

# 21

---

## **BASES MOLÉCULAIRES ET STRUCTURALES DES FONCTIONS DU VIVANT**

*Président de la section*

Sylvain BLANQUET

*Membres de la section*

Christiane BRANLANT

Dominique BURNOUF

Nathalie DECLERCK

Gilbert DELEAGE

Zorah DERMOUN

Anne DIETRICH

Marc DREYFUS

Alain FILLOUX

David J. S. HULMES

Alain LESCURE

Daniel LOCKER

Hugues LORTAT-JACOB

Annie MOUGIN

Josette SCHMITT

Annette TARDIEU

Mariella TEGONI

Vincent VILLERET

Philippe WALTER

### **PRÉAMBULE**

L'intitulé de la section 21, « bases moléculaires et structurales des fonctions du vivant », indique de façon particulièrement claire que le cœur des activités de la section se situe dans l'élucidation, l'analyse et la compréhension des aspects moléculaires et structuraux nécessaires à l'établissement d'une fonction et ce, quels que soient le niveau d'organisation et la complexité nécessaires à l'établissement de cette fonction. D'un côté la compréhension des propriétés des briques élémentaires gagne en profondeur, de l'autre le nombre de complexes macromoléculaires de grande taille, dont l'étude peut aujourd'hui être abordée, s'élargit. Ces recherches vont de pair avec un élargissement considérable des moyens d'études, de leur qualité, de leur automatisation.

Cet ancrage « structure-fonction » particulièrement solide a toujours permis à la section 21 de développer des activités interfaciales importantes sans perdre son identité. Dans un contexte biologique en évolution rapide, en particulier en raison des progrès réalisés dans le séquençage des génomes, les priorités sont, elles aussi, en train de changer. La « génomique structurale » a remplacé la « biologie structurale ». Depuis elle devient même « intégrative ». La section 21 entend associer les outils les plus performants en provenance de physique,

chimie, biologie, mathématiques appliquées et informatique afin d'approfondir la compréhension des mécanismes moléculaires tout en s'orientant davantage vers des stratégies d'exploration systématique, qu'il s'agisse des codes de repliement des macromolécules ou de catalogues divers : liste des repliements existants en lien avec leurs fonctions possibles, liste des objets constitutifs d'un assemblage ou d'un organisme, liste de leurs modifications post-transcriptionnelles ou post-traductionnelles, liste de leurs interactions. Sur le moyen terme, c'est un mouvement vers la biologie systémique de la cellule qui est en cours, le défi étant d'élucider les structures des assemblages macromoléculaires au sein des cellules, de comprendre leur architecture, la dynamique liée à l'activité et d'assurer le suivi de la localisation sub-cellulaire tout en gardant l'information moléculaire fine. Une appropriation concomitante de nouvelles approches est en cours (protéomique, puces et chips diverses, études en solution ou sur molécules uniques, imagerie, etc.).

Parmi les grandes « fonctions », qui intéressent particulièrement notre section, on peut citer : division cellulaire, réplication et surveillance du génome, transcription, maturation et fonction des ARNs, traduction, repliement, maturation et adressage des protéines, catalyse, et aussi différenciation cellulaire, voies de signalisation, motilité, adhésion, etc. La section a également une longue tradition de recherche dans le domaine de la microbiologie, en particulier dans l'étude des mécanismes de la pathogénie et des systèmes de sécrétion. Parmi les nouveaux défis adressés à la communauté de la section 21, on trouve : le décryptage des régulations post-transcriptionnelles et traductionnelles, qui renforcent la place de l'ARN dans les mécanismes cellulaires, le développement de la glycobiologie, longtemps laissée de côté faute d'outils et de cadres conceptuels adaptés, et l'élucidation de nombreux mécanismes épigénétiques, dont certains aspects fonctionnels semblent aujourd'hui pouvoir être compris en termes de structure.

Enfin, l'étude des relations structure-fonction se doit de déboucher aussi sur la com-

préhension des dysfonctionnements et leur correction. Ces dysfonctionnements peuvent avoir une cause infectieuse ou génétique, ou peuvent résulter de désordres structuraux comme, par exemple, dans la maladie d'Alzheimer. Les préoccupations de la section 21 rejoignent celles de beaucoup de nos collègues biologistes et chimistes quand il s'agit de concevoir de nouveaux médicaments. À ce sujet, on peut attirer l'attention sur le cas des récepteurs membranaires qui sont des cibles pharmacologiques privilégiées. Les défis que la section 21 doit relever concernent aussi la compréhension à l'échelle moléculaire d'un nombre croissant de pathologies qui dérivent de risques sanitaires, chimiques ou microbiologiques liés à l'environnement.

## **1 – STRUCTURE – FONCTION – RÉACTIVITÉ DES MACROMOLÉCULES ET DE LEURS COMPLEXES**

### **1.1 REPLIEMENT, STABILITÉ, DYNAMIQUE DES « BRIQUES ÉLÉMENTAIRES »**

Qu'il s'agisse de protéines, d'acides nucléiques ou de glycanes, beaucoup reste à faire au niveau le plus simple, celui du repliement et de la stabilité des macromolécules, ce qu'on a pris l'habitude d'appeler les « codes de repliement ».

#### **Les acides nucléiques et le monde des ARNs**

Les découvertes des dernières années ont révélé une ampleur insoupçonnée du rôle des ARN dans la plupart des processus cellulaires,

avec en particulier, la découverte des ARN non codants impliqués dans la régulation de l'expression des gènes. On estime qu'au moins 30 % des gènes humains sont régulés de cette manière. Par ailleurs, la complexité des étapes de maturation des transcrits entraîne une diversification de l'expression génétique elle aussi jusque-là sous-estimée. Ainsi, la majorité des gènes humains ne codent pas pour une protéine unique mais pour un ensemble de protéines. Ces découvertes remettent l'ARN au premier rang des préoccupations des biologistes.

### **Structure et dynamique des ARN**

Comme les protéines, les ARN adoptent des structures tertiaires très complexes et, d'ailleurs, ils peuvent eux aussi posséder des propriétés catalytiques. Leur particularité est d'avoir des structures 3D extrêmement dynamiques, ceci du fait qu'ils ne sont constitués que de 4 éléments de base, capables d'interagir deux à deux. D'où de nombreuses conformations possibles pour une même molécule, un grand nombre de transitions structurales et une grande capacité à établir des interactions intermoléculaires transitoires lors de leur activité. Ces transconformations sont, de plus, souvent facilitées par l'intervention de protéines chaperons, d'ARN hélicases qui hydrolysent l'ATP et de protéines intrinsèquement dépliées, qui se structurent en même temps que leur partenaire ARN. À cela, il faut ajouter l'existence de complexes protéiques assurant la spécificité d'assemblage des ARN avec leurs partenaires protéiques.

Les règles thermodynamiques régissant la structuration des ARN ne sont que partiellement connues, aussi n'est-il pas possible de prédire la structure 3D, ni même 2D, sur le seul examen de leur séquence. La prédiction de ces structures requiert d'enregistrer des données, tant phylogéniques qu'expérimentales (analyses structurales en solution, RMN, diffraction X), afin de constituer une banque des motifs structuraux possibles, qui peuvent être adoptés par un ensemble fini de séquences.

Les structures 3D déjà établies pour les ARN ribosomiques faciliteront la constitution de telles banques. Il faudra enfin optimiser les méthodes informatiques de prédiction des structures 3D d'ARN à partir de ces banques.

### **Modifications post-transcriptionnelles des ARN**

Tous les ARN eucaryotes et beaucoup d'ARN procaryotes, sont produits sous la forme de précurseurs, qui subissent un ensemble de modifications pour être convertis en molécules actives (clivages pour la plupart, élimination des introns par épissage pour les ARN pré-messagers, et modifications de certaines bases et des riboses). Par exemple, plus de 100 modifications distinctes ont été identifiées chez les ARN de transfert. Certaines de ces modifications sont catalysées par une enzyme ou une particule ribonucléoprotéique, d'autres reposent sur une cascade de réactions enzymatiques. Ces nucléotides modifiés changent la stabilité des ARN et leur capacité d'interaction avec leurs partenaires, d'où l'importance de développer des méthodes rapides d'identification et d'étudier l'origine des modifications et les conséquences fonctionnelles des modifications.

### **Les protéines**

#### **Séquence et repliement**

Si les relations structure-fonction dans le monde des protéines sont étudiées depuis longtemps, il n'en demeure pas moins que des questions cruciales demeurent, encore aujourd'hui, sans réponse. L'exemple le plus frappant est, bien sûr, la très grande difficulté où nous sommes de prédire la structure à partir de la séquence et, réciproquement, d'énumérer les séquences compatibles avec un repliement donné. Nous ne sommes pas encore capables de prédire à coup sûr les conséquences structurales d'une mutation ponctuelle et, à fortiori, les conséquences en termes de modi-

fications de propriétés physico-chimiques (stabilité, solubilité, etc.) et fonctionnelles (activité, liaison de substrats, etc.). Le processus de repliement natif, les chemins et étapes de dénaturation-renaturation, sont d'autres questions fondamentales, non élucidées.

### **Aspects thermodynamiques ou cinétiques**

Beaucoup de développements récents concernent en particulier :

- l'étude (théorique, expérimentale) des paysages énergétiques, des processus de dénaturation-renaturation, des conditions de réversibilité (thermodynamique et cinétique) ;

- la comparaison de la structure 3D dans le cristal à la structure et sa dynamique en solution et dans la cellule, en conditions proches des conditions de fonctionnement/dysfonctionnement *in vivo* ;

- l'étude des interactions à courte distance qui déterminent les diagrammes de phase, la cristallogénèse, les propriétés spécifiques en situation de solution concentrée (« crowding ») ;

- l'analyse des interactions de contact qui permettent, *in vivo*, la formation d'assemblages fonctionnels et, *in vitro*, la cristallisation, etc.

### **Le cas des protéines membranaires**

Là encore, beaucoup reste à faire. Les protéines membranaires représentent environ 25 % du génome humain. Nous l'avons déjà dit, une grande majorité des cibles thérapeutiques sont des protéines de membrane. A titre d'exemple, les récepteurs couplés aux protéines G (RCPG), constituent la cible de 50 % des médicaments connus. Moins d'une centaine de RCPG a été exploitée à ce jour et les 400 RCPG restant (hors récepteurs olfactifs) constituent autant de cibles thérapeutiques potentielles. Les protéines membranaires sont largement sous-représentées dans les bases de données structurales (moins de 50 structures sur les 17 000 dans la PDB en 2006). La

grande majorité des structures résolues sont celles de protéines membranaires fortement exprimées naturellement. Les interactions moléculaires entre la chaîne peptidique et les lipides sont essentielles pour le repliement des domaines membranaires. Elles restent à ce jour largement inconnues et peu étudiées.

### **Modifications post-traductionnelles**

La mise en évidence de l'importance fonctionnelle des phosphorylations, méthylations, acétylations, ubiquitylations, aminoacylations et sumoylations des protéines est l'une des avancées récentes les plus spectaculaires. Les progrès ont bénéficié des possibilités offertes par la protéomique. Grâce aux modifications post-traductionnelles, une même séquence codante peut mener à un grand nombre de protéines aux propriétés physico-chimiques et fonctionnelles différentes. On parle déjà d'un « nouveau code » à déchiffrer.

