

Organisation, expression et évolution des génomes

Président

Thierry GRANGE

Membres de la section

Serge BOITEUX
 Nathalie BONNEAUD
 Florence CAMMAS
 Agamemnon CARPOUSIS
 Laurent DURET
 Gwennola ERMEL
 Alain EYCHENE
 Cécile FAIRHEAD
 Clotilde GIMOND
 Olivier JEAN-JEAN
 Christèle MAISON
 Bénédicte MICHEL
 Kim Heng NG-BONAVENTURE
 Béatrice PY
 Lluis QUINTANA-MURCI
 Stéphane RONSSERAY
 Catherine ROUSSEAU
 Jean-Pierre ROUSSET
 René TOCI
 Lucas WALTZER

Introduction

Le génome est au cœur du fonctionnement des organismes vivants, assurant la pérennité et la propagation des propriétés des cellules et des organismes. Il est l'objet d'étude au centre des missions de la section 22 intitulée "Organisation, expression et évolution des génomes". La section couvre les domaines disciplinaires étudiant la structure des génomes et leurs évolutions; les mécanismes contrôlant leur intégrité, leur propagation et leur expression; l'organisation chromatinienne, chromosomique et nucléaire des génomes et les conséquences épigénétiques de ces organisations; les fonctions contrôlant et contrôlées par les génomes, et les liens entre génotype et phénotype. Les chercheurs rattachés à la section 22 étudient généralement ces aspects en utilisant alternativement ou en combinaison des approches moléculaires, génétiques et génomiques, utilisent souvent des systèmes modèles unicellulaires ou multicellulaires mais aussi des approches *in silico*, et peuvent s'intéresser aussi bien à des approches d'étude des molécules uniques que visant à la compréhension de l'intégration fonctionnelle des événements multimoléculaires. L'expansion des séquençages des génomes et l'évolution très rapide des performances des méthodes de séquençage à très haut débit influencent fortement les domaines disciplinaires couverts par la section et élargissent très significativement son assise interdisciplinaire. La bioinformatique génomique et la modélisation arrivent progressivement au cœur de ces disciplines et l'interdisciplinarité associant les approches formelles et expérimentales devient un enjeu stratégique essentiel pour la vitalité des domaines scientifiques relevant de la section 22. La section 22 est consciente des recouvrements avec d'autres sections s'intéressant à l'étude moléculaire des régulations présidant aux grandes fonctions du vivant, en particulier les sections 21 et 23 à 26, et ainsi qu'aux aspects touchant plus particulièrement à l'évolution des génomes et à l'intégration des données génomiques au niveau populationnel qui la rapproche de la section 29 rattachée à l'INEE.

La section s'efforce de gérer ces recouvrements de manière à ce que la cohérence des champs disciplinaires puisse être assurée d'une façon prospective et réactive afin que l'organisation en section et en institut du CNRS permette et favorise les évolutions scientifiques aux frontières des structures.

Les enjeux scientifiques

Dans les domaines d'intervention de la section 22, l'analyse des données issues du séquençage des génomes joue un rôle important. Auparavant, le séquençage des génomes dépendait exclusivement de grosses infrastructures et la France n'a pas joué un rôle

prééminent dans ce domaine compte tenu de la difficulté rencontrée dans la mise en place d'un nombre suffisant d'infrastructures efficaces, en partie faute de moyens. Les techniques de séquençage évoluent vers des machines à très haut débit qui peuvent séquencer plusieurs centaines de Gigabases par semaine à des coûts financiers beaucoup moins élevés, en utilisant peu de matériel initial et en s'affranchissant des étapes de construction et de caractérisation d'une banque génomique. Ceci offre des opportunités d'utilisation du séquençage pour de très nombreuses applications qui influencent et influenceront profondément l'évolution des sciences des génomes et de la vie en général. Cette évolution n'a pas été suffisamment anticipée par les structures déjà existantes et un retard très important dans le domaine des grands centres de génomique a déjà été pris lorsque l'on regarde les évolutions déjà réalisées dans d'autres pays, même dans des pays que l'on n'a pas encore pris l'habitude de considérer au plus haut niveau scientifique comme la Chine.

D'autre part, les machines de nouvelle génération offrent la possibilité de décentraliser la production des données de séquences et de la rapprocher directement des laboratoires d'expérimentation biologique, et donc d'intégrer plus efficacement ces données dans le processus global de production de connaissances.

