

# Bases moléculaires et structurales des fonctions du vivant

## *Président*

Philippe WALTER

## *Membres de la section*

Philippe BENAS  
 Thierry BLASCO  
 Sandrine BOSCHI-MULLER  
 Dominique BURNOUF  
 Patrice GOUET  
 Stéphane GRIMALDI  
 Daniel LEVY  
 Jean-Claude MICHALSKI  
 Michaël NILGES  
 Chantal PICHON  
 Véronique PILLER-MICHEL  
 Pierre PLATEAU  
 Véronique RECEVEUR-BRÉCHOT  
 Pascale ROMBY  
 Françoise SCHOENTGEN  
 Laurence SERRE  
 Marie-Christine SLOMIANNY  
 Patrick TAUC  
 Vincent VILLERET

## 1. PRÉAMBULE

La section 21 étudie les bases moléculaires et structurales des fonctions biologiques des organismes vivants par des approches pluridisciplinaires aux interfaces avec la biologie, la physique, la chimie, les mathématiques et la bioinformatique. Des modèles biologiques qui appartiennent aux trois règnes du vivant sont utilisés.

Les thématiques scientifiques de la section 21 représentent des axes majeurs de l'ère post-génomique à laquelle nous sommes arrivés. Elles impliquent la détermination des structures des macromolécules biologiques seules et en complexe avec leurs partenaires, et l'étude de la dynamique et l'énergétique de leurs interactions allant de la molécule unique, à l'organisation spatio-temporelle des biomolécules dans leur contexte cellulaire, à leur métabolisme, aux propriétés physiques des cellules, à leurs interactions et communications. L'objectif ultime est de définir les relations complexes entre les différents acteurs cellulaires (le génome, les ARN, les protéines et les métabolites), de répertorier les règles et codes utilisés pour le repliement des macromolécules, pour l'assemblage des complexes stables ou dynamique, et générés par les modifications post-transcriptionnelles et post-translationnelles. Il s'agit aussi d'obtenir une description fine de la dynamique structurale des génomes et des réseaux d'interactions entre biomolécules qui mènent à une expression coordonnée des gènes permettant aux organismes vivants de se reproduire, de se défendre, de s'adapter, et d'évoluer. Une autre préoccupation de la section 21 est de mieux comprendre au niveau moléculaire et atomique les dysfonctionnements cellulaires ayant une origine infectieuse ou génétique, et le fonctionnement de divers pathogènes pour identifier des cibles thérapeutiques et cribler de nouveaux inhibiteurs.

De part ses objectifs, ses approches expérimentales originales et infrastructures puissantes, la section 21 se situe à la croisée de la plupart des domaines de la biologie et de la biochimie (sections 22-25, 28-30) et de nombreuses disciplines scientifiques (chimie, physique, mathématique, bioinformatique) qui requiert un échange constant dépassant les frontières des instituts du CNRS voir même des autres EPST.

Les objectifs et les enjeux de la section 21 sont présentés ci-dessous.

## 2. MACROMOLECULES BIOLOGIQUES : DE LA GENOMIQUE A LA STRUCTURE

### 2.1 Le monde des ARN : structure, fonction et mécanistique

Au même titre que les protéines, l'ARN exerce un rôle clé dans la régulation de l'expression des gènes dans tous les organismes vivants. Avec le développement des méthodes de séquençage haut débit, ou de puces de type « tiling arrays », il est maintenant possible d'obtenir une cartographie complète des ARN exprimés à un moment donné de la croissance cellulaire. Cependant l'implémentation de ces technologies haut débit a pris un retard considérable en France. Chez les eucaryotes supérieurs, plus de 50% des gènes seraient contrôlés par des ARN non codant (ARNnc). Certains de grande taille induisent des effets épigénétiques. Le recensement de ces grands ARN reste cependant incomplet et leur mécanisme d'action mal compris. Par ailleurs, une myriade de miRNA intervient par le biais d'une machinerie complexe (RISC) pour contrôler de manière coordonnée l'expression des gènes au niveau post-transcriptionnel et une dérégulation de l'expression des microARN (miARN) est à l'origine de pathologies. D'autres petits ARNnc (piwiRNA) protègent le génome des cellules germinales contre la transposition. Dans les plantes et d'autres organismes, les siARN (ARN interférant) sont un moyen de défense contre l'infection virale alors que les virus ont co-évolué des mécanismes pour contrecarrer ces siARN. Dans les bactéries, pas moins de 5-10% du génome génèrent des ARN fonctionnels qui interviennent en réponse aux changements de l'environnement, contribuent à la pathogénicité et protègent la bactérie contre l'infection par des bactériophages. Par ailleurs, des ARNm exercent la fonction de régulateur comme ARN catalytique, modulent la stabilité d'autres ARN, ou modifient leur structure (thermosenseur ou riboswitch) pour affecter la transcription, la traduction du gène en aval.

L'impact des ARN régulateurs dans les relations hôte-pathogène et dans diverses pathologies humaines replace l'ARN au premier rang des préoccupations actuelles. Cependant, nous sommes encore loin d'avoir répertorié l'ensemble des riborégulateurs et trouver les cibles directes et les fonctions qu'ils régulent reste un défi actuel. Cette étape limitante est un pré-requis pour déterminer leur mode d'action et les réseaux de régulation dans lesquels ils interviennent, ce qui est essentiel pour la modélisation des différents niveaux d'intégration des signaux d'une cellule.

La compréhension des processus cellulaires et des mécanismes de régulation requiert la connaissance de la structure des ARN et de leur dynamique. A ce titre, la cristallographie des rayons X, la RMN, et la cryomicroscopie électronique (cryo-EM) ont permis des avancées spectaculaires avec la résolution de la structure du ribosome et de divers complexes ARN-protéines et ARN-métabolites. L'analyse et la prédiction de la structure des ARN restent un champ de recherche actif. Basé sur les structures cristallographiques de divers ARN, un ensemble

de motifs structuraux constitués d'appariements non canoniques ont été inventoriés, ce répertoire reste encore incomplet. Par ailleurs, les techniques de simulation de dynamique moléculaire apportent des informations sur l'hydratation des ARN et sur le rôle d'agents structurants (ions), ce qui contribue à une meilleure compréhension du repliement et de la stabilité de ces derniers. Ces questions liées à la dynamique peuvent être maintenant abordées grâce au développement de la RMN, de la cryo-EM, de la méthode SAXS, de la cartographie en temps résolu et des approches sur « molécules uniques ». Ces dernières méthodes (FRET, FCS...) ont été utilisées avec succès sur l'ADN. En revanche, il est encore difficile de marquer les ARN de grande taille par des fluorophores. D'autres défis technologiques concernent la visualisation des ARN dans une cellule vivante unique et en temps réel, et le développement des micromanipulations (étirements) d'ARN qui miment les perturbations mécaniques auxquelles la molécule est soumise. Ces approches permettront d'étudier l'impact de la vitesse de transcription sur le repliement de l'ARN en cours de synthèse, et d'identifier des structures transitoires qui peuvent moduler la réponse génétique. Il faut donc soutenir des plateformes d'imagerie performantes dédiées à ces analyses.

La plupart des ARN utilisent des cofacteurs comme les protéines pour exercer leur fonction et permettre leur adressage vers des séquences spécifiques. Les progrès récents dans les domaines de la génomique, la caractérisation structurale de complexes ARN-ligands et la bioinformatique ont permis des avancées significatives sur les mécanismes de reconnaissance ARN-ligands/cibles et sur l'évolution des fonctions de régulation par l'ARN à l'échelle moléculaire. Cependant, il reste difficile de comprendre comment une structure d'ARN peut acquérir de novo une fonction régulatrice et comment son implication dans un assemblage supramoléculaire, dans un réseau d'interactions ou dans une voie métabolique peut affecter son évolution. Les ARN restent des objets biologiques d'étude difficiles et les règles de reconnaissance ARN:protéines ou ARN:ARN parcellaires.

### 2.2 Repliement, stabilité et dynamique des protéines

#### *Les protéines : Séquence et repliement*

Malgré tous les progrès dans la compréhension des relations structure-fonction des protéines, il n'est pas encore possible de prédire la structure d'une protéine à partir d'une séquence donnée, et inversement. De telles prédictions reposent aujourd'hui sur des comparaisons de séquences primaires et de structures répertoriées au sein de bases de données. D'autre part, les connaissances récemment acquises sur les protéines intrinsèquement désordonnées et sur les amyloïdes montrent que les mécanismes impliqués sont extrêmement complexes. Ces protéines adoptent des structures différentes aussi bien selon les propriétés physico-chimiques de leur environnement (température, force ionique, pH...) ou selon le partenaire avec lequel elles s'associent, dans le cas d'interactions multiples. Déchiffrer le code du repliement tridimensionnel reste donc un défi scientifique

majeur pour les biologistes. Pour cela, l'identification et la caractérisation des intermédiaires et des étapes le long des différents chemins de repliement possibles sont nécessaires.

