET. Parmi ces 190 marqueurs, 176 ont été intégrés à la carte (> 92%; figure 2). Les autres, en raison d'un taux de rétention s'éloignant trop de la norme, ont été écartés. En effet, seuls sont gardés les marqueurs dont le nombre de positifs est de  $p \pm 2\sqrt{(p - [p^2/n])}$  où  $p$  est le nombre moyen de positifs par marqueur et  $n$  le nombre d'aliquots constituant le panel. L'analyse statistique de la répartition des marqueurs s'est effectuée à l'aide de la suite logicielle RHMap (26). Dans un premier temps, un premier regroupement de marqueurs a été opéré par une analyse 'deux à deux' en fixant un LOD score seuil minimal de 8 pour visualiser les associations entre marqueurs. Les marqueurs ont été placés par rapport à leurs voisins par analyse multipoint au sein de chaque groupe de liaison généré à un LOD score de 8. Par la suite, les groupes ont été disposés les uns par rapport aux autres en choisissant un LOD score de 4.

Parmi les 176 marqueurs ainsi cartographiés, 133 se sont placés sans ambiguïté les uns par rapport aux autres (>75%; figure 2). Toutefois, 43 autres marqueurs (partie gauche de la carte; figure 2), montrant une association suffisamment forte avec un ou plusieurs marqueurs déjà localisés pour être acceptée comme réelle, n'ont pas pu être placés de manière formelle en raison d'une moindre

association avec les marqueurs situés dans l'entourage immédiat. Pour cette raison, ils ont été indiqués par des barres flottantes.

Notons également que si la carte a été réalisée avec 176 marqueurs (densité de 1 marqueur/340 kb), des résultats similaires auraient été obtenus en n'utilisant que 75 marqueurs judicieusement choisis ce qui porte la densité minimale à 1 marqueur/800 kb).

### **3. Comparaison de la carte HAPPY avec les autres cartes disponibles**

Afin de valider notre carte, nous l'avons comparée aux données disponibles dans les banques informatiques. Actuellement, de nombreuses données issues de la cartographie par hybrides d'irradiation ou par études familiales sont disponibles. Nous avons choisi de comparer en priorité nos résultats avec ceux obtenus par hybrides d'irradiation sur le panel de cartographie G3 qui a été obtenu d'une irradiation cellulaire à 10000 Rad (30). La taille moyenne des fragments obtenus est d'environ 4 Mb, ce qui est très proche de la taille de fragments que nous avons sélectionné pour la constitution de notre panel (5,5 Mb). 48 marqueurs communs ont été placés sur les deux cartes (figure 3). Nous observons que l'ordre général de ces marqueurs est respecté dans ces deux cartes, puisque seules 4 inversions ont été mises en évidence. Ces inversions deux à deux ne sont observées que sur des marqueurs proches sur le génome (distance estimée inférieure à 500 kb, ce qui est proche de la résolution moyenne des cartes générées avec ce panel (<http://www-shgc.stanford.edu>) soit plus de 300 kb). Dans la région située immédiatement au dessus du centromère, trois marqueurs présentent une position relativement divergente. Ceci peut être mis en relation avec le fait que la cartographie par hybrides d'irradiation peut entraîner des biais dans la résolution des régions entourant le centromère. Ce fait peut aussi expliquer les différences de distances séparant les marqueurs dans ces régions (figure 3). Notons également que la position des marqueurs que nous avons obtenu est en accord avec la carte génétique disponible dans les banques de données, si l'on compare la position de nos marqueurs par rapport à 13 marqueurs polymorphes assignés à cette portion de chromosome (figure 3).

## **DISCUSSION**

### **1. Développement d'une nouvelle approche de cartographie**

Nous montrons donc ici une nouvelle façon d'aborder l'amplification *ad infinitum* d'une quantité faible d'ADN génomique —environ 1 équivalent génome haploïde, soit 2 pg dans le cas du génome humain—nécessaire à la constitution d'un panel utilisable pour la cartographie HAPPY. La technique mise au point, réalisée en deux étapes de PCR permet une amplification de plus de 90% du contenu génétique. De plus, cette approche nous a permis de cartographier un nombre important de marqueurs sur une portion du chromosome 2 humain. Nos résultats sont en accord avec toutes les données décrites dans la littérature.

