

***INSTITUT THEMATIQUE
MULTI-ORGANISMES***

***Bases Moléculaires et
Structurales du Vivant***



Etat des lieux et orientations stratégiques

Janvier 2011

Abréviations

- ACS : *American Chemical Society*
- ADN : Acide désoxyribonucléique
- AERES : Agence d'évaluation de la recherche et de l'enseignement supérieur
- AFM : *Atomic force microscopy* (microscopie de force atomique)
- AllEnvi : Alliance "Alimentation, Eau, Climat, Territoires"
- AMM : Autorisation de Mise sur le Marché
- ANRS : Agence Nationale de Recherche sur le Sida
- ANRT: Association Nationale de la Recherche et de la Technologie
- ARC : Association pour la Recherche sur le Cancer
- ARN : Acide ribonucléique
- ARNi : ARN interférants
- ARNm : ARN messenger
- ARN Xist : *ARN X inhibitory specific transcript*
- ASBMB : *American Society for Biochemistry and Molecular Biology*
- ASM : *American Society for Microbiology*
- ATIP-AVENIR : Action Thématique et Incitative sur Programme (CNRS) - appel à projets de recherche jeunes chercheurs (Inserm)
- AVIESAN : Alliance nationale pour les sciences de la vie et de la santé
- BAG : *Block Allocation Group*
- BioNMR : *Laboratory for Biomolecular NMR Spectroscopy*
- BioQuant Heidelberg : *Quantitative Analysis of Molecular and Cellular Biosystems*
- CDD : Contrat à durée déterminée
- CEA : Commissariat à l'énergie atomique et aux énergies alternatives
- ChIP on chip : *Chromatin ImmunoPrecipitation on Chip*
- ChIP-seq : *chromatin immunoprecipitation (ChIP) combines with sequencing*
- CLIP : *cross-linking and immunoprecipitation technique*
- CNRS : Centre national de la recherche scientifique
- CPER : Contrat de Projets Etat-Région
- DIM : Domaine d'Intérêts Majeur
- DNA : *Deoxyribonucleic acid*
- DSV : Direction des Sciences du Vivant au CEA
- EC : Enseignant-Chercheur
- EMBL : *European Molecular Biology Laboratory*
- ENS(P) : Ecole Normale Supérieure (*de Paris*)
- ENSC : Ecole Normale Supérieure de Cachan
- EPIC : Établissement public à caractère industriel et commercial
- EPST : Établissement public à caractère scientifique et technologique
- ERC : *European Research Council*

- ESRF : *European Synchrotron Radiation Facility*
- ESFRI : *European Strategy Forum on Research Infrastructures*
- ESPCI - ParisTech : École Supérieure de Physique et de Chimie Industrielle de la ville de Paris
- ETP : Equivalent Temps Plein
- FCS : *Fluorescence Correlation Spectroscopy* (corrélation de fluorescence)
- FEBS : *Federation of the Societies of Biochemistry and Molecular Biology*
- FEDER : Fonds Européen de Développement Régional
- FISH : *fluorescent in situ hybridization* (technique d'hybridation fluorescente)
- FP7 : 7^{ème} programme cadre européen
- FRAP : *Fluorescence Recovery After Photobleaching* (fluorescence après photoblanchiment)
- FRET : *Fluorescence (or Förster) Resonance Energy Transfer* (transfert d'énergie entre molécules fluorescentes)
- FT-ICR : Spectrométrie de masse à transformée de Fourier
- FutuRIS- ANRT : Service de l'ANRT de suivi du Système Français de Recherche et d'Innovation
- GDR : Groupement de recherche
- GFP : *green fluorescent protein*
- GIS : Groupement d'intérêt scientifique
- HFSP : *Human Frontier Science Program*
- HTS : *High Throughput Screening* (Haut débit)
- I3 : *Integrated Infrastructure Initiative*
- IATOS : Ingénieurs, administratifs, techniciens, ouvriers et personnel de service dans l'Éducation Nationale
- IBCDE : Institut de Biologie Cellulaire, Développement et Evolution (ITMO AVIESAN)
- IBiSA : Infrastructures en Biologie Santé et Agronomie
- Ibitecs : Institut de biologie et de technologies du CEA à Saclay
- IBMSV : Institut des Bases Moléculaires et Structurales du Vivant (ITMO AVIESAN)
- IBPC-Paris : Institut de Biologie Physico-Chimique à Paris
- IDMS : *Integrated Database Management System*
- IF : *Impact factor* (Facteur d'Impact)
- IGGB : Institut de Génétique, Génomique, Bioinformatique (ITMO AVIESAN)
- ILL : Institut Laue-Langevin
- IMMI : Institut de Microbiologie et Maladies Infectieuses (ITMO AVIESAN)
- INC : Institut National de Chimie du CNRS
- INCa : Institut national du cancer
- INP : Institut National de Physique du CNRS
- INRA : Institut National de la Recherche Agronomique
- INRIA : Institut National de Recherche en Informatique et Automatique
- INSB : Institut National des Sciences Biologiques du CNRS
- Inserm : Institut national de la santé et de la recherche médicale
- INSTRUCT : *Integrated Structural Biology Infrastructure for Europe funded by the European Union's Seventh Framework Programme*

- IPGP : Institut de Physique du Globe de Paris
- Irtsv : Institut de Recherches en Technologies et Sciences pour le Vivant du CEA à Saclay
- ITA : Ingénieurs, Techniciens, Administratifs dans les EPST et EPIC
- ITMO : Instituts Thématiques Multi-Organismes
- iTRAQ : *isobaric tag for relative and absolute quantitation* (Quantification relative par marquage isotopique)
- ITS : Institut des Technologies pour la Santé
- JCR : *Journal Citation Reports*
- LMD : Licence-Master-Doctorat
- LURE : Laboratoire pour l'utilisation du Rayonnement Electromagnétique
- miARN : microARN
- MNHN : Muséum national d'Histoire naturelle
- NOD : *Nucleotid Oligomerisation Domains*
- OCCAM - Oxford : *Centre for Collaborative Applied Mathematics*
- OCT : *optical coherence tomography*
- OST : Observatoire des Sciences et des Techniques
- MoU : *Memorandum of understanding* (protocole d'intention)
- PACA : Provence Alpes Côte d'Azur
- PALM : *photo-activated localization microscopy* (Photo-activation – Localisation – Microscopie)
- PCRD : Programme Cadre de Recherche et de Développement européen
- PCV : Programme interdisciplinaire en physique et chimie du vivant
- PEPS : Projets de recherche Exploratoires Premier Soutien (CNRS)
- PICS : Projet International de Coopération Scientifique (CNRS)
- PIR : Programme Interdisciplinaire de Recherche (CNRS)
- PIRibio : Programme interdisciplinaire de recherches sur les systèmes moléculaires et cellulaires, et d'innovation biomédicale
- PNA : *Peptide Nucleic Acid*
- PSB : *Partnership for Structural Biology*
- ReNaBi : Réseau National de Bioinformatique
- RMN : Résonance Magnétique Nucléaire
- RNA : *Ribonucleic acid*
- RNP : RiboNucléoProtéiques
- RPE : Résonance Paramagnétique Electronique
- SAGE : *Serial Analysis of Gene Expression*
- SASELELX : Soutien aux Activités Scientifiques Françaises autour des Lasers à Électrons Libres Émettant des Rayons
- SAXS : *Small-angle X-ray scattering*
- SERS : technique de Fluorescence et Diffusion Raman exaltée de surface
- SFBBM : Société Française de Biochimie et Biologie Moléculaire
- SFEAP : Société Française d'Electrophorèse et d'Analyse Protéomique
- SFSM : Société Française de Spectrométrie de Masse

- siARN : *small interfering RNA*
- SILAC : *stable isotope labelling by amino acids in cell culture*
- smFRET : *Single Molecule Fluorescence Resonance Energy Transfer*
- SNRI : Stratégie Nationale de Recherche et d'Innovation
- SRM : *Selected Reaction Monitoring*
- STED : *Stimulated Emission Depletion microscopy*
- STORM : *3D Super-Resolution Microscopy Technique*
- SVSE5 : Programme Blanc ANR Physique, chimie du vivant et innovations biotechnologiques
- SWOT : de l'anglais *Strengths* (forces), *Weaknesses* (faiblesses), *Opportunities* (opportunités), *Threats* (menaces)
- TAP tag : Tandem Affinity Purification
- TGIR : Très Grandes Infrastructures de Recherche
- TIRFM : *total internal reflection fluorescence microscope*
- UPBL : *ESRF Upgrade Programme*
- UPMC : Université Pierre et Marie CURIE

Termes anglais en italique

SOMMAIRE

Abréviations	- 2 -
Rédacteurs et correspondants	- 11 -
Avant-propos	- 12 -
Résumé	- 12 -
Le champ BMSV et la SNRI	- 12 -
1. LES RECHERCHES RELEVANT DU DOMAINE BMSV	- 13 -
1.1. <i>Périmètre</i>	- 13 -
1.1.1. <i>Thèmes</i>	- 13 -
1.1.2. <i>Méthodes d'investigation</i>	- 14 -
1.2. <i>Secteurs de recouvrement avec les autres ITMO et place en BMSV</i>	- 14 -
1.2.1. <i>Médicament (avec ITS)</i>	- 14 -
1.2.2. <i>Imagerie (avec ITS et IBCDE)</i>	- 14 -
1.2.3. <i>Biologie des systèmes (avec IGGB)</i>	- 15 -
1.2.4. <i>Biologie synthétique (avec ITS et l'Alliance Allenvi)</i>	- 15 -
1.2.5. <i>Microbiologie, maladies infectieuses et Cancer (avec IMMI +Cancer)</i>	- 16 -
1.3. <i>Quelques points de repères concernant l'IBMSV</i>	- 16 -
1.3.1. <i>Mots-clés</i>	- 16 -
1.3.2. <i>Communautés concernées</i>	- 16 -
1.3.3. <i>Thèmes du domaine</i>	- 17 -
1.3.3.1. <i>Molécules et chimie du et pour le vivant</i>	- 17 -
1.3.3.2. <i>Aspects biophysiques du fonctionnement subcellulaire et des macromolécules biologiques</i>	- 17 -
1.3.3.3. <i>Contrôle et modélisation du vivant</i>	- 17 -
1.3.3.4. <i>Biologie synthétique</i>	- 17 -
1.3.4. <i>Objectifs poursuivis pour la rédaction du présent document</i>	- 18 -
1.4. <i>BMSV et défis de la SNRI</i>	- 18 -
2. ETAT DES LIEUX	- 19 -
2.1. <i>Laboratoires, équipes, personnels à l'IBMSV</i>	- 19 -
2.1.1. <i>Analyse quantitative</i>	- 19 -
2.1.1.1. <i>Grandes masses</i>	- 19 -
2.1.1.2. <i>Origines institutionnelles des personnels permanents</i>	- 21 -
2.1.2. <i>Analyse géographique</i>	- 24 -
2.1.2.1. <i>Principaux sites</i>	- 24 -
2.1.2.2. <i>L'Ile de France</i>	- 27 -
2.1.2.3. <i>Variations par site</i>	- 31 -

2.1.2.4.	Cartographie des localisations d'équipe par thème	- 32 -
2.2.	<i>Publications</i>	- 32 -
2.2.1.	Les revues de référence du domaine et leur impact.....	- 32 -
2.2.2.	L'impact des publications françaises en BMSV	- 35 -
2.2.3.	Conclusions.....	- 37 -
2.2.4.	Analyse quantitative et géographique	- 38 -
2.2.4.1.	Aspects quantitatifs.....	- 38 -
2.2.4.2.	Répartition géographique des publications	- 39 -
2.2.4.3.	Conclusions et liens avec les opérations campus	- 39 -
2.3.	<i>Financements du domaine BMSV</i>	- 40 -
2.3.1.	L'ANR.....	- 42 -
2.3.2.	Les financements de l'émergence : PEPS et PIR du CNRS	- 44 -
2.3.2.1.	PEPS	- 44 -
2.3.2.2.	PIR	- 44 -
2.3.3.	L'émergence des jeunes chercheurs et équipes	- 45 -
2.3.3.1.	Recrutements.....	- 45 -
2.3.3.2.	Jeunes équipes.....	- 45 -
2.3.3.2.1.	Programmes Atip/Avenir.....	- 45 -
2.3.3.2.2.	ANR Jeune chercheur(se).....	- 46 -
2.3.3.3.	Formations	- 47 -
2.3.4.	Conclusions.....	- 47 -
2.3.5.	Les financements européens et internationaux en BMSV.....	- 47 -
2.3.5.1.	Human Frontier Science Program.....	- 47 -
2.3.5.2.	L'European Research Council	- 48 -
2.3.5.3.	Autres financements européens	- 48 -
2.3.5.4.	Conclusions.....	- 49 -
2.4.	<i>Conclusions</i>	- 49 -
3.	<i>MOLECULES ET CHIMIE DU ET POUR LE VIVANT</i>	- 50 -
3.1.	<i>Définition du domaine</i>	- 50 -
3.2.	<i>Principaux défis scientifiques et technologiques</i>	- 50 -
3.2.1.	Exploiter la diversité moléculaire structurale et fonctionnelle	- 50 -
3.2.2.	Développer des sondes pour l'étude des interactions moléculaires.....	- 51 -
3.2.3.	Relever les défis de la chimie thérapeutique	- 52 -
3.3.	<i>Etat des lieux</i>	- 54 -
3.3.1.	Généralités.....	- 54 -
3.3.2.	Structures d'animation et de coordination.....	- 54 -
3.4.	<i>Analyse stratégique (SWOT)</i>	- 54 -
3.4.1.	Points forts.....	- 54 -

3.4.2.	Points faibles.....	- 55 -
3.4.3.	Opportunités.....	- 56 -
3.4.4.	Risques	- 56 -
3.5.	<i>Recommandations</i>	- 56 -
3.6.	<i>Propositions opérationnelles</i>	- 56 -
4.	<i>ASPECTS BIOPHYSIQUES DU FONCTIONNEMENT SUB-CELLULAIRE ET DES MACROMOLECULES BIOLOGIQUES</i>	- 57 -
4.1.	<i>Définition du domaine</i>	- 57 -
4.1.1.	La biologie structurale	- 58 -
4.1.1.	Optique pour la biologie et méthodes de la molécule unique.....	- 60 -
4.1.2.	Mesure de forces et d'interactions sur des systèmes vivants.....	- 62 -
4.2.	<i>Principaux défis scientifiques et technologiques</i>	- 62 -
4.3.	<i>Etat des lieux</i>	- 62 -
4.4.	<i>Analyse stratégique</i>	- 63 -
4.4.1.	Points forts.....	- 63 -
4.4.2.	Points faibles.....	- 63 -
4.4.3.	Opportunités.....	- 64 -
4.4.4.	Risques	- 65 -
5.	<i>CONTROLE ET MODELISATION DU VIVANT</i>	- 66 -
5.1.	<i>Définition du domaine</i>	- 66 -
5.2.	<i>Approches scientifiques et technologiques</i>	- 66 -
5.2.1.	Mécanismes moléculaires du contrôle qualitatif et quantitatif.....	- 66 -
5.2.1.1.	Caractérisation des « acteurs » moléculaires et des mécanismes de reconnaissance : macromolécules ; petites molécules	- 66 -
5.2.1.2.	Etude mécanistique de la régulation	- 68 -
5.2.2.	Analyse phénotypique du Vivant par le biais d'approches globales (« ~omiques ») .-	70 -
5.2.2.1.	Inventaire des « acteurs » moléculaires des systèmes biologiques	- 70 -
5.2.2.2.	Analyses dynamiques du Vivant.....	- 71 -
5.2.2.3.	Analyses dynamiques du Vivant.....	- 73 -
5.2.2.4.	Les défis de l'analyse à grande échelle.....	- 74 -
5.2.2.4.1.1.	Les défis liés aux méthodes de préparation et d'analyse des échantillons.....	- 74 -
5.2.2.4.1.2.	Les défis liés aux méthodes d'analyse, de structuration et d'intégration des données	- 74 -
5.2.2.4.1.3.	Métabolomique et Fluxomique.....	- 74 -
5.2.2.4.1.4.	Défis transverse : Intégration des données phénotypiques et modélisation....	- 75 -
5.3.	<i>Etat des lieux</i>	- 75 -
5.4.	<i>Analyse stratégique (SWOT)</i>	- 76 -
5.4.1.	Points forts.....	- 76 -

5.4.2.	Points faibles.....	- 76 -
5.4.3.	Opportunités.....	- 76 -
5.4.4.	Risques	- 77 -
5.5.	<i>Recommandations/Propositions opérationnelles</i>	- 77 -
5.5.1.	Recommandations générales	- 77 -
5.5.2.	Cas particulier des analyses « ~omiques ».....	- 78 -
5.5.2.1.	Protéomique	- 78 -
5.5.2.2.	Métabolomique	- 79 -
6.	<i>BIOLOGIE SYNTHETIQUE, GENIE BIOLOGIQUE ET BIOMIMETIQUE</i>	- 80 -
6.1.	<i>Définition du domaine</i>	- 80 -
6.2.	<i>Principaux défis scientifiques et technologiques</i>	- 80 -
6.2.1.	Biologie synthétique	- 80 -
6.2.1.1.	Protocellules	- 80 -
6.2.1.2.	Génomés minimaux.....	- 81 -
6.2.1.3.	Systèmes orthogonaux.....	- 81 -
6.2.1.4.	Circuits artificiels de régulation	- 82 -
6.2.1.5.	Ingénierie métabolique	- 82 -
6.2.1.6.	Considérations générales sur la biologie synthétique.....	- 82 -
6.2.2.	Biotransformations et ingénierie enzymatique	- 83 -
6.2.3.	Biocarburants	- 84 -
6.2.4.	Bioremédiation.....	- 85 -
6.2.5.	Biomimétisme.....	- 85 -
6.2.6.	Le défi sociétal	- 85 -
6.3.	<i>Etat des lieux</i>	- 86 -
6.4.	<i>Analyse stratégique</i>	- 86 -
6.4.1.	Points forts.....	- 86 -
6.4.2.	Points faibles.....	- 86 -
6.4.3.	Opportunités.....	- 87 -
6.4.4.	Risques	- 88 -
6.5.	<i>Recommandations</i>	- 88 -
6.6.	<i>Propositions opérationnelles</i>	- 89 -
7.	<i>AXES TRANSVERSAUX</i>	- 90 -
7.1.	<i>La bioinformatique</i>	- 90 -
7.2.	<i>Modélisations mathématiques et bases moléculaires du vivant</i>	- 96 -
7.2.1.	Spécificité de l'IBMSV en besoin d'outils de modélisation.....	- 97 -
7.2.2.	Outils de modélisation	- 97 -
7.2.3.	Recommandations/propositions	- 98 -
7.2.3.1.	Défis.....	- 98 -

7.2.3.2.	Opportunités.....	- 98 -
7.2.3.3.	Recommandation	- 99 -
7.3.	<i>Grands équipements relevant du domaine BMSV</i>	- 99 -
7.3.1.	Sources de lumière synchrotron	- 99 -
7.3.2.	Sources de neutron et laser	- 101 -
7.3.3.	Sources de champs magnétiques intenses pour la biologie structurale	- 101 -
7.3.3.1.	La RMN	- 101 -
7.3.3.2.	La FT-ICR.....	- 102 -
7.3.3.3.	Recommandations.....	- 102 -
7.4.	<i>Les réseaux de plateformes nationales distribuées et la place d'IBISA</i>	- 102 -
7.5.	<i>Les réseaux européens ESFRI : la place d'INSTRUCT en BMSV</i>	- 103 -
7.5.1.	INSTRUCT	- 103 -
7.5.2.	Autres programmes ESFRI concernant BMSV	- 104 -
8.	<i>RESUME DES CONCLUSIONS ET RECOMMANDATIONS</i>	- 105 -
8.1.	<i>Soutien financier de la recherche BMSV</i>	- 105 -
8.1.1.	Financements de la recherche en BMSV.....	- 105 -
8.1.2.	Plateformes d'exploration du vivant	- 105 -
8.1.3.	Grands instruments.....	- 105 -
8.2.	<i>Pratique de l'interdisciplinarité en BMSV</i>	- 106 -
8.3.	<i>Sites visibles en BMSV</i>	- 106 -
9.	<i>ANNEXES</i>	- 107 -
9.1.	<i>Analyse des périmètres</i>	- 107 -
9.1.1.	Collecte des données	- 107 -
9.1.2.	Statistiques.....	- 108 -
9.2.	<i>Données chiffrées globales</i>	- 109 -
9.3.	<i>Analyse bibliométrique</i>	- 110 -
9.3.1.	Procédure.....	- 110 -
9.3.2.	Présentation des Indicateurs Bibliométriques	- 110 -
9.3.3.	Production des indicateurs et présentation des résultats pour la France.....	- 111 -
9.3.4.	Comparaison Internationale des données bibliométriques 2006-2007.....	- 111 -
9.3.5.	Cartographies : répartition géographique des publications par région	- 112 -

Rédacteurs et correspondants

Ce document a été réalisé au cours du premier semestre 2010 sous l'égide de Thierry Meinzel (Directeur) et Carine Giovannangeli (Directeur-adjoint) avec les contributions de Corinne Brachet-Ducos (CNRS), Nicole Haeffner (Inserm), Amandine Péchalat (Inserm) et Franc Pattus (CNRS) et des experts scientifiques suivants.

Relais auprès des opérateurs de recherches: Gérard Déléris (INC-CNRS), Jérôme Garin (DSV-CEA), Carine Giovannangeli (Inserm), Thierry Meinzel (INSB-CNRS), Benoit Perthame (INRIA), Jean-Charles Portais (INRA/Universités), Felix Rey/Pedro Alzari (Institut Pasteur), Charles Simon (INP-CNRS).

Relais auprès des instances nationales : Pascale Romby, Anne Imberty, Jean-François Mouscadet, Catherine Royer, Jean Weissenbach

Experts scientifiques :

Thème 1 : Jean-Daniel Brion (Chatenay-Malabry), Françoise Gueritte (Gif-sur-Yvette), Anne Imberty (Grenoble), Eric Deprez (Cachan), Gérard Déléris (Bordeaux), Georges Massiot (Toulouse), Jean-François Mouscadet (Cachan), Jean-Serge Remy (Strasbourg), Thierry Durand (Montpellier), Francis Schuber (Strasbourg), Eric Westhof (Strasbourg).

Thème 2 : Yves Bourne (Marseille), Catherine Etchebest (Paris), Anne Houdusse (Paris), Bruno Kieffer (Strasbourg), Bruno Klaholz (Strasbourg), Gilles Labesse (Montpellier), Claudine Mayer (Paris), Eva Pebay-Peyroula (Grenoble), Felix Rey (Paris), Bruno Robert (Saclay), Catherine Royer (Montpellier), Laurence Salome (Toulouse), Terrence Strick (Paris). Avec une contribution de la section 11 du CNRS Interface Physique/Chimie/Biologie Rédacteur coordinateur Cécile SYKES (Paris) Co-rédacteurs : Laurent HELIOT (Villeneuve d'Ascq), Zohar MISHAL (Marseille), Jean-François LEGRAND (Strasbourg), Christine GOURIER (Paris)

Thème 3 : Jérôme Garin (Grenoble), Pascale Romby (Strasbourg), Benoit Perthame (Paris), Jean-Charles Portais (Toulouse). Avec une contribution de la section 21, de Mathias Springer et Marc Dreyfus (Paris) pour la partie 5.2.1.

Thème 4 : Jean Weissenbach (Evry), Denis Pompon (Gif)

Avant-propos

Ce document n'a pas encore atteint sa forme définitive et il comporte un certain nombre de lacunes ou de formulations inappropriées. Une version plus aboutie et complète de ce texte devra ensuite être débattue au sein de la communauté et des instances nationales.

Résumé

L'Institut « Bases Moléculaires et Structurales du Vivant » (IBMSV) concerne les chercheurs et leurs équipes dès lors qu'ils s'intéressent à l'étude des systèmes biologiques, de leur dynamique, des interactions et inter-conversions, depuis le niveau moléculaire jusqu'au niveau cellulaire, ainsi qu'à la compréhension de l'organisation de la cellule et à l'élaboration de composés chimiques permettant d'interférer avec son fonctionnement. Il comprend la caractérisation structurale et/ou fonctionnelle de macromolécules individuelles ou organisées en complexe. Il s'intéresse aussi aux édifices biologiques dans leur contexte cellulaire, aux propriétés physiques des cellules, à leurs interactions et communications. Dans ce contexte sont couverts des développements technologiques originaux tels que de nouvelles méthodes d'imagerie moléculaire et subcellulaire, de quantification, de modélisation, de synthèse de composés mimes ou de régulateurs artificiels. Le domaine concerné implique donc des recherches pluridisciplinaires associant biologie, physique, chimie, bioinformatique et mathématiques.

Le champ BMSV et la SNRI

La Stratégie Nationale de Recherche et d'Innovation (SNRI) vise à répondre aux enjeux sociétaux majeurs. Pour répondre à ces enjeux, quatre grands domaines de recherche prioritaires ont été identifiés en 2009. Le champ BMSV se situe pleinement et majoritairement dans un de ces axes : la santé, l'alimentation et le bien-être. Pour répondre aux enjeux de santé publique et de développement économique de cet axe, une des priorités de recherche affichées dans le cadre de la SNRI concerne la caractérisation du vivant, du gène aux écosystèmes et inclut notamment la compréhension des mécanismes biologiques à l'échelle moléculaire, un domaine au cœur des préoccupations du champ BMSV. En outre un accent particulier est mis sur le domaine de la modélisation du vivant qui est absolument nécessaire au développement d'approches prédictives et quantitatives ainsi que sur le rôle central des approches interdisciplinaires et des infrastructures en biologie. Ces aspects aussi sont essentiels dans les thématiques BMSV, comme les développements de ce texte le démontreront.

1. LES RECHERCHES RELEVANT DU DOMAINE BMSV

Deux aspects interconnectés peuvent globalement exprimer ce qui motive les recherches soutenues dans le cadre de l'IBMSV :

- *Comprendre, « visualiser » et quantifier des mécanismes permettant aux composants moléculaires d'œuvrer ensemble dans leur environnement cellulaire dans le cadre d'un processus biologique donné.*
- *Mettre en place des approches interdisciplinaires et manipuler des systèmes biologiques avec une précision moléculaire pour atteindre l'objectif ci-dessus.*

1.1. Périmètre

1.1.1. Thèmes

L'ITMO Bases Moléculaires et Structurales du Vivant (IBMSV) a pour ambition de s'intéresser aux mécanismes qui sous-tendent la dynamique spatiale et temporelle, la structure fine et les régulations des systèmes biologiques et de leurs composants moléculaires. Ces molécules du vivant peuvent être d'origine biologique ou non-naturelle, pour les composés qui interfèrent artificiellement avec les fonctions biologiques, intra- ou extra-cellulaires. La nature des molécules peut être simple (métabolites primaires et secondaires), plus complexe (lipides, polymères de sucres, protéines et leurs modifications, acides nucléiques) ou résulter d'assemblages complexes et hétérogènes (par exemple ribosomes, membranes, organelles ou particules virales). Les molécules effectrices artificielles correspondent aux marqueurs (biomarqueurs en santé, sondes en imagerie), inhibiteurs, médicaments et leurs métabolites (pharmacologie moléculaire). Les systèmes/modèles biologiques pertinents pour ces études peuvent appartenir à différents règnes du vivant, animal ou végétal.

Les thématiques scientifiques regroupent donc celles des équipes qui étudient les mécanismes fins du vivant. Il s'agit de reconstruire et d'analyser les structures en trois dimensions des molécules biologiques, de comprendre les transformations chimiques et biochimiques qu'elles sont capables de réaliser ou d'analyser de manière détaillée la manière dont les médicaments agissent sur les cellules vivantes. L'objectif est de décrire et d'agir sur les molécules complexes et remarquables que la nature a sélectionné pour assurer toutes les tâches dont la cellule à besoin pour survivre et croître : enzymes, pompes, moteurs, transporteurs, messagers... Il s'agit aussi de comprendre les régulations mises en jeu entre différents partenaires ou composants d'un même chemin métabolique (voies de transduction par exemple) et de parvenir à une description détaillée de la dynamique structurale et fonctionnelle des génomes et des réseaux d'interaction menant à une expression génique coordonnée. La compréhension formelle par modélisation *in silico* des équilibres et la prédiction des conséquences de déséquilibres induits par exemple lors d'une pathologie ou d'un traitement, sont des buts ultimes souvent poursuivis. La prédiction par la modélisation passe par une description expérimentale quantitative des phénomènes cellulaires étudiés.

1.1.2. Méthodes d'investigation

Ces études requièrent des approches multidisciplinaires issues de la Biologie, de la Chimie, de la Physique, des Sciences de l'ingénieur, de l'Informatique et des Mathématiques. Le champ BMSV correspond donc à de nombreuses équipes d'interfaces associant des chercheurs de différentes communautés et l'interdisciplinarité s'inscrit ainsi comme une constante du domaine. Les approches biophysiques permettent d'obtenir des « images » du fonctionnement des molécules, depuis le niveau atomique jusqu'à des niveaux plus intégrés comme ceux des machines cellulaires complexes. Les progrès dans les méthodes de détection permettent d'accéder aux informations relatives à un édifice moléculaire unique et à ses transitions dynamiques locales. La modélisation moléculaire permet d'unifier les données portant sur les propriétés chimiques, structurales et dynamiques des biomolécules pour produire des modèles cognitifs de leurs fonctions et interactions. La biochimie permet quant à elle de suivre, de comprendre et d'analyser toutes les transformations qu'assure la cellule : production d'énergie, fabrication de molécules, réparation des lésions, élimination des déchets... La chimie du vivant permet de mimer des ligands ou des réactions biologiques et de fournir des outils pour l'étude de fonctions biologiques. Les médicaments sont des molécules souvent synthétiques qui agissent sur les processus cellulaires en mimant, détournant ou contrecarrant des effets qu'accomplissent des molécules naturelles. Les informations biophysique et biochimique, sont souvent nécessaires pour comprendre la manière dont ils agissent et pour concevoir ou identifier de nouvelles molécules actives. Les biomathématiques concernent la modélisation dynamique ou quantitative des systèmes biologiques. L'approche bioinformatique est enfin une composante indispensable et transverse concernant le « mining » et la hiérarchisation des analyses réalisées à haut-débit (~omique, dosages dits « high-throughput ») ainsi que le traitement ad hoc des données acquises avec des systèmes à haute technologie et dédiés (voir méthodes biophysiques et reconstructions d'images 2- ou 3-D).

1.2. Secteurs de recouvrement avec les autres ITMO et place en BMSV

1.2.1. Médicament (avec ITS)

Le développement de médicaments comprend plusieurs phases, depuis la conception de nouvelles stratégies thérapeutiques, jusqu'à l'autorisation de mise sur le marché (AMM). L'ITS couvre les recherches permettant d'arriver plus rapidement à l'AMM alors que l'IBMSV comprend des aspects plus amont, tels que le développement rationnel de composés ou la conception de systèmes de criblage.

1.2.2. Imagerie (avec ITS et IBCDE)

La biologie structurale et la biophysique ont connu une évolution importante vers l'étude de systèmes de plus en plus complexes. Les avancées techniques liées à la biophysique de la molécule unique sont applicables aussi bien sur des systèmes biologiques purifiés *in vitro* que dans le contexte cellulaire. Ceci est vrai aussi bien pour les microscopies optiques à haute et à super-résolution, d'abord développées pour la molécule unique (ATPases, kinésine, etc...), la microscopie et tomographie électronique à haute résolution, ou la combinaison des deux (microscopie corrélative), et sont maintenant utilisées au niveau

de la cellule entière. Donc les biologistes structuraux sont passés de la molécule à la cellule. En même temps, une évolution inverse s'est opérée chez les biologistes cellulaires (BCDE), qui, eux, sont descendus en échelle, partant de la cellule vers la molécule. Ces deux communautés, ayant des cultures assez distinctes, sont en train de converger autour de ces microscopies avancées et des études passionnantes sont devenues possibles. C'est une rencontre très riche, et qui mérite une coordination étroite entre les deux ITMOS (BCDE et BMSV) au niveau national, tant les équipements sont importants et les compétences requises, interdisciplinaires.

L'imagerie en BMSV implique aussi un recouvrement certain avec l'imagerie médicale de l'ITMO TS, et ceci à au moins deux niveaux. Le premier est celui de la chimie des traceurs et des composés de contraste ciblant les tumeurs, organes ou cellules spécifiques. On se réfère à l'utilisation des sondes fluorescentes en chirurgie afin de mieux identifier tumeurs, organes, etc... Le second niveau d'interaction est celui des applications des avancées techniques aux organismes entiers, par exemple l'application des approches de microscopies à fluctuation dans les animaux. Les frontières entre la biologie structurale, la biophysique moléculaire et cellulaire, et la biophysique médicale sont assez floues. Les progrès dans tous ces domaines bénéficieraient d'une grande et profonde coopération entre tous les acteurs. De même qu'un partage trop formalisé des approches, travaux et intérêts nuirait certainement aux évolutions.

1.2.3. Biologie des systèmes (avec IGGB)

La biologie des systèmes correspond à une étude par essence quantitative de la dynamique des phénomènes/comportements biologiques. Ce domaine - qui présente des ambitions fortes dans le domaine prédictif - est multi-échelle et concerne tous les niveaux d'intégration du vivant. Pour l'IBMSV, la biologie des systèmes concerne les relations entre petites et grandes molécules au niveau cellulaire. La biologie systémique a pour but d'intégrer des données pour partie résultant d'analyse à haut débit et/ou de réconcilier des analyses de natures et d'origines très différentes. Ce domaine accorde une part importante aux approches computationnelles avec utilisation de modèles stochastiques et déterministes et des méthodes statistiques pour mettre en adéquation optimale, le modèle avec les données expérimentales. Enfin, la biologie des systèmes s'ancre fortement sur les bases de données de connaissance et la biologie prédictive empruntées à la bioinformatique (moléculaire, évolutive, structurale, métabolique..) qui constituent ainsi un pilier important du domaine.

1.2.4. Biologie synthétique (avec ITS et l'Alliance Allenvi)

La biologie synthétique tend à créer des organismes nouveaux par une combinaison rationnelle d'éléments biologiques standardisés qui sont découplés de leur contexte naturel. Il s'agit d'une extension de la biotechnologie, avec le but ultime d'être capable de concevoir et construire des systèmes biologiques fabriqués qui traitent l'information, manipulent les éléments chimiques, produisent de l'énergie, fournissent de la nourriture, maintiennent et améliorent la santé humaine et notre environnement. La biologie synthétique possède un côté théorique avec des éléments de langage provenant des circuits électriques.

Pour l'IBMSV, cette approche fédère l'ensemble des biotechnologies (trouver des voies de synthèses intéressantes au niveau industriel, faire produire un médicament, créer de nouveaux matériaux), participe à la recherche des origines de la vie et emprunte à la biologie des systèmes et à la bioinformatique.

1.2.5. Microbiologie, maladies infectieuses et Cancer (avec IMMI +Cancer)

L'analyse des publications croisées (voir & 2.2.2) démontre sans ambiguïté l'importance de ces thèmes comme sources d'inspiration thématique et d'application pour les équipes du domaine BMSV. Il en résulte une forte implication dans ces domaines. Il s'agit clairement des deux types de pathologies qui éveillent le plus d'intérêt. On notera aussi la prise en compte des pathologies liées au vieillissement et qui impliquent des états de repliement ou d'agrégation particuliers de certaines protéines (exemple du prion).

1.3. Quelques points de repères concernant l'IBMSV

1.3.1. Mots-clés

Biomolécule, effecteur, ligand, assemblage, macromolécule, modélisation, prédiction, biologie structurale, enzymologie, biophotonique, ciblage, thérapeutique, métabolomique, transcriptomique, protéomique, bioinformatique, pharmacologie, médicament, marqueur, sonde, régulation de l'expression, réparation, recombinaison, réplication, biophysique, molécule unique

1.3.2. Communautés concernées

- Commissions nationales spécialisées

CNRS : à l'INP : S5, S7 ; à l'INC : S11, S12, S16 ; à l'INSB : S21, S22, S23, S28, S30, S43

Inserm : CSS2, CSS3, CSS7

CEA: Services de Grenoble (Irtsv), Saclay (Ibitecs), Cadarache

- Panels ERC et AERES incluant des aspects BMSV

PE5 Materials and Synthesis: materials synthesis, structure-properties relations, functional and advanced materials, molecular architecture, organic chemistry

PE5_13 Environment chemistry

PE5_14 Coordination chemistry

PE5_16 Biological chemistry

PE5_18 Homogeneous and heterogeneous catalysis

PE5_20 Macromolecular chemistry,

PE5_21 Polymer chemistry

PE5_22 Supramolecular chemistry

PE5_23 Organic chemistry

PE5_24 Molecular chemistry

LS1 Molecular and Structural Biology and Biochemistry: molecular biology, biochemistry, biophysics, structural biology, biochemistry of signal transduction

LS1_1 Molecular biology and interactions

LS1_2 General biochemistry and metabolism

LS1_3 DNA biosynthesis, modification, repair and degradation

LS1_4 RNA synthesis, processing, modification and degradation

LS1_5 Protein synthesis, modification and turnover

LS1_6 Biophysics

LS1_7 Structural biology (crystallography, NMR, EM)
LS1_8 Biochemistry of signal transduction

LS2 Genetics, genomics, bioinformatics and systems biology

LS2_3 Proteomics
LS2_4 Metabolomics
LS2_5 Glycomics
LS2_10 Bioinformatics
LS2_11 Computational biology
LS2_12 Biostatistics
LS2_13 Systems biology
LS2_14 Biological systems analysis, modelling and simulation

LS7 Diagnostic tools, therapies and public health

LS7_2 Diagnostic tools (e.g. genetic, imaging)
LS7_3 Pharmacology, pharmacogenomics, drug discovery and design, drug therapy

LS9 Applied life sciences and biotechnology

LS9_1 Genetic engineering, transgenic organisms, recombinant proteins, biosensors
LS9_2 Synthetic biology and new bio-engineering concepts
LS9_10 Biomimetics

1.3.3. Thèmes du domaine

Sur la base du périmètre défini et des spécificités propres, il a été défini quatre thèmes couvrant l'ensemble des thématiques de l'IBMSV.

1.3.3.1. Molécules et chimie du et pour le vivant

Animateurs: Anne Imberty (Grenoble), Jean-François Mouscadet (Cachan)

1.3.3.2. Aspects biophysiques du fonctionnement subcellulaire et des macromolécules biologiques

Animateurs : Catherine Royer (Montpellier). Dans ce thème deux grands sous-thèmes ont été identifiés :

- *Méthodologies, biophysique (Thème 2.1)*
- *Représentation structurale des macromolécules et intégration biologique (Thème 2.2)*

1.3.3.3. Contrôle et modélisation du vivant

Animateurs : Pascale Romby (Strasbourg), Jérôme Garin (Grenoble), Jean-Charles Portais, (Toulouse) et Benoit Perthame (Paris). Dans ce thème, deux grands sous-thèmes ont été définis :

- *Mécanismes de régulation, interactions entre biomolécules intervenant dans les processus cellulaires (réparation, transport, transcription, traduction...) (Thème 3.1)*
- *Approches globales (spectrométrie de masse, protéomique, génomique, réseaux d'interactions) et métabolisme-métabolomique (Thème 3.2)*

1.3.3.4. Biologie synthétique

Animateur : Jean Weissenbach (Evry)

1.3.4. Objectifs poursuivis pour la rédaction du présent document

Il a été demandé aux responsables et groupes de travail des différents domaines, avec l'aide d'un groupe d'experts qu'ils auront consulté, de travailler le sujet dans le cadre d'un travail collectif visant à définir un schéma stratégique avec pour objectifs :

- de définir le périmètre du domaine BMSV et de préciser les sous-périmètres
- d'identifier les défis/enjeux scientifiques et technologiques du domaine
- de réaliser une analyse des force-faiblesses-opportunités-menaces (SWOT), à partir d'un état des lieux (équipes, données ressources humaines, données financières, données bibliométriques) et des réseaux/structurations existants.
- de proposer des recommandations pour l'évolution du domaine
- d'identifier les besoins et les modalités de fonctionnement de grandes infrastructures nationales ou européennes ouvertes et partagées permettant de renforcer la communauté en BMSV.

1.4. **BMSV et défis de la SNRI**

Dans le domaine « sciences du vivant », parmi les trois composantes du défi deux concernent plus spécifiquement BMSV.

- Compréhension et modélisation du vivant. BMSV est concerné par approches ~omiques (thème 3.2) et imagerie (thème 2) ; l'accent mis sur l'interdisciplinarité est au cœur des recherches pratiquées dans le domaine BMSV.
- Ingénierie du Vivant. Pour BMSV cela concerne les applications pour la santé (thème 1), les technologies du vivant (thème 4) et la biologie des systèmes (thème 3.1)
- Parmi les enjeux de recherche mobilisateur concernant BMSV, on peut citer la diversité du vivant (diversité biochimique: thème 1 et 3.2).
- Dans le domaine « Santé », recherche fondamentale : mécanismes moléculaires des maladies (cancer, pathologies transmissibles, maladies liées à l'âge).