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

Les protéines membranaires représentent environ 25% du génome humain. Elles participent à de nombreux processus cellulaires comme le transfert des signaux extracellulaires, l'entrée/sortie de solutés, les interactions avec les pathogènes etc., et constituent de ce fait des cibles thérapeutiques majeures. Dans ce contexte, leur caractérisation structurale est essentielle et représente un enjeu majeur pour la compréhension des mécanismes cellulaires au niveau des membranes et pour le développement de futures thérapies. La production et l'analyse structurale de ces protéines restent toutefois des étapes limitantes malgré les avancées spectaculaires récentes que constitue la résolution des structures cristallographiques des premiers récepteurs couplés aux protéines G (RCPG). Parallèlement à la cristallographie X et la RMN liquide, les approches de biologie structurale que sont la RMN du solide, la cristallographie électronique et les simulations moléculaires apportent les informations sur la structure et la dynamique de ces protéines dans le contexte membranaire. Finalement les avancées récentes en microscopies de molécules uniques, de force atomique, de cryo-EM et des microscopies optiques à haute résolution permettent d'accéder à la dimension supramoléculaire des complexes protéiques membranaires ainsi qu'à leur dynamique en systèmes isolés et dans la membrane.

### **Les protéines intrinsèquement désordonnées**

L'étude des protéines intrinsèquement désordonnées est un autre défi majeur des prochaines années en biologie structurale. On connaît désormais leur prévalence dans le monde du vivant. On sait en outre qu'elles remplissent des fonctions cruciales dans la cellule, que ce soit par leur capacité à moduler les distances entre différents domaines fonctionnels ou macromolécules ou par leur mécanisme de reconnaissance moléculaire particulièrement original (repliement induit). Elles sont ainsi impliquées dans de nombreux processus essentiels du cycle cellulaire ou viral : réplication, régulation de la transcription, maturation, apoptose, transduction du signal, métabolisme, etc.... Déterminer les mécanismes dynamiques et structuraux par lesquels les protéines intrinsèquement désordonnées remplissent leur fonction est donc d'une importance fondamentale pour la compréhension du développement de certaines pathologies graves tels que les cancers, les maladies neuro-dégénératives, les maladies cardiovasculaires, le diabète, etc. De telles études nécessitent des approches adaptées à l'étude des protéines intrinsèquement désordonnées, comme la RMN, la diffusion aux petits angles, ou le dichroïsme circulaire en synchrotron. Enfin, un défi plus difficile à relever sera d'explorer cette famille de protéines en tant que cibles thérapeutiques potentielles, par des méthodes de conception rationnelle de médicament ou de criblage de molécules, en s'appuyant notamment sur les récents progrès dans l'inhibition des interactions protéine-protéine.

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

Les connaissances structurales sur les protéines amyloïdes ont grandement évolué depuis quelques années. Dans la forme fibrillaire, quelles que soient leur séquence ou la maladie dans laquelle elles sont impliquées, les protéines amyloïdes possèdent toutes des propriétés structurales très similaires, et se replient en formant des feuillets  $\beta$  croisés dont les brins sont perpendiculaires à l'axe de la fibre. En revanche, à l'état natif, elles peuvent être riches en hélices  $\alpha$ , riches en feuillets  $\beta$ , contenir des hélices  $\alpha$  et des feuillets  $\beta$ , ou même être intrinsèquement désordonnées. Comprendre les conditions environnementales et les chemins de repliement suivis conduisant à la formation d'amyloïdes est aussi d'une importance primordiale dans la compréhension des mécanismes d'association et de formation des dépôts amyloïdes et de leurs pathologies. A l'heure où la population des pays industrialisés se fait de plus en plus vieillissante, il devient crucial de trouver des molécules capables d'inhiber la formation d'amyloïdes, notamment dans le cas des pathologies neurodégénératives comme les maladies d'Alzheimer ou Parkinson.

### **Les enzymes**

L'enzymologie moderne permet aujourd'hui d'appréhender au niveau atomique les mécanismes impliqués dans l'acte catalytique, la régulation ou la fixation des différents substrats/effecteurs. Elle s'appuie pour cela sur une approche multidisciplinaire aux frontières de la physique, de la chimie et de la biologie. Ces approches de relation structure-fonction utilisent: i) les techniques de cinétique classique et rapide (en mode continu ou interrompu) couplées aux techniques spectroscopiques et de génie génétique, et ii) la connaissance de la structure tridimensionnelle des enzymes sous forme native, ou en complexe et d'intermédiaires réactionnels et plus récemment iii) de leur dynamique en solution, données obtenues notamment par des techniques de RMN permettant de décrire les différents états conformationnels qu'adoptent une enzyme au cours de l'acte catalytique. Les approches sur molécule unique, lorsqu'elles sont possibles, permettent quand à elles d'obtenir des informations quand à l'hétérogénéité «catalytique» des préparations enzymatiques et aux forces mises en jeu. De plus, les approches de mécanique quantique et moléculaire combinées permettent aujourd'hui de prédire les étapes chimiques de la catalyse enzymatique, apportant un support important aux études expérimentales. La conjonction de ces méthodes avec celles de la modélisation moléculaire permet à l'ingénierie enzymatique de modifier les enzymes pour les rendre par exemple plus stables ou modifier leurs propriétés. Ces connaissances permettent d'utiliser les enzymes comme cibles de médicaments, comme biocatalyseurs en biotechnologie ou comme modèles d'étude de processus fondamentaux complexes relevant de la biologie ou de la chimie. Des développements importants sont attendus afin notamment de relever les défis posés par l'ère post-génomique. En effet, les données de séquence de génomes ainsi que le nombre de structures tridimensionnelles établies pour des protéines est grandissant. En revanche, l'identification de leur mécanisme d'action, en particulier, dans le cas des protéines à activité enzymatique,

représente toujours un goulot d'étranglement dans la progression des connaissances fondamentales et appliquées. Le domaine de l'enzymologie subit aussi un désintérêt et un risque de perte de savoir-faire alors que la connaissance des génomes/métabolomes donne accès à de nouvelles enzymes potentiellement utilisables pour la biocatalyse et la bioconversion. La France risque de connaître un retard pénalisant dans ce domaine si aucune action n'est entreprise rapidement.

### **Modifications post-traductionnelles**

La mise en évidence de l'importance fonctionnelle des phosphorylations, méthylations, acétylations, ubiquitylations, aminoacylations, sumoylation et glycosylations des protéines est une avancée récente des plus spectaculaires dans le domaine de la biologie. Du fait de modifications post-traductionnelles, une même séquence codante peut en effet mener à beaucoup de protéines aux propriétés physico-chimiques et fonctionnelles différentes. Les progrès ont bénéficié des nouvelles possibilités offertes par la protéomique, ainsi que des nouveaux développements instrumentaux dans les domaines de la spectrométrie de masse et de la RMN. Ces « nouveaux codes » engendrés par les modifications post-traductionnelles et qui régulent nombre de fonctions cellulaires restent à déchiffrer.

## **3. GRANDS COMPLEXES MACROMOLECULAIRES ET MACHINERIES FONCTIONNELLES**

### **3.1 Transcription**

L'activité du complexe transcriptionnel est fortement modulée, d'une part, par la formation de structures régulatrices sur l'ARN naissant et, d'autre part, par l'interaction de l'enzyme avec de nombreux facteurs de régulation. Le mécanisme d'intégration des signaux issus de structures régulatrices par l'enzyme est encore peu connu, ainsi que la dynamique du complexe transcriptionnel et de l'action des facteurs associés sur son activité qui est la première étape vers l'analyse de la régulation normale ou pathologique de la cellule. Un niveau plus élevé de complexité est la synchronisation de la transcription avec les autres mécanismes (réplication, réparation), et les connections avec les mécanismes de transmission de signaux extérieurs, afin de maintenir l'homéostasie cellulaire.

### **3.2 Maturation des ARN**

Un ARN subit généralement une série de modifications post-transcriptionnelles : élimination des extrémités 5' ou 3', addition de nucléotides en 3', épissage d'introns, modifications chimiques de bases et/ou de riboses. Les enzymes requises pour ces processus sont multiples et extrêmement variées. Elles peuvent se trouver au sein de complexes protéiques permettant de coupler, par exemple, modification chimique et épissage chez les eucaryotes.