L'étape la plus critique dans la technique du HAPPY mapping est l'étape d'amplification totale et homogène du génome. Paul Dear, qui a inventé cette technique s'est heurté à ce problème et l'a contourné en amplifiant le génome à l'aide de PCR utilisant des amorces ancrées sur des éléments répétés tout au long du génome (Alu, LINE) (IRS-PCR, pour *Interspersed Repetitive Sequence-PCR*). Ceci lui a permis de positionner un millier de marqueurs sur le chromosome 14 humain (28). Toutefois, la limite de l'IRS-PCR est de ne permettre d'amplifier et donc de cartographier que de marqueurs proches de séquences répétées et encadrés par elles. Ce type d'amplification est par conséquent restreint à une partie du génome (puisque seules les séquences encadrées par deux séquences alu ou LINE sont amplifiées) et ne permet de cartographier que certains STS et peu de gènes. De plus, il est nécessaire par cette approche de définir des conditions d'amplifications spécifiques de chaque organisme étudié. Afin d'être utilisable pour la cartographie HAPPY, la méthode d'amplification doit être à la fois quantitativement *i.e.* amplifier à un niveau suffisant une quantité faible d'ADN initial (un équivalent génome haploïde) et qualitativement *i.e.* s'effectuer sans perte notable de l'information génétique, acceptable pour être retenue.

Pour cela, diverses approches ont été envisagées. Les trois premières, issues de la littérature (31-33), n'ont apporté que peu de résultats. La méthode décrite par Telenius *et al* (31) utilise un oligonucléotide dégénéré sur 6 positions pour une amplification par PCR réalisée en une seule étape (DOP-PCR, pour *Degenerated Oligonucleotide Primed-PCR*). Si cette méthode a donné des résultats satisfaisants pour ces auteurs, il est important de rappeler que ceux-ci utilisaient pour matériel de départ 100 ng d'ADN génomique humain, soit 50000 équivalents génome haploïde environ ou 500 copies d'un chromosome isolé. Rappelons que la cartographie HAPPY ne s'effectue qu'à partir d'échantillons d'ADN dont la quantité est voisine d'un équivalent génome haploïde. En utilisant les conditions décrites par Telenius *et al.* nous n'avons jamais dépassé une amplification de 50% du matériel génétique initial. Ceci est en accord avec les résultats de Cheung et Nelson (34) qui montrent que plus la quantité d'ADN génomique utilisée comme matrice est faible, en dessous d'un seuil de 600 pg —soit 300 équivalents génome haploïde—, plus le risque de ne pas amplifier un locus est grand.

La seconde méthode décrite par Zhang *et al* (32) utilise un oligonucléotide entièrement dégénéré sur ses 15 positions (PEP, pour *Primer Extension Preamplification*). Ces auteurs ont évalué pouvoir amplifier, à partir du DNA contenu dans un spermatozoïde —soit un génome haploïde—, 80 % des séquences au moins 30 fois. En variant les conditions de PCR utilisées par Zhang *et al.* (nombre de cycles, températures d'hybridation, concentration en primer), nous sommes arrivés à amplifier près de 90 % de l'ADN matrice à un minimum de 200 copies. Toutefois, cette approche ne permet pas de réamplification supplémentaire et limite donc grandement le nombre de marqueurs pouvant être cartographiés. De plus, le niveau d'amplification ne permet pas une visualisation directe du produit d'amplification d'un marqueur donné mais oblige l'utilisation de deux couples d'amorces, le second étant niché dans le premier, et donc d'effectuer une amplification en deux étapes.