### **Les protéines intrinsèquement désordonnées**

Les régions naturellement flexibles ou « désordonnées » dans les protéines sont impliquées dans beaucoup de processus biochimiques importants, dont la reconnaissance moléculaire, la transduction du signal, etc. La stabilité, la solubilité et les capacités de reconnaissance spécifique de ces domaines structuraux posent des questions fondamentales d'un point de vue physico-chimique et thermodynamique. Leur analyse suppose la mise en place de méthodes pour suivre les transitions ordre-désordre des protéines (citons par exemple l'utilisation des couplages dipolaires résiduels obtenus par RMN, la diffusion aux petits angles, etc.).

### **Les structures de type amyloïde**

Les transitions ordre-désordre impliquées dans certaines pathologies neuro-dégénératives, comme la maladie d'Alzheimer, méritent

une attention particulière. L'un des défis majeurs de notre temps sera de comprendre la cause et l'évolution de ces maladies afin d'aider à diriger le développement de thérapies.

### **Les protéines enzymatiques**

L'enzymologie moderne ne peut plus se concevoir sans un appui sur la structure 3D des protéines, sous forme libre ou en association avec leurs substrats et/ ou cofacteurs. La connaissance de ces structures permet, en combinaison avec les approches plus classiques de l'enzymologie, de formuler des hypothèses sur les bases moléculaires de la spécificité des substrats et aussi sur la chimie de la catalyse. Ces hypothèses peuvent ensuite être mises à l'épreuve de l'expérience grâce à la mutagenèse dirigée et à l'étude structurale et fonctionnelle des variants produits. Notons l'importance de l'interface avec la chimie dans ce domaine, qui permet de produire des analogues de substrats pour l'obtention de complexes mimant un état de transition. Notons aussi, la montée en puissance de la RMN 3D dans ce domaine, qui permet d'accéder aux aspects dynamiques de la catalyse et de décrire les états conformationnels, parfois multiples, des systèmes enzymatiques. Enfin, l'application des outils de la mécanique quantique aux structures ainsi établies permet aujourd'hui une prédiction des étapes chimiques de la catalyse enzymatique, prédiction qui peut être validée expérimentalement, par exemple par des approches type stopped-flow ou par des approches thermodynamiques très fines basées sur la calorimétrie. Soulignons enfin la possibilité de suivre directement certaines réactions enzymatiques dans le cristal par diffraction des RX. L'ensemble de ces approches peut s'appliquer aux enzymes de nature purement protéique mais aussi aux structures catalytiques plus complexes, telles que les complexes ARN-protéines.

### **Enzymologie sur molécule unique**

On en parle beaucoup. Par rapport à l'enzymologie «classique», on perçoit les atouts suivants :

– i) l'observation des molécules individuelles permet de détecter d'éventuelles sous-populations, ou bien d'éliminer la fraction des molécules qui sont inactives. Elle permet de connaître la variance des paramètres physico-chimiques de la réaction ;

– ii) elle donne accès à des paramètres nouveaux : force et déplacement ;

– iii) elle isole l'acte catalytique des étapes de liaison préalables.

Tout cela est innovant du point de vue des connaissances fondamentales. En revanche, les difficultés expérimentales et les artefacts restent pour l'instant nombreux et les applications limitées.

### **La glycobioologie**

La glycosylation des protéines est une des modifications post-traductionnelles les plus fréquentes (plus de la moitié des protéines sont glycosylées) et mérite un chapitre à part.

Les monosaccharides constituent les briques de base d'enchaînements plus complexes formant des oligo- ou des poly-saccharides. Il s'agit de molécules, parfois de grande taille, pouvant comporter plus de 200 unités, organisées en différents types de combinaisons. Il existe donc une grande diversité moléculaire, liée à la structure même de ces molécules et à leur mode d'enchaînement : si 2 acides aminés (ou 2 nucléotides) peuvent donner 2 peptides (ou 2 oligonucléotides), 2 pentoses permettent de donner 20 disaccharides. Avec 3 unités, on générera 6 peptides ou 6 oligonucléotides différents, mais c'est déjà 1 056 structures oligosaccharidiques qu'il devient possible d'imaginer et ainsi de suite.

Ainsi, comme on l'a fait pour le code génétique et les protéines, il faudra déchiffrer le sens de ce nouveau code, dont on devine l'extrême complexité. Pour les glycannes, il s'agit en effet de comprendre les bases de leurs structures, de leurs potentiels d'interaction, de leurs fonctions biologiques et de l'information qu'ils contiennent ainsi que des

mécanismes par lesquels ils participent à la réactivité des macromolécules qui les portent. Ces molécules n'étant pas directement codées par des gènes, les outils traditionnels de la biologie moléculaire, qui ont révolutionné l'étude des acides nucléiques et des protéines, n'ont eu jusqu'à présent qu'un impact faible sur cette discipline, dont les progrès restent tributaires de développements méthodologiques et conceptuels à venir.

## 1.2 LES GRANDS COMPLEXES MACROMOLÉCULAIRES ET LEURS FONCTIONS

### Grands complexes et machineries moléculaires impliquées dans les processus informationnels

Les machineries moléculaires impliquées dans les processus informationnels (réplication, transcription, maturation et dégradation de l'ARN, traduction) sont des assemblages protéiques ou nucléoprotéiques de grandes tailles. La structure de quelques unes des machineries les plus anciennement identifiées, qui sont impliquées dans la transcription et la traduction, est maintenant connue.

#### Complexes transcriptionnels

La détermination de la structure à haute résolution de l'ARN polymérase II eucaryote a été une étape importante, faisant suite à de longues années d'études de cette enzyme. Un autre exemple, moins avancé, est celui du facteur multi sous-unités TFIIF, qui intervient à la fois dans le démarrage de la transcription et dans la réparation de l'ADN par excision de nucléotides et qui est décrit de plus en plus précisément. Actuellement, les travaux tant fonctionnels que structuraux dans ce domaine sont axés sur les facteurs protéiques qui régulent l'activité transcriptionnelle, avec en particulier, l'étude déjà bien avancée des récepteurs

nucléaires, qui ont des rôles très larges dans la régulation de l'expression des gènes et aussi dans la cancérogénèse.

#### Complexes traductionnels

Les progrès des connaissances sur la machinerie traductionnelle ont été fulgurants au cours des dernières années. Différentes structures de sous-unités ribosomiques ou de ribosomes entiers ont été élucidées par cristallographie à haute résolution. La structure à une résolution de 2,8 Å du ribosome de *Thermus thermophilus*, complexé à l'ARNm et aux ARNt correspondants, vient d'être publiée. Ces structures servent actuellement de base à l'étude d'assemblages reflétant les nombreuses étapes successives de la traduction. Nous allons ainsi pouvoir disposer d'une série de clichés du ribosome lors de son fonctionnement, qui pourront par exemple être utilisés dans des calculs de simulation pour mieux comprendre la dynamique du système. De telles études illustrent le potentiel des combinaisons d'approches structurales à différentes résolutions (cristallographie X, RMN, cryomicroscopie électronique, SAXS, SANS, spectrométrie de masse, etc.). Dans un futur proche, les avancées prévisibles sont la détermination de la structure 3D des sous-unités ribosomiques eucaryotes et l'acquisition d'une vision globale à l'échelle atomique des différentes molécules périphériques impliquées dans la synthèse protéique et de leurs interactions. On peut en espérer une élaboration plus rationnelle d'antibiotiques dirigés contre la synthèse protéique.

#### Complexes assurant la maturation des ARN

Ces complexes ont été découverts récemment. Chez les procaryotes, l'étude des complexes assurant la maturation de l'ARNr ou la dégradation de l'ARNm (le « dégradosome d'ARN ») n'est pas achevée, et la pertinence de *E. coli* comme organisme modèle fait actuellement débat. Chez les archées, l'établissement

de la structure 3D des sRNP assurant la modification des nucléotides a progressé rapidement. Ainsi, ces catalyseurs contenant des protéines et un ARN guide, qui définit la position à modifier, vont être abordés sous l'angle de l'enzymologie.

Chez les eucaryotes, l'élimination des introns des ARN pré-messagers est catalysée par les épissosomes. L'analyse structurale récente de ces machines par microscopie électronique représente un exemple frappant du potentiel de cette approche pour l'étude de macrostructures trop instables et dynamiques pour être cristallisées. Ces données, jointes à l'analyse de la structure 3D de composants et facteurs de l'épissage et à des études biochimiques et génétiques, permettront d'accéder au mécanisme fin de l'épissage et de sa régulation, ainsi qu'à ses interactions avec la transcription et le transport des ARNm.

Les macrocomplexes impliqués dans la maturation des ARN ribosomiques sont les moins bien définis et tout n'est pas encore compris dans les étapes de biogénèse des ribosomes qui mettent en jeu transcription, clivage, modification du pré-ARNr et assemblage des protéines ribosomiques. L'étude de la liaison entre les étapes post-transcriptionnelles de cette synthèse et la progression du cycle cellulaire ou l'initiation de la traduction constituent aussi des défis actuels.

### **Systèmes de contrôle de qualité**

Le rôle des machineries de dégradation des ARN et des protéines dans les systèmes de surveillance cellulaire qui permettent d'éliminer aussi bien les ARN messagers contenant des codons non sens prématurés, que les particules pré-ribosomiques défectueuses, représentent également un champ actuel de recherche.

D'une manière générale, la compréhension des processus informationnels chez les eucaryotes, dont les dérèglements engendrent de nombreuses pathologies, a des implications importantes en santé humaine. Il est également

à noter que les archées, dont les processus informationnels ressemblent pour certains points à ceux des eucaryotes, permettent de fournir un matériel biologique facilement manipulable pour les études structurales, contribuant ainsi à l'avancée de nos connaissances.

### **Complexes réplicatifs et complexes impliqués dans la surveillance du génome**

La découverte récente de l'existence, chez tous les organismes, de multiples ADN polymérases capables notamment d'incorporer des nucléotides en face de lésions bloquantes et/ou de réactiver une fourche de réplication bloquée, a profondément modifié les modèles de la réplication et de la mutagenèse, mais aussi ceux des mécanismes de traitement des cassures de l'ADN et des processus de recombinaison, par exemple des gènes des immunoglobulines. Cette découverte a également relancé la problématique de la dynamique des multiples partenaires au sein des réplisomes.

En effet, les différentes polymérases, mais aussi d'autres facteurs intervenant notamment dans les processus de réparation, interagissent tous avec des « anneaux de processivité » entourant l'ADN. Les mécanismes gouvernant leur accès à ces plateformes et leurs échanges sont encore mal connus. L'existence d'anneaux alternatifs suggère une diversité et une complexité encore plus grande de ces processus. Ces mécanismes d'adressage et d'échange, ainsi que leurs modes de régulation (notamment par les voies d'ubiquitination), sont des champs d'investigation très compétitifs qui ouvrent sur des domaines comme les cascades de régulation du cycle cellulaire (cell cycle checkpoints) et tous les processus de surveillance du génome. Le fonctionnement intégré de ces grands complexes multiprotéiques intervenant à la fois dans la réplication, la réparation du génome et la transcription est aussi concerné. Les études dans ces domaines ouvrent également la voie à l'identification de

nouvelles cibles pharmacologiques pour le traitement d'affections diverses comme le cancer, les infections septiques, les processus inflammatoires ou ischémiques.