2. ETAT DES LIEUX

L'analyse stratégique présentée dans ce document s'appuie sur un état des lieux, recensement des équipes du domaine et production scientifique par une analyse bibliométrique. Les méthodes et quelques données utilisées sont décrites en Annexes (#9).

2.1. Laboratoires, équipes, personnels à l'IBMSV

2.1.1. Analyse quantitative

2.1.1.1. Grandes masses

Tableau 1 : Nombre d'équipes et unités par thème

Thème	Nombres d'équipes	Nombre d'unités de recherche
1. Chimie	79	42
2.1 Biophysique	97	62
2.2 Structure	76	34
3.1 Mécanismes	101	44
3.2 ~omique	71	50
4. Synthétique	31	20
Totaux	455	154*

* ne correspond pas à la somme des lignes car une même unité peut abriter plusieurs équipes de thèmes différents

Le domaine BMSV se répartit de façon assez homogène entre les différents thèmes/sous-thèmes, le thème 4 étant le moins important en taille. Des équipes de différents thèmes/sous-thèmes se retrouvent souvent dans la même unité, avec une forte interpénétration des thèmes 2 et 3 qui coexistent souvent au sein de la même unité de recherche. Il existe quelques unités de recherche très spécialisées dans ce domaine (IBS/UVHCI-Grenoble, CBS-Montpellier, AFMB-Marseille, LEBS/VMS-Gif, IBPC-Paris, ARA-Strasbourg...).

Tableau 2 : Répartition des personnels dans chaque thème

Effectifs par domaine BMSV	Chercheurs & Enseignants-Chercheurs	Ingénieurs & techniciens	Doctorants	CDD-post-doctorants	Totaux
1. Chimie	320	131	107	88	647
2.1 Biophysique	310	114	178	101	702
2.2 Structure	331	208	176	196	910
3.1 Mécanismes	302	152	174	111	740
3.2 ~omique	154	142	74	69	439
4. Synthétique	91	30	23	29	172
Totaux	1507	777	732	594	3610

* personnes physiques

Tableau 3 : Composition des équipes BMSV par thème

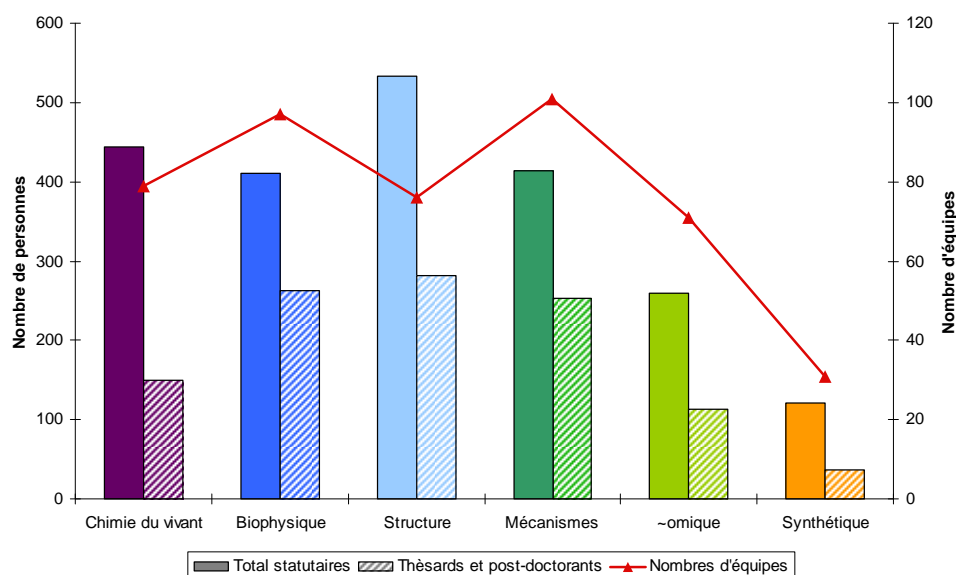
Thème	Personnes par équipe	Taux d'encadrement (permanents/total)	Nombre de permanents par équipe	Nombre d'EC par équipe	Nombre d'ITA par équipe	Rapport Chercheurs+EC /ITA
1. Chimie	8,2	0,68	5,6	2,1	1,7	2,4
2.1 Biophysique	7,2	0,56	4,1	1,2	1,2	2,5
2.2 Structure	12,0	0,56	6,7	1,1	2,7	1,5
3.1 Mécanismes	7,3	0,61	4,5	1,1	1,5	1,9
3.2 ~omique	6,2	0,69	4,3	0,6	2,0	1,2
4. Synthétique	5,5	0,74	4,1	1,1	1,0	3,3
Moyenne	7,9	0,62	4,9	1,2	1,7	1,9

* personnes physiques

Le domaine BMSV comprend donc 2300 permanents et 3600 personnes au total. Le chiffre de 3600 est à comparer avec les 12225 (dont 56% de chercheurs des organismes soit 6846) du domaine biologie fondamentale ou 26000 biologie-santé (voir rapport OST2008 tableau disponible en annexe, & 9); la communauté BMSV représente donc 20 % des personnels en biologie fondamentale. En résumé, les spécificités par rapport à la moyenne, par thème sont :

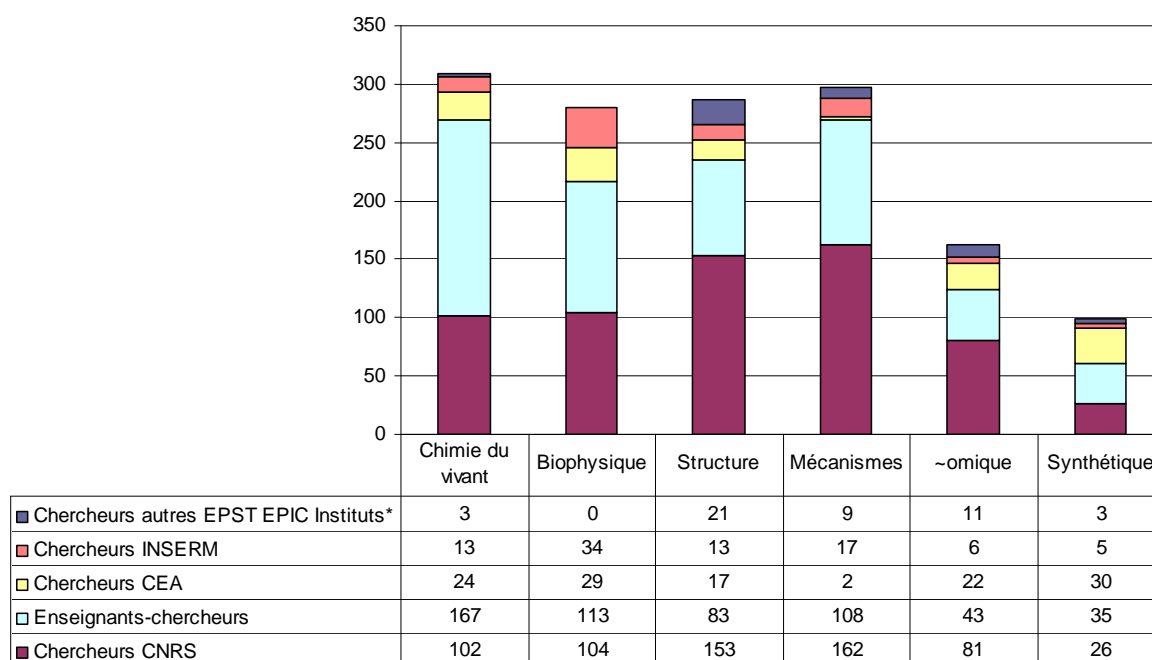
- 1. *Chimie* : forte implication des universités (2 fois plus d'EC)
- 2.1 *Biophysique* : peu d'ITA et plus de chercheurs
- 2.2 *Structure* : grosses équipes avec plus d'ITA et de non-permanents
- 3.1 *Mécanismes* : pas de particularités par rapport à la moyenne
- 3.2 *~omique* : beaucoup plus d'ITA par permanent, très peu d'EC
- 4. *Synthétique* : petites équipes composées de chercheurs et avec peu d'ITA et de non-permanents

Figure 1 : Types de personnels et équipes par thèmes



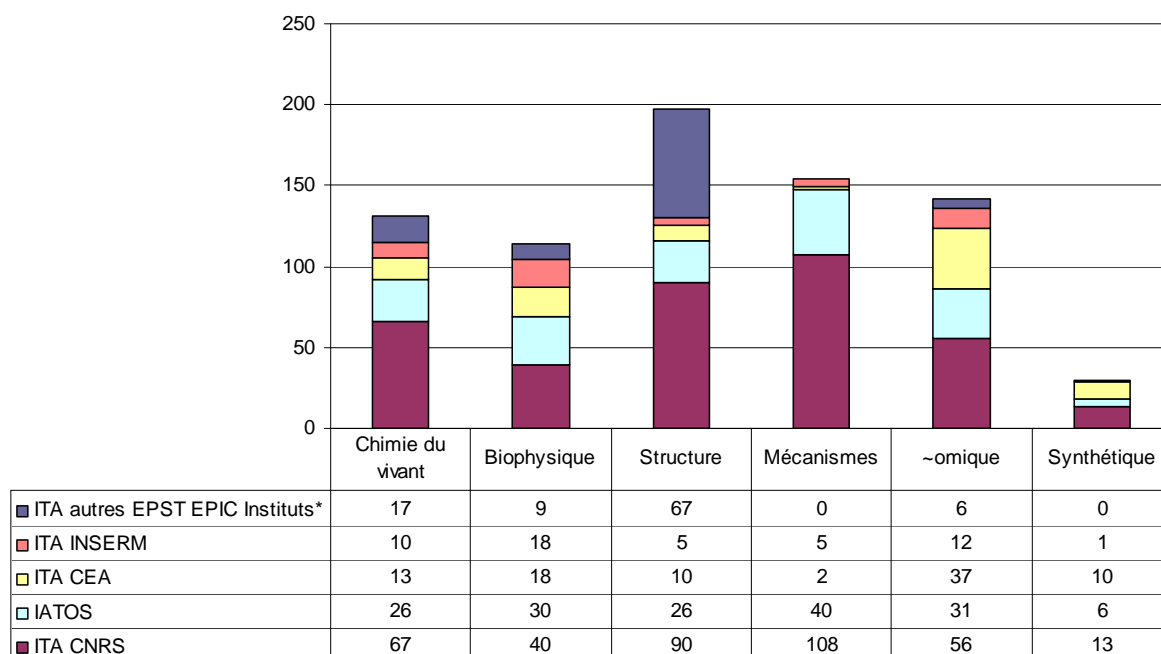
2.1.1.2. Origines institutionnelles des personnels permanents

Figure 2 : Chercheurs et enseignants-chercheurs



* Institut Pasteur (Paris), EMBL (Grenoble), INRA, INRIA, Institut Curie (Paris)

Figure 3 : Ingénieurs et techniciens



* Institut Pasteur (Paris), EMBL (Grenoble), INRA, INRIA, Institut Curie (Paris)

Il faut noter le très haut niveau de technicité généralement requis dans le domaine BMSV, sauf dans le domaine « mécanismes, 3.1 ».

Figure 4 : Effectifs et statuts des personnels des équipes relevant de l'IBMSV

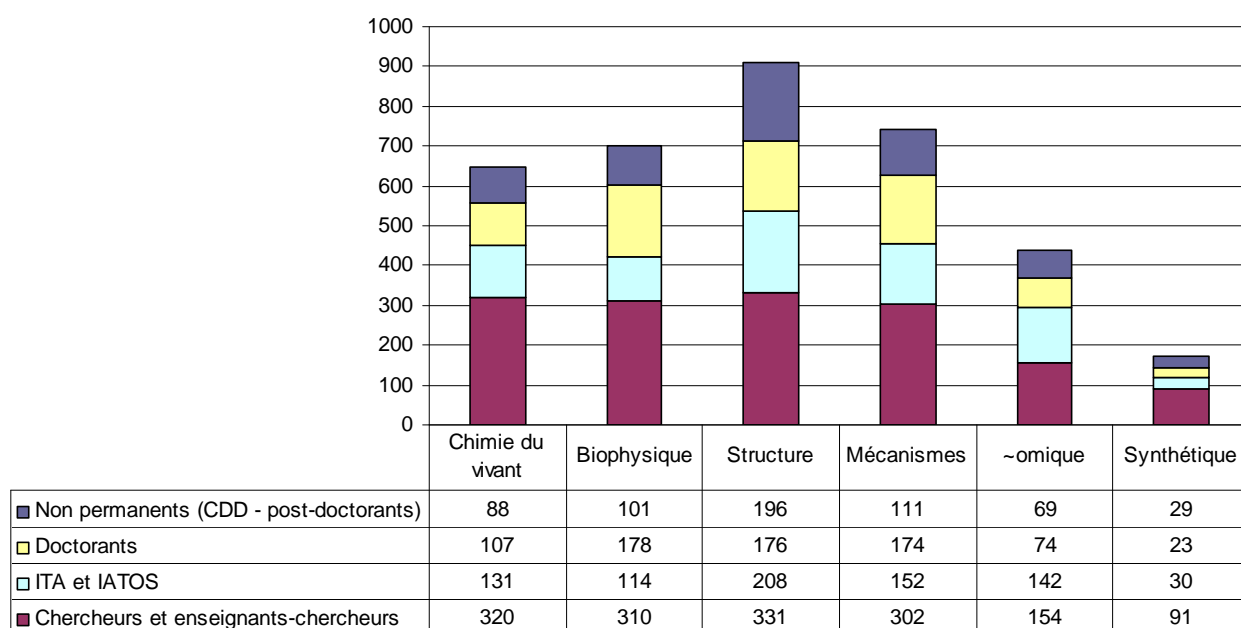


Tableau 4 : Synthèse des effectifs globaux

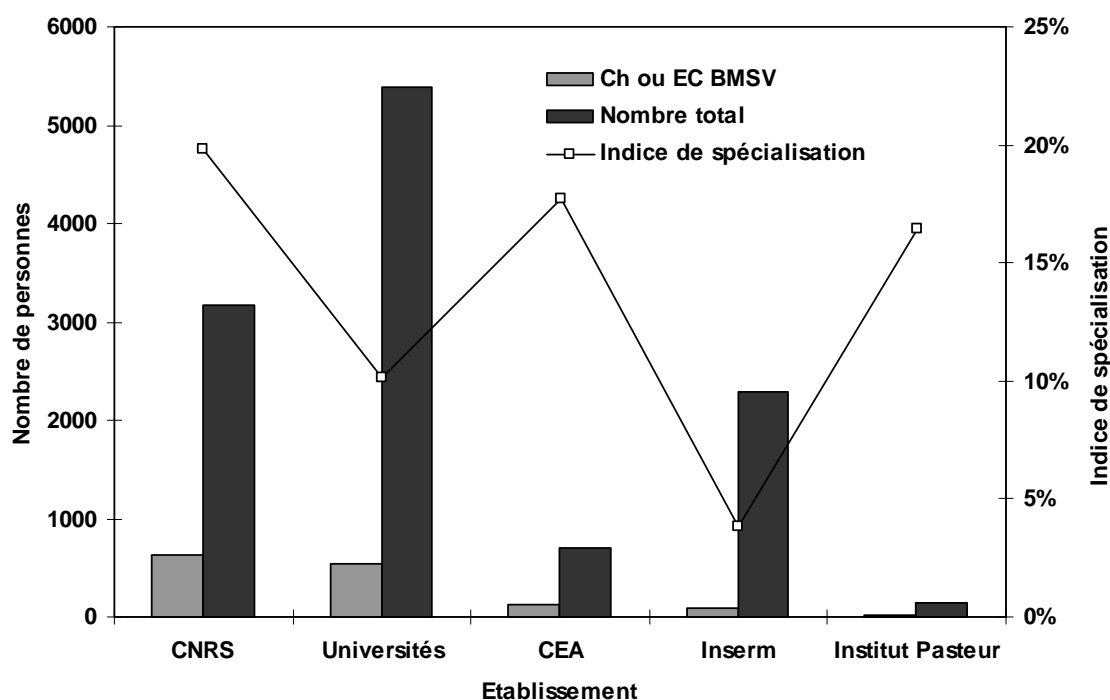
Effectifs	Chercheurs EPST EPIC*	Chercheurs Inserm	Chercheurs CNRS	Chercheurs tot EPST	EC	ITA Inserm	ITA CNRS	ITA EPST	IATOS	Post docs	Doctorants	Total statutaires	CDD	Total Chercheurs	Total non permanents	ITA autres EPST	Total ITA et IATOS	Thésards et post doctorants	Nombre d'équipes
Chimie	27	13	102	154	167	10	67	106	26	42	107	444	46	320	88	50	131	149	79
Biophysique	16	34	104	192	113	18	40	84	30	85	178	411	16	310	101	27	114	262	97
Structure	55	13	153	246	83	5	90	182	26	106	176	532	90	331	196	88	208	282	76
Mécanismes	11	17	162	194	108	5	108	112	40	78	174	413	33	302	111	0	152	253	101
~omique	33	6	81	134	43	12	56	111	31	39	74	260	30	154	69	43	142	113	71
Synthétique	33	5	26	62	35	1	13	24	6	14	23	121	15	91	29	10	30	37	31
Totaux	175	88	628	981	547	50	373	620	157	364	732	2181	230	1507	1326	197	777	1096	455

* non-CNRS ou Inserm

Figure 5 : Implication des établissements en BMSV

Mode de calcul :

CNRS = personnels de sections 16, 21-28, 30, CEA = DSV + chimie, Universités = uniquement biologie fondamentale non médicale ; Autres : EMBL, Ecoles, MNHN...



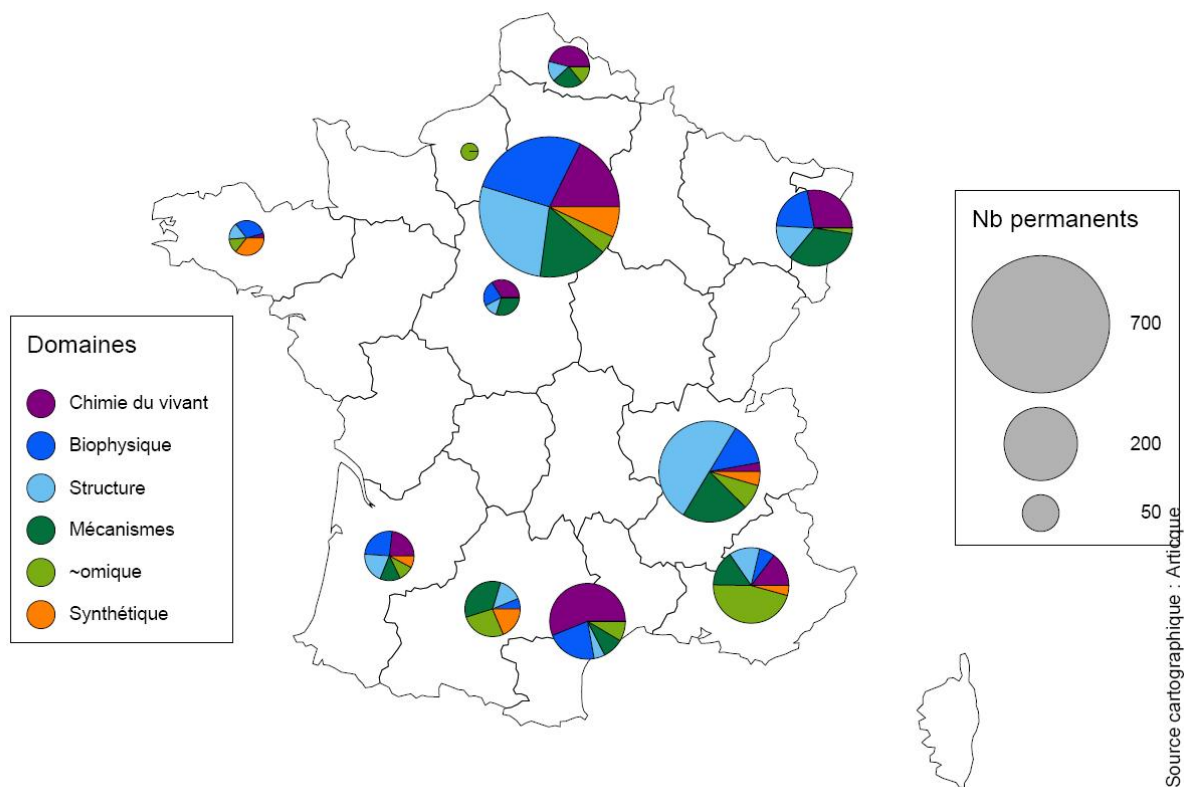
Etablissement	Ch /EC BMSV	Nombre total	Indice de spécialisation
CNRS	628	3171	20%
Universités	547	5379	10%
CEA	124	700	18%
Inserm	88	2295	4%
Institut Pasteur	24	146	16%
Autres	96	674	14%
Sommes	1411	11691	12%

Le CNRS et les universités sont très impliqués et omniprésents en BMSV. Les Universités sont plus spécialisées dans le thème 1, le CNRS dans les thèmes 2.2 et 3.1. Proportionnellement à ses effectifs en biologie-chimie, l'implication du CEA en BMSV est forte, surtout dans les thèmes 2.1 Biophysique et 4. Synthétique (Grenoble, Saclay). Comparativement, l'Inserm a une implication moindre et est surtout spécialisée dans le thème 21. Il faut noter les cas particuliers des deux URA de Pasteur (Paris) et de l'UVHCI-EMBL (Grenoble), l'implication de l'INRA et l'Inserm en 3.2 (~omique).

2.1.2. Analyse géographique

2.1.2.1. Principaux sites

Figure 6 : Répartition des personnels BMSV en France

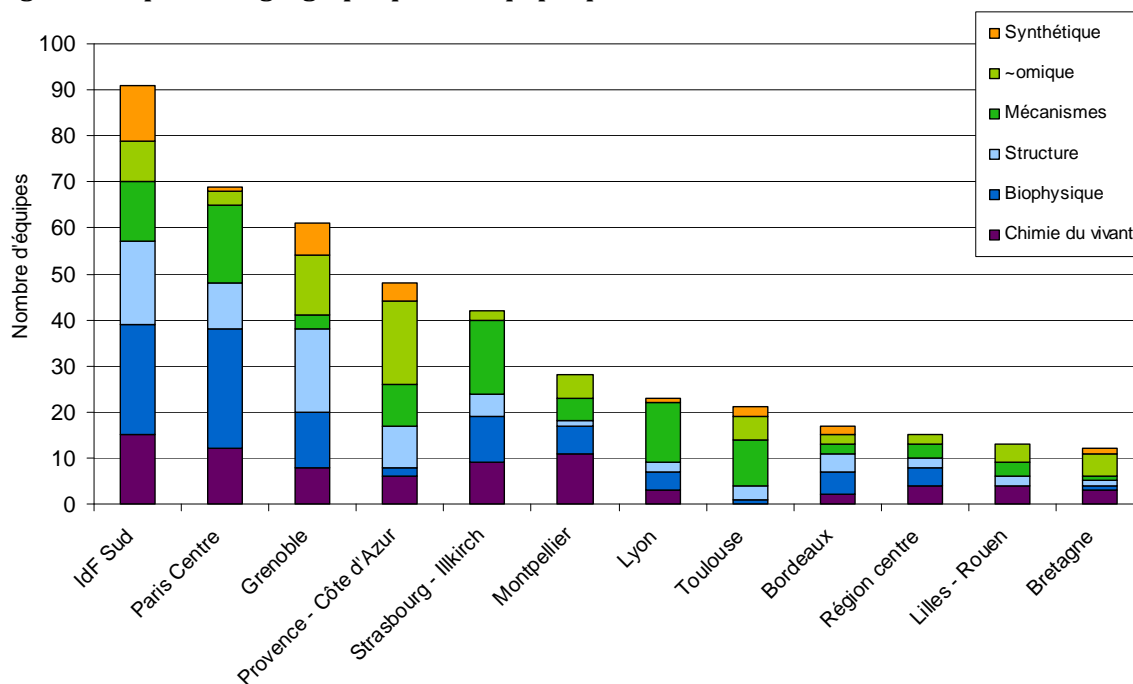


Cinq sites majeurs se distinguent dans le domaine BMSV :

- 1) Paris-Ile de France sud (33% des permanents)
- 2) Grenoble (15% des permanents)
- 3) Montpellier (11% des permanents)
- 4) Marseille-Cadarache (10% des permanents)
- 5) Strasbourg (9% des permanents)

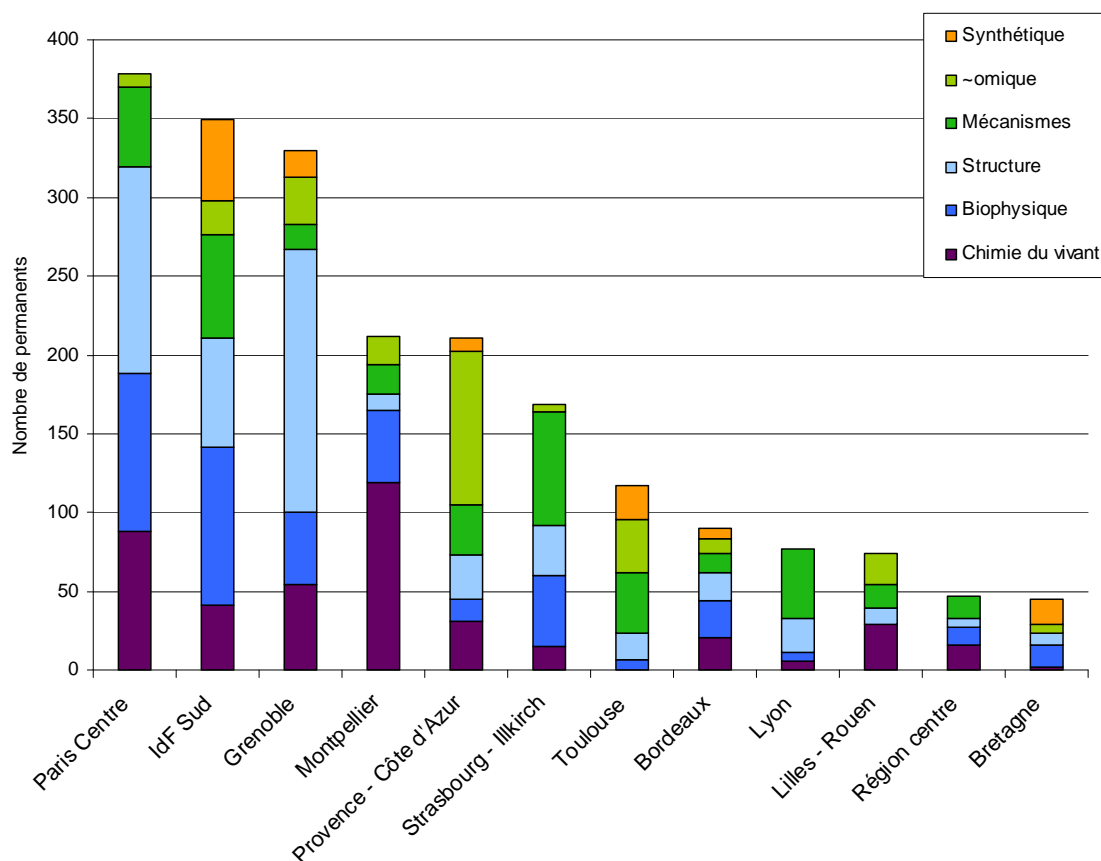
Globalement, on note un net déséquilibre vers le sud-est de la France de ces activités. On peut attribuer ce tropisme aux infrastructures européennes nécessaires pour le fonctionnement de la recherche (Grenoble, Strasbourg).

Figure 7 : répartition géographique des équipes par thème



- IdF Sud = Palaiseau - Saclay - Orsay - Bures Sur Yvette - Gif Sur Yvette - Cachan - Kremlin-Bicetre - Evry - Thiverval-Grignon - Jouy-En-Josas - Evry
- Provence - Côte d'Azur = Marseille - Valbonne - Nice - St Paul Les Durance - Marcoule - Cadarache

Figure 8 : répartition géographique des personnels par thème



Nombre de permanents	Paris Centre	Ile de France Sud	Grenoble	Montpellier	Provence - Côte d'Azur	Strasbourg - Illkirch	Toulouse	Bordeaux	Lyon	Lilles - Rouen	Région centre	Bretagne
Chimie du vivant	88	41	54	119	31	15	0	21	6	29	16	2
Biophysique	100	100	46	46	14	45	7	23	5	0	11	14
Structure	131	70	167	10	28	32	16	18	22	10	6	7
Mécanismes	51	65	16	19	32	72	39	12	44	15	14	0
~omique	8	22	30	18	97	5	34	9	0	20	0	6
Synthétique	0	51	17	0	9	0	21	7	0	0	0	16
TOTAL des thèmes	378	349	330	212	211	169	117	90	77	74	47	45

- IdF Sud = Palaiseau – Saclay – Orsay – Bures Sur Yvette - Gif Sur Yvette – Cachan - Kremlin-Bicetre – Evry
- Thiverval-Grignon - Jouy-En-Josas - Evry
- Provence - Côte d'Azur = Marseille – Valbonne – Nice - St Paul Les Durance - Marcoule - Cadarache

De ces données, les lignes de forces sont les suivantes :

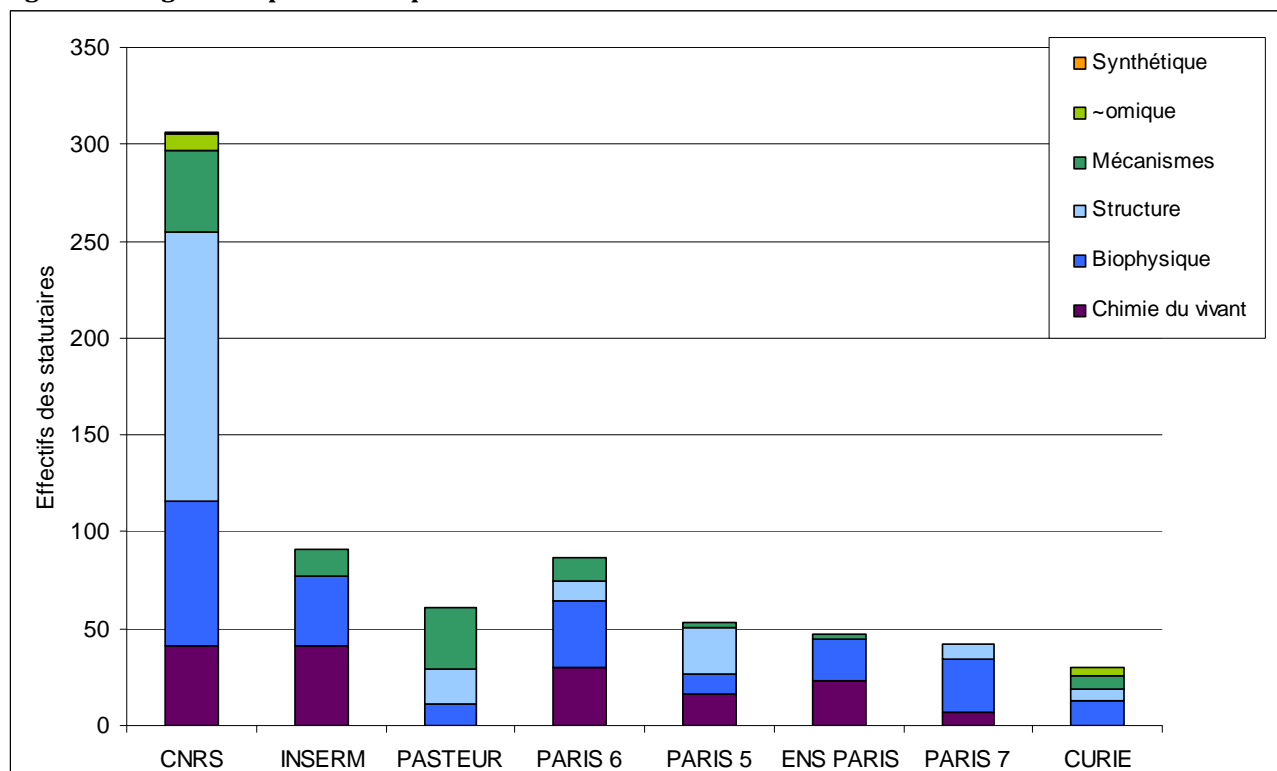
- Le plus gros site est Paris-Centre, fort en « biophysique » et « structure »
- Le sud de l'Île de France (Saclay) est le second site, très équilibré sur chaque thème, biophysique et mécanismes en tête
- Sur Montpellier la partie « Chimie » est dominante en BMSV
- A Grenoble, la partie « Structures » est très dominante et très localisée sur le polygone scientifique
- En région PACA, thème « -omique » est fort présent sur le centre CNRS de Joseph-Aiguier et au CEA à Cadarache (64 permanents sur 108)
- A Strasbourg, les thèmes « mécanismes » et « structure » prédominent
- A Lille-Rouen il ya une forte inclination sur les thèmes « mécanismes » et « ~omique »
- A Bordeaux et Lyon, deux grandes régions en biologie-chimie, les thèmes sont équilibrés mais la pénétration en BMSV est encore relativement faible.

Il faut noter que le thème « ~omique » est très visible sur Saclay, Toulouse, Grenoble, mais aussi en Bretagne et Région Centre.

2.1.2.2. L'Île de France

- Paris centre

Figure 9 : Origine des personnels permanents sur le site de Paris intra-muros



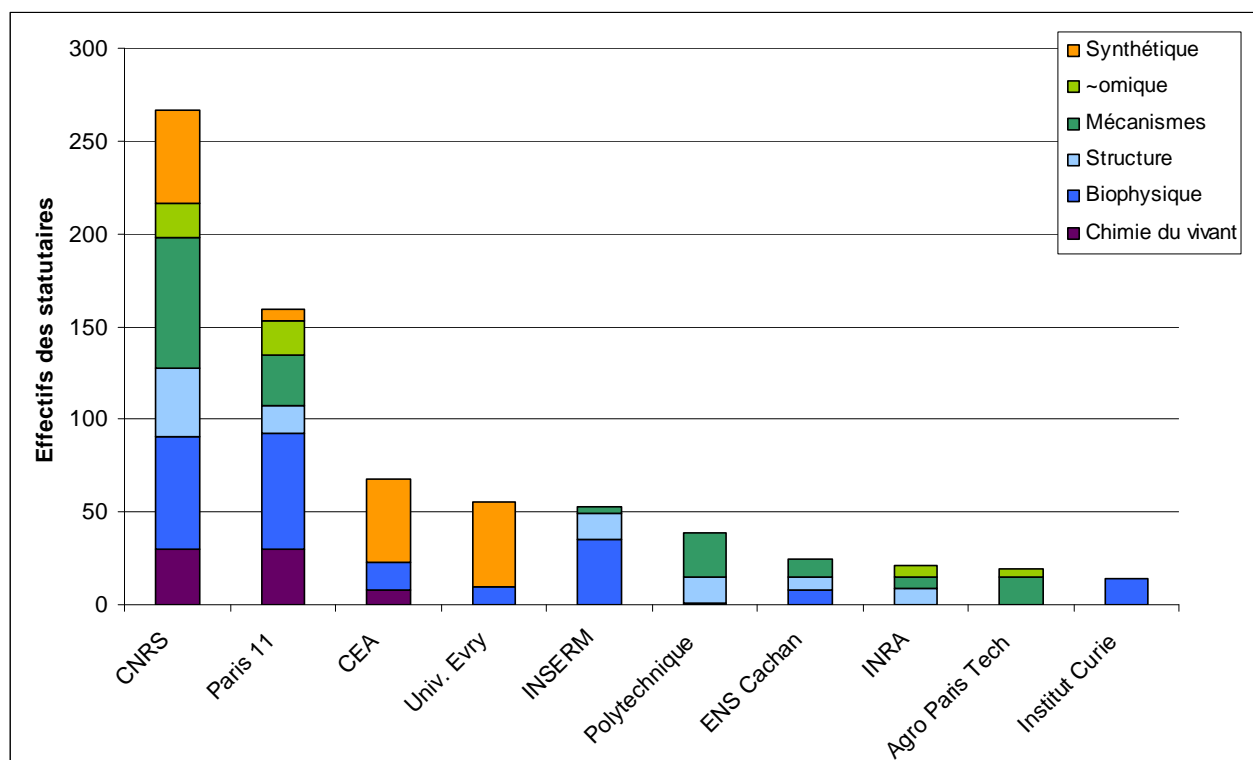
Sites parisiens	Enseignants chercheurs	Chercheurs CNRS	chercheurs Inserm	Chercheurs Pasteur	IATOS	ITA CNRS	ITA Inserm	ITA Pasteur	Total effectifs statutaires	Nombre d'équipes
Montagne Sainte Geneviève	29,5	21	2	0	30	3	4,6	0	90,1	31
Pasteur Necker	1	18	0	20	0	5	0	22	61	8
Paris 5	14,5	12	1	0	3,8	0	1	0	32,3	11
Paris 6, MNHN	15,5	6	3	0	3	1	1	0	29,5	7
Paris 7	11,25	0	7	0	0	0	5	0	23,25	3
Autres	1	3	0	0	1	3	0	0	8	1

Montagne Sainte Geneviève = Curie + ENSC + ENS + IBPC

Autres = ESPCI, Hôpital Saint Louis, IPGP, Paris rive gauche

- Ile de France Sud

Figure 10 : Origine des personnels permanents en Ile de France Sud

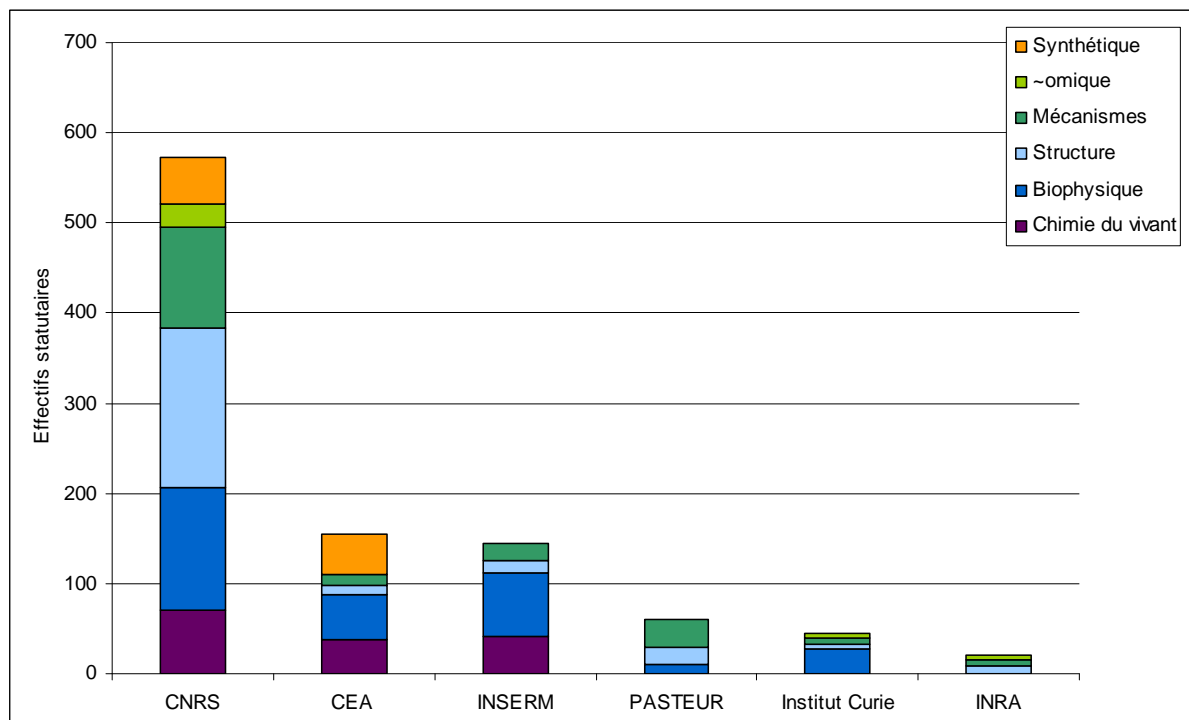


EFFECTIFS Statutaires	CNRS	Paris 11	CEA	Université Evry	INSERM	Polytechnique	ENS Cachan	INRA	Agro Paris Tech	Institut Curie
Chimie du vivant	30	30	8	0	0	0	0	0	0	0
Biophysique	61	62	15	10	35	1	8	0	0	14
Structure	37	15	0	0	14	14	7	9	0	0
Mécanismes	70	28	0	0	4	24	10	6	15	0
~omique	18	18	0	0	0	0	0	6	4	0
Synthétique	51	6	45	45	0	0	0	0	0	0
Totaux	267	159	68	55	53	39	25	21	19	14

Ecoles Paris Tech : Polytechnique, AgroParisTech, Telecom, Les Ponts, etc...