La structure, le mécanisme d'action des protéines et le rôle des nucléotides modifiés de l'ARN sont loin d'être complets.

### **3.3 Épissage**

L'épissage des pré-ARN messagers eucaryotes est catalysé par une machinerie nucléoprotéique, l'épissosome, qui se caractérise par une grande variabilité dans sa composition, dans sa structure et ses mécanismes d'activation. Pour arriver à une compréhension détaillée des contrôles qu'exerce une cellule sur l'épissage, il faut poursuivre la caractérisation du rôle des protéines de l'épissosome, ainsi que celui des modifications post-traductionnelles de ces protéines. Les recherches futures sur la structure de l'épissosome peuvent s'inspirer des succès obtenus sur le ribosome pour obtenir des préparations homogènes et disposer de produits chimiques bloquant spécifiquement certaines étapes de l'épissage. La description thermodynamique et cinétique de cette machinerie reste prioritaire.

### **3.4 Traduction**

Les connaissances sur la machinerie de traduction du message génétique ont fait des progrès spectaculaires ces dernières années, notamment avec l'élucidation de la structure à haute résolution du ribosome bactérien, archaéen et plus récemment de levure. Ces données structurales doivent maintenant être reliées aux aspects cinétiques et dynamiques du fonctionnement du ribosome, ainsi qu'aux mécanismes qui assurent la fidélité de la traduction. Ceci pourra être abordé par les techniques dites de « molécule unique » et les calculs par simulation. Les progrès récents sur la compréhension globale à l'échelle atomique du fonctionnement des différents acteurs de la traduction font de la machinerie traductionnelle un « système de référence » pour l'étude des machines cellulaires complexes. La comparaison des structures du ribosome des trois règnes du vivant laissent espérer l'identification de nouvelles cibles pour l'inhibition de la synthèse protéique, avec en perspective la mise au point de nouveaux antibiotiques ou de nouveaux anticancéreux.

### **3.5 Transport des ARN**

Chez les eucaryotes, les ARN doivent être transportés depuis le noyau où ils sont synthétisés vers le lieu de leur utilisation en transitant éventuellement par des sites où ils vont être maturés ou associés à des protéines. Ces différentes étapes peuvent être couplées. La localisation des ARNm dans certaines régions du cytoplasme est aussi fréquente, en particulier au cours du développement, et ceci est aussi le cas pour les bactéries. La caractérisation des machineries (locosome) qui permettent aux ARNm d'être assemblés dans des granules puis de voyager dans le cytoplasme est peu connue.

### **3.6 MACHINERIES DE DEGRADATION DES ARN**

La maturation nucléolytique des ARN et la dégradation des ARNm, des petits ARNnc ou des ARN défectueux sont effectuées par des ribonucléases, souvent regroupées au sein de particules complexes (dégradosome, exosome...).



L'importance de ces particules pour la cellule est soulignée par la multitude et la redondance des enzymes qui y participent. Ces dernières années ont été marquées par la compréhension des relations structure-fonction des exosomes et des ARN hydrolases. Mais de nombreuses questions restent en suspens : Quelles sont les spécificités et les substrats des RNases? Comment la dégradation des ARN est-elle régulée ? Quels sont les couplages entre maturation, repliement, traduction et dégradation des ARN ? Quels sont les signaux qui dégradent les ARN défectueux?

## 4 LE CONTROLE DE L'EXPRESSION DES GENES : STRUCTURE, MECANISTIQUE

### 4.1 Métabolisme, surveillance de génome et dynamique de la chromatine

Au cours de la dernière décennie, la découverte des ADN polymérases trans-lésionnelles a conduit à élaborer un nouveau modèle de réplication normale et mutagène, et à mieux comprendre les mécanismes moléculaires qui initient les processus tumoraux. La compréhension des mécanismes qui orchestrent l'échange permanent de partenaires pour assurer la réplication, la transition de la réplication vers la réparation, la régulation de l'accès de divers facteurs à la fourche de réplication, et la coordination des grandes fonctions dont l'ADN est le siège (réplication, réparation des lésions, recombinaison, transcription), est désormais en cours. Cette analyse devra s'intégrer dans une approche plus globale de l'étude de la dynamique de la chromatine. L'analyse de la dynamique des pools d'histones et de leur modification au cours de la réplication, de la transcription ou de la division cellulaire, ainsi que la découverte récente des histones variants, de leurs rôles différenciés dans l'ouverture, la stabilité et l'expression de la chromatine, souligne l'importance de ces processus dans la régulation du cycle cellulaire.

Si l'analyse de la structure, de la réactivité et de la dynamique de ces différents facteurs et complexes reste l'approche initiale nécessaire à la compréhension des processus moléculaires dans lesquels ils sont impliqués, la connexion des différents niveaux de régulation contribuera à une vision intégrée du fonctionnement cellulaire. En outre, la caractérisation de ces complexes et de leurs réseaux ouvre la voie à l'identification de nouvelles cibles pharmacologiques.

### 4.2 Contrôle de qualité des ARN

De multiples mécanismes de contrôle de qualité contribuent à façonner le transcriptome cellulaire et restreindre l'expression de l'information génétique inappropriée. Ils empêchent l'accumulation d'ARN non fonctionnels mais aussi contribuent à réprimer les ARN « parasites » issus de transposons, de régions répétées du génome ou de virus, et à réguler l'expression des gènes. Ces mécanismes de surveillance dépendent de protéines et complexes ribonucléoprotéiques qui ciblent les ARN

aberrants ou étrangers vers la dégradation nucléaire, ou cytoplasmique. La compréhension à l'échelle moléculaire de ces mécanismes de surveillance est fondamentale puisque la synthèse d'ARNm aberrants est à l'origine de nombreuses pathologies, ils peuvent aussi contribuer à l'évolution des génomes.

### 4.3 L'ARN et le contrôle post-transcriptionnel

La régulation post-transcriptionnelle (épissage des ARNm, traduction, stabilité) de l'expression des gènes contribue de manière significative à la complexité des génomes. Chez les eucaryotes, elle est requise au cours du développement, en réponse à divers stress, ou bien lors de la signalisation hormonale. Chez les bactéries, le contrôle de la traduction et de la dégradation des ARNm remodèle l'expression du protéome en réponse aux changements de l'environnement, à divers stress ou lors d'une infection. Dans la plupart de ces processus, la structure de l'ARNm a une fonction majeure pour favoriser ou empêcher la fixation des régulateurs. L'étude des règles de reconnaissance des ARNm cibles par ces régulateurs ont révélé récemment de nouveaux mécanismes d'initiation de traduction bactérienne.

### 4.4 L'ARN et la biologie synthétique

La biologie synthétique tend à créer des organismes nouveaux par une combinaison rationnelle d'éléments biologiques découplés de leur contexte naturel. Le but ultime est de concevoir et construire des systèmes biologiques fabriqués qui peuvent avoir une action bénéfique sur l'environnement et la santé humaine, et/ou générer des applications en biotechnologie. La plasticité de la structure de l'ARN permet d'envisager de telles applications encore peu étudiées en France. Les ARN thermosenseurs, ou bien encore les « riboswitch » ont été utilisés pour reprogrammer une bactérie. Les siRNA sont utilisés pour élucider la fonction des gènes chez les eucaryotes, et sont maintenant développés comme outils thérapeutiques.

### 4.5 Matrice extracellulaire et protéoglycannes

La matrice extracellulaire est constituée en grande partie de glycoprotéines (collagène, fibronectine, laminine) ainsi que de protéoglycannes. Ces constituants ont de nombreux domaines de liaison avec les cellules, facilitant leur adhésion et leur organisation en tissus et régissant la biodisponibilité de nombreux médiateurs solubles (facteurs de croissance, chemokines, cytokines), régulant ainsi leur activité biologique. Les protéoglycannes sont des macromolécules complexes composées d'une portion protéique portant des chaînes polysaccharidiques chargées négativement, les glycosaminoglycannes. Ils font l'objet d'un intérêt grandissant du fait qu'ils peuvent être considérés à la fois comme des cibles pharmacologiques potentielles et comme nouvelles molécules thérapeutiques. Leur étude est en plein essor

pour élucider les microdomaines oligosaccharidiques supports de la spécificité de reconnaissance par les autres acteurs protéiques. L'étude des mécanismes de leur biosynthèse constitue également un enjeu considérable.

