

Module optionnel X - Génomique/Protéomique (6 ECTS)

Bilan des Conférences proposées

réalisé par les étudiants d'IT2005 ayant suivi dans ce module

Lieu de l'enseignement :
Centre des Enseignements Institut Pasteur, 25-28 rue du Docteur Roux, 75015 Paris

Compte-rendu n°1 :

I/ INTRODUCTION

II/ RESUMES DES CONFERENCES

III/ APPORTS DE LA GENOMIQUE/PROTEOMIQUE A L'IMMUNOLOGIE ET AUX IMMUNOTECHNOLOGIES

IV/ LES LIMITES ET QUESTIONS EN SUSPEND

V/ LES CONFERENCES

Evolution des génomes viraux

Pr. Simon WAIN-HOBSON

Analyse Du Transcriptome : Bases théoriques et applications

Pr. Catherine NGUYEN

Analyse des complexes protéiques par les approches de protéomique.

Pr. Bertrand SERAPIN

Récents développements en génomique structurale

Pr. Arnaud DUCRUIX

Approches protéomique pour la mise au point de nouveaux médicaments anti-Herpès

Pr. Jean-Jacques DIAZ

Du génome à la cellule : comment prédire la fonction des gènes ?

Pr. Antoine DANCHIN

VI/ DISCUSSION

I/ INTRODUCTION

Qu'est-ce que la génomique ... Le génome est défini comme l'ensemble du contenu en ADN d'un organisme. Il contient toute l'information nécessaire pour la reproduction, le développement et la physiologie de celui-ci. La génomique étudie ces génomes, dans leur

organisation, leur évolution, leur expression pour enfin aboutir à la fonction des gènes qu'ils abritent. Permises par les progrès des biotechnologies et de l'informatique, la génomique vise à dresser un inventaire de l'ensemble des gènes d'un organisme, puis à étudier la fonction de ces gènes et leurs interactions.

Du séquençage des génomes... La toute première étape consiste à obtenir l'intégralité du patrimoine génétique d'un organisme : sa **séquence génomique**. Cette étape ne se limite pas à fournir la séquence "brute" d'un génome. Un important travail de recherche informatique est effectué pour prédire la position des gènes dans la séquence et leurs fonctions: c'est l'**annotation** du génome.

...à la génomique fonctionnelle... Mais le séquençage global d'un génome n'est pas une fin en soi. Reste ensuite à caractériser la fonction des milliers de gènes séquencés (activité, domaines, partenaires) puis établir leurs profils d'expressions (spatial, temporel). L'analyse fonctionnelle des gènes n'est pas nouvelle. La nouveauté réside dans le caractère global des démarches. Pour ce faire, de puissants outils informatiques et des méthodes biologiques automatisées, en constante évolution, permettent une analyse de plus en plus rapide des données. La stratégie de la **génomique comparative** permet d'approcher la fonction des gènes. Elle passe par la comparaison des génomes avec celui d'un organisme proche et pathogène : les gènes de pathogénicité seront *a priori* localisés dans les seules régions qui diffèrent entre les deux génomes.

Au-delà de la génomique : la protéomique ...Le "**protéome**" est l'ensemble des protéines codées par le génome. L'analyse des interactions et des modifications au cours du temps des différentes protéines codées par les gènes apporte des informations supplémentaires sur leurs fonctions.

II/ RESUME DES CONFERENCES

Evolution des génomes viraux, d'après le Pr. Simon WAIN-HOBSON

Il existe **deux mondes de virus**, ceux à **ARN** et ceux à **ADN**. Les virus à **ARN** représentent près de 80% des virus connus. Ils sont très instables et subissent très fréquemment des modifications de leur séquence. Les mutations jouent un rôle crucial dans leur évolution. Ces modifications de leur génome sont majoritairement favorables pour infecter les cellules et déjouer les barrières (*l'effet inverse existe aussi*). Des spécialistes ont mis en évidence que les virus à **ARN** évoluent peu, ils sont très conservés, un très faible nombre de substitutions est nécessaire pour qu'ils évoluent et s'adaptent.

Dans "*l'autre monde*", celui des virus à **ADN**, on dénombre près de 10^{31} formes de bactériophages qui sont extrêmement dynamiques. Ce monde des virus à **ADN** abrite une réserve dont la richesse reste encore à être explorée...

ANALYSE DU TRANSCRIPTOME : Bases théoriques et applications, d'après le Pr. Catherine NGUYEN

Le transcriptome se définit comme l'ensemble des **ARN messagers** d'une cellule à un moment donné. L'ère de la génomique apporte un changement d'**échelle** à notre analyse: quelques séquences d'**ARN** auparavant contre des milliers de séquences à la fois avec l'étude du transcriptome. Les messagers restent un matériel difficile à étudier avec les méthodes

traditionnelles qui nous limitent à une analyse dite "gène à gène" mais ne sont pas adaptées à l'envergure de projets d'analyse du transcriptome.

La technologie des puces à ADN (DNA arrays, DNA chips) est parfaitement adaptée, elle vise à augmenter le nombre de gènes analysés tout en réduisant la quantité d'ARN nécessaire. On distingue différentes approches (cf. conférence). De nombreuses applications sont en cours dans le secteur médical. L'analyse du profil d'expression de messagers devra permettre d'établir des **clusters de gènes prédictifs corrélés avec la survie**. Les spécialistes espèrent ainsi adapter des traitements spécifiques pour certains groupes de patients ...

Analyse des complexes protéiques par les approches de protéomique. D'après le Pr. Bertrand SERAPIN

La protéomique vise à identifier et quantifier l'ensemble des protéines prises dans un contexte biologique donné. Il ne faut donc pas considérer la protéomique comme l'étude du protéome mais plutôt **des protéomes** pris dans différents contextes biologiques.

L'analyse des protéines est complémentaire à celle du génome et du transcriptome. Il est important d'analyser les complexes protéiques. En effet les protéines n'agissent pas seules : ce sont des complexes protéiques qui les rendent biologiquement actives.

L'analyse des complexes protéiques par les approches de protéomique peut s'appliquer à l'immunologie. Il serait intéressant de comparer les profils d'expression de composants de ces complexes dans contextes pathologiques à différents stades de traitements afin d'identifier certaines variations relevantes.

Finalemment bien que les méthodes actuelles d'analyse du protéome ouvrent de nouvelles perspectives, elles n'ont pas encore la capacité de suivre la cadence imposée par la génomique à grande échelle.

Récents développements en génomique structurale d'après le Pr. Arnaud DUCRUIX

La génomique structurale s'inscrit comme une approche complémentaire aux programmes de séquençage systématique. Elle se situe à l'interface entre la génomique systématique et la protéomique. Cette discipline délivre une information supplémentaire : les **structures tridimensionnelles** des protéines solubles correspondants aux ORF du génome étudié. Cette approche bien que très prometteuse se heurte à certains problèmes techniques.

Parmi ces problèmes, un demeure récurrent. En effet le cristal n'est que la représentation **figée** d'une structure à un instant "*T*" obtenu dans des conditions « **optimale** ». Or les molécules sont en perpétuel mouvement, même à la surface membranaire, et dans des contextes pathologiques on peut penser que ce cristal « **consensus** » s'éloigne probablement de la réalité biologique.

Malgré ses limites la génomique structurale reste un domaine très prometteur, elle offre un panel d'applications variées. Pour exemple, dans le domaine thérapeutique elle pourrait détecter des **altérations tridimensionnelles (MISFOLDING)** associées à certaines pathologies, qui restent encore indétectable simplement par les séquences nucléiques et protéiques.

Approches protéomique pour la mise au point de nouveaux médicaments anti-Herpès D'après le Pr. Jean-Jacques DIAZ

Le virus Herpès simplex de type I (**HSV-I**) est un adénovirus dont le génome est constitué d'ADN bicaténaire.

Les stratégies courantes s'attaquent directement aux protéines virales. **Une nouvelle génération d'antiviraux** pourrait plutôt cibler les protéines cellulaires indispensables au développement du virus. Dans ce but une approche par la protéomique a permis de voir les

protéines cellulaires qui échappent à cette inhibition viro-induite et dont l'expression est même stimulée pendant l'infection.

Le traitement le plus utilisé actuellement est un antiviral nommé **ACYCLOVIR**. Cette étude a toute son importance car ce médicament s'avère inefficace face aux cas croissants de patients présentant des **souches résistantes à cet antiviral**.

Cette approche protéomique constitue un exemple concret d'application dans le domaine de l'immunologie infectieuse : elle a permis d'identifier des molécules inhibitrices de la voie de biosynthèse de polyamines qui constituent une nouvelle source de candidats thérapeutiques pour combattre les infections virales du HSV-1.

Du génome à la cellule : comment prédire la fonction des gènes ?

Pr. Antoine DANCHIN

La compréhension d'un génome dépend de la grandeur et de la justesse de **l'acquisition des données de séquences**. L'importance de cette acquisition est primordiale pour l'étude du génome en lui-même mais d'autant plus qu'actuellement et pour le futur ces données de séquences servent et serviront de référence dans d'autres approches comme le transcriptome et le protéome.

La pertinence et la qualité du travail d'analyse ne doivent pas être remplacé par un emballage qui peut aboutir à une **accumulation de données créant le flou parmi la masse et noyant ainsi l'information pertinente**. Ces approches doivent être encadrées par un concept central : **le concept de fonction**.

La science est une étude de relation entre objets et non une étude des objets

Ainsi les séquences nucléiques doivent être comparées à un modèle des plus réaliste possible (*génomique soustractive*).

Pour nous guider dans ces comparaisons il est intéressant d'explorer plusieurs types de voisinages : toute proximité dans les séquences, tous types de régulation commune, toute appartenance à un même complexe...

Finalement l'évolution passe par le cycle variations-sélections-amplifications. Ce schéma dynamique permet **l'émergence de la fonction**. Seulement cette émergence peut se présenter dans une fenêtre temporelle passée, présente comme future...

Finalement le transcriptome n'est qu'une approche, le travail qui suit reste le même.

Cette approche est puissante, mais elle permet simultanément de prendre en compte tant de paramètre dans l'étude que leur analyse en deviendra de plus en plus complexe.

III/ APPORTS DE LA GENOMIQUE/PROTEOMIQUE A L'IMMUNOLOGIE ET AUX IMMUNOTECHNOLOGIES

Ces études globales permettent d'étudier simultanément un nombre d'éléments beaucoup plus important. Avec cette multiplication de données complémentaires (*séquences génomiques, transcriptomes, protéomes*) nous sommes maintenant en passe de poser des questions biologiques en terme de **larges réseaux d'interactions globales** et non plus seulement de quelques interactions spécifiques. Ainsi l'accumulation de très larges collections

de données sur les réseaux transcriptionnels, les interactions protéiques, les complexes macromoléculaires ou les interactions génétiques, permettront d'envisager la **modélisation des interactions dynamiques** qui ont lieu dans la cellule et d'imaginer leur évolution.

De manière commune, l'immunologie comme la génomique/protéomique, sont des sciences d'études de **relations entre objets**. Le **concept d'étude** consiste à analyser de manière **comparative** les informations (profils d'expression protéiques, des messagers, ...) en fonction de différentes conditions physiopathologiques et de différents temps. Dans tous les cas cette comparaison doit se faire par rapport à une **référence** dont la justesse ou le **degré de réalisme** s'avère crucial pour l'analyse des résultats. Le choix du modèle demeure donc crucial quant à la pertinence des interprétations qui en découlent.

Pour nous guider dans ces comparaisons il est intéressant **d'explorer plusieurs types de voisinages** : toute proximité dans les séquences, tous types de régulation commune, toute appartenance à un même complexe...

Les modifications du protéasome :

Les bio-marqueurs

Il est possible de comparer des protéasomes pris :

- chez des sujets sains et d'autres développant la pathologie d'intérêt
- chez des malades avant pendant et après traitement
- sur des groupes de malades présentant des signes communs qui les démarquent des autres patients...

Le transfert des connaissances obtenues après analyse peut permettre de détecter une multitude de **"bio-marqueurs"** :

- des marqueurs détectant la **prédisposition** de sujets sains à développer une pathologie
- des marqueurs de **diagnostique** qui vont permettre de détecter l'apparition de la maladie : il est clairement reconnu que toute pathologie dépistée précocement offre les meilleures probabilités de guérison.

Ces marqueurs ont pour but de **détecter la maladie** ou les signes avant-coureurs. Dès que la maladie est en place il est possible de trouver d'autres types de bio-marqueurs qui vont permettre plutôt de **mieux comprendre la maladie**. Ils permettront de :

- caractériser par exemple **un type de tumeur** : ceci est utile pour établir des clusters de tumeur pour rendre encore plus spécifique l'analyse des résultats. Trop d'hétérogénéité risque de fausser les interprétations (*il faut comparer ce qui est comparable*)
- distinguer **différents stades** d'une même pathologie chronique

Dans le cas d'un traitement thérapeutique il est clair que les entreprises de biotechnologies proposant une immunothérapie seront intéressées par ce genre d'approche, notamment lors de leurs essais cliniques pour suivre au mieux l'évolution de la pathologie. C'est un moyen de mieux cerner l'action d'un médicament et surtout ses effets sur la population.

Ainsi pendant le traitement d'autres bio-marqueurs pourront être découverts :

- des marqueurs corrélés à la **survie et aux risques de rechute** permettront de prédire sur certaines **sous-populations** si le traitement sera plus ou moins efficace.

Avec ce genre d'approche on pourra prédire si tel ou tel traitement sera plus adapté et pourra alors être proposé au malade. Tous ces paramètres sont primordiaux pour les chercheurs du secteur des immunothérapies car ils renseignent sur le potentiel du produit, il sera possible de voir plus rapidement la **proportion de bons répondeurs** au traitement par rapport à la population totale.

Ces informations apporteront une aide précieuse aux médecins pour établir leur diagnostic et leur décision thérapeutique.

Un exemple concret actuel de projet d'étude clinique à grande échelle est actuellement en cours à l'Institut Curie, ils utilisent la technologie *genechips* d'**AFFYMETRIX**. Dans le traitement contre le virus HSV-1, cette approche a permis d'identifier une nouvelle génération d'antiviraux. La découverte s'est faite en analysant des profils d'expression protéiques avant et après infection.

Le polymorphisme

Au sein d'une même espèce nous devons apprendre à déchiffrer un message nécessairement bruité par la multitude et la complexité des degrés d'hétérozygotie dus au polymorphisme. Ce polymorphisme constitue un outil indispensable en génétique humaine pour l'étude de nombreuses pathologies. Outre les formes alléliques de nos gènes, différents polymorphismes sont retrouvés le long du génome :

- **RFLPs** (*restriction*),
- **VNTRs** (*micro et mini-satellites*)
- **SNPs** (*Single Nucleotide Polymorphism*).

La variabilité de l'efficacité d'un même traitement sur la population repose en grande partie sur les combinaisons de polymorphismes présentés par chaque individu. L'étude de son impact peut permettre de diviser les malades en sous classes de répondeurs plus ou moins sensibles à un traitement. Pour exemple, le génome humain présente quelque 10 millions de SNPs, le projet Québécois « HAPMAP » utilise une approche de génomique à grande échelle (Technologie ILLUMINA) pour regrouper ces SNPs en blocs aux propriétés communes. A terme les chercheurs n'auront qu'à en détecter quelques uns (considérés comme marqueurs) pour connaître tous les SNPs associés à un bloc particulier.

Les résultats de ce travail seront retentissant dans les approches pharmacogénomiques. Cela devrait permettre de définir des profils de réactivité à un traitement voir même de sensibilité à certains effets secondaires. On espère adapter au mieux ainsi les traitements pour "chaque individu".

Cette étude du polymorphisme à grande échelle pourra également être utilisé dans l'étude de certaines immunopathologies. Son champ d'application est très vaste.

Les modifications de conformation de protéines

Certaines pathologies sont indétectables par l'analyse des séquences nucléiques et protéiques. Il s'agit de protéines qui subissent des altérations plus ou moins discrètes au niveau de leur conformation tridimensionnelle. Imaginons une protéine X qui subit ce genre d'altération et un anticorps monoclonal dirigé contre l'épitope X1 dont la conformation n'est pas affecté par cette mutation. Il est clair que dans ce cas l'on ne soupçonnera pas cette altération dans la mesure où X est toujours reconnu par l'Ab. Cette situation devient très inquiétante si cette modification provoque des effets délétères. Ce phénomène est nommé **MISFOLDING** et il est responsable de nombreuses pathologies que l'on connaît et d'autres que l'on risque de connaître prochainement vu l'envergure du projet **ISGO** sur le site Pasteur Génopole. La plate forme technologique en place a pour objectifs d'automatiser la cristallogénèse, systématiser la renaturation et automatiser l'analyse des structures 3D afin de pouvoir pourquoi pas un jour s'attaquer aux protéines de plusieurs génomes.

Le secteur des immunotechnologies pourrait utiliser ces données pour concevoir de nouveaux anticorps spécifiques pour les épitopes conformationnels qui apparaissent exclusivement sur les formes protéiques ayant subi cette altération.

IV/ LES LIMITES ET LES QUESTIONS EN SUSPEND

Malgré tous ces progrès spectaculaires, les approches de génomique, transcriptomique et protéomique butent encore à plusieurs niveaux.

Les spécialistes ont du mal à reconnaître le "*bruit de fond*" dû en partie à l'imperfection de nos annotations ce qui perturbe alors l'information génomique. Ce problème restera récurrent dans le futur car l'évolution contribue au développement de ce "*bruit*".

Un autre point sombre reste la prédiction de la structure tridimensionnelle des protéines dans leur environnement cellulaire en fonction de leur simple séquence primaire.

Un autre niveau, chevauchant avec le précédent, est celui des interactions macromoléculaires qui sont au cœur même des problèmes fonctionnels. Actuellement l'analyse des complexes protéiques par la protéomique ouvre un nouveau champ dans cette direction (conférence du Pr. SERAPIN). On pourrait imaginer trouver ainsi de nouveaux anticorps neutralisants. Les cibles thérapeutiques correspondraient à des sites indispensables pour la formation de complexes protéiques impliqués dans des **désordres immunitaires** par exemple.

Communément à l'immunologie et à la génomique/protéomique, l'étude des protéines membranaires pose vraiment problème. La cristallisation des telles protéines n'est pas actuellement réalisable. De plus si un modèle de cristal est obtenu, la pertinence du modèle reste limitée dans la mesure où la conformation d'une protéine hors contexte membranaire n'a pas grand sens. Les rafts lipidiques sont impliqués dans les mécanismes immunitaires tel que la formation de la synapse immunologique mais actuellement les conformations tridimensionnelles adoptées par les nombreux composants impliqués restent obsolètes.

V/ LES CONFERENCES

Evolution des génomes viraux

Pr. Simon WAIN-HOBSON

L'histoire de la génomique débute avec les virus car les moyens disponibles à cette époque étaient adaptés pour l'étude de génome de plus petite taille. Ainsi En 1976 le premier génome de virus est alors séquencé, il s'agit du bactériophage M52, il mesure 3500 pb. Débute alors une série de travaux de séquençage : SV40 est séquencé en 1978, HBV l'année suivante, Poliovirus en 1981...SANGER publie en 1984 une séquence de taille jamais égalée jusqu'alors : 172 282pb. Enfin, en 1985 le virus HIV est séquencé, de taille bien inférieure mais dont les caractéristiques posent encore de sérieux problèmes à l'heure actuelle.