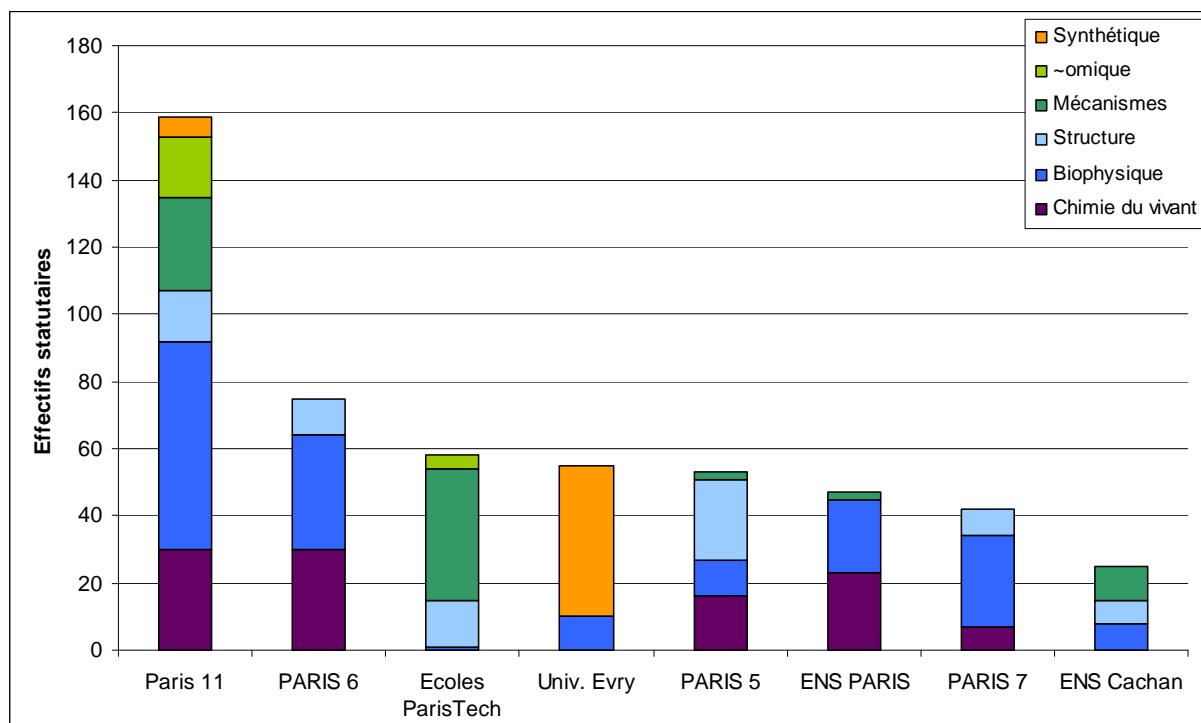
-Analyse globale de l'Ile de France

Figure 11 : Effectif permanents en Ile de France par organisme



EFFECTIFS Statutaires	CNRS	CEA	Inserm	PASTEUR	Institut Curie	INRA
Chimie du vivant	71	38	41	0	0	0
Biophysique	136	49	71	11	27	0
Structure	176	11	14	18	6	9
Mécanismes	112	12	18	32	7	6
~omique	26	0	0	0	4	6
Synthétique	52	45	0	0	0	0
Totaux	573	155	144	61	44	21

Figure 12 : Effectif permanents en Ile de France par université et grande école



EFFECTIFS Statutaires	Paris 11	Paris 6	Evry	Paris 5	ENS PARIS	Paris 7	ENS Cachan	ParisTech
Chimie du vivant	30	30	0	16	23	7	0	0
Biophysique	62	34	10	11	22	27	8	1
Structure	15	11	0	24	0	8	7	14
Mécanismes	28	0	0	2	2	0	10	39
~omique	18	0	0	0	0	0	0	4
Synthétique	6	0	45	0	0	0	0	0
Totaux	159	75	55	53	47	42	25	58

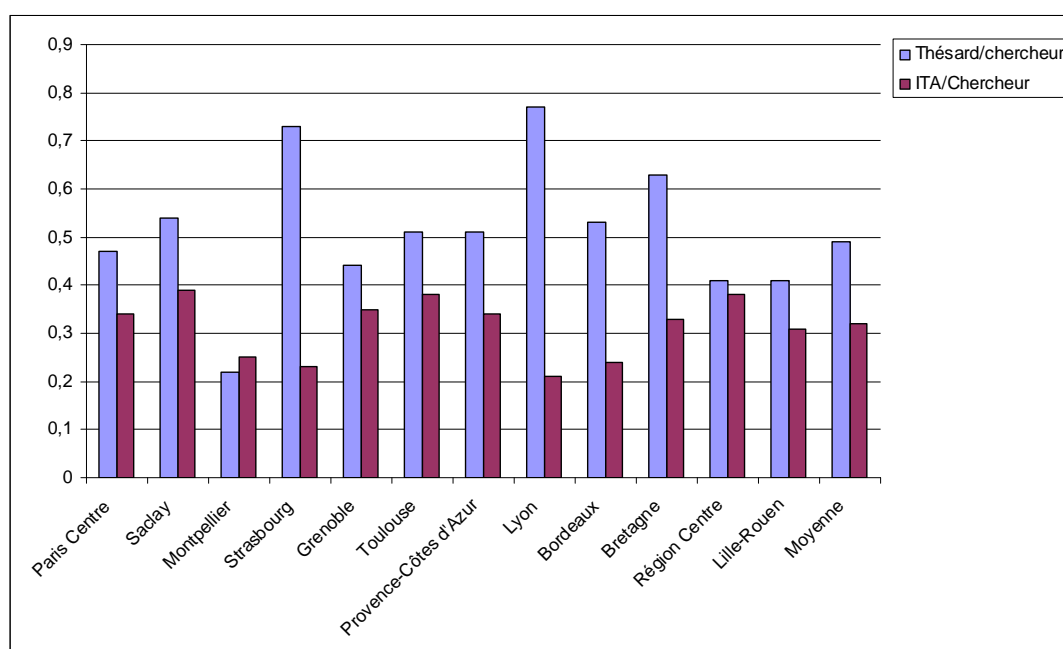
2.1.2.3. Variations par site

Tableau 5 : Variations par équipe et par site

Effectifs	Paris Centre	Ile de France Sud	Montpellier	Strasbourg	Grenoble	Toulouse	PACA	Lyon	Bordeaux	Bretagne	Région Centre	Lille-Rouen	Moyenne
Chercheurs/équipe	3,6	2,3	5,6	3,2	3,5	3,5	2,9	2,7	4,0	2,5	1,9	3,9	3,2
Permanents/équipe	5,5	3,8	7,6	4,1	5,4	5,6	4,4	3,3	5,3	3,8	3,1	5,7	4,8
Thésards/équipe	1,7	1,3	1,2	2,3	1,5	1,8	1,5	2,0	2,1	1,6	0,8	1,6	1,6
Thésard/chercheur	0,47	0,54	0,22	0,73	0,44	0,51	0,51	0,77	0,53	0,63	0,41	0,41	0,49
ITA/Chercheur	0,34	0,39	0,25	0,23	0,35	0,38	0,34	0,21	0,24	0,33	0,38	0,31	0,32

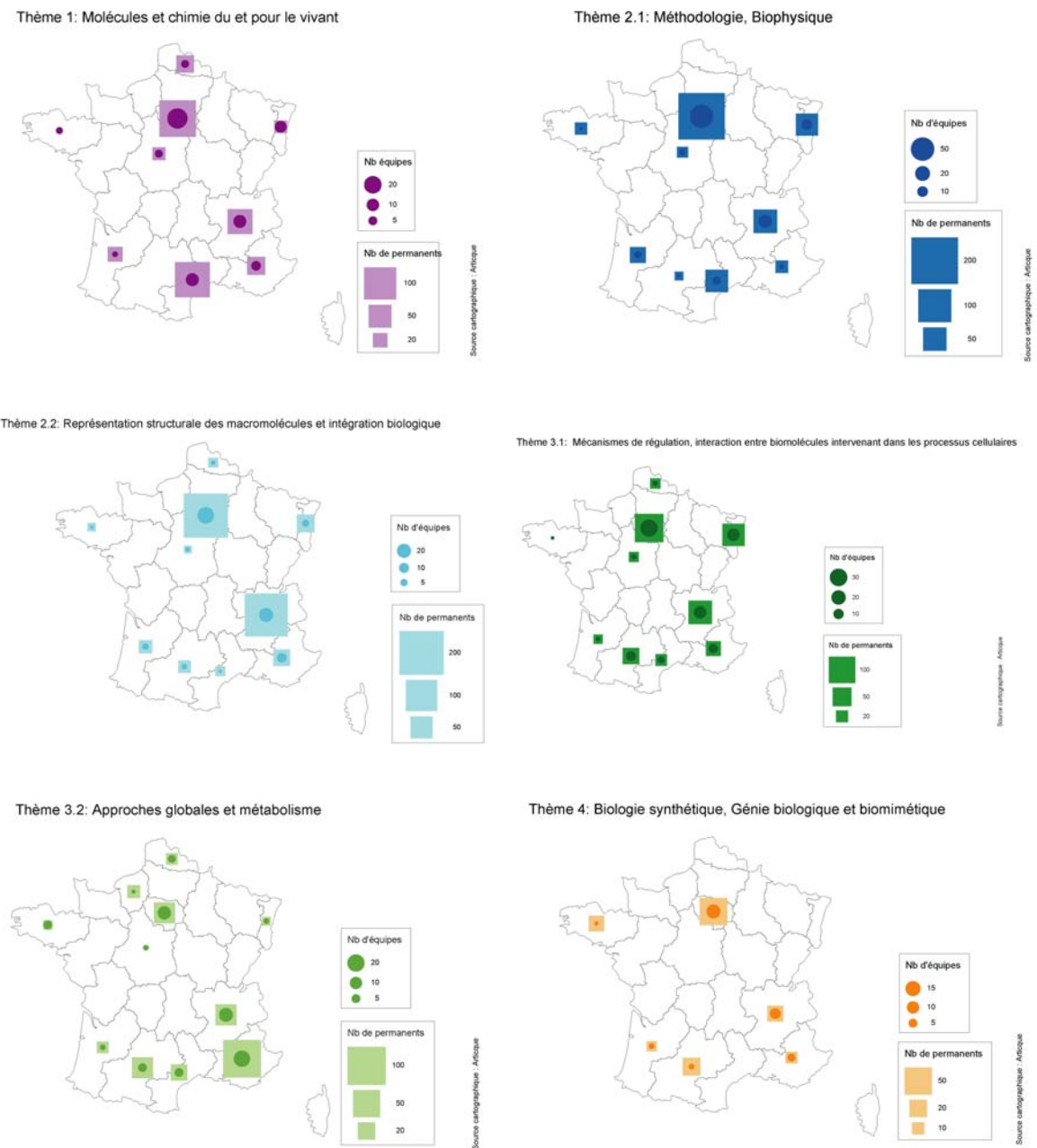
On observe de très petites équipes à Lyon, Région Centre et Saclay. Beaucoup d'ITA sont présents dans les équipes de Saclay, Toulouse, Grenoble, PACA et Paris-Centre. Ceci est en étroite relation avec la spécialisation (voir ci-avant). Moins d'ITA sont par contre présents à Lyon, Strasbourg et Montpellier. A noter qu'il semble y avoir très peu de thésards à Montpellier, et au contraire une très forte attractivité à Strasbourg et Lyon. Ceci pose le problème des écoles doctorales et de la nécessité au niveau national d'organiser une formation et des bourses dédiées financées par les organismes, CNRS en tête. Le CEA a déjà mis en place un programme de ce type.

Figure 13 : Thésards, ITA et chercheurs en BMSV



2.1.2.4. Cartographie des localisations d'équipe par thème

Figure 14 : Evaluation de l'importance des sites en nombre de personnels



2.2. Publications

2.2.1. Les revues de référence du domaine et leur impact

Tableau 6 : Les journaux de référence en BMSV et leur impact (pages suivantes)

Titre abrégé du journal ¹	Données issues du JCR 2010 ²					
	Nombre de citations	Impact Factor	Impact Factor à 5 ans	Immediacy Index	Nombre d'articles publiés	Demie vie de la citation
Journaux généralistes						
NATURE	443967	31.434	31.210	8.194	899	8.5
SCIENCE	409290	28.103	30.268	6.261	862	8.4
CELL	142064	31.253	30.149	6.126	348	8.8
NAT METHODS	5956	13.651	15.011	4.254	118	2.5
MOL CELL	33963	12.903	12.930	2.748	298	5.0
PLOS BIOL	12186	12.683	14.662	2.184	212	3.1
P NATL ACAD SCI USA	416018	9.380	10.228	1.635	3508	7.4
EMBO J	73804	8.295	8.904	2.100	291	8.5
EMBO REP	7823	7.099	7.265	1.683	142	4.4
NAT PROTOC	3366	4.170	4.175	1.321	156	2.0
Journaux spécialisés du domaine						
Thème 1						
<i>NAT REV DRUG DISCOV</i>	<i>10062</i>	<i>28.690</i>	<i>24.856</i>	<i>4.726</i>	<i>62</i>	<i>3.7</i>
NAT CHEM BIOL	3401	14.612	15.235	3.893	75	2.2
ANGEW CHEM INT EDIT	139534	10.879	11.025	2.657	1797	4.9
<i>MED RES REV</i>	<i>2426</i>	<i>8.907</i>	<i>8.558</i>	<i>2.935</i>	<i>31</i>	<i>5.8</i>
J AM CHEM SOC	318252	8.091	8.256	1.663	3242	7.3
<i>DRUG DISCOV TODAY</i>	<i>5570</i>	<i>6.618</i>	<i>7.432</i>	<i>0.788</i>	<i>132</i>	<i>3.</i>
CHEM BIOL	7423	5.603	5.743	1.272	125	5.7
ACS CHEM BIOL	867	5.149	5.149	1.351	77	1.8
J MED CHEM	45190	4.898	5.079	0.915	852	6.9
<i>CURR MED CHEM</i>	<i>7839</i>	<i>4.823</i>	<i>5.291</i>	<i>0.547</i>	<i>256</i>	<i>4.2</i>
<i>CURR TOP MED CHEM</i>	<i>3455</i>	<i>4.268</i>	<i>4.464</i>	<i>0.610</i>	<i>123</i>	<i>3.8</i>
<i>CURR OPIN DRUG DISC</i>	<i>1593</i>	<i>4.205</i>	<i>3.391</i>	<i>0.941</i>	<i>68</i>	<i>3.9</i>
CHEMBIOCHEM	5556	3.322	3.684	0.742	376	3.6
CHEMMEDCHEM	1100	3.150	3.150	0.645	200	1.8
BIOORGAN MED CHEM	14377	3.075	2.913	0.592	1015	3.9
BIOORG MED CHEM LETT	21446	2.531	2.419	0.667	1325	4.2
Thèmes 2.1, 2.2 & 3.1						
<i>NAT REV MOL CELL BIO</i>	<i>19628</i>	<i>35.423</i>	<i>34.221</i>	<i>7.238</i>	<i>84</i>	<i>4.0</i>
<i>ANNU REV BIOCHEM</i>	<i>16889</i>	<i>30.016</i>	<i>34.372</i>	<i>3.677</i>	<i>31</i>	<i>9.7</i>
<i>ANNU REV BIOPH BIOM</i>	<i>4500</i>	<i>17.049</i>	<i>16.667</i>	-	<i>0</i>	<i>8.5</i>
<i>NAT REV MICROBIOL</i>	<i>6095</i>	<i>14.310</i>	<i>16.894</i>	<i>3.939</i>	<i>82</i>	<i>3.2</i>
<i>TRENDS BIOCHEM SCI</i>	<i>14833</i>	<i>14.101</i>	<i>13.211</i>	<i>1.600</i>	<i>75</i>	<i>7.6</i>
<i>PROG LIPID RES</i>	<i>2741</i>	<i>11.237</i>	<i>13.358</i>	<i>1.538</i>	<i>26</i>	<i>6.4</i>
NAT STRUCT MOL BIOL	17436	10.987	10.799	2.764	165	6.1

<i>CURR OPIN STRUC BIOL</i>	8844	9.060	9.804	1.054	92	5.9
PLANT CELL	28874	9.296	10.458	1.572	229	6.3
NUCLEIC ACIDS RES	86787	6.878	6.968	1.635	1070	6.5
PLANT PHYSIOL	49108	6.110	6.650	1.246	509	8.1
J BIOL CHEM	407492	5.520	5.575	1.074	3761	7.6
STRUCTURE	10162	5.397	5.444	1.462	173	6.6
RNA	8344	5.018	5.228	1.177	254	4.6
BIOPHYS J	47035	4.683	5.036	1.090	1042	6.6
J LIPID RES	15572	4.409	4.361	1.142	260	7.7
BIOCHEM J	49310	4.371	4.251	1.290	462	>10.0
BIOINFORMATICS	30344	4.328	6.481	0.566	643	4.8
J MOL BIOL	67735	4.146	4.521	0.955	1054	8.7
BMC BIOINFORMATICS	8141	3.781	4.246	0.664	607	2.8
BIOCHEMISTRY-US	94645	3.379	3.402	0.649	1437	9.7
PHYS REV B	250465	3.322	3.284	0.885	5782	8.3
FEBS LETT	53419	3.264	3.261	0.647	660	8.
FEBS J	5609	3.139	3.421	0.520	517	2.6
ANAL BIOCHEM	36555	3.088	3.145	0.723	433	>10.0
BIOCHIMIE	5384	3.071	3.168	1.140	172	6.3
ACTA CRYSTALLOGR D	9365	2.943	1.827	0.581	148	9.7
Thème 3.2						
MOL CELL PROTEOMICS	7937	8.834	9.391	1.766	188	3.2
ANAL CHEM	77792	5.712	5.918	0.831	1221	7.3
J PROTEOME RES	8043	5.684	5.737	0.980	509	2.5
PROTEOMICS	12441	4.586	5.191	0.677	434	3.5
Thème 4						
NAT BIOTECHNOL	28188	22.297	25.329	7.730	100	5.3
NAT NANOTECHNOL	2927	20.571	20.588	5.097	93	1.6
MOL SYST BIOL	1741	12.243	12.633	1.815	65	2.2
MOL BIOSYST	846	4.236	4.527	0.721	122	2.2
LANGMUIR	76864	4.097	4.347	0.575	2026	5.9
J THEOR BIOL	12876	2.454	2.490	0.351	402	>10.0
Journaux thématiques multidisciplinaires						
ONCOGENE	56007	7.216	6.729	1.419	721	5.6
MOL CELL BIOL	69445	5.942	6.426	1.510	621	7.3
J VIROL	86021	5.308	5.135	1.183	1293	6.8
MOL MICROBIOL	31290	5.213	5.436	1.131	442	6.5
APPL ENVIRON MICROB	69589	3.801	4.535	0.570	1006	8.3
J BACTERIOL	58715	3.636	3.748	0.892	909	9.6
VIROLOGY	25469	3.539	3.278	0.711	558	9.0

¹ Les journaux de revue de synthèse sont indiqués en italique ; ² Base JCR 2008, Thomson

Comparativement à d'autres domaines comme la génétique, la biologie cellulaire, les neurosciences, le développement ou l'immunologie, il faut noter la valeur faible relative de tous les facteurs d'impact (IF) généralement situés entre 3 et 7 (moyenne 4,5). Il s'agit d'un phénomène récent avec resserrement des « valeurs » depuis quelques années avec des revues phares comme Journal of Biological Chemistry ou Biochemistry ayant des IF proches des journaux de moindre envergure comme ceux publiés par la FEBS ou la SFBBM (Biochimie).

2.2.2. L'impact des publications françaises en BMSV

Dans cette partie, il s'agissait d'évaluer l'impact des équipes françaises du domaine BMSV. Des difficultés méthodologiques importantes ont été rencontrées par la cellule en charge de ces aspects. Il s'avère en effet que les mots clés qui ont été progressivement incrémentés pour définir le domaine (voir annexe) ne permettent de couvrir que les deux tiers des publications émanant des >400 responsables d'équipe identifiés ci-avant. Ceci indique la difficulté à cerner cette communauté traditionnellement amenée à publier dans des journaux généralistes et d'autres considérés comme plus spécialisés (anti-infectieux, cancer, génétique etc...). Avec ces réserves, nous avons utilisé ces mots-clés pour évaluer le domaine.

Tableau 7 : Journaux du domaine BMSV représentés dans le Top 1% des citations

Journal	Nombre d'articles	Représentativité
Nature	14	7,7%
Science	10	5,5%
Plant Cell	9	5,0%
P Natl Acad Sci USA	7	3,9%
Cell	5	2,8%
J Am Chem Soc	5	2,8%
Nat Struct Mol Biol	4	2,2%
Biochimie	3	1,7%
Plant Physiol	3	1,7%
Total	60	33%

¹ Base 2006-2007, 181 articles retenus pour BMSV

Il existe effectivement pour le corpus national une forte corrélation avec les journaux de fort impact mondial. Il faut noter le cas de la biologie végétale très prégnante et visible en France dans ce domaine BMSV. La majorité des articles les plus cités (les deux tiers) sont publiés dans des journaux de tous les types. Noter le cas particulier de la présence de « Biochimie », dû à un biais de la base de données basée en partie sur le nom des chercheurs pour identifier le domaine. Ceci signifie que ce journal reste une référence en France, sans doute surtout pour les articles de synthèse.

Tableau 8 : Journaux les plus pertinents identifiés dans le Top 20%

Journal	Nombre d'articles cités	Nombre d'articles parus	Pertinence ¹	IF ₅
EMBO J	24	291	8,2%	8,90
Acta Crystallography D	10	148	6,8%	1,83
Mol Cell Proteomics	7	188	3,7%	9,39
Proteomics	12	434	2,8%	5,19
Mol Microbiol	10	442	2,3%	5,44
J Med Chem	18	852	2,1%	5,08
J Biol Chem	78	3761	2,1%	5,58
Nucleic Acids Res	21	1070	2,0%	6,70
Biochemistry-US	24	1437	1,7%	3,40
Oncogene	12	721	1,7%	6,73
J Bacteriol	15	909	1,7%	3,75
Mol Cell Biol	10	621	1,6%	6,43
J Mol Biol	16	1054	1,5%	4,52
Biochem. J	7	462	1,5%	4,25
Biophys J	15	1042	1,4%	5,04
J Virol	16	1293	1,2%	5,14
Appl Environ Microb	12	1006	1,2%	4,54
P Natl Acad Sci USA	41	3508	1,2%	-
J Am Chem Soc	37	3242	1,1%	-
Cell	3	348	0,9%	-
Anal Chem	8	1221	0,7%	5,92
Langmuir	13	2026	0,6%	4,35
Plos One	15	4414	0,3%	-
Angewandte Chemie	6	1797	0,3%	11,03
Total	430	32287	1,3%	5,67
Corpus	430	1572	27%	

¹ nombre d'articles / nombre d'articles publiés

Les journaux d'excellence du domaine sont Nature, Science, Plant Cell, P Natl Acad Sci USA, Cell, J Am Chem Soc, Nat Struct Mol Bio. En France, les journaux de référence de très haut niveau du domaine sont:

1. *Chimie*: J. Med. Chem., Angewandte Chemie

2.1 *Biophysique*: Biophys J.

2.2 *Structure*: Biochemistry, J. Mol. Biol. Acta Crystallography D

3.1 *Mécanismes*: EMBO J., Mol. Microbiol., J. Bacteriol., J. Virol, Mol Cell Biol, Appl Environ Microb, Oncogene, Nucleic Acids Res, Genes and Development, Molecular Cell, PloSPathogen

3.2 *~omique* : Mol Cell Proteomics, Proteomics, Anal Chem, J of Proteome Research

4. *Synthétique* : pas de masse critique, difficile de conclure.

Acta Crystallography D publie des papiers de référence des méthodes liées aux synchrotrons situés sur le sol français, haut débit et traitements de données cristallographiques et reste une référence dans le domaine Apparu de puis peu, Plos One est devenu le journal publiant le plus d'articles mais il est encore difficile d'en juger l'impact exact en BMSV.

2.2.3. Conclusions

Il existe une excellente adéquation avec les meilleurs journaux du domaine et ceux qui donnent une grande visibilité au secteur français, ce qui justifie la place de la France dans le domaine (voir plus loin). Les publications dans ces journaux doivent être considérées, malgré leur faible impact apparent (journal de haut niveau type J. Biol. Chem.). La communauté fait des choix en préférant publier chez les éditeurs « européens » lorsqu'il existe des « doublons » dans le domaine de prédilection d'un journal (Nucleic Acids Res./RNA; Proteomics/J. Proteome Res.; EMBO J./Mol. Cell). Incontestablement, il existe une grande fidélité des auteurs français aux journaux publiés sous l'égide des sociétés savantes américaines, comme l'ASM, ASBMB et ACS. Microbiologie (virologie-bactériologie) et cancer sont des domaines de forte implication de la communauté, ce qui implique un croisement avec les Itmo « Microbiologie et Maladies Infectieuses » et « Cancer ». La protéomique et la chimie médicinale sont des sous-domaines très visibles en France. On observe des « trous » dans des domaines d'émergence, comme la biologie synthétique et la « Chemical Biology ». Dans le domaine BMSV c'est donc plus le nombre et la diversité des journaux visés par les auteurs que le facteur d'impact respectif de chaque publication qui compte pour une évaluation pertinente de l'activité de la majorité des équipes.

2.2.4. Analyse quantitative et géographique

2.2.4.1. Aspects quantitatifs

Tableau 9 : Nombre de publication du domaine BMSV comparé aux principaux pays

Pays	Nb pub	Total citations ¹	Indice de citations moyen	h-index ²	Population	Nbr/pop	Budget biologie-santé (milliards d'€) ³	Rang/population
USA ⁴	80101	nc	nc	nc	307 212 123	0,26	47	6
Royaume-Uni	14001	209615	14,97	119	61 113 205	0,23	7,6	9
Allemagne	19700	256822	13,04	118	82 329 758	0,24	6,3	8
France	12754	151862	11,90	98	64 057 792	0,20	4,5 ⁵	10
Japon	18055	168041	9,30	94	127 078 679	0,14		13
Canada	9435	119325	12,65	92	33 487 208	0,28		5
Suisse	4551	76341	16,77	91	7 604 467	0,60		1
Italie	9884	109490	11,08	86	58 126 212	0,17		11
Pays-Bas	4787	71256	14,89	83	16 715 999	0,29		4
Australie	5251	64641	12,31	75	21 266 767	0,25		7
Chine	18127	121135	6,68	70	1 360 612 968	0,01		15
Danemark	2343	34553	14,75	67	5 484 723	0,43		3
Suède	4203	54300	12,92	66	9 059 651	0,46		2
Espagne	7009	72119	10,29	65	46 063 511	0,15		12
Corée du Sud	5005	38905	7,77	54	48 422 644	0,10		14

1 années 2006 et 2007

2 mesure du nombre « n » d'articles étant cité au moins n fois. Traduit à la fois le nombre et l'impact des articles. Très pertinent sur des grandes masses.

3 Les Echos, 26 Mai 2010 ; aucun lien avec données OST 2008 confirmé par G. Filliatreau, directrice)

4 les statistiques pour les USA n'ont pu être établies avec précision du fait du très grand nombre de publications (>50 000).

5 Le rapport FutuRIS -ANRT indique 2,7 M€ (voir annexe &9.2)

La mesure du h-index permet de tenir compte à la fois du nombre et de la visibilité relative de chaque communauté. Il s'agit d'une mesure pertinente dans le domaine BMSV. La France se situe au 4^{ème} rang mondial, devant le Japon. Si l'on tient compte de l'impact par rapport à la taille de populations, la France est 10^{ème}, loin derrière les pays du Nord de l'Europe (Suède, Danemark, Pays-Bas) et la Suisse, le Canada et l'Australie mais au même niveau que l'Allemagne et le Royaume-Uni. Dans le domaine BMSV, la France est donc une nation très performante.

2.2.4.2. Répartition géographique des publications

Figure 15 : carte des publications BMSV émanant d'adresses mentionnant une équipe française

Publications 2006-2007

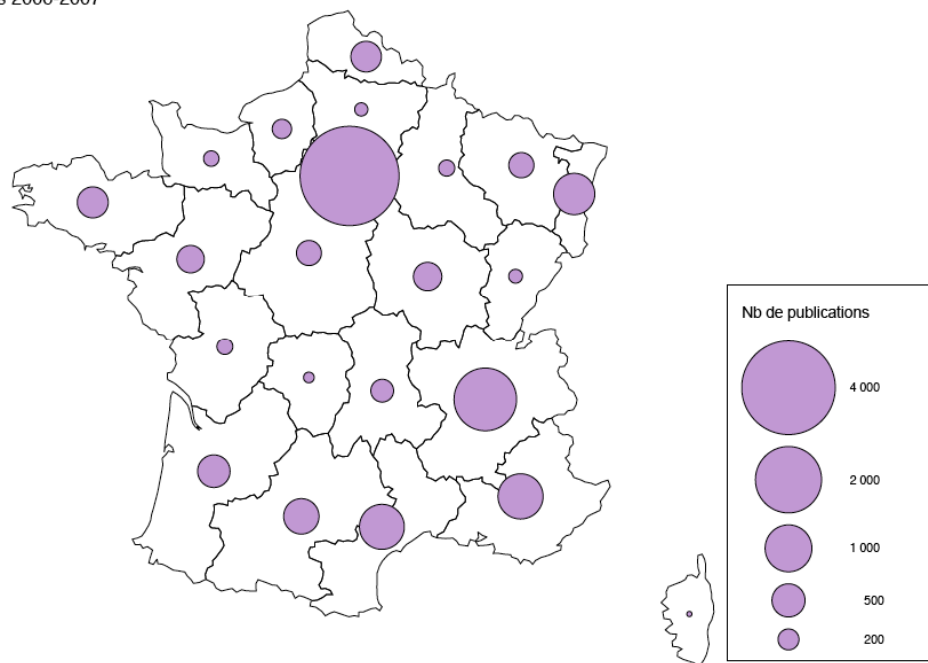


Tableau 10 : nombre d'articles/permanent par région

Ile de France	Grenoble-Lyon	Montpellier	Provence - Côte d'Azur	Strasbourg	Toulouse	Bordeaux
1,27	0,90	0,88	0,90	0,94	1,00	1,10

* le calcul tient compte de la somme des Chercheurs+EC+ITA ; période 2006-2007

La situation est très homogène en France. Cependant, la région parisienne présente une plus forte activité en termes de volume. Noter que le calcul ne tient pas compte des ETP mais du nombre d'EC. Des sites plus riches en EC en BMSV comme à Montpellier peuvent être moins bien caractérisés de ce fait. Ce type de données devrait être corrigé par l'impact des articles.

2.2.4.3. Conclusions et liens avec les opérations campus

La région parisienne démontre la plus grande visibilité en BMSV. Cependant, sa cohérence et sa visibilité sont à renforcer. Pour ce faire des actions régionales (création d'un DIM dédié) et l'opportunité des opérations (Campus de Saclay et Paris-Centre) sont à conjuguer. Les opérateurs d'excellence (Institut Curie, Institut Pasteur) déjà engagés dans le domaine devront contribuer dans ce cadre à mettre en place des programmes spécifiques dédiés (virologie structurale/bactériologie-résistance/cancer-cytosquelette) adaptés pour permettre à leur communauté BMSV de rivaliser dans le futur au niveau international.

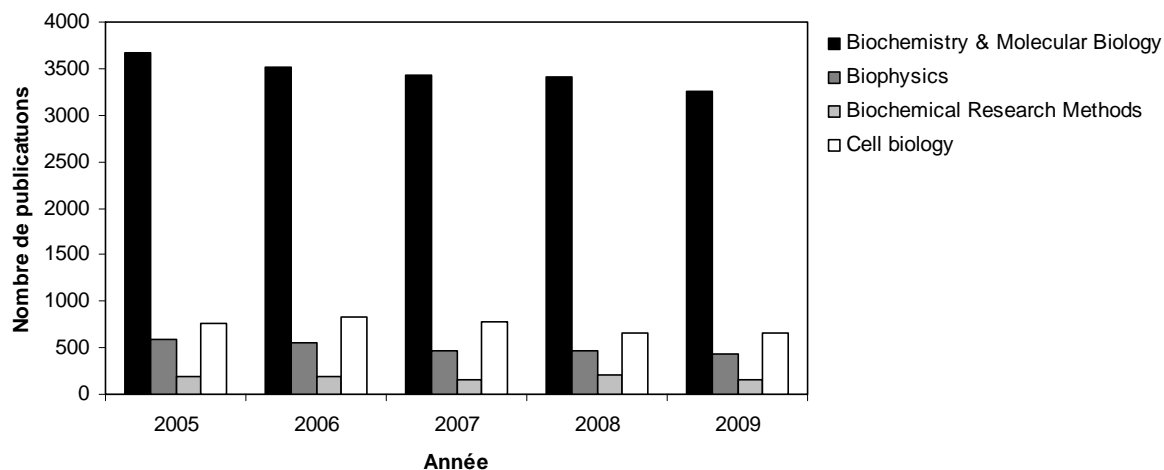
Tableau 11 : Données analytiques des publications françaises en BMSV*

Nb Pub Total	Total Cites	ICm	IFm	Nb Pub Top1%	Nb Pub Top10%	Nb Pub Top20%	Nb Pub Top50%	Nb Pub sous la médiane
10367	125460	12,10	4,57	181	1680	1572	3471	3464

* Données 2006-2007

Figure 16: Evolution du nombre de publication en BMSV

Etabli d'après le Web of Knowledge (Wok) 2010



On observe une lente érosion du nombre de publication mais il est difficile de donner du sens à cette analyse car il peut tout aussi bien s'agir de moins d'article dans des journaux de plus fort impact.

2.3. Financements du domaine BMSV

Ancré sur des approches essentiellement cognitives, ce qui l'oppose à la majorité des Itmos plus liés aux pathologies, le financement du domaine BMSV dépend très largement des financements publics (européens, nationaux et régionaux) et des fondations et associations caritatives. Au niveau national, outre les établissements de recherche, organismes et universités, les agences de financement sur projets ont une part essentielle au soutien des équipes, avec principalement l'ANR et dans une moindre mesure pour le domaine BMSV, l'ANRS et l'INCa. Concernant les financements régionaux, il est difficile d'obtenir une vision précise mais ils contribuent largement aux infrastructures.

Les contrats industriels concernent aujourd'hui essentiellement les équipes des domaines de la chimie du vivant, avec la conception de candidats médicaments ou d'outils de recherche. Le domaine contrôle et modélisation du vivant, devrait aussi naturellement déboucher sur des applications industrielles, avec la caractérisation de cibles ou de marqueurs ; de même la biologie synthétique devrait pouvoir intéresser des industriels. Cet aspect mérite une analyse détaillée qui ne sera pas présentée dans cette version du document.

S'il est difficile d'obtenir une vision consolidée des financements régionaux qui contribuent beaucoup aux infrastructures, il est certain que le financement par les fonds caritatifs ne contribue que marginalement à l'activité des équipes BMSV, même si certaines associations sont des partenaires

importants, comme l'ANRS dans le domaine de l'ARN ou l'ARC qui soutient les projets plus orientés « cancer » ou la FRM pour des projets en infectiologie. Les régions disposent depuis quelques temps de budgets conséquents alloués au soutien de la recherche mais leurs politiques sont différentes. Ces budgets permettent le soutien d'achats d'équipements, de bourses de thèses ou de chercheur postdoctoral. Par exemple, au regard des populations de chercheurs concernés et de leur visibilité incontestable, la région Ile de France – qui n'affiche aucun « Domaine d'Intérêt Majeur » proche de la communauté BMSV - finance peu de projets dans le domaine. Au contraire d'autres régions (PACA, Aquitaine, Rhône-Alpes), présentent une vraie dynamique dans ce domaine.

Seront détaillées et commentées dans la suite les principales sources de financement du domaine BMSV. (ANR et en amont PEPS et PIR et Jeunes chercheurs : ATIP-AVENIR et ANR Jeunes Chercheurs).

Tableau 12 : Budgets en BMSV, salaires

Effectifs par domaine BMSV	Chercheurs & Enseignants-Chercheurs	Ingénieurs & techniciens	Doctorants	CDD - post-doctorants	Totaux
Nombre*	1507	777	732	594	3610
Cout/chercheur (k€/an)	80	60	30	45	-
Cout total (M€/an)	121	46	22	27	220

* personnes physique

Pour BMSV, une projection suggère un coût salarial de 220 M€/an dont une partie prise en charge par le programme ; on peut considérer que les CDD représentent 70% du cout d'un projet ANR. A l'ANR, un projet représente en moyenne 600k€/projet/3ans et 38 projets/ans relèvent dans le domaine BMSV (8 M€/an). Au CNRS (le principal organisme concerné), le budget consacré en BMSV avoisine 12 M€. Les programmes collectifs (voir chapitre dédié) contribuent à hauteur d'environ 4 M€, auquel il faut ajouter la part du programme ministériel IBISA (4-6 M€/an sur les 13 contribuent au domaine BMSV). Avec la participation des autres établissements, on peut avancer un chiffre entre 250 et 300 M€ par an qu'il convient de rapprocher des 2,7 Milliards dédiés (voir Tableau 3 du rapport BIOMED Futuris-ANRT publié en mai 2008). Le cout salarial étant généralement considéré représenter 80-85% du cout consolidé d'un chercheur, ceci conforte cette analyse encore imprécise. Ceci indique que le domaine BMSV représente 10% des financements totaux pour une communauté représentant 20 % des chercheurs. On peut considérer dès lors que ce type de domaine est peu onéreux ou sous-financé par rapport à d'autres.

L'infrastructure en 2010 sur la base de 43 000 m2 bâtis (5E)

51 m2/personne, 50€/m2

Estimation sur un total de l'accueil de 850 personnes (25%)

Soit 2550 €/personnes :Pour 3610 personnes, 9,2 M€ de coût d'infrastructure

2.3.1. L'ANR

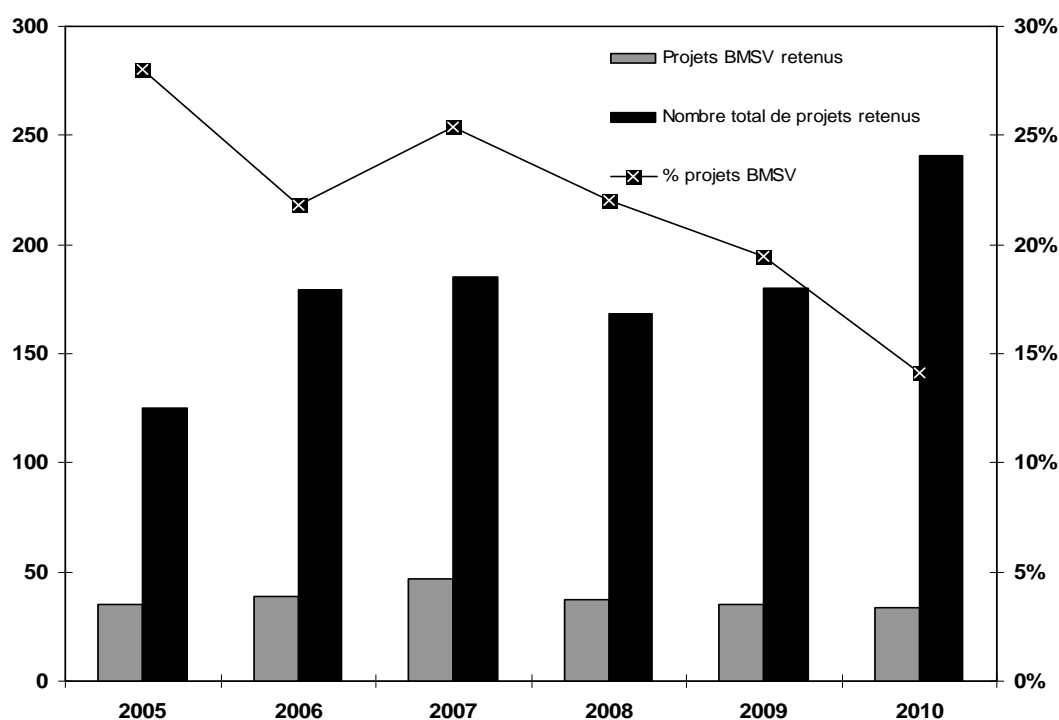
Depuis sa création en 2005, il est certain que, dans le domaine BMSV, l'Agence National de la Recherche (ANR) est devenue un des principaux financeurs, à côté des organismes et établissements de recherche. Depuis 2007, l'ANR finance des projets à hauteur de 200 M€/an en biologie-santé, incluant des appels à projets thématiques et le programme blanc et 110 M€ en chimie, les deux sous-thèmes qui correspondent le mieux au domaine BMSV). Le domaine Biologie-Santé représente environ 20 à 25% des actions de l'ANR. Les appels d'offre sont annuels.

Tableau 13 : Budgets pour des projets financés en 2007 pour la période 2007-10 dans le programme blanc

2007	Programme blanc (k€)	Programme blanc (%)	Programme Jeune chercheur(se) (k€)	Programme Jeune chercheur(se) (%)
Chimie	17562	14%	3545	15%
Agronomie-Ecologie	13543	11%	2330	10%
Biologie-santé	26263	21%	5583	24%
Total	124518	100%	23158	100%

Une étude détaillée des projets sélectionnés par l'ANR depuis sa création a été entreprise pour mesurer l'influence du domaine dans la recherche française (Figure 17). Ont été analysés, les projets financés dans les programmes blancs (biologie-santé, chimie, interdisciplinaires) et les appels d'offres thématiques, PCV (en 2006, 2007 et 2008) puis PIRibio (en 2009) ; l'identification du domaine BMSV a été réalisée par croisement de la nature du titre et des noms et origines des porteurs.