## **Chromatine et noyau ou les mécanismes épigénétiques**

L'activation ou la répression d'un ensemble de gènes dans une cellule sont généralement établies pendant un seul cycle cellulaire sous l'action de différentes voies de signalisation, notamment par la liaison des facteurs de transcription aux régions régulatrices des gènes. Pour transmettre l'information à ses cellules filles, une cellule doit mémoriser les changements d'expression des gènes au cours d'un cycle cellulaire ; cela doit pouvoir se faire par des mécanismes épigénétiques stables au travers des mitoses. D'après l'idée qui prévaut actuellement, cette mémoire serait gardée dans la structure de la chromatine avec deux états possibles : un état ouvert pour activer le gène, un état fermé pour le réprimer. Ces états sont caractérisés par des modifications soit au niveau de l'ADN, soit au niveau des protéines chromatinienne et des histones. L'information ou code histone apportée par ces dernières est une information supplémentaire par rapport à la séquence nucléotidique de l'ADN.

D'autres groupes importants de protéines non histones peuvent modifier la structure de la chromatine en formant des complexes de grande taille : certains maintiennent la répression des gènes cibles en créant des domaines chromatinien très compacts dans lesquels l'ADN est peu accessible, d'autres assurent le maintien de l'activation des gènes en remodelant la chromatine pour lui donner une structure ouverte accessible aux facteurs de transcription.

L'identification de toutes les modifications de la structure de la chromatine susceptibles d'établir un nouveau code (l'épigénome) et l'étude de leur dynamique, constituent un véritable défi dans le domaine. Les problèmes majeurs du contrôle épigénétique au cours du

développement restent à résoudre. Ils concernent notamment :

- i) la mise en place des états alternatifs des gènes : activité ou inactivité ;
- ii) le maintien de ces états après le passage de la fourche de réplication.

On peut espérer dans les années à venir mieux comprendre au niveau moléculaire le fonctionnement de ce code épigénétique et rendre compte au niveau mécanistique des nombreux phénomènes sortant du cadre de l'hérédité mendélienne classique observés actuellement dans la transmission des gènes.

## **Noyau interphasique et dialogue chromatinien**

La technique de fluorescence *in situ* (FISH) a permis de montrer que les chromosomes n'étaient pas localisés d'une façon aléatoire dans le noyau interphasique mais au niveau de territoires chromosomiques encore mal définis et pour le moins controversés. Au sein des chromosomes, différentes régions peuvent établir ou, au contraire, ne pas établir de contacts en fonction du moment du cycle cellulaire. Ce « dialogue » entre différentes régions des chromosomes est encore peu étudié ; quels sont le rôle et la dynamique de ces appariements ectopiques ? Le développement des techniques de l'imagerie *in vivo* devrait pouvoir apporter une réponse.

## **Contrôle de l'architecture cellulaire et de la motilité**

### **Cytosquelette**

Maintenant que la plupart des composants du cytosquelette ont été identifiés, le défi est de comprendre leur fonctionnement dans toute une série de processus, division cellulaire, migration, organisation et trafic intracellulaire. Il reste à comprendre au niveau moléculaire les mécanismes qui régulent l'assemblage et la dissociation des microtubules,



des filaments d'actine et des filaments intermédiaires, en utilisant les méthodes de la biologie structurale (y compris la tomographie électronique à 3D), la reconstitution *in vitro* et l'imagerie (FRET, FLIP, etc.). Il serait également important de comprendre le fonctionnement au niveau moléculaire des différents moteurs moléculaires (myosines, kinésines, dynéines) ainsi que les mécanismes de reconnaissance et de transport directionnel de cargos. De telles études pourraient avoir des retombées pratiques importantes dans le domaine des nano-biotechnologies et des nano-machines.

### **Matrice extracellulaire**

La matrice extracellulaire donne aux cellules une armature structurale. Elle influence également la migration cellulaire, l'adhésion, la différenciation et la prolifération. Il s'agit d'un réseau macromoléculaire complexe constitué de protéines structurales, d'enzymes, de facteurs de croissance et de cytokines. Les protéines structurales de la matrice (collagènes, élastine, fibronectine, laminine, fibrilline, etc.) sont de grande taille. Elles sont faites de domaines structuraux multiples et s'auto-assemblent en divers niveaux d'organisation (fibres, réseaux, etc.). Comme les protéines du cytosquelette, elles sont difficiles à étudier par les méthodes classiques à haute résolution (cristallographie, RMN). De nouvelles approches sont nécessaires pour comprendre les structures et assemblages des protéines de la matrice, comme par exemple la cryo-tomographie électronique ou la diffusion des rayons X aux petits angles. Cette compréhension devrait déboucher sur d'importantes applications dans le domaine, par exemple, de l'ingénierie des tissus (biomatériaux) ou celui des maladies dégénératives (ostéoarthritis).

### **Nanomachines**

Une meilleure compréhension des mécanismes d'auto-assemblage du cytosquelette et des protéines de la matrice extra-cellulaire devrait avoir des retombées importantes dans

la conception de nouvelles structures capables de s'auto-assembler, avec des applications en nanobiotechnologies (par exemple dans l'adressage des médicaments). Les mécanismes par lesquels les cellules répondent et réagissent à leur environnement (différenciation, migration, etc.) présentent aussi un intérêt considérable. La possibilité de fabriquer des substrats nanostructurés avec des propriétés physico-chimiques définies sera particulièrement utile dans ce domaine.

## **Communications intercellulaires**

### **Interactions cellule-matrice**

Un domaine de recherche particulièrement actif est l'étude des mécanismes par lesquels les cellules réagissent avec leur environnement extracellulaire. Ce domaine a des implications importantes dans, par exemple, le cancer (différenciation cellulaire, adhésion, migration et prolifération). Différents récepteurs de surface ont été identifiés (intégrines, héparines, protéoglycannes, etc.) mais on sait peu de choses sur les interactions qui contrôlent la transmission du signal au niveau moléculaire (intérieur-extérieur et vice-versa) et, donc, sur le comportement cellulaire. De nouvelles approches seront nécessaires pour caractériser les structures des complexes macromoléculaires impliqués et leur dynamique à la fois dans le cytosol (interactions avec les protéines du cytosquelette) et à l'extérieur de la matrice extra-cellulaire (adhésions focales).

### **Interactions cellule-cellule**

Au-delà des interactions cellule-matrice, il faut appréhender les mécanismes d'interactions cellule-cellule et leur régulation par des protéines de surface comme les cadhérines; pour cela, on doit rassembler davantage de données sur les jonctions intercellulaires (comme les desmosomes). De nouvelles avan-

cées méthodologiques comme les spectroscopies à force atomique permettront d'accéder à une mesure de telles interactions à l'échelle de la molécule unique.

### **Structure des complexes multiprotéiques impliqués dans les voies de signalisation**

Chez les animaux et les végétaux, les processus physiologiques tels que mouvements, reproduction, métabolisme, développement embryonnaire et différenciation cellulaire sont régis par des communications extra- et intracellulaires. La diversité de ces communications s'explique par l'association de nombreuses molécules avec des récepteurs différents, chacune d'elles transmettant un message à une ou plusieurs voies de signalisation. Celles-ci aboutissent *in fine* à l'expression de divers gènes dans une cellule. Dans les années à venir, seule la collaboration d'équipes de domaines différents (de la chimie à la génétique) et le développement d'approches globales pour étudier le Transcriptome, le Protéome, l'Interactome, etc. permettront de comprendre l'organisation et les supports moléculaires aboutissant à l'activation d'une série de gènes spécifiques à partir de messages reçus par la membrane cellulaire. Il faudra aussi trouver comment les différentes voies de signalisation sont coordonnées et régulées via des interactions moléculaires. Un des défis sera de déterminer la structure des complexes protéines-protéines, protéines-petites molécules et protéines-ADN pour comprendre la modulation de l'activité des gènes en réponse aux différents messages. Un autre sera de déterminer le rôle de la dynamique des complexes dans la mise en œuvre de leur fonction. Les retombées pourront être intéressantes non seulement pour la recherche fondamentale mais aussi pour la recherche appliquée : au niveau pharmaceutique par exemple, cela permettra de cibler la recherche de petites molécules capables d'orienter une voie de signalisation dans un sens ou dans un autre.

### **L'interactome ou les communications intra-cellulaires**

Ces dernières années, l'étude des interactions protéines-protéines a donné naissance à une masse importante d'informations sous la forme de banques d'interactions (interactome). Ces données sont un préliminaire obligé pour qui voudra accéder à une vision systémique de la cellule. Elles doivent à l'avenir permettre :

- i) de décrire d'un point de vue quantitatif et qualitatif l'organisation de l'interactome ;
- ii) de fournir un modèle (automate) du fonctionnement de la cellule. Les données disponibles actuellement sont encore insuffisantes et encombrées de résultats non validés *in vivo*.

Il faut développer dans ce domaine de nouveaux outils destinés à croiser rapidement les résultats obtenus avec d'autres techniques. Aujourd'hui, nous disposons principalement de celle du double hybride. Il faudra également tenir compte du fait que les interactomes étudiés représentent la plupart du temps une vue statique des interactions et que la localisation cellulaire des interactions est encore inconnue. Dernier point et non des moindres, il est nécessaire de modéliser ces réseaux d'interactions si l'on veut qu'ils contribuent à une approche prédictive des fonctions cellulaires.

Cette thématique s'applique aussi bien aux cellules eucaryotes qu'aux cellules procaryotes et peut également être étendue à la communication inter-cellulaire ou reliée au métabolome. En effet, en microbiologie, on s'intéresse de plus en plus à des consortia de microorganismes, dont certains peuvent être cultivés en culture pure. Il serait intéressant de connaître le réseau intercellulaire qui s'établit par exemple dans un biofilm ou dans un consortium de microorganismes dans un biotope complexe reconstitué.

Le contrôle des réseaux d'interaction par la conception de modulateurs de réponses permettra de mieux comprendre, voire de maîtriser, la réponse cellulaire globale.

## Vers le « glycome » et la « glycomique » ?

Les glycannes jouent un rôle très général qui concerne l'expression, le repliement, le trafic, la localisation et la durée de vie des protéines auxquelles ils sont liés. Ces molécules, largement réparties dans le vivant, sont impliquées dans de nombreux processus de reconnaissance et dans la formation de complexes multimoléculaires. Elles contribuent à donner à la cellule (et à l'organisme) les bases de sa robustesse et de son adaptabilité mais également celles de sa complexité.