Il existe **deux mondes de virus**, ceux à **ARN** et ceux à **ADN**. Les virus à ARN représentent près de 80% des virus connus, leur taille n'excède pas 30Kb. Les virus à ADN sont bien plus grands, ils peuvent atteindre le Mégabase. Un écart de 3 log sépare le taux d'erreur des virus à ARN (2-3 par cycle) comparé aux virus à ADN (0.0046). On pourrait expliquer en partie cette différence par une surveillance moins accrue dans le cas du système de correction impliquant les RNases.

Si l'on réalisait une "*Echelle de Richter pour les virus*", le VIH serait en tête. Il s'agit d'un virus à ARN, qui présente des taux de fixation de mutation des plus élevés. Les mutations jouent un rôle crucial dans l'évolution des rétrovirus. Pour exemple, le génome du HIV fait 9200pb et il a été montré qu'en 15 ans celui-ci pouvait subir près de 9000 recombinaisons (3 recombinaisons par cycles et 200 cycles annuels). La reverse transcriptase du VIH crée des **mutations ponctuelles** de l'ordre de 0.25/génome/cycle et environ 700

hypermutations /génom/cycle. Il faut cependant ajouter que ces recombinaisons sont majoritairement favorables pour infecter les cellules et déjouer les barrières mais l'effet inverse existe aussi.

Comme l'on vient de le voir les génomes de virus à ARN sont très instables et subissent très fréquemment des modifications de leur séquence. Pour le cas du VIH une même cellule peut d'ailleurs être le siège de plusieurs formes de ce virus : des expériences décrites dans un système in vitro dénombrent jusqu'à 4 variants du provirus dans une seule et même cellule T-CD4⁺. Ils indiquent 9 séquences d'ADN différentes dans ces cellules. Cette dernière observation révèle qu'il y a en moyenne plus de génomes que de provirus intégrés. Ceci se traduit entre autre par la co-existence de formes intégrées et non intégrées du virus.

Ainsi une même cellule subit de multiples infections, typiquement des centaines de virions sont nécessaires pour l'infecter. Une étude montre d'ailleurs qu'une cellule infectée peut contenir jusqu'à 50-700 formes et que 500-4000 RNA virus sont nécessaires pour obtenir ce profil d'infection.

Enfin, des spécialistes ont mis en évidence que les virus à ARN évoluent peu, ils sont très conservés, un très faible nombre de substitutions est nécessaire pour qu'ils évoluent et s'adaptent. Finalement malgré leur fort potentiel de variabilité, les virus à ARN changent plus qu'ils ne s'adaptent.

Dans "*l'autre monde*", celui des virus à ADN, on dénombre près de 10³¹ formes de bactériophages. Cette population de virus est extrêmement dynamique mais encore si mystérieuse : pour exemple, 50% des gènes de mycobactériophages ne trouvent aucune similarité avec des séquences déjà connues. Finalement le monde des virus à ADN abrite une réserve dont la richesse reste encore à être explorée.

ANALYSE DU TRANSCRIPTOME : Bases théoriques et applications

Pr. . Catherine NGUYEN

L'ère de la génomique apporte un changement d'**échelle** à notre analyse: quelques séquences d'ARN auparavant contre des milliers de séquences à la fois avec l'étude du transcriptome. Cette approche s'accompagne d'un changement de **vision** : l'analyse devient dynamique, la possibilité de travailler à si grande échelle permet de tenir compte des facteurs **spatiaux** et **cinétiques**.

Le transcriptome se définit comme l'ensemble des ARN messagers d'une cellule à un moment donné. Une cellule de mammifère contient 10-30 pg d'ARN total, lui même composé des ARN ribosomiaux (85%), ARN de transfert (10-15%) et des ARNm qui ne représentent que 1-5%. Autrement dit l'ARNm faiblement abondant, plus fragile que l'ADN, reste un matériel difficile à étudier. Les techniques traditionnelles nous renseignent parfaitement sur les aspects qualitatifs et quantitatifs de l'expression des messagers : Northern blot, Dot blot, RT-PCR. Cependant ces approches se limitent à une étude dite "*gène à gène*" mais ne sont pas adaptées à l'envergure de projets d'analyse du transcriptome.

La technologie des puces à ADN (DNA arrays, DNA chips) est parfaitement adaptée, elle vise à augmenter le nombre de gènes analysés tout en réduisant la quantité d'ARN nécessaire. On distingue différentes approches selon le type de :

- sonde : *oligo longs obtenus par PCR, oligo courts synthétisés chimiquement,*
- support : *membrane de nylon, lame de verre, puce avec silicium,*
- marquage réalisé : *colorimétrie, radioactivité, fluorescence ...*

Des entreprises phares (AFFYMETRIX, ILLUMINA) proposent des technologies de puces optimisées mais dont le prix est encore peu abordable pour les laboratoires publics. La dernière version puce humaine d'**AFFYMETRIX** mise sur le marché en septembre 2003 porte **1 300 000 oligonucléotides** différents de **25 mer** de long (taille maximum pour éviter les formes dégénérées). Le principe repose sur l'existence de séquences dites **consensus** qui servent de références et doivent être spécifiques à un gène donné. Pour exemple une "**genechip array**" peut contenir **500 000 puits** contenant chacun des oligomères.

De nombreuses applications sont en cours dans le secteur médical. L'institut Curie possédant des **banques d'échantillon des 15 dernières années** s'est allié à la technologie AFFYMETRIX pour mener une étude clinique à grande échelle. Ce sont les premières puces à être suffisamment puissante pour permettre l'analyse de plusieurs dizaines de milliers de messagers dans une même cellule. Le but du projet est de mettre en évidence des signatures génétiques de certains cancers pouvant servir au développement de tests diagnostics. L'analyse du profil d'expression de messagers devra permettre d'établir des **clusters de gènes prédictifs corrélés avec la survie**. Des classifications pertinentes parmi l'hétérogénéité des tumeurs facilitera l'identification de marqueurs génétiques prédictifs du risque des rechutes après le traitement initial de cancer du sein et de l'œil. En fonction de ces paramètres, un traitement plus adapté pourra être proposé.

Finalement le transcriptome n'est qu'une approche, le travail qui suit reste le même. Cette approche est puissante, mais elle permet simultanément de prendre en compte tant de paramètres dans l'étude que leur analyse en deviendra de plus en plus complexe. Cependant ces analyses permettront de mieux comprendre la maladie et d'en identifier des cibles. On pourrait rêver d'adapter des traitements spécifiques pour chaque patient ...

[Analyse des complexes protéiques par les approches de protéomique.](#)

Pr. Bertrand SERAPIN

La protéomique vise à identifier et quantifier l'ensemble des protéines prises dans un contexte biologique donné. A la différence de la génomique il est nécessaire de tester beaucoup de conditions pour mettre certaines protéines en évidence. Il ne faut donc pas considérer la protéomique comme l'étude du protéome mais plutôt **des protéomes** pris dans différents contextes biologiques.

L'analyse des séquences nucléiques bien que fastidieuse semble plus simple que l'analyse de séquences polypeptidiques. En effet une autre dimension est à prendre en compte : la **structure tridimensionnelle** des protéines qui dépend elle-même de multiples **propriétés physico-chimiques** qui peuvent varier selon les **conditions physiopathologiques**.

Cette approche requiert diverses méthodes pour séparer, identifier et analyser ces protéines.

L'électrophorèse bidimensionnelle est couramment utilisée, elle s'appuie sur les différences de charge et de masse, aujourd'hui elle permet de séparer jusqu'à plusieurs centaines de protéines en une seule étape. Elle demeure une méthode assez fiable pour estimer leur abondance et leurs modifications post-traductionnelles.

La spectrométrie de masse (MALDI-TOF ou MS/MS) permet une analyse plus poussée, elle permet l'identification des protéines. En pratique celles-ci subissent une digestion enzymatique, les fragments peptidiques générés sont ionisés en phase aqueuse (MALDI) ou bien en solution (MS-MS) puis propulsés dans un confinement sous vide. L'identification se fait par comparaison du spectre peptidique avec des références. Le **pouvoir**

et la justesse de cette étape sont véritablement **dépendants de la qualité des banques de données** donc finalement de la bonne caractérisation des génomes de références.

Les protéines n'agissent pas seules, ce sont des complexes protéiques qui les rendent biologiquement actives. Il est donc important d'analyser ces complexes en terme de :

- composition (*type de sous-unités et leur stoechiométrie*),
- patron des sous unités interagissant (*les éventuelles modifications de ses partenaires dans le temps ou selon les contextes biologiques*)
- structure tridimensionnelle (*X-ray, RMN ...c à d génomique structurale*)
- localisation, concentration cellulaire, fonction ...

L'assemblage des complexes est souvent coopératif, ce serait une erreur de se limiter seulement à des interactions binaires.

La technologie TAP (Tandem Affinity Purification) permet de purifier des complexes protéiques rapidement et avec fiabilité. Il s'agit d'une méthode de purification par chromatographie d'affinité nécessitant 2 étapes. La stratégie consiste à produire une protéine X fusionnée à la fois avec une protéine A (affine pour les IgG) et un domaine CBD (affin pour la calmoduline). Deux éluions successives sont ainsi nécessaires. La technique s'avère prometteuse en terme de sensibilité, rapidité et fiabilité en comparaison avec les immunoprécipitations courantes. Son association avec l'électrophorèse bidimensionnelle et la spectrométrie de masse permet de caractériser simultanément des "*pattern*" de partenaires protéiques au cœur de complexes. A la différence des approches par "*protein chips*" où les sondes sont spécifiques individuellement pour chaque facteur, cette combinaison incluant **TAP** permet de **détecter des facteurs jusqu'alors inconnus**.

L'analyse des complexes protéiques par les approches de protéomique peut s'appliquer à l'immunologie. Il serait intéressant de comparer les profils d'expression de composants de ces complexes dans contextes pathologiques à différents stades de traitements afin d'identifier certaines variations relevantes.

Finalement bien que les méthodes actuelles d'analyse du protéome ouvrent de nouvelles perspectives, elles n'ont pas encore la capacité de suivre la cadence imposée par la génomique à grande échelle.

Récents développements en génomique structurale

Pr. Arnaud DUCRUIX

La biologie structurale vise à éclairer des structures moléculaires et supramoléculaires à l'échelle atomique pour mieux cerner les relations structure-fonction sous-jacentes.

La génomique structurale s'inscrit comme une approche complémentaire aux programmes de séquençage systématique. Elle se situe à l'interface entre la génomique systématique et la protéomique. Cette discipline délivre une information supplémentaire : les **structures tridimensionnelles** des protéines solubles correspondants aux ORF du génome étudié.

Les méthodes employées sont diverses, il s'agit en général de projets pluridisciplinaires :

- La bioinformatique facilite l'analyse du génome et la sélection des ORFs

- la biologie moléculaire, la biochimie permettront d'exprimer, produire, purifier les protéines d'intérêt. La biochimie analytique permet de dégager les principales caractéristiques de la protéine (Mw, pI, glycosylations, hydrophobicité...)
- les outils de microscopie haute résolution associés à l'expérience des biophysiciens et cristallographes (rayons X, RMN...) permettront d'établir la résolution des structures tridimensionnelles.

Cette approche bien que très prometteuse si elle pouvait être automatisée, se heurte à certains problèmes techniques : en effet, chaque technique possède ses limites donc selon les cas certaines sont plus appropriées. Ainsi les structures figées de taille limitée seront plutôt étudiées par Rayons X, alors que la RMN permet d'approcher des structures dynamiques. Le soucis majeur se pose pour les protéines membranaires, toute approche structurale tridimensionnelle est presque impossible, seule la microscopie donne une vision bidimensionnelle.

Ce dernier problème n'est pas des moindres puisque ce microenvironnement membranaire abrite près de 30% des protéines totales. De plus il est clair d'après la littérature que la connaissance plus poussée de certaines de ces protéines présentes dans les rafts pourrait nous éclairer fortement quant aux mécanismes impliqués dans certains phénomènes immunologiques (reconnaissance antigénique, contact cellulaire ...). Un autre point sombre à l'heure actuelle reste l'étude de structures désorganisées ou qui s'organisent en présence d'un partenaire.

Parmi ces problèmes, un autre demeure récurrent. En effet le cristal n'est que la représentation **figée** d'une structure à un instant "**T**" obtenu dans des conditions "**optimales** ." Or les molécules sont en perpétuel mouvement, même à la surface membranaire, et dans des contextes pathologiques on peut penser que ce cristal "consensus" s'éloigne de la réalité biologique.

Malgré ses limites la génomique structurale reste un domaine très prometteur, elle offre un panel d'applications variées. Entre autre, concernant la problématique "*relation structure fonction*", des modèles d'insertions de structures dans des cavités ou poches sont en cours (ex : insertion dans la poche ATP).Egalement, dans le domaine thérapeutique elle pourrait détecter des **altérations tridimensionnelles** associées à certaines pathologies, qui restent encore indétectable simplement par les séquences nucléiques et protéiques. Pour répondre aux attentes d'une automatisation de ce genre d'approche, un programme international est en cours **ISGO** (***I**nternational **S**tructural **G**enomics **O**rganisation*). En Ile-De-France le site Pasteur Génopole a mis en place une plate forme technologique dont les objectifs sont d'automatiser la cristallogenèse, systématiser la renaturation et automatiser l'analyse des structures 3D afin de pouvoir pourquoi pas un jour s'attaquer aux protéines de plusieurs génomes.

[Approches protéomique pour la mise au point de nouveaux médicaments anti-Herpès](#) Pr. Jean-Jacques DIAZ

Le virus Herpès simplex de type I (**HSV-I**) est un adénovirus qui fait 200 nm de diamètre. Son génome est constitué d'ADN bicaténaire. Sa période d'incubation dure de 7 à 10 jours. Lorsque HSV-I infecte des cellules permissives, il induit la synthèse de transactivateurs transcriptionnels, des protéines de phase précoce puis tardive. Ceci aboutit à l'expression de son génome viral. En parallèle la synthèse des protéines saines subit une inhibition de type viro-induite.

Les stratégies courantes s'attaquent directement aux protéines virales. Une nouvelle génération d'antiviraux pourrait plutôt cibler les protéines cellulaires indispensables au développement du virus. En effet les protéines cellulaires dont l'expression persiste ou croît durant l'infection, sont probablement nécessaires au bon développement de celle-ci. Dans ce but une approche par la protéomique a permis de voir les protéines cellulaires qui échappent à cette inhibition viro-induite et dont l'expression est même stimulée pendant l'infection.

Le traitement le plus utilisé actuellement est un antiviral nommé **ACYCLOVIR**. Cette étude a toute son importance car ce médicament s'avère inefficace face aux cas croissants de patients présentant des souches résistantes à cet antiviral.

Une équipe de chercheurs a étudié les modifications transitoires du protéome de cellules infectées par HSV-1. L'étude est réalisée *in vitro*. Des cellules infectées par le virus **HSV-1** sont incubées en présence de **méthionine** ⁵¹ **S**. Une cinétique d'expression des protéines cellulaires est suivie à divers temps post-infection. Leur étude a permis de détecter 28 protéines échappant à cette régulation négative viro-induite. Dans les virus les polyamines tels que la spermine ou spermidine sont connues pour favoriser l'empaquetage de l'ADN viral en neutralisant les charges négatives de l'ADN. Une analyse plus précise révèle que ces 28 protéines appartiennent à la voie de biosynthèse des polyamines. Dans une expérience réalisée sur ces mêmes cellules infectées, ils ont montré qu'en présence d'un inhibiteur de cette voie (**SAMDC** : *Polyamine synthesis pathway*) ils empêchaient la réplication virale. (*Une expérience annexe montre d'ailleurs que l'addition de spermine restaure la réplication virale*).

Cette approche protéomique constitue un exemple concret d'application dans le domaine de l'immunologie infectieuse : les molécules inhibitrices de la voie de biosynthèse de polyamines constituent une nouvelle source de candidats thérapeutiques pour combattre les infections virales du HSV-1. La découverte de telles cibles thérapeutiques s'avère très utile dans la mesure où elle a permis de combattre 50% des souches virales résistantes au traitement classique de l'antiviral **ACYCLOVIR**. Ce nouveau type de traitement réduit donc également l'émergence de souches virales résistantes.

Finalement dans cette direction l'approche par la protéomique permettra de mieux cerner comment un virus peut détourner la machinerie cellulaire à son compte et ainsi envisager de nouveaux traitements plus adaptés.

[Du génome à la cellule : comment prédire la fonction des gènes ?](#)

Pr. Antoine DANCHIN

Comment définir la vie ? Elle est composée de trois processus de base. Un métabolisme indispensable à celle-ci, une compartimentation plus ou moins élaborée et un transfert d'information véhiculé selon trois langages distincts : réplication (ADN), transcription (ARNm) et traduction (protéine).

La compréhension d'un génome dépend de la grandeur et de la justesse de **l'acquisition des données de séquences**. L'importance de cette acquisition est primordiale pour l'étude du génome en lui-même mais d'autant plus qu'actuellement et pour le futur ces données de séquences servent et serviront de référence dans d'autres approches comme le transcriptome et le protéome. Le séquençage du génome représente un travail fastidieux et bien souvent après une couverture de 10X on considère cela suffisant. Ceci est regrettable, il faudrait vraiment une couverture complète même si en général pour terminer cela est toujours long...

Comprendre un génome... c'est savoir l'analyser

Beaucoup de données de séquences restent et resteront partiellement exploitées ou exploitables à un instant "t". En augmentant ces données il arrive que l'on trouve la réponse à des questions laissées en suspens depuis cet instant "t". La question est de savoir si cette réponse n'était pas finalement accessible tout simplement dès cet instant "t" ?

En d'autres termes la pertinence et la qualité du travail d'analyse ne doivent pas être remplacé par un emballage qui peut aboutir à une **accumulation de données créant le flou parmi la masse et noyant ainsi l'information pertinente**. Ces approches doivent être encadrées par un concept central : **le concept de fonction**.

La science est une étude de relation entre objets et non une étude des objets.

Ainsi les séquences nucléiques doivent être comparées à un modèle des plus réaliste possible. **La génomique soustractive** s'appuie sur des comparaisons entre ce qui est présent par rapport à ce qui est absent et/ou ce co-varie. Pour des applications thérapeutiques on utilisera des modèles de "*références*" que l'on comparera aux conditions pathologiques.

Pour nous guider dans ces comparaisons il est intéressant d'explorer plusieurs types de voisinages : toute proximité dans les séquences, tous types de régulation commune, toute appartenance à un même complexe...

Organisation du génome... de moins en moins de hasard...

L'analyse des génomes a montré que l'ordre des gènes dans le chromosome n'est pas aléatoire, mais correspond à l'architecture de la cellule. On peut également se demander si l'organisation des gènes autour d'une fourche de réplication est régit par le hasard ? Les gènes les plus proches de l'origine de réplication sont les plus répliqués or il se trouve que chez E. coli ils correspondent aux gènes indispensables pour la survie et l'adaptation : gènes de maintenance, du métabolisme, et de résistance. Leur organisation spatiale est donc régit dans ce cas selon leur niveau d'importance.

Le concept de la fonction

Il est clair que le "*non connu*" avec les gènes ne veut pas forcément dire "*pas de fonction*". Ainsi la compartimentation cellulaire n'est-elle pas à l'origine d'une caractéristique des génomes qui avait beaucoup surpris au moment de sa découverte, à savoir la présence de nombreux gènes sans fonction connue ? Dans ce contexte l'étude du métabolisme du soufre a permis d'éclaircir la situation concernant de nombreux gènes dont la fonction était restée jusqu'à présent inconnue.

Finalement l'évolution passe par le cycle variations-sélections-amplifications. Ce schéma dynamique permet **l'émergence de la fonction**. Seulement cette émergence peut se présenter dans une fenêtre temporelle passée, présente comme future.

Dorénavant dans quelle fenêtre doit-on se placer concernant certains gènes aux fonctions inconnues ? Si vous pensez comme moi l'avenir nous réserve encore bien des surprises...