Figure 17 : Part des projets BMSV sélectionnés par l'ANR



Regroupe les projets sélectionnés dans le cadre des programmes blancs (chimie, biologie-santé et interdisciplinaires) et des programmes thématiques PCV 2006-2007-2008 et PiriBio 2009, fusionné au programme blanc en 2010.

De 2005 à 2010, une moyenne de 21 %, soit 38 projets en BMSV sur un total de 191 projets financés, portent chaque année sur un thème explicitement relatif au domaine BMSV. Globalement, ce chiffre reflète bien la taille de la communauté (20-22%, voir Tableau 3). Cependant, le nombre de projets sélectionnés est constant, malgré le doublement du nombre de projets financés depuis 2005. On observe donc un recul progressif et significatif – d'un facteur 2 en 6 ans – du soutien du domaine BMSV par l'ANR. Avec un changement des modes d'évaluation, 2010 apparaît comme l'année la plus noire avec seulement 14,1% de projets en BMSV financés, ce qui mérite quelques commentaires.

Depuis 2006, dans les programmes PCV/PIRiBio, 50% de projets retenus relevaient du domaine BMSV. Il en est de même en 2010 pour le sous-comité SVSE5 qui incluait le même domaine thématique. Néanmoins, il apparaît que les autres comités thématiques du programme blanc biologie-santé financent de moins en moins les projets en BMSV. En effet, on est passé de 27 projets financés en 2005, à 15 en 2007, 9 en 2009, puis enfin 6 en 2010, alors que le nombre total de projets financés a doublé. Parmi les projets, les comités « chimie » ont sélectionné depuis 2005 de 7 à 18% de projets en BMSV (moyenne 12%, soit 5 nouveaux projets/an ; en 2010, comités SIMI3 et surtout SIMI7). Ce nombre semble très faible au regard de la large communauté concernée (Thème 1). Les autres comités des programmes blancs comme Agronomie-Ecologie (SVSE7) ne sélectionnent virtuellement aucun projet de ce type, ce qui est très surprenant. En conclusion, on observe donc une très forte diminution du soutien aux projets avec un positionnement clair en BMSV (c'est à dire, titre ou porteur) par l'ANR, plus de 25% en 2005 et moins de 15% en 2010. Il est important de souligner que l'analyse présentée ici n'inclut pas le soutien au domaine BMSV via des équipes qui participent en tant que partenaires à des projets qui globalement peuvent

émarger dans un autre domaine principal, car ce type de données n'est aujourd'hui pas rendu public. Ce type de participation des équipes en BMSV pourrait s'expliquer par une tendance à aller vers une biologie plus intégrée et ne reflèterait pas un déclin du domaine mais seulement un affichage moins affirmé. Une autre explication au moindre soutien du domaine BMSV pourrait être la difficulté à évaluer des projets interdisciplinaires comparativement à d'autres et donc au final un moindre taux de sélection.

Pour résumer, ces données sont importantes pour le domaine BMSV et son devenir ; il est donc nécessaire d'identifier les causes réelles de ce constat pour permettre à la communauté de relever les défis scientifiques mentionnés plus avant. Il faudrait évaluer la prise en compte de mieux en mieux valorisée des approches de biologie intégrée à l'échelle cellulaire ou multicellulaire aux dépens de celles de biologie purement moléculaire ou structurale.

2.3.2. Les financements de l'émergence : PEPS et PIR du CNRS

Ces financements concernent des programmes du CNRS destinés à financer l'émergence et à donner les moyens d'obtenir quelques validations nécessaires pour une demande à l'ANR.

2.3.2.1. PEPS

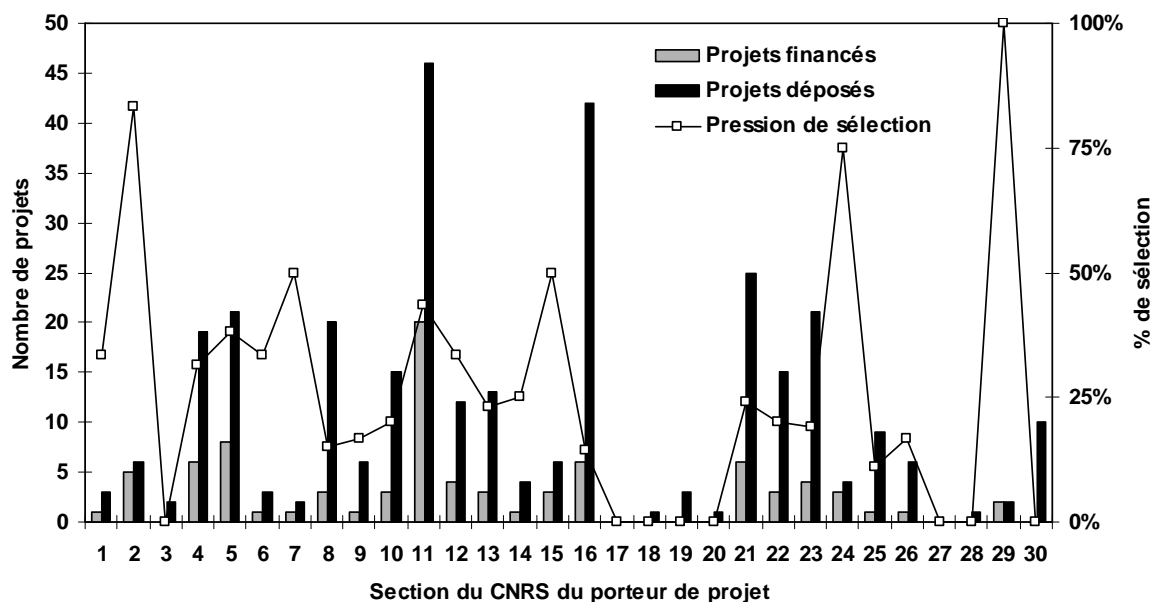
Les PEPS correspondent à des programmes exploratoires mis en place en 2008 par l'INSB du CNRS avec un financement de 15 k€ par projet sur un mode rapide de sélection (budget 700 k€/an). L'un des 5 thèmes identifiés et qui correspond au domaine BMSV est Biochimie/Biologie Moléculaire/Pharmacologie. Dans ce cadre, 28 projets du domaine (9 à 10 par an) ont été financés sur un total de 117 soit 24% et pour 624 demandes déposées. Comme le taux de sélection est identique dans chaque thème (environ 20%), il y a donc beaucoup plus proportionnellement de projets déposés dans ce cadre que dans d'autres. Ceci traduit un défaut de financement par ailleurs et/ou le constat que les projets dans ce domaine sont risqués.

2.3.2.2. PIR

Les programmes PIR du CNRS correspondent à des financements d'environ 50 k€ pour des projets en interdisciplinarité (budget total en 2010 de 16 M€/an). Parmi ces programmes, le programme Physique/Chimie du Vivant (aide à la prise de risque) concerne la communauté BMSV. Les projets impliquent généralement deux équipes, l'une de biologistes, l'autre de chimistes ou de physiciens (budget 2,5 M€/an). Comme les PEPS, soumission, évaluation et réponses sont rapides. Le domaine BMSV est principalement présent dans les sections, 5, 11, 12, 16, 21, 23, 30. L'observation des intitulés projets retenus (à défaut de celui des projets présentés) montre une très forte orientation en biologie cellulaire/neurosciences ou nanotechnologies. On remarque que les porteurs issus des sections 16, 21 et 23, représentent 28% des porteurs de projets mais sont moins bien financés avec une moyenne de 18% contre une moyenne générale de 27%. Les porteurs issus de la section 11, avec un taux de succès de 44% représente 25 % des récipiendaires sur des projets souvent en lien avec l'imagerie cellulaire. Dans ce type de programmes une des difficultés réside dans le choix des comités de programme à couvrir de façon

équilibrée les différentes disciplines et ne pas biaiser la sélection des projets. Pour le domaine BMSV, il apparaît que ce programme ne joue pas le rôle dynamisant qu'il devrait jouer.

Figure 18 : Le domaine BMSV dans les programmes PIR du CNRS (période 2007-2010)



2.3.3. L'émergence des jeunes chercheurs et équipes

2.3.3.1. Recrutements

Il est possible de résumer la situation comme suit :

- CNRS :

S16 : 3-4 recrutements dans le domaine BMSV

S21 : 6 recrutements/an dans le domaine BMSV

S23 : 2 recrutements/an dans le domaine BMSV

Une dizaine de postes de nouveaux postes de chercheurs sont mis au concours dans le domaine chaque année. Globalement, il y a maintien du nombre de postes depuis 5 ans.

- Inserm concerne un à deux postes en CSS2 et CSS7

- CEA, aucune information

- Dans les universités, il est difficile de disposer d'une vision consolidée générale des recrutements ; les postes sont généralement issus de la 64e section (mais aussi 28e et 32e).

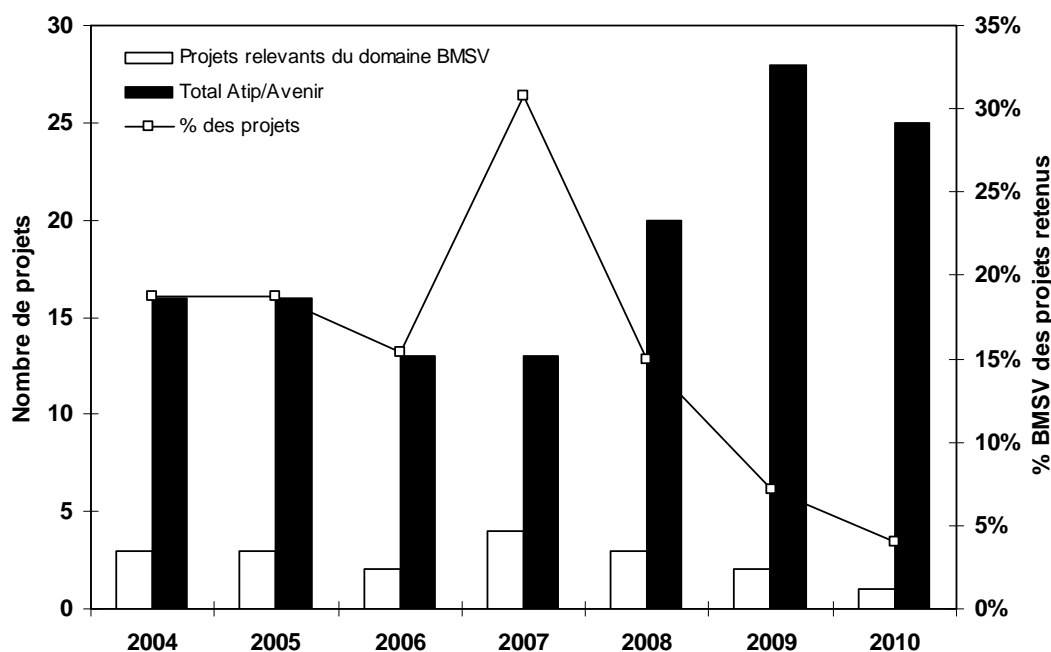
2.3.3.2. Jeunes équipes

2.3.3.2.1. Programmes Atip/Avenir

Les données de la Figure 19 ne tiennent pas compte du programme Avenir entre 2004 et 2008 (données en attente). La moyenne sur 6 ans est de 13%, fortement en érosion. On observe un recul fort avec 1-2 candidats sur 25 par an depuis deux ans. Le domaine BMSV est donc aussi en perte de vitesse

aussi chez les jeunes chercheurs. Pour le programme phare ATIP/Avenir instauré en 2009, la situation est très préoccupante pour le domaine BMSV. Il faut noter qu'en 2010, 8 candidats présélectionnés pour l'audition sur 51 affichaient un projet dans le domaine BMSV mais un seul porteur a été retenu à la fin. Il faut signaler enfin que l'Institut de Chimie du CNRS finance aussi des programmes de type ATIP. Si leur nombre est faible (2-4/an), il apparaît que le nombre de projets relevant du domaine BMSV est de 1-2/an. Il s'agit d'une proportion très importante au regard de ce que représente cette activité au sein de cet institut. Il s'agira de comprendre le sens exact de ce mouvement.

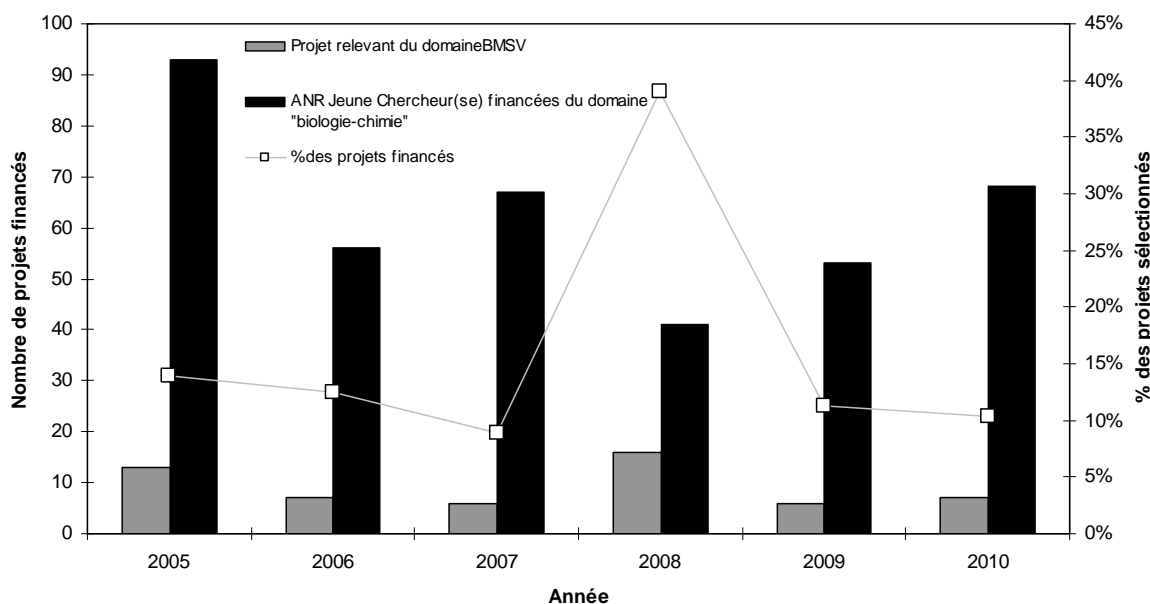
Figure 19 : Visibilité du domaine BMSV dans les programmes Atip/Avenir



2.3.3.2.2. ANR Jeune chercheur(se)

En moyenne, 14,6% des dossiers BMSV financés relèvent du domaine « Biologie – Chimie ». Il s'agit d'une proportion stable au cours du temps mais en retrait de la communauté concernée (20%).

Figure 20 : Visibilité du domaine BMSV dans la sélection des ANR Jeune Chercheur(se)



2.3.3.3. Formations

Dans les universités, on observe un recul de l'intérêt des étudiants pour les approches de sciences dures nécessaires pour la maîtrise des approches naturellement multidisciplinaires du domaine BMSV. Dans les grandes écoles, creuset naturel pour ce type de thématique, il existe un manque flagrant de formations et de filières d'attraction des étudiants potentiellement intéressés. Il est certain que le manque d'une tradition de la formation par la recherche dans les grandes écoles renforce cet aspect.

2.3.4. Conclusions

Dans le domaine BMSV, il faut rester très attentif au domaine de l'émergence des jeunes chercheurs et se méfier des critères d'évaluation basés sur la bibliométrie, notamment dès lors qu'il s'agit d'arbitrages entre domaines très différents de la biologie. On devra aussi se défier aussi de la perte de compétence en décourageant certaines communautés dans les approches qu'elles maîtrisent (enzymologie-cinétique, chimie médicinale...) mais qui tendent à disparaître dans la communauté faute de soutien plus fort des opérateurs financiers et des formateurs.

2.3.5. Les financements européens et internationaux en BMSV

2.3.5.1. Human Frontier Science Program

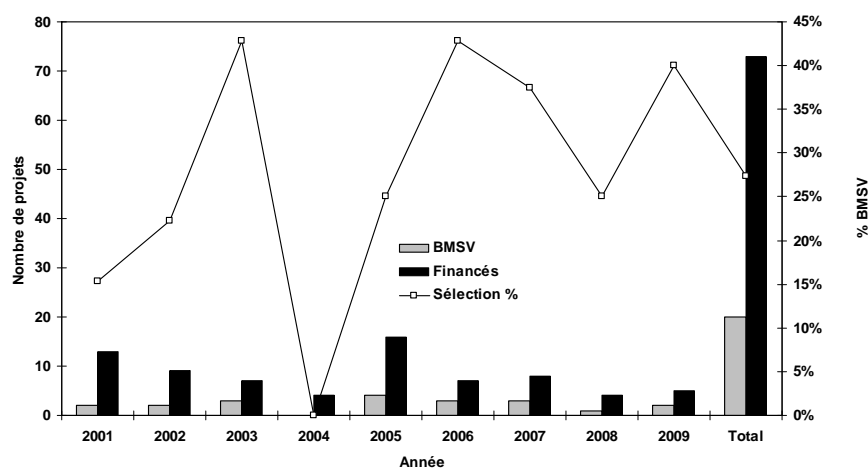
Le Human Frontier Science Program (HFSP) est actuellement la seule agence internationale de financement de la recherche fondamentale en biologie. HSFP met l'accent sur les approches nouvelles et interdisciplinaires qui implique des échanges scientifiques dépassant les frontières nationales et disciplinaires. Créée il y a 21 ans sous l'impulsion du Japon qui en est le principal financeur et par les pays du G7, cette agence regroupe actuellement 13 pays membre plus une représentation de l'Union

européenne et distribue environ US\$59 million par an. Si la France y consacrait 2,1 M US\$ en 2008, son taux de financement généralement dépasse du double, son investissement (4,5 M US\$ en 2008). Le programme de HSFP se divise en deux grands types de financement de la recherche :

- **les subventions :**

- les **Young Investigators' Grants** (5 projets BMSV sur 28)
- **Program Grants** d'un maximum de \$450,000 par an 27% (20/73)

Figure 21 : Porteurs français en BMSV dans le cadre des HFSP Program Grants



- **les bourses**

- **Long-term Fellowships** d'une durée maximum de trois ans de formation postdoctorale dans un autre pays (max \$US 45,000 par an) 1 bourse sur 16 financée dans le cadre BMSV depuis 2005
- **Cross-Disciplinary Fellowships:** des bourses transdisciplinaires destinées à des postdocs non biologistes pour recevoir une formation en biologie. 3 bourses sur 25 financées dans le cadre BMSV depuis 2005
- **Career Development Awards:** Aide au retour des boursiers « Long-Term » et des boursiers « Cross-Disciplinary » qui retournent dans leur pays d'origine. 6 bourses sur 24 financées dans le cadre BMSV depuis 2005

En tout, 9 bourses sur 65 en cinq ans relèvent de BMSV, soit une proportion de 14%.

2.3.5.2. L'European Research Council

Du fait de sa création récente, nous ne disposons pas de recul suffisant car depuis 2008 seulement, on ne peut analyser un état stationnaire. Pour 2008-2009, 5 sur 60 ERC senior et junior soit 8% des récipiendaires en biologie-santé (LS1 + interdisciplinaire).

2.3.5.3. Autres financements européens

11 projets sur 63 PICS internationaux financés à l'INSB en 2010 relèvent du domaine BMSV.

2.3.5.4. Conclusions

Les chercheurs BMSV participent fortement aux programmes de structuration européenne. Parmi eux la communauté strasbourgeoise est très dynamique.

2.4. Conclusions

Globalement, on assiste depuis 5 ans à une érosion lente des financements et supports nationaux du domaine dont l'ampleur peine à dépasser 15% pour une population de plus de 20%. Comme exemples de recommandations :

- un sous-domaine spécifique de l'ANR blanc devra être mis en place dès 2011 avec ce contour précis BMSV pour permettre à la communauté de s'identifier clairement dans ce large domaine. La mise en place de programmes thématiques à l'ANR de type « Chemical Biology », Biologie synthétique ou Protéomique, pourrait aider dans ce sens.

- les opérateurs de recherche, CNRS en tête, devront impérativement compenser par des financements récurrents plus importants dans ce domaine traditionnellement moins doté que la génétique par exemple.

- la communauté BMSV doit aussi réaliser une autocritique et rechercher un renouvellement de sa manière de présenter ses questionnements ou ses collaborations pour l'adapter aux évolutions les plus modernes de la biologie intégrative et non plus seulement structurale ou de la « chemical biology »; pour ce faire des outils comme les programmes INSTRUCT et EU-Openscreen devraient permettre à la communauté de mieux y réfléchir (voir & 7.5)

3. MOLECULES ET CHIMIE DU ET POUR LE VIVANT

3.1. Définition du domaine

La conception d'outils moléculaires qui permettent d'explorer le vivant ou de le contrôler, dans un but thérapeutique par exemple, est de plus en plus rationnelle. Ce faisant, elle devient très transdisciplinaire et fait appel, outre aux techniques de la chimie, à la physico-chimie structurale mais aussi à la biochimie, la pharmacologie cellulaire et à l'immunologie.

Les molécules d'intérêt biologique sont d'origines naturelle ou synthétique. De par leur variété et leur complexité, les constituants de base (peptides, lipides, glucides) ainsi que les macromolécules (protéines, acides nucléiques, polysaccharides, et autres constituants), permettent aux cellules d'assurer leurs tâches nécessaires au maintien, à la croissance et à la réplication des organismes. Le dérèglement de leur synthèse ou de leurs interactions est la cause première de nombreuses pathologies. Les molécules naturelles, d'hémisynthèse ou de synthèse sont des outils permettant d'interférer avec les processus biologiques pour les caractériser (sondes, marqueurs, modulateurs, agents de contraste pour l'imagerie) mais aussi pour rétablir un fonctionnement altéré lors de pathologies. Elles constituent des outils permettant aux techniques de visualisation d'étudier la nature hétérogène et pléiomorphe de nombreux complexes macromoléculaires impliqués dans les processus du vivant (activation de récepteurs, transmission du signal, etc.). Enfin, elles sont essentielles pour la nanochimie (fonctionnalisation, vectorisation, développement de nano-objets et nanomachines).

3.2. Principaux défis scientifiques et technologiques

3.2.1. Exploiter la diversité moléculaire structurale et fonctionnelle

Les organismes vivants ont développé une diversité moléculaire exceptionnelle associée à une diversité d'activités biologiques. Les métabolites secondaires d'origine végétale, marine ou isolés de microorganismes ont prouvé leur intérêt comme source de médicaments, mais également comme outils pour explorer les systèmes biologiques. Les substances naturelles offrent aussi aux chimistes des modèles à partir desquels de nouvelles molécules peuvent être élaborées par synthèse totale ou synthèse biomimétique. Elles sont aussi utilisées en tant que matières premières, précurseurs, châssis moléculaires dans la conception de structures originales aux applications potentielles multiples. Concernant les bio-macromolécules, la caractérisation des protéomes est en pleine expansion. A l'opposé, les domaines des lipidomes et des glycomes n'ont été que très peu explorés et de grandes avancées devraient être réalisées dans les prochaines années ouvrant de nouvelles voies approches thérapeutiques ou fondamentales. Enfin, la caractérisation de la diversité fonctionnelle des génomes demeure un enjeu comme en témoigne l'émergence du rôle des ARN non-codants.

Au niveau structural, l'étude de la diversité de structure et des interactions des macromolécules reste un objectif. C'est par exemple le cas du repliement de l'ARN et des interactions ARN-ARN qui ont une grande importance dans les interactions hôte-pathogène et les régulations de l'expression des gènes. Les

complexes transitoires qui impliquent plusieurs protéines et des acides nucléiques comme les complexes ARN-protéine(s) sont particulièrement difficiles à caractériser et représentent le défi majeur à venir.

La diversité moléculaire s'accompagne de propriétés physicochimiques qui peuvent également être exploitées. L'ADN est le siège de modifications chimiques et d'interactions diverses avec un ensemble de partenaires. Les défis actuels concernent l'exploitation des propriétés physico-chimiques de l'ADN et le développement d'outils en vue d'études structurales et fonctionnelles de l'ADN cellulaire. Le principe de nanomachines et de nanomatériels à ADN est maintenant bien établi, mais il reste cependant à exemplifier et à optimiser. Concernant la « nanochimie », l'ADN en double hélice ou structuré est un « réacteur chimique » de choix (« DNA Template synthesis », « click chemistry ») qui mériterait d'être mieux exploité pour catalyser des réactions de couplage *in vitro* et même cellulaire. Parallèlement, la synthèse d'ADN long est un enjeu, tant pour la biologie synthétique que pour la biologie moléculaire plus standard ; si elle est aujourd'hui possible, elle est encore largement améliorable. D'autre part, les études structurales et fonctionnelles de l'ADN cellulaire nécessitent le développement de ligands spécifiques, notamment de synthèse, en tant que sondes et régulateurs artificiels, en particulier pour la caractérisation des structures d'ADN inhabituelles ou des séquences répétées. Ces études impliquent, par exemple, le développement de stratégies originales de fonctionnalisation de l'ADN y compris de gènes *in situ*, notamment pour l'imagerie ou la modification contrôlée de séquences génomiques. La nécessité de tels outils ira sans doute en croissant car la possibilité de modéliser et prédire les interactions ADN-protéines à l'échelle d'un génome, révèle de nouvelles fonctions qui devront être caractérisées et validées biologiquement.

3.2.2. Développer des sondes pour l'étude des interactions moléculaires

L'étude des interactions entre deux macromolécules ou entre une macromolécule et son ligand exige un ensemble de méthodes d'analyse performantes. Les méthodes classiques permettent la caractérisation des paramètres structuraux (rayons X, RMN), cinétiques (résonance plasmonique de surface, cinétique rapide/stopped-flow...) et thermodynamiques (microcalorimétrie de titration). La conception de nouveaux outils moléculaires orientera les développements technologiques vers des sensibilités accrues et une miniaturisation plus importante grâce à l'utilisation des nano-objets tels que les puces (à ADN, à protéine, à sucre) et les capteurs.

L'imagerie moléculaire permet la caractérisation et la quantification de processus biologiques aux échelles cellulaire et moléculaire. Au niveau cellulaire, la combinaison unique de résolutions spatiale et temporelle, d'un aspect non invasif et non destructeur des cellules et des organismes, fait de la fluorescence, la technique de choix actuelle pour l'analyse des molécules uniques comme des interactions. Les interactions protéine/protéine, les changements conformationnels, l'activité de gènes et les synthèses et turn-over protéiques sont autant de phénomènes biologiques qui deviennent accessibles. Si la diversité des protéines fluorescentes naturelles doit poursuivre son évolution, les méthodes de marquages *in cellulo* fondées sur le développement d'étiquettes génétiques comme la méthode FISH, sont appelées à croître. Le développement de sondes à deux photons permet également de s'affranchir en partie de l'auto-fluorescence, réduit les dommages photo-induits et le blanchiment et autorise une pénétration tissulaire

plus profonde. Les techniques s'étendent à présent à la manipulation de l'activité de protéines ou leur adressage intracellulaire, en utilisant des chromophores photactivables/inactivables. Concernant les génomes, des outils de synthèse spécifiques pour les études structurales et fonctionnelles de l'ADN cellulaire seront également nécessaires. Cela implique le développement de stratégies originales de fonctionnalisation de l'ADN y compris de gènes *in situ*.

Les développements futurs se concentreront sur l'augmentation de la résolution spatiale et de la pénétration intracellulaire. Au niveau des assemblages complexes, les progrès porteront sur l'utilisation d'une illumination structurée ou de déplétion de l'émission. Pour la pénétration, l'excitation multiphotonique dans l'infrarouge autorise déjà des profondeurs de l'ordre de 500 µm à 1 mm. La tomographie sur tissus vivants devrait améliorer encore ces performances. Dans le futur, le développement de méthodes de ciblage plus efficaces permettrait également de combiner imagerie cellulaire et thérapie au sein d'un seul composé, utilisable en théranostique, enjeu majeur de la pharmacologie actuelle.

Outre les sondes protéiques et nucléiques, d'autres molécules du vivant peuvent être utilisées pour caractériser des interactions spécifiques. C'est le cas de lipides zwitterioniques stabilisant/déstabilisants, ou lipides ubiquitaires, mimant ou s'adaptant à une multiplicité d'environnement dont la conception permettra de déterminer les différents assemblages lipides-protéines présents dans les membranes biologiques.

3.2.3. Relever les défis de la chimie thérapeutique

La chimie thérapeutique joue un rôle décisif dans la découverte de molécules actives en vue de parvenir à de nouveaux médicaments. Les métabolites secondaires d'origine végétale et marine ou isolés de microorganismes représentent, encore aujourd'hui, une des principales sources de principes actifs dans différents domaines thérapeutiques. Les substances naturelles offrent aussi aux chimistes des modèles à partir desquels de nouvelles molécules peuvent être élaborées par synthèse totale ou synthèse biomimétique. Les approches et outils classiques de la synthèse organique restent toujours d'actualité dans le cadre d'une recherche thérapeutique d'avenir. Du point de vue synthétique, après les succès tout relatifs du criblage haut débit et de la chimie combinatoire, l'ère post-combinatoire se porte notamment à présent sur l'approche de découverte par fragments et sur l'exploitation de la diversité des macromolécules.

Les glucides (oligosaccharides et glycoconjugués), en interagissant avec des récepteurs protéiques spécifiques, participent à la communication intercellulaire. La synthèse chimique ou chemoenzymatique de carbohydrates, qui demeure une thématique ardue, devrait trouver des applications importantes dans le domaine des agents anti-infectieux ou antitumoraux et dans celui des chaperones chimiques et des vaccins. Dans le domaine des lipides, à titre d'exemple, les isoprostanes jouent un rôle physio-pathologique dans les maladies cardio-vasculaires, pulmonaires et neurodégénératives tandis que les prostaglandines, prostacylines et thromboxanes demeurent d'actualité dans l'inflammation et la régulation des métabolismes.

Un des défis majeurs de la chimie thérapeutique concerne le ciblage des interactions protéine-protéine, qui régulent les fonctions essentielles de la cellule et dont le dérèglement est responsable de pathologies humaines, soit par perte d'interactions, soit à la suite de stabilisation excessive. L'identification de petites molécules modulatrices capables d'interagir sélectivement avec ces systèmes dynamiques constitue un défi dont la preuve de concept a été déjà apportée (antitumoraux pro-apoptotiques...). Ces approches bénéficieront d'avancées en modélisation moléculaire portant sur la caractérisation *in silico* des propriétés des cibles et des ligands des effecteurs ou sur l'exploration de leurs interactions en prenant en compte la flexibilité des partenaires (induced-fit docking..).

Une autre approche porte sur le contrôle du repliement et de la stabilité des protéines dont les défauts sont à l'origine de graves dysfonctionnements. La réponse cellulaire aux protéines mal repliées (surexpression de chaperon moléculaire par exemple) est activée dans de nombreuses pathologies comme les cancers, les infections virales et des pathologies neurodégénératives. La conception de régulateurs chimiques des systèmes responsables de cette machinerie cellulaire constituera indubitablement une des enjeux importants de la recherche thérapeutique du futur proche.

Le domaine des vaccins a bénéficié des avancées considérables en immunologie moléculaire et cellulaire et du récent décryptage du mode d'action des adjuvants, permettant de concevoir à présent des vaccins sur des bases plus rationnelles et moléculaires. Dans ce domaine émergent, le rôle joué par la chimie apparaît essentiel pour identifier, caractériser et synthétiser de nouveaux épitopes, mais aussi pour concevoir et préparer des immuno-adjuvants innovants non toxiques et non inflammatoires tels que les analogues des ligands des « Toll-like Receptors » (lipopeptides, glycolipides, ligands des « Nucleotid Oligomerisation Domains » (NOD), etc...). Le vaccin de demain sera constitué par l'association de ces molécules néosynthétisées et fera appel à de nouvelles générations de vecteurs pour les véhiculer/cibler vers les cellules immunocompétentes. Des techniques imaginatives de bioconjugaison devront être développées : couplages chimio- et régio-sélectifs dans des conditions douces ou encore, ciblage actif à l'aide de ligands de récepteurs spécifiques de cellules présentatrices d'antigènes.

Dans le domaine des oligonucléotides, l'ARN interférence et le saut d'exon constituent deux modèles thérapeutiques à très fort potentiel. De nombreux travaux ont été réalisés sur la chimie des acides nucléiques. Cependant, dans le domaine thérapeutique, ces molécules nécessitent toujours une vectorisation. Les chimistes doivent développer des molécules intégrant les bases nucléiques et les fonctions de protection (aux nucléases, aux protéines), de ciblage et de pénétration cellulaire, de manière à passer du modèle d'administration « particulière » au modèle « moléculaire ». De même, le concept de nano-cargo supramoléculaire serait intéressant à développer. Au niveau synthétique, les défis portent sur le développement de nouvelles synthèses automatisées permettant d'intégrer plusieurs fonctions chimiques biomimétiques, sur l'association de ces techniques de synthèse automatisées avec les techniques de synthèse conventionnelles, sur le développement de la chimie de conjugaison des nano-objets (fullerènes, nanobagues, nanotubes), ainsi que les techniques analytiques correspondantes.

La contribution de la chimie thérapeutique ne se limite pas uniquement à l'élaboration de candidats médicaments. Elle s'étend à la conception de systèmes moléculaires de délivrance, capables de traverser les barrières de l'organisme ou encore à l'édification de nanoparticules capables d'apporter le principe actif jusqu'à sa cible, non sans avoir déjoué les systèmes de reconnaissance immunologique. Dans ce cadre, de nouveaux défis apparaissent. La thérapie génique notamment est toujours en manque de

vecteurs efficaces *in vivo*. Les principaux défis portent sur la biodisponibilité, l'efficacité intrinsèque des vecteurs et les voies d'administration notamment per os. Au niveau synthétique, le défi est de réunir dans un très faible nombre de composants chimiques l'ensemble des fonctionnalités biologiques nécessaires, ce qui implique de développer des réactions de conjugaison multiples, sélectives (pas de réactions croisées) et efficaces en phase aqueuse. L'association de plusieurs types de liaisons/réactions (ponts disulfures, esters, éthers, amides, carbamates, chimie click,...) en est la condition du succès.

3.3. Etat des lieux

3.3.1. Généralités

Les pôles principaux se retrouvent en Ile de France, Alsace (Strasbourg), Languedoc-Roussillon (Montpellier), Midi-Pyrénées (Toulouse), PACA (Marseille, Nice), Rhône-Alpes (Lyon et Grenoble) et Aquitaine (Bordeaux). Voir aussi la carte de France détaillée en Figure 14.

3.3.2. Structures d'animation et de coordination

Structurations locales :

- Pôles de compétitivité: Medicen, Lyon Biopôle (tous deux à vocation mondiale) et Alsace Biovalley
- Sur 77 IFR, une dizaine structurent un axe biomolécules/médicament et la grande majorité l'incluent en partie (on peut citer par exemple : *Innovation Thérapeutique : du Fondamental au Médicament*, *l'Institut des Sciences du Médicament*, *l'Institut Fédératif de Recherches Multidisciplinaires sur les Peptides (IFRMP)*, *Institut de Recherches en Technologies et Sciences pour le Vivant (iRTSV)*, *l'Institut Interdisciplinaire des sciences du Vivant des Saints Pères*)

Structurations nationales :

- GDR BioChiMar (Biodiversité et Chimiodiversité Marines),
- GDR Bioinformatique Moléculaire
- GDR Imagerie fonctionnelle du vivant
- GDR Médicaments Photoactivables - Photochimiothérapie (PHOTOMED)
- GR Anticorps et ciblage thérapeutique

Outils communs de recherche :

La chimiothèque nationale et le réseau de chimiothèques locales et de plateformes de criblage (Strasbourg, Paris, Grenoble, Toulouse, Marseille).

3.4. Analyse stratégique (SWOT)

3.4.1. Points forts

- Les recherches sur les substances naturelles ont une base historique forte en France, notamment depuis l'action de Pierre Potier et ses réussites dans le domaine de la pharmacologie humaine. Elles s'appuient

traditionnellement sur les ressources naturelles des DOM-TOM mais ont su évoluer vers les approches biosynthétiques modernes où le chemin de biosynthèse des produits naturels rares est étudié et reproduit en laboratoire, évitant ainsi le pillage d'environnements fragiles. La France conserve un potentiel important de recherche académique dans ce domaine.

- L'interface chimie biologie est également bien organisée. Elle a su mettre en place des outils performants pour la communauté, tels que la chimiothèque nationale et les plates formes de criblage. Des initiatives existent également pour fédérer celles-ci au niveau national et européen.

- L'intérêt de l'industrie du médicament est évidemment fort pour ce domaine. La France possède plusieurs entreprises pharmaceutiques dont l'une parmi les premières mondiales. Des structures susceptibles de constituer des passerelles existent comme l'Alliance pour la Recherche et l'Innovation des Industries de Santé. De plus, quelques initiatives ont été prises pour créer des structures mixtes public-privés permettant de mettre en commun des ressources chimiques et/ou techniques. Dans le domaine des biotechnologies, des GDRs favorisent le développement des biomolécules (anticorps, vecteurs).

- L'interface chimie biologie a su aussi mettre en place des collaborations fortes avec les équipes de biologie structurale. Les projets regroupant chimistes, biochimistes et structuralistes (RMN, cristallographie ou modélisation) obtiennent d'excellents résultats.

- La mutualisation des ressources dans le domaine de la biologie structurale et de l'imagerie est bien développée grâce aux plateformes nationales (IBiSA) et régionales. L'interface physique-biologie se structure également favorisant la pénétration rapide des nouvelles méthodes d'imagerie dans les laboratoires grâce notamment aux GDRs comme « Imagerie fonctionnelle du vivant ».

3.4.2. Points faibles

- La chimie thérapeutique souffre d'une absence de coordination : les laboratoires académiques développent des molécules potentiellement actives sur des cibles d'intérêt thérapeutique mais l'absence de structure permettant d'une part la caractérisation des activités biologiques et d'autre part la détermination des paramètres pharmacocinétiques et toxicologiques représente une réelle difficulté pour un transfert vers l'industrie avant les stades cliniques

- Le niveau d'interaction entre les sociétés pharmaceutiques et les laboratoires académiques reste limité particulièrement pour la mise en place de contrats de recherche sur sujets exploratoires.

- Le savoir faire en biochimie computationnelle (bioinformatique, modélisation moléculaire, docking, screening virtuel, chimiométrie..) est fort en France mais apparaît très dispersé, peu visible et manquant d'une coordination au niveau national.

- Le domaine de l'enzymologie subit un désintérêt et une perte de savoir-faire alors que la connaissance des génomes/métabolomes donne accès à de nouvelles enzymes utilisables pour la biocatalyse et la bioconversion. Un retard certain est en train d'apparaître en France dans ces domaines.

3.4.3. Opportunités

Les progrès en génomique et protéomique, l'émergence de nouvelles technologies à l'interface chimie-biologie, et les développements instrumentaux récents en imagerie moléculaire et cellulaire et en biologie structurale constituent autant d'opportunités pour concevoir et valider de nouvelles molécules actives. Le développement des nanotechnologies apporte une dimension nouvelle pour l'étude des molécules biologiques et de leurs modes d'action. Nanopuces, nanofluidiques, nanoréacteurs et autres outils miniaturisés à l'extrême permettent d'une part la caractérisation des propriétés et des interactions de molécules sans avoir à les produire à grande échelle et d'autre part l'étude de molécules uniques.

3.4.4. Risques

- En matière d'enseignement, les formations en biochimie et en chimie sont en perte de vitesse et n'ont pas su renouveler suffisamment leur image en s'orientant vers les sciences de la santé, la biocatalyse ou l'environnement. Le faible nombre d'étudiants pourra être une inquiétude dans les années à venir.

- Le manque de reconnaissance et de financement spécifique pour les projets situés à l'interface chimie-biologie au niveau national et européen n'encourage pas suffisamment les laboratoires à se lancer dans des projets ambitieux.

3.5. Recommandations

La prise en compte de la complexité de la formation des complexes moléculaires à l'origine des fonctions du vivant devient un enjeu majeur dans le but d'élaborer de nouvelles stratégies pharmacologiques. La question de la structuration de la recherche aux interfaces est de ce fait cruciale. Celle-ci devrait porter sur la mise en réseau de masses critiques de chercheurs autour de projet communs associant étude des réseaux d'interactions, caractérisation de ces interactions (biophysique), validation cellulaire (imagerie) et synthèse de régulateurs chimiques (sondes ou agents pharmacologiques).

- Mettre en place d'une coordination efficace avec l'ITMO TS pour une continuité efficace entre la chimie biologie et le développement de médicaments.
- Soutenir le développement de projets visant à converger les efforts de chimistes, modélisateurs et biochimistes pour le développement rationnel de molécules actives vers des cibles identifiées.
- Au delà de la constitution de réseaux par l'intermédiaire de programmes de recherche interdisciplinaire, des regroupements régionaux de chercheurs et la mise en commun de moyens d'analyse puissants constituent un objectif à moyen terme.

3.6. Propositions opérationnelles

- Engager au niveau national une réflexion sur un « chemin molécules actives » qui précéderait le « chemin médicament » en train de se mettre en place dans l'ITMO TS.