La décentralisation du séquençage représente aussi un enjeu stratégique de formation des scientifiques à la nouvelle biologie interdisciplinaire qui est en train d'émerger et qui allie analyse bioinformatique et formelle à la biologie expérimentale. Un accompagnement significatif en moyens humains et financiers de cette décentralisation nous paraît stratégique pour tirer le meilleur parti de la qualité actuelle du tissu national de laboratoires de biologie expérimentale d'analyse des génomes et permettre ainsi à la France de tenir encore son rang scientifique dans ce domaine stratégique pour la biologie fondamentale, comme pour les biotechnologies ou la biologie médicale.

Organisation et expression des génomes

L'analyse mécanistique et fonctionnelle de l'organisation chromatinienne, chromosomique et nucléaire des génomes et de leur expression a progressé très significativement ces dernières années et des avancées importantes ont été réalisées grâce d'une part à la caractérisation systématique d'un grand nombre d'enzymes affectant l'organisation chromatinienne, les modifications de l'ADN (méthylation) et l'expression des gènes aussi bien au niveau transcriptionnel que post-transcriptionnel, d'autre part à l'analyse globale des produits de l'activité de ces enzymes : modifications de la chromatine ou de l'ADN, ou ARNs ou protéines exprimées. Les progrès dans la caractérisation systématique des enzymes résultent essentiellement du grand nombre d'équipes qui combinent des approches biochimiques et génétiques.

Ces approches génétiques ont pu être réalisées beaucoup plus systématiquement dans de nombreux systèmes modèles, en particulier cellulaires, grâce à la découverte des ARN interférents. Les progrès et la généralisation des techniques de spectrométrie de masse ont aussi contribué

pour beaucoup à la caractérisation quasi exhaustive des partenaires protéiques des multiples complexes enzymatiques. Il y a encore certainement un grand nombre d'activités qui restent à découvrir et la pérennisation de ces directions de recherche est clairement justifiée. Le tissu des laboratoires français qui s'intéressent à ces aspects est de qualité et de nombreuses équipes ont contribué significativement aux progrès du domaine. Il paraît important de leur permettre de poursuivre à ce niveau, de se régénérer, et d'intégrer les analyses globales dans leurs problématiques. Les analyses globales de modification de la chromatine, de méthylation de l'ADN, et des populations d'ARNs exprimés ou associés à des compartiments ou des complexes protéiques se sont fortement développées ces dernières années, tout d'abord grâce à l'utilisation des puces à ADN (transcriptome, ChIP-chip, MeDip, ...). Ces approches ont récemment fait un saut quantitatif et qualitatif très important suite à l'application des techniques de séquençage à très haut débit (RNA-seq, ChIP-seq, MethylC-Seq, 5C, ...) qui permettent d'augmenter considérablement la précision et l'exhaustivité des résultats. Cette évolution ouvre des perspectives passionnantes d'intégration des données et de modélisation des phénomènes biologiques et moléculaires qui représente un véritable défi pour le domaine, car elle va impliquer une évolution profonde des compétences au sein des équipes de recherche. L'interdisciplinarité entre les sciences expérimentales et formelles est appelée à devenir un des cœurs d'activité des domaines scientifiques couverts par la section 22. Ces approches sont amenées à modifier considérablement notre compréhension de l'organisation chromosomique ainsi que notre image des processus épigénétiques affectant le contrôle de l'expression des gènes, ainsi que la réparation, la recombinaison et la réplication du génome.

Les connaissances obtenues ces dernières années ont permis de mettre en évidence les rôles essentiels de la chromatine et des ARNs régulateurs dans le contrôle de l'activité des génomes. Le mot « épigénétique » est passé rapidement d'un statut périphérique à central, perdant au passage un peu de précision dans la définition de son sens, puisque maintenant il s'applique aussi bien aux phénomènes et aux molécules « autour » du génome et affectant sa fonction qu'aux mécanismes d'hérédité non classique qui reposent sur ces mêmes phénomènes. Les progrès des connaissances dans le rôle de l'organisation chromatinienne et de ses modifications dans le contrôle de l'expression génétique sont trop nombreux pour être listés et détaillés ici.

Les enjeux pour l'avenir concerne aussi bien la mise en évidence de nouvelles activités enzymatiques que la compréhension de leurs fonctions, mais aussi la compréhension des structures chromatiniennes d'ordre supérieur et des liens entre ces structures chromatiniennes et l'architecture nucléaire. La compréhension de ces phénomènes se heurte à des problèmes méthodologiques car ils impliquent des échelles supramoléculaires qui sont mal décrites par les approches microscopiques. Les progrès déjà réalisés dans le domaine de la chromatine et de la mémoire épigénétique ont permis de mettre en évidence clairement le rôle des perturbations de ces modifications dans un grand nombre de pathologies et particulièrement au cours de la cancérogénèse. Ceci a permis à la fois de développer des approches diagnostiques, mais

aussi thérapeutiques (des drogues interférant avec la méthylation de l'ADN sont en phase III d'essais cliniques contre certaines tumeurs, par exemple). La caractérisation des enzymes affectant l'expression génétique et dont le dysfonctionnement est avéré dans des situations pathologiques ouvre la voie pour le développement de molécules thérapeutiques, ce qui montre clairement l'importance que la compréhension fondamentale de ces mécanismes peut avoir pour le développement de nouvelles approches thérapeutiques.