Une autre approche, très voisine de celle que nous avons développée, s'effectue en deux étapes et s'appuie sur un oligonucléotide dégénéré sur les 9 résidus situés dans sa partie 3' et comportant une séquence fixe en 5' utilisée pour la seconde étape d'amplification (T-PCR, pour *Tagged-PCR*; 33). Toutefois, si cette approche a été montrée efficace de façon qualitative et quantitative —en terme de

quantité initiale d'ADN à amplifier— ces deux aspects, cruciaux pour la cartographie HAPPY, n'ont jamais été montrés pour une même expérience. En effet, lors de la démonstration “qualitative”, aucune indication sur la quantité de matériel initial utilisé n'est fournie. De plus, Grothues *et al.* montrent une amplification d'1 pg d'ADN génomique démontrant ainsi la possibilité d'amplifier une très faible quantité d'ADN, toutefois, aucun aspect qualitatif de l'amplification n'est ici mis en évidence. Par ailleurs, il est clairement libellé dans le brevet US5,731,171 que la limite inférieure d'amplification a été de 17 équivalents génome haploïde. De surcroît, cette approche s'appuie sur l'utilisation de deux enzymes différentes, la première étant une polymérase thermosensible alors que la seconde est thermostable et utilisée pour la PCR. Un autre inconvénient non négligeable dans la méthode décrite dans US5,731,171 est le risque de contamination par de l'ADN étranger. En effet, lors de l'étape d'amplification à faible stringence, chaque cycle nécessite l'introduction ‘manuelle’ de DNA polymérase. Le nombre de cycle devant être accru pour des quantités très faibles d'ADN, le risque de contamination s'en trouve d'autant augmenté. Des conditions proches de celles “succinctement” décrites par Grothues *et al.* ont été employées. Nous nous sommes aperçu que la première étape d'amplification est très sensible à de nombreux paramètres tels que la concentration en primer, le nombre de cycles de PCR, la température d'hybridation, ainsi que la longueur de la chaîne dégénérée en 3'. Le choix de la séquence fixée en 5' de l'oligonucléotide est quant à lui très important pour la seconde étape d'amplification. Cette seconde phase est elle aussi très dépendante de la quantité de *primer* utilisé et du nombre de cycles de PCR effectués.

En conclusion, notre approche *Full Genome Mapping* est à ce jour la seule décrite permettant à la fois d'amplifier à l'infini des quantités très faibles d'ADN sans occasionner de perte notable de séquence. La cartographie présente un intérêt indéniable par rapport aux autres techniques de cartographie précédemment décrites. En effet, elle permet l'utilisation de tout type de marqueurs —qu'ils soient polymorphes ou non— et elle n'est pas perturbée par la présence d'ADN étranger, comme c'est le cas pour la cartographie par hybrides d'irradiation. Il est important de rappeler ici qu'un panel utilisable pour la cartographie HAPPY est simple à produire puisqu'il ne requiert que quelques jours voire quelques semaines. En outre, il est adaptable à n'importe quel niveau de résolution, selon la taille des fragments générés et sélectionnés (coupure par irradiation  $\gamma$  plus ou moins forte, cassure mécanique *etc...*) et peut même aboutir au clonage moléculaire des fragments d'intérêt en vue d'un séquençage.

## **2. De l'intérêt de l'étude des génomes de pongidés**

En conclusion, ce projet débouche tout d'abord sur une problématique passionnante, celle des relations phylogéniques entre les grandes espèces de primates. L'établissement de cartes de synténie fines devrait développer le concept de l'évolution et de la structure des génomes, qui n'est pas nécessairement parallèle à l'évolution des gènes, et qui met en avant les problèmes de la cohérence d'un génome dans sa fonctionnalité.

Certaines affections génétiques ont été attribuées à la fragilité de certaines régions chromosomiques. La localisation précise des points de rupture de synténie chez les primates supérieurs peut favoriser la détection de zones de fragilité jusqu'à présent méconnues et nous apporter des éléments sur les mécanismes de remaniements.