VI/ DISCUSSIONS

1. Origine de la génomique...

Alors que les objets physiques n'obéissent qu'à des lois, les êtres vivants ont en plus une histoire. Celle-ci est portée par le matériel héréditaire dont le support est l'ADN.

Connaître ce matériel héréditaire et son fonctionnement c'est donc lire l'histoire des êtres vivants, comprendre leur complexité et, finalement, appréhender ce qui nous distingue du monde inanimé. Cette approche est en passe de devenir possible à une échelle inaccessible auparavant grâce à **l'analyse des génomes**.

La génomique permet d'envisager un accroissement considérable d'échelle. Non seulement on pourra connaître la séquence standard du génome d'une espèce, par exemple l'homme, mais pourquoi pas aussi celle des individus qui la composent.

2. L'Analyse du génome

Bien souvent à cause de leur taille ou des difficultés inhérentes à leur complexité, tous les génomes ne sont pas nécessairement séquencés de manière complète. Ainsi dans un séquençage aléatoire, la couverture C est donné par le rapport $C = NL/G$ (L : longueur moyenne de chaque lecture, N : nombres de lectures et G pour la taille du génome). Une couverture standard est de l'ordre de 10X. Cependant avec une couverture incomplète il demeure toujours des trous ou zones de basse qualité.

Un génome séquencé contient en trace l'histoire de ses ancêtres. Et pourquoi pas également des indices sur ses possibilités évolutives futures? Même si cela était vrai, saurions-nous vraiment un jour interpréter pleinement ce texte génomique ? Combien d'informations nous échappent ?

Ces informations peuvent nous échapper soit parce qu'elles sont tout simplement absentes des banques de données (trous dus à une couverture insuffisante pour les détecter) ou bien parce que nos critères de sélection les excluent de l'analyse : par exemple si le critère est la taille minimale d'une séquence nucléique, les gènes de séquence courte peuvent être alors écartés de l'étude. Ainsi chez la levure, même dix ans après avoir séquencé son génome, on en est encore à modifier le nombre de gène car on en avait oublié quelques centaines, surtout les plus courts.

3. Haut débit et pertinence

Ce n'est pas tant le haut débit des expériences mais le caractère complet de la connaissance apportée qui définit une étude génomique. Evidemment automatisation et haut débit sont souvent nécessaires pour atteindre cette complétude mais ils demeurent l'outil et non le but recherché.

4. Haut débit et réseaux d'informations

Ces études globales permettent d'étudier simultanément un nombre d'éléments beaucoup plus important. Avec cette multiplication de données complémentaires (**séquences génomiques, transcriptomes, protéomes**) nous sommes maintenant en passe de poser des questions biologiques en terme de larges réseaux d'interactions globaux et non plus seulement de quelques interactions spécifiques. Ainsi l'accumulation de très larges collections de données sur les réseaux transcriptionnels, les interactions protéiques, les complexes macromoléculaires ou les interactions génétiques, permettront d'envisager la **modélisation des interactions dynamiques** qui ont lieu dans la cellule et d'imaginer leur évolution.

5. Haut débit et convergence de nouvelles technologies

La convergence de technologies vient au service de ces approches globales. En effet cette accélération des recherches a été rendue possible par un développement spectaculaire de technologies plus performantes et automatisables. Certaines ont été apportées à la biologie par

d'autres disciplines comme l'électronique, la micromécanique, la robotique, la chimie, la physique est bien sûr l'informatique.

Compte-rendu n°2 :

Le domaine qui s'étend de la génomique à la protéomique s'intéresse à l'étude du génome, aux produits de son expression (ARN puis protéines) et à la dynamique des différents modules (ADN et protéines) entre eux dans la régulation de la transcription, de la traduction, du métabolisme cellulaire,...

Nous venons de voir les différentes techniques qui permettent de caractériser un génome d'étude. L'étude *in silico* ou étude informatisée d'un génome permet de localiser les gènes le long du génome séquencé : sont recherchés la présence de signaux de réplication (TATAAT-box, séquence de terminaison Rho-indépendante chez les bactéries), des ORF, des promoteurs, des accepteurs et donneurs du mécanismes d'excision des introns pour déterminer les parties codantes. Seulement, la seule analyse des gènes d'un génome ne permet pas de prévoir sa fonction. Pour étudier la fonction et le profil d'expression de chaque gène il est nécessaire d'étudier les produits dérivés, les ARN (ARNt, ARNr et ARNm). Cette étude est l'objet de l'analyse du transcriptome

L'analyse du transcriptome est notamment utilisé en virologie (cas du VIH) pour suivre la progression de la maladie. La technique microarray est très utilisée : sur un support nylon, les clones de cDNA correspondants aux ARNm codants des protéines virales contenus dans le sérum d'un patient par exemple. Ces cDNA sont hybridés avec des cDNA marqués dont la séquence matche avec un cDNA codant un peptide viral par exemple ou bien avec des molécules protéiques médiateurs de la réponse immunitaire (lymphokines, chimiokines,...), etc. . Ainsi peut-on suivre par exemple la charge virale d'un patient suite à une trithérapie. Les clones de cDNA hybridés aux sondes sont quantifiés et analysés par informatique. Cette méthode remplace la migration sur gel d'ARN ou d'ADN amplifié par méthode PCR. Cependant, si l'analyse concerne un échantillon de composition simple alors cette méthode est bien moins compétitive que l'analyse par PCR en temps réelle. De plus, on est limité par le nombre de gènes régulés. Toutefois si l'échantillon est de composition complexe, alors l'analyse sur puce à ADN est un avantage non négligeable car elle permet d'analyser un grand nombre d'ARNm.

Par ailleurs, nous venons de voir dans les conférences précédentes que l'analyse du transcriptome ou de l'ADN génomique des bactéries étudiées permet une comparaison des génomes bactériens. Par exemple, pour typer la bactérie *Helicobacter pylori*, bactérie incriminée dans la gastrite, les chercheurs utilisent des plaques microarray à partir desquelles

ils analysent les produits de PCR de chacun des ORF de chaque souche d'*Helicobacter pilori* ou isolat clinique : ils marquent de façon aléatoire chacun de ces ORF avec des octamères de nucléotides fluorescent qui s'hybrident aux ORF. Après analyse, on identifie les gènes spécifiques à chaque isolat bactérien.

Nous venons de voir également qu'il est possible de détecter la présence d'espèce de bactérie dans un échantillon en suivant l'ARN ribosomique 16S. Par le même procédé que précédemment, sur puce à ADN, des sondes spécifiques de l'ARN 16S permettent de détecter certaines espèces de bactéries pathogènes.

Néanmoins, il n'existe pas toujours de corrélation directe entre présence de l'ARNm et abondance de la protéine codée. Il est donc important d'étudier l'expression de la protéine qui en dérive. Cette technique peut être utilisée à des fins de diagnostics. La protéomique est l'étude quantitative et qualitative des protéines exprimées dans un tissu à un moment donné et dans des conditions particulières. Nous venons de voir dans les conférences précédentes que les méthodes RMN et cristallographie sont des techniques permettant l'analyse d'une protéine (anticorps ou autres peptide) ou d'un complexe protéique purifié. Ces méthodes donne une structure 3D de la protéine étudiée. La RMN, permet également d'étudier la dynamique de la protéine car l'objet étudié est en solution alors que la cristallographie nécessite que l'objet soit cristallisé, ce qui rend impossible l'étude dynamique. De plus, en RMN, l'analyse est limitée par la taille de l'objet étudié.

Pour outrepasser ces difficultés, la bioinformatique est une alternative non négligeable. En effet, elle permet à partir de données sur la séquence protéique et la nature des acides aminés qui la compose, de prédire la conformation en trois dimensions. Elle peut également prédire les fonctions de cette protéine dans une chaîne métabolique, dans un processus de régulation cellulaire. Ces prédictions sont le point de départ pour des recherches expérimentales faites sur la protéine et sa fonction. Par cette technique, il est possible d'étudier la conformation d'un anticorps ou d'un paratope dans l'organisme où on a injecté un anticorps thérapeutique par exemple et de prévoir les interactions avec sa cible en incluant différents paramètres environnementaux.

Par ailleurs, la société CIPHERGEN a mis au point un kit de détection de protéine (anticorps,...) qui allie la chromatographie d'affinité à la spectrométrie de masse. L'échantillon analysé (produits biologiques humain tels que sérum, salive, urine) est déposé sur une surface chromatographique contenant des molécules liants telles que des récepteurs, des anticorps, ...marqués par fluorescence. Les protéines contenues dans l'échantillon (anticorps, ligands tels que des chimiokines, lymphokines par exemple) vont se lier aux

molécules liants de la surface suivant leur propriété physico-chimique. L'analyse de la fluorescence va permettre d'identifier exactement la nature de la protéine liée. Ensuite, les protéines fixées vont être analysées en spectrométrie de masse : le temps de vol de chacune de ces protéines liées va permettre de préciser la masse moléculaire de chacune. Cette technique est limitée par le nombre de molécules pouvant se lier sur la surface chromatographique.

Enfin, il est possible de suivre le profil d'expression d'un marqueur protéine à la surface membranaire d'une cellule de l'immunité par cytométrie de flux, en vue, par exemple, de suivre l'évolution d'un sous-groupe lymphocytaire au cours d'une maladie ou suite à un traitement thérapeutique. Nous avons vu, dans le cadre du suivi des patients atteints de SIDA, que cette analyse permet par exemple de suivre le profil d'expression des récepteurs à l'IL-7, une lymphokine impliquée dans la lymphopoïèse au niveau des tissus lymphoïdes centraux et périphériques, ainsi que dans l'activation et la prolifération et la survie des lymphocytes T naïfs.

Annotation in silico de séquences génomiques bactériennes

Le séquençage de génome procaryote puis eucaryote a fourni une masse de données considérable qu'il est crucial d'annoter pour une meilleure gestion de l'information. L'annotation d'une séquence génomique consiste à localiser les gènes sur un génome d'étude (annotation syntaxique), puis à assigner une fonction biologique à chaque gène (annotation fonctionnelle) et enfin, à élucider les relations moléculaires entre chaque module, ADN et protéine (annotation relationnelle).

L'annotation syntaxique consiste à repérer les gènes directement sur le génome en recherchant des séquences signales du génome bactérien (promoteur de la transcription, signal de terminaison de la transcription Rho indépendant) ou indirectement via des signaux du transcrit (Ribosome Binding Site ou RBS, codon start AUG, codon stop UAG) : ces signaux sont caractérisés par leur motif consensus, établi à partir de séquences bien définies dont on sait qu'elles contiennent le signal à caractériser. Celui-ci est appelé jeu d'apprentissage : il est généralement issu de résultats expérimentaux. Ces signaux peuvent être également repérés en évaluant la position d'une base à une position « i » du motif via un « tableau poids-position ». Sinon, l'utilisation d'algorithmes s'avère une alternative pour inférer le consensus de la séquence signale à caractériser. Mais l'identité des bases ne suffit plus lorsqu'il s'agit de signaux structuraux (cas de la boucle de terminaison de la transcription Rho indépendante).

L'étape de prédiction des gènes fait suite à l'étape de recherche de signaux. Elle consiste à prédire des régions codantes en recherchant les plus grandes phases ouvertes de lecture (Open Reading Frame ou ORF, située entre deux codons de terminaison de la traduction) d'au moins 300 nucléotides pour que la probabilité qu'elle soit dû au hasard soit faible. On peut ainsi plus facilement déterminer le codon de démarrage de la traduction. Sinon, d'autres modèles statistiques étudiant la fréquence d'apparition d'oligonucléotides spécifiques permettent de discriminer les régions codantes des non codantes.

L'annotation fonctionnelle consiste principalement à aligner la séquence de la protéine prédictive avec les banques de protéines déjà connues (type de programme : BLAST). La séquence doit présenter au moins 50% d'identité avec au moins 80% des séquences de protéines de la banque pour opérer à un transfert de similarité de la fonction biologique

présumée. Toutefois, on a recours aux comparaisons avec des banques de motif pour déterminer l'architecture quaternaire éventuelle de la protéine (ex : banque PFAM).

L'annotation relationnelle cherche à identifier les relations entre les objets génomiques en explorant les voies de régulation, le voisinage chromosomique, les voies métaboliques. Un voisinage est caractérisé par le rassemblement d'objets génomiques (gènes, protéines) dans un espace donné. Le cas le plus évident est l'opéron dont les gènes proches les uns des autres sont tous reliés fonctionnellement et co-régulés. L'analyse des similarités des gènes ou des protéines entre plusieurs espèces bactérienne permettent également d'explorer un autre type de voisinage : celui de l'espace de similarité de séquence à l'origine de la définition des familles de paralogues (spéciation d'un gène au cours de l'évolution) et d'orthologues (duplication d'un gène au cours de l'évolution). Ces familles sont décrites par les bases de données telles que COG et WIT. Cette analyse permet de caractériser des groupes de syntonie bactérienne : ce sont des groupes de gènes dont les séquences, les positions sur le chromosome, leur distances les uns par rapport aux autres sont conservés chez plusieurs espèces bactérienne durant l'évolution.

L'intérêt de cette conférence a été de prendre conscience des bioinformatiques comme un moyen complémentaire des études *in vitro* et *in vivo* pour appréhender les relations entre les différents objets de la génomiques. De plus, elle aide l'expérimentateur car elle est capable d'émettre des modèles prédictifs qui ne seront vérifiable que par l'expérimentation. Cette méthode pourrait être mise à profit dans les études d'interactions et les conséquences prédictives lié à la vaccination pour mieux comprendre et anticiper les effets.

Approches protéomiques pour la mise au point de nouveaux médicaments anti-herpès

Le virus de l'herpès est un virus enveloppé possédant une capsid e écosaédrique qui renferme de l'ADN double brin : celui-ci se compose d'une région courte (US) et d'une région longue (UL) liées covalamment. Entre l'enveloppe et la capsid e, il existe une zone tégumentaire qui contient des protéines virales telles que VP16 déversées dans la cellule au début de l'infection. Il existe deux types de virus de l'herpès : le virus labial, HSV1, et le virus génital, HSV2.

Le virus infecte les cellules épithéliales permissives de la bouche puis remonte l'axone qui les innervent jusqu' au corps cellulaire où il reste latent. Un stress (UV, température,...) induit sa migration inverse vers ces cellules épithéliales où il se réplique activement. Dans la cellule infectée, le virus exprime séquentiellement des transactivateurs transcriptionnels α activée par VP16, des enzymes de la réplication de l'ADN (expression précoce) et des protéines de structure virale (expression tardive). Son entrée dans la cellule entraîne l'inhibition ou l'activation de protéines cellulaires. Celles-ci sont analysées en protéomique pour les identifier et les utiliser comme cibles thérapeutiques pour développer des antiviraux innovants.

Ces traitements sont cruciaux car les manifestations herpétiques sont en recrudescence : en effet, l'augmentation du nombre de personne immunodéprimées, du nombre de greffe d'organe et de moelle osseuse, favorise l'implantation du virus qui peut entraîner la mort. Le seul traitement actuel est l'administration d'Acyclovir, un analogue de nucléoside qui inhibe la réplication virale. Or, il a été constaté chez ces patients une résistance au traitement dû à l'apparition de souche virale résistante (dans 5 à 30% des cas). D'où un besoin médical important pour contrer ces résistances.

Pour évaluer le taux de protéines stimulées par l'entrée du virus dans la cellule, on infecte une lignée de cellules épithéliales par le virus : cette infection entraîne un processus de syncytialisation des cellules visible au microscope. Un marquage métabolique court à la méthionine³⁵S à différents temps permet d'évaluer la radioactivité incorporée dans les protéines. Les protéines sont ensuite séparées par gel polyacrylamide bidimensionnel.

Seules 28 protéines sur les 183 analysées sont maintenues voire stimulées : la SAMDC (décarboxylase) impliquée dans la voie de synthèse des polyamines (spermine et spermidine) intervenant dans la synthèse des acides nucléiques est principalement activée.

Pour tester les effets induits par l'inhibition de SAMDC, on ajoute l'inhibiteur en présence du virus sur une culture de cellules épithéliales : on parvient à inhiber la réplication du virus car les formes syncytiales disparaissent. Cette inhibition est levée si on injecte aux cellules de la spermine et/ou spermidine : ces résultats montrent effectivement que l'inhibiteur est bien spécifique de la SAMDC et que la réplication virale est SAMDC-dépendante.

Pour visualiser les remaniements intranucléaires et le détournement des protéines viro-induites, les cellules sont transfectées pour coder des protéines marquées à la GFP. La protéine US11 (Uniq Short) spécifique du virus a permis de suivre les migrations intracellulaires du virus. Ainsi, certaines zones d'exclusions chromatiniennes sont observées à l'endroit où le virus se réplique. Ensuite, les virions s'accumulent au niveau du cytoplasme et entraînent la mort cellulaire par nécrose. Suite à ces observations, des molécules inhibitrices de la réplication ont été testées. Le dextran fixe la protéine US11 qui empêche le virus de passer la membrane nucléaire. US11 interagit également avec la protéine HIPK2 qui adopte alors une nouvelle conformation stimulable par UV. Cette stimulation entraîne l'apoptose des cellules viro-induites. Il serait donc intéressant de développer des molécules qui ciblent HIPK2, et empêchent sa fixation avec US11. En complément de cette méthode, on peut suivre la dynamique intranucléolaire des protéines fusionnées à la GFP lors de l'infection par des techniques de FRAP et de FLIP. En effet, en 1960, on a mis en évidence que le nucléole est le lieu de biosynthèse des ribosomes et de séquestration de certaines protéines à certains moments du cycle cellulaire pour réguler des processus du cycle biologique. On a notamment remarqué que la nucléoline est complètement réorganisée au moment de l'infection et que la protéine B23 reste dans le noyau et se relocalise dans les régions de réplication du virus.

Ce travail amorcé, l'équipe de recherche a réalisé une analyse protéomique poussée du nucléole pendant l'infection versus des cellules non infectées. La méthode a donc consisté à purifier le nucléole (dépourvu de membranes) et les protéines de la fraction nucléolaire sont ensuite extraites. Celles-ci sont séparées en gel bidimensionnel et analysées par spectrométrie de masse. On élabore ensuite des cartes protéomiques. Les protéines dont la synthèse est activée ou inhibée sont séquencées, leur fonction est analysée par bioinformatique.

Cette conférence a été l'occasion d'un tour d'horizon sur les moyens d'identifier les protéines cruciales pour la réplication du virus (marquage GFP, spectrométrie de masse, établissements de cartes protéomiques) : elles sont incontournables pour identifier de nouvelles cibles thérapeutiques pour une immunothérapie ou pour discriminer et suivre un sous-groupe de population cellulaire infecté par le virus.

Génomique et mitochondrie

La mitochondrie est apparue au moment de la rencontre d'une archaebactérie ou d'une cellule eucaryote avec une α protéobactérie (symbiose). La cellule résultante est capable d'utiliser l'oxygène via ses mitochondries. Elle réduit l' O_2 en H_2O . Par ailleurs, elle synthétise des radicaux libres via le cycle de l'ubiquinone. Elle permet également de réguler la concentration calcique intracellulaire. Enfin, elle est le siège de la synthèse de complexes fer-soufre qui constituent des coenzymes essentielles pour la vie de la cellule. Certaines cellules eucaryotes sont dépourvues de mitochondries et fabriquent leur complexe fer-soufre dans des mitosomes (cas de *Gardia intestinalis*) ou bien elles possèdent des machineries d'adressage dans des hydrogénosomes (cas de *Trichomonas vaginalis*). Ces trois organelles ont probablement une origine commune endosymbiotique. Mais à la différence de la mitochondrie, les

hydrogénosomes ainsi que les mitosomes ont perdu leur génome et n'ont pas de chaîne respiratoire.