## 4.6 Complexes multiprotéiques dans les voies de signalisation

Dans toutes les cellules vivantes, les mécanismes biologiques fondamentaux tels que la croissance, la division, la différenciation, l'apoptose, le métabolisme, les réponses à l'environnement et au stress sont régis par des communications entre l'extérieur et l'intérieur de la cellule. De nombreuses molécules servant de stimuli s'associent aux récepteurs membranaires pour transmettre un message à une ou plusieurs voies de signalisation qui amplifient le signal ainsi reçu pour aboutir in fine à l'expression des gènes adaptés à la réponse cellulaire. Les mécanismes mis en jeu se traduisent par exemple des cascades de phosphorylations qui activent séquentiellement les enzymes constitutives de la voie. Les cascades de signalisation, souvent interdépendantes, sont régulées et modulées de façon complexe par des petits ligands, des protéines et des ARN. Les défis actuels sont la caractérisation complète des acteurs impliqués dans ces voies de signalisation, 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, et de suivre la dynamique (dans l'espace et dans le temps) des complexes requise pour leur fonction. L'enjeu est d'importance car la plupart des grandes pathologies humaines (cancer, maladie d'Alzheimer) sont dues à des dérèglements des voies de signalisation cellulaire. D'ores et déjà, des molécules-médicaments (notamment en cancérologie) sont conçues pour réguler les voies de signalisation cellulaire. A terme, nous pouvons espérer concevoir de petites molécules capables d'orienter la réponse de la cellule dans le sens choisi par l'expérimentateur ou le clinicien.

## 5. L'INTERACTOME, RESEAUX DE REGULATION, COMMUNICATIONS INTRA-CELLULAIRES : VERS LA BIOLOGIE DES SYSTEMES

L'interactome est défini comme l'ensemble des interactions physiques entre molécules dans une cellule. Les informations ainsi collectées sont un préliminaire pour accéder à une vision systémique de la cellule. La compréhension d'une fonction biologique passe aussi par la modélisation des différents réseaux d'intégration, et la simulation des réseaux coordonnés de régulation et d'interactions entre biomolécules.

En pratique l'étude des interactions entre biomolécules est réalisée à partir d'éléments biologiques variables tels que des organelles (mitochondries, phagosomes), des cellules individuelles ou d'organismes entiers (bactéries, levure, drosophile), ce qui donne naissance à une masse importante d'informations sous la forme de banques d'interactions. Un autre type d'interactome très étudié concerne les interactions protéine-ADN (réseau formé

par les facteurs de transcription et leurs gènes cibles). Il sera aussi essentiel de prendre en compte la contribution quantitative des ARN régulateurs dans les phénomènes biologiques, ainsi que les interactions de type protéine-ARN (réseaux formés par des régulateurs post-transcriptionnels et leurs ARNm cibles) et ARN non codant-ARNm (réseaux formés par les ARN régulateurs et leurs cibles).

Dans tous les cas, les données actuelles ont été acquises essentiellement par des méthodes permettant un criblage à haut débit. Les interactions mises en évidence sont généralement binaires et statiques, et un bon nombre d'entre elles restent à valider in vivo. Dans l'avenir, il faudra établir la hiérarchie des interactions cataloguées. Celle-ci dépend de plusieurs facteurs : l'affinité entre les partenaires, leur localisation (espace), la séquence des événements qui conduit à leur interaction (temps) et l'état physiologique de la cellule (présence ou non d'autres acteurs moléculaires à un instant donné). Il faudra aussi combiner les résultats obtenus avec les données régulièrement obtenues par les autres domaines de la biologie intégrative (génome, transcriptome, protéome, métabolome), de tels développements n'iront pas sans la mise en œuvre d'outils informatiques et mathématiques spécialement dédiés et adaptés.

S'il reste du chemin à parcourir avant de décrire d'un point de vue qualitatif et quantitatif l'organisation d'une cellule et de fournir un modèle rendant compte globalement de son fonctionnement, il n'en demeure pas moins qu'on voit dès aujourd'hui apparaître d'intéressantes applications à l'interactome. Par exemple la fusion d'un interactome viral à l'interactome humain a permis de simuler une infection virale ; la comparaison des réseaux d'interactions dans six maladies neurodégénératives a mis en évidence des protéines-clés communes dans ces maladies. En étendant l'interactome à l'étude des interactions inter-cellulaires on peut encore imaginer d'autres applications en médecine (développement, renouvellement et régénération des tissus) et en microbiologie (réseau qui s'établit dans un biofilm ou dans un consortium de microorganismes dans un biotope reconstitué). Enfin, la synthèse de modulateurs de réponses conçus à partir de réseaux d'interactions devrait constituer des outils permettant de mieux comprendre, voire de maîtriser, la réponse cellulaire globale.

## 6. LA GLYCOBIOLOGIE

La glycosylation concerne au moins 80% des protéines et jusqu'à 30% des lipides d'une cellule. Les monosaccharides constituent les briques de base d'enchaînements plus complexes formant des oligo- ou des poly-saccharides organisés en différents types de combinaisons, parfois de grande taille, comme dans le cas des glycosaminoglycannes (>200 unités), et pouvant exister à l'état libre ou conjugué (glycannes). Leur grande diversité moléculaire est liée à la structure de ces molécules et à leur mode d'enchaînement : si 2 acides aminés (ou 2 nucléotides) peuvent donner 2 peptides (ou 2 oligonucléotides), 2 hexoses permettront de donner 20 disaccharides. Avec 3 unités on générera 1056 structures oligosaccharidiques. Cette extrême diversité est le support même du «glycocode» et fait des glycannes des molécules «informatives» par excellence. Chaque glycoconjugué existe en fait sous forme d'un mélange complexe de «glycoformes» souvent douées

d'activités biologiques différentes. Ainsi, à l'instar du code génétique et des protéines, il devient essentiel de déchiffrer le sens de ce glycode. Ces molécules n'étant pas directement codées par des gènes, les outils traditionnels de la biologie moléculaire n'ont eu jusqu'à présent qu'un impact faible sur cette discipline. La glycobiologie comprend aussi l'étude des enzymes impliquées dans la biosynthèse des monosaccharides et des nucléotides-sucres, la biosynthèse ou le catabolisme des glycanes (glycosyltransférases et glycosidases) ainsi que l'étude des protéines reconnaissant les molécules sucrées (récepteurs membranaires, lectines). Enfin, les composés glycosidiques sont impliqués aussi bien dans la structuration des cellules, que dans la communication intercellulaire ou encore certaines pathologies d'origine exogène ou endogène.

Les glycanes constituant des composants importants de la biodiversité, les connaissances accumulées jusqu'à maintenant sur la glycosylation des eucaryotes supérieurs et principalement des mammifères, doivent s'accompagner de l'étude de la glycosylation des bactéries, des parasites, des insectes et des plantes pour une meilleure compréhension des fonctions biologiques à travers l'évolution. La glycomique appliquée à l'étude des micro-organismes se révèle aussi comme une source extraordinairement riche d'enzymes utilisables pour la synthèse chimioenzymatique de glycoconjugués.

## 6.1 Structure, fonction, réactivité et régulation des glycoconjugués

La « glycomique fonctionnelle » et la « glycobiologie moléculaire » représentent à plusieurs titres un véritable défi. Après avoir identifié le répertoire glycanique d'un objet (protéine, cellule, tissu, etc.) il s'agira d'établir les bases moléculaires des relations structure-fonction des oligosaccharides. L'extrême diversité des glycoconjugués et leur complexité structurale rendent difficile l'obtention de matériel homogène à partir de sources naturelles. La synthèse chimique et/ou chimioenzymatique représente un moyen pour décrypter leur rôle informatif. La flexibilité conformationnelle représente aussi une difficulté majeure pour la mise en œuvre des approches classiques de biologie structurale.

Au-delà de la description des structures oligosaccharidiques et de leurs propriétés, la recherche de partenaires cellulaires et des voies de régulation mises en œuvre représente un enjeu important. Les progrès en cours dans la compréhension du métabolisme des oligosaccharides permettent de mieux appréhender la façon et les conditions dans lesquelles la cellule peut modifier le niveau d'expression et la structure de ses glycanes et les conséquences de telles modifications.