L'utilisation des grands singes comme modèle d'étude, si elle a été relativement peu développée jusqu'à maintenant, peut apporter de nombreuses réponses à des questions encore non élucidées. En effet, les grands primates sont moins sensibles que les humains à certaines pathologies, telles que certaines formes de cancer, l'hépatite B, la malaria ou le SIDA. Ainsi, les faibles différences observées entre leurs génomes présentent un intérêt dépassant la simple étude de l'évolution. Une grande question émerge: "Lesquels de nos gènes font de nous des humains?"

La description de gènes dont la fonction pourrait varier entre les espèces de grands primates reste encore marginale. Une tactique utilisable afin d'identifier les différences fondamentales est l'étude des chromosomes et de leur structure. Il est bien connu maintenant que si les grands singes et l'homme présentent des caryotypes proches, un certain nombre de remaniements ont eu lieu depuis que ces espèces se sont différenciées. Ces remaniements impliquent une relocalisation de certains gènes dans un environnement génétique pouvant influencer sur leur expression. Par exemple, un fragment du chromosome 4 humain a subi chez le chimpanzé, une inversion péracentrique (10). Ce fragment du chromosome 4 contient, dans une région de la rupture de synténie, le gène codant un facteur de transcription, AF4, dont on sait que certaines mutations entraînent des formes aiguës de leucémie chez l'homme. Le fait que les chimpanzés soient plus résistants à certaines formes de cancer, notamment les leucémies, est peut être à mettre en relation avec ce remaniement chromosomique qui altérerait l'expression d'AF4 et ainsi les protégerait de la leucémie.

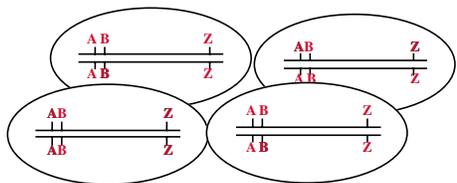
L'approche par *Full Genome Mapping*, telle que nous la décrivons, peut être d'un intérêt capital pour les efforts de cartographies préalables, nécessaires à la mise en place des projets de séquençage de génomes ou à l'assemblage de *shot-guns* si, comme la tendance actuelle l'indique, des *shot-guns* directs sont réalisés sur des génomes de plus en plus grands. De plus, sa résolution facilement adaptable en fait un outil de choix pour la cartographie fine de petits génomes.

## **BIBLIOGRAPHIE**

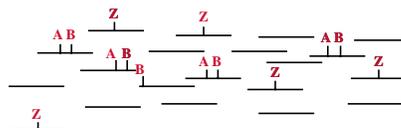
- (1) Weber JL et Myers EW (1997) *Human whole-genome shotgun sequencing*. Genome Res, **7**, 401-409
- (2) Lander ES et Botstein DB (1989) *Mapping mendelian factors underlying quantitative traits using RFLP linkage maps*. Genetics, **121**, 185-199
- (3) Dutrillaux B (1979) *Chromosomal evolution in primates: tentative phylogeny from Microcebus murinus (prosimian) to man*. Hum Genet, **48**, 251-314
- (4) Cremer T et al. (1988) *Detection of chromosome aberrations in metaphase and interphase tumor cells by in situ hybridization using chromosome-specific library probe*. Hum Genet, **80**, 235-246
- (5) Lichter P et al. (1988) *Delineation of individual human chromosomes in metaphase and interphase cells by in situ suppression hybridization using recombinant libraries*. Hum Genet, **80**, 224-234
- (6) Jauch A et al. (1992) *Reconstruction of genomic rearrangements in great apes and gibbons by chromosome painting*. Proc Natl Acad Sci USA, **89**, 8611-8615
- (7) Luke S et Verma RS (1993) *The genomic synteny at DNA level between human and chimpanzee chromosomes*. Chromosome Res, **1**, 215-219
- (8) Muller S et al. (1996) *Comparative fluorescence in situ hybridization mapping of primate chromosomes with Alu polymerase chain reaction generated probes from human/rodent somatic cell hybrids*. Chromosome Res, **4**, 38-42
- (9) Tarzami ST et al. (1997) *Unique genomic sequences in human chromosome 16p are conserved in the great apes*. Mol Gen Genet, **253**, 512-514
- (10) Nickerson E et Nelson DL (1998) *Molecular definition of pericentric inversion breakpoints occurring during the*