La taille du génome mitochondrial varie de 2,4 à 6 kb mais il n'y a pas de relation directe entre la taille de ce génome et le nombre de gènes mitochondriaux. Sur l'échelle de l'évolution, il y a eu un transfert de gènes de la mitochondrie vers le noyau via des ARN de transfert et/ou via un processus global de transfert et d'insertion dans l'ADN nucléaire. Ce transfert a été mis en évidence par la présence dans le noyau de séquences datant de 4 à 6 millions d'année.

Pourquoi un transfert privilégié de la mitochondrie vers le noyau ? Pourquoi tous les gènes mitochondriaux n'ont pas été transférés au cours de l'évolution ? La chaîne respiratoire mitochondriale contient deux protéines membranaires, CoxI et Cytb. Ces protéines sont adressées à la membrane mitochondriale sous forme dépliée (du fait de leur hydrophobicité) par des protéines chaperons : cette hydrophobicité est donc un frein au transfert vers le noyau. Pour étudier la biogenèse des complexes protéiques mitochondriales, on procède à un arrêt de la traduction par Cycloheximide puis lyse cellulaire. On montre ainsi par migration sur gel agarose que certains ARNm nucléaires à localisation mitochondriale sont traduits sur des polysomes au voisinage des mitochondries et d'autres sont traduits librement sur une chaîne de polysome. Une approche plus globale a déterminé par des puces à ADN quels sont les gènes nucléaires qui émettent des ARNm à destination de la mitochondrie. Les ARNm associés aux polysomes et ceux associés aux mitochondries sont marqués, rétrotranscrits et hybridés sur des puces à ADN. En établissant un score, le MLR (Mitochondrial Localisation of RNA), il en résulte une forte proportion de messagers associés au polysomes au voisinage de la mitochondrie et une faible proportion d'ARNm associé aux polysomes libres dans le cytoplasme. Des expériences complémentaires de marquages à la GFP ont permis de mieux suivre le trajet de messagers nucléaires vers la mitochondrie.

Les signaux d'adressage des messagers à destination de la mitochondrie ont été recherchés. Par des expériences de délétions successives de la séquence 3'UTR de gènes (cas du gène ATP3) qui adresse le messenger à la mitochondrie ou par des expériences de substitution du 3'UTR de ATP3 par un 3'UTR hétérologue ont montré que cette région 3'UTR est cruciale pour l'adressage de ces ARNm.

Il a également été montré que les ARNm codants de grands ensembles moléculaires mitochondriaux comme l'ATP synthase sont traduits au voisinage de la mitochondrie. Alors pourquoi certains ARNm sont préférentiellement traduits dans le cytoplasme et d'autres au voisinage de la mitochondrie ? Par puce à ADN, PCR quantitative, il a été montré que les gènes nucléo-mitochondriaux d'origine procaryotiques (dans l'évolution) codent pour des éléments du cœur des complexes tels que l'ATP synthase et sont traduits au voisinage de la mitochondrie.

Cette conférence a été l'occasion de voir des procédés tels que la puce à ADN, outil crucial notamment en immunologie...

Laboratoire de cristallographie et RMN biologiques

La biologie structurale est l'ensemble des méthodes qui permettent d'accéder aux composantes atomiques et moléculaires des constituants cellulaires et des organismes vivants et de comprendre leur fonctionnement atomique et la relation structure/fonction. Ces méthodes sont au nombre de deux : la diffraction aux rayons X qui permet une représentation 3D de l'objet quelque soit sa taille. Cependant celle-ci nécessite que l'objet soit cristallisé, elle n'utilise pas de lentilles à rayons X (obligation d'utiliser la transformée de Fourier) et la structure doit être figée. La seconde méthode est la RMN : il s'agit de représenter en 3D, comme la cristallographie, l'objet analysé qui est en solution. Cette méthode permet de

travailler en présence d'atomes marqués pour la visualisation des échanges en solution. Cette méthode est utilisable en biologie analytique pour savoir si une protéine est bien repliée. Mais elle est limitée par la taille de l'objet (échantillon de l'ordre du mmolaire, ce qui entraîne un coût non négligeable) et nécessite un bon rendement signal/bruit. De plus, les protéines étudiées doivent rester solubles.

Pour travailler dans les meilleures conditions, il faut des quantités importantes de protéines purifiées dans l'échantillon (~dizaines de mg) : pour ce faire il faut exprimer la protéine dans un vecteur d'expression particulier : E. coli, levure, baculovirus, système d'expression cell-free tel que le système de Roche. Ensuite, la protéine doit être pure : la protéine est taggée pour être fixée spécifiquement après passage en colonne d'affinité. Après fixation, la protéine est libérée par une protéase qui coupe le lien entre l'étiquette et la protéine (ex d'étiquette : la glutathion S transférase). Par ailleurs, il est nécessaire d'obtenir un haut rendement.

Plusieurs analyses sont ensuite effectuées pour tester la nature de(s) protéines purifiées. Les tests biochimiques (HPLC, spectrométrie de masse,...) effectués sur les protéines permettent de déterminer la pureté de la protéine. On vérifie également, le poids moléculaire, le point isoélectrique, la glycosylation, l'organisation de la protéine en domaines oligomériques. Tous ces paramètres sont testés sur les protéines à purifier. Par ailleurs, on teste la solubilité de la protéine: il s'agit de la quantité de matière qu'on peut dissoudre dans un volume donné. Cette solubilité est fonction de la température, du pH, du solvant : à partir de ces paramètres on peut construire une courbe de saturation divisée en 2 (parties sous-saturation et sur-saturation). Pour une analyse en RMN, il est indispensable d'être en sous-saturation, et en sur-saturation pour une analyse par cristallographie. L'allure de la courbe est spécifique de la protéine analysée.

Autre domaine faisant appel à la cristallographie et à la RMN : la génomique structurale. Elle fait appel aux compétences des microbiologistes, des biologistes moléculaire et cellulaire et des biochimistes, des cristallographes et RMN, ainsi que des bioinformaticiens. Ces derniers sont chargés d'analyser le génome et de sélectionner les ORFs en vue d'étudier les gènes présents sur un génome d'étude. Des programmes de génomiques structurales sont lancés aux USA (start-up TRIAD Therapeutics, San-Diego, travaux sur la kinase), au Japon et en Europe (France : plateforme de cristallogénèse à l'Institut Pasteur dont l'objectif est de mettre en place du clonage à haut débit, faire de l'expression et de la purification protéique, du contrôle qualité et de la cristallisation automatique et de l'imagerie du cristal ; projet levure à Orsay : 6213 ORFs dans le génome).

Cette conférence nous a exposé les différents points qui caractérisent la méthode de purification protéique et d'analyse par la méthode RMN. Cette méthode peut-être employée dans la purification d'un anticorps, de chimiokines, lymphokines,... destiné à la recherche fondamentale ou au domaine thérapeutique. Des problèmes subsistent lorsque l'on produit des molécules thérapeutiques dans un modèle bactérien : ce sont les problèmes de glycosylation. En effet, un mauvais profil de glycosylation peut être à l'origine de réponse immunogène rencontrés dans le cas de l'administration d'un anticorps ou adjuvant thérapeutique. La RMN permet de mettre en évidence ce problème.

Les Archaeobactéries et la génomique

Le premier caractère constitutif qui distingue les Archaeobactéries des bactéries est la présence d'une monocouche membranaire qui résiste à des températures allant jusqu'à plus de 105°C contrairement aux bactéries munies d'une membrane plasmique constituée d'une bicouche lipidique sensible aux températures extrêmes. Cette différence tient également de la stéréochimie des lipides. Le carbone central du glycérol composant les lipides membranaires est asymétrique chez les Archaeobactéries contrairement aux bactéries. Enfin, dans les lipides membranaires des Archaeobactéries, les acides gras sont toujours reliés au glycérol par une liaison éther. Au point de vue phylogénétique, ce groupe est divisé en deux branches : les

Euryarchaeota (ou Euryotes, Euryos pour « divers ») et les Crenarchaeota (ou Crenotes, Crenos pour « source, origine »).

Le premier séquençage d'une Archaeobactérie a eu lieu en 1996. La génomique a confirmé l'existence de hits au niveau des gènes entre les Archaeobactéries et les bactéries (exemple : 17% des gènes de *Thermophilus* sont retrouvés chez la bactérie). Il peut y avoir eu transfert entre les bactéries et les Archaeobactéries (exemple : les Archaeobactéries ont appris à respirer en présence d'oxygène au contact des bactéries via un transfert de gènes).

Cependant, de nombreux gènes sont similaires à ceux présents chez les eucaryotes. En effet, leur ARN polymérase et les sous-unités qui les constituent sont semblables à celle des eucaryotes. Vice-versa, les bactéries expriment des protéines ribosomales qui ne sont pas retrouvées chez les Archaeobactéries et les eucaryotes.

Un des problèmes majeurs est la culture de ces Archaeobactéries. Certaines ne vivent que dans un environnement particulier: c'est le cas des Archaeobactéries méthanogènes qui vivent au cœur des champs agricoles, dans l'intestin et les gencives. Par ailleurs, toutes les Archaeobactéries sont anaérobies strictes ce qui complique leur étude. D'autres Archaeobactéries sont halophiles (culture entre 4 et 5M de NaCl chez les Euryotes), thermoacidophile (culture entre 50 et 70°C à pH 0-1 pour les Euryotes, et 70 à 90°C à pH 2-3 pour les Crenotes), hyperthermophiles (culture entre 80 et 110°C dans les sources chaudes sous-marines et terrestre). Cependant, plus la température est élevée plus la perméabilité aux protons et aux ions est élevée : ceci entraîne la mort des Archaeobactéries car leur métabolisme énergétique est entravé, ce qui pose des difficultés pour leur culture.

Au point de vue génomique, les Archaeobactéries possèdent un petit génome circulaire présentant un gène tous les 1000 paires de bases. Les régions intergéniques sont courtes et il n'y a pas d'introns. Cette structure est similaire aux bactéries. Pourtant la machinerie de réplication ne ressemble en rien à celle des bactéries (DnaA, DnaB, DnaG chez les bactéries n'ont pas d'homologues chez les Archaeobactéries). Par contre, les enzymes de la réplication telles que CDC6, Mcm, Dnapol, PCNA, primase, RPA sont retrouvées à la fois chez les Eucaryotes et les Archaeobactéries.

Pour localiser l'origine de réplication (Ori C) ainsi que la séquence de terminaison de la réplication (Ter C), les outils d'analyse *in silico* ont été utilisés. Le brin leader correspond au brin riche en G comparé au brin retardé. En intégrant le biais en GC entre deux points arbitraires du génome, on peut théoriquement déterminer la position de l'Ori C et de Ter C.

Ensuite, pour déterminer la séquence nucléotide où se fixe les constituants du complexe enzymatique de la réplication chez les Archaeobactéries, les chercheurs immunoprécipitent un constituant protéique du complexe (ex : utilisation d'un anticorps anti-CDC6). L'ADN est traité à la nucléase pour détruire l'ADN nu (i.e. sans protéines fixées à l'ADN). L'ADN resté fixé à la protéine CDC6 est donc cross-linké. L'ADN est hybridé avec une sonde et transféré sur une puce à ADN. Ainsi, par un code couleur, on peut déterminer avec précision si CDC6 s'est fixé à son site de fixation sur l'ADN et déterminer précisément la séquence nucléotidique reconnue par CDC6.

Cette conférence a été l'occasion de montrer comment utiliser l'outil anticorps en association avec la puce à ADN pour mettre en évidence des sites de reconnaissance par des protéines sur l'ADN.

Transferts génétiques horizontaux et plasticité des génomes bactériens

La variation génétique ou plasticité génétique est le résultat de transfert de gènes qui s'accompagne de la présence d'enzymes de transposition et de recombinaison.

Il existe une corrélation entre la capacité adaptative d'une bactérie vis-à-vis de son environnement et la taille du génome : plus le génome est grand plus la capacité adaptative est élevée (les bactéries parasites obligatoires et intracellulaires ont de petits génomes). La taille

du génome est également lié au nombre de séquence d'insertion ou IS : plus le génome est grand moins il y a d'IS.

Le génome bactérien évolue par le fruit des mutations, des duplications de gènes et de l'acquisition de gènes exogènes par transfert horizontal. La part d'ADN exogène chez la bactérie est de l'ordre de 15%. L'amplitude du transfert varie en fonction de la niche bactérienne : les bactéries de l'environnement acquièrent de nombreux gènes par transfert.

La variation génomique au sein de l'espèce *E. Coli* est de l'ordre de 20% : la plupart de l'ADN additionnel incorporé suite à un transfert est souvent hébergé dans les prophages.

La détection de l'ADN exogène se fait par analyse comparative du génome A avec un génome de référence B (A et B étant des génomes d'espèce contemporaines), par incongruence phylogénétique et par détections des bornes caractéristiques des différents mobiles (ou ADN transféré). L'analyse comparative consiste d'une part à examiner le pourcentage de G et de C d'un génome A comparé avec celui d'un génome B puis d'évaluer les ruptures de la composition en G, C entre ces deux génomes : lorsqu'il existe une rupture alors cela signe une origine exogène. Pour savoir si cet ADN exogène est d'origine bactériophagique on évalue la composition en A, T du génome sachant que l'ADN phagique est plus riche en A et en T que le génome bactérien. L'incongruence consiste à étudier l'ARN 16S (protéine ribosomique sujette à transfert) dont la topologie très solide permet de dresser un arbre phylogénétique. La recherche des bornes consiste à rechercher des séquences répétées inversées en orientation inverse ou directe en association avec des gènes de transposases ou de recombinases (tyrosine ou sérine recombinase). On sait que ces séquences ont moins de G et de C que la moyenne du génome.

Les mécanismes de transferts d'ADN exogènes sont la transduction ou échange de matériel génomique bactérien par infection d'un phage ayant incorporé de l'ADN bactérien d'une bactérie A vers une bactérie B de la même espèce. Dans ce cas l'ADN exogène peut s'incorporer au génome par recombinaison homologue. Toutefois des phages d'une espèce bactérienne peuvent infecter une autre espèce pour transmettre l'ADN exogène qui peut s'incorporer au génome via des séquences d'insertion IS ou des transposons. La transformation est un mode très efficace mais rare de transfert d'ADN exogène. Enfin, la conjugaison par contact entre une bactérie donneuse et une bactérie receveuse est un mode de transfert très efficace et très répandue qui peut avoir lieu sur de longue distance. Peuvent être véhiculés par ce dernier mode de transfert des plasmides (tels que RK2 et RP4) et transposons dits conjugatifs. L'ADN exogène s'intègre ensuite par recombinaison homologue ou via les transposons.

Cette conférence nous a fait prendre conscience que le génome bactérien est doué d'une grande plasticité et qu'il est possible d'utiliser cette propriété à des fins de productions de molécules biologiques. Par exemple, il est possible de transférer à la bactérie telle que *E. coli* un plasmide contenant un gène d'intérêt (une chimiokine qui sera utilisée comme adjuvant dans le traitement immunothérapeutique d'une maladie) dont l'expression serait sous contrôle d'un promoteur inductible par l'ajout d'une substance chimique (tel que l'IPTG couramment employé en laboratoire) additionné d'un gène de résistance à une drogue. En présence d'inducteur on déclenche l'expression et la production de la molécule biologique d'intérêt.

Compte-rendu n°3 :

CONFERENCES SUIVIES

Introduction à la génomique et à la protéomique : aspects fondamentaux et technologiques

Odile OZIER-KALLOGEROPOULOS

1. Annotation *in silico* des séquences génomiques bactériennes
Claudine **MEDIGUE**
2. Transferts génétiques horizontaux et Plasticité des génomes bactériens
Didier **MAZEL**
3. Génomique comparative des levures : cycles sexuels et régions subtélomériques
Cécile **FAIRHEAD**
4. Analyse des complexes protéiques par les approches protéomiques
Bertrand **SERAPHIN**
5. Approches protéomiques pour la mise au point de nouveaux médicaments anti-herpes
Jean-Jacques **DIAZ**
6. Récents développements en génomique structurale
Arnaud **DUCRUIX**
7. Analyse « transcriptome » : bases théoriques et applications
Catherine **NGUYEN**
8. Etude de la réplication et de l'instabilité génomique sur molécules uniques d'ADN
Aaron **BENSIMON**
9. L'évolution des génomes viraux
Simon **WAIN-HOBSON**
10. Phylogénomique des Archaea
Patrick **FORTERRE**
11. Reconstruction du génome d'un vertébré ancestral
Hugues **ROEST-CROLLIUS**
12. Comment la génomique peut-elle aider à comprendre le pouvoir pathogène de la bactérie *Helicobacter pylori*
Hilde **DE REUSE**
13. Du génome à la cellule : comment prédire la fonction des gènes ?
Antoine **DANCHIN**

APPLICATIONS EN IMMUNOLOGIE

1. Protéomique clinique et diagnostique : illustration par l'approche SELDI-TOF
Rémi **Bocquentin**
2. Etude du transcriptome des cellules T CD4 et CD8 chez les sujets infectés par le VIH
François **BOUTBOUL**
3. ISEApeaks, une stratégie d'étude globale du répertoire lymphocytaire T : application à l'étude du neuropaludisme
Encarnita **FERRANDIZ** & Adrien **SIX**
4. Tissus Micro-Arrays (TMA) : application à l'étude des réponses anti-tumorales

NOTE DE SYNTHÈSE

La génomique est la science qui étudie les génomes des organismes. Elle comprend les études de la localisation des gènes sur les chromosomes, de leur nombre dans un organisme et de leur fonction, mais également l'étude de l'influence d'un gène sur un autre et l'étude de l'activation ou de la suppression des gènes. La génomique couvre également la recherche des gènes responsables des maladies.

La génomique est née dans les années 1970, grâce au développement de nouvelles techniques de séquençage de l'ADN, dont notamment la technique mise au point par Frederick Sanger. En 1977, le premier génome d'un organisme, le virus ψ x174, a été entièrement séquencé. Il faudra ensuite attendre les années 1990 pour voir les génomes des levures et des bactéries entièrement séquencés (*Haemophilus influenzae* en 1995), et plus particulièrement 1997 où le séquençage d'*Escherichia coli*, initié en 1987, arrive à terme. Depuis lors, l'engouement pour la génomique est tel qu'au 1er novembre 2005, on comptait 275 génomes procaryotes et 39 génomes eucaryotes séquencés, et 804 génomes procaryotes et 547 eucaryotes en cours de séquençage.

Pourquoi séquence t-on les génomes ? Il y a trois grandes raisons à ces séquençages. La première est d'approfondir les connaissances sur l'arbre de la vie, c'est-à-dire essayer de comprendre l'évolution des espèces vivantes depuis l'apparition de la vie sur Terre. Cet "objectif" ne représente que 3% des séquençages. Dans 47% des cas, le séquençage d'un génome donné a un but médical : on séquence ainsi les bactéries pour trouver les causes de leurs virulences. Enfin, dans 48% des cas, les séquençages ont un intérêt biotechnologique : on peut ainsi séquencer des bactéries pour l'industrie agro-alimentaire (le fromage contient des bactéries), les bactéries étant également des usines de fabrication des acides aminés utilisés dans l'alimentation. Les deux derniers points évoqués expliquent la faible proportion de séquençage de génomes d'Archae. En effet, ces derniers n'ont aucun intérêt en biotechnologie et ne sont pas responsables de maladie. Par contre, leur séquençage est essentiel pour comprendre l'évolution de la vie.

