

Le maintien de l'intégrité des génomes, la création de la diversité génétique, l'organisation et l'évolution des génomes

Les problèmes de l'organisation, l'expression et l'évolution des génomes font l'objet d'études intenses et nombreuses en France et dans le monde. Des avancées considérables ont été réalisées au cours des dernières années, des questions fondamentales sont en voie d'être résolues, et de nouvelles questions ont émergé. Le maintien de l'intégrité du génome au cours des divisions cellulaires est évidemment crucial alors que le brassage du matériel génétique au fil des générations est une source majeure de création de la diversité génétique. Tous les organismes possèdent des systèmes élaborés pour réparer et/ou tolérer les lésions de l'ADN, que celles-ci soient spontanées ou provoquées par des agents chimiques ou physiques présents dans l'environnement. Ces questions sont abordées par une communauté scientifique qui se reconnaît sous le vocable 3R pour Réplication-Recombinaison-Réparation. Les questions d'organisation de la chromatine chez les eucaryotes ou du nucléoïde chez les procaryotes sont également abordées par plusieurs équipes françaises reconnues internationalement. Les éléments transposables, une composante majeure des génomes, et la régulation de leur transposition, ont également des conséquences importantes sur l'évolution des génomes.

"LES 3R"

Le point des connaissances

Les bactéries ont longtemps constitué les modèles d'étude majeurs de la réplication et de la réparation de l'ADN (modèle du réplicon, initiation de la réplication, caractérisation des ADN polymérase et des protéines du réplisome...). De nouvelles protéines associées à ces mécanismes ont été découvertes, et des liens ont été établis entre ces protéines ainsi qu'avec des complexes membranaires, des voies de signalisation, ou le métabolisme des sucres. Une régulation épigénétique du cycle cellulaire bactérien a par ailleurs été mise en évidence via un système de méthylation de bases à des sites spécifiques par des ADN-méthylases. Ces dernières régulent la réplication, la ségrégation des nucléoïdes, la transcription de gènes spécifiques, et jouent un rôle dans la virulence de divers pathogènes.

L'étude des Archaea a permis l'identification d'un grand nombre de protéines impliquées dans la réplication et la réparation, et mis en évidence le rôle des motifs structuraux et non de séquence. Les diverses ADN polymérase semblent avoir des rôles spécifiques (réplication du brin continu, du brin retardé, allongement d'amorce ARN pour générer des amorces ARN-ADN). La question de la régulation de l'initiation de la réplication dans un contexte de cycle cellulaire est posée. Une étude *in silico* de l'environnement des gènes de réplication dans 27 génomes d'Archaea suggère un lien entre réplication, biosynthèse du ribosome et traduction. De nouvelles nucléases putatives et des hélicases impliquées dans les 3R ont également été mises en évidence. Les Archaea sont donc à la fois un modèle d'étude simplifié pour l'étude des 3R chez les eucaryotes, mais constituent également un réservoir de nouvelles familles de protéines impliquées dans le métabolisme de l'ADN.

Chez les organismes multicellulaires, de multiples origines de réplication sans consensus de séquence sont utilisées de façon précoce ou tardive tout en étant conservées à chaque division. Le pattern des origines de réplication précoces et tardives change pendant le développement ou la différenciation, et nous ignorons ce qui détermine cette régulation dans l'espace et dans le temps. Moins de 0,1% des 50000 origines de réplication utilisées au cours du cycle cellulaire chez les mammifères ont été caractérisées. Il a été proposé que les origines de réplication chez les organismes multicellulaires soient définies par un état épigénétique permettant l'établissement de domaines de chromatine stables et autonomes de réplication et de transcription. La formation des complexes de pré-réplication est un nœud de régulation positive et négative du cycle cellulaire chez tous les eucaryotes. Le désengagement des complexes de pré-réplication constitue une étape importante dans le contrôle de la division cellulaire, couplée à la différenciation. La caractérisation des protéines spécifiques qui s'assemblent en complexe sur la chromatine a constitué une des grandes avancées dans la compréhension de la réplication et dans les mécanismes qui empêchent toute re-réplication au cours du même cycle cellulaire.