Un deuxième niveau de contrôle de l'expression génétique et qui peut se trouver aussi au cœur des mécanismes épigénétiques concerne l'ARN. Là aussi, les progrès des connaissances fondamentales ouvrent des perspectives thérapeutiques à court et moyen terme. Il semble que les petits ARNs non-codants aient un rôle quasi universel et très conservé au cours de l'évolution. Ils ont des fonctions classiques (biogenèse des ribosomes, spliceosome) et contrôlent également l'expression génique par le système d'ARN interférence qui cible la stabilité des ARN messagers. Ils sont également impliqués dans certains cas dans des modifications de la structure chromatinienne. Ils sont impliqués dans des processus aussi variés que le développement embryonnaire, le contrôle des réarrangements géniques, le contrôle de la transposition et autres. Chez les procaryotes, ce domaine est également en plein essor et intervient dans l'expression génique, même s'il n'utilise pas les mêmes voies que chez les eucaryotes. La maturation post transcriptionnelle des ARN et particulièrement l'épissage permet une diversification importante du génome exprimé et joue un rôle clef dans la complexité des patrons d'expression des protéines, tout particulièrement chez les vertébrés, où le nombre d'introns contenus dans la plupart des gènes de protéines est considérable.

Des progrès importants ont permis l'identification et la caractérisation de nombreux facteurs de cette machinerie. Néanmoins, beaucoup reste encore à faire pour comprendre comment cette machinerie extrêmement complexe s'assemble de manière spécifique au niveau des sites d'épissage et catalyse l'élimination des introns ainsi que dans la compréhension des liens unissant épissage et organisation chromatinienne du génome qui apparaissent de plus en plus clairement. De nouvelles pistes thérapeutiques sont actuellement développées ciblant les facteurs d'épissage. De nombreuses découvertes concernent aussi les édifices moléculaires qui contrôlent le devenir de l'ARNm dans le cytoplasme - traduction, stockage et dégradation. Ainsi, depuis 4 ans, des avancées majeures ont été réalisées dans l'inventaire des ARN qui interviennent dans l'expression des gènes (découverte des ARN non codants longs en plus des micro-ARN), et dans la diversité des protéines qui lient l'ARN. Ces découvertes ont amené une meilleure connaissance de la variété des assemblages ribonucléoprotéiques (RNP) impliqués dans les processus contrôlant l'épissage des ARN messager, leur stabilité, leur dégradation et leur accumulation sous forme de granules nucléaires ou cytoplasmiques. Les laboratoires du CNRS ont été précurseurs dans nombre de ces domaines - remodelage de la chromatine, structure des RNP, régulation de l'épissage et de la dégradation des ARN messagers, ARN non codants - et ont acquis une solide réputation internationale. Le soutien constant du CNRS dans ces domaines permettra d'affermir cette

réputation et d'exploiter ces découvertes pour aboutir à une compréhension intégrée de l'expression des gènes et de la plasticité du génome.

Il n'y a pas que l'expression du génome dont le dysfonctionnement intervienne dans un grand nombre de maladies dont les cancers, la stabilité et la plasticité des génomes joue aussi un rôle important. De nombreuses données montrent que la transformation tumorale est associée à l'instabilité génomique. Stress réplicatif, déficit de réparation de l'ADN, recombinaison désordonnée jouent tous des rôles importants dans l'étiologie des anomalies génétiques associées au cancer et sont observables dans les tissus pré-néoplasiques. Pour étudier en détail les mécanismes de la stabilité, de la plasticité génomique, et de la surveillance du génome, l'utilisation des systèmes modèles est là aussi indispensable. Il est maintenant clair que l'organisation chromatinienne joue un rôle pour l'ensemble des «3R» (Réplication, réparation, recombinaison), comme pour l'expression génétique, comme le montre, par exemple, la découverte récente qu'une enzyme de méthylation de la chromatine jouait un rôle clef dans la détermination de la position des points chauds de recombinaison méiotique. La compréhension fondamentale de ces mécanismes et de la manière dont ils sont liés aura donc des retombées aussi bien cognitives que thérapeutiques.