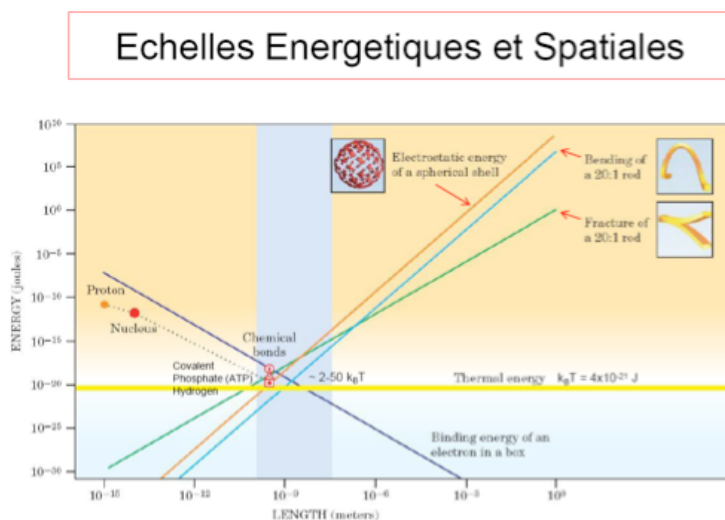
- Organiser un état des lieux de la chimie/biochimie computationnelle pour décloisonner les activités de modélisation moléculaire, chimiométrie, screening virtuel et donner une visibilité forte à ce domaine.

4. ASPECTS BIOPHYSIQUES DU FONCTIONNEMENT SUB-CELLULAIRE ET DES MACROMOLECULES BIOLOGIQUES

4.1. Définition du domaine

Ce champ d'investigation des molécules du vivant se concentre sur la caractérisation des mécanismes moléculaires et de leur fonctionnement. Une telle caractérisation implique de déterminer la structure des biomolécules, seules et en complexe avec leurs partenaires, d'étudier la dynamique et l'énergétique de leurs interactions et leurs réactions, ainsi que leur organisation spatio-temporelle dans la cellule (cytoplasmique, nucléaire, membranaire, ou associée à des organelles spécifiques).

De manière fascinante, les échelles de l'énergie et de l'espace en biologie convergent juste au dessus du niveau de l'énergie thermique et au carrefour entre les comportements macroscopiques régis par la mécanique classique, et les comportements de l'infiniment petit, régis par la mécanique quantique.



R. Phillips & S. Quake, Physics Today, Mai 2006

Ainsi, les molécules du vivant sont faites pour tirer profit des deux régimes, transformant l'énergie solaire en liaisons chimiques, et en utilisant cette énergie pour produire les molécules de sa structure et les forces nécessaires à la reproduction et à la survie. L'élucidation des mécanismes permettant aux biomolécules d'opérer une telle combinaison, et la relation entre leurs séquences (i.e. le code génétique), leur structure, leur dynamique et leur énergétique

constituent le but principal de la biophysique moléculaire.

Hautement interdisciplinaire, cette étude détaillée des molécules biologiques s'appuie sur des méthodes venant de la physique, de la chimie, des mathématiques et de l'informatique. Celles-ci comprennent des méthodes structurales à l'échelle atomique telles que la RMN et la cristallographie aux rayons X et aux neutrons, et à l'échelle mésoscopique telles que la microscopie électronique, la microscopie à force atomique ou encore la diffraction des neutrons ou des rayons X aux petits angles. Les informations dynamiques et énergétiques, y compris des mécanismes très fins, concernent le plus souvent des spectroscopies (optiques et de résonance magnétique) et des approches de molécule unique, avec des approches bien établies (patch-clamp) et des nouvelles méthodes de molécule unique basées sur des mesures de force (pinces optique et magnétique) ou sur des mesures optiques (smFRET, SERS). Les

nouvelles microscopies optiques à super-résolution (PALM, STORM, STED), les microscopies/spectroscopies à fluctuation, et les diverses méthodes de suivi de particule unique en 2 et 3D permettent de faire des mesures quantitatives *in cellulo* des concentrations, des interactions, de la localisation et de la dynamique des biomolécules individuelles et de leurs complexes. Ainsi, la résolution au niveau de l'Angstrom (0,1 nm) se confond avec la résolution nanométrique.

D'ailleurs, c'est sur le terrain de la biophysique quantitative et structurale abordée par les méthodes d'imagerie et microscopie avancée, et les techniques de la molécule unique qu'existe un recouvrement important entre l'ITMO BMSV et l'ITMO BCDE. Il sera important de coordonner les efforts dans ces domaines entre ces deux ITMOs. Il existe aussi un grand recouvrement avec l'INP et la section 11 de l'INC, car beaucoup des chercheurs impliqués dans le développement de ces méthodes sont des physiciens ou physico-chimistes. Cette partie thématique qui relève de la section 11 (*Inspiration de la matière molle: auto-assemblage, systèmes hors-équilibre, optique pour les systèmes biologiques ou « biophotonique », forces/mécanique, microfluidique, etc*) s'est considérablement développée dans les 5 dernières années. L'interface avec la Biologie constitue maintenant un des principaux thèmes de la section 11, avec chaque année des postes qui y sont attribués, allant de l'optique pour la biologie ou des études physiques *in vivo* s'appuyant sur des concepts développés en matière molle (voir § 6). En effet, ce champ d'investigation, en France plus qu'ailleurs, implique des acteurs de communautés assez distinctes, travaillant depuis longtemps de manière séparée, mais qui semblent actuellement en cours de convergence. (Paragraphe de la section 11 de l'INC- CNRS).

4.1.1. La biologie structurale

L'objectif principal de la biologie structurale reste la détermination de l'organisation atomique en 3 dimensions des molécules biologiques. Classiquement basée sur les techniques de diffraction aux rayons X et la Résonance Magnétique Nucléaire (RMN) liquide à haut champs et multi-dimensionnelle, la Biologie Structurale est tributaire de gros équipements (synchrotrons et spectromètres à haut champ magnétique, et aujourd'hui cryo-microscopes électroniques) pour accomplir ses objectifs. Le résultat des études structurales d'une biomolécule est une carte tri-dimensionnelle de l'emplacement le plus probable de ces atomes constitutifs. Ces cartes permettent, pour une enzyme par exemple, de proposer des schémas de réaction chimique, de comprendre les effets de mutations, souvent liées à des pathologies humaines, et servent comme bases primordiales pour la conception, par les chimistes biologiques et médicinales, de molécules pouvant agir positivement ou négativement sur l'activité de l'enzyme. Cette branche de la biologie structurale (Structure-Based Drug Design) a véritablement pris son élan lorsque ce concept a été intégré dans les programmes de recherche de la « Grande Pharma », et dont le précurseur est la société Genetech. Ce sous-domaine intègre, la RX et la RMN, mais aussi ce que l'on peut appeler la biologie computationnelle (prédiction de structure, criblage de ligands *in silico*) très développé outre Atlantique, mais quoique de haut niveau, toujours assez désorganisée en France.

Le domaine de la biologie structurale des protéines membranaires restent peut-être le "challenge" le plus important actuellement. Enormément de protéines membranaires, qu'elles soient humaines ou bactériennes sont des cibles de programmes de développement de molécules thérapeutiques. Parmi ces protéines cibles se trouvent les canaux ioniques et les récepteurs couplés aux protéines G et les protéines

MDR (Multi-Drug Resistant), ainsi que des porines et autres protéines bactériennes. Or ces protéines sont particulièrement récalcitrantes aux études structurales. Leur cristallisation est très délicate et leur étude par RMN est extrêmement laborieuse, nécessitant des spectromètres liquides à très haut champs ou des spectromètres solides (aussi à très haut champ). Ces travaux impliquent surtout une connaissance profonde de la physico-chimie des interactions protéine-lipide pour la préparation d'échantillons susceptibles de donner des résultats en cristallisation ou en RMN. La France dispose de plusieurs groupes experts dans ce domaine difficile (soit RMNistes associés surtout à l'INC, soit cristallographes, associés à l'INSB), quoique l'aspect relativement aléatoire des recherches a beaucoup freiné le soutien matériel apporté dans ce domaine.

Des programmes de recherche sur d'autres cibles très importantes, telles les virus et les protéines virales, le ribosome et les récepteurs nucléaires sont bien représentés et internationalement connus en France. Cibler les interfaces entre protéines dans des complexes fonctionnels est apparue comme nouvelle stratégie thérapeutique. Cette approche nécessite des informations structurales, non pas sur des protéines isolées, mais sur leurs des complexes comportant deux voir plusieurs partenaires. Pour caractériser la structure et la dynamique de ces interfaces et les cribler, une combinaison de techniques impliquant la RX, mais aussi la RMN et le DXPA ou SAXS (Diffusion des rayons X aux Petits Angles) est nécessaire. Le SAXS est très utilisé aussi pour caractériser des changements conformationnels. Par ailleurs, les protéines ne représentent pas les seules cibles biomoléculaires thérapeutiques; ainsi avec la démonstration récente de leurs rôles clefs dans de multiples processus biologiques, les ARN constituent une classe de molécules très importantes, mais difficile aussi à étudier, étant donné leur dynamique importante. La biologie structurale ne doit pas être réduite uniquement à son utilité dans la recherche de médicaments; elle sert aussi tout simplement à comprendre les mécanismes du vivant. Ainsi, la biologie structurale des plantes est aussi bien représentée en France, et des études sur les systèmes modèles issus des 3 règnes de la vie (4 avec les virus) nous apportent des informations générales qui ont des répercussions dans beaucoup de domaines d'intérêt pour le public, mais aussi le privé (Biotechnologies, Bio-ingénieries, par exemple).

Depuis plus d'une décennie déjà, la biologie structurale a dépassé les limites de la structure de la molécule ou même des complexes moléculaires pour s'attaquer aux édifices de plus en plus grands et dans leurs environnements naturels. Cette approche relativement récente, dite « intégrative » se poursuit avec des combinaisons de techniques, par exemple rayons X ou RMN avec la microscopie électronique, ou encore la microscopie électronique avec la microscopie optique, appelée microscopie corrélative, ou la tomographie cellulaire et les mesures optiques de molécule unique *in cellulo*. Les objets de ces études, les grands édifices cellulaires (ribosome, polymérases, protéasome, virus, complexes multi-enzymatiques, signalosome, et autres « somes ») sont les usines où un grand nombre des fonctions clefs et des réactions chimiques essentielles sont accomplies. Pouvoir visualiser ces machines dans toute leur complexité, apportera une compréhension très profonde sur ces mécanismes, et nous permettra d'aller encore plus loin dans la conception de stratégies pour moduler ces fonctions dans des buts de santé, de biotechnologie et de biologie synthétique. Actuellement trois grands axes d'avancement se distinguent l'augmentation de la résolution spatiale accessible en microscopie électronique, le développement des approches de visualisation ces édifices dans leur contexte cellulaire par tomographie électronique et le dernier concernant l'automatisation instrumentale et l'analyse d'image. On s'y trouve à la croisée des chemins des biologistes, qui sont hautement intéressés par ces nouvelles approches, avec des applications par

exemple en développement cellulaire, trafic intracellulaire cancérologie et organisations membranaires. En parallèle, se développent différentes approches qui apportent un complément d'échelle entre le moléculaire et le cellulaire comme les microscopies à force atomique d'une part et les microscopies synchrotron d'autres part. Le prochain défi consistera à intégrer les aspects dynamiques des mécanismes de fonctionnement de ces machines. C'est ici que la biologie structurale rencontre la biophysique.

4.1.1. Optique pour la biologie et méthodes de la molécule unique

Au niveau moléculaire, les méthodes de molécule unique ont permis de découvrir les mécanismes très fins du fonctionnement des biomolécules. L'apport principal de la biophysique et l'imagerie de la molécule unique est celui de dépasser les effets de moyenne des ensembles de molécules afin d'accéder aux étapes individuelles de fonctionnement des moteurs moléculaires qui sont totalement désynchronisés au niveau d'une population.

Les nouvelles techniques de l'optique permettent surtout une imagerie plus précise (10 à 80 nm de résolution, moins de signal/bruit) d'un plus petit nombre de molécules, ou une imagerie de plus en plus performante pour l'étude des systèmes dans des conditions *in vivo*, avec ou sans marquage fluorescent. Certaines techniques optiques comme la corrélation de fluorescence (FCS) permettent de suivre en temps réel la mobilité de molécules, *in vitro* et *in vivo*, avec une résolution spatiale de l'ordre du pixel de l'image (qui peut atteindre quelques nanomètres avec une caméra sensible et un objectif X100) alors que les techniques maintenant classiques (microscopie confocale, « spinning disk ») ne peuvent aller au dessous d'une fraction de micron, la résolution optique. La microscopie de force atomique (AFM) permet maintenant de visualiser des molécules uniques et leurs changements de configuration à une échelle spatiale de l'ordre d'une dizaine de microns, et une échelle temporelle bien inférieure à la seconde. Un couplage entre les différentes microscopies devrait permettre dans les années à venir de bien compléter les techniques de caractérisation du vivant. La microscopie corrélative, couplant l'optique avec la cryomicroscopie permet une visualisation du vivant dans des conditions de congélation *in situ* et évitant les artefacts dus au traitement des échantillons par des fixateurs. Par contre, la microscopie électronique ne permet pas d'aborder la dynamique des systèmes.

L'imagerie biologique a connu au cours de la décennie passée des développements majeurs en lien avec une conception de plus en plus intégrative et fonctionnelle de la biologie. L'étude des dynamiques et interactions moléculaires passent par leurs suivis cinétiques au cours de processus physiologiques naturels ou induits. Les lasers et l'optique non-linéaire, l'apparition des sondes fluorescentes de type protéines fluorescentes *in vivo* (la GFP, « green fluorescent protein »), a permis des progrès spectaculaires pour l'imagerie en molécule unique, mais aussi en cohorte de cellules. L'enjeu actuel est de réussir à corréler l'information obtenue par les différentes microscopies de fluorescence (FRET : transfert d'énergie entre molécules fluorescentes, FCS : spectroscopie de corrélation de fluorescence, anisotropie, FRAP : redistribution de fluorescence après photoblanchiment) en développant des dispositifs multimodaux. L'ambition sera ensuite de réaliser ces études au sein d'organismes vivants. Ces techniques doivent être également couplées à l'AFM et la microscopie électronique.

L'utilisation de protéines photoactivables (soit fluorescentes, soit dont l'activité biologique est activée) couplée aux dispositifs très sensibles d'imagerie (PALM : photo-activated localization microscopy,

STORM : « stochastic optical reconstruction microscopy », STED : « stimulated emission depletion microscopy ») permettra d'atteindre des résolutions spatiales de l'ordre de 5 à 20 nm et temporelle de l'ordre de 30 ms. Les années à venir devraient voir l'émergence de techniques basées sur le contraste intrinsèque d'objets biologiques non fluorescents, comme l'holographie, l'OCT : « optical coherence tomography », la microscopie plasmonique. Ces approches entre physique et matière molle « vivante » rentrent pleinement dans les spécificités de la section 11.

Les techniques de mesure de force sur molécule unique (spectroscopie de force) se sont considérablement développées ces dernières années avec des appareils basés sur des pinces optique ou magnétique capables d'exercer des forces du piconewton à quelques nanonewtons. Alors qu'il y a dix ans l'ADN était traité dans la communauté de la matière molle comme un polymère modèle qu'on pouvait micromanipuler, ce qu'on ne pouvait pas faire avec un polymère synthétique, la spectroscopie de force a permis récemment des avancées spectaculaires dans la compréhension des mécanismes d'interaction des acteurs agissant sur l'ADN, pour sa transcription, et aussi sa topologie. Depuis quelques années maintenant ces techniques sont adaptées aux protéines, et comment leurs changements de conformations sont possibles sous l'effet de l'application d'une force. L'application de la spectroscopie de force sur des protéines est la mécanosensibilité, i.e. comment des molécules transmettent au reste de la cellule les informations mécaniques. Souvent, il s'agit de sites cryptiques qui sont « révélés » sous l'action d'une force, et qui déclenchent ensuite une signalisation cellulaire.

Les forces d'assemblage moléculaires peuvent aussi être mesurés par des systèmes microfluidiques dans lesquels une protéine d'intérêt est immobilisée sur une surface, et ses interacteurs injectés dans un flux permettant, par exemple par des marquages fluorescents des molécules observées par des techniques optiques comme la réflexion totale de fluorescence (TIRF), et de caractériser la dynamique d'interaction. (Fin de la contribution du rapport de conjoncture section 11).

A titre d'exemple, nous savons grâce à ces approches de molécule unique, que l'extraordinaire efficacité de l'ARN Polymérase est due à un mécanisme de Roue à Criquet Brownien, dont le Criquet consiste en une transition ordre-désordre d'une partie de la protéine contrôlée par l'appariement du nucléotide correcte. Ce mécanisme permet à la fois une vitesse et une sélectivité extraordinaires. Il a aussi été démontré par un groupe français et un groupe américain par deux méthodes molécule unique distincte (les pinces magnétiques et le transfert de fluorescence de la molécule unique, smFRET) que l'initiation abortive de la Polymérase implique une entrée de l'ADN transcrit à l'intérieur du site actif de l'enzyme. De même les mécanismes de fonctionnement du ribosome, la kinésine et de plusieurs hélicases ont été élucidés par des approches de molécule unique.

Biologie structurale computationnelle Au-delà des approches expérimentales mentionnées ci-dessus, la biologie moléculaire computationnelle et théorique s'applique à tous ces niveaux d'organisation et de dynamique des biomolécules, permettant des prédictions de structures des biomolécules et de leurs ligands et d'appréhender les forces responsables de la dynamique structurale et des interactions biomoléculaires. Ainsi, les systèmes biomoléculaires étudiés en biologie structurale s'étendent sur de multiples échelles d'espace, allant des mécanismes de repliement d'un ARN ou d'une protéine individuelle, au fonctionnement des grosses « usines » cellulaires. Avec le pouvoir computationnel croissant, et ces approches multi-échelle, tout atome et gros grain, il est aujourd'hui possible de simuler la dynamique et les interactions moléculaires dans un contexte « cellulaire » ou encore d'étendre l'échelle de temps de la

nanoseconde à la microseconde et au-delà. Ces études visent à comprendre aussi bien des biomolécules « solubles » que membranaires, les machines moléculaires (protéasome, ribosome, polymérase, et des moteurs comme les myosines, les kinésines, etc) et les grosses structures sub-cellulaires (e.g., corps de Cajal, hétéro- et eu-chromatine, microdomaines membranaires). De nouvelles approches computationnelles d'amarrage et d'identification de ligands possibles, soulignent les interactions fortes entre la biophysique moléculaire et la chimie du et pour le vivant. Les informations issues des études en biologie structurale et biophysique moléculaire sont indispensables, non seulement pour la compréhension en soi, mais elles posent les bases pour la conception des nouvelles approches biotechnologiques et thérapeutiques.

4.1.2. Mesure de forces et d'interactions sur des systèmes vivants

4.2. Principaux défis scientifiques et technologiques

Le défi principal dans ce domaine sera l'intégration des approches aux systèmes moléculaires de complexité croissante dans des environnements au plus près du vivant. Il s'agit de l'intégration à deux niveaux : (i) l'intégration au niveau de l'utilisation et du développement des différentes technologies afin de rendre possible des synergies multi-résolutives dans l'espace et dans le temps, et (ii) l'intégration multi-échelle de données hétérogènes (structurales mais aussi fonctionnelles) dans le but d'intégrer les aspects atomiques et cellulaires, voir tissulaires.

Afin que leurs résultats représentent des avancées significatives dans la compréhension du vivant ou dans l'évolution des techniques pour ces approches, les équipes travaillant dans ce domaine doivent être constituées de personnels formés (individuellement ou collectivement) aux interfaces (biologie, physique, et chimie/biochimie, mathématiques et informatique). Ceci nécessite la création de programmes de formations interdisciplinaires pour les doctorants et post-doctorants.

Ces équipes ont également besoin d'avoir accès à des équipements importants (e.g. synchrotrons, centres de calcul) ou des appareillages lourds (RMN haut champ, cryo-microscopie électronique, bancs SAXS). Pour les approches molécule unique et microscopies et spectroscopies, l'équipement est moins lourd, mais toutefois plus important que la norme en biologie (lasers, microscopes, détecteurs rapides, etc...).

De plus, toutes ces approches nécessitent du développement s'appuyant sur des ateliers en électronique, optique, mécanique, micro-fluidique, et des développements en informatique (traitement de données multi-échelle, traitement d'image).

Enfin, ces équipes ont besoin d'évoluer dans un environnement (biochimie préparatrice, culture cellulaire, etc.) permettant la préparation rapide et efficace des échantillons pour les études biophysiques ou structurales des édifices biomoléculaires de plus en plus complexes.

4.3. Etat des lieux

Voir analyse en 2.

4.4. Analyse stratégique

4.4.1. Points forts

- **Un nombre important de groupes très performants en biologie structurale et biophysique, en particulier de la molécule unique**, reconnus au plus haut niveau international et un **terreau de jeunes chercheurs très prometteurs** et actifs.

- **L'organisation de la recherche** en France, autour des équipes dans des centres/laboratoires avec les organismes CNRS, INSERM, CEA, etc, pour tutelle minimisée l'isolement des chercheurs et augmente les possibilités d'interaction et de collaboration. En principe, un tel environnement est plus adapté que l'organisation anglo-saxonne autour des individus pour favoriser la recherche interdisciplinaire.

- L'existence d'IBISA (précédé par les Génopôles), un programme national de plateforme et de gros instruments est un atout précieux pour un domaine qui s'appuie fortement sur les approches physiques.

- Communauté interdisciplinaire forte et structurée en réseau au niveau national (GDR) avec des « points chauds » de développements

- L'existence des postes permanents d'ingénieurs de recherche est aussi d'une importance capitale pour des domaines très techniques tels la biophysique et la biologie structurale. En effet, cette catégorie de personnels confère un avantage compétitif très significatif par rapport au système anglo-saxon.

- Une organisation de la recherche en biologie structurale en France favoriserait la synergie entre laboratoires et avec des centres portant des développements et mettant à disposition des ressources et un savoir-faire d'un niveau international (notamment à travers le réseau INSTRUCT).

- compétences d'amont forte en physique, mathématique et matière molle

- bonne reconnaissance internationale en général

- bonne collaboration biochimie-chimie-physique dans le cadre d'un système intégré (au niveau des chercheurs sur le terrain).

- de plus en plus de projets avec une bonne implication de chaque côté (biologie/physico-chimie).

4.4.2. Points faibles

- Un des points critiques pour la biophysique et la biologie structurale en France est la **faiblesse des formations et filières universitaires** dans le domaine. Le repli général vers les disciplines principales (physique, chimie, biologie) qui a accompagné la restructuration en LMD des cursus universitaires a été dévastateur pour les disciplines d'interface, et en particulier la biophysique et la biologie structurale. Couplée à la tendance générale des étudiants de délaisser les filières scientifiques, cette évolution a fortement diminué le nombre de jeunes entrant dans les laboratoires pour y accomplir leurs thèses.

- **La faiblesse globale de la recherche industrielle** s'appuyant sur les méthodes de biophysiques dans les filières de la biotechnologie et la biopharmacie rend le choix d'une formation dans ce domaine assez aléatoire. Ceci a pour effet de diminuer encore plus l'attractivité de cette filière en France, alors

qu'ailleurs dans le monde, de telles formations sont très recherchées par les laboratoires privés. Cette faiblesse freine aussi le transfert de technologie et le développement de nouveaux produits et procédés en biotechnologie et en bio-pharma.

- **La capacité limitée des organismes de recherche à reconnaître et à pratiquer une réelle interdisciplinarité** reste un contrepoids négatif important qui tend à annuler les avantages pour l'interdisciplinarité qui pourraient découler de notre organisation en équipe et en centres.

- **La complexité du système de recrutement de chercheurs**, couplé au niveau assez dérisoire des salaires et la difficulté dans l'acquisition d'une subvention de démarrage, rend aléatoire l'identification et le recrutement de chercheurs experts dans un domaine spécifique de la biophysique ou la biologie structurale.

- **La complexité des règles de transfert de technologie** public-privé reste un frein monumental au développement et aux applications pratiques des résultats dans notre domaine.

- A un niveau de granularité plus faible, la France manque singulièrement de **microscopistes électroniques**. Par ailleurs, la **biologie moléculaire computationnelle et théorique** reste assez **désorganisée** par rapport à la situation outre atlantique.

- Trop faible accessibilité des appareils d'optique pour l'ensemble des laboratoires, nécessité de plus d'intérêt pour les développements suscités par les biologistes : actuellement, les biologistes attendent trop que le produit soit commercial pour l'utiliser.

- Trop faible implication des biotechs en France

4.4.3. Opportunités

- **Les évolutions du système** de recherche en France pourraient aboutir à un modèle qui serait particulièrement bien conçu pour les études interdisciplinaires. Pour ce faire, il serait important de prendre en compte les forces et faiblesses mentionnées ci-dessus dans la conception de nouveaux modes d'organisation. Par exemple, il serait peut-être avantageux de privilégier la création de postes au niveau ingénieur dans les laboratoires de biophysique et biologie structurale.

- **La création de centres interdisciplinaires**, rapprochant les biologistes et les chercheurs en biophysique moléculaire et cellulaire permettrait une application plus étendue des nouvelles approches développées et mises en œuvre par ces derniers. L'articulation de tels centres avec le tissu local (instituts, départements universitaires, etc) doit permettre dans des cas bien spécifiques de faire émerger des projets et des synergies qui ne verraient pas le jour autrement.

- Les avancées en biophysique et la biologie structurale étant étroitement liées aux avancées et capacités en instrumentation physique pour la biologie, il est essentiel de continuer à **développer une**

politique rationnelle d'équipement pour la biologie, en soutenant et en élargissant les programmes d'IBiSA, et en les coordonnant avec les actions de l'ANR et des organismes, des universités et des régions. Dans ce cadre, le projet INSTRUCT représente une grande opportunité. Il est important de créer ces centres ouverts et visibles qui allient développements technologiques et accès des avancées à la communauté..

Néanmoins, toutes les excellentes équipes françaises de biologie structurale et de biophysique n'ont pas vocation à fonctionner en plateau technique national de développement et de service. De très grand standing international, beaucoup d'entre elles sont hautement spécialisées dans les systèmes biologiques qu'elles étudient (virus, muscle, tuberculose, ARN, protéines de signalisation, etc.). Il est crucial pour toutes ces équipes de biophysique et de biologie structurale d'avoir accès aux technologies des grands centres, mais aussi de pouvoir bénéficier d'équipements de routine et de proximité pour faire avancer leurs travaux. La création d'un programme national d'instrumentation partagée alliant les acteurs locaux (Universités, Régions, Instituts (Curie, Pasteur, etc) et nationaux (ANR ou organismes) permettrait de soutenir cette recherche de très haut niveau.

- Mettre en valeur les approches interdisciplinaires et intégratives en créant un programme ou une section ANR spécifique à cette interface

4.4.4. Risques

Le risque principal réside dans **l'immobilisme** qui aurait pour effet de rendre impossible des évolutions positives qui pourraient combler les points faibles de la France dans notre domaine. Dans ce cas, le reste du monde nous dépassera rapidement dans ce domaine d'avenir.

Etant donné le caractère hautement interdisciplinaire, la volonté de tous les acteurs de contribuer à l'émergence de l'excellence doit être forte et constante. Il existe un risque important d'échec de dialogue entre les acteurs multiples et variés.

5. CONTROLE ET MODELISATION DU VIVANT

5.1. Définition du domaine

Les récents progrès technologiques en biologie ont conduit à un changement d'échelle en ce qui concerne la caractérisation des macromolécules biologiques. Ainsi, les techniques de séquençage à haut débit ont permis de déchiffrer de nombreux génomes provenant des trois règnes vivants : bactéries, *archaea* et *eukarya*. Les techniques de mesure de l'expression des gènes à grande échelle (puces à ADN, SAGE, spectrométrie de masse, etc...), et l'identification des interactions protéine-protéine ou acide nucléique-protéine (ChIP on chip, ChIP-seq, CLIP, méthode de double et triple hybride, TAP tag, etc...) donnent accès à une cartographie complète des gènes exprimés dans une cellule à un moment donné, et aux réseaux d'interactions entre macromolécules. L'élucidation des mécanismes moléculaires de régulation génique et de l'assemblage dynamique des protéines fait partie des grands enjeux de l'ère post-génomique. Le respect de la programmation génétique, de l'expression différentielle et/ou coordonnée des gènes au sein des cellules provenant d'organismes différenciés ou unicellulaires, et des réseaux d'interactions protéine-protéine et protéine-acides nucléiques, sont autant d'éléments indispensables à leur développement correct.

De nombreux biologistes ressentent aujourd'hui le besoin d'intégrer les connaissances acquises sur les acteurs moléculaires du vivant afin d'analyser les propriétés du système dans son intégralité, que ce soit au niveau d'un complexe multi-protéique, d'un compartiment intracellulaire, de la cellule, du tissu, de l'organe ou de l'individu. Pour cela, la caractérisation fine d'un système biologique passe par la production et l'analyse de données quantitatives, spatiales et temporelles. Ces données peuvent être produites en partant d'un système « normal » ou d'un système perturbé par des variations génétiques, des ligands chimiques, des stress biotiques ou abiotiques, afin d'analyser les conséquences phénotypiques de ces perturbations. Pour mener à bien de tels projets, les approches à grande échelle et la modélisation/simulation sont nécessaires.

5.2. Approches scientifiques et technologiques

5.2.1. Mécanismes moléculaires du contrôle qualitatif et quantitatif

5.2.1.1. Caractérisation des « acteurs » moléculaires et des mécanismes de reconnaissance : macromolécules ; petites molécules

Les processus fondamentaux d'une cellule (réplication, transcription, traduction, transport/localisation, dégradation,...) sont souvent régulés pour conduire à la répression et/ou l'activation coordonnée et temporelle de l'expression des gènes. Ces réseaux de régulation confèrent aux cellules la faculté de percevoir l'environnement, d'interagir ou de se synchroniser avec d'autres cellules, de proliférer, ou de se différencier. Parallèlement aux mécanismes de régulation, les mécanismes de surveillance du génome, et du contrôle de qualité des ARN et protéines contribuent à restreindre l'expression de l'information génétique inappropriée. Toutes ces propriétés d'une cellule ou d'un

organisme et leur évolution dépendent principalement d'interactions spécifiques entre biomolécules qui se produisent à différents niveaux de l'expression des gènes et de la synthèse des protéines.

Déterminer l'ensemble des interactions physiques entre biomolécules dans une cellule, défini par le terme interactome, est un préliminaire pour accéder à une vision systémique de la cellule. Des efforts considérables ont été réalisés sur divers organismes pour cartographier les interactions protéine-protéine, et plus récemment les interactions protéine-ADN conduisant à l'établissement des réseaux formés entre les facteurs de transcription et leurs gènes cibles. Avec le développement des méthodes haut débit de type transcriptomique (§ 5.2.2), il est apparu qu'une grande partie du génome est transcrite générant un nombre insoupçonné de nouveaux ARN dans les trois règnes du vivant. Même si nous sommes encore loin d'avoir répertorié l'ensemble des ARN et de leurs cibles, leur mécanisme d'action et leurs fonctions, les ARN sont au même titre que les protéines des acteurs clés de la régulation génique. Outre les interactions protéine-protéine et protéine-ADN, il devient donc essentiel de prendre en compte la contribution quantitative des ARN ainsi que les interactions protéine-ARN (maturation de l'ARN, transport et localisation, régulation post-transcriptionnelle), ARN non-codant-ARNm (réseaux formés par les ARN régulateurs et leurs ARNm cibles) et métabolites-ARNm (particulièrement dans les bactéries).

Un autre aspect important de la biologie concerne les modifications de l'ADN, des ARN et des protéines qui influent sur leur activité, sur la réponse génétique et sur la fidélité de la traduction. Un grand nombre d'enzymes de modification a été caractérisé mais leur mécanisme d'action est encore peu étudié. Les modifications chimiques de l'ADN génomique perturbent souvent l'expression des gènes et sont à l'origine de phénomènes épigénétiques (e.g., empreinte parentale). Les modifications des protéines comme la glycosylation exercent des fonctions multiples comme la régulation de l'activité de la protéine, l'étiquetage pour la dégradation, l'ancrage à la membrane, l'intégration dans une cascade de signalisation, l'adressage à un compartiment cellulaire, et l'identité immunologique. La formation de certaines modifications de protéines (phosphorylation, ubiquitinylation, sumoylation) est perturbée par divers pathogènes (virus, bactéries), et est à l'origine des dérèglements cellulaires. L'état de la chromatine est généralement dicté par les modifications post-traductionnelles des histones et une interdépendance entre la méthylation de l'ADN et des histones peut exister. La compréhension des mécanismes de reconnaissance au niveau cinétique et thermodynamique des complexes enzyme-cible, la structure des complexes, ainsi que la régulation des voies de biosynthèse ne sont donc pas à négliger.

Les interactions les plus souvent caractérisées sont généralement statiques et stables. Cependant, dans une cellule, elles sont souvent transitoires et apparaissent de façon hiérarchisée. Cette chronologie des événements dépend de l'état physiologique de la cellule ou de l'organisme. Leur compréhension nécessite la connaissance de multiples facteurs qui incluent la localisation cellulaire des acteurs, leur concentration intracellulaire, leur modification éventuelle, leur structure, l'aspect temporel de leur expression, la cinétique de formation des complexes et leur affinité. Les informations quantitatives sur les paramètres cinétiques et les concentrations des molécules *in vivo* sont souvent méconnues. Les données de l'interactome devront être par ailleurs complétées par les autres domaines de la biologie intégrative (génomique, transcriptome, protéome, métabolome ; voir §5.2.2). De telles informations n'iront pas sans la mise en œuvre d'outils informatiques et mathématiques spécialement dédiés et adaptés qui permettront de modéliser et étudier la dynamique des réseaux de régulation à l'échelle de la cellule (voir §7.1).

Un autre défi qui est actuellement en développement, concerne l'analyse de la régulation sur cellule unique. Elle permet d'aborder entre autre, la régulation stochastique des gènes (voir §7.1). Longtemps considéré comme un facteur limitant pour la cellule et donc souvent qualifié de "bruit", son intérêt évolutif est maintenant reconnu. La compréhension de ce phénomène et de ses conséquences passe nécessairement par l'étude de ses causes moléculaires. La stochasticité est souvent due à la rareté de certains évènements moléculaires du processus d'expression génique (et donc associée à la faible quantité de certaines macromolécules : facteurs de transcription, ARNm, protéine) et à la dynamique des promoteurs et de la machinerie moléculaire impliquée dans la régulation : temps de résidence très variés, coopération/compétition entre macromolécules, recrutement cyclique des composants. Par ailleurs, l'expression d'un gène peut être aussi variable d'une cellule à l'autre et les mécanismes moléculaires qui causent ces différences peuvent apporter un gain fonctionnel, e.g. adaptation plus facile d'une population de bactéries à divers stress. Ces analyses en cellule unique requièrent le développement de techniques sophistiquées comme par exemple la microfluidique, les techniques d'imagerie cellulaire super-résolutive (voir § 4.1.2), ainsi que des analyses quantitatives basées sur des modèles mathématiques (voir § 7.1).

La description du point de vue qualitatif et quantitatif de l'organisation d'une cellule et d'un modèle qui rend compte de son fonctionnement global aura sans aucun doute de multiples conséquences. Elle devrait aider à identifier l'origine des désordres pathologiques et les perturbations d'un organisme causées par divers agents pathogènes (virus, bactérie, parasite,...). Les technologies à haut débit (expression et inactivation des gènes) combinées aux analyses bioinformatiques des génomes (prédiction de nouvelles fonctions/voies métaboliques) devraient conduire à l'identification de gènes cibles, qui pourront avoir un intérêt fondamental biomédical ou biotechnologique. L'étude de la biodiversité métabolique, de l'adaptation et des mécanismes de résistance des microorganismes aux conditions extrêmes seront la source d'applications biotechnologiques, comme la recherche de nouvelles sources d'énergie ou de meilleures conditions pour le développement durable (bioremédiation, interactions bactéries/environnement) (voir § 6.2). A un niveau plus complexe, la détermination des mécanismes de transmissions des signaux entre cellules et de leurs interactions aura des applications en médecine (e.g., régénération de tissus), et en microbiologie (e.g., réseaux conduisant à la formation d'un biofilm, analyse des populations de microorganismes cohabitant dans un biotope donné).

5.2.1.2. Etude mécanistique de la régulation

La conjonction des études génétiques, biochimiques, cellulaires et structurales a conduit ces dernières années à l'élucidation des mécanismes moléculaires et de leur régulation de divers aspects fondamentaux du vivant, comme la transcription des ARN, la traduction et la dégradation des ARNm où les équipes françaises sont bien représentées. Par exemple, les progrès récents sur la compréhension globale à l'échelle atomique du fonctionnement des différents acteurs de la traduction, dont le ribosome, font de la machinerie traductionnelle un « système de référence » pour l'étude des machines cellulaires complexes, mais aussi pour appréhender les mécanismes de régulation au niveau de la traduction, et les processus de couplage entre traduction et stabilité des ARN.

De nombreuses équipes se sont aussi attachées à comprendre les interactions entre les voies de signalisation, les réseaux de régulation transcriptionnelle et les mécanismes épigénétiques qui impliquent un changement héritable du profil d'expression d'un gène sans modification de sa séquence. L'objectif est d'obtenir une vision intégrée de l'organisation du génome (compartimentation et dynamique des fonctions nucléaires, structure de la chromatine et mécanismes de régulation épigénétique et transcriptionnelle) à l'échelle de la cellule. Il est maintenant clairement établi que les processus épigénétiques peuvent impliquer des complexes ribonucléoprotéiques dont l'ARN est souvent de grande taille (e.g. l'ARN Xist et l'inactivation du chromosome X). Le recensement de tels grands ARN avec une localisation nucléaire reste cependant incomplet et leur mécanisme d'action mal compris.

L'une des découvertes majeures est certainement l'identification de la machinerie d'interférence et d'une myriade de nouveaux petits ARN non codants (siARN, miARN...). Les siRNAs, majoritairement observés chez les plantes et les eucaryotes inférieurs, exercent des fonctions de défense et de maintien de l'intégrité du génome contre les transposons, transgènes et virus. Dans la levure ou les plantes supérieures, les siRNAs participent à la formation d'hétérochromatine proche de régions répétées du génome. Les miRNAs, codés par les génomes des eucaryotes, régulent de manière coordonnée la traduction et/ou la stabilité de multiples ARNm qui codent, en outre, des facteurs de régulation de la transcription. Chez l'homme, il est suspecté que près de la moitié des ARNm cellulaires seraient régulés par un ou plusieurs miRNAs. Outre leur implication dans le développement et la différenciation cellulaire, ils contribuent à la régulation des réponses hormonales et à la réponse de divers stress d'origine biotique ou abiotique. Certains clusters de gènes de miARNs se trouvent positionnés à proximité de *loci* connus pour agir comme suppresseurs de tumeurs, comme oncogènes, ou bien encore impliqués dans des maladies génétiques (e.g. Prader Willis). Il est ainsi possible que la dérégulation de la synthèse des miRNAs serait un des facteurs déterminant du cancer. Ces petits ARN sont maintenant utilisés comme outils pour étudier la fonction des gènes mais s'avèrent être aussi des outils thérapeutiques prometteurs (voir § 6.2). Dans les bactéries, même si le nombre d'ARN régulateurs est plus réduit, pas moins de 5-10% du génome génèrent des ARN non-codants fonctionnels. Ces ARN 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 en modulant la stabilité d'autres ARN non-codants, en répondant à des signaux extérieurs physiques (température) ou à la concentration de métabolites (riboswitch) qui modifient leur structure pour perturber la terminaison de la transcription, la traduction ou la stabilité. La structure des ARN bactériens contribue de manière active à la régulation. A cause du couplage transcription-traduction-stabilité, la vitesse de transcription et de traduction peut influencer sur le repliement de l'ARN et sur l'expression génique (voir § 4.2). 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 mais aussi des complexes ribonucléoprotéiques qui sont impliqués dans toutes les étapes du devenir des ARNm. Trouver les cibles directes requiert l'utilisation combinée de techniques biochimiques, génétiques, d'imagerie cellulaire (voir § 4.1.2), de protéomique et de séquençage haut débit (§ 5.2.2). Cette étape limitante est un pré-requis pour déterminer leur mode d'action et les réseaux de régulation dans lesquels ils interviennent. Ces réseaux devront être intégrés dans un schéma global des régulations qui feront appel à la biologie des systèmes (voir § 7.1).

Outre la caractérisation des régions régulatrices (situées au niveau du génome, des ARNm, ou des protéines), et des facteurs agissant en *trans* (protéine, ARN, métabolites...), il est aussi essentiel de soutenir les études mécanistiques afin de caractériser à l'échelle moléculaire les mécanismes de régulation. Cet aspect nécessite le développement et l'utilisation d'une combinaison d'approches expérimentales *in vivo* et *in vitro* pour définir la structure des complexes formés entre les différents acteurs de la régulation (voir § 4.1) et d'en identifier les conséquences fonctionnelles sur l'expression des gènes et la synthèse des protéines.

La découverte de l'ARN interférence dans les plantes et le nématode démontre à quel point il est essentiel de continuer à soutenir l'étude de ces processus biologiques à l'échelle moléculaire dans divers organismes unicellulaires (bactérie, *archaea*, levure) et pluri-cellulaires (plante, invertébré, eucaryote supérieur), et des relations hôte-pathogène afin d'obtenir une vision évolutive des processus de régulation, des mécanismes d'adaptation et de défense des organismes vivants.