La réplication n'est pas un processus uniforme et continu. Plusieurs voies de sauvetage de fourches bloquées pour assurer le réassemblage d'un réplisome ont été décrites, et une grande diversité de polymérase translésionnelles ont été découvertes et caractérisées dans les trois domaines du vivant. Des études récentes ont permis de comprendre comment s'effectuent le choix et le recrutement des polymérase translésionnelles au niveau des fourches bloquées. De nombreux points de contrôle permettent d'éviter l'entrée de cellules contenant un ADN endommagé en phase S ou l'activation d'origines de réplication tardives, laissant ainsi aux cellules le temps de réparer fidèlement les lésions et empêchant l'accumulation de mutations.

Ces dernières années, on peut noter une mise en évidence de liens de plus en plus étroits entre réparation et réplication, mais également entre réparation et transcription, les régions d'ADN activement transcrites étant réparées prioritairement. Les réseaux d'interaction faisant intervenir des protéines de réparation et des facteurs impliqués dans la signalisation et la régulation du cycle cellulaire sont progressivement élucidés.

Des approches pointues en imagerie permettent de suivre l'action des protéines en temps réel et de déterminer leur cinétique de recrutement au niveau des lésions et leur co-localisation transitoire avec d'autres protéines de réparation. Les nouvelles approches (ChIP, ChIP-chip assays) permettent d'analyser le recrutement de protéines de réparation au niveau de l'ADN endommagé à l'échelle de génomes entiers. Le séquençage haut débit permettra d'élucider au niveau moléculaire les causes précises de maladies génétiques liées à des défauts de réparation de l'ADN. Par ailleurs, plusieurs équipes travaillent à la détermination de la structure des protéines de réparation par cristallographie ou RMN.

La problématique de la stabilité du génome est une thématique centrale de la méiose. En effet, lors de la méiose, un nombre élevé d'événements de recombinaison est induit, assurant la ségrégation en première division. Ces événements sont initiés par l'induction programmée de nombreuses cassures double brin, normalement réparées par recombinaison avec le chromosome homologue. Plusieurs observations suggèrent que d'autres modes de réparation de ces cassures peuvent conduire à des réarrangements chromosomiques à l'origine de pathologies chez l'homme, avec des implications importantes puisqu'ils seront transmis à la descendance.

Un deuxième aspect d'instabilité du génome concerne la ségrégation des chromosomes. Plusieurs études ont montré qu'une partie importante des trisomies était due à des problèmes de ségrégation des chromosomes en première division de méiose, ceux-ci liés à des modifications de la distribution et/ou de la fréquence des événements de recombinaison et de la stabilisation des chiasmas.

On voit ainsi que l'étude des 3R est d'une importance majeure pour analyser et comprendre ces phénomènes complexes. Une des propriétés des mécanismes de recombinaison et ségrégation des chromosomes en méiose est sa grande conservation évolutive. Ainsi, les grandes lignes des mécanismes mis en œuvre et plusieurs protéines clés impliquées dans ces processus sont conservées chez les champignons, plantes et animaux. Une des particularités du domaine est la complémentarité des études menées sur différents organismes, pour lesquels des stratégies et outils expérimentaux distincts peuvent être développés.

Les perspectives

Les bactéries et les eucaryotes unicellulaires constituent des modèles de choix pour repousser la frontière de nos connaissances dans le domaine des 3R, du fait de leur manipulation génétique facile, de leur temps de génération rapide et de l'accès à la séquence complète et annotée de centaines de génomes de microorganismes.

On peut cependant noter un intérêt croissant pour les études chez les organismes multicellulaires, en particulier les mammifères. La mise en place des origines de réplication, sa logique, son importance dans la régulation de l'entrée en division, dans le développement de l'embryon et la différenciation, dans les checkpoints et le stress cellulaire, sont des thèmes actuels de recherche. Au niveau moléculaire, le contrôle épigénétique des origines est une voie principale de recherche. La caractérisation et la localisation des facteurs impliqués dans la reconnaissance des origines, leur organisation structurale dans les foyers nucléaires de réplication, leur implication dans les pathologies de la prolifération, dont le cancer, sont des thèmes récurrents concernant les protéines de réplication. Ces dernières constituent une

famille de protéines qui devraient être d'un très grand intérêt dans le diagnostic et la thérapie anticancéreuse. Elles s'avèrent maintenant des marqueurs largement supérieurs aux marqueurs conventionnels pour identifier précocement une cellule compétente à la prolifération.