Par la suite, la génomique a donné naissance à la protéomique, qui désigne la science qui étudie l'ensemble des protéines d'un organisme. Dans la pratique, la protéomique s'attache à identifier les protéines extraites d'une culture cellulaire ou d'un tissu, leurs localisations dans les compartiments cellulaires ainsi que leurs modifications post-traductionnelles. Elle permet de quantifier les variations de leurs taux d'expression en fonction du temps, de leurs environnements, de leurs états de développement, de leurs états physiologiques et pathologiques ou de l'espèce d'origine. Elle étudie aussi les interactions que les protéines ont avec d'autres protéines, avec l'ADN ou l'ARN, ou encore avec diverses substances. La protéomique fonctionnelle étudie les fonctions de chaque protéine. La protéomique étudie enfin les structures primaires, secondaires et tertiaires des protéines.

La génomique s'appuie sur l'utilisation de nombreuses technologies parmi lesquelles les méthodes de séquençage, les outils de bioinformatique, les puces à ADN, les puces à protéines, la méthode du double-hybride, la spectrométrie de masse, la cristallographie, la biologie moléculaire et la biochimie des protéines, la génétique, la biochimie, la physiologie...

Mais quelles sont ses applications ? La génomique est une science qui peut-être appliquée à tous les domaines et à tous les niveaux de la biologie. Parmi ces applications, nous pouvons citer :

- ∞ L'arbre de vie : depuis son énoncé, la théorie de Darwin s'est progressivement imposée et il est à l'heure actuelle communément admis que tous les êtres vivants descendent d'un ancêtre commun. Avant l'apparition de la génomique, les travaux de classification des espèces étaient limités à l'échelle "macroscopique" : comparer leurs morphologies, leurs comportements et leur distribution géographique. La découverte du support de l'hérédité puis de l'universalité du code génétique et la disponibilité des séquences d'ADN et de protéines ont ouvert de nouvelles perspectives : c'est au niveau moléculaire que se recherchent à présent les "liens de parenté" unissant les différents organismes. Pour cette recherche, les outils informatiques permettent une analyse des différents gènes d'un génome entièrement séquencé mais également l'étude de gènes homologues chez un très grand nombre d'organismes et d'espèces différents. Dans tous les cas, des algorithmes de comparaison et d'alignement de séquences sont sollicités. Le principe est simple : il s'agit d'aligner plusieurs séquences codantes homologues, la "distance" entre ces séquences (c'est-à-dire leur degré de similitude) est alors proportionnelle à la distance relative des organismes correspondants sur l'arbre de vie. Cette nouvelle approche vient compléter les méthodes plus anciennes de phylogénie et les enseignements apportés par des disciplines comme la paléontologie, l'éthologie ou des études d'embryogenèse.

- ∞ Le domaine thérapeutique : la génomique peut être appliquée à un grand nombre de disciplines ayant pour but de comprendre le mécanisme des maladies. En effet, que ce soit dans le cas du cancer, d'infections bactériennes, virales ou parasitaires, ou encore dans le cas de maladies génétiques, ces maladies vont se traduire à un moment donné par un dérèglement de l'expression de certains gènes et/ou protéines. De plus, la génomique est un outil essentiel pour chercher de nouvelles cibles thérapeutiques. Nous pouvons citer, par exemple, la recherche des facteurs de virulence d'une bactérie pathogène. Le fait de connaître les protéines impliquées dans la pathogénicité d'une bactérie permet d'émettre l'hypothèse que si on les inactive, on inactive en même temps la virulence de la bactérie et de ce fait l'infection qui lui est associée. La génomique et la protéomique peuvent être également utilisées pour identifier des marqueurs spécifiques d'une maladie donnée (par exemple, la variation dans l'expression d'une protéine) et de ce fait en faire des marqueurs pour les diagnostics cliniques. La génomique joue un rôle important pour le diagnostic et le pronostic d'un cancer. Elle a d'abord ouvert le champ à la connaissance de facteurs prédisposant à certains cancers : mutations des gènes BRCA dans le cancer du sein, des gènes APC dans les cancers digestifs, etc. La génomique appliquée à la cancérologie a pour but d'élargir au maximum l'éventail des gènes de prédisposition en vue de les rassembler sur une même puce à ADN et ainsi en faire un outil indispensable pour le dépistage, le diagnostic et le suivi thérapeutique des patients. L'un des nouveaux axes de recherche de la génomique en cancérologie est la pharmacogénomique. En effet, dans une maladie où le taux global de succès des traitements reste faible, de nouvelles cibles thérapeutiques sont nécessaires. Il est également nécessaire de reconnaître les sous-groupes de patients répondant, ou non, à tel ou tel traitement. Par exemple, pour le cancer du sein, le nouvel anticorps monoclonal Herceptine n'est actif que si la tumeur mammaire traitée exprime le gène du récepteur correspondant (Her2neu). La génomique peut également être utilisée pour identifier de nouveaux gènes synthétisant une protéine d'intérêt mais aussi les gènes responsables de la résistance de certaines bactéries aux antibiotiques.

- ∞ Le domaine biotechnologique : la génomique tient une place importante dans le monde de l'industrie alimentaire. On peut prendre comme exemple la génomique végétale, actuellement en plein essor. Cette science a pour but d'isoler, de cloner et de manipuler des gènes utiles pour les plantes de grande culture afin d'améliorer leur qualité et leur résistance aux maladies, aux insectes et aux stress environnementaux.

L'UE d'analyse scientifique Génomique/Protéomique, constitué de conférences tenues par des professionnels du domaine ainsi que des cours d'application à l'immunologie, nous a permis de nous tenir au fait sur l'état actuel de la génomique et de la protéomique, de connaître les derniers développements de l'étude des génomes et de réfléchir aux implications de la connaissance des génomes dans tous les aspects de la biologie. Pour cela, les avancées de génomique et de protéomique nous ont été illustrées à l'aide d'exemples concrets et de points sur les recherches en cours.

Annotation *in silico* des séquences génomiques bactériennes

Claudine MEDIGUE, Génoscope

L'annotation *in silico* (c'est-à-dire réalisée à l'aide de logiciels informatiques) des séquences génomiques bactériennes, a pour objectif de donner un sens au texte génomique obtenu après séquençage.

Cette annotation se réalise en trois étapes, décrites ci-après :

- L'annotation syntaxique permet l'identification des gènes d'un organisme, c'est-à-dire de trouver leur localisation précise sur la séquence du génome. On détermine ainsi les ARNs, les signaux de régulation, les signaux de transcription (séquences de début et de fin), les signaux de traduction (codon d'initiation, séquence de fixation du ribosome, codon de terminaison), les CDSs (phase de lecture entre le codon START et le codon STOP), les ORFs (phase ouverte de lecture entre deux codons STOP) et les répétitions. Ceci se fait grâce à des alignements de séquences, à l'utilisation d'une matrice poids/position (donne la fréquence d'apparition d'une base donnée à une position donnée du motif) ou encore à l'utilisation combinée de la matrice poids/position avec une matrice de distance entre les acides aminés (pour une position donnée, on tient ainsi compte à la fois de la fréquence d'apparition des acides aminés mais aussi des substitutions possibles entre les acides aminés). On peut également utiliser la chaîne de Markov (modèle mathématique) pour prédire les séquences codantes, et les logiciels GeneMark et Glimmer pour calculer la probabilité des gènes sur les six phases de lecture.
- L'annotation fonctionnelle permet, dans un second temps, d'assigner une ou plusieurs fonctions biologiques à chaque gène hypothétique, et ce par comparaison de leur séquence avec celles de gènes qui ont une fonction déjà connue. Pour cela, on utilise des banques de données généralistes comme NCBI NR ou SwissProt, mais également des banques de données spécialisées comme PFAM, SMART, COGs ou encore EcoGene. On peut aussi prédire les structures 2D et 3D des protéines pour essayer d'établir une liaison entre les classes de repliements et la fonction de la protéine.
- Pour finir, l'annotation relationnelle consiste à identifier les relations entre les objets génomiques mis en évidence au cours des étapes précédentes. Ces relations peuvent être de natures très variées : interactions physiques protéines/ADN, protéines/ARN et protéines/protéines, réseaux de régulation de l'expression génique, voies métaboliques,

synténies (deux gènes d'un même organisme sont en synténie s'ils sont portés par un même chromosome) conservées entre organismes (les deux orthologues des gènes en synténie dans le génome A sont aussi en synténie dans le génome B). Ces relations seront mises en évidence grâce à des logiciels tels que le programme GénOLink.

L'intérêt du séquençage de ces génomes bactériens ainsi que de l'annotation des séquences obtenues est directement lié aux types de bactéries étudiées. En effet, le séquençage de bactéries pathogènes conduit à l'identification des gènes impliqués dans la virulence, ce qui peut amener à la fabrication de nouveaux antibiotiques ou vaccins (ex. *Yersinia*, *Listeria*). Le séquençage de bactéries modèles nous permet d'étudier des domaines biologiques clés aidant à la compréhension de mécanismes cellulaires (ex. *Escherichia coli* K12). Enfin, le séquençage de bactéries de l'environnement nous permet d'explorer la biodiversité et les fonctions biologiques développées au sein de différents écosystèmes, ce qui peut aboutir à la découverte de nouvelles voies métaboliques (ex. bactérie de type *Annomix* qui métabolise l'ammoniac).

Transferts génétiques horizontaux et Plasticité des génomes bactériens

Didier MAZEL, Institut Pasteur

A l'heure actuelle, 257 génomes bactériens sont déjà séquencés et 521 sont en cours de séquençage. La taille des génomes déjà séquencés varie entre 0,580 Mb (*Mycoplasma genitalium*) et environ 10 Mb (*cyanobacteria* et *actinobacteria*). Il y a une corrélation directe entre la capacité adaptative des bactéries et la taille de leur génome : par exemple, les bactéries parasites intracellulaires (environnement constant et riche) ont de petits génomes. Il y a également une variation dans le contenu et dans l'organisation de ces génomes bactériens, même entre espèces très proches. Cette variation dans la plasticité, qui a pour conséquence l'évolution des bactéries, se fait par trois voies différentes qui co-existent et agissent souvent en synergie : mutation dans le génome, duplication interne et acquisition génétique par transfert horizontal. Dans ce résumé, nous allons décrire brièvement les transferts horizontaux.

L'amplitude des transferts (quantité d'ADN pouvant être transférée) varie selon les espèces bactériennes. En effet, la part d'ADN étranger (porté ou non par des éléments mobiles comme les phages ou les transposons) peut aller jusqu'à 15% pour les gros génomes. Pour les petits génomes, ce pourcentage est beaucoup plus variable.

Ces transferts peuvent se faire via trois mécanismes différents : la transduction (transfert d'ADN bactérien d'une bactérie à l'autre par l'intermédiaire d'un bactériophage) qui peut être généralisée (transfert qui concerne n'importe quelle partie du génome) ou spécialisée, la transformation (pénétration dans une cellule bactérienne, naturellement transformable, d'ADN nu) et la conjugaison (processus spécialisé qui implique un transfert unidirectionnel d'ADN d'une cellule donatrice à une cellule réceptrice, par un mécanisme requérant un contact spécifique). Les principales caractéristiques de ces mécanismes sont indiquées dans le tableau ci-après :

	Transduction	Transformation	Conjugaison
Vecteur	Phages	ADN	Plasmides ou Transposons
Spectre d'hôtes	Intra-espèce	Intra et inter-espèce	Intra et inter-espèce
Efficacité	Importante	Importante	Variable
Distribution	Ubiquitaire	Restreinte	Ubiquitaire

L'identification de ces transferts peut se faire de quatre façons différentes, à savoir :

- L'analyse comparative de génomes. Elle se fait par comparaison entre les différents arbres phylogénétiques, ce qui va refléter les acquisitions mais également les pertes de gènes.
- L'étude des incongruences phylogénétiques c'est-à-dire l'étude du positionnement des gènes. Pour cela, on regarde le positionnement d'un gène donné dans deux génomes bactériens différents.
- L'identification des bornes des éléments mobiles. On détermine ainsi les insertions au niveau de gènes très conservés (ARNt), la présence de séquences répétées dans la même orientation ou inversées et la présence de gènes codant des recombinases ou des transposases.
- L'observation d'une déviation dans la composition en GC dans l'usage des codons. En effet, lorsque l'on observe une chute dans le pourcentage en GC, un transfert horizontal de gènes est souvent suspecté car en général, l'ADN exogène transféré est plus riche en AT que le génome endogène.

Analyse des complexes protéiques par les approches protéomiques

Bertrand SERAPHIN, CGM CNRS

L'analyse des complexes protéiques (interaction entre au moins deux protéines dont la stoechiométrie est connue) est beaucoup plus difficile que celle des génomes de part la structure même des protéines. En effet, les analyses ne peuvent pas être standardisées car la structure 3D des protéines les rend plus ou sensibles aux conditions et variations des propriétés physico-chimiques. De plus, elles sont limitées à la quantité de protéines que l'on a car leur amplification n'est pas possible. L'analyse se fera alors soit par des méthodes indirectes via l'étude du transcriptome (ensemble des ARNm d'une cellule à un moment donné) ou en développant des réactifs spécifiques (anticorps, protéines partenaires), soit par des méthodes directes via l'analyse du protéome (ensemble des protéines codées par une espèce dans des conditions et à un temps donnés).

L'étude du protéome se fait en trois étapes. Dans un premier temps, on réalise l'analyse qualitative des protéines, c'est-à-dire que l'on cherche à déterminer les protéines exprimées dans des cellules données, à un temps et dans des conditions connus. La seconde étape permet de les quantifier via des analyses quantitatives. Enfin, l'analyse fonctionnelle permet d'attribuer une fonction à chacune de ces protéines. Pour cela, on réalise des tests d'activité, on étudie leurs modifications (phosphorylation, glycosylation...) et les interactions entre les protéines. L'analyse qualitative peut se faire par des méthodes de spectrométrie de masse (ex. Maldi Tof) qui permet de déterminer la masse moléculaire des peptides présents dans un échantillon. Une fois cette masse déterminée, l'identification de la protéine se fera par des recherches dans les bases de données. On peut également utiliser la spectrométrie de masse en tandem qui permet d'obtenir la séquence en acides aminés du peptide et de remonter à la protéine grâce à ces mêmes banques.

La plus grande majorité des protéines n'agissent pas seules mais plutôt sous formes de complexes protéiques qui sont responsables de l'activité biologique. Ces complexes peuvent être purifiés à l'aide de diverses techniques comme la purification sur colonne d'affinité (il faut connaître les propriétés du complexe), l'immunoprécipitation grâce à un anticorps dirigé contre un composant du complexe ou encore la technique de Pull Downs. Cette dernière

consiste à marquer une sous unité connue du complexe puis à la mettre en présence du mélange contenant les autres sous unités. Le complexe va alors se reformer avec la sous unité marquée ce qui va permettre de l'isoler. On peut également utiliser la technique dite TAP (Tandem Affinity Purification) qui permet de purifier des complexes protéiques en utilisant des étiquettes et qui présente l'avantage de pouvoir purifier des complexes peu abondant dans la cellule. Les méthodes de double hybride et de puces à protéines (quantification possible des protéines) sont aussi utilisées pour purifier ces complexes. La structure 3D des protéines pourra être étudiée par cristallographie et le repliement de la protéine sera étudié par analyse aux rayons X. L'analyse fonctionnelle se fera notamment par l'étude de mutants (la mutation dans une sous unité du complexe entraînera la perte de l'activité biologique) ou par l'étude de la dégradation des ARNm spécifiques de ces protéines.

Les méthodes précédemment décrites ont pour but de caractériser les complexes protéiques. Pour un complexe donné, on pourra ainsi caractériser son activité, sa composition, la stoechiométrie de ses sous unités, sa structure, sa concentration et sa localisation cellulaire mais également ses modifications post-traductionnelles, ses partenaires (autres complexes, petits métabolites...) mais aussi les sous unités responsables des interactions.

Approches protéomiques pour la mise au point de nouveaux médicaments anti-herpes

Jean-Jacques DIAZ, INSERM

Le simplex Virus de type I (HSV-1) ou virus de l'Herpès, est un virus à ADN double brin de 150Kb, codant environ 80 protéines différentes. Au niveau de sa structure, il est composé de deux régions, UL (Unique Long) et US (Unique Short), bordées par des éléments répétés inversés. Il possède une enveloppe, synthétisée à partir de la membrane nucléaire, et est empaqueté dans une capsid virale. Entre la capsid et l'enveloppe se trouve la zone tégumentaire contenant les protéines virales jouant un rôle lors de l'infection (elles sont déversées dans le cytoplasme de la cellule hôte au début de l'infection).

Le HSV-1 infecte les jeunes enfants par des contacts physiques avec leurs proches. Cette primo-infection passe en générale inaperçue. Les particules virales infectent les cellules épithéliales, les axones qui les innervent puis les corps des neurones regroupés au niveau des ganglions trijumeaux. Le virus pénètre dans les noyaux des cellules et y reste latent en attendant un stress (émotion, accouchement, air de la montagne, stress thermique...) qui le réactivera. Cette réactivation entraîne la réplication de quelques particules virales qui vont retourner dans les cellules épithéliales initialement infectées. Le système immunitaire est alors activé et jugule l'infection (au niveau de la bouche, du haut du visage, des yeux...).

Il existe également un simplex virus de type II : le HSV-2, responsable de l'Herpès génital. Son cycle d'infection est le même que pour le HSV-1 à l'exception de son mode de transmission qui se fait par contact sexuel.

Lors de l'infection, le génome viral s'exprime et la protéine VP16 est produite. Cette protéine va interagir avec des facteurs transcriptionnels et induire la transcription des gènes très précoces (les gènes α). Les produits de ces gènes vont trans-activer les gènes précoces (gènes β codant des enzymes) puis les gènes tardifs (gènes γ codant des protéines de structure et des protéines régulatrices) responsable de la lyse cellulaire. La protéine VP16 n'est pas fonctionnelle dans les neurones, les gènes très précoces ne sont donc pas activés et le virus peut ainsi rester en latence. Les protéines α sont responsables d'un shut-off primaire des protéines cellulaires et les protéines β d'un shut-off secondaire des protéines cellulaires et virales. Néanmoins, il y a certaines protéines qui échappent à cette inhibition cellulaire comme les HSP ou les protéines de stress. Ces protéines sont peut-être nécessaires à la

réplication virale, dans ce cas, après leurs identifications (fonction et nature de la protéine) par des approches protéomiques globales, elles pourraient devenir des cibles thérapeutiques possibles pour le développement de nouveaux anti-viraux innovants. Ces anti-viraux n'auraient pas comme fonction d'inhiber les protéines virales, comme les traitements existants à l'heure actuelle (ex. l'Acyclovir : analogue de nucléoside, inhibant la réplication de l'ADN viral), mais plutôt d'inhiber des protéines cellulaires indispensables à la réplication du virus. Il est indispensable de mettre au point de nouveaux traitements anti-viraux car les infections par les virus HSV-1 ou HSV-2, sont en constante augmentation. Leurs cibles privilégiées sont les personnes immunodéprimées (SIDA, leucémies, greffe), ils sont ainsi responsables de la mort de près de 700 personnes par an en Europe. De plus, des souches virales résistantes à l'Acyclovir commencent à émerger et il n'existe aucun traitement de substitution (environ 5% de souches résistantes).

Les recherches actuelles portent sur l'étude des modifications transitoires du protéome au cours de l'infection virale. Un candidat médicament est en cours de développement. C'est un inhibiteur de la SAMDC (S-adenosyl methionine décarboxylase, enzyme de la voie de synthèse des polyamines). Chez l'animal, cette molécule permet l'inhibition de la réplication du virus. D'autres candidats nucléolaires et ribosomiques sont en cours d'études. Affaire à suivre.