- evolution of humans and chimpanzees*. Genomics, **50**, 368-372
- (11) Ried T *et al.* (1993) *Comparative high resolution mapping of human and primate chromosomes by fluorescence in situ hybridization*. Genomics, **18**, 381-386
- (12) Arnold N *et al.* (1995) *Comparative mapping of DNA probes derived from the V K immunoglobulin gene regions on human and great ape chromosomes by fluorescence in situ hybridization*. Genomics, **26**, 147-150
- (13) Zimonjic DB *et al.* (1997) *Fluorescence in situ hybridization analysis of keratinocyte growth factor gene amplification and dispersion in evolution of great apes and human*. Proc Natl Acad Sci USA, **94**, 11461-11465.
- (14) Copeland NG *et al.* (1993) *A genetic linkage map of the mouse: current applications and future prospects*. Science, **262**, 57-66
- (15) Lundin LG (1993) *Evolution of the vertebrate genome as reflected in paralogous chromosomal regions in man and the house mouse*. Genomics, **16**, 1-19
- (16) Rettenberger G *et al.* (1995) *Visualization of the conservation of synteny between human and pigs by heterologous chromosome painting*. Genomics, **26**, 372-378
- (17) O'Brien SJ *et al.* (1993) *Anchored reference markers for comparative genome mapping in mammals*. Nature Genet, **3**, 103-112
- (18) Archibald AL *et al.* (1995) *The PiGMaP consortium linkage map of the pig (Sus scrofa)*. Mammalian Genome, **6**, 157-175
- (19) Creighton P *et al.* (1994) *A genetic linkage map of the bovine genome*. Nature Genet, **6**, 227-235
- (20) Schuler GD *et al.* (1996) *A gene map of the human genome*. Science, **274**, 540-546
- (21) Deloukas P *et al.* (1998) *A Physical Map of 30,000 Human Genes*. Science, **282**, 744-746
- (23) Gyapay G *et al.* (1996) *A radiation hybrid map of the human genome*. Hum Mol Genet, **5**, 339-346
- (24) Cox DR *et al.* (1990) *Radiation hybrid mapping: a somatic cell genetic method for constructing high-resolution maps of mammalian chromosomes*. Science, **250**, 245-250
- (25) Dear PH *et al.* (1993) *Happy mapping: linkage mapping using a physical analogue of meiosis*. Nucl Acids Res, **21**, 13-20
- (26) Boehnke M *et al.* (1991) *Statistical methods for multipoint radiation hybrid mapping*. Am J Hum Genet, **49**, 1174-1188
- (27) Slonim D *et al.* (1995) *RHMAPPER: An interactive computer package for constructing radiation hybrid maps*. Cold Spring Harbor Meeting on Genome Mapping and Sequencing, May 10-14, 1995.
- (28) Dear PH *et al.* (1998) *A high resolution metric HAPPY map of human chromosome 14*. Genomics, **48**, 232-241
- (29) Piper MB *et al.*, (1998) *A HAPPY map of Crypto sporidium parvum*. Genome research, **8**, 1299-1307
- (30) Stewart EA *et al.* (1997) *An STS-based radiation hybrid map of the human genome*. Genome Res, **7**, 422-433
- (31) Telenius H *et al.* (1992) *Degenerate Oligo-Primed PCR: General amplification of target DNA by a single degenerate primer*. Genomics, **13**, 718-725
- (32) Zhang L *et al.* (1992) *Whole genome amplification from a single cell: Implications for genetic analysis*. Proc. Natl. Acad. Sci. USA, **89**, 5847-5851
- (33) Grothues D *et al.* (1993) *PCR amplification of megabase DNA with tagged random primers (T-PCR)*. Nucleic Acids Res, **21**, 1321-1322
- (34) Cheung VG *et al.* (1996) *Whole genome amplification using a degenerate oligonucleotide primer allows hundreds of genotypes to be performed on less than one nanogram of genomic DNA*. Proc. Natl. Acad. Sci. USA, **93**, 14676-14679