En ce qui concerne la méiose et la recombinaison, les thèmes abordés sont la différenciation germinale, le contrôle épigénétique lors de la gamétogenèse, la recombinaison, la ségrégation des chromosomes, la progression du cycle cellulaire. Les travaux de recherche pour aborder ces thèmes impliquent des approches complémentaires en génétique, biologie moléculaire et cellulaire, avec notamment une forte contribution de l'imagerie. Ces technologies sont particulièrement adaptées pour aborder plusieurs questions importantes encore sans réponse comme le processus de recherche d'homologie, la régulation des crossing over le long des chromosomes, le rôle de l'organisation des chromosomes dans le noyau, les contrôles épigénétiques.

Les forces en France

Les chercheurs Français qui s'intéressent au métabolisme de l'ADN se retrouvent tous les deux ans depuis 1995 au colloque des 3R qui réunit plus de 250 participants, notamment des jeunes chercheurs. Les laboratoires sont répartis dans tous les grands centres de recherche en région parisienne comme en province (Strasbourg, Toulouse, Marseille, Grenoble, Montpellier, Orléans, Brest...). Cette communauté travaille au sein du CNRS, de l'Inserm, de l'INRA, du CEA et de l'Université, et est regroupée dans deux GDR (un GDR "réplication", et un GDR "recombinaison et méiose").

Des perspectives nouvelles et des besoins nouveaux s'ouvrent avec le développement de nouvelles technologies en microscopie (imagerie en temps réel et microscopie à superrésolution), et en biologie moléculaire avec les analyses d'expression des gènes, de la chromatine, d'interaction des protéines à l'échelle du génome (Next Generation Sequencing).

On peut s'inquiéter de la disparition progressive de savoir-faire en matière de manipulation de microorganismes difficiles à cultiver, mais qui restent des sources de découvertes de nouveaux mécanismes et de nouveaux gènes assurant le maintien de l'intégrité de leurs génomes dans des environnements extrêmes ou des conditions de stress. Il est impératif de veiller à maintenir la diversité des modèles d'études et la complémentarité des approches expérimentales, sans céder à des effets de mode.

Les éléments transposables, composante majeure des génomes, source d'innovations pour l'évolution, révélateurs du contrôle épigénétique

L'accumulation des données de séquençage a mis en relief la place massive des éléments transposables et des rétrovirus endogènes dans les génomes, dont ils constituent une source majeure de plasticité. Ils sont en effet présents chez tous les organismes vivants, et représentent chez certains eucaryotes la plus grande partie du génome (la moitié du génome humain résulte de rétrotransposition). Ils jouent un rôle moteur dans la plasticité des génomes et la structuration de la chromatine. Du fait de leur aptitude à transposer, ils sont à l'origine de mutations, pouvant conduire à des maladies, y compris chez l'homme. Ils sont à l'origine de nombreux cas d'instabilités génétiques, ils peuvent induire différents types de réarrangements chromosomiques, et contribuent au brassage génétique. Ils constituent le moteur de transferts interspécifiques du matériel génétique. Ils sont ainsi des acteurs de premier plan dans la création de la diversité génétique, base de la biodiversité. Ils participent à la régulation de l'expression des génomes en apportant des séquences régulatrices, ou par des mécanismes épigénétiques récemment découverts, faisant notamment intervenir des petits ARN. Ils constituent donc une voie d'entrée originale et privilégiée pour l'étude des régulations épigénétiques comme le montre un nombre sans cesse croissant d'articles de référence concernant le contrôle par des ARN interférents de diverses catégories.

Avancées récentes et perspectives

Ces dernières années ont été marquées par des avancées importantes dans la compréhension des mécanismes moléculaires de la transposition, concernant aussi bien les transposons ADN (transposant directement sous forme ADN) que les rétrotransposons (transposant par l'intermédiaire d'un ARN). Les mécanismes d'intégration des transposons ADN et des séquences rétrovirales (qui sont similaires) sont maintenant finement disséqués, y compris au niveau de la molécule unique. Le mécanisme de transposition des rétrotransposons sans LTR (les plus abondants chez les mammifères), dans lequel l'intégration est couplée à la transcription inverse, est de mieux en mieux compris.