D'importants progrès ont également été faits dans la connaissance des processus qui permettent à la cellule de s'adapter aux variations de son environnement, aux changements programmés de son état de prolifération ou de différenciation, ou à la signalisation des stress affectant le génome. Pour répondre aux modifications de leur environnement, les cellules eucaryotes utilisent des cascades de phosphorylation qui gouvernent les réponses aux hormones, aux facteurs de croissance, aux stress et à la disponibilité en nutriments, et qui permettent la transduction de signaux externes vers les machineries transcriptionnelles et traductionnelles. Ces voies de signalisations contrôlent aussi bien la synthèse protéique et la croissance cellulaire que la transcription ciblée des gènes gouvernant, entre autres, le métabolisme énergétique, l'adipogenèse ou l'angiogenèse. Leur dérégulation est à l'origine de pathologies prolifératives ou métaboliques. Notre compréhension de ces voies de signalisation et de leurs rapports au vieillissement, à l'apport en nutriments et aux stress liés aux carences et à l'environnement est un atout pour l'élaboration de nouvelles molécules bioactives et pour l'exploration de nouvelles approches thérapeutiques.

Bactériologie, microorganismes eucaryotes et génomes

La microbiologie englobe l'étude des procaryotes et des eucaryotes unicellulaires. Elle touche à plusieurs domaines non-exclusifs dont la plupart ont une composante inter-disciplinaire, et qui peuvent être classés en (1) étude des machineries moléculaires, (2) microbes et santé, (3) microbes et environnement, (4) génomique et métagénomique des microbes. Les microbes sont les outils de choix pour analyser les réactions biologiques grâce à la possibilité d'en isoler les acteurs et les étapes, et

parce qu'ils permettent l'accès à l'ensemble des réactions impliquées dans le fonctionnement et la multiplication d'un organisme vivant.

La microbiologie est un domaine privilégié pour décrypter les mécanismes moléculaires responsables des réactions biologiques dans les organismes vivants. Les machineries moléculaires des procaryotes et des eucaryotes unicellulaires sont simplifiées par rapport à celles des organismes plus complexes, de par la limitation des assemblages multi-protéiques et des modifications post-traductionnelles. Ces machineries assurent des fonctions universelles, qui correspondent à des activités biologiques et enzymatiques fondamentales, présentes dans tous les organismes vivants. Un exemple de concept issu de l'étude des organismes unicellulaires et maintenant généralisé à tous les organismes vivant est le rôle important des « accidents de réplication » dans l'instabilité des génomes. D'autre part, l'accessibilité et la simplicité des machineries moléculaires des organismes unicellulaires a permis le développement de techniques de microscopie et d'imagerie, et de techniques d'analyse en molécule unique sophistiquées et performantes. A côté de ces techniques qui représentent une interface avec la physique, le crible de chimiothèques bloquant l'action d'enzymes particuliers est un exemple d'interface avec la chimie. Le CNRS est, de par sa mission de développement des connaissances, l'organisme scientifique où ces études moléculaires fondamentales doivent être soutenues. Le décryptage précis des mécanismes moléculaires qui soutiennent le vivant constitue également le passage obligé du développement de la biologie synthétique dont les enjeux aussi bien fondamentaux qu'appliqués sont potentiellement immenses. Dans ce cadre, les microorganismes seront sans aucun doute les modèles expérimentaux de choix pour les projets de création « ex nihilo » de systèmes biologiques complexes.

Le rôle du CNRS est également de favoriser l'extension des études des phénomènes fondamentaux vers les microorganismes procaryotes ou eucaryotes qui jouent un rôle important en pathogénicité ou dans le fonctionnement des écosystèmes. L'élargissement des études moléculaires à une plus grande variété d'espèces microbiennes, pathogènes ou commensales permet de définir les bases fondamentales de la pathogénicité. On peut citer par exemple la découverte des mécanismes par lesquels certaines bactéries pathogènes intègrent de l'ADN exogène et donc évoluent, ou les études des mécanismes d'assemblage des systèmes d'adhésion utilisés par les procaryotes pour l'infection des cellules eucaryotes cibles. La description de la régulation des facteurs de pathogénicité est également une composante importante qui doit s'appuyer sur les études réalisées dans les microorganismes les mieux connus.

Le rôle des microorganismes dans le fonctionnement et l'évolution des écosystèmes doit être abordé sous l'angle de la génomique ou de la métagénomique, avec une interface importante avec la bioinformatique, la génétique des populations et l'écologie. Les progrès des techniques de séquençage ouvrent de nouvelles voies de compréhension de l'évolution dans différents habitats et pour analyser la biodiversité au sein d'un écosystème (e.g., projets « métagénome » pour analyser la diversité microbienne ou virale). L'exploration de la biodiversité microbienne par séquençage révèle aussi l'existence

de nouvelles activités enzymatiques qui pourront être caractérisées. Enfin de nouvelles fonctions moléculaires, sources de nouveaux concepts, émergeront comme par le passé de l'étude des microorganismes vivant dans des conditions extrêmes.