Analyse « transcriptome » : bases théoriques et applications

Catherine NGUYEN, INSERM

Le génome est défini comme l'ensemble du contenu en ADN (et donc l'ensemble des gènes) d'un organisme. Il contient toute l'information nécessaire pour la reproduction, le développement et la physiologie d'un organisme. La génomique (étude des génomes, de leur organisation et de leur évolution, ainsi que de l'expression et de la fonction des gènes) consiste à cartographier et séquencer les génomes, à déterminer la structure des gènes (promoteur, ARNm, introns, exons...) mais cela ne suffit pas pour identifier la fonction d'un gène. Pour cela on utilise les techniques de génomique fonctionnelle qui vont déterminer les fonctions des gènes (activités, domaines, partenaires) et établir leurs profils d'expression (tissus, pathologies...).

L'étude du transcriptome (ensemble des ARNm d'une cellule à un moment donné) fait partie de la génomique. Les ARNm représentent entre 1 à 5 % des ARN totaux contenus dans une cellule, il est donc assez difficile de les étudier. Cependant quelques techniques sont à notre disposition comme le Northern-blot (technique basée sur le même principe que le Southern-blot mais sans digestion enzymatique car ce sont des ARN qui sont étudiés) qui permet d'apprécier la distribution d'un ARN dans les tissus, d'étudier son abondance relative, de déterminer sa taille et de détecter les intermédiaires de maturation et les différentes formes d'épissage de l'ARN, ou le Dot-blot (technique permettant de quantifier un ARN ou un fragment d'ADN donné sans séparation préalable sur un gel d'électrophorèse) qui permet d'obtenir des spots d'ARN mais qui ne permet pas de déterminer sa taille.

La technologie de puces à ADN, basée sur la technique d'hybridation, a été récemment développée pour optimiser l'étude du transcriptome. Cette technique consiste à immobiliser sur une matrice (nylon, lame de verre ou silicium), des sondes à savoir des oligonucléotides longs ou des produits de PCR (de 100 à 10 000 par matrice) et de les mettre en contact avec une solution contenant des ADNc marqués (colorimétrie, radioactivité, fluorescence), obtenus à partir d'une RP-PCR sur des ARN totaux extraits d'une cellule. Si une sonde est spécifique de l'ADNc, il y a hybridation. Les signaux d'hybridation seront détectés, selon le type de marquage, par mesure radiographique ou par fluorescence, puis quantifiés.

Il existe deux types de puces à ADN : les macroarrays et les microarrays (dont dérive les puces Affymetrix). La technologie des macroarrays consiste à déposer sur des membranes Nylon des produits PCR à une densité d'environ 40 dépôts/cm². Ces membranes sont ensuite hybridées avec des cibles radioactives marquées au ³³P. Les signaux d'hybridation sont détectés à l'aide d'un phospho-imager, quantifiés à l'aide de logiciels d'analyse d'images puis analysés.

La technologie des microarrays consiste à déposer sur des lames de verre des produits PCR ou des oligonucléotides longs (50mer-70mer) à des densités pouvant atteindre 6000 dépôts/cm². Ces lames sont ensuite hybridées avec des cibles fluorescentes (Cy3 et Cy5). Après hybridation, les signaux d'hybridation sont détectés à l'aide d'un microscope confocal. L'utilisation de deux fluorochromes différents permet de déterminer les signaux d'hybridation de deux souches distinctes au cours d'une seule expérience.

Les techniques de transcriptomiques peuvent être utilisées pour découvrir de nouveaux gènes impliqués dans différents mécanismes (criblage à grande échelle, criblage quantitatif différentiel) ou pour étudier les profils d'expression (typage moléculaire). Par exemple, en cas de cancer, on étudie les profils d'expression dans le but de faire des classifications pertinentes des tumeurs hétérogènes, de définir et valider des marqueurs de diagnostics et de pronostiques, de comprendre la maladie en identifiant des cibles, de développer de nouvelles drogues plus efficaces et/ou plus spécifiques mais aussi en vue d'adapter des traitements spécifiques pour chaque patient.

Comment la génomique peut-elle aider à comprendre le pouvoir pathogène de la bactérie *Helicobacter pylori*

Hilde **DE REUSE**, Institut Pasteur

Helicobacter pylori est une bactérie Gram négatif, de forme spiralée ou incurvée, flagellée, microaérophile et capnophile, isolée et cultivée en 1982 par les deux médecins australiens Barry Marshall et Robin Warren (Prix Nobel de physiologie et médecine en 2005). Elle est présente à la surface de la muqueuse gastrique, où elle est capable de survivre et de persister en dépit de l'extrême acidité des sucs gastriques acides et d'une forte réponse immunitaire. Cette bactérie peut provoquer des gastrites (inflammation de la muqueuse) en altérant la qualité du mucus de l'estomac. Ces gastrites peuvent être chroniques et, dans certains cas, contribuer à la formation d'un ulcère gastrique ou duodénal (segment initial de l'intestin grêle). Elles sont fréquemment retrouvées avant ou au moment de l'apparition d'un cancer de l'estomac, ce qui laisse supposer son influence dans l'origine de la tumeur maligne. Elle a ainsi été reconnue comme la première bactérie impliquée dans la genèse d'un cancer, le cancer de l'estomac, deuxième cause de décès dans le monde et première cause de décès dans les pays en développement. L'infection est plus fréquente dans les pays en développement (80 à 90%) que dans les pays industrialisés (25 à 30%). La bactérie se transmet directement d'homme à homme par voie orale, l'infection étant acquise dans la jeune enfance, le plus souvent au cours d'une transmission intrafamiliale. La nature et le degré de l'infection dépendent du génotype de la bactérie mise en cause, mais également du background génétique de l'hôte, du mode de vie et du régime alimentaire de l'hôte (la cigarette et une nourriture très salée sont des facteurs associés à des formes plus graves de l'infection). Néanmoins, dans la grande majorité des cas, ces infections se soignent plutôt bien grâce à un traitement d'une semaine composé de deux antibiotiques (l'amoxicilline et la clarithromycine) associés à un anti-acide (permet aux antibiotiques d'être efficaces).

Différentes approches ont été mises en œuvre pour étudier la bactérie *Helicobacter pylori* dont la génomique fonctionnelle (étude des interactions protéine-protéine, crible des gènes essentiels *in vivo* à la bactérie, étude des transcrits) et comparative (étude des variations de l'expression du génome en fonction d'un environnement donné), l'étude des interactions des bactéries avec les cellules épithéliales et immunitaires, ainsi que des voies de signalisation qu'elles induisent, l'analyse des réponses immunitaires associées aux infections...

C'est dans ce contexte qu'en 1995 et 1999, les génomes de deux souches d'*Helicobacter pylori* ont été entièrement séquencés ce qui a permis d'accroître les connaissances sur le métabolisme, la physiologie, les gènes régulateurs... de cette bactérie. Cela a également permis d'analyser la biodiversité entre les différentes souches d'*Helicobacter pylori* et d'identifier les différents facteurs de virulence ou d'adaptation, en vue de trouver de nouveaux candidats médicaments ou vaccins.

Certains facteurs de virulence ont ainsi été mis en évidence, comme l'activité uréase (rôle dans la résistance à l'acidité gastrique), la mobilité (les flagelles et la forme spiralée de la bactérie lui permettent de pénétrer dans le mucus gastrique), des systèmes de résistance au système immunitaire hôte (mimétismes moléculaires), des facteurs d'adhésion...

En parallèle, l'étude comparative des transcriptomes de deux souches d'*Helicobacter pylori* cultivées dans des conditions différentes (par exemple à pH7 et à pH5 pour étudier la réponse à l'acidité), a permis d'identifier une série de gènes up-régulés ou down-régulés, impliqués dans la pathogénicité de la bactérie, ce qui en fait des cibles thérapeutiques de choix.

Les travaux de génomique fonctionnelle et de génomique comparative se poursuivent en vue de mettre en évidence de nouveaux facteurs de virulence, de nouvelles cibles thérapeutiques et de développer des antibiotiques alternatifs dans le cas d'une résistance aux traitements précédemment décrit.

Compte-rendu n°4 :

SOMMAIRE

➤ Résumés des conférences :

- I. Récents développement en génomique structurale
- II. Analyse des complexes protéiques par les approches protéomiques
- III. Approches protéomiques pour la mise au point de nouveaux médicaments anti-herpès
- IV. Comment la Génomique peut elle aider à comprendre le pouvoir pathogène de la bactérie : *Helicobacter pylori* ?
- V. Génomique comparative des levures : cycles sexuels et régions subtélomériques
- VI. L'évolution des génomes viraux

➤ Synthèse

RESUMES DES CONFERENCES

I. Récents développement en Génomique Structurale

Conférence du 25/11/05 par Arnaud Ducruix-UMR 8015 CNRS, Paris

Depuis longtemps en biologie on s'intéresse à l'étude de la structure des molécules ayant pour but de comprendre le fonctionnement à l'échelle atomique et la relation entre la structure et la fonction. Ceci fait parti de la **Biologie structurale**.

Plus récemment on a assisté dans ce domaine au développement de la **Génomique structurale** et de la **Protéomique**. La génomique structurale consiste en la détermination des ORFs d'un génome, d'en purifier les protéines correspondantes et d'analyser la **structure tridimensionnelle** soit par diffraction aux rayons X (**crystallographie**) ou par **RMN** (résonance magnétique du noyau) à haut champ ou la microscopie. La protéomique se définit par l'étude des protéines exprimées à un temps donné.

De nombreux vecteurs d'expressions sont utilisés en pharmacutique, biotechnologies et ce sont des outils essentiels. Le plus connu c'est **E. coli**, mais d'autres connaissent un développement comme les **levures** intéressantes au point de vu des modifications post-traductionnelles. En effet beaucoup de paramètres sont à prendre en compte pour la qualité d'une protéine pour mieux apprécier la relation structure/fonction : glycosylation, pI, nombre de ponts disulfures,...

La génomique structurale a comme objectif d'améliorer la connaissance sur la **relation structure fonction** et les **interactions moléculaires**. Une des stratégies est la cristallographie. On travaille sur le cristal en solution d'une protéine purifiée qui par la diffraction aux rayons X permet de déterminer les distances interatomiques. L'informatique permet ainsi de donner un squelette de la molécule en 3D : de plus, à haute résolution on distingue les atomes. Les paramètres essentiels à contrôler sont les conditions dans lesquelles on obtient le cristal : type de solvants, température, pH,...

De même la RMN permet une analyse structurale claire d'une protéine : « l'excitation magnétique » de la protéine génère la protonisation des atomes d'hydrogène et d'azote, le traitement en spectres 1D ou 2D, selon que l'on analyse un ou deux atomes, donne des informations sur la composition en amines de la chaîne principale. Donc la combinaison de ces stratégies permet de « dessiner » la structure d'une molécule via le traitement des données par **l'informatique** qui modélise la structure tridimensionnelle déterminée. De nombreux programmes d'études en génomique structurale sont répartis mondialement : aux Etats-Unis, Canada, Japon et Europe (Allemagne, France). En France, c'est le réseau national des Génopoles qui en est chargé dont les principaux projets sont l'étude de *M. tuberculosis* à l'Institut Pasteur, l'étude des récepteurs nucléaires à Strasbourg et Grenoble,...

L'étude des **interactions protéiques** est primordiale pour comprendre les mécanismes de signalisation cellulaire à la fois en immunologie mais aussi au niveau des voies

métaboliques. L'étude de la voie de l'insuline en est un exemple. En effet l'étude d'un complexe : grb14/PIR par RMN et diffusion au rayons aux petits angles (DXPA) a permis de déterminer les caractéristiques structurales de PIR. L'intéressant dans cette analyse fut la découverte surprenante d'une protéine très flexible, sans repliement (très peu de structures secondaires), et de faible compacité qui a une activité fonctionnelle connue *in vivo*. Or généralement on pensait que pour la fonction d'une protéine le repliement et la compacité entre autres étaient déterminantes. Donc la génomique structurale a mis en évidence une nouvelle famille de protéines : les **IUPs (Intrinsically Unstructured Proteins Family)** ce qui démontre l'outil puissant qu'elle peut être.

Pour conclure, le développement de ces techniques ouvrent de **grandes perspectives** dans l'étude des protéines : structure, fonction, interactions,... mais aussi des **avancées technologiques** telle que l'automatisation de la cristallographie et comme dernier point la production à grande échelle de protéines de plusieurs génomes. De plus, la génomique structurale est un outil d'avenir dans le **criblage des molécules**, futurs médicaments. Enfin, cette conférence est d'un grand intérêt en Immunotechnologies car ces techniques s'applique notamment dans l'étude des protéines des voies de signalisation des lymphocytes, ou de la structure des anticorps,...

II. Analyses des complexes protéiques par les approches protéomiques *Conférence du 23/11/05 par Bertrand Séraphin – CGM CNRS, Gif sur Yvette*

Pour pouvoir analyser des complexes protéiques il existe différents outils dont certains en récents développements. Tous ceux-ci sont soumis aux difficultés que constitue l'étude des **complexes protéiques** et qui sont principalement le caractère tridimensionnel des protéines et les différentes propriétés physico-chimiques. Quels sont les moyens d'analyse en protéomiques présenté ici ? On distingue la **spectrophotométrie de masse** qui est une étape d'identification directe, **l'étude des interactions protéiques** regroupant différentes techniques détaillées plus loin et enfin **l'analyse fonctionnelle** de telle ou telle protéine identifiée.

Tout d'abord, pour identifier les peptides des protéines d'intérêt on dispose de la spectrophotométrie de masse : **MALDI (Matrix assisted laser desorption/ionization)**. Sous un même principe il y a différentes techniques. En effet, cela consiste en une excitation d'un échantillon protéique par un laser qui conduit à l'ionisation ; les ions libérés sont acheminés vers des détecteurs et les signaux électrostatiques sont traités informatiquement pour déterminer la masse des peptides libérés. Cet outil est sensible et puissant connaissant des évolutions constantes : MALDI-TOF, MALDI-TOF-TOF,...

Dans la connaissance de la composition des complexes protéiques il est essentiel de découvrir quels sont les interactions existantes entre les protéines et donc mettre en évidence l'existence d'une protéine jusque là inconnue. Cela est possible par différents approches : l'identification de partenaires protéiques par la technique des **doubles hybrides** dans un système de levure et la technique des « **protein chips** ». Mais pour pouvoir analyser de manière la plus précise un complexe il faut une étape de **purification** , étape critique et essentielle car d'elle dépend en partie la qualité et la sensibilité de l'identification. Une technique dite **TAP : Tandem Affinity Purification** permet de « tagger » une protéine d'intérêt par une construction moléculaire constituée de la protéine A d'un site de clivage spécifique à une protéase et de la calmoduline « binding domain ». On est capable de produire ces protéines « tagguées » chez la levure qui une fois extraites sont purifiées en deux étapes : une première élution par affinité via des billes avec des IgG et une seconde élution par affinité via des billes couplées à la calmoduline.

Cette technique a **plus d'avantages** que l'immunopurification classique : elle est simple, rapide à mettre en place, on peut travailler avec une faible quantité d'échantillon protéique, diminuer les interaction non spécifiques,... et procéder directement à différentes analyses : spectrophotométrie de masse, gel SDS-PAGE,... D'autant plus que la stabilité et la structure protéique sont mieux préservées par des conditions d'élutions plus « douces » permettant de garder l'intégrité du complexe la plus fidèle possible. L'entreprise Cellzome basée à Heilderberg en Allemagne a montré par ses travaux dans ce domaine les perspectives importantes dans l'application pour la découverte de nouveaux médicaments.

A un autre niveau, un moyen permettant de mettre en évidence des partenaires spécifiques dans un complexe protéique est la technologie du double hybride ; technique qui a fait ses preuves notamment dans la connaissance de facteurs protéiques intervenant dans la transcription,... Enfin en comparaison avec la génomique, en protéomique la technologie des « puces » est utilisée pour identifier les protéines exprimées par un organisme ou une cellule.

Avec le développement de ces technologies comme la méthode de purification TAP on crée des stratégies d'analyse qui **combinent la génomique et la protéomique** . Cela est possible dans l'analyse de complexes protéiques du métabolisme de l'ARNm,... on peut alors combiner la spectrophotométrie de masse et la technique des micro arrays sur l'ARN. Ainsi, cette présentation de l'application de la protéomique dans **l'identification des complexes protéiques** est intéressante dans le développement des connaissances en immunologie des mécanismes et interactions de facteurs protéiques clés du système immunitaire.

III. Approches protéomiques pour la mise au point de nouveaux médicaments anti-herpès

Conférence du 23/11/05 par Jean-Jacques Diaz – INSERM U369, Faculté de Médecine Lyon

Dans la population les cas d'infection par le virus de **l'herpès simplex de type 1** (HSV-1) sont toujours en augmentation. A l'heure actuelle le traitement de routine administré est l' **Acyclovir** qui est un inhibiteur de l'ADN viral. Cependant on observe l'émergence de souches résistantes c'est pourquoi il est nécessaire de **développer de nouveaux médicaments anti-herpétiques**.

Pour cela il faut découvrir des **cibles thérapeutiques**. On s'intéresse ici aux **protéines cellulaires** dont la synthèse est maintenue au cours de l'infection. Plus précisément l'étude a été faite sur la voie de biosynthèse des polyamines et les protéines du nucléole.

L'approche est de suivre au cours du temps les modifications transitoire du protéome cellulaire durant l'infection. On suit qualitativement l'expression des protéines cellulaires par électrophorèse bidimensionnelle effectuée 3, 6, puis 9h après l'infection (incorporation de Methionine S35). Après analyse on a observé 28 protéines dont la synthèse n'est pas inhibée par le virus.

Parmi ces protéines, il en existe une dont la synthèse est particulièrement bien maintenue. C'est la SAM décarboxylase qui appartient à la voie de biosynthèse des polyamines. L'étude de cette protéine a permis de montrer qu'en inhibant son action par des inhibiteurs spécifiques on évitait la réplication virale.

Ainsi les **inhibiteurs de la voie de synthèse des polyamines** constituent de nouveaux médicaments possibles pour le traitement anti-herpès.

L'analyse histochimique a démontré que l'infection par HSV-1 entraîne des **modifications dans l'architecture du noyau**, notamment au niveau inter chromosomique. Ces observations ont motivé l'étude protéomique au niveau du noyau. Des expériences d'électrophorèse sur gel ont mis en évidence dans le nucléole 213 protéines. Parmi celles-ci on a identifié certaines fonctions : chaperonnes, protéines ribosomales, ... On a pu cribler de façon plus précises 39 protéines par gel 2D.

De même on détermine si des modifications de la synthèse de ces **protéines nucléolaires** sont induites durant l'infection. En effet il a été démontré que la synthèse des protéines ribosomales est modifiée ; on a mis en évidence une protéine : **US11** dont la synthèse est stimulée par l'infection. Cette protéine est un modèle pour l'étude des

mouvements intracellulaires. Cette protéine jouerait un rôle dans le transport du nucléole au cytoplasme.

De plus par électrophorèse on a cherché des partenaires de cette protéine et on a purifié un complexe dans lequel on a identifié un partenaire : **HIPK-5 (Homeodomain Interacting Protein Kinase 2)**. Il agit dans la cascade de signalisation de la réponse anti-virale qu'est l'apoptose dont US11 serait un inhibiteur. Donc il serait intéressant d'empêcher l'interaction entre ces deux protéines et de stimuler l'action de HIPK-5.