Longtemps considérés pour leurs aspects structuraux et de soutien (la cellulose de la paroi végétale, le peptidoglycane de la paroi bactérienne, les glycosaminoglycanes de la matrice extracellulaire chez les animaux) les glycannes interviennent de façon très importante dans tous les grands processus de régulation biologique ou les mécanismes de reconnaissance au niveau moléculaire. Ce faisant, les glycannes participent à la plupart des grandes fonctions physiologiques (développement embryonnaire, adhésion, migration et différenciation cellulaire, régulation du système immunitaire, régulation de l'activité de nombreuses protéines, des récepteurs, des voies de signalisation, etc.). Enfin, leurs modifications ont été observées dans un grand nombre de pathologies (maladies infectieuses, désordres inflammatoires, maladies auto immunes, maladies neurodégénératives et amyloïdes, troubles cardiovasculaires, cancer, etc.). Une étude détaillée de la glycosylation s'avère ainsi pertinente pour comprendre le fonctionnement normal ou pathologique des objets biologiques.

Récemment le « glycome » a été défini comme le registre complet des molécules glucidiques d'un organisme, d'un tissu ou d'une cellule donnée, et la « glycomique » comme l'étude de la structure et des fonctions des glycannes ainsi que des enzymes participant à leur biosynthèse ou à leur dégradation. La glycoprotéomique est, quant à elle, une méthode de protéomique différentielle qui vise à élucider le registre complet des glycoprotéines et

des glycoformes présentes dans un tissu ou une cellule, dans un état physiologique donné. Elle propose l'identification de la protéine et l'établissement de l'hétérogénéité glycannique au niveau des différents sites de glycosylation. L'établissement du glycome suppose, bien entendu, l'étude structurale des glycannes, mais aussi la compréhension des mécanismes de leur biosynthèse (par exemple au cours de la différenciation, de la prolifération, de la migration, ou pour faire face à des organismes pathogènes ou des changements du milieu) et de leurs fonctions. Il s'agit là d'un enjeu nouveau et de grande importance.

En raison de la sophistication structurale des glycoconjugués (nature des monosaccharides, types de liaisons, séquences, anomérie, présence d'aglycones tels que sulfate, phosphate...), les approches de glycoprotéomique font appel à une grande diversité de procédés parmi lesquels la spectrométrie de masse et la RMN occupent une place importante.

## 2 – MICROBIOLOGIE

L'étude de la structure, de la fonction et de la réactivité des macromolécules et de leurs complexes génère une connaissance fondamentale pour la compréhension des mécanismes moléculaires qui sont à l'origine de tout le monde vivant.

Dans de nombreux cas, on l'a vu, le dysfonctionnement de ces macromolécules, ou le mauvais assemblage de leurs complexes, entraîne des pathologies graves chez l'homme. Les laboratoires de la section 21 utilisent de nombreux modèles cellulaires, allant des procaryotes à toutes espèces d'eucaryotes. Néanmoins, dans ce paragraphe, nous avons mis en exergue l'intérêt d'aborder l'étude des édifices macromoléculaires complexes chez les microorganismes. Si ces organismes modèles simples permettent d'avancer rapidement dans l'analyse, leur étude répond également à un enjeu sociétal fort dans le domaine de la

santé publique et, ce, au moment où l'on voit resurgir des problèmes graves liés à l'émergence ou la ré-émergence de maladies infectieuses.

L'étude des micro-organismes a été et continue d'être à l'avant-garde d'une grande variété d'avancées fondamentales importantes pour l'élucidation des mécanismes moléculaires et cellulaires essentiels à la vie. La microbiologie a ainsi contribué à la biologie cellulaire, la biochimie, l'enzymologie, etc. La relative facilitée d'investigation des systèmes modèles, microbes et virus, permet l'utilisation optimale de beaucoup d'approches.

Si la microbiologie souffrait quelque peu d'un manque de reconnaissance dans les années 90, elle est à présent à un tournant clé de son développement. Les raisons en sont variées et complexes, mais apparaissent évidentes, par exemple sous l'angle de la santé publique. En effet, les maladies infectieuses tuent à ce jour plus de 14 millions d'êtres humains chaque année, 90% de ces populations vivant dans les pays en voie de développement. Entre 1972 et 1997, parmi les 1 450 drogues introduites sur le marché durant cette période, seules 13 drogues développées pour traiter les infections (principalement parasitaires) qui affectent spécifiquement les populations pauvres ou marginalisées ont été mises sur le marché. Les résistances aux antibiotiques sont en constante augmentation et on redoute dans un avenir proche l'apparition de souches extrêmement, voire totalement, résistantes à tous les antibiotiques connus (cas notamment de la tuberculose). Dans nos pays, l'impact des maladies nosocomiales est également en constante augmentation. Cet impact est particulièrement inquiétant à la lumière des résistances aux antibiotiques développées par ces souches pathogènes. Enfin, la possibilité d'utiliser des bactéries ou des virus pathogènes pour des actions terroristes implique le développement de nouvelles stratégies thérapeutiques. Plus que jamais, la compréhension des mécanismes de résistance aux antibiotiques et aux antiviraux et la découverte de nouvelles cibles thérapeutiques doivent être poursuivies.

Face à ces problèmes de santé publique, la recherche internationale en microbiologie et en virologie connaît un essor nouveau, avec l'apparition de nouveaux champs d'investigation. La microbiologie cellulaire, en particulier, constitue une nouvelle approche mêlant étroitement microbiologie, biochimie, biologie structurale et biologie cellulaire, établissant des passerelles entre les mondes de la microbiologie et de la biologie cellulaire traditionnelles. Cette discipline cherche également à comprendre comment les micro-organismes communiquent entre eux, avec leur environnement ou encore avec leurs cellules hôtes. Ces processus de communication sont à la base de la pathogénie bactérienne. La découverte de nouveaux médicaments nécessite comme prérequis la compréhension moléculaire des dialogues s'instaurant entre pathogènes et cellules-hôtes.

## 2.1 VIRUS

Parmi les machines moléculaires à l'œuvre dans le vivant, les virus font l'objet d'une attention particulière, en raison des nombreuses pathologies dont ils sont responsables et de leur importance dans le domaine de la santé. De plus, les virus servent encore aujourd'hui de systèmes modèles pour l'élucidation du fonctionnement de nombreuses grandes machineries cellulaires, qu'ils détournent à leur profit. Les mécanismes d'entrée virale dans les cellules-hôtes sont des processus complexes faisant notamment intervenir des interactions spécifiques avec des récepteurs cellulaires (protéines membranaires de la membrane plasmique, glycosaminoglycane, etc.), l'utilisation des voies d'endocytose cellulaire et la fusion membranaire, pour délivrer le matériel génétique viral dans le cytosol. La formation des complexes de réplication des génomes viraux font non seulement intervenir des protéines virales spécifiques (transcriptase reverse, ARN ou ADN polymérase, intégrase, hélicase, etc.) mais aussi des protéines cellulaires spécifiques. Dans certains cas, les protéines virales impliquées dans la formation des com-

plexes de réplication sont des protéines membranaires associées à des réseaux qui dérivent du reticulum endoplasmique et qui sont induits par l'expression des protéines virales. La formation et la libération des particules virales exploitent les différentes voies du trafic cellulaire. Toutes ces étapes impliquent la formation d'assemblages supramoléculaires, formés de protéines et d'acides nucléiques. Une connaissance approfondie des structures atomiques de ces assemblages est non seulement essentielle pour comprendre les mécanismes moléculaires de la réplication virale, mais aussi pour l'identification de nouvelles cibles thérapeutiques, la conception d'antiviraux et la conception de vaccins.

### **Virus pathogènes animaux et humains**

Nous assistons à un regain d'intérêt envers ces virus. La première raison en est l'explosion du VIH dans les années 80. Une autre cause est sans doute la succession des épidémies dues à des virus (ré)émergents : pas moins d'un événement épidémique significatif par an (virus de West-Nile, du SRAS, de la grippe aviaire, du Chikungunya, etc.). Bien malin qui peut aujourd'hui prédire les pandémies à venir.

Dans ce contexte, la biologie et la génomique structurales engrangent les structures et contribuent à fournir, *ab initio*, des médicaments anti-viraux. Pour cela, elles étendent leur champ d'action non plus uniquement aux protéines structurales, uniques « stars » des années passées, mais aussi aux enzymes et autres protéines non-structurales, et s'attachent à démêler les modes d'assemblage et le fonctionnement de ces machines.

### **Virus infectant d'autres organismes, amibes, procaryotes et archaea**

Ce que nous connaissons du monde des virus n'est que la partie émergée d'un iceberg. Les virus de bactéries (ou bactériophages) sont toujours abondamment étudiés. De nouveaux

venus dans ce monde sont les virus infectant les archaea ou les amibes. Ces virus, dont la taille du matériel génétique peut être supérieure à celle d'une bactérie, posent la question de leur origine et de la relation virus/organismes dans l'apparition du monde à ADN.

## **2.2 BACTÉRIES ET LEVURES**

Ces dernières années, nous assistons à une augmentation très rapide de l'information structurale concernant les protéines impliquées dans des processus pathogènes d'origine infectieuse. Les études structurales dédiées aux processus d'adhésion, de reconnaissance hôte-pathogène, ou encore dédiées aux éléments structuraux constitutifs des systèmes de sécrétion bactériens, ont souligné l'explosion, au niveau international, de la microbiologie structurale et cellulaire et ont permis d'importantes avancées dans notre compréhension de la pathogénie bactérienne.

On voit bien l'importance des protéines membranaires et des complexes multiprotéiques membranaires pour permettre à la bactérie de s'adapter et d'interagir avec l'environnement ou l'hôte. La caractérisation de ces processus complexes, à l'interface de la microbiologie et de la biologie cellulaire, nécessite l'étude structurale de complexes macromoléculaires (incluant des complexes membranaires), de leur dynamique d'assemblage et d'interaction (tant à l'interface des membranes que dans les cellules hôtes) ou encore le repliement sélectif, suivant leur localisation, des protéines sécrétées et/ou injectées. Il s'agit donc d'un secteur qui nécessite tout particulièrement l'intégration de toutes les approches de la biologie structurale autour d'une thématique fonctionnelle. La biochimie, qui prépare et purifie les complexes, est bien sûr un élément clé de l'analyse structurale.

Les microorganismes s'adaptent très activement à leurs hôtes, mais également à leur environnement. Cette adaptation se déroule via des mécanismes moléculaires complexes. Les deux plus importants mécanismes sont

d'une part la régulation de la transcription bactérienne par les facteurs anti-sigma, et d'autre part, la régulation par les systèmes à deux composants. La coordination de l'expression des facteurs de virulence avec la régulation des fonctions vitales de la bactérie, qui se produit lorsque la bactérie s'associe avec une cellule hôte, est essentielle et les systèmes de signalisation à deux composants sont communément associés à ces processus. Ils constituent des cibles intéressantes pour la conception de nouveaux antibiotiques. Enfin, on doit citer les mécanismes dits de « Quorum Sensing » qui permettent aux bactéries de percevoir la densité de leur population et de déclencher l'expression massive de nombreux facteurs de virulence. Ce système de régulation fait intervenir la diffusion et la perception de petites molécules. Ces molécules permettent d'engager un dialogue cellulaire entre bactéries d'une même espèce ou d'espèces différentes. L'identification de ces molécules et la conception d'antagonistes font également l'objet de développements thérapeutiques intéressants.