### 1/ Inclusion des lymphocytes dans l'agarose



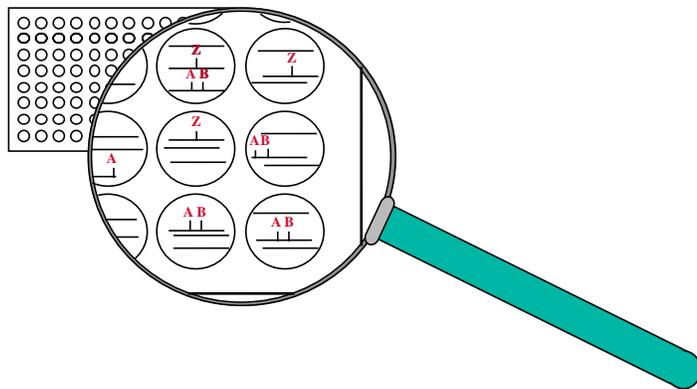
### 2/ Lyse cellulaire et Fragmentation de l'ADN



### 3/ Séparation des fragments d'ADN par électrophorèse en champs pulsés

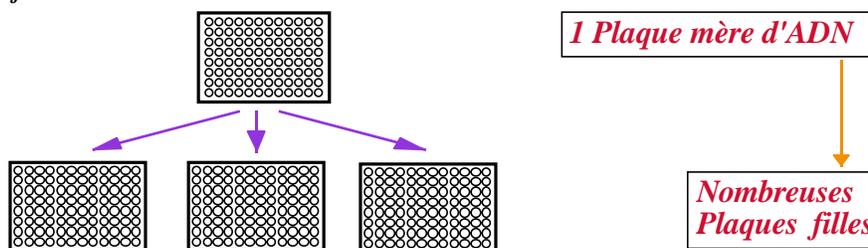


### 4/ Répartition de l'ADN sur plaque de microtitration (1 eq. génome haploïde / puits)

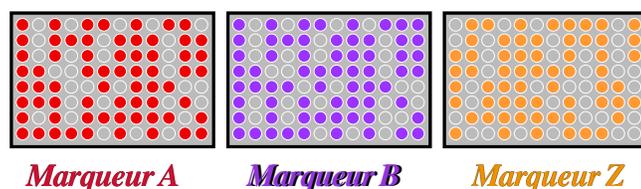


### 5/ Amplification de la plaque mère par PCR

Répartition en plaques filles