Évolution des génomes, Génomique comparative et paléogénomique

On a pu croire qu'après le séquençage du génome humain la course aux génomes allait ralentir. On assiste en réalité à une forte accélération. Cette accélération, facilitée par l'apparition de nouvelles techniques de séquençage rapides et peu coûteuses, est due à l'intérêt de comparer les génomes et d'explorer les divergences évolutives à différentes échelles, depuis les études intra-spécifiques jusqu'aux analyses regroupant les grands domaines du vivant. A ce jour, environ 1000 génomes de bactéries, archées ou eucaryotes sont entièrement séquencés et disponibles publiquement. Les progrès réalisés dans la longueur des lectures des nouvelles techniques de séquençage ainsi que l'élaboration de nouvelles stratégies d'assemblage de génome adaptées aux données issues de ces machines permet d'envisager une progression spectaculaire du nombre de génomes séquencés dans les années à venir. Grâce à ces données et aux nouvelles approches « phylogénomiques », c'est à dire se fondant non pas sur l'analyse de quelques gènes mais sur l'intégralité du génome, il devient envisageable d'élucider la phylogénie du vivant. Ce large échantillonnage taxonomique est particulièrement utile pour l'annotation des génomes (i.e., l'identification des gènes et autres éléments fonctionnels) par analyse comparative de séquences. Ces données offrent également une opportunité unique pour analyser la variabilité des répertoires de gènes et ainsi mieux comprendre l'adaptation des espèces à leur environnement (et notamment l'évolution de la virulence chez des organismes pathogènes). Le séquençage massif ouvre la voie à la « génomique des populations » – c'est à dire à l'analyse de la variabilité génétique au sein d'une population à l'échelle génomique. Ainsi, le projet de séquençage de 1000 génomes humains permettra non seulement de connaître l'histoire et la dynamique des populations de notre espèce mais également d'identifier les régions du génome soumises à pression de sélection (et donc impliquées dans l'adaptation de l'homme à son environnement) et d'analyser les mécanismes mutationnels à l'origine de la diversité génétique. Ces données serviront également de référence pour identifier des mutations impliquées dans des pathologies. Ainsi, le séquençage de génomes complets est une approche très prometteuse pour l'identification des mutations ponctuelles ou des aberrations chromosomiques impliquées dans des cancers.

La reconstitution des événements évolutifs à partir des génomes présents nécessite parfois des interpolations qui peuvent être erronées suite à l'extinction de lignage ou à des approximations erronées des mécanismes génétiques. La possibilité d'apporter des données complémentaires permettant de tester la validité des modèles d'interpolation utilisés en déterminant des portions de séquence des génomes anciens provenant de traces fossiles sort maintenant du cadre de la science-fiction. Il a ainsi été

possible d'obtenir des couvertures très honorables des génomes du mammoth, de l'homme de Néanderthal et d'un autre hominidé apparenté, le Dénisovien. L'analyse du génome du Néanderthalien a permis de mettre en évidence les portions du génome humain sous pression sélective intense depuis la divergence des lignées de l'homme moderne et du Néanderthalien. Elle a aussi permis de montrer l'existence de flux génétiques passés entre ancêtres des humains modernes et Néanderthaliens et Dénisoviens qui éclairent sous un jour nouveau l'histoire évolutive du génome humain. La mise en évidence de pressions sélectives s'exerçant sur des portions du génome, et la capacité de dater ces pressions, permettent d'éclairer la fonction postulée des gènes impliqués et donc contribuera aussi à une meilleure compréhension des relations entre génotype et phénotype dans des situations normales et pathologiques. On peut aussi s'attendre, par exemple, à ce que l'analyse de génomes ancestraux d'organismes domestiqués par l'homme, en utilisant des fossiles datant du début de cette domestication, permette d'aider à l'identification de régions génomiques responsables de caractères d'importance agronomique.