Si nous avons illustré ci-dessus l'apport de nouvelles approches moléculaires à la compréhension des interactions entre les microorganismes et leur hôte, des études similaires visant à comprendre les mécanismes d'adaptation des microorganismes à d'autres environnements naturels sont toutes aussi importantes. Les bactéries ont colonisé toutes les niches écologiques, elles s'adaptent sans cesse et régénèrent inlassablement un environnement, qui, dans le cas des extrémophiles, peut être particulièrement inhospitalier. La richesse du métabolisme bactérien sera certainement encore la source de nouveaux concepts et de nouvelles applications pour les bioconversions et l'environnement. On peut ici citer quelques-uns des objectifs dans ce domaine. Il peut s'agir de :

- i) la mise en évidence de la biodiversité métabolique aérobie ou anaérobie, et de l'adaptation et de la résistance des microorganismes aux conditions extrêmes ;

- ii) la mise en évidence d'édifices supra-moléculaires impliqués dans les voies bioénergétiques chez des organismes extrémophiles ;

- iii) la recherche de nouveaux biocatalyseurs issus d'organismes résistants à des environnements extrêmes pour des applications biotechnologiques (Bioénergies).

Les mécanismes d'adaptation et de résistance au stress (oxydant, métaux lourds) sont susceptibles d'applications industrielles (mécanismes originaux de protection contre le stress oxydant, bioremédiation des métaux lourds toxiques, biolixiviation, biohydrogène).

## 2.3 BIODIVERSITÉ

### Métagénome, métaprotéome ?

Les microorganismes jouent un rôle primordial dans l'établissement des conditions environnementales qui existent aujourd'hui sur Terre. Ils sont les principaux acteurs des cycles géochimiques et de dégradation des déchets de la planète par l'intermédiaire de leurs activités enzymatiques. L'analyse en milieu défini des protéines microbiennes correspondant à différents écosystèmes, la métaprotéomique, constitue un champ d'étude émergent important dans l'étude du fonctionnement d'un écosystème microbien. Jusqu'à présent, les techniques post-génomiques ont été limitées principalement à l'étude de cultures pures et ne reflètent donc pas l'expression des gènes d'un mélange microbien complexe qu'on peut trouver dans un écosystème particulier. Grâce aux données de séquences métagénomiques et au nombre croissant de génomes individuels séquencés, les techniques post-génomiques, comme l'analyse protéomique, peuvent désormais être appliquées à des communautés complexes. Le métaprotéome alors obtenu reflètera l'activité microbienne du consortium. L'aspect intégratif de la métaprotéomique permettra une meilleure compréhension des fonctions microbiennes et constitue un outil pour identifier des gènes fonctionnels de première importance.

Toutes les études en « ome » feront appel au développement d'outils de modélisation

pour intégrer la masse importante d'informations quantitatives obtenues par ces approches globales avec des données phénotypiques et physiologiques. On parle alors de biologie systémique.

### **Nouvelles approches structurales et fonctionnelles des voies métaboliques représentatives de la biodiversité des microorganismes**

À l'heure actuelle, un déséquilibre important existe entre le flot d'informations qui est obtenu par l'annotation automatique et/ou experte de génomes complets et les informations biochimiques, fonctionnelles et structurales visant à alimenter ces attributions par des preuves expérimentales. De plus, ces informations expérimentales comportent un biais vers les organismes modèles qui sont particulièrement bien étudiés, ce qui résulte souvent dans des attributions fausses ou une absence totale d'attribution de gènes, voire groupes de gènes de fonction non encore rencontrée. Si on considère que, sur un génome, moins de la moitié des gènes annotés ont une fonction connue, cela indique qu'à ce jour nous ne connaissons que la fonction des gènes qui sont nécessaires et suffisants aux bactéries pour survivre dans les conditions de laboratoire. Il est évident que de nombreuses autres fonctions seront uniquement révélées dans une niche écologique plus complexe ou au sein d'un hôte. L'analyse bioinformatique et comparative des génomes complets de microorganismes adaptés à des milieux ou à des niches extrêmes et peu explorées (milieux marins, extrémophiles, etc.) nous montre qu'il existe des voies métaboliques encore peu connues, qui représentent parfois jusqu'à 60 % du génome. Certains laboratoires s'investissent dans des méthodes post-génomiques, telles que la transcriptomique, la protéomique, ou encore des méthodes génétiques. Mais aucune de ces techniques « post-génomiques » ne peut apporter l'information sur la fonction biochimique précise des gènes en question. D'autres approches sont donc nécessaires pour préciser le rôle de gènes appartenant à

des voies métaboliques peu ou pas connues. À cet égard, une stratégie valide consiste à se baser sur les acquis techniques des projets de génomique structurale « haut débit », et de les implanter au sein de laboratoires à « moyen débit » (i.e. clonage et surexpression recombinante de nombreuses cibles en parallèle, etc.), en les combinant à des études fonctionnelles et biochimiques poussées, en cherchant dans ce cadre les gènes prédits par analyse bioinformatique comme appartenant à des voies métaboliques nouvelles ou peu explorées (gènes colinéaires avec des gènes de fonction connue, etc.). Enfin, le criblage à haut débit pour l'étude des interactions hôte-pathogène, comme les modèles *Caenorhabditis elegans* ou amibe, permettra d'accéder rapidement à des déterminants de la pathogénie bactérienne (on peut parler de « pathogénomique »).

Il est finalement important de réaliser, que les séquences de génomes bactériens s'accumulent dans les bases de données publiques à un rythme très élevé. Pour une utilisation efficace de ces données, il faut souligner que la génomique comparative est à la base de l'annotation des génomes et de la prédiction fonctionnelle des gènes (similarité, coévolution, etc.). Les questions sont simples : par exemple, pourquoi deux bactéries phylogénétiquement très proches présentent-elles un caractère pathogène différent, à quoi peut-on attribuer cette différence ? La génomique comparative contribue donc directement à l'identification de gènes cibles, qui, ici, auront un intérêt biomédical, mais, selon la question, prendront un éclairage fondamental ou biotechnologique.

### **Origines de la vie, biosphère**

En conclusion, il est bon de rappeler que quatre cinquièmes de l'histoire de la vie sur notre planète étaient exclusivement microbiens. Les écosystèmes microbiens ont, pendant environ 3 milliards d'années, transformé l'environnement d'une terre hostile à la vie multicellulaire en celui d'une planète propice à l'existence d'organismes variés. La connais-

sance de ce passé microbien représente un but majeur dans la quête de l'humanité pour connaître ses origines les plus anciennes.

La mise en évidence de la biodiversité métabolique, de l'adaptation et de la résistance des microorganismes aux conditions extrêmes ainsi que le décryptage des métabolismes fondamentaux sont donc des enjeux importants pour le département des Sciences du Vivant. Ils peuvent également être la source d'applications biotechnologiques essentielles, comme la recherche de nouvelles sources d'énergie (bio-conversions, biomasse, production de biogaz et d'hydrogène, etc.) ou celle de meilleures conditions du développement durable (bioremédiation, interactions bactéries/ environnement, etc.).

## **3 – ANALYSE STRUCTURALE, ÉTAT DE L'ART, DÉFIS ET BLOCAGES**

### **3.1 LES APPROCHES EN BIOLOGIE STRUCTURALE**

La biologie structurale, qui repose principalement sur trois grandes méthodologies expérimentales : cristallographie, RMN et microscopie électronique, est en profonde mutation. On peut noter l'importance croissante de la bio-informatique structurale et de la dynamique moléculaire, l'explosion de nouvelles approches comme la spectrométrie de masse et la RMN du solide, le développement parallèle des études en solution (fluorescence, SAXS, etc.) ou en cellules (imageries diverses), le développement de nouveaux outils pour l'automatisation, la miniaturisation et le haut débit (robots, puces et autres nanotechnologies), la mise en place de plates-formes, la banalisation des TGE-TGI, etc. Beaucoup d'avancées dans ces domaines sont le fait de nos collègues relevant de la section 16. On en

trouvera une description plus complète dans le document de cette section.

### **Cristallographie**

Même si pour des protéines solubles et assez compactes, la cristallographie est devenue presque routinière, le champ des investigations structurales est encore immense. Les moyens actuels permettent d'aborder des protéines, acides nucléiques ou complexes de molécules, plus difficiles à produire et manipuler (cas des complexes transitoires), pour lesquels les cristaux sont difficiles à obtenir ou restent très petits, et qui nécessitent encore des développements. La robotisation de la production des échantillons (production de protéines pures et obtention de cristaux) initiée par la génomique structurale permet à la fois la parallélisation des expériences (exploration de différentes constructions pour une même protéine par exemple) et la miniaturisation (nanométrique pour les essais de cristallisation) et augmente donc le taux de réussite. La qualité incomparable des données obtenues par l'utilisation du rayonnement synchrotron et la possibilité de déterminer les phases associées aux rayons diffractés (utilisation du signal anomal en choisissant la longueur d'onde) font du synchrotron un instrument de routine. La stratégie actuelle est de préparer l'expérience sur une source de RX conventionnelle (test de congélation des cristaux, caractérisation des cristaux et recherche de dérivés d'atomes lourds si la taille le permet) puis d'enregistrer les données quantitatives sur une source de rayonnement synchrotron. Les accès simplifiés aux sources de rayonnement synchrotron par la mise en place de lignes dédiées à la cristallographie des protéines, les développements faits sur ces lignes (automatisation du changement de cristal, de la stratégie d'enregistrement et du traitement des données) rendent ces expériences accessibles à des « structuralistes » qui ne sont plus nécessairement des cristallographes de formation.

L'étape limitante se situe non plus tant dans l'approche structurale proprement dite que dans l'expression des macromolécules et



complexes macromoléculaires. L'expression hétérologue de ces molécules peut se révéler délicate. Elle reste cependant indispensable pour une reconstitution *in vitro* des structures macromoléculaires. Une autre étape limitante en diffraction X reste la cristallisation.

Les protéines de membrane dont la structure a pu être résolue jusqu'ici se caractérisent par leur abondance tissulaire et leur stabilité exceptionnelle (exemple de la rhodopsine bovine). Pour accéder à de nouveaux modèles expérimentaux il est nécessaire de développer :

- i) des approches haut-débit permettant de cribler un grand nombre de systèmes d'expression (différents vecteurs de clonage et d'expression) ;

- ii) des méthodes adéquates pour la production (différents microorganismes ou souches d'expression hétérologues et conditions de croissance), la purification, la re-solubilisation en cas de produit en corps d'inclusion, et la cristallisation (méthodes classiques ou cristallisation en phase cubique, présence d'agonistes ou antagonistes) d'un grand nombre de protéines membranaires.

Dans le criblage des conditions de cristallisation, l'utilisation de robot utilisant de très petits volumes (gouttes de 50 à 200 nL) apparaît essentielle pour étendre le nombre d'essais compatibles avec des taux d'expression ou de re-solubilisation faibles.