De nombreuses applications découleront de ces études fondamentales. Ainsi la conception d'analogues de monosaccharides amènera à les utiliser comme glycomarqueurs in cellulo et in vivo et comme inhibiteurs spécifiques de certaines voies de biosynthèse des glycanes. Ces études ouvrent sur de nouvelles perspectives dans le diagnostic ou la compréhension des mécanismes physiopathologiques de plusieurs maladies d'origine infectieuse ou liées à l'inflammation,

désordres cardiovasculaires, cancers, etc.... Bon nombre de protéines recombinantes issues des biotechnologies sont des glycoprotéines (anticorps, hormones, facteurs de coagulation...) dont le contrôle des processus de glycosylation s'avère essentiel pour assurer leur efficacité thérapeutique (le premier médicament de nature oligosaccharidique mis sur le marché, un pentasaccharide anti-thrombique, est issu de la recherche française). De même, la diversité des propriétés mécaniques et dynamiques des polysaccharides et 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 encore la génération de structures biomimétiques. Nul doute qu'au regard de leurs applications variées dans des domaines aussi divers que la pharmacie, l'agroalimentaire, la cosmétique, les années à venir verront un essor considérable des « glyco-médicaments » et des « glycotecnologies » en général.

## 6.2 Vers le « glycome » et la « glycoproteomique » ?

Les glycanes exercent des fonctions multiples pour l'expression, le repliement, le trafic, la localisation et la durée de vie des protéines auxquelles ils sont liés. Ces molécules universelles 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. Les glycanes participent ainsi à la plupart des grandes fonctions physiologiques (développement embryonnaire, adhésion, migration et différenciation cellulaire, etc.) et leurs modifications ont été observées dans un grand nombre de pathologies (maladies infectieuses, auto-immunes, neurodégénératives et amyloïdes, désordres inflammatoires, troubles cardiovasculaires, cancer, maladies génétiques du métabolisme telles que les « Congenital Disorders of Glycosylation ») (voir ci-dessus). 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 glycanes ainsi que des enzymes participant à leur biosynthèse ou leur dégradation. La « glycoproteomique » est, quant à elle, une méthode de protéomique différentielle qui vise à élucider toutes les glycoprotéines et glycoformes présentes dans un tissu ou une cellule, dans un état physiologique donné. L'établissement du glycome suppose l'étude structurale des glycanes, mais aussi la compréhension des mécanismes de leur biosynthèse et de leurs fonctions.

En raison de la sophistication structurale des glycoconjugués (nature des monosaccharides, types de liaisons, séquence, anomérie, présence de sulfate, phosphate) les approches de la glycomique font appel à une grande diversité de procédés parmi lesquels la spectrométrie de masse et la RMN occupent une place prépondérante. Un effort important est également fait pour la conception de « glycopuces » qui constituent un outil de choix pour l'élucidation des spécificités de reconnaissance des différents récepteurs et lectines. La cristallographie



et la modélisation moléculaire devraient également contribuer à apporter des éclairages nouveaux sur les rôles conformationnels des glycanes et les mécanismes de reconnaissance sucres-protéines. Notons enfin que la glycomique comme l'ensemble des autres méthodes d'approches à haut-débit (protéomique, transcriptomique) reste tributaire du développement de la bioinformatique.

## 7. LES VIRUS

Les virus ont une taille généralement comprise entre 30 et 300 nm et sont incapables de se diviser. Aussi doivent-ils détourner la machinerie cellulaire d'organismes vivants (eucaryotes, bactéries, archées) afin de se répliquer. La multiplication de particules virales dans les cellules infectées peut entraîner leur mort ou leur multiplication incontrôlée pour les virus oncogènes. Les virus sont souvent associés à des risques de pandémie. Des recherches sont indispensables pour guider la conception de nouveaux médicaments, qui peuvent cibler le génome viral, les protéines virales (associées à la réplication, aux voies de régulation, aux mécanismes d'entrée, à l'assemblage) ou des partenaires cellulaires clés.

Jusqu'à présent, la cristallographie aux rayons X permettait d'étudier des virus à l'échelle atomique, tandis que la microscopie électronique permettait d'étudier des virus enveloppés non-cristallisables à moindre résolution. Les progrès réalisés en cryo-microscopie électronique (microscopes TEM-FEG, automatisation du traitement) conduisent aujourd'hui à des reconstructions 3D à une résolution quasi-atomiques (3,3 Å pour l'aquareovirus, un virus non enveloppé), tandis que le développement de la tomographie modifie nos connaissances sur les assemblages non-symétriques. Ces informations peuvent être utilisées en dynamique moléculaire pour modéliser des changements de conformation, ou en bioinformatique pour construire des bases de données de phénotypes mutants. La RMN et la diffusion des rayons X et des neutrons aux petits angles donnent des informations essentielles sur la dynamique des protéines virales, et sur des régions non-structurées susceptibles d'interagir avec différents partenaires cellulaires. Ces zones de flexibilité constituent de nouvelles cibles thérapeutiques.

### 7.1 VIRUS PATHOGENES

Les virus sont à l'origine de maladies humaines graves causées par des virus enveloppés à ARN (dengue, ebola, VIH...) ou ADN (hépatite B, Herpesviridae...) et chez les animaux de graves épidémies virales, causent des dégâts économiques importants. Des efforts considérables ont été réalisés pour mieux comprendre les mécanismes de fusion membranaire associés à l'entrée virale. Les recherches en cristallographie aux rayons X ont ainsi établi la notion de protéines de fusion de classe I, dont le modèle est l'hémagglutinine du virus de la grippe, de classe II chez les flavivirus et alphavirus et de classe III chez les rhabdovirus et les herpesvirus. Ces résultats permettent d'élaborer de nouvelles stratégies antivirales. Les recherches sur les protéines de réplifications des virus à ARN (hélicases, polymérases des virus de la dengue, de l'hépatite C, West-Nile...) ont considérablement avancé grâce à des programmes de génomique structurale qui doivent être poursuivis.

## 7.2 Utilités des virus et bactériophages

Le monde des virus est vaste. Les bactériophages et les baculovirus d'insectes ont joué un rôle décisif dans le développement de la biologie moléculaire et continuent d'être étudiés. Ils constituent des systèmes modèles pour l'étude des mécanismes moléculaires qui régulent les interactions hôte-pathogène. Aujourd'hui, de grands espoirs sont fondés sur l'utilisation en thérapie génique de vecteurs viraux d'eucaryotes (adénovirus, herpesvirus, rétrovirus...) rendus inoffensifs pour transférer des gènes comme médicaments. La découverte de virus géants (mimivirus, mamavirus...) infectant les amibes et contenant de l'ARN en plus de leur ADN génomique a rapproché les virus du monde du vivant. Parallèlement, les études menées sur les virus infectant les archées ont révélé des virions de formes atypiques (bouteille, citron). Ces virus se rencontrent dans les sources chaudes et constituent des réservoirs de nouveaux repliements protéiques utilisables en biotechnologie. Ils peuvent également permettre de mieux comprendre les mécanismes d'évolution et de transferts de gènes, qui ont conduit à la naissance de notre monde à ADN.

## 8. LA BIODIVERSITE DES MICROORGANISMES

### 8.1 Les atouts de la microbiologie

La microbiologie englobe l'étude des procaryotes, des archées et des eucaryotes unicellulaires. Les bactéries et les levures restent des modèles de choix pour décrypter la structure et la fonction des machineries moléculaires universelles, pour analyser au niveau moléculaire les interactions entre biomolécules impliquées dans les réactions biologiques, les voies de signalisation et les réseaux de régulation, qui relie le génotype au phénotype. Les bactéries offrent la possibilité d'étudier la dynamique des interactions et des réseaux de régulation dans une population homogène ou hétérogène (biofilm). Un grand nombre de bactéries exercent des rôles clés dans l'écosystème mais aussi provoquent des infections humaines, qui en milieu hospitalier, posent un problème de santé publique à cause de l'émergence de résistances aux antibiotiques. L'élucidation de la structure des facteurs de virulence (protéines de surface et membranaires), des mécanismes de communication entre cellule («quorum sensing»), des mécanismes de régulation des facteurs de virulence, et de résistance aux antibiotiques peut conduire à de nouvelles stratégies anti-microbiennes. Une analyse comparative des réactions biologiques des bactéries commensales, pathogènes stricts ou opportunistes est utile pour comprendre les mécanismes d'expression de la pathogénicité. C'est sans doute par les bactéries et les organismes unicellulaires, que la biologie synthétique pourra se développer en France.