Génétique humaine

Génétique humaine et maladies génétiques ont toujours été parmi les secteurs d'intervention de la section 22 qui est historiquement la section de génétique. Ces thématiques touchent à beaucoup de disciplines telles que clinique, épidémiologie, biologie cellulaire, physiologie, etc... et font donc appel aux expertises d'autres sections des sciences de la vie et de l'INSERM notamment. Les efforts dans ces domaines sont très largement soutenus par des fondations (LNCC, ARC, AFM, FRM, etc...), et l'industrie pharmaceutique. Actuellement, ces recherches vont de la recherche des gènes morbides et de leurs mutations à la caractérisation moléculaire et cellulaire des mécanismes impliqués jusqu'aux applications thérapeutiques (par transfert de gènes ou mise en évidence de cibles médicamenteuses par exemple). Ces efforts impliquent dans la plupart des cas des réseaux collaboratifs importants. A titre d'exemple, des progrès notables ont été obtenus dans de grands groupes de maladies tels que maladies neuromusculaires (myopathies), neurosensorielles (surdités), diabète et obésité. Des progrès importants dans l'efficacité d'association de gènes ou de régions génomiques avec des pathologies ont été réalisés récemment grâce aux approches de génotypage à très haut débit, qui reposent actuellement sur des puces à ADN. Le séquençage à très haut débit commence à être utilisé dans ce contexte, bien que son utilisation ne se substituera pas avant longtemps aux approches de criblage par puces d'un grand nombre d'individus (patients et contrôles de même famille) nécessaires pour chaque pathologie.

Toutefois, les éclairages apportés par le séquençage d'un grand nombre d'individus de différentes populations humaines aura certainement un impact sur le choix des SNPs à interroger selon les populations considérées, voire sur l'interprétation des données. Il paraît important à cause de cela de maintenir le lien entre les diverses analyses génomiques que ce soit génomique fonctionnelle ou génomique des populations, avec la génomique des maladies génétiques.

D'autre part, l'évolution à la baisse prévisible des coûts de séquençage entrainera certainement une augmentation importante du nombre de situations pathologiques qui seront analysées par ces approches, ce qui rendra de plus en plus importants les liens entre recherche génomique fondamentale et appliquée. Un certain nombre de chercheurs s'intéressant aux aspects pathologiques des maladies génétiques sont rattachés au CNRS et à la section 22, et bien qu'une partie de leur activité puisse relever d'autres organismes, il paraît important de maintenir leur lien avec les approches génomiques plus cognitives qui sont bien représentées au sein de la section 22. Il nous paraît donc que la génétique humaine, particulièrement celle liée aux approches génomiques, a toujours sa place au sein de cette section.

Bioinformatique

L'analyse globale des génomes requiert le développement d'outils informatiques importants. A ce titre, la biologie s'est rapprochée récemment des sciences formelles (mathématiques, physique, informatique) pour traiter les énormes masses de données produites, mais des efforts sont à poursuivre pour que la concertation entre ces disciplines conduise à un meilleur rendement de données exploitables en analyse et/ou en modélisation, notamment dans la reconstruction dynamique des réseaux de régulation. Comme discuté auparavant, les progrès des techniques de séquençage ont ouvert la voie à de nombreux champs d'application au-delà du séquençage de génomes. Désormais le séquençage « haut débit » permet d'analyser et quantifier l'expression des gènes (RNAseq), de détecter les interactions ADN-protéine (ChIPseq) ou ARN-protéine, de quantifier la méthylation de l'ADN (MethylC-Seq), d'analyser les processus de réplication, de recombinaison, l'organisation des chromosomes dans le noyau (5C), etc. Ainsi il devient possible non seulement d'étudier la diversité génétique (entre espèces ou entre individus), mais également les modifications épigénétiques des chromosomes et leurs conséquences sur l'expression des gènes. Bref, ces nouvelles techniques révolutionnent l'étude de l'organisation et du fonctionnement des génomes. Par ailleurs, les progrès techniques ne sont pas limités au séquençage. On assiste à la multiplicité des «-omics» (proteomics, metabolomics etc). Toutes ces nouvelles approches ont transformé la biologie moléculaire moderne d'une science "pauvre en données" en une science "riche en données". Cette nouvelle donne représente un défi pour la bioinformatique. D'une part, le volume de données à traiter pose des difficultés algorithmiques importantes. D'autre part, les changements dans la nature même des données impose de nouveaux développements (e.g., les méthodes statistiques développées pour l'analyse de données de puce à ADN ne sont pas directement transposables aux données de RNAseq; les outils d'alignement de séquences couramment utilisés – tels que BLAST – ne sont pas adaptés aux nouveaux types d'analyse de séquence). Les enjeux sont considérables car ces nouvelles données permettent d'envisager, dès à présent, des analyses intégrées allant des séquences complètes des génomes aux conséquences phénotypiques de mutations, en passant par les aspects structuraux et fonctionnels sur les différents acteurs cellulaires. Face à ce volume croissant de données complexes et hétérogènes, l'intégration

des données couplée à des analyses bioinformatiques comparatives et prédictives est cruciale pour réaliser la description étendue de la fonction d'un gène et de la compréhension de son rôle non seulement au niveau moléculaire, mais également aux niveaux supérieurs des voies cellulaires, des complexes macromoléculaires, de la cellule ou de l'organe.