## RMN

La RMN multidimensionnelle permet d'aborder la structure, la dynamique et les interactions de protéines et acides nucléiques. Elle est particulièrement adaptée à l'étude de macromolécules flexibles et souvent difficiles à cristalliser et celle des ARN de petite taille. La présence, sur le territoire national, de nombreux spectromètres à haut champ et de 3 spectromètres à 800 MHz, l'amélioration de la sensibilité par les sondes cryogéniques et des techniques multidimensionnelles, ainsi que le recours au marquage isotopique  $^{15}\text{N}$ ,  $^{13}\text{C}$

et  $^2\text{H}$ , permettent maintenant de résoudre assez facilement la structure de protéines allant jusqu'à des masses moléculaires de 30 kD (par exemple, études de protéines ou d'ARN de très grande taille par des expériences TROSY). C'est en combinant des méthodes de marquage, l'utilisation de milieux orientants, les derniers développements méthodologiques et l'utilisation de spectromètres à champ de plus en plus élevé, qu'on peut maintenant espérer travailler sur des systèmes dont la masse moléculaire avoisine les 100 kD. La RMN du solide, en pleine expansion et moins sensible à la taille des objets étudiés, ouvre aussi de nouvelles perspectives, surtout pour l'étude de protéines de grande taille, celle de protéines membranaires ou fibrillaires et celle de complexes de protéines. C'est la RMN du solide qui a déjà permis de caractériser le comportement de certaines protéines fibrillaires importantes dans le cadre de l'étude de maladies dégénératives telles que l'encéphalopathie spongiforme ou la maladie d'Alzheimer.

En plus de la structure, la RMN renseigne sur les propriétés dynamiques des molécules. Ces informations sont importantes pour comprendre la fonction. Avec le développement de champs magnétiques plus élevés, de sondes cryogéniques, et de méthodes d'alignement partiel, avec également l'étude de la dynamique des protéines à l'état solide, la RMN jouera un rôle central pour étudier la structure et la dynamique des protéines à l'échelle atomique.

## Microscopie et cryo-microscopie

La microscopie électronique (ME), longtemps réservée à la visualisation d'architectures assez grandes, comme les fibres musculaires par exemple, révèle aujourd'hui de nouvelles potentialités dans l'obtention d'images directes d'assemblages moléculaires à moyenne résolution. Les résolutions obtenues par la cryo-microscopie électronique sont de l'ordre de 10 Å. Elles peuvent parfois descendre plus bas et révéler les structures secondaires, en particulier dans les systèmes membranaires.

Combinées à des traitements d'images puissants, la ME permet de remonter à l'architecture tri-dimensionnelle d'assemblages dont la masse est supérieure à quelques centaines de milliers de Dalton. Les assemblages symétriques s'y prêtent particulièrement bien, et les résultats les plus percutants sont obtenus pour des virus. Ces dernières années, plusieurs exemples combinant l'enveloppe de virus obtenue par microscopie électronique avec la structure à haute résolution des protéines individuelles constituant le virus, ont illustré la possibilité d'obtenir des modèles «pseudo-atomiques» des assemblages. Des études similaires ont été faites sur des filaments d'actine décorés de têtes de myosine, montrant ainsi, à l'échelle presque atomique, l'assemblage des fibres musculaires et les acides aminés impliqués dans le mécanisme fonctionnel de la contraction musculaire. La microscopie électronique peut se pratiquer sur des assemblages, des molécules isolées ou des cristaux bi-dimensionnels. L'approche sur des cristaux 2D peut avoir quelques applications dans l'étude des protéines membranaires, en complément des approches 3D. Elles ne représente néanmoins pas vraiment une alternative au 3D. Les progrès dans le traitement des images devrait ouvrir d'autres possibilités dans les prochaines années (assemblages de symétrie plus basse, obtention de modèle à partir d'un nombre d'images réduit). Les développements récents et les premiers résultats de tomographie sur cellules entières montrent l'avenir de ces approches.

De même, la microscopie à force atomique semble être un outil d'avenir adapté à l'étude de protéines membranaires et peut-être à l'observation de membranes natives.

Durant les dix dernières années, l'AFM est apparue comme un outil très performant en biologie structurale. Elle permet l'observation de la surface d'échantillons biologiques dans un tampon physiologique avec des résolutions latérale et verticale qui peuvent atteindre quelques angströms.

L'AFM permet d'observer aussi bien des structures biologiques complexes que des molécules uniques. Elle a permis des avancées dans l'étude des membranes biologiques,

natives ou artificielles, et de leurs constituants lipidiques et protéiques. C'est un outil complémentaire de la microscopie électronique, de la RMN et la cristallographie RX. L'AFM permet également d'imager, dans leur état fonctionnel, d'autres structures biologiques comme des complexes nucléoprotéiques. Elle peut servir d'instrument de nano-dissection et de manipulation à l'échelle moléculaire. Son utilisation en mode «spectroscopie de force» s'est beaucoup développée ces dernières années et est utile pour évaluer les forces d'interactions intra- ou intermoléculaires.

Les approches par RMN et microscopie restent cependant encore difficiles à automatiser (au niveau des expériences et du traitement des données) et, par conséquent, sont encore moins accessibles à des utilisateurs peu spécialisés.

## Spectrométrie de masse

Les demandes de la biologie ont été un moteur puissant dans les développements de la spectrométrie de masse, aussi bien du point de vue de l'instrumentation (méthodes d'ionisation et de fragmentation, analyseurs hybrides) que des méthodologies associées, en particulier au niveau des techniques séparatives. Associée à l'électrophorèse bidimensionnelle ou à d'autres techniques de séparation, elle permet de comparer les profils d'expression protéiques normaux vs pathologiques, mutants vs. sauvages, stressés vs. non stressés, etc.

En dépit d'avancées instrumentales récentes (analyseur à haute résolution, analyseur de type trappe linéaire basée sur la technique du triple quadripôle, etc.) des développements sont encore nécessaires afin d'élargir les champs d'application en protéomique ainsi que dans les diverses disciplines «omiques».

Il paraît indispensable de développer : l'identification des protéines dans des mélanges complexes, l'identification des modifications post-traductionnelles sur des protéines isolées ou sur des mélanges complexes, le séquençage de novo, des stratégies pour la

spectrométrie de masse quantitative, ainsi que des techniques de préparation d'échantillon en amont. Le point de passage obligé reste encore aujourd'hui la formation d'une nouvelle génération de jeunes scientifiques rompus au travail d'interface.

### **Le cas sensible de la « bio-informatique structurale »**

C'est une discipline devenue incontournable, tant dans le développement d'algorithmes que dans la manipulation de la masse d'information que les projets de génomique/protéomique nous fournissent. Cette discipline n'a, à l'heure actuelle, qu'une visibilité très réduite en France.

La recherche en bioinformatique structurale doit impérativement être développée, notamment en ce qui concerne :

- l'amélioration des méthodes d'alignements multiples à bas taux d'identité (structures 2D et 3D des protéines) ;
- l'organisation en modules des protéines – lego protéique – Familles protéiques ;
- la prédiction et l'amélioration de la solubilité, l'accessibilité au solvant, la cristallogénèse ;
- la définition de domaines structuraux pour la détermination de structures 3D (« domain-design ») ;
- la prédiction de repliements (« threading ») et de topologies ;
- la simulation et la dynamique des interactions moléculaires (ADN, ARN, protéines, membranes) ;
- la production d'algorithmes d'amarrage moléculaires (« docking ») ;
- la modélisation moléculaire automatique à grande échelle ;
- les bases de données structurales, la classification et l'intégration d'informations structurales dans les bases de données biologiques.

Les simulations de dynamique moléculaire sont un outil de choix pour l'étude de la flexibilité des assemblages macromoléculaires. Il est également possible d'étudier de larges mouvements grâce à la dynamique moléculaire dirigée. Par ailleurs, l'une des voies de développement de la dynamique moléculaire est son incorporation aux processus d'arrimage moléculaire (docking) afin de mieux prendre en compte la flexibilité du ligand. Enfin, la dynamique moléculaire permet de simuler le comportement des protéines membranaires dans le contexte de la membrane. Toutefois, ce type de simulations est très exigeant en temps de calcul et, ce, malgré le développement d'outils de calcul intensif au cours des dernières années. L'un des défis majeurs des méthodes de simulation est le développement de nouvelles méthodes pour la prise en compte du milieu environnant (eau, solvant, membrane, etc.) permettant ainsi de réduire les temps de calcul.

### **Approches en solution et dans la cellule**

Elles sont en plein développement.

#### **Diffusion des rayons X (SAXS) ou des neutrons (SANS) aux petits angles**

Cette technique est particulièrement puissante pour détecter et analyser les changements conformationnels en solution en fonction des conditions environnementales (température, force ionique, pression, etc.) ou en présence d'effecteurs, pour étudier le repliement et la stabilité, pour suivre l'évolution cinétique. Les développements en cours concernent en particulier l'obtention des enveloppes moléculaires des macromolécules. Cette approche peut être avantageusement combinée avec les déterminations de structure à plus haute résolution.

Un autre développement très prometteur est celui de l'étude des interactions en solution, à courte et moyenne distance, qui permet de

remonter aux forces en jeu. Cette approche s'est révélée particulièrement utile pour comprendre les bases physico-chimiques de la cristallisation des macromolécules et pour aborder la problématique des transitions de phase. Elle pourrait être utilisée davantage pour étudier les systèmes micellaires ou colloïdaux actuellement en développement pour des applications en vectorisation, microfluidique ou nanotechnologies.

**Fluorescence et techniques spectroscopiques (dichroïsme, FTIR) in vitro et in vivo : fonctionnement intracellulaire, etc. fluorescence résolue en temps, ondes évanescentes, etc.**

L'utilisation de plus en plus large des techniques de fluorescence (intensité, énergie, anisotropie, transfert, etc.) pour l'étude des interactions en biologie s'explique par leur facilité d'emploi, associée à une grande sensibilité et versatilité. Les signaux de fluorescence reflètent, de façon particulièrement fine, l'environnement immédiat de la sonde qui change en fonction de la dénaturation, de la liaison de substrats, ou des interactions entre protéines et protéines-membranes. La sensibilité du signal de fluorescence permet de travailler avec des concentrations faibles (picomolaire), et donc avec peu de matériel, ce qui est appréciable pour les études à l'équilibre thermodynamique qui demandent une variation systématique des conditions expérimentales (titrations, association-dissociation, etc.). Il est aussi possible de travailler en « molécule unique ». De plus, comme les échelles de temps impliquées sont du domaine de la nanoseconde, on peut étudier tous les phénomènes cinétiques dans des échelles de temps au dessous de la microseconde (repliement, fixation de substrats, dissociations, etc.). La fluorescence est également adaptée aux applications à haut débit. Les sondes peuvent être « intrinsèques » quand il s'agit de suivre la fluorescence d'acides aminés ou d'acides nucléiques, ou « extrinsèques » et, là, la variété est infinie, depuis l'UV jusqu'à l'infra rouge. Enfin, avec le développe-

ment à la fois de sondes adéquates et de détections sophistiquées, c'est une approche qu'on peut de plus en plus utiliser *in cellulo*, voire *in vivo*. Ce sont quelques unes des raisons pour lesquelles la fluorescence est en train de prendre une place de choix pour l'étude des relations structure-fonction, depuis la biophysique jusqu'à la biologie cellulaire.