### 6/ Analyse de la répartition des marqueurs dans chaque puits de la microplaque



- PCR effectuée à l'aide d'amorces spécifiques  
- Analyse sur gel d'agarose

### 7/ Calcul de la fréquence d'association des marqueurs (LOD score)

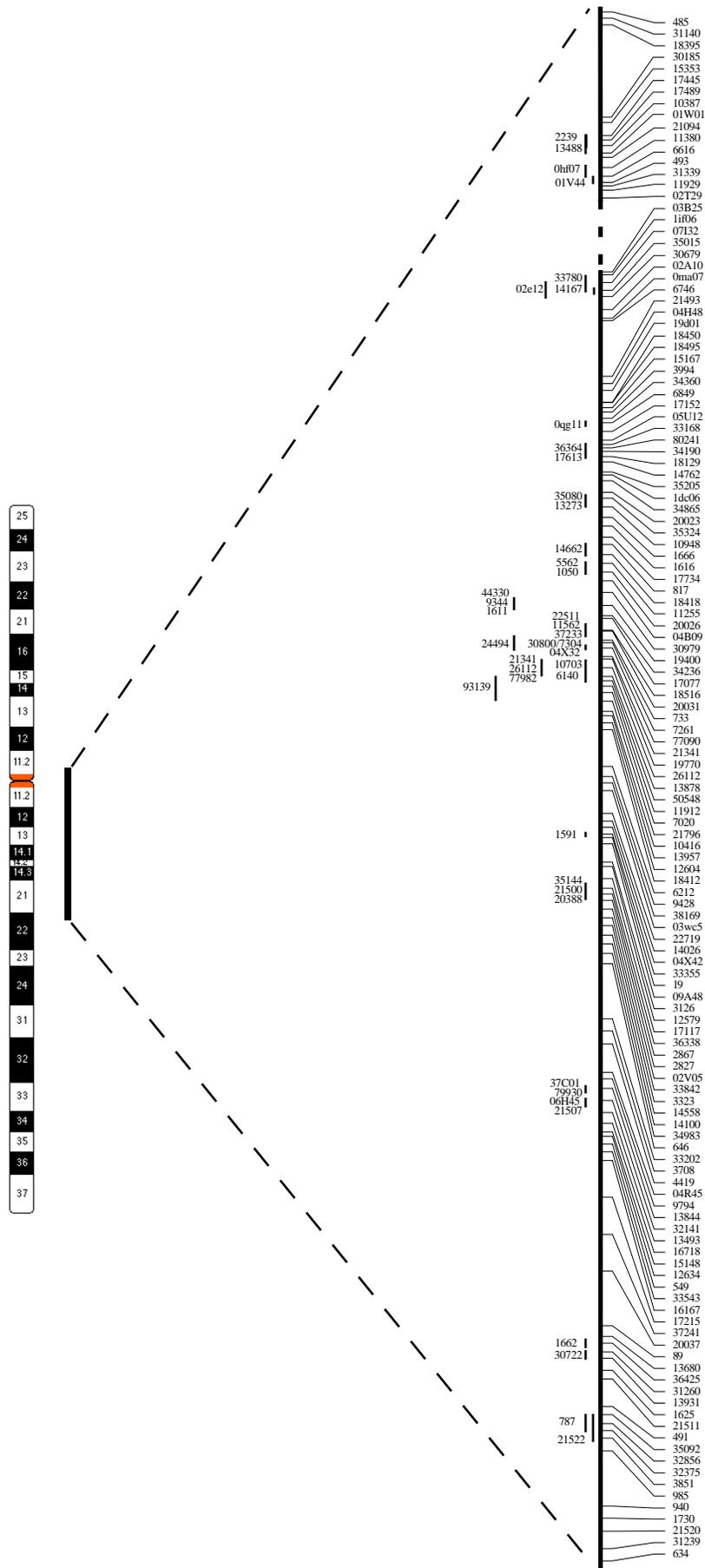
La fréquence d'association des marqueurs dans un même puits dépend de la distance qui les sépare:  
Deux marqueurs proches sont souvent présents ensemble  
- cas des marqueurs A et B  
Deux marqueurs éloignés sont associés aléatoirement  
- cas des marqueurs A et Z ou B et Z