Il paraît donc un enjeu essentiel de réaliser des progrès dans l'analyse bioinformatique aussi bien en ce qui concerne le développement de nouveaux algorithmes, que dans l'analyse des données issues des approches expérimentales. Une attention particulière doit être portée à l'efficacité des transferts de problématiques et d'analyses et d'interprétations des données entre bioinformaticiens et biologistes expérimentaux. Les progrès les plus importants ont été réalisés à l'étranger dans des centres qui avaient su faire cohabiter, sur des sites avec une unité géographique, les différentes compétences afin que s'établisse une véritable culture de l'interdisciplinarité et afin qu'émerge à l'interface des scientifiques avec des compétences doubles qui assurent l'efficacité optimale des flux d'informations interdisciplinaires. Les structures et leurs périmètres d'influence doivent être conçus de façon à favoriser cette interdisciplinarité et à éviter que la bioinformatique formelle continue à maintenir une distance importante avec la biologie expérimentale, distance qui paraît préjudiciable à l'évolution des domaines liés à la génomique. L'association à la section 22 d'une composante bioinformatique génomique forte dans le cadre d'un futur redécoupage des contours des sections paraît un enjeu stratégique devant être sérieusement considéré.

Biologie des systèmes, Biologie synthétique

Dans les domaines scientifiques couverts par la section 22 un changement de fond est en train de s'opérer. En effet, la vision de plus en plus claire des limites du réductionnisme pour résoudre les problèmes complexes (relation génotype-phénotype, métabolisme et physiologie...) implique la nécessité d'aborder ces problématiques avec des outils nouveaux. La biologie des systèmes semble être l'alternative émergente mais cette discipline n'a pas encore pleinement fait ses preuves ni atteint sa maturité. Les années à venir seront déterminantes pour l'essor de cette discipline et pour le positionnement des groupes français qui participeront à ce glissement méthodologique. Cette discipline est toujours en cours d'installation et il est important de soutenir les programmes de qualité l'utilisant. Comme pour la bioinformatique, et d'une façon qui y paraît intimement liée, il faut privilégier les structures qui feront cohabiter intimement approches expérimentales de génomique, analyses et développement bioinformatiques et modélisation formelle. Il faut encourager les initiatives tout en prêtant la plus grande attention à la qualité scientifique des intervenants et des réalisations effectuées. L'interdisciplinarité nécessite des personnalités capables de s'aventurer aux marges de leurs frontières de compétence mais le succès de cette interdisciplinarité requiert que leur compétence cœur soit forte.

La Biologie synthétique paraît aussi être un débouché de fort potentiel qui pourrait émerger de l'interdisciplinarité entre génomique expérimentale, bioinformatique, modélisation

et biologie des systèmes. En effet, non seulement les coûts de séquençage des génomes semblent suivre la loi de Moore appliquées à la croissance exponentielle des capacités de calcul des ordinateurs, mais aussi ceux de la synthèse chimique d'ADN. Les technologies développées pour la synthèse in situ de puces à ADN ont ainsi été adaptées pour produire des séquences utilisables pour des approches enzymatiques, ce qui rend possible la synthèse de très grands fragments d'ADN. La possibilité de rendre concrète la production d'organismes virtuels offre de multiples possibilités, allant de la synthèse de molécules complexes ou de biocarburants par des organismes reprogrammés, à des organismes ou des cellules senseurs dont les propriétés s'adapteraient aux fluctuations de l'environnement de la manière souhaitée. Parmi les tentatives de prédiction futuriste existant dans la littérature, on trouve par exemple la proposition de créer des cellules qui seraient sensibles à la proximité de cellules cancéreuses et qui libéreraient alors des toxines les ciblant. Afin de préparer le terrain à l'émergence de ce domaine, il paraît essentiel de permettre et de favoriser les rassemblements de compétence autour des approches expérimentales et formelles de biologie génomique.