Parmi les champs émergents, on peut citer la biophotonique, qui est l'introduction en imagerie cellulaire photonique et électronique de la 4<sup>e</sup> dimension : le temps. On peut ainsi réaliser une microscopie en temps réel sur des cellules vivantes, en utilisant des molécules marquées par différents fluorochromes qui ne fluorescent que lorsqu'ils sont excités : FRAP (Fluorescence recovery after photobleaching), FLIP (Fluorescence loss in photobleaching), photoactivation. Ces techniques pratiquées sur cellules vivantes permettent d'accéder aux cinétiques et au trafic des molécules marquées ou à la translocation séquentielle de facteurs d'un compartiment à un autre.

L'accès à l'analyse d'interactions spatiales et temporelles entre protéines à l'intérieur d'une cellule unique peut se faire par FRET (fluorescence resonance energy transfer), FLIM (fluorescence lifetime imaging microscopy) et/ou FCS (fluorescence correlation microscopy). Ces techniques permettent de visualiser et quantifier l'expression de gènes, les trafic intracellulaires et les activités enzymatiques associées dans une cellule individuelle en temps réel et nécessitent un matériel innovant : microscope performant, caméras numériques (electron-multiplying CCDs) et surtout un système informatique d'analyse très rapide.

Les nouvelles approches, telles que la microscopie à force atomique ou la microscopie confocale combinée au FRET, permettront de comprendre l'organisation des protéines dans une membrane native et de suivre leurs changements de conformation. En principe, la spectroscopie infrarouge (ATR-FTIR) combinée au dichroïsme linéaire permet également de déterminer l'orientation d'hélices transmembranaires dans des bicouches de lipides et constitue un outil intéressant pour la caractérisation des interactions protéines-lipides et

l'analyse des réarrangements d'hélices lors d'un changement de conformation.

**Études cinétiques au niveau de la molécule unique** (appliquées à l'étude de la recombinaison homologue, l'étude de polymérasas, hélicases, etc.)

Grâce à l'exercice d'une force d'intensité choisie sur une bille paramagnétique ancrée à une molécule unique d'un acide nucléique, on peut enregistrer les changements mécaniques auxquels la molécule est soumise, notamment en fonction de l'action de moteurs moléculaires comme de celle d'enzymes. On mesure alors des constantes cinétiques grâce à un microscope couplé à une caméra et à un ordinateur comportant les logiciels adéquats à l'analyse. La principale limitation est la mise au point de plates-formes expérimentales dédiées.

### **3.2 PLATES-FORMES, TRÈS GRANDS ÉQUIPEMENTS (TGE), TRÈS GRANDES INFRASTRUCTURES (TGI)**

#### **Les différents types de plates-formes sur le territoire national**

Ces plates-formes, qui ont grand besoin d'être renforcées, concernent d'abord les grandes techniques de la biologie et de la génomique structurale qui viennent d'être évoquées. Mais elles portent aussi sur les techniques en développement : spectrométrie de masse, puces, micro-fluidique, nanotechnologies, etc. ou sur le soutien au développement et à l'interopérabilité de bases de données conviviales et accessibles à tous.

Dans le contexte des programmes à grande échelle comme ceux de génomique structurale qui entrent dans le champ de la section, il reste aussi à développer des outils d'annotation fonctionnelle à partir des struc-

tures, des bases de données de connaissance intégrée.

Les plates-formes de production de protéines ont été conçues en vue du « haut débit ». La plupart d'entre-elles sont pour l'instant adaptées uniquement à *E. coli*. Dans ce cas, les premières étapes de clonage et de tests d'expression peuvent être pratiqués à un coût de main-d'œuvre réduit. Une fois les conditions d'expression identifiées, on retombe dans un schéma classique d'expression/purification. Les coûts liés à l'étape exploratoire sont faibles, bien qu'ayant demandé des investissements significatifs ; ils sont protéine-indépendants. Les coûts de la seconde étape en main-d'œuvre sont importants, et sont protéine-dépendants. Le travail des plates-formes haut-débit est difficilement applicable aux protéines isolées car il faut attendre d'avoir accumulé un certain nombre de cibles (multiple de 12 par exemple). Alternativement, le nombre de tests sur une protéine peut être augmenté (construction, souche, température, induction, chaperon, etc.).

L'expression de protéines *in vitro* utilisant des extraits cellulaires (« cell free ») est encore trop peu développée en France. Cette approche est prometteuse pour l'expression de protéines membranaires, généralement toxiques pour les bactéries. Il faut noter le grand potentiel des extraits provenant de cellules eucaryotes (e.g., germes de blé) pour l'expression de protéines d'intérêt en thérapeutique humaine.

#### **TGE-TGI**

D'un point de vue technique, nous disposons d'outils très performants (ESRF, SOLEIL) pour la cristallographie (données haute résolution et/ou résolues dans le temps, phasage simplifié). De plus, les progrès considérables réalisés dans l'automatisation des différentes étapes (de l'enregistrement des données aux méthodes de remplacement moléculaire), font que de plus en plus de structures sont déterminées en un laps de temps très bref. L'étape limitante se situe non plus tant dans l'approche structurale proprement dite que

dans la préparation des macromolécules et complexes macromoléculaires.

L'apport du rayonnement synchrotron en génomique structurale ne se limite pas à la cristallographie. Par exemple, LURE et l'ESRF ont changé la physionomie des études en solution par SAXS. Les développements en cours auprès des centres de rayonnement synchrotron concernent également le dichroïsme circulaire, la fluorescence, l'infra-rouge. On peut ajouter à cette liste les montages à micro-faisceaux, qui permettent des études particulièrement poussées sur les fibres (muscle, collagène, etc.).

Les TGI en biologie structurale incluent également la RMN, avec 3 grands centres en France équipés en Très Haut Champ.

## 4 – INTERFACES

### 4.1 INTERFACE « HISTORIQUE » PHYSIQUE-CHIMIE

#### Avec la physique

L'interface avec la physique est en grande partie implicite et concerne l'appropriation des concepts « physiques » utiles aux progrès de la biologie.

Les collaborations les plus récentes portent sur l'utilisation des nanomachines et des nanotechnologies, des micromanipulations, de la microfluidique (ou microphysique) et sur toutes les approches « molécules uniques » et l'imagerie.

#### Avec la chimie

L'interface avec la chimie est en train d'évoluer en un nouveau champ disciplinaire, la « biologie chimique » (voir le document de la section 16). Nos collègues de la chimie s'ap-

proprient de plus en plus les molécules et macromolécules du vivant et entendent utiliser les concepts et les outils dont ils disposent pour analyser au niveau moléculaire et atomique, voire quantique, les processus chimiques intégrés dans les systèmes biologiques. L'un des défis est d'utiliser ces connaissances à des fins thérapeutiques. Parmi les préoccupations partagées par nos deux sections, on peut noter :

- la mise en œuvre des concepts « chimiques » de dynamique, réactivité, biocompatibilité, etc. ;

- le développement des méthodes : cristallographie, RMN, spectrométrie de masse, etc. ;

- l'amélioration de molécules à visées thérapeutiques et de leur vectorisation ;

- le domaine des biotechnologies et de la valorisation avec l'élaboration de nouvelles sondes, biomarqueurs, biomatériaux, dispositifs biomimétiques ou biodégradables, prototypes, etc.

### 4.2 INTERFACES EN ÉMERGENCE : AVEC L'INFORMATIQUE, AVEC LA BIOLOGIE CELLULAIRE

#### Avec l'informatique

L'utilisation de l'informatique en biologie est relativement récente (par rapport à la physique par exemple) et peut avoir un caractère utilitaire (moyens de calculs, graphique 3D, web services, SGBD). En termes de recherche, si l'interface entre la biologie et l'informatique a conduit au développement de la bio-informatique dont on peut décliner plusieurs branches :

- i) recherche méthodologique ;

- ii) bases de données, logiciels et webciels ;

- iii) application et bioanalyse informatisée, il faut souligner que, dans le futur, l'informatique pour la biologie devra développer de

nouvelles méthodologies pour résoudre des problématiques biologiques complexes.

Le passé montre comment certains algorithmes particulièrement utilisés en biologie, par exemple les réseaux de neurones et les algorithmes génétiques, ont été « bioinspirés ».

### Avec la biologie cellulaires

L'arrivée à maturité de grands projets de génomique structurale (Japon, US, etc.) engendre la détermination d'un nombre exponentiel de structures. L'une des conséquences est que l'impact d'un projet en génomique structurale provient de plus en plus de la plus-value fonctionnelle obtenue sur la base de la structure, par une connaissance approfondie des systèmes biologiques étudiés (comportement dans divers milieux, diversité des systèmes d'expression, des formes mutées, connaissance des partenaires, etc.). Par conséquent, les échanges entre biologistes cellulaires et structuralistes s'amplifient. Les laboratoires de biologie structurale s'attaquent à des systèmes de complexité croissante, les laboratoires de biologie cellulaire s'ouvrent aux approches structurales. Ce type d'interface en émergence, à fort potentiel, va dans le sens d'une intégration de plus en plus poussée des approches fonctionnelles et structurales. Cela devrait également permettre aux étudiants d'acquérir une meilleure formation interdisciplinaire.

## 4.3 INTERFACE SANTÉ – SÉCURITÉ SANITAIRE

La société civile est de plus en plus sensible aux risques liés à l'exposition aux agents pathogènes, chimiques ou biologiques. Dans ce contexte, plusieurs secteurs relèvent en partie des compétences de la section 21 : les mécanismes d'intoxication et de détoxification (pesticides, métaux), l'analyse des gènes de susceptibilité et de leurs produits, la définition

de marqueurs spécifiques, le lien de ces marqueurs avec des maladies (ré)émergentes, etc.

## 4.4 AVEC LA CLINIQUE

Parmi les domaines émergents, sont découverts des déterminants génétiques et/ou moléculaires d'un grand nombre de pathologies, à caractère infectieux ou non.

## 4.5 BIOTECHNOLOGIES – VALORISATION

En plus du domaine « partagé » avec nos collègues chimistes, on peut citer :

- l'exploitation de plus en plus large du « métagénome » environnemental comme source d'activités enzymatiques nouvelles, ou plus efficaces. Par exemple, les génomes des organismes, caractérisés ou non, issus de biotopes froids conduiront à des enzymes *a priori* actifs à basse température, avec un intérêt économique évident ;

- une autre approche prometteuse pour l'amélioration des enzymes est le « DNA shuffling », ou échange aléatoire *in vitro* de séquences issus de gènes orthologues d'organismes différents ;

- l'évolution dirigée.

## 4.6 BIOLOGIE SYSTÉMIQUE

Ce terme recouvre les approches qui tentent à modéliser les différents niveaux d'intégration, par exemple depuis la séquence jusqu'à l'expression du génome, etc. ainsi qu'à simuler les réseaux divers en interaction, etc.

Ce type de modélisation/simulation doit rester fondé sur des données fiables, si possible

quantitatives. L'approche réductionniste est donc complémentaire des approches à caractère systémique. La communauté de la 21 est déjà engagée dans la représentation de la complexité des assemblages moléculaires et des réseaux cellulaires et inter-cellulaires. À ce titre, la section 21 participe au mouvement « intégratif » souhaité par la direction du CNRS.