### 8/ Déduction de la carte



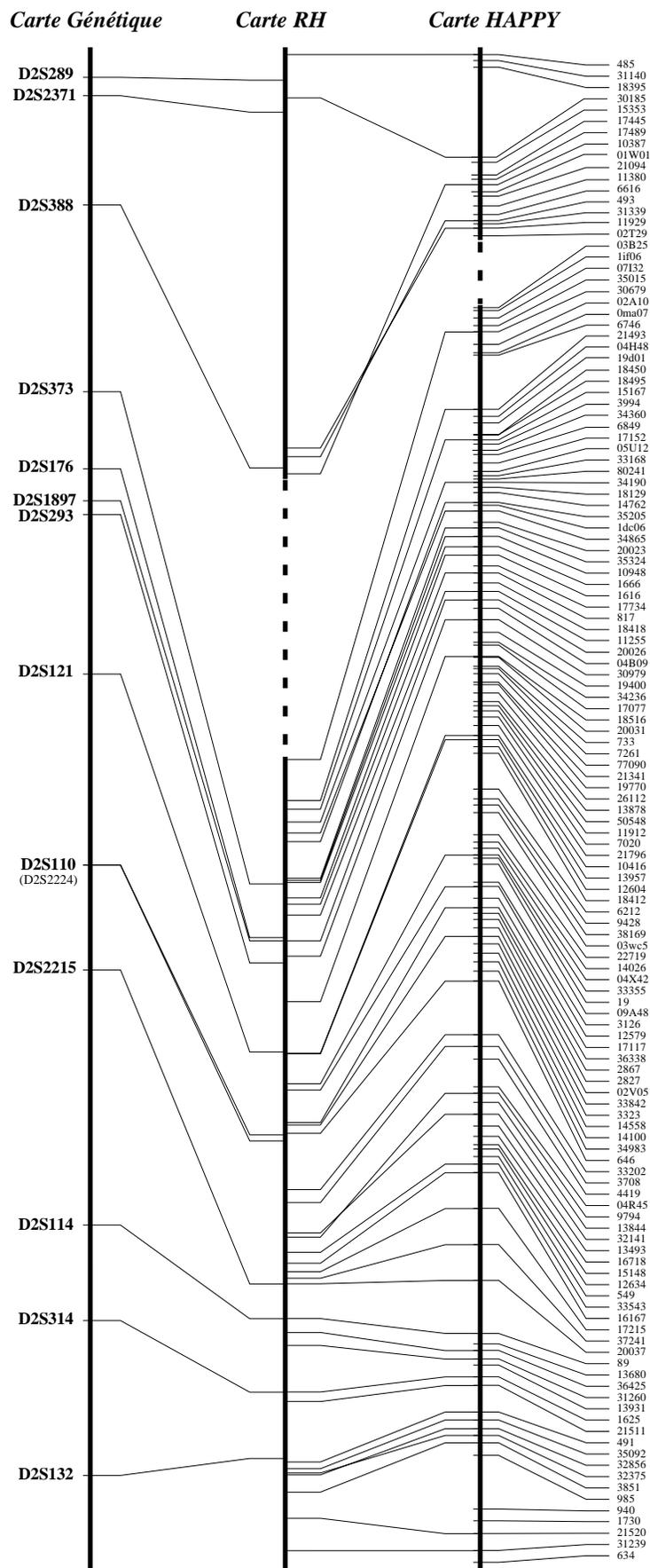
Figure 1: Principe de la cartographie HAPPY

Les différentes étapes de la cartographie HAPPY sont ici détaillées. L'ADN génomique est préparé à partir de lymphocytes circulant (1), fractionné (2) et les fragments sont sélectionnés à une taille donnée (3) puis déposés dans une plaque de microtitration à raison de 1-2 pg par puits (4). Ceci constitue le panel de cartographie qui est amplifié par PCR (5). Chaque plaque fille est alors utilisée pour cartographier un marqueur (6). La fréquence de coségrégation des marqueurs permet de déterminer leur position respective et d'établir une carte physique (7-8).



**Figure 2: Résultats de la cartographie HAPPY**

La région du chromosome 2 humain couverte est indiquée sur la gauche de la figure. La carte HAPPY a été générée en associant les marqueurs deux à deux avec un LOD score supérieur à 4. Les marqueurs indiqués sur la partie droite de la carte ont été disposés sans ambiguïté les uns par rapport aux autres. Les marqueurs "flottants" repérés sur la partie gauche de la carte présentent une association forte (LOD score supérieur à 6) avec les marqueurs qui leur font face, mais ne présentent qu'une faible association avec les marqueurs adjacents. La localisation du centromère est indiquée par un trait pointillé.



**Figure 3:** Comparaison des résultats de cartographie

Trois cartes sont représentées sur cette figure. La carte génétique (gauche) comporte les marqueurs polymorphes localisés par études familiales. La carte créée par hybrides d'irradiation (milieu) résulte de la localisation de marqueurs à la fois polymorphes et non polymorphes sur le panel G3 (Deloukas *et al.*, 1998). La carte HAPPY (droite) provient de notre étude. Les marqueurs disposés sur deux cartes différentes sont représentés par des traits fins joignant les deux cartes. La localisation du centromère est indiquée par un trait pointillé.