Le transfert horizontal de l'information génétique a un impact important sur l'évolution des génomes et la spéciation. Il a été récemment montré chez les bactéries que des intégrons sont impliqués dans la dispersion de déterminants de résistance aux antibiotiques et de facteurs de virulence. La régulation de ces mécanismes est dépendante du système SOS, ce qui souligne leur lien avec l'adaptation aux conditions environnementales. La découverte de l'induction ou de la modulation de la transposition en réponse aux conditions physiologiques ou à des modifications du milieu, chez les procaryotes comme chez les eucaryotes, suggèrent que les éléments transposables jouent un rôle clé dans la réponse des organismes aux stress et aux changements environnementaux.

De nombreuses études ont été conduites sur l'évolution des éléments transposables et leurs relations phylogénétiques, et sur les conséquences de leur présence dans les génomes. L'étude des mécanismes de rétrotransposition a, par exemple, permis de comprendre comment le génome humain a été façonné par l'activité de l'élément L1, en montrant que près de 45% de ce génome résulte de la machinerie enzymatique qu'il apporte et qui est responsable non seulement de sa propre dissémination, mais aussi de celle des séquences Alu et d'autres éléments, ainsi que des "rétropseudogènes processés".

Si ces éléments sont d'abord des parasites génomiques qui se propagent du fait de leur pouvoir invasif, leur présence influe sur le génome de l'hôte, et réciproquement. Les découvertes récentes de la génomique et l'étude du contrôle épigénétique montrent qu'il n'est plus possible de considérer indépendamment le génome "hôte" et ses "parasites": ainsi, les signaux de régulation, la méthylation de l'ADN, et la structuration de la chromatine, sont

souvent associés à ces éléments. Diverses fonctions, parfois vitales, comme la recombinaison V(D)J, essentielle dans la réponse immunitaire, la réorganisation programmée du macronoyau chez les ciliés, ou les fonctions codées par les syncytines, nécessaires à la formation du syncytiotrophoblaste, dérivent d'éléments transposables.

La transposition étant souvent un événement délétère, elle est étroitement contrôlée. Les mécanismes qui répriment les éléments transposables sont longtemps restés énigmatiques. Les études les plus récentes ont mis en évidence de nouvelles voies de régulation impliquant des petits ARN (siRNA, miRNA, piRNA). Dans de nombreux cas, les travaux pionniers sur les mécanismes de RNA silencing, qui sont universels chez les eucaryotes, ont été conduits à travers l'étude du contrôle des séquences mobiles. La découverte de ce nouvel univers n'en est qu'à ses débuts, et fait l'objet d'une compétition internationale intense.

Les forces en France, les nécessités

La France possède la seconde communauté (après les Etats-Unis) étudiant la transposition sous ses différents aspects. Les équipes françaises ont participé à toutes les avancées mentionnées ci-dessus, et pour nombre d'entre elles en position de leadership et de façon pionnière.

Cette communauté est reconnue comme telle sur le plan international et attire de nombreux jeunes chercheurs de talent. Elle est régulièrement invitée aux congrès internationaux, et organise elle-même en France des conférences, nationales et internationales, qui comptent parmi les plus importantes. Elle est structurée sous la forme d'un GDR regroupant pratiquement tous les acteurs du domaine. Plus de 25 équipes, dont un bon nombre sont largement reconnues sur le plan international, s'intéressent aux éléments transposables. Elles sont réparties sur l'ensemble du territoire, notamment dans la région parisienne, et à Montpellier, Lyon, Toulouse, Clermont-Ferrand, Perpignan, Nice, qui figurent parmi les centres les plus connus dans le monde.

Comme dans beaucoup d'autres domaines, des efforts considérables sont à faire pour que cette communauté puisse affronter l'ère du séquençage massif et pour être en mesure d'intégrer ses retombées dans cette recherche, qui se situe au coeur de la compréhension du génome, de son organisation, de son expression et de son évolution. L'un des axes qui devraient se voir décerner une attention particulière est celui des mécanismes épigénétiques de contrôle de l'expression du génome dans lequel l'étude des éléments transposables s'est révélée être d'une particulière richesse. Les recherches plus "classiques" doivent en parallèle être soutenues, car les vraies avancées en sont souvent issues, et qu'il n'y a pas de réel progrès provenant des études systématiques si on ne dispose pas de phénotypes renseignant sur fonctions biologiques qu'elles sont susceptibles de révéler.