### 8.2 Découverte de nouvelles voies métaboliques

La très grande majorité des espèces vivantes non



caractérisées à ce jour sont les organismes unicellulaires. Ces espèces représentent un réservoir énorme de diversité peu exploré malgré une impressionnante étendue d'habitats colonisés, de processus biochimiques et moléculaires mis en œuvre et de variations génomiques. Pourtant les microorganismes sont des éléments clés des cycles de la matière; ils sont à la base de nombreux processus écologiques dominants et sont responsables d'un grand nombre des maladies humaines, des animaux et des plantes. Ils sont à l'origine de la majorité des composés bioactifs commercialisés à ce jour. L'achèvement du séquençage de plusieurs centaines de génomes bactériens a marqué le début de l'ère post-génomique. L'approche comparative de ces génomes a permis leur annotation et la prédiction fonctionnelle des gènes (similarité, conservation de leur position relative, co-évolution, etc.), mais elle est malheureusement limitée par le nombre encore trop réduit d'informations biochimiques, fonctionnelles et structurales. Les processus biologiques connus se limitent à un nombre réduit de microorganismes. L'analyse du métabolome (discipline en plein essor depuis 4-5 ans) permet maintenant d'analyser l'ensemble des petites molécules (métabolites, coenzymes, cofacteurs, etc.) produit par un système biologique reflétant les capacités du métabolisme cellulaire. Combinées aux analyses bioinformatiques des génomes séquencés, ces approches devraient conduire à la caractérisation de nouvelles fonctions/voies métaboliques et d'identifier des gènes cibles, qui pourront avoir un intérêt fondamental biomédical ou biotechnologique. En liaison avec d'autres sections CNRS (22, 29), la biodiversité bactérienne peut maintenant être abordée par la métagénomique pour révéler de nouvelles activités enzymatiques, et de nouvelles fonctions des ARN.

### 8.3 Origines de la vie, biosphère

Il est bon de rappeler que quatre cinquièmes de l'histoire de la vie sur notre planète étaient exclusivement microbiens. Tout au long de cette histoire, les écosystèmes microbiens se sont adaptés aux contraintes géochimiques fluctuantes et ont vice versa profondément influencé voire parfois renversé les conditions géochimiques de la terre. La biologie et la géochimie de notre planète s'influencent donc mutuellement de façon profonde et constituent un ensemble fonctionnel indissociable.

En conséquence, une large diversité métabolique a dû évoluer afin que les micro-organismes puissent exploiter toutes les ressources de leur environnement. Sur la terre actuelle, caractérisée par une prévalence d'habitats fortement oxygénés, cette diversité métabolique est retrouvée principalement chez des espèces vivant dans des niches écologiques particulières (souvent appelées «extrêmes»). Cependant, cette flexibilité et pluri-potence de métabolismes reflètent la variabilité (spatiale et temporelle) de l'environnement géochimique sur fond duquel la vie a évolué et à la transformation duquel elle a participé. Une bonne connaissance de la biodiversité des métabolismes et plus particulièrement de leurs mécanismes fondamentaux, de leurs substrats et des unités protéiques impliquées est donc essentielle pour comprendre l'histoire de la vie sur notre planète depuis ses origines.

La mise en évidence de cette biodiversité métabolique, de l'adaptation et de la résistance des micro-organismes aux conditions extrêmes ainsi que le décryptage des métabolismes fondamentaux sont donc des éléments

clés de la compréhension de la vie sur Terre, voire au-delà. 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.) Finalement, une compréhension profonde des anciens bouleversements environnementaux induits par des changements d'écosystèmes permettra de mieux appréhender les effets à long terme de la dégradation anthropogénique actuelle de la biodiversité.

## 9. LA BIOLOGIE STRUCTURALE INTEGRATIVE ET SES GRANDS EQUIPEMENTS

### 9.1 La RMN

La RMN est un outil extrêmement versatile dans l'étude des macromolécules biologiques. Elle est unique dans ses capacités pour étudier la structure, la dynamique et les interactions moléculaires à une échelle atomique.

La biologie structurale se trouve aujourd'hui à un "turning point" car la détermination de structures à l'échelle atomique par les moyens classiques (rayon X, RMN) a atteint sa maturité. Le développement de la biologie structurale se concentrera sur la détermination de structures qui présentent les plus grandes difficultés (ex. protéines membranaires, agrégats peu structurés, machinerie moléculaire...). Pour la RMN, surtout dans l'état liquide, le développement de méthodes de détermination de structure, et leur application, devraient en conséquence devenir moins importants. La part des approches hybrides (en combinaison avec la microscopie électronique à particule unique, par exemple) sera plus significative. Cependant l'un des avantages de la RMN est la possibilité de caractériser la dynamique de macromolécules et leurs interactions. Les développements récents de RMN « in cellulo » permet d'étudier les biomolécules dans les cellules intégrales ou dans un extrait de cellule où des processus cellulaires comme les modifications post-transcriptionnelles des ARN ou post-traductionnelles des protéines peuvent être caractérisés avec une précision sans précédent.

La RMN du solide jouera sans doute un rôle majeur dans les cas de détermination de structure difficiles. Des avancées déterminantes ont été faites dans l'application de cette technique à de structures de plus en plus grandes. L'avantage spécifique de la méthode est qu'elle peut être appliquée à des structures non-cristallines comme les fibrilles.

### 9.2 Synchrotron

L'imagerie utilisant une source de rayonnement synchrotron se développe au niveau international ainsi qu'en France, autour de l'accélérateur SOLEIL sur le plateau de Saclay et du synchrotron européen de Grenoble (ESRF). Des intensités et des flux de photons X considérables sont en effet émis par ces synchrotrons de 3ème génération.

Associés aux progrès techniques sur les onduleurs et les optiques X, ils ont permis la production d'ondes X cohérentes. Ces avancées contribuent au développement d'applications des sources synchrotron pour l'imagerie biologique de grands complexes protéiques, de compartiments intracellulaires, de cellules entières, pour la nano-médecine et la pharmacocinétique, à des résolutions allant jusqu'à la dizaine de nanomètres. On peut distinguer les modes d'imagerie directe et les modes d'analyse et de cartographie chimique. L'imagerie directe par microscopie de transmission X ou de diffraction cohérente apporte des informations 2D/3D d'échantillons biologiques jusqu'à 10 microns d'épaisseur. Ces approches sont complémentaires et corrélables avec la cryotomographie électronique. L'imagerie chimique par microscopie X à balayage, de fluorescence et d'infrarouge, permet de cartographier au sein d'une cellule des éléments chimiques d'intérêt. Ces approches sont corrélables aux données recueillies de NanoSIM ou de cartographie par microscopie électronique. Enfin, la spectroscopie synchrotron UV visible (e.g. ligne DISCO de SOLEIL) doit permettre une spectroscopie visible de cellules entières vivantes, résolue dans le temps à la nanoseconde.

### 9.3 Biocristallographie

L'importance de la cristallographie aux rayons X pour l'étude des relations structures/fonctions des macromolécules biologiques a été particulièrement mise en relief en 2009, avec l'obtention du prix Nobel de Chimie pour l'étude du ribosome et le développement associé d'antibiotiques. Cependant, le champ des recherches structurales reste immense et la cristallographie doit répondre à des problèmes biologiques de plus en plus complexes comme l'étude des protéines intrinsèquement désordonnées, les gros assemblages moléculaires ou les protéines membranaires qui sont difficiles à produire et à cristalliser. Les progrès en robotisation et en miniaturisation (robot de cristallisation nanogoutte, ferme à cristaux) permettent de multiplier et d'optimiser la recherche des conditions de cristallisation pour une quantité de protéine limitée. La puissance et la qualité des lignes de lumière de type « microfocus », accessibles sur les synchrotrons de dernière génération, permettent d'obtenir une information atomique à partir de données enregistrées sur des cristaux de quelques dizaines de microns. L'utilisation de robots passeurs d'échantillons a permis de multiplier le nombre de cristaux testés, tandis que l'automatisation des procédures de changement de longueur d'onde sur les lignes synchrotron a grandement facilité l'utilisation des méthodes de phasages par utilisation du signal anomal. Les nouveaux pixel-détecteurs contribuent à améliorer la qualité du signal mesuré. Enfin, l'amélioration de la puissance des stations de calculs et l'utilisation de nouveaux programmes basés sur des méthodes de maximum de vraisemblance a augmenté considérablement la vitesse et qualité du phasage et de l'affinement des structures cristallographiques.

Une nouvelle façon d'envisager la cristallographie a émergé de ces nouveaux développements et changé profondément les habitudes des cristallographes qui doivent dorénavant adopter une approche pluridisciplinaire intégrant la biologie moléculaire, la biochimie aussi bien que les données obtenues par d'autres méthodes

biophysiques, comme les données à basse résolution obtenues par la diffusion aux petits angles (SANS et SAXS) et la microscopie électronique. La combinaison de cet ensemble de données expérimentales conduira à des résultats structuraux de tout premier ordre.