5.2.2. Analyse phénotypique du Vivant par le biais d'approches globales (« ~omiques »)

5.2.2.1. Inventaire des « acteurs » moléculaires des systèmes biologiques

- **Protéomique.** Le terme « protéome » désigne l'ensemble des protéines codées par un génome. Plusieurs avancées technologiques récentes expliquent la montée en puissance spectaculaire de la protéomique observée au cours des dix dernières années: (i) l'introduction de la spectrométrie de masse dans le domaine de la microanalyse des protéines et des peptides ; (ii) le séquençage et l'annotation de très nombreux génomes, ainsi que l'accession à des banques de données protéiques de mieux en mieux annotées ; (iii) le développement d'outils informatiques et bioinformatiques adaptés aux besoins générés par la protéomique. Ces avancées permettent désormais d'identifier en quelques heures les centaines, voire milliers, de protéines qui sont présentes dans un échantillon biologique. Après une digestion de l'échantillon par une protéase spécifique (généralement la trypsine), le processus d'identification est réalisé en mettant en correspondance les spectres de fragmentation expérimentaux des peptides produits et analysés par le spectromètre de masse avec ceux qui peuvent être prédits à partir des séquences existant dans les banques de données protéiques.

- **Métabolomique.** Il s'agit d'un domaine partagé avec l'ITS. L'ITS est concerné par l'aspect « biomarqueurs » ; l'IBMSV est concerné par l'ensemble des mesures et déterminations relatives à la détection de petits composés d'abondance relative très variable, voire de traces. On peut par extension extrême assimiler la mesure des métaux liés aux macromolécules comme faisant partie de ce domaine.

Le terme « métabolome » désigne l'ensemble des petites molécules (métabolites, coenzymes, cofacteurs, etc.) présentes dans un système vivant ou dans un échantillon biologique. Par extension, l'analyse du métabolome représente l'analyse du métabolisme à l'échelle d'un système biologique (cellule, tissu, organisme). Il s'agit d'une discipline récente (initiée au début des années 2000), en plein essor depuis 4-5 ans, qui s'appuie sur une convergence d'outils analytiques (parmi lesquels émergent la RMN et la spectrométrie de masse de type LC-, GC-MS ou FT-ICR), la bioinformatique (« data mining », identification des molécules, banques de données, etc.) et les mathématiques (statistiques, modélisation

des réseaux métaboliques). La fluxomique représente l'analyse des vitesses réelles des réactions biochimiques (flux métaboliques) *in vivo*. Il n'existe pas de méthode permettant une mesure directe – in situ – des flux métaboliques. La méthode indirecte la plus développée (¹³C-fluxomique) s'appuie sur des stratégies de marquage isotopique (isotopes stables, e.g. ¹³C) et sur des modèles mathématiques pour déterminer la distribution des flux dans les différentes voies métaboliques. Ces distributions de flux représentent la mesure quantitative du fonctionnement des réseaux métaboliques. Les flux sont le résultat final de l'ensemble des interactions entre les acteurs du métabolisme –enzymes, métabolites, etc – et des interactions régulatrices sur le métabolisme qui se déroulent dans un contexte biologique et physiologique particulier (résultante globale de l'interaction entre le génome et l'environnement).

5.2.2.2. Analyses dynamiques du Vivant

Une cellule est un système hautement dynamique : ainsi, la nature et la localisation de certains de ses composants, de même que les interactions entre composants, sont modifiées en réponse à de nombreux facteurs. L'étude quantitative des modifications du protéome, du transcriptome et du métabolome d'une cellule, fournit des informations extrêmement précieuses puisqu'elle permet d'appréhender, au niveau moléculaire, la manière dont la cellule réagit en permanence à son environnement. Un enjeu majeur réside aujourd'hui dans l'intégration de l'ensemble des informations issues de ce type d'analyses « ~omiques » dans le but d'améliorer nos connaissances des grands mécanismes qui régissent le fonctionnement des systèmes biologiques. Outre leur apport dans le domaine des sciences fondamentales, ces connaissances sont indispensables pour aborder certaines problématiques ayant trait à la biologie synthétique (voir § 6.).

- **Protéomique.** L'étude de la dynamique d'un système biologique par le biais d'analyses protéomiques quantitatives et temporelles reste aujourd'hui un challenge. Il s'agit d'associer aux répertoires protéiques des informations d'abondance pour chaque protéine identifiée.

Aujourd'hui, il existe une grande variété de stratégies possibles pour réaliser des analyses permettant d'accéder à la mesure relative ou absolue de la concentration des protéines dans un échantillon biologique. Ces stratégies font appel à des méthodes différentes, méthodes qui ne cessent d'évoluer et de s'enrichir de nouvelles approches. L'innovation s'appuie sur des progrès réalisés dans un vaste champ de disciplines qui s'étend de la biochimie aux statistiques, en passant par les technologies d'analyse par spectrométrie de masse. Ce sont à la fois la nature du matériel biologique et le type de question posée qui conduisent à utiliser une stratégie plutôt qu'une autre. Schématiquement, on peut distinguer deux types de stratégies: les études protéomiques comparatives menées à large échelle, et généralement sans *a priori*, et les études quantitatives ciblées sur quelques protéines d'intérêt, protéines sur lesquelles les analyses sont focalisées.

Les études comparatives à large échelle sont par essence naïves. Ce type d'approche a le mérite de permettre la découverte d'acteurs moléculaires impliqués dans les mécanismes biologiques étudiés en comparant entre plusieurs échantillons les intensités des signaux produits par le spectromètre de masse pour chaque peptide ; l'intensité de chacun de ces signaux est en effet directement lié à la quantité du peptide correspondant. Ce type d'approche, qui fait souvent appel à des méthodes de marquage

isotopique différentiel des protéines (SILAC, iTRAQ, ¹⁵N), peut par exemple être mis en œuvre pour analyser la dynamique d'un système biologique dans le cadre d'une étude cinétique (réponses à un ligand de récepteur, à une modification du métabolisme énergétique), ou pour décrypter les conséquences d'une mutation. Dans un contexte clinique, ces mêmes approches naïves peuvent être utilisées pour comparer des échantillons biologiques issus de personnes saines et de patients atteints d'une pathologie ; elles conduiront alors à la découverte de candidats biomarqueurs de la pathologie.

Les analyses protéomiques quantitatives ciblées sont utilisées pour suivre plus particulièrement le comportement d'un nombre limité de protéines (quelques dizaines). Pour cela, il est fait appel à des méthodes de spectrométrie de masse de type « SRM » (« Selected Reaction Monitoring ») qui permettent l'obtention de données quantitatives d'excellente qualité, voire de données de quantification absolue si l'analyse est réalisée en présence des standards internes appropriés pour la quantification : peptides (AQUA), polypeptides (QconCATs) ou protéines (PSAQs) isotopiquement alourdis.

- Métabolomique/Fluxomique

L'analyse du métabolome s'appuie sur plusieurs stratégies expérimentales qui répondent à deux objectifs majeurs : 1) le phénotypage (haut-débit) du vivant pour l'analyse de perturbations globales et pour la recherche et l'identification de biomarqueurs et 2) l'analyse fonctionnelle du métabolisme.

Les approches de phénotypage s'appuient sur des analyses biochimiques rapides (empreintes ou profils métaboliques) - réalisées le plus souvent par RMN ou spectrométrie de masse - couplées à des outils statistiques permettant de classer des échantillons sur la base de l'identification de facteurs discriminants (chimométrie). La nature chimique du ou des facteurs discriminants (biomarqueurs) peut être déterminée grâce à des méthodes analytiques de résolution structurale (MS, RMN, etc.) ou par des approches de différences spectrales. Ces approches sont très largement utilisées pour la recherche sans *a priori* de biomarqueurs à une échelle qui peut être très globale (organisme entier, écosystèmes microbiens, etc.).

Les approches d'analyse fonctionnelle ont pour objectif, l'analyse fine, qualitative ou quantitative, des réseaux métaboliques. Elles répondent à trois objectifs principaux:

(1) Identifier le métabolome, c'est-à-dire identifier d'une part, l'ensemble des métabolites présents dans le système étudié (approches de métabolomique ciblée ou globale, bases de données et outils bioinformatiques pour l'identification) et, d'autre part, identifier les réactions métaboliques réellement en œuvre dans un contexte particulier (approches de marquage isotopique, méthodes de reconstruction métabolique, etc).

(2) Quantifier le métabolome : méthodes de métabolomique quantitative soit globales (approches IDMS, etc.), soit ciblées (profils métaboliques), permettant de mesurer les concentrations absolues des métabolites et leurs variations, et méthodes d'analyse des flux métaboliques (¹³C-fluxomique, modèles métaboliques quantitatifs, etc) pour quantifier l'activité réelle des réactions et des voies métaboliques.

(3) Mesurer la réponse métabolique globale à des perturbations pouvant être génétiques ou environnementales. Il s'agit ici de mettre en œuvre les approches mentionnées ci-dessus pour obtenir une mesure fine, dynamique, de la réponse du métabolome et du fluxome à des perturbations pouvant être de nature très variées (modifications génétiques, changements environnementaux, action d'effecteurs

biologiques, biotiques ou abiotiques, effet de pathogènes, de virus, etc). Des stratégies complémentaires, comme la RMN *in vivo* ou *in situ*, peuvent également apporter des informations dynamiques.

- **Cribles siRNA et HTS.** Bien que les cribles HTS et siRNA s'inscrivent aujourd'hui dans le cadre des approches à grande échelle, le contexte de leur utilisation est différent des approches « ~omiques » puisqu'il s'agit généralement d'utiliser des banques de molécules dans le but d'identifier celles qui sont capables d'induire un phénotype donné.

L'utilisation de banques d'ARN interférants permet de réaliser des transfections massivement parallèles d'ARNi dans des îlots de cellules déposées sur des lames de verre afin d'étudier de manière systématique et à haut débit les conséquences phénotypiques de la déplétion d'un gène. La chemogénomique, elle, utilise des petites molécules pour perturber systématiquement la fonction des protéines et déterminer leur fonction. On distingue la chemogénomique inverse, où le criblage est utilisé pour sélectionner les petites molécules capables de se lier à des protéines purifiées et de perturber leur fonction (ou leur interaction avec d'autres protéines), et la chemogénomique directe où des petites molécules capables de provoquer des phénotypes cellulaires particuliers (stabilisation des microtubules, état de phosphorylation d'une histone, etc.) sont recherchées. En plus de leur intérêt thérapeutique potentiel, les petites molécules présentent beaucoup d'avantages pour la recherche. Ce sont en effet des outils pour sonder les phénomènes dynamiques (contrôle temporel) qui permettent aussi bien une activation qu'une inactivation rapides des voies ciblées (effet réversible) ; de plus, elles peuvent permettre un contrôle dose-dépendant des fonctions biologiques.

5.2.2.3. Analyses dynamiques du Vivant

Les défis des analyses à grande échelle sont multiples ; ils concernent principalement la nature et la qualité des données générées, les volumes de données extrêmement importants qui seront produits par de telles approches, et enfin la fouille, l'intégration et l'exploitation de ces données. A ces défis génériques qui existent pour l'ensemble des approches « ~omiques », s'ajoutent des efforts à mener dans chaque domaine particulier.

5.2.2.4. Les défis de l'analyse à grande échelle

5.2.2.4.1.1. Les défis liés aux méthodes de préparation et d'analyse des échantillons

La protéomique s'appuie largement sur la Biochimie. En effet, lorsque l'analyse protéomique vise à explorer les grandes fonctions de la cellule afin d'en décrypter les mécanismes moléculaires, il est indispensable de « cibler » le matériel qui sera étudié en fonction de la problématique abordée. Pour cela, il est fait appel à des méthodes de fractionnement qui permettent d'enrichir le matériel biologique qui est le plus pertinent pour étudier ces fonctions. Dans le cadre des approches protéomiques à grande échelle, le principal objectif est l'analyse descriptive et dynamique des architectures et complexes protéiques impliqués dans les grandes fonctions cellulaires. Cet objectif implique d'être capable de relever un certain nombre de défis, parmi lesquels : (1) le développement et la mise en œuvre de stratégies qui permettent d'isoler systématiquement les complexes protéiques d'intérêt, par exemple en ayant recours à l'étiquetage des protéines ; (2) le développement de méthodes permettant la caractérisation des principales modifications post-traductionnelles portées par les protéines ; (3) le développement de méthodes permettant l'analyse systématique de protéines présentant des signatures, des fonctions ou des propriétés physico-chimiques particulières : protéines membranaires, protéases actives, protéines ubiquitinyllées, etc (4) La maîtrise des méthodes d'analyses protéomiques quantitatives d'échantillons biologiques complexes, qu'il s'agisse de méthodes comparatives ou de méthodes de quantification absolue.

5.2.2.4.1.2. Les défis liés aux méthodes d'analyse, de structuration et d'intégration des données

Les données générées par les approches à haut débit sont aujourd'hui largement sous-exploitées, ce qui limite l'impact des approches « ~omiques ». Cet état de fait, qui nuit à l'image des approches dites « à grande échelle », est principalement dû à un manque de moyens dans les domaines de l'informatique et de la bioinformatique, la bioinformatique étant ici définie comme l'ensemble des approches mathématiques, algorithmiques et statistiques appliquées à la génomique. Un défi majeur pour la Protéomique consiste à être capable d'intégrer la somme d'informations associées aux données de manière à pouvoir ensuite effectuer des requêtes gouvernées par la question biologique qui a motivé la réalisation des analyses protéomiques. Dans ce contexte, le développement d'expertises et d'outils permettant d'accéder à cette dimension intégrative doit être une priorité pour les années à venir.

5.2.2.4.1.3. Métabolomique et Fluxomique

La métabolomique et la fluxomique s'appuient sur des stratégies analytiques rendues complexes par la nature même du métabolome: très grande diversité physico-chimique des métabolites (un sucre ne peut être ni extrait ni analysé de la même façon qu'un acide gras), par des gammes de concentrations qui peuvent être très différentes (de 10^{-10} à 10^{-1} M), et par des demi-vies des métabolites très courtes (de l'ordre ou inférieures à la seconde). L'analyse la plus complète possible du métabolome passe donc par une convergence d'approches analytiques, de façon à identifier et quantifier les métabolites. Les principaux outils sont la RMN et la spectrométrie. Par ailleurs, l'analyse du métabolisme ne se limite pas à

l'analyse des métabolites, mais inclut la nécessité d'identifier quelles sont les différentes réactions et voies métaboliques qui fonctionnent réellement dans un contexte particulier et comment celles-ci interagissent (en incluant les aspects de compartimentation) et se coordonnent entre elles, dans le temps (notamment pour répondre à des changements environnementaux ou à des effecteurs biologiques), ce qui fait appel à des approches supplémentaires. Du fait de son émergence très récente, les défis sont nombreux et se situent aux différents stades du processus analytique :

- Développement de méthodes d'échantillonnage et des méthodes analytiques permettant d'accéder à un nombre de plus en plus importants de métabolites. Cet aspect est critique pour la métabolomique quantitative
- Développement et convergence des méthodes analytiques (notamment RMN et MS) pour résoudre la complexité métabolique et accéder qualitativement puis quantitativement à toute la diversité physico-chimique du métabolome,
- Développement des approches de métabolomique quantitative.
- Développement des approches de fluxomique, dont le champ d'application doit être étendu. Le développement de la fluxomique en états non stationnaires constitue également un enjeu important en vue de caractériser finement la réponse métabolique en termes de redistribution des flux.
- Développement des outils bioinformatiques, à la fois pour le stockage et le traitement des données et des meta-données, mais aussi pour l'identification des molécules.
- Développement des outils mathématiques, que ce soit en statistiques ou en outils de modélisation des réseaux métaboliques.
- Développement du haut-débit, aussi bien en métabolomique qu'en fluxomique.

5.2.2.4.1.4. Défis transverse : Intégration des données phénotypiques et modélisation

Pour atteindre cet objectif, il est nécessaire d'être capable d'intégrer la somme d'informations associées aux données de manière à pouvoir ensuite effectuer des requêtes gouvernées par la question biologique qui a motivé la réalisation des analyses. C'est à cette seule condition que l'on pourra repérer des protéines d'intérêt, par exemple parce qu'elles présentent des profils de variation d'abondance parfaitement corrélés dans le cadre de l'analyse cinétique d'un mécanisme. Une analyse récente illustre parfaitement la nature des informations qu'il est nécessaire de prendre en considération, ainsi que les niveaux de structuration de ces informations qu'il faut atteindre pour espérer à terme proposer des modèles réalistes des systèmes biologiques étudiés.

5.3. Etat des lieux

Voir la partie & 2 du document.

Les laboratoires principaux sont localisés en Ile de France, Strasbourg, Illkirch, Montpellier, Toulouse, Marseille, Lyon, Grenoble, Bordeaux. Voir aussi la carte de France détaillée en Figure 14.

5.4. Analyse stratégique (SWOT)

5.4.1. Points forts

- L'organisation de la recherche permet de fédérer des groupes sur des projets à long terme, autour de plateformes IBiSA bien armées au niveau technologique et animées en général par des chercheurs et des ingénieurs de recherche. Ceci constitue une expertise stable qui parie sur la durée.
- Bonne visibilité des équipes au niveau international sur les mécanismes de régulation de l'expression génique et de contrôle de qualité, sur le rôle de l'ARN dans la régulation et ses applications.
- Plates-formes certifiées ISO ; 5 en protéomique.
- Un ensemble d'investissements réalisés et de moyens humains conséquents (voir l'état des lieux).
- Des approches qui semblent arriver à maturité.

5.4.2. Points faibles

La communauté souffre du retard sur la mise en place de plateformes de séquençage haut débit et de personnel compétent en bioinformatique pour l'analyse des données (voir rapport de l'IGGB).

Pour ce qui concerne spécifiquement la protéomique :

- Important éclatement des compétences (nombreuses petites plateformes)
- Peu de structures ayant acquis une visibilité internationale dans le domaine
- Interactions encore trop faibles avec les experts du domaine de l'informatique et de la bioinformatique.
- Très peu de programmes de recherche dans le domaine de la biologie des systèmes
- Présence faible dans les projets européens.

Pour la métabolomique

- Une communauté en construction, et des plateformes qui émergent.
- Peu de laboratoires ayant acquis une visibilité internationale dans le domaine.
- Aucun programme spécifique n'a pour l'instant appuyé le développement de cette discipline en France.
- Offre de formation très faible, que ce soit en formation initiale (un seul Master 2 de métabolomique en France, à Toulouse), ou continue (initiatives du Réseau Français de Métabolomique-Fluxomique).

5.4.3. Opportunités

- Pour l'aspect mécanistique et régulation, une communauté structurée autour de groupes thématiques très actifs de la Société Française de Biochimie et de Biologie Structurale (SFBBM), et des sociétés savantes internationales comme la FEBS, ou la société Américaine RNA.
- Une communauté structurée grâce à IBiSA. Existence d'un réseau Protéomique pour coordonner certaines initiatives, notamment dans le domaine de l'animation scientifique.
- Des approches méthodologiques qui arrivent à maturité.
- Des domaines d'expertise diversifiés et complémentaires entre les principales plates-formes françaises

- Une communauté bien organisée, avec notamment l'existence de deux sociétés savantes : la Société Française d'Electrophorèse et d'Analyse Protéomique (SFEAP), et la Société Française de Spectrométrie de Masse (SFSM)
- Une communauté de mieux en mieux structurée grâce à IBISA.
- Existence d'un réseau Protéomique pour coordonner certaines initiatives, notamment dans le domaine de l'animation scientifique.

Pour la métabolomique

- Une communauté qui s'est structurée autour de la formation du Réseau Français de Métabolomique et Fluxomique : (https://www.bordeaux.inra.fr/ifr103/reseau_metabolome/accueil.htm), lieu d'échange, d'animation scientifique et de coordination d'initiatives. L'action de ce réseau a été récemment appuyée par IBISA.

5.4.4. Risques

- Etant donné le coût des analyses, recherche permanente de fonds pouvant conduire à une multiplication de projets de recherche souvent dispersés sur le plan thématique, et de durée et d'envergure limitées.
- Difficultés à assurer la maintenance et la jouvence des parcs instrumentaux du fait de leur coût très élevé.
- Difficultés pour certaines plates-formes à mettre en œuvre des approches à grande échelle du fait d'un environnement informatique non adapté.
- Incapacité pour certaines plates-formes à répondre aux « Guidelines » de plus en plus lourdes imposées par les meilleures revues du domaine, et donc à valoriser correctement les résultats de leurs travaux.

5.5. Recommandations/Propositions opérationnelles

5.5.1. Recommandations générales

- La recherche fondamentale peut bénéficier de la recherche appliquée et/ou médicale et inversement. Ces disciplines devraient donc être menées de manière conjointe et en concertation puisque les dérégulations de l'expression des gènes conduisent à des pathologies humaines. Des développements biotechnologiques pourraient aussi émerger de l'étude mécanistique de la régulation.
- Nécessité de renforcer les thématiques autour de l'ARN et des RNP (où nous manquons d'information sur les mécanismes de reconnaissance ARN-ARN, ARN-protéine) qui sont au premier rang des préoccupations actuelles de la recherche fondamentale au niveau international (i.e., programmes prioritaires aux USA, en Allemagne...).

- Une partie des préoccupations de l'ITMO BMSV et de l'ITMO IGGM se recouvrent (régulation de l'expression des gènes, transcriptomique et séquençage haut débit). Il sera important de coordonner les efforts et faciliter les échanges d'idées et des savoir-faire entre ces deux ITMOs.

- Nécessité de renforcer la communication auprès des jeunes étudiants, et la formation pour rendre nos disciplines plus attractives.

- Renforcer les soutiens pour l'organisation de workshops ou colloques internationaux, comme par exemple, les Conférences Jacques Monod (CNRS) ou les formations INSERM ouvertes sur l'international. Il est à noter que les congrès soutenus par Cold Spring Harbor sur la régulation post-transcriptionnelle sont maintenant organisés tous les deux ans à Heidelberg (Allemagne), montrant une forte dynamique des Allemands dans l'organisation de tels événements.

5.5.2. Cas particulier des analyses « ~omiques »

5.5.2.1. Protéomique

Aujourd'hui, la masse d'informations qui peut être rendue accessible grâce à la Protéomique fait de ce domaine un outil d'investigation majeur pour étudier le fonctionnement de la cellule. Un défi à relever en France serait d'avoir la capacité de systématiser les études protéomiques à grande échelle afin d'être en mesure de répondre aux attentes de la communauté des Biologistes. Un tel défi passe par une évolution significative en termes d'organisation et de formation. Ainsi, nous proposons la mise en place d'une infrastructure constituée de quelques centres nationaux de ressources en Protéomique, à l'image des infrastructures qui sont actuellement financées à l'échelle de l'Europe dans le cadre du FP7. Cette infrastructure, qui réunirait au début les trois ou quatre plates-formes françaises les plus performantes, serait rapidement créée. Elle serait chargée :

(1) De coordonner la réponse à un appel d'offre « Protéomique » qui serait mis en place par l'ANR. Les projets soumis dans le cadre de cet appel d'offres seraient expertisés par un conseil scientifique international. Les coûts de réalisation des projets seraient pris en charge par l'enveloppe budgétaire de l'ANR. L'appel d'offres aurait comme mission d'inciter la proposition de projets ambitieux susceptibles d'augmenter la visibilité de la France dans le domaine de la Protéomique. En particulier, l'appel d'offres pourrait fonctionner comme un outil permettant d'inciter le recours à des approches multidisciplinaires menées sur certains modèles biologiques, et de favoriser quelques thématiques de recherche prioritaires, thématiques qui seraient proposés par les ITMOs. En cela, cet appel d'offres pourrait ressembler au programme SystemX lancé en 2008 par la Suisse dans le but de favoriser la montée en puissance de la Biologie des systèmes.

(2) De mettre en place un fonctionnement en réseau, en mettant l'accent sur le développement coordonné d'expertises complémentaires sur les différentes plates-formes. Ainsi, les projets sélectionnés dans le cadre de l'appel d'offres ANR seraient attribués aux différentes plates-formes en fonction de la nature du projet et des spécificités de chaque plate-forme.

(3) De se coordonner avec le Réseau National de Bioinformatique (ReNaBi), notamment dans le but de favoriser l'émergence de projets dans le domaine de la Biologie des systèmes.

(4) De gérer la formation et l'animation scientifique sur le plan national. Des initiatives fortes dans le domaine de la formation, notamment la formation des personnels statutaires, permettraient de

progressivement amener certaines plates-formes à maîtriser l'ensemble des « outils » nécessaires à la réalisation d'analyses protéomiques à grande échelle.

5.5.2.2. Métabolomique

La métabolomique (et à fortiori la fluxomique) sont des approches récentes qui sont encore en phase d'émergence et de développement, aussi bien au plan international qu'en France. Ainsi, les premières labellisations de plates-formes de métabolomique sont intervenues seulement au cours des 2-3 dernières années. Il apparaît nécessaire de soutenir leur développement.

(1) De soutenir le développement des plates-formes, dont la mise en place est récente, sur le plan instrumental (acquisition et renouvellement) et en termes de compétences (biochimiques, analytiques, bioinformatique, mathématiques).

(2) De coordonner la réponse à un appel d'offre « Métabolomique » mis en place par l'ANR. Les projets seraient expertisés par un conseil scientifique international. Cet appel d'offres aurait comme mission d'inciter la proposition de projets ambitieux susceptibles d'augmenter la visibilité de la France dans le domaine de la Métabolomique. Les coûts seraient pris en charge par l'enveloppe budgétaire de l'ANR.

(3) De s'appuyer sur le Réseau Français de Métabolomique et Fluxomique pour favoriser la structuration de la communauté.

(4) Développer la formation. Cela concerne d'abord la formation initiale : reconsidérer l'enseignement du métabolisme, souvent basé sur un apprentissage rébarbatif des voies métaboliques plutôt que sur l'étude des principes d'organisation, de fonctionnement et de régulation des processus métaboliques. Il est aussi important de développer l'offre de formation au niveau du master, et d'intégrer celle-ci dans un cadre qui peut être soit biologique, soit analytique, soit encore bioinformatique. La formation continue doit être également développée, et les opérateurs de recherche doivent être des acteurs importants dans ce cadre.

6. BIOLOGIE SYNTHETIQUE, GENIE BIOLOGIQUE ET BIOMIMETIQUE

6.1. Définition du domaine

La biologie synthétique sera vue ici comme la possibilité de constituer des systèmes biologiques cellulaires ou acellulaires à des fins de recherche ou d'applications qui n'ont pas encore été observés dans la nature et qui sont supposés ne pas pouvoir apparaître comme un simple résultat de l'évolution naturelle. Elle ne se distingue donc pas radicalement du génie génétique qui en constitue une pratique initiale et est comme elle, une ingénierie de systèmes vivants. La biologie synthétique a cependant pour ambition de concevoir des systèmes dont les composants ont des propriétés exhaustivement définies. C'est sans doute l'intégration de la dimension génomique, avec un système de stockage de l'information décrit au nucléotide près qui nous fait entrevoir des possibilités nouvelles pouvant déboucher sur des systèmes radicalement distincts des systèmes biologiques actuels. La biologie synthétique sera étendue ici à l'ensemble des systèmes biologiques élaborés artificiellement en vue d'applications à l'ensemble des biotechnologies dites blanches, même lorsque ces applications ne nécessitent pas *a priori* le franchissement de barrières évolutives. La biologie synthétique concerne aussi les domaines de la santé (techno santé) et de l'agriculture (biotechnologies « vertes ») qui ne seront pas considérés ici. Il s'est aussi avéré utile et cohérent d'inclure dans le domaine certains aspects liés, comme la biocatalyse et les biotransformations. L'inventaire des capacités de transformation chimiques du vivant par biocatalyse (biosynthèses et biodégradations) et la modification de ses capacités par évolution dirigée, brassage de gènes (reshuffling), mutagenèse ciblée demeurent à ce jour au cœur des réalisations et des besoins des biotechnologies blanches. La chimie biomimétique alternative à la biocatalyse a aussi été incluse.

6.2. Principaux défis scientifiques et technologiques

6.2.1. Biologie synthétique

La biologie synthétique se caractérise par une pluralité d'objectifs dans des champs pouvant être très éloignés les uns des autres. Sont souvent mentionnés comme les défis majeurs de la biologie synthétique, l'obtention de systèmes biologiques avec génomes minimaux in fine synthétiques, les protocellules, les systèmes orthogonaux aux machineries de stockage et d'expression de l'information génétique, les circuits et réseaux de gènes, l'ingénierie métabolique et les interfaces entre biologie et nanotechnologies. S'ajoute à ces défis scientifiques, la question particulièrement délicate de l'acceptabilité de la biologie synthétique par la société.

6.2.1.1. Protocellules

L'objectif est ici de créer des cellules synthétiques de novo par conception rationnelle à partir de constituants chimiques organiques et minéraux définis. Ces cellules artificielles devraient in fine être dotées de propriétés d'auto-assemblage, d'auto-réparation, d'autopoïèse voire de capacités d'évolution. Il

est généralement admis que pour atteindre cet objectif, la manière la plus réaliste à ce jour, serait d'héberger dans un contenant physique (e.g.: une vésicule lipidique) des molécules portant l'information (DNA, RNA, PNA) et une machinerie métabolique qui par des processus chimiques fournit l'énergie, régule et régénère les composants cellulaires nécessaires au fonctionnement de l'ensemble.

Il va sans dire qu'il s'agit là d'un domaine particulièrement sensible sur le plan de son impact sociétal et bien que l'on soit très loin du but, les craintes suscitées sont manifestes (voir ci-dessous risques).

6.2.1.2. Génomes minimaux

La réalisation de génomes minimaux, initiée aux Etats-Unis essaie en premier lieu (par approche top down) de définir le nombre de fonctions minimales nécessaires à la vie d'un organisme et d'utiliser de tels génomes comme plateformes d'accueil et d'expression de nouvelles fonctions. Ce genre d'approche se pratique avec les systèmes modèles les plus étudiés (*Escherichia coli*, *Bacillus subtilis*, *Saccharomyces cerevisiae*) ou ayant déjà des génomes fortement réduits naturellement (Mycoplasmes) et s'appuie sur les connaissances étendues disponibles sur ces organismes. Il n'y a que peu de tentatives similaires en Europe. La démarche consiste à supprimer de manière séquentielle des gènes non essentiels en supplémentant le milieu avec les métabolites nécessaires. Certains résultats tendent à montrer que des génomes de taille réduite semblent procurer une plus grande productivité pour des applications industrielles. Mais ces résultats restent controversés et parfois contredits.

La synthèse complète d'un génome illustre l'alternative (bottom up) qui aurait peut-être une plus grande flexibilité pour la mise en œuvre. Le tour de force réussi de la synthèse du génome complet de *Mycoplasma mycoides* suivie de son introduction et propagation dans le contenant cytoplasmique hôte de *Mycoplasma capricolum* avec perte concomitante du génome de la cellule réceptrice apporte la preuve de la validité et de la généralité de cette approche. L'objectif consistant à disposer d'une plateforme minimale prête à recevoir des éléments supplémentaires devient donc un objectif réaliste. Il n'en reste pas moins que le génome réintroduit est une copie quasi conforme du génome original de *M. mycoides* et qu'il reste maintenant à supprimer les gènes inutiles ou à remplacer les gènes essentiels par des gènes de substitution distincts des gènes natifs et optimalement profilés pour l'application envisagée.

6.2.1.3. Systèmes orthogonaux

Les systèmes orthogonaux sont des machineries remplissant des fonctions particulières implantées dans un contexte biologique hôte, qui opèrent de manière partiellement autonome par rapport au contexte. Cet enclavement est aujourd'hui envisagé pour deux types d'usage. (1) Il pourrait être utile pour exprimer une information génétique dont le code est altéré par rapport au code de l'hôte (universel) afin de réserver l'usage d'un ou plusieurs triplets pour d'autres acides aminés non naturels tout en maintenant le code universel pour le reste du système. (2) Il est aussi envisagé de stocker l'information sous forme d'un acide nucléique non naturel pour empêcher sa dissémination en dehors du système hôte. Dans cette utilisation, l'hôte serait capable d'utiliser et de répliquer cette information mais serait strictement dépendant de précurseurs n'existant pas dans la nature et fournis par l'utilisateur. Dans ces systèmes aussi, il y aurait "cohabitation" entre supports naturels et synthétiques de l'information biologique.

La réalisation de systèmes orthogonaux est un objectif à long terme et en est aujourd'hui aux premières étapes de développement.

6.2.1.4. Circuits artificiels de régulation

Des circuits artificiels régulant l'expression des gènes et contrôlés par l'expérimentateur ont été développés depuis une dizaine d'années. Ces circuits synthétiques émulent la dynamique des processus d'expression de systèmes biologiques et sont parfois considérés comme des analogues de circuits électroniques. Des réseaux de gènes artificiels peuvent être conçus à partir de composants génétiques modulaires bien caractérisés, compatibles entre eux et avec le système hôte. Des associations de réseaux dans des réseaux biologiques distincts imitant des écosystèmes ont aussi été expérimentées. Outre les applications potentielles, ils éclairent d'une nouvelle manière le fonctionnement des réseaux de contrôle/régulation naturels. Les premières réalisations illustrent aussi le chemin qui reste à parcourir concernant les possibilités et combinatoires à développer ainsi que les nombreux éléments de régulation biologiques encore inconnus à identifier.

6.2.1.5. Ingénierie métabolique

L'ingénierie métabolique d'un système biologique consiste en la modification de voies métaboliques existantes ou en l'élaboration de voies métaboliques nouvelles pour produire des métabolites non physiologiques ou optimiser la production de métabolites naturels. C'est la pratique la plus ancienne de la biologie synthétique, directement issue du génie génétique. Il existe déjà un nombre significatif de réalisations de ce type. Le domaine de l'ingénierie métabolique n'a été qu'effleuré à ce jour, en partant en général de systèmes peu éloignés des objectifs finaux, notamment un rendement économique suffisant, pour que de petites retouches puissent s'avérer satisfaisantes.

Mais de nouvelles perspectives s'ouvrent aujourd'hui qui visent à introduire des modifications de plus en plus radicales ou de plus en plus subtiles pour ajuster le système aux finalités toujours plus exigeantes du concepteur. Il peut notamment être envisagé de remplacer les voies métaboliques centrales par des voies totalement artificielles. Certains de ces prototypes pourraient par exemple présenter un rendement énergétique supérieur aux systèmes naturels et servir de plateforme pour toutes sortes d'utilisations biotechnologiques.

La plupart des réalisations actuelles de la biologie synthétique reposent sur des raisonnements élaborés tenant compte des connaissances sur les réseaux métaboliques mais restent pour l'essentiel empirique.

Les domaines d'application actuels et futurs sont :

- chimie de synthèse de la chimie fine au vrac en passant par la chimie de base.
- biocarburants/bioénergies
- biomatériaux
- bioremédiation

6.2.1.6. Considérations générales sur la biologie synthétique

La biologie synthétique est souvent présentée comme l'utilisation d'une boîte à outils pouvant être assemblés en vue d'obtenir un système capable d'exécuter une tâche plus ou moins sophistiquée

présentant un intérêt cognitif ou pratique. Cette présentation délibérément simpliste est aussi à modérer par le fait que les outils existants ne sont simplement pas profilés pour les tâches envisagées : nous en sommes à l'âge de pierre de la biologie synthétique.

Les difficultés rencontrées lors des premières réalisations d'ampleur de la biologie synthétique montrent à quel point les défis sont considérables et les moyens proposés pour les surmonter souvent profondément naïfs. L'émergence de cette discipline prédite par la vision de Szibalsky est en fait assez lente du fait de la difficulté à adapter le vivant à la réalisation d'outils comme nous les concevons dans les autres domaines technologiques.

D'une manière générale, on envisage aujourd'hui de réaliser des plateformes (châssis) généralistes multi-usages à partir de systèmes biologiques aux comportements exhaustivement paramétrisés. Cette vision demeure cependant optimiste au vu du degré de connaissance que nous avons des systèmes les mieux étudiés dont les comportements demeurent fortement imprédictibles. La stabilité dans le temps des systèmes reprogrammés demeure aussi un obstacle. La logique d'un système biologique naturel ou même notablement re-profilé demeure celle de la reproduction au moindre coût et un tel système aura toujours tendance à échapper aux contraintes imposées au fur et à mesure des générations. Les constructions actuelles ne représentent que des compromis entre ces deux logiques; le système biologique tentera toujours d'échapper aux contraintes imposées dans la mesure où elles s'opposent à sa logique. Pour surmonter cet obstacle, il est envisagé de développer des procédés répliatifs avec une fidélité supérieure aux systèmes naturels.

6.2.2. Biotransformations et ingénierie enzymatique

Les biotransformations constituent un domaine industriel diversifié et largement établi, qui a oscillé dans le passé entre des périodes de stagnation et d'accélération spectaculaires. On considère aujourd'hui que plus de 500 produits commerciaux sont fabriqués à partir de procédés incluant l'utilisation d'enzymes. Le domaine des biotransformations repose largement sur la biocatalyse qui constitue aussi une solution à une grande partie des défis de la chimie verte. Parmi les avantages nous n'insisteront que sur la stéréosélectivité/stéréospécificité de plus en plus requise dans la synthèse de molécules à activité biologique. Il subsiste cependant des limitations à l'utilisation plus généralisée de la biocatalyse :

- a) la spécificité de la biocatalyse limite le choix des substrats possibles pour un enzyme donné.
- b) le répertoire des réactions enzymatiques existantes reste limitée.
- c) la stabilité et la longévité des enzymes sont limitées.
- d) la productivité volumétrique des réactions enzymatiques est faible notamment en raison de la faible solubilité dans l'eau de certains substrats.
- e) les enzymes sont souvent dépendants de cofacteurs chers, souvent labiles et difficiles à manipuler.
- f) L'étroitesse des conditions réactionnelles est occasionnellement un facteur limitant de rendement.
- g) l'activité des enzymes, optimale en milieu aqueux peut aussi bien se révéler avantageuse que désavantageuse, notamment en cas de faible solubilité dans l'eau du substrat.
- h) Les enzymes sont assez sensibles à des effets inhibiteurs par le substrat ou le produit. i) Le temps de développement d'un procédé biocatalytique ou de son amélioration est trop long.

Il existe de nombreuses possibilités de surmonter les limitations actuelles dans des domaines tels que:

- augmentation du répertoire des activités biocatalytiques : recherche d'activités enzymatiques non décrites à partir de la masse considérable des données de séquence par des approches systématiques et exploitation intensive du phénomène de promiscuité de substrat.
- recyclage des cofacteurs en recourant à des bioconversions in vivo ou par d'autres moyens tels que utilisations de membranes, nanotechnologies)
- amélioration des stabilité, efficacité, versatilité des biocatalyseurs en exploitant plus systématiquement les possibilités de promiscuité de substrat et les techniques d'amélioration par mutagenèse dirigée et brassage génique (gene shuffling)
- génie des procédés enzymatiques : introduction des nanotechnologies, utilisation de solvants supercritiques, nanofibres
- génie des procédés biologiques : ingénierie métabolique des systèmes biologiques

La biocatalyse qui a connu des débuts laborieux et chaotiques a son avenir devant elle.

Les conditions théoriques et pratiques pour des avancées importantes dans ces domaines sont réunies. La biocatalyse devrait donc jouer un rôle déterminant dans la conversion de l'industrie chimique vers des pratiques plus "vertes". Sa migration vers des processus plus in vivo avec le recours de la biologie synthétique est appelée à se développer considérablement.

6.2.3. Biocarburants

Un biocarburant est un carburant produit à partir de matériaux organiques non fossiles, provenant de la biomasse. Par extension la notion de biocarburant pourrait être étendue à la génération d'énergie électrique par des systèmes biologiques utilisant des sources primaires d'énergie renouvelable (lumière en particulier).

Il existe actuellement deux filières principales : l'huile et ses dérivés (biodiesel) et l'alcool fabriqué à partir d'amidon, de cellulose ou de lignine hydrolysés. D'autres formes moins développées voire simplement au stade de la recherche existent aussi : carburant gazeux (biogaz carburant, dihydrogène), carburant solide. On distingue les biocarburants de première et de seconde génération. Parmi les définitions on compte celles qui distinguent les carburants issus de produits alimentaires des carburants issus de source ligno-cellulosique. D'autres définitions reposent sur les moyens utilisés les produire ou distingue ceux issus des cultures agricoles à vocation générique, de ceux issus de cultures à vocation strictement énergétique. Les biocarburants fabriqués à partir de micro algues constituent une « troisième génération car d'un point de vue théorique sont beaucoup plus efficaces à produire que les oléagineux terrestres rendant ainsi envisageable une production de masse sans déforestation massive ni concurrence avec les cultures alimentaires.

L'enjeu pour la biologie synthétique est la synthèse à haute efficacité de biocarburants les plus nobles possibles avec un enjeu particulier en ce qui concerne l'hydrogène ou la production directe de courant électrique. Dans ces deux derniers cas, les systèmes biologiques naturels sont soit peu performants, soit inadapté à une production de masse. La biologie synthétique serait de nature à apporter des solutions

originales pouvant significativement contribuer à ces énergies renouvelables et non polluantes qui sont par ailleurs dénuée d'impact en matière de CO2.