Pour conclure l'approche protéomique a permis de cribler des protéines de choix comme cible thérapeutiques qui permettent d'envisager le **développement de molécules innovantes anti-herpès**. On pourrait ainsi moins privilégier le traitement conventionnel qui induit des souches résistantes. Enfin, étant donné la découverte de ces nouvelles cibles on pourrait développer de nouveaux anticorps qui amélioreraient leur criblage et pourquoi pas améliorer le diagnostic. Cette application de la protéomique rend compte des possibilités d'applications en immunologie.

IV. [Comment la génomique peut aider à comprendre le pouvoir pathogène de la bactérie *Helicobacter pylori*](#) Conférence du 9/12/05 par Hilde De Reuse – Institut Pasteur

Cette bactérie a été récemment découverte en 1982 par Barry Marshall et Robin Warren qui en 2005 ont été récompensés par un prix Nobel de Physiologie et de Médecine. C'est une découverte d'importance; en effet ce micro-organisme ayant pour réservoir l'estomac chez l'homme et le primate est responsable dans la population mondiale de nombreuses **pathologies du système digestif** : gastrites, ulcères,...avec des cas moins fréquents de lymphome et de carcinome dus à des formes sévères de gastrite. De nombreux facteurs rendent susceptible le développement d'une infection par *H. pylori* qui sont la qualité de vie et le régime alimentaire d'où des cas plus nombreux dans les pays pauvres ou en développement, mais aussi le fond génétique de l'hôte et le génotype de la bactérie.

L'étude par la **génomique** par la technique de **micro-arrays** permet de mieux connaître les mécanismes de **pathogénicité** de cette bactérie. Les **facteurs de virulence** sont un cas d'étude de choix. Quels sont ces facteurs? Tout d'abord la capacité de colonisation de *H. pylori* dans l'estomac est possible par la production d'ammoniac qui permet de résister à l'acidité. De plus, *H. pylori* est mobile. Aussi cette bactérie est capable de **résister à la réponse immunitaire de l'hôte** (catalase, mimétisme moléculaire,...). Concernant la

pathogénicité, elle induit l'inflammation et la dégradation du tissu gastrique. La cytotoxine sécrète est VacA et dans le génome on trouve l'îlot de pathogénicité *cag*.

En 1997 et 1999 le génome de deux souches : 26695 et J99 à été séquencé. Pourquoi utiliser la génomique? On souhaite mieux connaître les gènes du métabolisme, la physiologie...Mais aussi analyser la **biodiversité des souches** et le plus important: trouver des **nouvelle cibles thérapeutiques** et développer alors des médicaments ou un vaccin. La **génomique comparative** de différents souches cliniques permet de mettre en évidence une diversité génétique parmi les souches de *H. pylori*. Or cela n'est pas suffisant, d'autres étapes d'analyses sont nécessaires à travers la génomique dite fonctionnelle. En effet, il y a l'identification des interactions entre les protéines exprimées par un génome: c'est **l'interactome** qui est possible par la technique des doubles hybrides en utilisant les levures comme système d'expression. Un autre type d'étude : la "**STM Technology**" a pour but d'identifier les gènes essentiels à la bactérie *in vivo* (étude chez la gerbille). Enfin, on s'intéresse aux changements au niveau des transcrits exprimés, soit le **transcriptome**. Deux études *in vitro* chez *H. pylori* dans des conditions proches de l'estomac ont été réalisé et le transcriptome analysé par micro-arrays.

Ainsi, l'étude de *Helicobacter pylori* par la génomique ouvre de grandes perspectives pour la découverte de cibles thérapeutiques nouvelles, en l'occurrence de nouveaux facteurs de virulence. **L'approche fonctionnelle de la génomique** est puissante dans le sens où elle permet de mieux connaître les mécanismes qui entrent en jeu durant une infection : colonisation, résistance au système immunitaire... Ce travail effectué sur cette bactérie montre l'outil d'avenir que peut être la génomique dans la compréhension des **mécanismes hôte pathogène**.

V. Génomique comparative des levures : cycles sexuels et régions subtélomériques

Conférence du 16/11/05 par Cécile Fairhead - Génétique Moléculaire des levures, Département Structure et Dynamique des Génomes, Institut Pasteur, Paris

Cette conférence a pour sujet l'étude et la comparaison des **gènes du cycle sexuel** et des **régions subtélomériques** entre différentes espèces de levures. Les levures sont des champignons microscopiques unicellulaires faisant parties des **Hémiascomycètes** ; c'est un phylum aussi ancien et varié que les autres du vivant.

Les études sur le génome des levures ont permis de constater que de manière générale la taille et le nombre de gènes sont conservés entre les espèces sauf pour une qui a évolué plus différemment : *Y. lipotica* avec une taille des chromosomes plus grande. Au niveau de la

reproduction sexuée on distingue deux types de cellules haploïdes : une cellule portant le **gène MAT α** et une cellule portant le **gène MAT a** . Ces gènes codent pour des facteurs de transcriptions qui déterminent le sexe de la levure. La fécondation de ces deux cellules donne un zygote α/a qui après méiose génère une tétrade de cellules a ou α .

Au sein des Hémiascomycètes on observe une **diversité de la sexualité** due aux variations entre espèces parmi ces gènes. En effet, des levures comme *S. cerevisiae* ont un cycle « haplo-diplontic » ou simplement « haplontic » dans le cas de *K. lactis* alors que *A. gossypii* n'est qu'une cellule haploïde et est incapable de croisement, elle a perdu sa sexualité. De plus il semblerait qu'il y ait une corrélation entre la perte de la sexualité et la pathogénicité. Un point intéressant est la capacité de changement sexuel parmi certaines levures. Ceci est possible grâce à des cassettes de séquences : **MAT-like et HML-like, HMR-like** qui elles sont « silencieuses ». Une **conversion génique** peut avoir lieu par coupure au niveau de sites spécifique d'une endonucléase : HO qui permet ainsi le changement de type sexuel. Donc l'étude des gènes de la sexualité des levures a mis en évidence des différences et l'importance de cette sexualité comme facteur clé dans **l'évolution** du monde des levures.

La seconde partie qui concerne l'étude des régions subtélomériques présente l'organisation, le type de gène et leur rôle. Ces régions sont en amont des régions télomériques des chromosomes, qui contiennent des **gènes non essentiels** et qui se caractérisent par un grand **polymorphisme**. L'étude comparative entre levures montre une variation d'une espèce à l'autre. En effet, les gènes présents dans ces régions sont impliqués dans **l'adaptation de la levure à son milieu spécifique**. Chez *S. cerevisiae*, par exemple, beaucoup de ces gènes ont un rôle dans le **métabolisme des sucres**,... Aussi, l'étude de l'assemblage des gènes chez *K. lactis* a permis de révéler des gènes dupliqués, des gènes paralogues et de même des gènes du métabolisme des sucres. On a donc une organisation diverse avec des gènes issus de **duplications** ou étant des **paralogues** d'autres gènes ou des gènes « uniques ». Dans les régions subtélomériques il y a une évolution plus rapide que dans le reste du génome : les **gènes** sont dits **de contingence**.

Une expérience de cassure de l'ADN chez *S. cerevisiae* par intégration d'une cassette par un plasmide entraîne la perte de croissance chez 95% des levures, or 5 % sont capables de réparer cette cassure. En fait au niveau des subtélomères, on trouve plusieurs **mécanismes de réparation de l'ADN**: conversion génique, recombinaison homologue... Ceci apporte à la levure un moyen de survie lors d'une cassure de l'ADN dans ces régions mais toutefois, il ne semblerait pas que cela soit aussi efficace.

Ainsi, dans ces régions subtélomériques on observe une organisation des gènes particulières et une vitesse de mutation importante qui permet aux levures de s'adapter à leur milieu. Cette **plasticité** explique l'évolution du génome entre espèces et individus d'une même espèce.

VI. L'évolution des génomes viraux

Conférence du 5/12/05 par Simon Wain-Hobson – Institut Pasteur

Dans cette conférence on présente les caractéristiques des génomes viraux en terme d'évolution. Depuis 1976 de nombreux génomes de différents virus ont été séquencés, le premier étant le virus MS2 par Fiers. Dans le monde des virus il faut distinguer les virus à ARN et les virus à ADN, selon la classification de Baltimore on sépare les virus à ARN, ADN simple brin ou double brin.

Ce qui caractérise les virus à ARN c'est leur taux élevé de mutation par génome et par cycle par rapport aux microorganismes à ADN comme les bactériophages... En fait, ces virus n'ont pas d'activité de Proof Reading et ne corrigent donc pas les erreurs au niveau du génome. Aussi, les virus à ARN constituent environ 80% des virus dans le monde : cette prédominance semblerait être due à cette « capacité » de mutation. A titre de comparaison ces virus fixent un million de fois plus vite les mutations que les mammifères.

Une étude sur la recombinaison chez le virus HIV a été effectuée. Lors d'une infection, une cellule peut être infectée par différents provirus générant ainsi des nouveaux provirus recombinants. Sur deux patients on détecte par FISH différents provirus dans un LTCD4 : chez l'un 3 provirus sont détectés et chez l'autre 4 après analyse de la séquence. Ces observations expliquent la complexité du virus en l'occurrence dans la structure de l'enveloppe : on a une mosaïque de protéines. Ceci rend plus difficile la détection par les anticorps. Donc HIV n'est pas clonal.

On se pose la question de savoir quel bénéfice retire le virus de ces recombinaisons. Or la taille du génome des virus à ARN n'excède pas 10kb. Ce génome compact fait que la syntenie est plutôt conservée : grande conservation en terme d'organisation. La petite taille du génome serait un frein à l'évolution, en fait ces virus changent plus qu'il ne s'adapte à leur hôte.

Par contre pour les virus à ADN on n'observe pas les mêmes caractéristiques. En fait la majorité des formes de vie sont des **bactériophages** et toutefois on a très peu étudié les génomes de ce monde viral : la taille du génome est plus grande avec une organisation « mosaïque » des gènes c'est à dire des gènes uniques et des gènes homologues à des gènes bactériens... D'ailleurs l'analyse du génome de Acidianus two-tailed virus (ATV) a permis de

déterminer la présence de gènes homologues à ceux d'archaebactéries. Tout ceci amène de plus en plus à l'hypothèse que **certains virus dériveraient en fait de bactéries**. En terme d'évolution ce type de génome est plus apte à générer des réarrangements, des transductions de gènes, des paralogues...

Ainsi, cette conférence montre en quoi les caractéristiques d'un génome détermine en partie comment évolue les virus dans leur milieu que ce soit les virus à ARN, rétrovirus et à ADN. De plus, ce thème abordé peut totalement servir en Immunologie car cela nous apporte des informations sur comment les virus évoluent au sein de l'hôte et nous amène à des nouvelles voies à explorer dans la compréhension des infections par HIV,...

SYNTHESE

Il n'est pas surprenant de dire que la génomique et la protéomique constituent une grande révolution en biologie. Ces « outils » puissants ont amélioré les connaissances et la manière d'analyser les génomes : séquences, organisation et disposition des gènes,... et les protéines correspondantes. A travers les différentes conférences présentées dans ce module, on a pu constater la diversité des applications et approches dans ce domaine. La preuve en est leur utilisation au niveau médical, de l'agroalimentaire ou dans l'étude de la biodiversité.

Tout d'abord, la génomique rassemble de nombreux sujets abordés. En effet, on s'intéresse à l'ensemble des gènes qui constituent un génome donné. Le premier grand pas technique fut le séquençage d'un génome entier qui est toutefois soumis à des difficultés selon la taille et la complexité du génome. A partir de 1977 (méthode de Sanger) jusqu'à aujourd'hui de nombreux programmes de séquençages ont vu le jour, et continueront à se faire grâce notamment en France au réseau des Génomopôles dont l'Institut Pasteur fait partie. A titre d'illustration : 276 séquences de procaryotes et 39 eucaryotes ont été publiées à ce jour contre 804 et 547 séquences qui sont elles en cours d'étude.

D'autre part, l'apport de la génomique est important dans la connaissance et l'identification de la biodiversité à travers le monde bactérien, des levures et des virus. On compte ainsi comprendre le développement, la physiologie de ces organismes, en s'intéressant aussi aux mécanismes d'évolution des génomes. Dans un autre domaine c'est-à-dire au niveau de la recherche biomédicale, l'application de la génomique permet des avancées dans l'étude des mécanismes de pathogénicité de microorganismes pathogènes (virus, bactéries...), ou encore dans l'étude de diverses pathologies : cancers... Ceci permettant d'apporter des solutions nouvelles en trouvant de nouvelles cibles thérapeutiques : identification de nouveaux facteurs... et en lien avec la thérapeutique d'améliorer les outils diagnostiques.

D'ailleurs la protéomique et la génomique connaissent un développement pour atteindre cet objectif ; c'est le cas en oncologie dans l'identification des tumeurs, point développé plus loin. De manière spécifique à la protéomique une évolution cruciale c'est l'étude de la relation structure fonction des protéines appuyés par des avancées technologiques en spectrophotométrie de masse, dans la purification des protéines (méthode TAP...) dans l'identification de partenaires (double hybrides...). Surtout on se tourne de plus en plus vers l'automatisation des techniques comme la cristallographie, la spectrophotométrie... dans le but d'avoir un processus reproductible, rapide et plus sensible ; cette automatisation est primordiale pour le milieu industriel.

La génomique et la protéomique nécessitent un outil essentiel : la bio informatique. Il faut pouvoir traiter plus facilement et rapidement les données apportées par l'étude d'un génome (annotation *in silico* de séquences génomiques bactériennes...), d'une protéine... Par exemple, dans le cas de l'analyse des complexes protéiques, la modélisation en 3D de la structure obtenu est importante pour comprendre les interactions moléculaires... ceci est de plus en plus utilisés par l'industrie pharmaceutique qui souhaitent ainsi mieux cribler les futurs médicaments en simulant par l'informatique les interactions entre le médicament et sa « cible ».

La diversité des techniques et approches en génomique est présente de même en protéomique. La technique « phare » c'est la technologie des puces à ADN ou à ARN. Elle permet de donner le profil d'expression des gènes d'un génome : savoir quels gènes sont éteints ou activés... Le transcriptome a l'avantage de donner de manière plus précise quels gènes sont exprimés par l'analyse de l'expression des ARN messagers correspondants ; on sait donc pour un temps et un état donné quels ARN messagers sont exprimés. En amont de tout cela la biologie moléculaire est nécessaire : PCR...

La génomique et la protéomique ne s'excluent pas, elles sont complémentaires. En fait, la génomique s'intéresse à « l'information » et la protéomique peut en tirer la fonction. L'avancée la plus intéressante est sans doute l'application de la protéomique dans la recherche médicale. La protéomique connaît un développement plus important de part l'application directe qu'elle apporte en terme médical. En effet, la protéomique permet de s'intéresser à la fonction par l'analyse de l'expression des protéines à un temps et un état donné. Les ProteinsChips basés sur le principe des puces à ADN... permettent d'identifier les protéines exprimées par un tissu. Ce sont des puces qui sont couvertes par des Anticorps qui capturent les protéines d'un échantillon. La combinaison des techniques de techniques comme la chromatographie d'affinité, la spectrophotométrie de masse ajoute à l'identification et la

caractérisation des protéines identifiées. La but principal est la découverte d'un bio-marqueur c'est-à-dire une protéine caractéristique par exemple d'une tumeur donné, correspondant à un stade de développement donné... En fait pour résumer on a ainsi le « phénotype » d'une cellule ou d'un tissu.

Ceci est plus puissant que la génomique car on améliore les perspectives de traitement par la détermination directe d'une cible thérapeutique qui permet le développement d'un traitement. A cela est rattaché l'amélioration du diagnostique par la création de nouveaux outils supportant la thérapeutique afin de mieux adapter le traitement.

Où se placent l'Immunologie au milieu de la génomique et de la protéomique ? Les applications dans ce domaine sont nombreuses. Le développement des immunotechnologies en l'occurrence en recherche biomédicale a besoin de la génomique et de la protéomique. Les anticorps monoclonaux sont le meilleur exemple du développement des outils de thérapies « immunologiques ». L'obstacle rencontré est l'amélioration de la détermination d'un antigène cible de choix, ce qui est d'autant plus vrai en oncologie... et ces domaines peuvent apporter des solutions.

Au delà de la recherche biomédicale, la génomique et la protéomique sont des outils adaptés pour la connaissance du système immunitaire : étude des voies d'activations lymphocytaires, études des interactions hôte-pathogène, découverte d'antigènes pour le développement de vaccins...

Finalement, on se rend compte que entre la génomique/protéomique et l'Immunologie les concepts sont assez proches. La première a pour but de mieux identifier et comprendre comment fonctionne, réagit un organisme ou une cellule à différents niveaux. La deuxième va exactement dans ce sens.

Compte-rendu n°5 :

Biogénèse des mitochondries : une approche génomique.....	45
Du génome à la Cellule : comment prédire la fonction des gènes ?	46
Annotation in Silico de séquences génomiques bactériennes	45
Analyses des complexes protéiques par des approches protéomiques.....	47
Approches protéomiques pour lamise au point de nouveaux médicaments anti-herpès...	48
Comment la génomique peut aider à comprendre le pouvoir pathogène de la bactérie Helicobacter pylori ?	49
Conclusion.....	50

Biogenèse des mitochondries : une approche génomique.

Claude Jacq

La biogenèse des mitochondries met en jeu des mécanismes qui ont été sélectionnés au cours de l'évolution pour assurer le correct assemblage des mitochondries. La théorie de l'origine endosymbiotique des mitochondries rend leur biogenèse encore plus intrigante, car la machinerie de transport de protéines des mitochondries a du subir une ré-évolution après le transfert de la plupart des gènes de l'ancêtre bactérien au génome de la cellule hôte. Les mitochondries ont, au cours de l'évolution, perdus une grande partie de leurs gènes par des mécanismes de transfert actif de l'ADN mitochondrial vers le noyau, comme l'indique les nombreuses traces 'fossiles' d'ADN mitochondrial dans l'ADN nucléaire, les séquences dites 'Nuclear Mitochondrial DNA' (NUMTs). La plupart des NUMTs dans le génome humain forment des introns insérés au milieu de gènes.

Une approche transcriptomique globale a permis de séparer les gènes nucléaires codant des protéines en deux catégories : celle des gènes traduits dans le cytoplasme par des complexes ribosomiaux, les polysomes dits « libres », et celle des gènes traduits par des polysomes associés aux mitochondries. Pour identifier les gènes qui envoient leurs ARNm au voisinage de la mitochondrie, les polysomes libres ou associés aux mitochondries ont été purifiés à partir de cellules de levure, puis les ARNm de ces deux fractions ont été extraits pour produire des molécules d'ADNc. Ces ADNc ont ensuite été marqués à l'aide de sonde deux sondes fluorescentes : Cy3 pour la préparation d'ADNc obtenue à partir des polysomes libres, et Cy5 pour la préparation obtenues à partir des polysomes liés aux mitochondries, puis hybridés à des micorarrays. L'analyse des rapports Cy5/Cy3 pour chacun des gènes a permis de leur attribuer un score de localisation mitochondriale des ARNm (MLR), compris entre 1 et 100, et de déduire la distribution intracellulaire de ces ARNm : un score MLR élevé, pour un gène codant une protéine mitochondriale, indique une probabilité élevée de trouver son ARNm au voisinage d'une mitochondrie. Cette méthode a révélé que plus de la moitié des gènes codant des protéines mitochondriales ont un score supérieur à 70, alors que les protéines du cytoplasme ont en majorité un score faible.