### 9.4 Diffusion des rayons X ou des neutrons aux petits angles

La diffusion aux petits angles connaît aujourd'hui un essor considérable. Elle permet effectivement d'apporter des réponses à des questions majeures de la biologie structurale vers l'étude de systèmes de plus en plus complexes. Avec des domaines d'applications très variés, cette technique permet d'appréhender les propriétés structurales et fonctionnelles de protéines souvent flexibles, mais pas exclusivement : interactions avec un substrat ou un inhibiteur, assemblages macromoléculaires, protéines intrinsèquement désordonnées, etc. Les derniers développements de cette technique la rendent particulièrement puissante en combinaison avec d'autres méthodes structurales, comme par exemple la RMN (Paramagnetic Relaxation Enhancement, couplages résiduels dipolaires...), ou la cristallographie. La mise au point de nouvelles lignes de lumières dédiées à la diffusion aux petits angles sur les synchrotrons situés sur le sol français vont indubitablement contribuer à l'essor de cette technique dans la communauté française : l'ESRF a ouvert la ligne ID14-EH3 exclusivement pour les biologistes, tandis que la ligne SWING de SOLEIL développe de nouveaux environnements échantillon (HPLC, passeur automatique d'échantillons...) afin d'améliorer la qualité des mesures et diminuer le volume d'échantillon requis. La diffusion des neutrons aux petits angles disponible à l'Institut Laue Langevin (ILL, Grenoble) est également appelée à jouer un rôle majeur dans les années à venir pour l'étude de systèmes biologiques complexes, notamment avec la méthode de variation de contraste, ce qui permet de renforcer ou de masquer le signal provenant d'une partie du complexe. Les neutrons constitueront également un outil de choix pour étudier la dynamique moléculaire à l'échelle de la picoseconde, et l'hydratation au sein des systèmes biologiques.

### 9.5 La cryomicroscopie et la microscopie de force atomique

Les progrès récents de la cryomicroscopie électronique (cryo-EM) permettent aujourd'hui d'acquérir des images tridimensionnelles de l'échelle moléculaire à haute et moyenne résolution par cristallographie électronique et par analyse de particules isolées, jusqu'à l'échelle cellulaire par tomographie. La cristallographie électronique, principalement appliquée aux cristaux 2D de protéines membranaires apporte les informations indispensables de structure dans un environnement membranaire à des résolutions intermédiaires et pour de rares très bons cristaux 2D à des résolutions atomiques. L'analyse dite « de particules isolées » bénéficie d'intenses développements de préparation d'échantillons, d'automatisation des microscopes et algorithmiques qui permettent aujourd'hui de construire des modèles quasi-atomiques ab initio

de virus à haute symétrie, ou pseudo-atomiques après recalage de structure de protéines constitutives obtenues par RMN ou cristallographie des rayons X. Des résolutions subnanométriques permettant de déterminer la structure secondaire sont également maintenant accessibles pour des assemblages moléculaires sans symétrie de plusieurs centaines de kiloDalton difficilement cristallisables comme ceux impliqués dans l'expression des gènes. Les développements algorithmiques s'attachent également aux recalages de structures atomiques obtenues par RMN ou cristallographie dans les enveloppes de microscopie électronique donnant accès aux interactions au sein de larges complexes. Cette approche non cristallographique permet également d'analyser les variabilités de conformations locales au sein des assemblages protéiques. Les progrès attendus dans le traitement des images concernent l'analyse de protéines de petite taille et/ou à faible symétrie, la variabilité conformationnelle et la résolution. La cryotomographie électronique révèle quant à elle la complexité de cellules entières ou de larges complexes à des résolutions de quelques nanomètres. Des efforts importants sont réalisés dans la segmentation des volumes et le recalage de macrocomplexes protéiques. Enfin la cryomicroscopie est de plus en plus souvent impliquée dans les approches corrélatives multi-échelles avec la cristallographie des rayons X et la RMN, et les microscopies optiques et de tomographies X. La cryo-EM couplée à des approches de cinétique rapide pour la formation de gros complexes (par exemple ribosome) offre la possibilité de visualiser les différentes étapes de l'assemblage des particules ou complexes.

Depuis une quinzaine d'années, la microscopie à force atomique s'est également fortement développée et apparaît comme une approche indispensable à l'analyse de molécules de structures connues, protéines membranaires, complexes nucléoprotéiques, virus mais dans un environnement contrôlable, proche de conditions physiologiques. C'est un outil complémentaire de la microscopie électronique, de la RMN et la cristallographie RX. Elle a permis des avancées significatives dans l'organisation supramoléculaire de protéines et de lipides de membranes natives spécialisées (comme celles de la photosynthèse), en cristaux 2D ou systèmes reconstitués. Son utilisation en mode «spectroscopie de force» permet d'aborder les forces d'interactions intra- ou intermoléculaires protéiques, ou les forces d'interaction entre protéine et acides nucléiques.

## **9.6 Fluorescence et techniques associées pour l'étude des relations structure fonction et des interactions moléculaires**

La fluorescence et techniques associées sont largement exploitées in vitro pour étudier les relations structure-fonction des molécules d'intérêt du fait de la facilité de leur mise en œuvre et la sensibilité de la fluorescence. La variation fine du signal de fluorescence en fonction de l'environnement immédiat de la sonde permet de déterminer de manière sensible les changements de conformation dus à une dénaturation ou une interaction moléculaire. La fluorescence est exploitée pour des études

classiques telles celles à l'équilibre thermodynamique que celles plus sophistiquées du suivi de la molécule unique ou encore pour des applications à haut débit.

Le développement de sondes et d'appareils performants (microscopie confocale, caméra electron-multiplying CCDs et algorithmes d'analyse rapides) a permis de faire un bond dans l'exploitation de la fluorescence en microscopie optique. La possibilité d'effectuer des études en temps réel sur des cellules vivantes et l'utilisation de techniques telles que le FRAP (Fluorescence recovery after photobleaching), FLIP (Fluorescence loss in photobleaching) ou la photoactivation permettent de déterminer des cinétiques, le trafic intracellulaire des molécules marquées. Les interactions spatiales et temporelles de protéines d'intérêt peuvent être résolues par des techniques de FRET (fluorescence resonance energy transfer), FLIM (fluorescence lifetime imaging microscopy) et/ou FCS (fluorescence correlation microscopy). Par l'utilisation de la microfluidique couplée à une analyse optique comme la réflexion totale de fluorescence (TIRF), il est possible aussi de suivre les forces d'assemblage moléculaires entre biomolécules.

Les techniques de microscopie à fluorescence «super-résolutives» ont été récemment mises au point. Elles utilisent une illumination mise en forme de telle sorte qu'il est possible d'atteindre une résolution dépassant celle de la microscopie optique classique par au moins un ordre de grandeur (de 0.2 à 0.5  $\mu\text{m}$  à quelques dizaines de nanomètres). L'observation des détails auparavant non résolus de structures cellulaires a démontré la grande promesse de la microscopie de fluorescence super-résolutive à élucider les processus biologiques au niveau cellulaire et à l'échelle moléculaire. C'est ainsi le cas des techniques microscopiques STED (stimulated-emission-depletion), PALM (Photo-activated localization microscopy) et STORM (Stochastic Optical reconstruction microscopy). Ces deux dernières permettent la localisation individuelle de sondes fluorescentes. L'exploitation de ces techniques et le développement de sondes adéquates permettront de définir des paramètres jusqu'alors inaccessibles avec l'imagerie photonique sur des cellules fixées ou vivantes. D'autres méthodes (optique non linéaire) n'utilisant pas de marqueurs sont complémentaires même si elles n'ont pas cette sensibilité (SHG Génération de Seconde Harmonique, THG Génération de Troisième Harmonique, CARS Coherent Anti-Stokes Raman Scattering).

En conclusion l'acquisition d'une image multidimensionnelle en microscopie optique requiert la maîtrise de nombreuses techniques : photonique (lasers), optique (microscopes), électronique (détection) et informatique (traitement des données). Seule une approche interdisciplinaire pourrait permettre le développement d'une nouvelle instrumentation afin d'améliorer la résolution spatiale.

## **9.7 La résonance paramagnétique électronique**

La Résonance Paramagnétique Électronique (RPE) est une technique centrale de caractérisation des propriétés de la matière. Dans le domaine des Sciences du Vivant, elle est particulièrement bien adaptée pour l'analyse au niveau moléculaire des systèmes biologiques, de leurs interactions et de leur fonctionnement.