Recommandations

Ce rapport de conjoncture a largement discuté en quoi les différents domaines scientifiques évalués par la section 22 étaient influencés dans leurs évolutions récentes et à venir par l'avènement des technologies de séquençage à très haut débit et l'analyse bioinformatique des données. Ces mutations technologiques, méthodologiques et conceptuelles représentent un enjeu stratégique pour permettre à la France de maintenir une position d'importance dans le paysage scientifique mondial dans les domaines de la génétique, la génomique et la microbiologie, et de se placer à un niveau respectable dans les domaines en émergence de la biologie des systèmes et la biologie synthétique. Afin de répondre à cet enjeu, nous recommandons de favoriser la généralisation de l'accès à ces techniques et à l'analyse bioinformatique des données afin qu'elles deviennent des outils de routine des approches de biologie expérimentale, au même titre que la biologie moléculaire au début des années 1980. Il nous semble qu'il faille profiter de l'opportunité de décentralisation de la génomique rendue possible par les technologies actuelles et futures pour faciliter la diffusion de ces approches conceptuelles et permettre l'adoption de cette nouvelle culture par le plus grand nombre possible de biologistes expérimentaux. Ceci signifie qu'il ne faille pas reculer devant les efforts financiers et les investissements humains que cela représente afin de permettre de faire cohabiter à la fois au moins un grand centre fort de génomique et de multiples structures périphériques. Il est clair que les investissements matériels devront être renouvelés régulièrement car il est facile de prédire que les évolutions technologiques à venir correspondront à des évolutions d'importance qu'il ne faudra pas manquer. Toutefois, le changement culturel que représente l'adoption de modes de raisonnement formels reposant sur les outils informatiques et mathématiques est d'importance pour la majeure partie des biologistes et prendra nécessairement du temps. Le maintien de la compétitivité nationale dans des pans entiers de la biologie repose sur la rapidité avec

laquelle ces approches seront adoptées par un grand nombre de biologistes et de la facilité avec laquelle ces biologistes expérimentaux seront à même de collaborer de manière fructueuse avec des biomathématiciens et bioinformaticiens. Afin de favoriser la diffusion de ces approches dans les laboratoires de biologie expérimentale, il faut rapprocher de ces laboratoires les centres de production de séquences et les équipes de bioinformatique et de biomathématiques. Ceci nécessite clairement de distribuer sur un nombre significatif de sites des investissements assez conséquents en moyens financiers et humains, mais aussi de créer des structures qui permettent de favoriser ces rapprochements. Un des moyens d'atteindre cet objectif est d'associer plus étroitement les instances d'évaluation des biologistes et des bioinformaticiens de la génomique, un rapprochement qui paraît aussi se justifier compte-tenu du rapprochement des domaines scientifiques correspondants.

Le dynamisme et la compétitivité de la recherche biologique en France repose clairement sur une évolution des moyens financiers qui lui permette de faire face à l'importante évolution à la hausse des coûts des technologies et des investissements consentis par les autres pays développés comme par les pays émergents. Bien que les sciences biologiques représentent une part très importante du personnel du CNRS, la part des moyens attribués par le CNRS à ces disciplines n'a pas évolué significativement depuis plusieurs dizaines d'années alors que le domaine a subi déjà plusieurs mutations profondes et que les moyens alloués à ce domaine ailleurs sur la planète ont profondément augmenté. Cette situation a déjà affaibli le rôle joué par le CNRS dans ces disciplines, et si rien n'est fait pour inverser cette situation dans les années à venir, ce rôle finira par devenir marginal. Il nous paraît que cette évolution, même si elle peut paraître acceptable à un certain nombre de biologistes, risque de compliquer très significativement en France le développement de la biologie interdisciplinaire alliant génomique, mathématique et bioinformatique, car le CNRS est un acteur majeur regroupant un nombre important de ces compétences en son sein. De ce point de vue, la section 22 regarde avec inquiétude la diminution régulière des postes ouverts au recrutement CR dans ses domaines de compétence, diminution qui lui paraît excessive compte tenu du poids important de cette section qui compte parmi celles qui ont le plus grand nombre de chercheurs rattachés. Alors que le domaine correspondant est en pleine mutation, la sous-représentation des recrutements frais par rapport à la base importante des laboratoires rattachés ne peut qu'entraîner le déclin du rôle du CNRS dans des thématiques scientifiques d'avenir et lui faire manquer l'émergence d'un domaine fortement interdisciplinaire.

L'évolution actuelle des sources de financement va clairement dans le sens d'une diminution des moyens d'intervention du CNRS et, via l'ANR et le grand emprunt, à une redistribution des moyens vers un petit nombre d'équipes et de sites au détriment d'un grand nombre d'équipes performantes. On observe une évolution récurrente vers un accroissement déraisonnable des écarts de financements alloués aux différentes équipes, officiellement justifiée par une différence de performance des équipes considérées. Toutefois, il paraît à notre instance d'évaluation que les critères d'évaluation utilisés pour justifier ces écarts de financement peuvent être discutés

et que d'autres instances d'évaluation pourraient aboutir à des constats sensiblement différents. Particulièrement, les écarts de financements ne sont bien souvent pas proportionnels aux écarts de qualité scientifique qui peuvent être évalués. Cette politique qui revient en fait à étouffer progressivement une proportion significative d'équipes performantes pour faire des économies budgétaires porte en elle une diminution inéluctable de la qualité scientifique globale de la recherche française car il signifie une diminution du nombre des centres de formation des chercheurs performants de demain: si on enlève le bas de la pyramide, son sommet perdra de l'altitude.