Dans un deuxième temps, les séquences 3'UTR d'ARNm codant des protéines mitochondriales, dont les gènes ont obtenu un score MLR supérieur à 84, ont été marquées *in vivo* à l'aide de protéines de fusion contenant la GFP, ce qui a permis d'observer *in vivo* la localisation de ces transcrits au voisinage des mitochondries. Enfin, par une analyse bioinformatique des séquences des gènes étudiés par microarray à l'aide du logiciel BLAST, il a été possible de démontrer que les gènes à haut score MLR, donc traduits au voisinage des mitochondries, ont en majorité une origine procaryotique et codent des éléments du cœur des complexes protéiques mitochondriaux.

Cette conférence portait sur un thème très 'fondamental', dont l'intérêt a été d'aborder la génomique mitochondriale, dont on parle relativement, et d'enrichir notre culture générale scientifique.

Annotation *in silico* de séquences génomiques bactériennes

Claudine Médigue

Les progrès de la génomique, à la fin du 20^{ème} siècle ont permis de séquencer le génome de nombreux organismes. Ce ne fut cependant qu'une première étape, car le plus difficile restait encore à faire : comprendre la signification des « textes génomiques » séquencés, c'est-à-dire

identifier les séquences codantes ou régulatrices, caractériser leur fonction et comprendre comment elles interagissent. L'annotation des génomes englobe ces trois niveaux d'analyse : l'annotation syntaxique, l'annotation fonctionnelle et l'annotation relationnelle, en s'appuyant sur des modèles mathématiques et l'outil bioinformatique.

L'annotation syntaxique concerne l'identification de zones d'intérêt sur la séquence génomique. Il s'agit typiquement de la recherche des zones codant potentiellement des protéines (CDSs) ou des ARNt, et de la recherche de signaux de régulation de l'expression génétique. L'annotation syntaxique des gènes codant des protéines des génomes procaryotes ne pose plus de réelles difficultés, mais l'identification de signaux de régulation ou de « petites » structures régulatrices d'ARN reste un problème pertinent.

L'annotation fonctionnelle concerne l'attribution d'une ou plusieurs fonctions biologiques aux signaux détectés au niveau précédent. Il s'agit par exemple d'attribuer un rôle fonctionnel aux protéines codées par des gènes détectés ou de caractériser la fonction d'une séquence régulatrice. Lorsqu'il n'existe pas de données expérimentales associées à une séquence polypeptidique, la stratégie classique consiste à lui attribuer, par analogie, la fonction de séquences fortement similaires identifiées par criblage des bases de séquences. Les résultats d'une telle stratégie sont des hypothèses de travail qu'il convient de valider expérimentalement. Réalisée automatiquement, cette stratégie d'assignation de fonctions présente de nombreuses limites. Par exemple, il est nécessaire d'évaluer au cas par cas la pertinence de la similarité entre les séquences comparées. D'autre part, cette stratégie est totalement dépendante de la qualité des données présentes dans les bases de données utilisées lors du criblage (problème de propagation d'erreurs d'annotations).

Enfin, l'annotation relationnelle concerne l'identification des relations existant entre les séquences caractérisées individuellement aux deux niveaux précédents. Elles peuvent être impliquées dans un processus cellulaire commun (voie métabolique), appartenir à des réseaux d'interactions géniques ou bien être en interaction protéine-protéine.

Il y a aujourd'hui trois grands projets dans lesquels s'inscrivent des projets d'annotation de génomes : le séquençage du génome de bactéries pathogènes, pour le développement de nouveaux antibiotiques ou de nouveaux vaccins, le séquençage du génome de bactéries modèles, pour des études de biologie fondamentale et le séquençage du génome de bactéries de l'environnement pour étudier la biodiversité. Cette conférence démontre que la génomique en elle-même reste un domaine d'avenir à la fois scientifique et technologique, puisque de nombreuses innovations en matière d'outils bioinformatiques, notamment naissent initialement des besoins de la recherche en génomique, et aboutissent parfois à des projets d'entreprises mais il faut reconnaître que ces outils innovants restent souvent dans le domaine public.

Du génome à la Cellule : comment prédire la fonction des gènes ?

Antoine Danchin

Qu'est-ce que la Vie ?

Trois processus sont nécessaires pour définir un organisme vivant. Les deux premiers sont le métabolisme (flux permanent qui isole certaines espèces chimiques de l'environnement, les intègre, les modifie, et y rejette d'autres espèces), et la compartimentation (il n'y a pas

d'organisme vivant sans un intérieur et un extérieur). Le troisième, le transfert de l'information, suppose l'existence d'une mémoire, le génome, et repose sur deux lois centrales : la loi de complémentarité (règle de réécriture qui permet la conservation de l'information génétique dans le même langage lors de la réplication et de la traduction) et la loi du code génétique (transfert de l'information lors de la traduction entre deux niveaux, les acides nucléiques et les protéines). La fonction est ici un concept central, à partir duquel s'est construite la théorie de l'évolution de Darwin, (Variation (+stabilisation)/ Sélection / Amplification) : tout système soumis à ces processus évolue, créant de nouvelles fonctions, qui seront capturées au sein de structures, elles-mêmes codées par des séquences.

Le séquençage d'un génome ne permet pas en lui-même d'accéder à la fonction des gènes. L'annotation syntaxique d'un génome séquencé permet dans un premier temps de localiser les séquences codantes ou opératrices, et ça n'est qu'ensuite que l'on va caractériser la fonction des gènes identifiés par la démarche de génomique comparative, méthode *in silico* de comparaison des génomes d'espèces voisines. Il s'agit de rechercher des similitudes avec d'autres séquences connues dans le même organisme (séquences paralogues) ou dans d'autres espèces (séquences orthologues). Cette méthode permet, à l'aide de logiciels (ex : *DNA strider*, *Blast* et *Fasta*) de rechercher des motifs connus, de les comparer à des banques de motifs structuraux et fonctionnels, etc. On peut ainsi identifier des phénomènes de polymorphisme normal ou pathologique, prédire les structures 2D et 3D des protéines ou des modifications post-traductionnelles. La génomique dite « fonctionnelle » va faire la synthèse de ces résultats produits par la démarche *in silico* aux résultats que produit la "biologie humide" (génie génétique, clonage, mutagenèse dirigée, puces à ADN).

Cette conférence, dont le thème principal avait déjà été abordé par Claudine Médigue (voir page 45), a permis d'introduire, dans ce cycle très technique, des notions de Biologie Théorique, ce qui a permis de rappeler que la Biologie est une démarche intellectuelle qui, au même titre que la philosophie, participe de la Pensée, que l'on appelle également Science.

Analyses des complexes protéiques par des approches protéomiques.

Bertrand Seraphin

De nombreuses protéines interagissent entre elles pour former des complexes. Une partie des études protéomiques cherche donc à révéler ces réseaux d'interaction.

Pour purifier des complexes protéiques, il existe une méthode standard (biochimique), une méthode utilisant l'interaction anigène-anticorps (mais les méthodes de dissolution des complexes AG-AC dégradent l'ensemble du complexe à purifier) et la technique « pull down », qui consiste à remplacer un des partenaires du complexe par un analogue muni d'une étiquette. Cette technique permet de définir un protocole standard, quelle que soit la protéine cible, une fois celle-ci étiquetée.

Une méthode de purification de complexes protéiques particulièrement adaptée à la protéomique a été mise au point sur ce principe: la méthode TAP ("Tandem Affinity purification") Elle consiste à greffer sur une des protéine d'un complexe l'étiquette [Protéine A] + [site de clivage par la protéase TEV] + [domaine de liaison à la calmoduline]. Cette méthode est basée sur deux étapes successives de chromatographie d'affinité, d'abord sur une colonne contenant des billes couplées à une IgG anti-protéine A, puis, après élution par action de la protéase TEV, sur une deuxième colonne contenant des billes couplées à la calmoduline

et en présence d'ions Ca^{++} . La qualité des complexes purifiés par cette méthode est grandement améliorée, en comparaison avec les techniques standards, car cette technique permet de « voir » aussi les protéines liées au complexe par des interactions faibles. Une fois purifiées par la méthode TAP, les sous-unités des complexes protéiques d'intérêt sont identifiées par spectrométrie de masse.

Cette stratégie permet des études de complexes à grande échelle et permet en particulier, de caractériser le processus de dégradation des ARNs messagers, processus reste peu connu même si il contribue de manière importante au contrôle de l'expression des gènes. On peut aussi réaliser un double marquage, ou « Split TAP tag », qui consiste à marquer deux protéines d'un même complexe à l'aide d'étiquettes différents, ce qui permet de distinguer les formes « complètes » du complexe, contenant les deux partenaires, des autres formes incomplètes.

Cette technique représente un outil de plus dans l'étude des complexes à l'échelle du protéome entier. Cependant, la connaissance de la composition de complexes protéiques, même au niveau cellulaire, ne sera pas suffisante pour comprendre leur fonction. Le prochain niveau de l'analyse sera la définition de l'agencement interne des sous-unités dans les complexes dans un premier temps avant d'aller vers des analyses structurales plus détaillées. Néanmoins, cette approche peut dans certains cas représenter une alternative très intéressante à la technique d'immunoprécipitation.

Approches protéomiques pour la mise au point de nouveaux médicaments anti-herpès.

Jean-Jacques Diaz

Le virus HSV est un pathogène majeur qui peut développer deux types d'infection chez l'homme. Dans le monde, un milliard d'individus serait atteints d'infections herpétiques, dont 90% des malades du SIDA. Les traitements disponibles sont à base d'Acyclovir, un analogue de nucléoside qui inhibe la synthèse d'ADN viral, mais on constate l'émergence de résistance chez 5 à 30% des souches, d'où le besoin d'identifier de nouvelles cibles thérapeutiques pour combattre le virus.

Le virus HSV comprend deux types : HSV-1, responsable de l'herpès buccal et d'un tiers des cas d'herpès génital, et HSV-2, responsable de l'herpès génital. Lors de l'infection lytique, dans les cellules épithéliales, le génome viral exprime l'ensemble de ses 80 gènes, de manière séquentielle et ordonnée. L'expression du programme lytique aboutira à la mort cellulaire et à la production d'une nouvelle génération de particules infectieuses. Certaines de ces particules vont infecter les neurones sensoriels, où le génome viral s'établira en latence. Le génome viral peut rester en latence pendant des années, avant qu'un stress ne vienne le réactiver, ce qui provoquera le redémarrage d'une infection lytique et la réapparition du virus à la surface de l'organisme.

Des études des modifications transitoires du protéome de la cellule hôte pendant l'infection ont été menées. L'expression des gènes du virus de l'herpès de type 1 (HSV-1) s'accompagne d'une interruption sélective de la synthèse des protéines des cellules infectées. Tandis que la synthèse de la grande majorité de protéines cellulaires est inhibée juste après l'infection, plusieurs protéines cellulaires continuent à être synthétisées, même en phase tardive de l'infection. Puisque ces protéines cellulaires peuvent intervenir dans le cycle de vie du virus, des analyses ont été entreprises par électrophorèse bidimensionnelle sur gel de

polyacrylamide (2-D PAGE) par pour évaluer la proportion de ces protéines cellulaires. Des cellules humaines ont été infectées par HSV-1. À différents temps après infection, des protéines ont été marquées à la ³⁵S-Méthionine, puis des mesures ont été prises juste après le marquage.

Le taux de synthèse d'un ensemble de 183 protéines cellulaires de la cellule hôte, dont des protéines ribosomales, a été mesuré pendant l'infection, après séparation par électrophorèse 2-D PAGE. Comme prévu, HSV-1 induit une forte inhibition de la synthèse protéique chez la cellule hôte juste après l'infection. Cependant, la synthèse des protéines ribosomales de base et de celle d'une proportion élevée du sous-ensemble de protéines cellulaires analysées est soutenue ou stimulée pendant l'infection HSV-1, dont notamment les protéines de choc thermique et les protéines de la voie de biosynthèse des polyamines. Ces résultats permettent de penser que des inhibiteurs des protéines de cette voie de biosynthèse sont peut-être de bons candidats pour de nouveaux traitements contre le virus HSV-1.

Cette conférence illustre bien les premières étapes du processus de « Drug Discovery », élément central dans la plupart des créations d'entreprises de biotechnologies.

Comment la génomique peut aider à comprendre le pouvoir pathogène de la bactérie *Helicobacter pylori* ?

Hilde de Reuse

Helicobacter pylori (*HP*) est une e-proteobactérie à Gram - qui peuple l'estomac des primates, dont environ une personne sur deux à l'échelle de la planète. Elle est responsable de la plupart des infections bactériennes chez l'homme et provoque chaque année chez 0.3% des adultes et 3% des enfants de moins de cinq ans des infections aiguës du tube digestif, qui deviennent chroniques et peuvent dans certains cas évoluer en ulcères ou provoquer des lymphomes ou des adénocarcinomes, selon la souche bactérienne, le fond génétique de l'hôte et son hygiène de vie. Le traitement actuel contre l'infection par *HP* comprend deux antibiotiques (amoxicilline et claitromycine) associés à un inhibiteur de pompe à proton.

Le séquençage du génome de deux souches pathogènes chez l'homme (Souches J99 et 26695) a permis d'initier des études de génomique comparative par microarray et de post-génomique et qui ont pour but d'identifier des cibles thérapeutiques, des antigènes potentiels pour des vaccinations, mais aussi des biomarqueurs de pathogénicité. Ainsi, le génome de *HP* renferme un cluster de 30 gènes (Pathogenicity Island, ou PAI) responsables de la pathogénicité de la bactérie (gènes *cag*). Les protéines Cag induisent l'apoptose des cellules épithéliales et la sécrétion d'IL-8 qui déclenche une réaction inflammatoire. Il existe aussi plusieurs régions variables, les zones de plasticité (PZ), résultats de transferts de gènes horizontaux, qui regroupent des gènes essentiels pour l'adaptabilité de la bactérie et qui subissent de nombreux réarrangements. Ces gènes codent des transposases, des endonucléases, des protéines de la membrane externes ou des protéines impliquées dans la biosynthèse des lipopolysaccharides, et sont une source de variabilité qui confèrent à *HP* un avantage sélectif. Néanmoins, aucune corrélation entre le contenu du génome des souches bactériennes et une pathologie en particulier n'a pu être établie sur la base de l'étude de ces seuls deux génomes.

Outre des études de transcriptomique et de protéomique, des études d'interactomiques ont été menées afin de dresser une cartographie des interactions protéiques par la méthode des double-hybrides. Des banques génomiques des *HP* ont été constituées chez la levure, en utilisant des vecteurs renfermant un gène A pour une première banque, et des vecteurs

renfermant un gène B pour une deuxième banque, de telle sorte que toutes les protéines bactériennes exprimées dans la banque levures A sont hybridées avec la protéine A, et toutes les protéines bactériennes exprimées dans la banque levures B sont hybridées avec la protéine B. Le croisement des levures des deux banques permet de détecter l'interaction des deux protéines A et B par un système de gène rapporteur. L'interaction entre les protéines A et B n'étant rendue possible que par une interaction entre les deux protéines bactériennes hybridées, l'expression du gène rapporteur indique une interaction entre deux protéines bactériennes. Ces travaux ont permis l'élaboration d'une véritable plateforme d'interactomique, qui est à l'origine de la création de la société Hybrigenics.

Cette conférence montre un bon exemple de naissance d'une entreprise de Biotechnologies issue en premier lieu d'un progrès scientifique, puis d'une innovation technologique.

Conclusion

La génomique et les disciplines de 'post-génomiques' occupent aujourd'hui une place incontournable dans les sciences de la Vie et de la Santé, depuis la fin du 20^e siècle et l'apparition de nouveaux outils de séquençage des génomes, bien plus performant que ce qui existait auparavant, réduisant ainsi considérablement le facteur temporel, et rendant possible le séquençage de génomes entiers. Une fois ces génomes séquencés, néanmoins, est apparu le besoin de décrypter l'information génétique contenue dans ces génomes, pour exploiter ces données à des fins scientifiques ou médicales. D'où le développement exponentiel des disciplines de post-génomiques.

La transcriptomique étudie les profils d'expression spatio-temporelle des gènes, en terme de transcription en ARN messagers, notamment en réponse à un stimulus, comme à un traitement, ou dans un contexte pathologique. La transcriptomique permet donc se restreindre au premier niveau d'expression des gènes de façon précise, ce qui est très utile lorsque l'on cherche à détecter l'activation d'un gène, sans pour autant s'intéresser à sa fonction. C'est le cas lorsque le gène étudié est connu, et que l'on étudie sa régulation. C'est aussi le cas, à l'inverse, lorsque l'on veut identifier l'ensemble des gènes qui répondent à un stimulus, avant de pouvoir les caractériser dans un second temps.

De façon assez semblable, la protéomique s'attache, elle, à étudier les profils d'expression des protéines, produits finaux de l'expression des gènes. L'avantage ici, réside dans la stabilité relative des molécules étudiées, les protéines, par rapport aux ARNm's, qui sont particulièrement sensibles. De plus, se placer au niveau final de l'expression d'un gène permet de s'affranchir du biais d'analyse que constitue la différence possible, et même et fréquente entre la transcription d'un gène et la traduction de son messenger, voire l'expression de sa protéine à partir de vésicules de stockage, dans certains cas, comme celui des récepteurs transmembranaires.

Aussi, de nouvelles disciplines, de plus en plus spécialisées, apparaissent au fur et à mesure. C'est le cas de l'épigénomique, qui étudie les modifications de l'ADN qui n'affectent pas sa séquence, de l'interactomique, qui étudie les réseaux d'interactions protéiques au sein de la cellule, mais c'est également le cas de la métabolomique, qui étudie l'ensemble du métabolisme à l'échelle de la cellule et de la métabonomique, étude du métabolisme à l'échelle de l'organisme.

Toutes ces disciplines apportent leur lot de nouveaux outils de recherche, et profitent à l'ensemble des différents domaines de la Biologie. En immunologie, ces outils permettent par exemple de suivre la réponse de des lymphocytes à un immunogène par une approche transcriptomique, ou d'identifier, à l'aide de puces à ADN ou à ARN, des biomarqueurs spécifiques d'une pathologie par une approche protéomique, qui serviront de cibles potentielles pour des immunothérapies, ou qui permettront, par un suivi de leur expression dans un contexte pathologique d'établir des diagnostics, ou bien qui serviront d'indicateurs pour suivre la réponse à un traitement, dans des problématiques de développement de médicaments.

Un point essentiel à souligner également, à l'issue de ce cycle de conférences est l'importance du phénomène de conversion de progrès scientifiques en innovations technologiques puis en projets de création de start-ups de biotechnologies. Le sujet a presque toujours été au minimum suggéré, voire abordé par au cours des interventions. La génomique et les domaines qui en découlent ont permis la création de nombreuses entreprises de services ou de « Drug discovery » utilisant des technologies issues de la recherche en bioinformatique et en post-génomique (plateformes de transcriptomique, protéomique, interactomique, etc.).

Il reste, semble-t-il, un saut culturel à franchir car les chercheurs semblent encore méfiants vis-à-vis de la question du transfert de technologies, sans doute par méconnaissance du sujet. Il réconfortant de constater que cette question est omniprésente, car le parcours Immunotechnologies répond au besoin de ces futurs créateurs d'entreprises, qui auront besoin de trouver des scientifiques maîtrisant des techniques de pointes, tout en étant capable de s'adapter au fonctionnement en mode projet des start-ups de biotechnologies. Le secteur des biotechnologies ne semble émerger que maintenant, avec très peu d'entreprises cotées en bourse, de nombreuses structures de très taille qui peinent à grandir. A ce titre, j'ai le sentiment que les étudiants et futurs diplômés du Master Immunotechnologies ont de l'avance sur le marché, en France, tout du moins, où il n'est pas encore mûr, et qu'ils auront à la fois les compétences recherchées et l'expérience requise lorsque le marché sera près à les accueillir, très bientôt.