## 5 – DÉFIS

La section 21 a une longue tradition de recherche dans l'étude des bases moléculaires des fonctions du vivant. Ces études, plus que jamais au centre des défis futurs de la biologie, amènent, outre une compréhension des processus moléculaires fondamentaux à l'œuvre à tous les niveaux d'organisation du vivant, une meilleure perception des événements moléculaires à la base de pathologies diverses, infectieuses et non infectieuses. En particulier, ces études contribueront à l'invention de nouvelles thérapies. Quelques-uns des grands défis actuels, pertinents pour la section, sont explicités ci-dessous.

### 5.1 COMPRÉHENSION DES BASES MOLÉCULAIRES DE DIVERS DYSFONCTIONNEMENTS ET PHÉNOMÈNES PATHOLOGIQUES

Nous vivons dans une société où le décryptage des génomes (et les téléthons) met en avant toute une série de maladies qui proviennent directement de « défauts » moléculaires ou macromoléculaires, des pathologies qui dérivent d'une mutation ponctuelle, des maladies orphelines. Dans le même temps, le vieillissement de la population nous montre, avec la maladie d'Alzheimer, l'importance des défauts de repliement et/ou d'association, moléculaires encore.

Ce n'est pas tout. Nous vivons dans une société de plus en plus préoccupée par les risques qui l'entourent. Il peut s'agir de risques sanitaires liés à l'environnement, provenant d'agents chimiques (pesticides, amiante, etc.), bactériologiques (légiionellose, bactéries opportunistes résistantes aux antibiotiques, etc.), viraux (HIV, HCV, HBV, etc. cas récents de la grippe aviaire ou du Chikungunya). Il peut s'agir de risques alimentaires (vache folle). Suivant les taux et la durée de l'exposition, le résultat peut être une infection bénigne, ou le développement d'un cancer, parfois bien longtemps après le début de l'exposition. Dans la plupart des cas, la compréhension des phénomènes au niveau moléculaire est loin d'être acquise ; les déterminants moléculaires des mécanismes de résistance (aux infections aussi bien qu'aux traitements), les phénomènes moléculaires à l'origine de l'apoptose, etc. sont encore à élucider.

Dans ce contexte, l'un des enjeux majeurs pour la section 21, à la fois un enjeu fondamental et un enjeu de société, est la compréhension à l'échelle moléculaire, voire atomique, des phénomènes pathologiques avec en perspective des solutions thérapeutiques originales. La section 21 soutient depuis des années les études des relations structure-fonction, des bases moléculaires et structurales des fonctions du vivant. L'enjeu est aujourd'hui d'en comprendre aussi les dysfonctionnements et de proposer des stratégies correctrices.

### 5.2 LA NOUVELLE BIOLOGIE DE L'ARN

Seuls 2% du génome humain codent pour des protéines ; dans les 98% restants, de nouveaux gènes d'ARN régulateurs sont identifiés chaque jour. Les petits ARN non codants (ARNnc) sont susceptibles d'intervenir dans la mise en silence transcriptionnelle des gènes par la modification des histones et/ou de l'ADN, comme cela a été montré chez *Schizosaccharomyces pombe* ou la Drosophile. C'est



aussi le cas pour de très grands ARNnc chez les vertébrés : un exemple très important est l'inhibition de l'activité d'un des deux chromosomes X chez les organismes femelles. Chez les procaryotes, les petits ARN régulateurs interviennent dans la régulation de l'expression de gènes de virulence et constituent de futures cibles thérapeutiques de la lutte contre les infections bactériennes. Chez les eucaryotes, ils jouent aussi un rôle dans la régulation post-traductionnelle : toute une série d'ARNnc (miARN, siARN) interviennent dans la dégradation des ARNm ou la mise en silence de la traduction. Ces petits ARNnc peuvent aussi faire partie intégrante de complexes catalytiques, qui modifient des ARN, ou jouer de multiples rôles dans la plupart des processus biologiques, y compris la synthèse des télomères. Ils peuvent avoir eux-mêmes une activité catalytique. Cette liste n'est sans doute pas exhaustive : l'importance biologique des ARNnc est loin d'être complètement cernée, mais elle se révélera sans doute considérable. Aussi le développement d'approches nouvelles pour l'identification des ARNnc et surtout de leurs cibles biologiques est-il un important défi pour l'avenir, y compris pour la santé humaine. Par ailleurs, il est maintenant clair que les étapes post-transcriptionnelles de l'expression génétique permettent de diversifier énormément la relation gène-protéine chez les eucaryotes, notamment à travers l'épissage alternatif. Nombre de maladies génétiques semblent liées à des mutations altérant ces processus de diversification. Comprendre ces défauts permettra de définir des stratégies de correction, ce qui, là encore, constitue un défi majeur. Ainsi qu'on l'a dit, la compréhension des mécanismes de synthèse, maturation et dégradation des ARN est l'une des préoccupations de la Section. Les ARN sont actuellement considérés comme de nouvelles cibles et de nouveaux outils thérapeutiques de haut intérêt. On peut en effet espérer bloquer de manière sélective l'expression des gènes par l'approche siRNA ou par les approches qui en dérivent.

### **5.3 STRUCTURE – FONCTION – RÉACTIVITÉ ET RÉGULATION DES GLYCOCONJUGUÉS**

La « glycomique fonctionnelle », représentée à plusieurs titres un véritable défi. Après avoir identifié le répertoire glycanique d'un objet (protéine, cellule, tissu, etc.) il s'agira en effet d'établir les bases moléculaires des relations structure-fonction des oligosaccharides à partir d'approches disciplinaires très variées, y compris les outils haut-débit (glyco-chips). La complexité structurale de ces molécules rend difficile l'obtention de matériel parfaitement défini à partir de sources naturelles. La synthèse chimique et/ou chimio-enzymatique représente un moyen, souvent nécessaire, pour décrypter le rôle informatif de ces molécules. La flexibilité conformationnelle représente aussi une difficulté majeure pour la mise en œuvre des approches structurales.

Au-delà de la description des structures oligosaccharidiques et de leurs propriétés, la recherche des partenaires cellulaires et des voies de régulations mises en jeu représente un enjeu important. On peut citer le rôle de différents oligosaccharides dans les fonctions de co-récepteur, dans l'activation des voies de signalisation, dans le trafic et l'adressage des molécules, etc. Les progrès en cours dans la compréhension des voies de biosynthèse des oligosaccharides permettent de mieux comprendre la façon et les conditions dans lesquelles la cellule peut modifier le niveau d'expression et la structure de ses glycannes, afin de mieux comprendre les conséquences de telles modifications. Une analyse transcriptomique devrait permettre d'aborder la question importante des mécanismes de régulation au niveau cellulaire de la synthèse, de la séquence et de la dégradation des oligosaccharides.

L'obtention d'oligosaccharides de structure caractérisée, la compréhension des mécanismes de reconnaissance aux niveaux cinétique et thermodynamique, la structure des complexes de ces molécules avec des protéines cibles, l'analyse des voies de biosyn-

thèse, représentent un axe fort de développement, permettant une meilleure compréhension de mécanismes fondamentaux, et le développement de molécules d'intérêt dans de nombreux domaines.

Ces études ouvrent déjà de nouvelles perspectives dans le diagnostic ou la compréhension des mécanismes physio-pathologiques de plusieurs maladies, ou d'envisager l'utilisation d'oligosaccharides comme outils d'intervention dans différentes situations : maladies infectieuses ou liées à l'inflammation, désordres cardiovasculaires, etc. Rappelons que le premier médicament de nature oligosaccharidique mis sur le marché, un pentasaccharide anti-thrombotique, est issu de la recherche académique, puis industrielle, française. De même la diversité des propriétés mécaniques, dynamiques ou physico-chimiques des polysaccharides, leur capacité à former des assemblages aux propriétés particulières (hydrogels, nano-objets par exemple), permettent d'envisager des dispositifs pour l'encapsulation, la vectorisation, la distribution ou la libération de molécules d'intérêt, ou la génération de structures biomimétiques.

## **5.4 LES PROTÉINES MEMBRANAIRES : CIBLES PRIVILÉGIÉES DE MÉDICAMENTS**

Les protéines membranaires contrôlent tous les échanges entre cellules ou entre compartiments cellulaires et sont impliquées dans de nombreux processus biologiques (signalisation cellulaire, jonction synaptique, photosynthèse, systèmes de transport d'électrons, bioénergétique, transport de solutés, reconnaissance et internalisation d'organismes pathogènes, résistance aux antibiotiques, etc.). Ces protéines sont donc des cibles pharmacologiques privilégiées (60% à 70% des médicaments actuels agiraient sur des protéines membranaires) et la connaissance de leur structure est une étape limitante pour la conception

de nouveaux médicaments. Cependant, ces protéines sont difficiles à produire, à manipuler et à cristalliser, car toutes ces opérations impliquent des mélanges protéines-détergents-lipides aux comportements physico-chimiques complexes.

Depuis la résolution de la structure d'un centre photo-réactionnel bactérien en 1985, une centaine de structures de protéines membranaires ont pu être abordées avec succès. Cette progression est loin du succès obtenu pour les protéines solubles. Sur environ 100 protéines membranaires présentes dans la Protein Data Bank, 10 structures ont été résolues en France. Il convient également d'ajouter les polypeptides membranaires dont les structures 3D ont été résolues essentiellement par RMN, notamment ceux impliqués dans les ancrages membranaires (en particulier les insertions membranaires post traductionnelles – « tail anchored proteins » et « in-plane membrane anchored proteins ») et dans les interactions intramembranaires. Les blocages principaux se situent à la fois au niveau de la surexpression de ces protéines (adressage difficile, volume accessible dans la membrane limité, toxicité pour l'organisme utilisé) et au niveau de leur stabilisation (purification, caractérisation fonctionnelle ou biophysique, et cristallisation). Les programmes de génomique structurale ont permis la mise en place de plates-formes automatisées à haut-débit qui, sans pouvoir se substituer au savoir-faire des biologistes, vont toutefois faciliter la production, purification et stabilisation de ces protéines. Pour tenir compte de la nature physico-chimique de ces protéines (surfaces hydrophobes et hydrophiles et interactions spécifiques avec des lipides), de nouveaux outils biophysiques et de nouvelles approches, basées par exemple sur une meilleure compréhension et exploitation du milieu amphiphile, devront être mis en place. Ainsi, la RMN du solide et les simulations par dynamique moléculaire *in silico*, permettront d'aborder l'étude des interactions protéines-lipides et celle des mécanismes de repliement pour une meilleure connaissance de la structure de ces protéines dans leur environnement naturel.

## 5.5 CONCLUSION

Si tous ces défis se situent au cœur des préoccupations de la section 21, la compréhension aux niveaux moléculaire et structural des fonctions du vivant reste un prérequis dans tous les secteurs de la biologie. Ce sont des

recherches de ce type qui fourniront le socle indispensable au développement, par exemple, de nouvelles thérapies contre les pathologies infectieuses (d'origine bactérienne, virale) et non infectieuses (cancers, etc.). On peut également en attendre l'émergence d'une vision systémique des fonctions du vivant.