Technique de haute sensibilité et sans limitation de masse des objets étudiés, la RPE permet l'analyse des contenus en métaux et en cofacteurs métalliques ainsi que la détection des radicaux, qu'ils soient associés à un mécanisme enzymatique ou issu d'un processus d'altération (stress oxydant, irradiation). Les progrès rapides d'insertion de sondes paramagnétiques (spin label) et de codage isotopique alliés aux forts développements technologiques et méthodologiques actuels ont considérablement élargi ses champs d'applications. En effet, l'essor remarquable ces dernières années des techniques de RPE impulsioneuse multidimensionnelle de haute résolution (Spectroscopie de Modulation de l'écho de spin électronique – ESEEM), d'analyse des interactions intercentres (Double résonance électron-électron - DEER), combinées à l'approche multifréquence (de 3 à 300 GHz) permettent l'obtention d'informations structurales comprises entre 0.1 nm (détection du magnétisme nucléaire) et 10 nm (double résonance électron-électron). Ces approches modernes qu'il convient de développer en France ouvrent la possibilité d'étudier à l'échelle moléculaire de nombreux processus biologiques, telles que les métabolismes bioénergétiques, la sensibilité moléculaire aux facteurs de stress (stress oxydant, thermique, chimique), l'action d'inhibiteurs, l'élaboration de stratégie d'ingénierie enzymatique, la cartographie des interactions moléculaires entre partenaires ou le suivi des transitions structurales associées par exemple à la régulation, la signalisation et l'assemblage.

## 9.8 Spectrométrie de masse

La spectrométrie de masse est un outil de choix pour un grand nombre d'études qualitatives et quantitatives dans de nombreux domaines. Son application à la biologie a nécessité des méthodologies spécifiques de préparation d'échantillon, incluant une séparation préalable ou en ligne des constituants. La SM concerne des domaines aussi variés que la protéomique, la métabolomique, la glycomique, les études d'interactomes, et plus généralement les analyses qualitatives ou quantitatives de composés spécifiques.

Les couplages de systèmes chromatographiques avec un spectromètre de masse ont subi de profondes évolutions durant ces deux dernières décennies, avec l'introduction de techniques d'ionisation douces comme l'électronebulisation (Electrospray) ou la désorption laser assistée par matrice (MALDI), indispensables pour l'étude de composés fragiles, instables, ou de poids moléculaire élevé. La séparation est effectuée soit en amont, de plus en plus souvent de manière automatisée, soit en ligne. Par exemple, en protéomique, les solutions retenues sont généralement l'électrophorèse ou la nanoLC, et souvent un couplage des deux techniques, dans cet ordre. Les résultats obtenus permettent d'identifier les espèces présentes, ou par comparaison, entre profils d'expression protéique. Une application majeure est l'identification de biomarqueurs dans une approche clinique.

Les avancées seront avant tout méthodologiques incluant l'identification des protéines et leurs modifications post-traductionnelles, l'identification des ARN et leurs modifications post-transcriptionnelles dans des mélanges complexes, l'analyse descriptive et dynamique des complexes protéine-protéine, et ribonucléoparticules, et la

mise au point de stratégies pour la spectrométrie de masse quantitative. On notera cependant que les développements instrumentaux récents en imagerie par spectrométrie de masse, tant en terme de résolution spatiale (de l'ordre du micron), que spectrale (détermination en routine de formule brute par haute résolution) laissent augurer d'un réel essor de la discipline pour la réalisation de cartographies spécifiques. Les approches à haut débit sont encore sous-exploitées à l'heure actuelle, principalement à cause du manque d'outils bioinformatiques qui permettraient d'accéder à une dimension plus intégrative.

## 9.9 LA «BIO-INFORMATIQUE STRUCTURALE»

La bioinformatique structurale est concernée par des approches informatiques permettant de prédire et analyser les structures spatiales des macromolécules. Des avancées considérables ont été faites dans la prédiction de la structure tridimensionnelle de petites protéines à partir de leur séquence primaire seule. Cependant, dans la plupart de cas, la modélisation à partir d'une structure homologue reste l'approche privilégiée. L'analyse de la structure 3D, même prédite, peut apporter des informations sur la fonction d'une macromolécule qui ne peuvent pas être déduites directement de la séquence primaire. L'analyse combinée des données structurales et génomiques deviendra certainement plus importante dans un avenir proche.

Il devient de plus en plus important de replacer les structures de macromolécules biologiques dans un contexte cellulaire. Ceci nécessite des approches de modélisation à plusieurs échelles : l'échelle atomique (rayons X, RMN) doit être combinée avec des informations provenant de la (cryo-) microscopie électronique, de la tomographie électronique, de la microscopie à super-résolution, ou bien d'autres expériences biochimiques (co-purification, mutagenèse, ...). La modélisation de ces systèmes complexes est aussi un lien vers la biologie systémique, qui souvent ne prend pas les données structurales en compte.

Pour obtenir une image cohérente à partir de ces données hétérogènes et parfois contradictoires, il faut des approches de modélisation adaptées. En conséquence, la modélisation moléculaire prend une place de plus en plus centrale dans la biologie structurale. La dynamique moléculaire est un outil irremplaçable, pour la génération de ces structures de complexes, et pour leur analyse fonctionnelle. Des systèmes de plus en plus grands sont traités par dynamique moléculaire, ce qui nécessite le développement de logiciels plus performants et d'ordinateurs adaptés (parallélisme massif). Des avancées significatives ont été faites au niveau du développement de simplifications type «gros-grains».

## 10. Conclusion et recommandations

Comme le montre ce rapport de prospective, les équipes du CNRS ont été souvent précurseurs dans plusieurs domaines et ont acquis une reconnaissance internationale incontestable, on citera par exemple la structure spectaculaire du ribosome eucaryote, de complexes ribosomiques fonctionnels et de protéines membranaires, la découverte de nouvelles fonctions et mécanisme d'action des ARN, la découverte de nouveaux virus et de leurs structures... La compréhension au niveau moléculaire

et atomique des processus biologiques et des relations structure-fonction des biomolécules à tous les niveaux d'organisation du vivant reste une préoccupation majeure de la section 21. Ces données biologiques fiables (structure des biomolécules et de leurs complexes, interactome...) seront un apport conséquent à la «biologie des systèmes» qui doit amener la dimension intégrative et quantitative de la dynamique des comportements biologiques d'un organisme ou d'une cellule. Ainsi, l'un des objectifs de la section 21 sera d'évoluer vers la biologie intégrative et l'étude de systèmes biologiques de plus en plus complexes étudiés dans des environnements proche de la cellule.

Les avancées en biologie moléculaire et structurale sont souvent liées aux avancées technologiques. Les nombreux défis de la section 21 (i.e., structure-fonction des objets biologiques complexes comme les protéines membranaires, les glycoprotéines, les ARN, les protéines non structurées, et les machineries complexes, l'analyse des interactions transitoires, l'étude des structures des biomolécules dans leur contexte naturel, la découverte de nouvelles fonctions enzymatiques...) requièrent une évolution et/ou adaptation constante des outils technologiques et des instruments en biologie structurale, en imagerie, en biophysique et en bioinformatique (exploitation des données expérimentales provenant des approches haut débit de la protéomique, du métabolome, du transcriptome, interactome, plateforme de criblage haut débit). Un grand nombre de ces instruments et outils sont onéreux en coût d'investissement et d'exploitation et demande une assistance technique importante, qui nécessitent un support financier récurrent et qu'il faut continuer à promouvoir sans doute à travers les programmes IBISA ou les projets européens comme INSTRUCT.

La section 21 se préoccupe de la diminution des postes de jeunes chercheurs, d'une baisse de moyens provenant de l'ANR pour la biologie moléculaire et structurale et d'une répartition du grand emprunt vers un petit nombre de laboratoires par rapport aux nombreuses équipes qui sont hautement performantes sur les systèmes biologiques qu'elles étudient. La compétitivité et le dynamisme des équipes en biologie moléculaire et structurale, repose sur l'évolution significative des moyens mais aussi sur la mise en valeur de l'interdisciplinarité alliant biologiste, enzymologiste, chimiste, physicien, mathématicien et bioinformaticien autour de thèmes forts. Le CNRS est certainement bien placé pour promouvoir ce dialogue mais ceci pourrait aussi se faire à travers des actions spécifiques de l'ITMO BMSV « Bases Moléculaires et Structurales du Vivant » auquel se reconnaît pleinement la section 21.