6.2.4. Bioremédiation

- La bioremédiation vise la capture, la défonctionnalisation, la destruction ou le recyclage après transformation de composés organiques ou minéraux. Il peut s'agir par exemple de la décontamination de déchets organiques du sol ou des eaux au moyen d'utilisation de d'organismes naturels ou issus de l'ingénierie. Les contaminants ne sont pas forcément détruits (ce qui peut s'avérer impossible dans le cas de composés minéraux) mais peuvent être simplement capturés ou transformés sous un état moins toxique.

- La bioaugmentation met en œuvre des micro-organismes afin de traiter des zones touchées (sol et eau) par divers polluants carbonés, azotés ou phosphorés. Les microorganismes déjà présents dans les sols ou l'eau, ne sont généralement pas capables de dépolluer, c'est pourquoi l'utilisation de microorganismes extérieurs est nécessaire. Il est souvent nécessaire que des modifications génétiques soient apportés aux souches inoculées, afin d'améliorer ou de permettre la dépollution.

-La phytoremédiation est basée sur l'utilisation des plantes pour dépolluer un sol. En effet, certaines plantes sont capables de fixer dans leurs cellules des polluants présents dans les sols contaminés. Elles ont la propriété d'accumuler et de tolérer des niveaux extrêmement élevés de métaux, par exemple, dans leurs tissus et dans leurs parties aériennes. Dans ce domaine aussi l'ingénierie métabolique peut permettre une amélioration significative des performances.

L'existence de polluant dit peu ou non dégradables (comme les dioxines par exemple) va nécessiter le dessin par biologie synthétique de catalyseurs (enzymes, structures biomimétiques) ou de molécules ou structures de capture adaptés qui n'existent pas forcément dans la biodiversité naturelle.

6.2.5. Biomimétisme

Le biomimétisme vise à mimer les mécanismes et fonction du vivant en mettant en œuvre des composants moléculaires ou des organisations différentes.

L'objectif peut être :

- d'exploiter des mécanismes de la nature pour les appliquer dans différents domaines technologiques.

- de faciliter l'étude scientifique de la nature en reproduisant en laboratoire des comportements autrement noyés dans la complexité du réel.

- d'accéder à de nouvelles structures et interfaces moléculaires capables de reproduire certaines fonctions du vivant

- de s'inspirer de l'organisation et du fonctionnement des êtres vivants pour mieux y intégrer l'organisation et les technologies humaines

6.2.6. Le défi sociétal

Il est souvent considéré que les règlementations établies pour le génie génétique restent valides pour la biologie synthétique et qu'elles sont suffisantes. L'écho médiatique donné à certaines réalisations

comme la réintroduction d'un chromosome entier dans un autre hôte cellulaire ou l'ambition affichée de vouloir reconstituer des systèmes "vivants" à partir de composants "inertes" sont de nature à susciter des inquiétudes généralisées. Il faudrait éviter de se retrouver dans une situation de blocage similaire à celle engendrée par les OGM. A cet égard le terme même de biologie synthétique qu'il faudrait peut-être réserver aux systèmes orthogonaux et aux proto-cellules a tous les atouts pour susciter des inquiétudes rationnelles.

Or il importe d'obtenir un soutien public éclairé, sans lequel il ne sera pas possible d'espérer des financements et un cadre réglementaire raisonnables.

6.3. Etat des lieux

Voir la partie & 2 du document et la figure 14 en particulier.

6.4. Analyse stratégique

6.4.1. Points forts

Il y a très peu de forces dans les tissus de recherche nationaux. Mais cette situation est assez générale en Europe. Les forces reposent sur le savoir-faire et les acquis de quelques équipes académiques qui ont une stature internationale indéniable et des réalisations parfois remarquables. Ces équipes sont surtout présentes dans le domaine de la biocatalyse et du génie métabolique. Il existe une communauté constituée dans le domaine de la biocatalyse. Pour la biologie synthétique il s'agit de quelques initiatives isolées.

Sur le plan des entreprises on trouve les industriels dans la biotransformation, les amidonniers, quelques entreprises de Biotechnologie.

Les formations d'ingénieurs dans le domaine des biotechnologies (Clermont-Ferrand, Toulouse, Strasbourg) et des bioprocédés et de l'agronomie sont de bonne qualité.

On note aussi ça et là un intérêt de quelques chercheurs non-biologistes pour des aspects autour des réseaux de régulation et des circuits artificiels.

Bonne collaboration biochimie-chimie-physique dans le cadre d'un système intégré

6.4.2. Points faibles

- Biologie synthétique

Le domaine de la biologie synthétique est un domaine naissant qui ne s'est pas encore structuré. En France, il n'a pas véritablement attiré de biologistes à ce jour, mais les compétences sont aussi peut-être plus à recruter à partir d'autres domaines.

S'agissant d'un domaine requérant des compétences très diverses, il a du mal à s'établir en dehors de communautés constituées ayant leurs problématiques établies. En France, plus qu'ailleurs, les barrières culturelles entre disciplines scientifiques sont difficiles à franchir.

La formation des biologistes à des problématiques de la biologie synthétique ou même biotechnologiques est quasi inexistante en dehors des écoles d'ingénieurs en biotechnologie.

La plupart des biologistes fondamentalistes reste focalisée par la problématique "comment fonctionne ou dysfonctionne le vivant, comment pallier aux dysfonctionnements, comment améliorer les fonctionnements, etc.". L'utilisation du vivant ou l'imitation du vivant à d'autres fins leur est en général peu familière et ne les attire pas. La technologie de l'ADN recombinant est d'abord devenue un outil de connaissance et le génie génétique appliqué est pratiquement sorti du champ.

Parfois l'application biologique est difficile à évaluer, en particulier sur la mise au point de capsules microniques pour médicaments.

- Biocatalyse

Il existe aussi une grande fracture entre chimistes et biologistes. Alors que l'interface chimie/biologie fonctionne bien en direction de problématiques biologiques susceptibles d'intéresser des chimistes, l'inverse marche plutôt mal.

Il existe une énorme fracture entre biologistes et chimie de synthèse.

La chimie repousse et rebute les biologistes, dont les cursus en chimie sont de plus en plus minimalistes voire totalement indigents. Or une cellule est une entité chimique, avec un foisonnement d'activités chimiques sans égal ailleurs dans la nature.

L'enseignement de la biochimie elle-même est de plus en plus réduit en France (ne pas décourager les étudiants).

A cela s'ajoute au sein même de la communauté des chimistes un scepticisme ancien et bien établi envers la biocatalyse.

- Monde industriel

Les entreprises de biotechnologie du domaine sont peu nombreuses (<10) et n'ont pas la taille critique. Il serait aussi important de réunir des industriels et les chercheurs publics afin de proposer des modes d'interaction qui résulteraient en **la création d'emplois scientifiques dans le secteur privé** en biophysique (avec une utilisation plus astucieuse et encadrée des crédits impôts-recherche).

6.4.3. Opportunités

La biologie synthétique peut se présenter comme une voie à explorer pour nombre de défis industriels, environnementaux de notre époque. C'est un domaine en émergence, il n'est pas trop tard pour s'y lancer.

Certaines formations (écoles d'ingénieurs, écoles de biotechnologie) peuvent facilement se développer et s'appuyer sur des pluridisciplinarités (écoles de chimie, INSA, agro)

De nombreuses technologies (séquençage, ~omiques, génie génétique, nanotechnologies, méthodes analytiques de la chimie, bioinformatique) sont de plus en plus performantes et matures pour être exploitées efficacement.

L'autonomisation des universités est une opportunité unique pour créer des synergies interdisciplinaires

6.4.4. Risques

On observe un manque de réactivité dans la communauté des biologistes. Le domaine a aujourd'hui le vent en poupe avec la publication de résultats spectaculaires qui se veulent autant de preuves de concept, mais les réalisations avec une utilité pratique manquent. Le domaine ne doit pas devenir un champ d'action réservé pour théoriciens de la biologie des systèmes où les applications sont considérées avec condescendance, comme on sait si bien le faire en France (pas seulement certes).

Le cloisonnement des chapelles académiques dans les milieux universitaires et les organismes de recherche, travers bien français est plus qu'une menace.

L'absence de vision claire dans le domaine de la propriété intellectuelle peut-être un frein pour un engagement important des industriels.

La plus grande menace est évidemment la question de l'acceptation sociale, si on rajoute les nanotechnologies au cocktail, on rentre de plain-pied dans le royaume des savants fous.

6.5. Recommandations

Les clés sont d'abord à rechercher dans les formations :

- Lancer une ou deux formations écoles de biotechnologie.
- Renforcer la chimie dans les écoles de biotechnologie. Arrêter de massacrer la biochimie et la chimie dans les formations universitaires de premier cycle de biologie et créer une véritable filaire mixte avec une solide formation de chimie et de physique dans quelques universités
- Constituer une véritable communauté interdisciplinaire couvrant l'ensemble du territoire académique et industriel et ayant des connections étroites avec les acteurs étrangers et en particulier européens. La communauté doit se constituer par le bas mais avec un soutien efficace et affiché et une écoute venant du haut pour éviter une certaine cacophonie. La présence discrète des organismes peut-être un élément rassurant.
- Si la masse critique manque en France se coller le plus possible à des initiatives et projets européens. Certains projets, comme par exemple, la construction de souches sûres et sans risques pour l'environnement demandent des compétences diverses et des approches multidisciplinaires et peuvent être des projets à long terme. Ce genre de projet aurait aussi le mérite de montrer que les risques sont pris en compte et ne sont pas mésestimés.
- Seule une véritable interdisciplinarité opérant au quotidien peut déboucher sur des réalisations significatives. Profiter de la réorganisation du tissu de recherche pour constituer quelques équipes mixtes chimistes, biologistes moléculaires, modélisateurs, génie des procédés, spécialistes de microfluidique et nanotechnologies ...
- Montrer aux chimistes que la biocatalyse alliée à la biologie synthétique peut développer des solutions industrielles crédibles.
- Etre à l'écoute des industriels (pas seulement les chimistes) et savoir ce qu'ils attendent de la recherche académique dans ces domaines. Créer une ou deux tasks forces sur un objectif industriel ou environnemental concret regroupant des industriels et académiques.

6.6. Propositions opérationnelles

1) Commencer par organiser des états généraux de la discipline. Peut-être les concevoir de façon assez large (en incluant délibérément la biocatalyse et la bioremédiation) avec toutes les parties prenantes, Organismes, universités, ANR, industriels etc...Le lancement d'un appel à projets par l'ANR peut aussi aider à identifier les acteurs potentiels.

2) La biologie synthétique connaît des débuts laborieux.

Deux scénarios peuvent être envisagés pour sortir de la situation actuelle.

a) Disposer d'un lieu où on réfléchit en terme de biologie synthétique, autour d'un leader international reconnu. Recruter quelques chefs d'équipes à partir d'un appel d'offre international. Concentrer des moyens à cet endroit sur un ou des projets validés par un comité scientifique international. Une telle opération n'est pas sans risque, mais il paraît difficile de constituer des masses critiques en une pluralité de lieux.

Une coopération étroite des organismes impliqués est indispensable. Ce lieu devrait être un campus universitaire ayant si possible déjà des engagements dans une telle option.

b) L'excellence ne se décrétant pas, on peut alternativement fédérer autour de projets communs avec un financement spécifique fort (et des ressources en personnel) quelques unités de compétences complémentaires pour constituer d'emblée une masse critique délocalisée. L'ensemble pourra alors servir d'attracteur et les noyaux les plus actifs pourraient atteindre une masse critique.

7. AXES TRANSVERSAUX

7.1. La bioinformatique

Le mot «Bioinformatique » remonte aux années 80-90. La bioinformatique est une « interdiscipline » qui couvre le développement de nouvelles méthodes informatisées d'analyse de données biologiques ainsi que leur mise à la disposition des biologistes pour traiter leurs données. Progressivement, le domaine couvert s'est élargi pour devenir immense et est intrinsèquement multi/interdisciplinaire. La bioinformatique ne peut élargir à un seul ITMO et concerne principalement 2 ITMO dans Aviesan. Tout d'abord, dans le cadre de l'ITMO "Génétique, génomique et bioinformatique", il s'agira principalement du développement de méthodes et d'algorithmes concernant le séquençage massif (NGS) et l'analyse à grande échelle des génomes (analyse globale dans la métagénomique), l'assemblage, l'annotation, la recherche et/ou la prédiction de gènes, l'analyse d'épissage alternatif, d'EST et la comparaison de génomes) ; l'analyse de séquences (alignements multiples, recherche de séquences ou de motifs dans les séquences, la classification, la prédiction de fonction ou de localisation) ; la phylogénie (procédures sur données moléculaires, modèles d'évolution, validation statistique, génétique des populations). Ensuite, l'ITMO "Bases moléculaires et structurales du vivant" sera principalement concerné par l'analyse structurale (la prédiction et la comparaison des structures, les études du repliement, le « docking » ou amarrage moléculaire, la conception de nouvelles protéines, l'évolution dirigée des structures, la prédiction de sites fonctionnels) et par les développements méthodologiques en particulier pour l'acquisition et de traitement de données d'expression (microarrays), de protéomique (spectrométrie de masse), les classifieurs, les réseaux de régulation, l'annotation structurale). Enfin, dans certaines domaines et approches tous les ITMO sont concernés de près ou de loin. Citons par exemple la biologie des systèmes (modélisation des systèmes cellulaires, des cascades de signalisation, des réseaux d'interactions, de régulation, voies métaboliques, modélisation de propriétés biologiques), les bases de données utilisées quotidiennement par tout biologiste, les applications cliniques ou épidémiologiques, les données biométriques ainsi que l'écologie et le suivi et la modélisation des populations. Enfin, il faut souligner l'implication d'organismes ou d'instituts CNRS impliqués dans d'autres alliances dans les secteurs de la fouille de données et de textes (extraction de connaissance à partir de textes, l'organisation et distribution des données; les bases de données et leur interopérabilité, les ontologies (entrepôts de données, services Web, intégration de bases, ontologies biologiques), les outils de visualisation les grilles de calculs et le « cloud computing ».

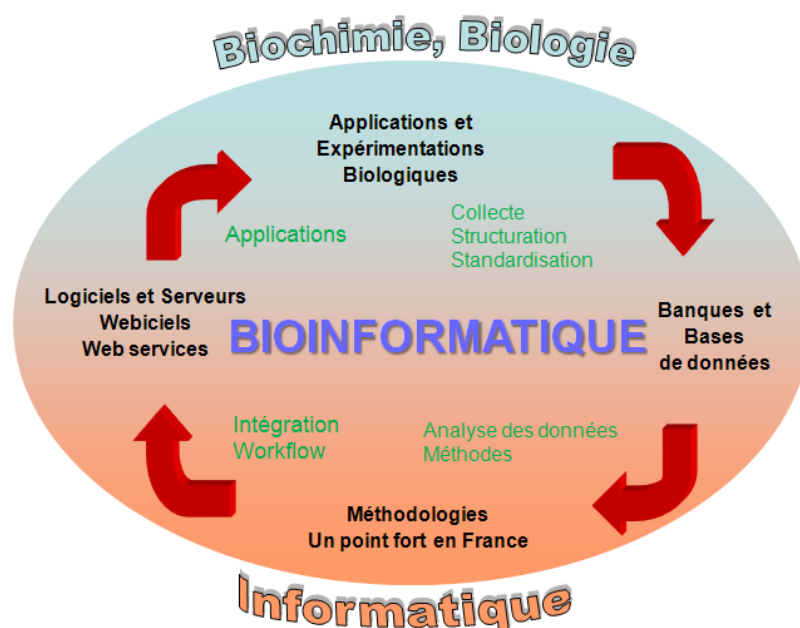
Le problème aigu du stockage aujourd'hui et demain...

Les données biologiques sont très particulières, elles sont hétérogènes par nature ce qui pose le problème de leur organisation. Leur qualité est souvent médiocre ce qui induit la propagation des erreurs par les systèmes automatiques. La sémantique et la représentation des données ne sont pas uniques (par exemple la notion de gène dépend de qui la présente) et le format est souvent associé aux producteurs de données ce qui pose le problème des standardisations et des vocabulaires contrôlés. De nouveaux types de données émergent constamment ce qui pose le problème de leur intégration globale. Enfin, et non des

moindres propriétés, le volume des données est en pleine expansion. Par exemple le programme 1000 génomes humains séquencés du « Welcome Trust» a généré en 6 mois l'équivalent de 20 fois le contenu en séquences amassé depuis 30 ans dans GenBank (100Gbases)! Cette complexité (et cette mouvance qui pourra atteindre 100 Gb par semaine voire plus encore) en fait un domaine difficile à appréhender pour l'informatique et donne naissance à des formalisations complexes et/ou à des « problèmes combinatoires insurmontables ». Ainsi il est prévu pour 2020, l'équivalent de 1000 racks de stockage à EBI.

Un cercle vertueux fragile...

Le cercle vertueux de la bioinformatique (représenté ci-après) peut s'appliquer à tous les domaines et objets de la biologie (voies métaboliques, organisme, cellules, molécules, protéines, génomes, séquence, images, structures, protéomes, ligands, etc.). Dans ce cercle, dont le point de départ est l'expérimentation biologique, il apparaît clairement l'intrication de la recherche, de l'ingénierie logicielle, des bases de données, du « carrossage » indispensable des outils développés et de la connaissance biologique. En bioinformatique, aucune activité de ce cercle est plus noble qu'une autre, tout est utile et surtout indissociable. Une erreur fondamentale pour Aviesan serait de considérer que seule l'application compte parce qu'elle représente le besoin. Ainsi, vouloir limiter la bioinformatique aux applications biologiques, c'est prendre le risque de ne pas savoir traiter les données de demain faute d'une recherche fondamentale en bioinformatique forte aujourd'hui. A l'opposé, considérer que la recherche en bioinformatique est pour partie le domaine réservé de l'Informatique c'est prendre le risque de développer des outils non adaptés (ou même carrément inutiles). En revanche, la recherche en informatique peut prendre comme modèle la biologie, ainsi cette « informatique bioinspirée » qui se nourrit de la complexité des processus biologiques comme les réseaux de neurones ou les algorithmes génétiques. Le concept de « programmation biologique » a d'ailleurs été illustré dès 1994 par les travaux de Leonard et Adleman qui ont réalisé des calculateurs biologiques basés sur l'appariement des bases ADN pour résoudre un problème NP complet. Il faut considérer la bioinformatique aujourd'hui comme une interdiscipline voire demain comme une discipline qui prendra tout son sens dans la biologie des systèmes, la biologie synthétique et les « omics ».



Une structuration réelle de la recherche en bioinformatique...

La recherche en Bioinformatique s'effectue dans nombre d'organismes de recherche : universités, CNRS, INRA, INRIA, INSERM, CEA, Institut Pasteur, Institut Curie, etc. Depuis 2003, l'ensemble de cette recherche en bioinformatique est animée et fédérée dans le GDR 3003 "Bioinformatique moléculaire" (www.gdr-bim.u-psud.fr) qui fédère un grand nombre des équipes bioinformatique de tous les organismes présents dans AVIESAN (et au-delà), dans les différentes disciplines : Biologie, Informatique, Mathématiques, Physique principalement. Ce GDR comprend 106 équipes, 1020 personnes dont 540 permanents. Ce GDR vient d'être renouvelé en 2009. Une difficulté récurrente en bioinformatique est l'évaluation décalée de chercheurs vraiment interdisciplinaires (souvent d'excellente qualité) dont les sections ou commissions dont ils dépendent n'ont pas la compétence et/ou l'ouverture nécessaire pour une évaluation pertinente. Ces sections ou commissions ont d'ailleurs beaucoup de difficulté à effectuer ces recrutements tant il reste vrai qu'une section favorise naturellement le cœur de sa discipline, laissant volontiers le recrutement de « l'interface » à l'autre partenaire ou section. La motivation de chercheurs interdisciplinaires pas toujours reconnus a quand même permis l'irrigation des laboratoires de recherche par des chercheurs qui ont répondu à un besoin d'échange de compétences entre des disciplines parfois éloignées pour faire avancer la science sur quelques sujets pointus relevant de la biologie. Ce besoin demeure d'actualité et une procédure de recrutement et d'évaluation de chercheurs interdisciplinaires en bioinformatique ayant fait ses preuves reste encore à mettre en place. Cette communauté de bioinformaticiens se réunit depuis 2000 chaque année au cours de JOBIM (Journées Ouvertes de Biologie, Informatique et Mathématiques) qui réunit environ 500 acteurs (chercheurs et doctorants) du domaine dans une ville différente. Pour le 10^{ème} anniversaire, cette réunion revient dans sa ville de début (Montpellier).

La France bien placée...*

Les USA produisent 57% des publications en bioinformatique de l'échantillon. La France occupe le 4^e rang mondial juste devant le Japon et le Canada, derrière l'Angleterre (siège de l' « European Bioinformatics Institute ») et l'Allemagne (Heidelberg - EMBL).

Place de la France

Exclusif	France	Allemagne*	Angleterre*	Suisse	Japon	Espagne	Canada	USA	Total
Publications ¹	393	829	713	101	310	189	289	3749	6573
Point H	36	50	59	22	32	28	29	103	-
Citations	5429	12315	17359	2608	5210	7062	3959	92705	146647

*EMBL

** EBI

Non exclusif	France	Allemagne	Angleterre	Suisse	Japon	Espagne	Canada	USA	Total
Publications	542	1120	1026	185	400	297	417	4268	8255
Point H	43	58	67	27	36	34	36	109	-
Citations	10707	19731	26777	3630	11196	8548	6381	106989	193959

**Il est très difficile d'estimer ou de fournir des indicateurs de performance pour la recherche en bioinformatique qui est une discipline d'interface. Afin de mesurer l'impact et le positionnement relatif en bioinformatique, la procédure suivante a été utilisée. Deux revues seulement ont été prises en compte : « Bioinformatics » (IF 2008= 4.328) et « BMC Bioinformatics » (IF 2008 =3.781) pour toute l'étude. Il en*

ressort une très large sous estimation des valeurs collectées mais ce choix est justifié par le fait que même si beaucoup de journaux publient des articles relevant de la bioinformatique (Nucl. Acids Res. J. Proteome Res., Proteins, Nature Structural Biology,...), ces 2 journaux font référence dans le domaine et il n'y est publié a priori que des travaux de bioinformatique. Et l'objectif ici n'est que de comparer différents pays. Le nombre de publications exclusif (par exemple la France seule) ou non exclusif (par exemple la France avec d'autres pays) a été calculé ainsi que le nombre de citations et le point H. (source Juin 2010.)

Plateformes bioinformatiques

Les plateformes bioinformatiques labellisées par le comité RIO puis IBiSA sont fédérées au sein d'un réseau RENABI (Réseau National de BioInformatique) qui pourrait se structurer administrativement et constituer le nœud national français d'ELIXIR. Sous l'impulsion d'IBiSA, le réseau RENABI a vu sa mission s'élargir pour une fédération englobant les projets émergents et les bases de données spécialisées. Ainsi, les acteurs de la bioinformatique qui sont immergés dans les laboratoires de biologie sont associés aux demandes de financements, aux formations spécifiques et aux animations mises en place par les plateformes. Avec l'appui d'IBiSA (Infrastructures en Biologie, Santé, Agronomie), le réseau a entamé une restructuration des plateformes selon le schéma matriciel géographique/thématique. Une restructuration a été proposée en 2009 autour de 7 PF « majeures » dont le périmètre peut être élargi.

Rôle des plateformes

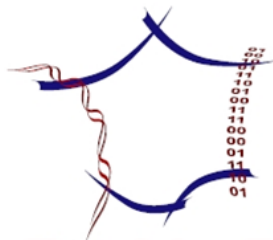
1. La mise à disposition de ressources génériques comme la diffusion de logiciels, de base de données, l'accès web ou authentifié avec la nécessité de maintenir (mise à jour, implémentation et intégration, sauvegarde, sites miroirs) et d'exploiter une infrastructure adaptée.

2. La réalisation de prestations à façon avec notamment une activité d'ingénierie dédiée sur projets. Pour cela, il faut souligner la nécessité d'un comité scientifique et celle d'« environnémenter » les CDD recrutés (ANR, PCRD, autres...) par du personnel statutaire sur les plateformes labellisées. Enfin, il faut noter la difficulté de couvrir des champs thématiques très larges.

3. Le maintien et le développement de ressources originales de très forte visibilité et d'intérêt collectif :

Ces développements assurent la visibilité et la spécificité des plateformes.

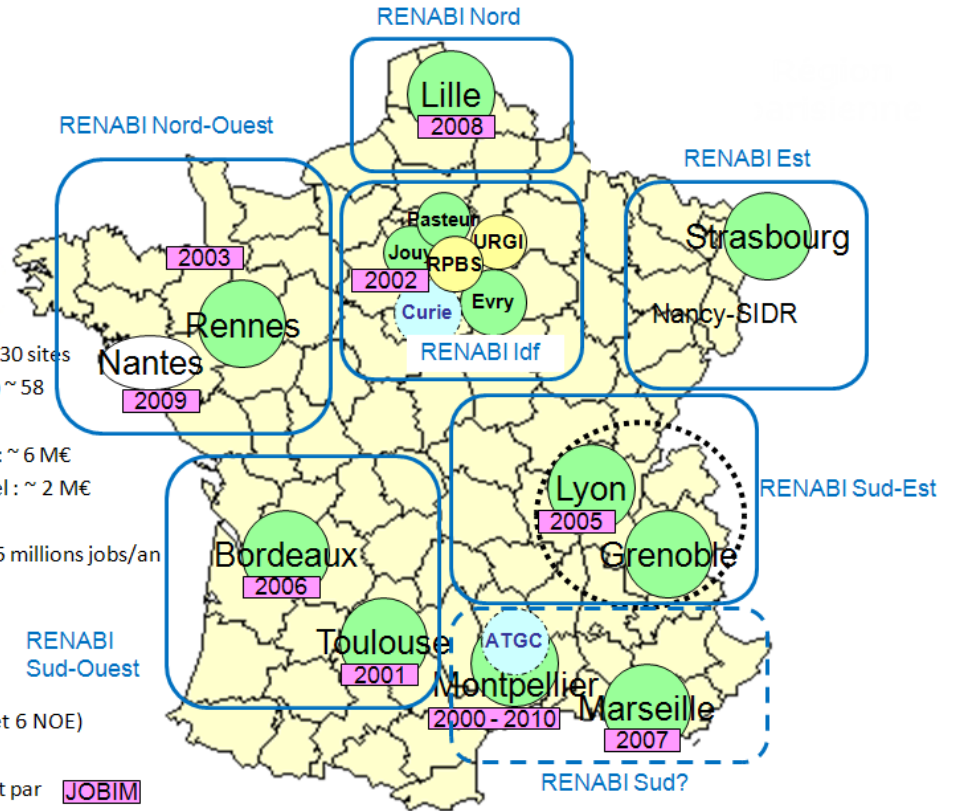
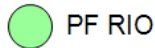
Le paysage des plateformes bioinformatiques est organisé autour des grandes régions.



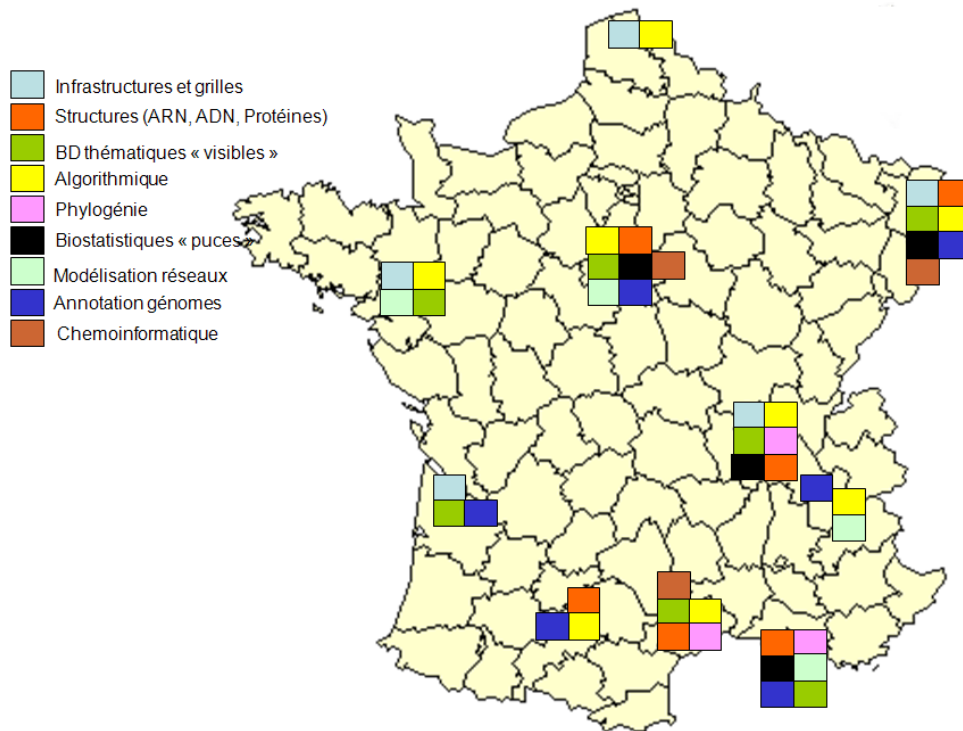
ReNaBi

- ✓ 7-8 PF RENABI 12-14 PF RIO ~25-30 sites
- ✓ Permanents (labos de recherche) ~ 58
- ✓ Non permanents ~ 39
- ✓ Coût annuel de fonctionnement : ~ 6 M€
- ✓ Coût annuel d'équipement annuel : ~ 2 M€
- ✓ BD européennes ~ 15
- ✓ Utilisation de serveurs Web: ~ 2,5 millions jobs/an
- ✓ Publications 2006-2007 >120
- ✓ Applications 2006-2007 >1500
- ✓ Thèses soutenues : 25
- ✓ Doctorants : 45
- ✓ Projets européens 25 (dont 5 IP et 6 NOE)

500 personnes réunie annuellement par **JOBIM**
 Journées Ouvertes Biologie Informatique Mathématiques



En plus de l'organisation régionale, les plateformes ont chacune une panoplie de compétences et d'expertises qui doivent absolument être mutualisée. Un premier bilan des expertises thématiques a été effectué début 2009 et est destiné à évoluer. Cette distribution thématique viendra se superposer à la distribution géographique.



Les plates-formes majeures doivent être dotées de ressources humaines pérennes et non pas fonctionner sur une succession de CDDs. En effet, la capitalisation de l'expertise est vitale pour être en mesure de faire face à la quantité de données produites par les nouvelles technologies de séquençage et autres « ~omics ».

Deux types de demandes pourraient être soumis à IBISA :

1) Pour les PF majeures, des demandes «classiques» récurrentes : équipement en adéquation avec les projets de la PF majeure, CDD si besoin, et surtout des postes statutaires dont la demande remonte aux tutelles : au minimum 3 ETP devraient être injectés pour assurer le fonctionnement d'un seul site.

2) D'autres projets scientifiques transverses vont recouvrir un domaine thématique ou des « compétences ». De tels projets nécessitent une recherche « technologique » et le développement des applications correspondantes. Ils peuvent faire l'objet d'un soutien financier par le GIS IBISA (équipements et CDD « méthodologiques »).

Infrastructures

Une réunion récente (30 avril 2009) sur la bioinformatique avec les différents représentants des organismes et des acteurs du domaine a fait apparaître la difficulté d'identifier les infrastructures nécessaires. Aucune infrastructure unique ne peut satisfaire actuellement la demande en biologie. Au niveau des calculs « haute performance », il ressort que les très grands équipements sont peu adaptés à la biologie qui a plus besoin de « cluster de calculs » que de machines vectorielles. Dans ce contexte, devant l'explosion combinatoire des traitements à effectuer, on peut imaginer une « parallélisation » par les données (distribution, grille de calculs). Ainsi, l'initiative « Institut des grilles » du CNRS constitue une infrastructure matérielle mais on peut facilement prédire que ces architectures (encore non opérationnelles en biologie) atteindront leurs limites et ne seront adaptées qu'aux besoins de calculs très massifs et très répétitifs. De plus, la granularité incertaine et inégale des calculs rend difficile la prévision et la distribution des calculs (la gridification des algorithmes est difficile et ne concerne pas que BLAST ou FASTA!). Par ailleurs, au niveau des données, les « grilles » sont très loin d'être fonctionnelles et sont encore peu adaptées à la biologie. Dès lors, de nouveaux algorithmes de déploiement des données et des heuristiques de traitements restent à découvrir et à développer. En attendant, il convient de doter les laboratoires de biologie qui s'impliquent dans la bioinformatique de moyens informatiques (et des ressources humaines afférentes) suffisants pour intégrer les méthodes et réaliser les interfaces web et les « web services » indispensables aux biologistes « end user » pour atteindre les données. Ces moyens sont complémentaires des grands centres de calculs qui auront la mission d'offrir la capacité de traitement et de stockage à grande échelle, le « scale-up » pour des analyses liées à la biologie massive comme le séquençage massif, les projets de métagénomique ou de « médecine prédictive ». Dans ces domaines, les besoins de stockage vont augmenter plus rapidement que les capacités de stockage. Une architecture distribuée sur un réseau est peut-être une solution envisageable à condition de résoudre les problèmes spécifiques posés par les données biologiques. Il s'agit là d'un défi pour l'Informatique.

Une organisation rationnelle de l'infrastructure se décline grossièrement en 3 couches : le niveau centre national, le niveau des plateformes de bioinformatiques et enfin l'utilisateur final i.e le biologiste. La couche des centres nationaux n'est en général pas visible pour le biologiste utilisateur. Cependant, elle peut parfois lui être nécessaire. Pour cela, il faut que les plateformes de bioinformatique servent

d'interface. Considérer qu'augmenter les équipements de calcul et de stockage sur les grands sites sans prévoir leur interface avec les plateformes bioinformatiques revient à augmenter le nombre de voies sur une autoroute sans augmenter le nombre de bretelles d'accès. De même, il faut optimiser les algorithmes (portage de code, etc..) et réguler de manière pertinente les flux pour ne pas se retrouver sur une autoroute à 6 voies avec un vélo ou à l'inverse un 1er août sur une route départementale parce que l'autoroute est devenue inaccessible.

La standardisation et l'interopérabilité

De même, comme souligné dans l'introduction, la standardisation et l'interopérabilité des bases de données et des systèmes est un enjeu majeur dans la biologie moderne. A ce propos, l'initiative SIDR (Standards-based Infrastructure with Distributed Resources) qui mobilise les compétences de l'INIST à Nancy devrait, à terme, permettre d'équiper les laboratoires avec un système d'échange d'informations qui leur soit adapté. L'organisation des données du projet prend en compte les deux grandes stratégies qui existent pour la collecte et l'annotation des données, celles-ci consistant soit à construire de très grandes bases de données qui ne se concentrent pas sur des domaines biologiques en particulier, soit à produire des bases de données de petite taille, hautement spécialisées qui contiennent des données pointues avec leurs spécifications. Ces bases de données spécialisées présentent un intérêt majeur pour les personnels de recherche qui les élaborent. Cependant, si on les construit indépendamment les unes des autres, elles risquent de comporter d'importantes zones d'intersection qui ne peuvent être utilisées en réalité parce que ces bases ne sont pas standardisées. Voilà l'objectif principal du projet SIDR : construire des bases de données à la fois hautement spécialisées et interopérables. Pour cela, il est judicieux de s'emparer de la question de la gestion informatisée des données biologiques produites par les laboratoires Français de recherche.

Le défi des LIMS...

Il est urgent de mettre en place des LIMS (Laboratory Information Management System) associés au cahier de laboratoire électronique qui permettront, à terme, à tous les laboratoires d'alimenter de manière pertinente les bases de données avec des données de qualité. Il n'y aura pas de grand projet de biologie systémique (ou intégrative) ou de médecine personnalisée sans la mise en place de ces outils devenus essentiels à la Biologie. La France (à travers Aviesan) ne fera pas l'économie d'une réflexion prospective sur les moyens que les organismes devront mettre en place ensemble pour évaluer, adapter, acheter ou voire développer les systèmes de gestion d'expériences en biologie. En investissant dans ce secteur, la recherche française en biologie sera plus performante et fera des économies d'échelle.

En conclusion, la Bioinformatique en France est forte, structurée tant au niveau de la recherche que des plate-formes et concerne tous les acteurs de la biologie.

7.2. Modélisations mathématiques et bases moléculaires du vivant

Les besoins d'outils mathématiques en BMSV relèvent d'aspects variés suivant l'activité des équipes. On peut les classer en quatre grands thèmes qui relèvent souvent aussi de l'informatique et de la physique. Ils forment la méthodologie générale de la « modélisation mathématique ».

- Analyse des données (outils statistiques, analyse et traitement des images et des signaux)
- Modélisation des processus (déterministe ou non)
- Simulation numérique (qualitative ou prédictive)
- Optimisation ou contrôle de ces processus

La compréhension quantitative des données permet de modéliser le processus (le réduire à des éléments principaux permettant l'expliquer les observations). L'objet est alors un modèle mathématique (système algébrique ou logique, équations différentielles, processus stochastique, équations aux dérivées partielles...) que l'on peut simuler sur ordinateur et reproduire facilement afin de le confronter aux expériences. Lorsque le modèle est suffisamment riche et prédictif, il peut servir à optimiser le processus biologique ou éventuellement le contrôler dynamiquement.

7.2.1. Spécificité de l'IBMSV en besoin d'outils de modélisation

Les spécificités de cet ITMO concernent les champs d'application où, même si la molécule joue un rôle central, les échelles sont multiples (de l'atome à la cellule), les données hétérogènes (structurales, fonctionnelles). Tout comme la chimie computationnelle est un domaine bien établi, la biologie moléculaire computationnelle émerge. Certains aspects de la biologie synthétique pourraient nécessiter de 'dégrossir' avec différents outils de simulation les performances (rendements) d'organismes (ou de ses composantes). On peut citer les champs d'applications suivant qui apparaissent dans les différents textes définissant cet ITMO

- Biologie moléculaire computationnelle
- Réseaux de régulation moléculaires et leur dynamique
- Organisation spatio-temporelle de la molécule
- Organisation spatio-temporelle des molécules dans la cellule
- Analyse statistique des données (~omiques), leur structuration et la recherche d'information pertinente

Mais il faut insister sur le fait qu'avoir une approche 'modélisation' relève souvent plus de la méthodologie, de l'état d'esprit d'un domaine, que de la nécessité immédiate. L'enjeu est donc de préciser domaine par domaine, des besoins dans la boîte à outils existante et d'en développer de mieux adaptés au domaine des BMSV.

7.2.2. Outils de modélisation

Les outils de modélisation sont nombreux et sans doute encore mal définis. Il s'agit d'un enjeu sur le long terme d'arriver à les préciser et à en développer de nouveaux dans le cadre de cet ITMO. On peut les classer en grands thèmes :

- Stabilité, bistabilité, multistabilité dans les systèmes et réseaux, robustesse
- Rôle de l'aléatoire dans le fonctionnement des réseaux (théorie des catastrophes ?)
- Equation master de la biochimie et algorithmes de Monte-Carlo (Gillepsie)
- Dynamique biomoléculaire (*ab initio*, réductions)
- Equations différentielles et aux dérivées partielles pour la concentration moléculaire, la probabilité de présence (spatialisées)
- Analyse multi-échelle et asymptotique

Les spécificités inhérentes au monde biologique proviennent de la difficulté de quantifier les valeurs pour les modèles (constantes cinétiques, compositions biochimiques) et surtout de leur variabilité intrinsèque. Pour ces raisons, un modèle donnant une réponse qualitative est souvent un progrès intéressant en biologie et l'obtention de modèle prédictif est plutôt rare.

Les propriétés de stabilité (bistabilité, multistabilité) sont fondamentales dans les modèles relevant de cette ITMO. En effet, compte tenu de la variabilité, ils doivent être robustes¹. D'une part les coefficients ne doivent pas influencer les propriétés qualitatives des solutions (stabilité structurelle). D'autre part elles ne doivent pas dépendre trop fortement de l'état initial. Ceci impose que les états permanents (équilibres stables en général) soient bien séparés et suffisamment nombreux pour rendre compte des comportements observés. C'est pourquoi les systèmes bistables jouent un rôle central et sont largement observés dans les réseaux de régulation biomoléculaires².

Le bruit et les effets stochastiques jouent un rôle important dans les transitions d'un état stable à l'autre et le comportement global de la cellule et de la population. Il peut s'agir d'un résultat évolutif pour échapper aux « pédations ». Ces bruits sont souvent d'origine environnementale mais peuvent aussi être générés par le réseau d'interaction RNA/protéine/ADN lui-même³. Dans les réseaux de régulations moléculaires, on observe souvent une grande variété de protéines et, de ce fait, un faible nombre de protéines d'un type donné et dont la répartition spatiale est aléatoire. C'est aussi une cause de stochasticité qui motive l'utilisation des « chemical master equation », des algorithmes de Gillespie et des nombreuses tentatives pour aller au delà.

7.2.3. Recommandations/propositions

7.2.3.1. Défis

La création d'une véritable interdisciplinarité s'appuie sur l'échange de culture et de méthodologie entre sciences biologiques et physique/informatique/mathématique, etc. L'un des enjeux fondamentaux est de trouver des structures de recherche permettant l'émergence de cet échange. Des individus 'modélisateurs' isolés dans des équipes de biologistes n'est certainement pas une solution viable et il faut donc fédérer ces chercheurs. On peut donner de nombreux exemples à l'étranger où des centres interdisciplinaires ont été conçus (BioQuant à Heidelberg, OCCAM à Oxford, MathBiosciences Institue à Ohio state University...).

7.2.3.2. Opportunités

Contrairement au monde anglo-saxon, la France n'a pas une tradition d'interdisciplinarité mathématique/biologie mais un développement rapide peut s'observer ces dernières années (l'INRIA y a joué un rôle très important). Les encadrements doctoraux sont maintenant établis dans ce domaine. Ils resteront toutefois limités si les débouchés industriels ne se développent pas.

¹ Gunawardena J (2010) Systems biology. Biological systems theory. Science 328: 581-582

² Tian T, Burrage K (2004) Bistability and switching in the lysis/lysogeny genetic regulatory network of bacteriophage lambda. J Theor Biol 227: 229-237.

³ Schultz D, Ben Jacob E, Onuchic JN, Wolynes PG (2007) Molecular level stochastic model for competence cycles in Bacillus subtilis. Proc Natl Acad Sci U S A 104: 17582-17587.

7.2.3.3. Recommandation

Envisager la création d'une structure d'échange interdisciplinaire permettant de développer des programmes de recherche et de rencontres entre chercheurs issus de différentes disciplines.

7.3. Grands équipements relevant du domaine BMSV

Les grandes installations pluridisciplinaires jouent un rôle majeur en ce qui concerne l'apport des méthodes de la physique et de la chimie en biologie, en rapprochant en un même lieu les diverses communautés. En moins d'une vingtaine d'années, ces infrastructures sont devenues un outil majeur indispensable pour la communauté BMSV relevant essentiellement des domaines 2.2 (structure) et 3.2 (~omique). Les demandes sont souvent gérées collectivement par « BAG » (Block Allocation Group) et présentées dans le cadre d'appels d'offres compétitifs. L'attribution d'un temps d'utilisation de ces infrastructures est entièrement environnée (accueil, voyage et utilisation de l'infrastructure).

Les 9 BAG français :

1. Marseille (Sulzenbacher)
2. Lyon/Toulouse/Bordeaux (Dautant)
3. Lille/Orléans/Roscoff (Czjzek)
4. Downtown Paris (Benas)
5. Institut Pasteur (Bentley)
6. Gif/Orsay "campus" (IBBMC, LEBS, ISV and INRA) (Graille)
7. CNRS(Gif)-CEA(Saclay)-Curie(Paris) (Bressanelli)
8. Strasbourg/Nancy/Palaiseau (Moras)
9. IBS Grenoble

Généralement, les coûts indirects liés à l'utilisation de la mise à disposition de ces grands équipements financés au niveau national ou européens sont mal pris en compte par les communautés qui en bénéficient, qui n'intègrent généralement pas ces « avantages » dans leur coût de fonctionnement d'équipe.

Les grandes infrastructures de recherche analytiques que sont les synchrotrons et les sources de neutrons sont aujourd'hui des outils indispensables à la détermination structurale des molécules biologiques et sont utilisées par de très nombreuses équipes de ce domaine. De plus, de nouvelles techniques liées à la diffusion aux petits angles et à la dynamique sont aujourd'hui très utilisées pour l'analyse multi-échelle aussi bien temporelle que spatiale des systèmes biologiques.

7.3.1. Sources de lumière synchrotron

Le paysage français des synchrotrons est organisé autour du synchrotron SOLEIL (St Aubin, Essonne), qui sera sous peu complètement opérationnel et de l'European Synchrotron Radiation Facility (ESRF) instrumentation internationale dont la France est l'un des pays membres. Le domaine BMSV représente environ 25% des utilisations de ces synchrotrons. Le programme européen (ELISA) permet à la

communauté française de financer ses participations aux autres synchrotrons européens comme DIAMOND en Grande-Bretagne, DESY en Allemagne, ELLETRA en Italie, PSI en Suisse, etc...

L'ESRF résulte d'une coopération scientifique européenne entre dix-neuf pays (plus Israël) qui participent conjointement au financement et à la maintenance de cette source de rayons X. La contribution de la France au budget annuel de l'ESRF est de 27.5%, cette contribution étant également partagée entre le CNRS et le CEA. Le rayonnement synchrotron de l'ESRF se caractérise par une forte brillance et des possibilités remarquables pour l'étude des macromolécules biologiques. L'ESRF dispose d'un panel de lignes de lumière dédiées à la biologie structurale comme ID14-1, ID23-1 et ID29, trois lignes de lumière très intense dont la longueur d'onde est ajustable ; ID14-1 et ID14-2, deux lignes de lumière à longueur d'onde fixe ; ID23-2, ligne bien adaptée aux petits cristaux avec son faisceau « microfocus » (10 µm) et ID14-3, la ligne BIOSAXS ouverte depuis octobre 2008 pour les études en solution aux petits angles. A celles-ci, il faut ajouter la ligne CRG française BM30A (alias FIP pour « French beamline for investigation of proteins ») qui peut explorer une large gamme de longueurs d'ondes et offre des outils novateurs pour l'automatisation des expériences.

Dans les mois qui viennent, l'ESRF va se réorganiser pour augmenter sa puissance. Dans ce projet, les lignes pour la biologie structurale seront regroupées géographiquement au sein d'un « village ». C'est le projet UPBL10/MASSIF qui comprendra trois faisceaux dédiés à des études à haut débit et hautement automatisées. Sur les lignes opérationnelles depuis une douzaine d'année, l'ESRF réserve environ 400-500 « shifts » à des utilisateurs Français regroupés en « Block Allocation Groups » (BAG) (8 en France), pour une durée de temps de faisceau de 8 h/shift. Chaque shift représente un coût d'environ 4 k€ et 3 projets sur 5 sont financés. Indirectement, la communauté BMSV est donc collectivement subventionnée à hauteur de 1,6 à 2 M€/an. Le nombre de demandes à l'ESRF décroît naturellement depuis la mise à disposition récente et progressive du synchrotron français SOLEIL à Saclay. Ce synchrotron, installé sur le plateau de Saclay, propose plusieurs lignes déjà dédiées à la biologie structurale : Proxima-1, SMIS, SWING et DISCO. Pour l'heure, ce synchrotron monte encore en puissance et la ligne Proxima-2 dédiée aux microcristaux sera opérationnelle début 2011.

Tableau 14 : SOLEIL

Demandes par ligne*	Projets soumis	Projets acceptés	Shifts demandés	Shifts alloués
DISCO_Imaging	14	11	171	129
DISCO_APEX+SRCD	22	15	261	132
Proxima 1 STD cycle courant	13	10	42	23
Proxima 1 BAG cycle courant	4	4	57	15
Proxima 1 BAG cycle précédent	10	10	120	111
SMIS	30	24	363	230
SWING	58	28	311	141
SAMBA (EXAFS)	49	15	624	144

*Période Juillet-décembre 2010

7.3.2. Sources de neutron et laser

La situation est la même pour la diffraction de neutrons qui permet des analyses en faisant varier le contraste et ainsi annuler le signal de certaines phases du mélange. Le laboratoire Léon Brillouin, source nationale est complété par l'ILL, source européenne dont la France est membre avec les anglais et les allemands. Un programme européen NMI3 joue un rôle similaire à ELISA. Une synergie grenobloise existe entre les sources synchrotron et de neutrons, regroupés en un même lieu, avec l'EMBL.

Pour ce qui concerne les lasers à rayons X, la France participe à la construction du laser Xfel (fel : free electrons lasers) à Hambourg qui devrait offrir de nombreuses opportunités nouvelles pour les biologistes français. Ainsi, dans le cadre du PEPS SASELELX : (Soutien aux Activités Scientifiques Françaises autour des Lasers à Électrons Libres Émettant des Rayons ; <http://www.cnrs.fr/inp/spip.php?article225>), l'INstitut de Physique (INP) du CNRS, en partenariat avec l'INstitut de Chimie (INC), lance un appel ouvert à projets destiné à faciliter l'engagement de la communauté scientifique française dans des actions de recherche autour des lasers à électrons libres dans le domaine des rayons X. Cet appel est ouvert aux chercheurs de toute discipline.

7.3.3. Sources de champs magnétiques intenses pour la biologie structurale

7.3.3.1. La RMN

La TGIR-RMN est le premier TGIR impliquant une activité dans le domaine de la biologie de type BMSV. Il s'agit d'une structure distribuée composée d'équipes implantées dans des laboratoires répartis sur le territoire national où sont installés des spectromètres RMN tous munis d'aimants pompés allant de 750 MHz WB (Wide-Bore) à 1 GHz. Leur acquisition a été cofinancée par le CNRS, opérateur national traditionnel du domaine. Dominique Massiot est directeur de cette fédération sans mur (FR 3050). Les implantations des spectromètres très hauts champs du TGIR concernent Lyon (CRMN), Grenoble (IBS), Lille (GSF), Orléans (CEMHTI), Gif (ICSN), Bordeaux (IECB). Chacun des sites s'est engagé à proposer un accès à la communauté (INC, INSB, CEA, Universités,...) à hauteur de 30 % du temps d'expérience sur la base d'appels d'offre permanents avec décision rapide. Cet accueil combine expertise et instrumentation (voir : <http://www.cnrs.fr/inc/recherche/tge/reseaurmn/accesTGE.htm>). Les activités liées au domaine BMSV relèvent généralement de l'analyse structurale de macromolécules par RMN liquide ou solide. Dans ce contexte il est important de noter qu'un spectromètre 800 MHz sera installé bientôt au LISBP (unité mixte INSIS-CNRS/INRA/INSA/UPS) pour une activité dédiée à la métabolomique. Avec le CRMN à Lyon, ce constitue sera le second à présenter ce type de spécificité en France et pourrait intégrer le TGIR à terme. Le budget de ce TGE qui regroupe investissements et fonctionnement, depuis 2005, est en moyenne de plus de 2,5 M€/an au CNRS. Il faut noter que ce budget ne couvre environ un tiers du budget total nécessaire, budget complété par d'autres co-investisseurs (CPER, FEDER, CEA, Ministère, INRA, universités).

Enfin, depuis 6 ans le Centre RMN de Lyon et l'IBS à Grenoble sont intégrés dans une infrastructure européenne intégrée de type « I3 » regroupant en outre les centres RMN historiques situés à Florence et

Francfort. Le renouvellement est actuellement en cours sous forme de contrat BioNMR. Un éclaircissement concernant le fonctionnement des réseaux européens et du TGIR RMN est souhaitable.

7.3.3.2. La FT-ICR

Un projet d'association des laboratoires disposant de spectromètres de masse de type FT-ICR (aimant de 7 à 18 T) est en cours de montage dans le cadre d'un TGE du CNRS (6 sites). Dans le domaine BMSV, quatre sites présentent une activité dans le domaine « ~omique ». Trois d'entre eux sont localisés dans la région parisienne (Ecole Polytechnique, ESPCI et UPMC), le dernier étant localisé sur le site lillois de Villeneuve d'Ascq. Si chacun de ces quatre sites affiche des compétences dans le domaine de la protéomique, le site parisien de l'UPMC se caractérise par les développements qu'il réalise dans le domaine de la métabolomique en collaboration avec le site du CEA de Saclay. Ce réseau pourrait inclure les plateformes IBiSA disposant de technologies hybrides de type Orbitrap.

7.3.3.3. Recommandations

Les liens scientifiques avec les grands instruments – en particulier les synchrotrons de Saclay et Grenoble - doivent être renforcés. Il semblerait opportun d'y adosser plus de chercheurs et ingénieurs issus des organismes et universités pour y développer des projets ambitieux ; ce type d'initiative potentielle était pratiquée par l'ex-Laboratoire de rayonnement synchrotron à Orsay (LURE, actuellement en fin de démantèlement) avec des chercheurs ayant une activité partielle sur le site et dans des laboratoires de recherche adossés.

7.4. Les réseaux de plateformes nationales distribuées et la place d'IBiSA

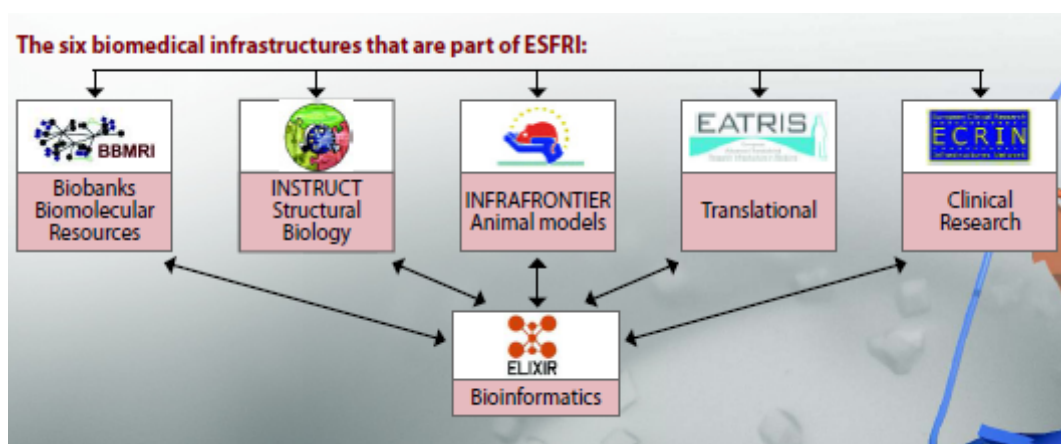
Les programmes IFR apparus en 1997 ont largement contribué à mettre en place ces éléments dans ce sens. Leur fin est programmée depuis 2010 date de création des dernières structures pour 4 ans. C'est certainement le programme mis en place depuis mai 2007 par le GIS IBiSA (Infrastructures en Biologie, Santé et Agronomie; <http://www.ibisa.net/>) qui permet désormais de renforcer la visibilité de ces ensembles très disparates (budget total d'environ 13-17 M€/an). Ce GIS rassemble déjà une analyse collective multi-tutelle des dossiers présentés dans tous les domaines du vivant (hors RMN). La sélection d'accompagner certains projets associant service et développement technologique sur la base de projets scientifiques clairement élaborés et évalués nous permet de disposer d'une vue très robuste des plateformes les plus performantes et qu'il convient de renforcer encore. Pour le domaine, il est important de renforcer encore ce dispositif, pour permettre à quelques plateformes d'émerger incontestablement au niveau national et international. Pour ce faire des programmes de recherche intégrés pourraient être sélectionnés en particulier dans le domaine ~omique (protéomique, métabolomique) sous forme de « bags » afin de gérer les activités des quelques plateformes de très haut niveau et robotisées permettant de mettre en place des appels d'offre nationaux ambitieux et associant des collectifs. Pour les plateformes de biologie structurale

(production de protéines, cristallisation, biophysique, robotisation des tests), il est également souhaitable de rationaliser leur dissémination.

7.5. Les réseaux européens ESFRI : la place d'INSTRUCT en BMSV

Dans le cadre du 7^e programme cadre européen (FP7), le rôle des infrastructures de recherche (European Strategy Forum on Research Infrastructures, ESFRI) est de promouvoir une approche cohérente des politiques concernant les infrastructures de recherche, et d'agir en incubateur pour les négociations internationales sur des initiatives concrètes. A l'heure actuelle, une cinquantaine de projets dans toutes les disciplines ont été financés pour des phases dites préparatoires ; ces projets sont à des stades différents de maturation. L'IBMSV est concerné par 4 infrastructures sur 9 de ce type, identifiées dans le domaine biomédical. Ceci démontre non seulement le travail de la communauté BMSV au niveau européen mais aussi l'absolue nécessité de ce type d'infrastructures pour que les équipes du domaine relèvent les défis scientifiques du futur.

Figure 22 : Infrastructures de type ESFRI dans le domaine biomédical en 2009



En 2010, trois autres projets devraient démarrer leur phase préparatoire (EMBRF pour les stations marines, EU-Openscreen et EuroBioimaging)

7.5.1. INSTRUCT

INSTRUCT pour « Integrated Structural Biology Infrastructure for Europe » correspond à un programme initié en 2007, avec fin de la phase préparatoire prévue en mars 2011. Il s'agit de mettre en place une série de « Core Centres » and « Associate Centres » permettant de disposer d'un réseau pan-European distribué d'infrastructures pour développer une biologie structurale intégrative (« different resolution levels into specific cellular contexts, with a temporal component, to underpin biomedical issues »). Des ateliers ont été mis en place en production de protéines, production à grande échelle de virus et développement de méthodes de purification pour l'analyse structurale, spectrométrie de masse pour protéines natives et complexes, Centre informatique et de calcul pour la biologie structurale intégrative, RMN solide, traitement de l'Image en microscopie électronique. En France, les nœuds seront

localisés à Grenoble via le Partnership for Structural Biology (PSB, un MoU associant les tutelles nationales et internationales comme l'EMBL et l'ESRF, <http://www.psb-grenoble.eu/>) et Strasbourg-Illkirch au Centre de Biologie Intégrative (un projet de construction et de rassemblement de compétence à l'interface entre l'IGBMC et l'ICS). Outre ces deux centres proposant une approche résolument intégrative de la biologie structurale, les autres « core centers » sont situés à Oxford (miniaturisation-automatisation), Florence (centre européen de RMN), Munich (imagerie et microscopie électronique), à l'Institut Weizmann (structure et dynamique de cibles d'intérêt biomédical) et Francfort (biophysique intégrative des complexes, avec emphase sur les membranes). On note la place importante de la France dans ce dispositif qui s'appuie sur deux sites (CBI, Grenoble et PSB, Strasbourg) et un « Integrated Infrastructure Initiative (I3) » dédié, pour participer au financement des projets et à la sélection des projets.

Figure 23 : Formulation actuelle du projet d'I3 associé à INSTRUCT

Projet en cours EU Framework Programme 7 Capacities: Research Infrastructures Theme

Porteur Dr. Matthias Wilmanns, EMBL-Hamburg Germany.

*Work Programme area Support for existing research infrastructures: Integrated Infrastructure Initiative (I3)
Infrastructures for high-resolution structural biology with Xrays*

High-resolution structural biology by X-ray crystallography offers virtually unlimited opportunities to determine the molecular structures of the most challenging macromolecular complexes found in living organisms. Biological X-ray crystallography is used in hundreds of laboratories in virtually all EC members and associated states. The community nowadays critically depends on access to centralized facilities that are mostly provided by or associated with leading synchrotron and laser sites. The project under this topic should therefore aim at providing key integrated infrastructures for high-resolution structure determination of biological macromolecules by X-ray crystallography, coupled with leading complementary methods. The project will be exploited through tight interactions between selected facilities and the Pan-European structural biology research community. The project will have tight links to other ongoing initiatives for existing infrastructures in structural biology with different focus, and with the biomedical ESFRI project "INSTRUCT" for coordinated and integrated future infrastructures in structural biology. The project should be driven in a bottom-up mode by leading members of the user community to ensure optimal facility access as well as networking and training.

Pour conclure, il semble plus qu'opportun d'associer à ces efforts une ligne TGE/TGIR claire pour permettre de consolider dans le temps ces efforts.

7.5.2. Autres programmes ESFRI concernant BMSV

Il s'agit d'Elixir en bioinformatique, porté par l'EBI (EMBL, Hinxton). Deux autres projets sont en cours de démarrage de leur phase préparatoire

- EU-Openscreen consiste en un projet de très grande infrastructure distribuée, Chemical Biology FR-Openscreen (Chimiothèque-Criblage-Chemoinformatique) et ressources de diversité moléculaire. Encore une fois, Strasbourg joue un rôle important dans le dispositif.
- EuroBioimaging pour les imageries moléculaires, cellulaires et biomédicales.

8. RESUME DES CONCLUSIONS ET RECOMMANDATIONS

8.1. Soutien financier de la recherche BMSV

8.1.1. Financements de la recherche en BMSV

Les analyses présentées ici débouchent sur un constat d'alarme avec un tarissement progressif apparent des sources nationales de financement, et ce malgré une visibilité incontestable du domaine. Le domaine n'est globalement pas fortement soutenu malgré sa très grande visibilité au niveau international. Des actions urgentes doivent donc être menées pour inverser cette tendance. Par exemple, des recommandations devraient concerner un affichage thématique à l'ANR et un soutien des opérateurs de recherche.

Pour ce qui concerne les aspects internationaux, la situation est moins préoccupante. Il faut veiller attentivement au maintien de l'implication financière et intellectuelle de la France via les ministères dans ces programmes (HFSP, ERC, PCRD) non seulement lors de l'établissement de la programmation mais aussi dans son évaluation.

8.1.2. Plateformes d'exploration du vivant

Le domaine BMSV est très fortement dépendant d'outils technologiques onéreux : imagerie (y compris l'imagerie super-résolutive, biophysique (Raman, AFM...), cristallisation/robotisation, RMN à haut champ, RPE, spectrométrie de masse de type FT-ICR et technologie hybride, plateforme de criblage à haut débit. Ces instruments d'analyse du vivant représentent non seulement des coûts d'investissement qu'il faut régulièrement renouveler (3-5 ans) mais aussi des coûts d'exploitation importants. Il est indispensable de stabiliser ces outils. Pour le faire, une évolution naturelle est de les placer dans des plateformes mutualisées pour plusieurs utilisateurs, sur un site donné et de les entourer par une assistance technique constante (ingénieurs d'études et de recherche).

Le programme IBiSA présente la légitimité incontestable pour continuer à animer et organiser ces réseaux. Pour le faire, une évolution naturelle est de les placer dans des plateformes mutualisées pour plusieurs utilisateurs sur un site donné et de les entourer par une assistance technique (ingénieurs d'études et de recherche), en exploitant l'existence du GIS IBiSA.

8.1.3. Grands instruments

L'ITMO BMSV soutient aussi fortement la politique des TGE/TGIR pour ce qui concerne le soutien dans la durée des infrastructures en biologie-santé, particulièrement celles impliquées dans les réseaux ESFRI et IBiSA pourrait jouer un rôle majeur dans la montée en puissance de ce système. Enfin, l'accès et les liens scientifiques avec les grands instruments tels que les synchrotrons et les RMN à haut champ doivent être encore renforcés.

8.2. Pratique de l'interdisciplinarité en BMSV

Le domaine BMSV est par essence multidisciplinaire. Sa pratique au quotidien et son évaluation requièrent donc la mise en place de compétences multiples, et une large ouverture d'esprit qu'il faut veiller à entretenir dans le futur. Il s'agit donc d'un effort de formation des nouvelles générations à cet état d'esprit.

L'interdisciplinarité du champ BMSV nécessite donc de mettre en œuvre des formations adaptées aujourd'hui très rares et de réfléchir aux procédures et critères d'évaluation qui doivent être nécessairement différents des autres champs de la biologie.

8.3. Sites visibles en BMSV

La création de campus visibles dans le champ BMSV devrait permettre à la communauté de rivaliser dans le futur au niveau international. Dans cette réflexion de structuration la région parisienne, avec les forces en présence à Paris-Centre et Paris-Sud, démontre une grande visibilité en BMSV et mérite une attention particulière.

9. ANNEXES

9.1. Analyse des périmètres

9.1.1. Collecte des données

Etablissement d'un document de travail (tableau Excel) dont les ITEMS ont été définis en fonction des objectifs de la collecte : rédaction par les experts d'un document prospectif et en vue d'établir une base de données de compétences :

Ingenieurs + techniciens INSERM	Institut du CNRS
Ingenieurs +techniciens CNRS	Tutelle EPST / EPIC
Ingenieurs +techniciens total EPST	Autres tutelles EPST ou EPIC
Ingenieurs +techniciens non EPST	Tutelle universitaire ou grande école
Post doctorants	Ville
Doctorants	Région ou DR CNRS
Total statutaires	Institut
Total CDD	Sigle unité DESP/ylg
Total chercheurs	Intitulé unité et acronyme
Commission /section ?	Nom directeur labo
ITMO principale	Intitulé équipe ou groupe
ITMO secondaires	Mots clés ou activités groupe ou équipe
Commentaires	Nom responsable équipe
AERES note globale	chercheurs CEA - INRA - ou autre EPST
Données financières (financement)	chercheurs INSERM
Equipements	chercheurs CNRS
Structures d'animation	chercheurs total EPST
Formations spécifiques	enseignants-chercheurs non EPST et/ou PH

La collecte des unités a varié suivant les instituts concernés :

- Pour l'INSERM : un tableau a été communiqué par l'INSERM comprenant des équipes indiquées en Techno pour la santé correspondent plutôt au périmètre BMSV
- Pour le CNRS (non ISB) : un tableau des laboratoires que l'INSB-CNRS reconnaît comme présentant des activités dans le domaine des sciences de la vie et de la santé = 101 unités communiquées
- Pour le CEA : des intitulés d'unités communiquées sur proposition des experts
- Pour le CNRS ISB : il a été effectué une extraction complète de Labintel des unités appartenant en 1er ou non à la section 21.
- Pour l'INRA : les données proviennent d'un fichier de résultat d'évaluation AERES des vagues C et D du périmètre de la cellule 5E du CNRS (opérateur et agence de moyen)
- Pour certaines unités CNRS des détails sur les équipes d'intérêt sont aussi venues de se fichier AERES

Pour trouver les intitulés des équipes, leurs responsables et leurs effectifs précis :

- Utilisation de Labintel Consultation (en recherche avancée avec les codes d'accès agent CNRS)
- Exploitation des organigrammes reçus de la cellule « Structures biologiques-Pharmacologie--Enzymologie » de l'Institut des sciences biologiques
- Exploitation des sites Internet des unités de recherche
- Impression de tous les organigrammes reçus ou trouvés pour une exploitation plus aisée

9.1.2. Statistiques

Utilisation d'Excel

Utilisation de « Cartes & données de Articque » pour les répartitions géographiques

Pour établir les ratios d'effectifs :

Nombre total d'unités concernées

Nombre total d'équipes concernées

Nombre de statutaires (chercheurs, enseignants-chercheurs, ingénieurs, techniciens) par unités et par équipes

Nombre de doctorants et post doctorants (et non permanents) par unités et par équipes

Répartition géographique

Tutelles

Rappel des ratios à établir pour les effectifs :

Doctorants et post doctorants / statutaires

Nombre de statutaires et autres ratios pour chacun des 4 sous domaines

Nombre de statutaires et autres pour chacune des tutelles

Répartition géographique pour l'ITMO et pour chacun des 4 sous domaines et par rapport aux tutelles

(résultats en % statutaires, tutelles et part relative effectifs statutaires au sein des unités par organisme.)

Statistiques reprises pour chaque sous-domaine

9.2. Données chiffrées globales

Complément du Tableau 2 : Données issue des rapports OST et Futuris

Rapport OST

Tableau 1-2-15

France : chercheurs du secteur public civil en personnes physiques (pp) – nombre et répartition par catégorie de personnel (2005) selon la discipline

Discipline	Chercheurs du secteur public civil (pp) (2005)			Nombre (pp)
	Répartition (%)			
	Enseignants-chercheurs	Chercheurs	Ensemble	
Sciences de la vie	56,0	44,0	100,0	28 591
Biologie fondamentale	43,9	56,1	100,0	12 225
Recherche médicale	72,7	27,3	100,0	13 659
Biologie appliquée-écologie	26,6	73,4	100,0	2 707
Sciences de la matière	49,4	50,6	100,0	41 757
Chimie	46,6	53,4	100,0	6 893
Physique	38,7	61,3	100,0	6 610
Sciences de l'univers	34,4	65,6	100,0	4 709
Sciences pour l'ingénieur	52,2	47,8	100,0	18 925
Mathématiques	72,3	27,7	100,0	4 620
Sciences humaines et sociales	76,8	23,2	100,0	27 672
Toutes disciplines	59,1	40,9	100,0	98 020

données MEN-MESR-DÉPPC2 et Coopération, traitements et indicateurs OST

rapport OST-2008

- 17 204 enseignants-chercheurs et 3 481 chercheurs (soit un total de 20 685 personnes), dont la discipline de rattachement n'est pas identifiée, ne sont pas pris en compte
- sont inclus dans les chercheurs : les enseignants-chercheurs, les chercheurs des organismes, les ingénieurs de recherche, les chercheurs sur contrats
- ici, le personnel des établissements d'enseignement supérieur non rattachés au ministère chargé de l'Enseignement supérieur est pris en compte (par exemple : écoles des mines, écoles vétérinaires...)
- les organismes de recherche sont le CNRS, l'Inra, l'Inserm, l'Ifremer, etc., des GIP tel le CNRG, des fondations comme l'Institut Pasteur et l'Institut Curie (voir la nomenclature en fin d'ouvrage)

Données du Rapport Futuris – ANRT : La recherche publique dans le domaine biomédical en France Analyse quantitative et éléments de diagnostic ETUDE BIOMED, mai 2008

Tableau 9 : Personnels chercheurs et enseignants-chercheurs des institutions publiques de recherche – répartition par et entre institutions, tous domaines confondus et domaine biomédical (2005)

	Nombre en Biomédical (1)	Nombre total dans l'institution	Part de l'institution dans le total Biomédical (%)	Part du Biomédical dans l'institution (%)
<i>Personnes physiques</i>				
Universités	11 316	53 098	63,5	21,3
CNRS	3 238	11 606	18,2	27,9
Inserm	2 161	2 161	12,1	100,0
Inra	357	1 825	2,0	19,5
IRD	88	752	0,5	11,7
Pasteur	317	317	1,8	100,0
CEA	346	5 843	1,9	5,9
Autres (2)	0	863	0,0	0,0
TOTAL	17 823	76 465	100,0	
<i>Equivalents temps plein recherche</i>				
Universités	4 170	2 5060	39,1	16,6
CNRS	3 238	11 606	30,3	27,9
Inserm	2 161	2 161	20,2	100,0
Inra	357	1 825	3,3	19,5
IRD	88	752	0,8	11,7
Pasteur	317	317	3,0	100,0
CEA	346	5 843	3,2	5,9
Autres (2)	0	863	0,0	
TOTAL	10 677	48 427	100,0	

Données Coopération, OST novembre 2006 et institutions; traitements Futuris

(1) Biomédical = sciences de la vie + médecine et odontologie

(2) Autres établissements inclus dans l'étude de l'OST

9.3. Analyse bibliométrique

La recherche des publications (articles, lettres et revues et proceedings papers) a été effectuée à partir de la base de l'ISI. Cette base est constituée de l'ensemble des articles provenant de 5800 journaux scientifiques (Science citation index ; SCI) indexés depuis 1945, de 1700 journaux de sciences sociales (Social sciences CI) indexés depuis 1956 et de plus de 8000 revues (Arts and humanities CI) indexées depuis 1975.

Cette base constitue aujourd'hui la base internationale la plus complète (Medline n'indexe que 4500 journaux depuis 1966).

9.3.1. Procédure

- A. Le corpus de publications a été établi à partir de mots-clés établis en 2 temps :
 - 1) Mots clés fournis par les experts de l'Institut,
 - 2) Recherche affinée et complétée par une recherche de mots clés par une analyse manuelle sur les articles des meilleurs des journaux de spécialité. Le corpus obtenu est analysé pour éliminer les publications hors domaines soit sur le titre du journal soit sur le titre de la publication.
- B. Des Tests de validité du corpus obtenu ont été effectués : à partir des meilleurs journaux de la spécialité (classés par IF dans la base du JCR) : JMB, Nature Structural Biology, Nature Chemical Biology, Protein Structure functions and bioinformatics...
- C. Afin d'établir une comparaison et vérifier nos requêtes, nous avons chargé les publications des chercheurs identifiés dans l'Institut.
Sur les 1257 publications des chercheurs récupérées à partir de leurs noms, 768 (61%) se trouvent dans notre corpus de publications constitué à partir des mots-clés.

9.3.2. Présentation des Indicateurs Bibliométriques

Nous avons établis différents tableaux de données à l'aide des corpus de publications : données bibliométriques de la France, comparaisons internationales, productivité et cartographies (répartition géographiques des sous domaines en nombre de publications, réseaux de collaborations des organismes d'un Institut).

La complexité à définir des mots-clés et la transversalité de cet Institut nous ont obligés à procéder de deux manières :

- ❖ afin de pouvoir faire des **comparaisons internationales** nous nous sommes servies des requêtes pour produire les indicateurs pour l'Allemagne, l'Angleterre, les Pays-Bas, la Suisse, la Suède, le Danemark, la Chine, le Japon, la Corée, l'Italie, l'Espagne, l'Australie, le Canada et les Etats-Unis.

Pays	Nb pub	Total citations ¹	Indice de citations moyen	h-index ²
	0	0	0	0



Nous n'avons pas accès aux données sur les citations lorsque le nombre de publications est supérieur à 10 000 (représenté par NC dans le tableau).

9.3.5. Cartographies : répartition géographique des publications par région

Pour cette cartographie, le logiciel utilisé est Arctique4 : plusieurs cartes ont été établies à partir des données fournies sous forme d'un tableau Excel par l'Institut.

Mots clés et requêtes pour les analyses bibliométriques :

#15 ts=((synthesis same (peptide* or carbohydra* or oligosacch* or lipid*)) or (synthesis and (agonist* or antagonist*))) and cu=france and py=2006-2007 AND Document Type=(Article OR Letter OR Proceedings Paper OR Review)

Databases=SCI-EXPANDED, SSCI, A&HCI, CPCI-S, CPCI-SSH Timespan=All Years

14 ts=((biol* or protein* or RNA* or DNA* or peptid* or lipid* or glyco* or carbohydr*) and (synchrotron or crystal* or NMR or SPECTROSCOPY or spectrometry or maldi* or NANO-ELECTROSPRAY* or CRYSTAL* or spectro or "X-ray" or "synchrotron" or imag*))and py=2006-2007 and cu=france AND Document Type=(Article OR Letter OR Proceedings Paper OR Review)

Databases=SCI-EXPANDED, SSCI, A&HCI, CPCI-S, CPCI-SSH Timespan=All Years

13 ts=(biogenesis or binding-site or "active site" or "ligand-binding domain*" or "binding domain*" or conformational-change or (enzyme and structure) or conformational-change* or conformational-dynamic* or "enzyme-substrate complex" or (crystal* and (protein* or enzym*)) or "PROTEIN CRYSTALS" or "MULTIVALENT LIGAND*" or (molecular motor* and (kinesin* or dynamic* or motion)) or KINESIN-RELATED PROTEIN or "ACTIN DYNAMICS" or (actin and propulsion) or (actin* and motility)) and cu=france and py=2006-2007 AND Document Type=(Article OR Letter OR Proceedings Paper OR Review)

Databases=SCI-EXPANDED, SSCI, A&HCI, CPCI-S, CPCI-SSH Timespan=All Years

12 ts=((NMR and protein*) or (protein* and (unfold* or folding*)) or "protein dynamic*" or (iron-sulfure and (protein or cluster*)) or (cluster same protein) or "protein* voltammetry" or (relaxation and protein*) or PROTEIN-STRUCTURE or SECONDARY-STRUCTURE* or "quaternary structure*" or (protein and assembly)) and cu=france and py=2006-2007 AND Document Type=(Article OR Letter OR Proceedings Paper OR Review)

⁴ www.arctique.com

Databases=SCI-EXPANDED, SSCI, A&HCI, CPCI-S, CPCI-SSH Timespan=All Years

11 So=(JOURNAL OF BIOMOLECULAR STRUCTURE & DYNAMICS or Journal Of Molecular Graphics & Modelling or Proteins-Structure Function And Genetics or STRUCTURE or Journal Of Molecular Recognition or BMC STRUCTURAL BIOLOGY or Journal of proteome research or PROTEIN EXPRESSION AND PURIFICATION or RNA-A PUBLICATION OF THE RNA SOCIETY or PROTEINS-STRUCTURE FUNCTION AND BIOINFORMATICS or PROTEIN SCIENCE or PROTEOMICS or JOURNAL OF MOLECULAR STRUCTURE or nature structural and molecular biology) and cu=france and py=2006-2007 AND Document Type=(Article OR Letter OR Proceedings Paper OR Review)

Databases=SCI-EXPANDED, SSCI, A&HCI, CPCI-S, CPCI-SSH Timespan=All Years

10 ts=((protein same model*) or (protein and atom*) or (protein* and structur*) or (drug and "Bioactive molecule*") or (molecul* and conformation*) or (enzym* and kinetic*) or (kinetics same aggregation) or (ATOMIC-FORCE and ("cell membrane*" or lipid* or protein*))) and cu=france and py=2006-2007 AND Document Type=(Article OR Letter OR Proceedings Paper OR Review)

Databases=SCI-EXPANDED, SSCI, A&HCI, CPCI-S, CPCI-SSH Timespan=All Years

9 ts=((structural insights and (molecule* or macromolecul* or protein*)) or ("self assembl*" and protein*) or ("Protein Fold*" or "protein misfold*") or ("protein architectur*" or "PROTEIN STRUCTUR*" or "nucleotide excision repair") or protein-protein or "protein-protein interaction*" or "protein-fragment complementation assay" or "protein interactome" or (ANGSTROM* and protein*)) and py=2006-2007 and cu=france AND Document Type=(Article OR Letter OR Proceedings Paper OR Review)

Databases=SCI-EXPANDED, SSCI, A&HCI, CPCI-S, CPCI-SSH Timespan=All Years

8 ts=((nanopore* and biol*) or "protein misfolding" or "protein cargo" or "structural biology" or (modelling same protein*) or PHYSICAL INTERACTOME) and cu=france and py=2006-2007 AND Document Type=(Article OR Letter OR Proceedings Paper OR Review)

Databases=SCI-EXPANDED, SSCI, A&HCI, CPCI-S, CPCI-SSH Timespan=All Years

7 TS=((oligosaccharide* or glycan* or carbohydrat* or polysacch* or lipid* or peptid*) and (structure* or conformation*)) and py=2006-2007 and cu=france AND Document Type=(Article OR Letter OR Proceedings Paper OR Review)

Databases=SCI-EXPANDED, SSCI, A&HCI, CPCI-S, CPCI-SSH Timespan=All Years

6 ts=("Triplex-forming oligonucleotide*" or MicroRNA* or riboswitch* or "modelling RNA" or "molecular dynamics" or "deep sequencing" or "rna folding" or "three-dimensional structur*" or (crystal* same (dna or RNA)) or "Nucleic-Acid Structure" or (structure same mRNA) or (RNA* and complex) or "RNA* folding" or (assembly same rna) or (rna* same protein*)) and py=2006-2007 and cu=france AND Document Type=(Article OR Letter OR Proceedings Paper OR Review)

Databases=SCI-EXPANDED, SSCI, A&HCI, CPCI-S, CPCI-SSH Timespan=All Years

5 TS=("Metabolic pathway*" or metabolomics or "metabolic network" or glycomics or "ligand binding modes" or "target protein" or "MOLECULAR DOCKING" or "FLEXIBLE DOCKING" or "ligand docking" or "pharmacophoric map*" or "computational molecular biology" or (enzyme and characterisation) or (receptor* same (dimer* or polymer*)) or proteom* or (snrmp or "small nuclear ribonucleoprotein") or (kinetic* and domain* and protein*) or palindromic repeats or "real-time biosensors" or "single-molecule sensors" or enzyme* catalysis or "substrate delivery" or "Dynamic nucleosomes" or (protein* and photo)) and py=2006-2007 and cu=france AND Document Type=(Article OR Letter OR Proceedings Paper OR Review)

Databases=SCI-EXPANDED, SSCI, A&HCI, CPCI-S, CPCI-SSH Timespan=All Years

4 TS(("STRUCTURAL same PREDICTION) or (structur* same ribosom*) or (structural plasticity and protein*) or ("mechanistic features" and protein*) or ("Structural systems" and biology) or (modelling protein*" or "subcellular localization") or ("structural characterization") or ((structur* or conformation*) same kinase) or ("structural basis" and (protein*or cell*)) or (structur* same function*) or "molecular basis" or (fold* same rna*)) and cu=france and py=2006-2007 AND Document Type=(Article OR Letter OR Proceedings Paper OR Review)

Databases=SCI-EXPANDED, SSCI, A&HCI, CPCI-S, CPCI-SSH Timespan=All Years

3 ts=((mitochondr* and structur*) or (mitochondr* and crystal*) or (mitochondr* and complex*)) and cu=france and py=2006-2007 AND Document Type=(Article OR Letter OR Proceedings Paper OR Review)

Databases=SCI-EXPANDED, SSCI, A&HCI, CPCI-S, CPCI-SSH Timespan=All Years

2 ts=((Conformation* and "drug design") or ("allosteric regulation") or ("structural motif*") or (Biological and ("complex morphology" or "chemical composition")) OR (molecul* and (subunit architecture or oligomeric structure))) and cu=france and py=2006-2007 AND Document Type=(Article OR Letter OR Proceedings Paper OR Review)

Databases=SCI-EXPANDED, SSCI, A&HCI, CPCI-S, CPCI-SSH Timespan=All Years

1 ts=((rna polymerase or dna polymerase or dna nucleoprotein or dna duplexe* or mrna export or chromatin arrangement* or nuclear architecture or intersubunit rearrangement or allosteric modulatos) or (DNA same (repair or transcription or replication)))and py=2006-2007 and ad=france AND Document Type=(Article OR Letter OR Proceedings Paper OR Review)

Databases=SCI-EXPANDED, SSCI, A&HCI, CPCI-S, CPCI-SSH Timespan=All Years