

**AMBASSADE DE FRANCE AUX ETATS-UNIS**  
MISSION POUR LA SCIENCE ET LA TECHNOLOGIE  
CONSULAT GENERAL DE FRANCE A BOSTON ET A SAN-FRANCISCO

**NOUVELLES INITIATIVES EN  
PROTEOMIQUE ET GENOMIQUE STRUCTURALE**

**Christiane BRANLANT, Jacques HAIECH, Jacques JOYARD  
Jean ROSSIER, Florent SOUBRIER  
Pierre MICHEL, Stéphane ROY**

**Janvier 2002**

**Rapport de**

**Christiane BRANLANT<sup>1</sup>**

Directrice adjointe du Département des Sciences de la Vie du CNRS

**Jacques HAIECH<sup>2</sup>**

Conseiller à la Direction de la Recherche, Ministère de la Recherche

**Jacques JOYARD<sup>3</sup>**

Directeur adjoint de la Génopole Rhône-Alpes

**Jean ROSSIER<sup>4</sup>**

Conseiller INSERM

**Florent SOUBRIER<sup>5</sup>**

Comité Scientifique du CNG

**Pierre MICHEL<sup>6</sup>**

Attaché pour la Science et la Technologie

**Stéphane ROY<sup>7</sup>**

Attaché pour la Science et la Technologie

---

<sup>1</sup> [christiane.branlant@maem.uhp-nancy.fr](mailto:christiane.branlant@maem.uhp-nancy.fr)

<sup>2</sup> [jacques.haiech@recherche.gouv.fr](mailto:jacques.haiech@recherche.gouv.fr)

<sup>3</sup> [jjoyard@cea.fr](mailto:jjoyard@cea.fr)

<sup>4</sup> [jean.rossier@espci.fr](mailto:jean.rossier@espci.fr)

<sup>5</sup> [florent.soubrier@chups.jussieu.fr](mailto:florent.soubrier@chups.jussieu.fr)

<sup>6</sup> [pierre.michel@ambafrance-us.org](mailto:pierre.michel@ambafrance-us.org)

<sup>7</sup> [stephane.roy@consulfrance-sanfrancisco.org](mailto:stephane.roy@consulfrance-sanfrancisco.org)

### Résumé

*La protéomique est l'étude de l'ensemble des protéines et plus particulièrement de leurs fonctions. Cette nouvelle vague d'innovation technologique, qui dans sa définition la plus large regroupe la génomique structurale, revêt une importance capitale pour la communauté biomédicale: les pathologies sont souvent le résultat de dysfonctionnements des protéines et ces dernières sont aussi des cibles pour les actions thérapeutiques. La protéomique augmente donc notre compréhension des mécanismes moléculaires des maladies et améliore le diagnostic et les thérapies. Plusieurs initiatives récentes qui ont pris naissance à la fin de l'année 2000 et au cours de l'année 2001 illustrent l'importance de ce domaine de recherche aux Etats-Unis et la protéomique aura sans aucun doute, des retombées commerciales importantes.*

*En se basant sur des exemples pris dans les régions de Boston, San Francisco et San Diego, nous avons cherché à analyser les initiatives conduites pour étudier le protéome à la fois dans un cadre académique et dans le contexte plus large du développement de ce secteur économique. Il ressort de cette réflexion les points suivants :*

- *La protéomique doit répondre aux défis de l'industrialisation et de la modification profonde du concept de la recherche en biologie. Le développement de l'automatisation et de la robotisation des unités de productions permet d'ores et déjà un changement d'échelle très net dans l'analyse des protéines*
- *Les programmes développés aux Etats-Unis, initiatives publiques ou entreprises industrielles, sont beaucoup plus ambitieux que ce qui existe en France. Le modèle de développement, largement utilisé au cours des années précédentes par le secteur de la génomique, semble pouvoir être repris pour la protéomique avec le développement d'outils technologiques d'analyse du protéome et/ou la création de bases de données.*
- *La vitalité entrepreneuriale des Etats-Unis est toujours évidente, même si la prise de risque des financiers n'est pas toujours apparue corrélée au contenu scientifique des projets présentés. Toutes les entreprises ont pour objectif le criblage de molécules sur cellules, parfois sur des protéines et les cibles potentielles choisies sont alors très semblables, quelque soit la compagnie.*
- *Les entreprises font du développement technologique. Certaines apportent des solutions originales, mais il est à craindre que les coûts des produits soient alors très élevés et ne soient pas disponibles pour la recherche académique. Ces entreprises prennent en compte la multidisciplinarité de manière plus importante qu'en Europe et en France*
- *Les initiatives publiques ne semblent pas inquiéter les sociétés de biotechnologie qui misent sur la génomique structurale pour développer un business model. Ces sociétés pensent que la vitesse qui résulte de l'automatisation de chaque étape attirera la clientèle des groupes pharmaceutiques à la recherche de structures de protéines.*
- *La disponibilité des données structurales du génome n'est pas encore résolue. Si tout le monde s'accorde dans le milieu académique pour dire que les données devront être publiques, il y a une demande européenne et japonaise pour étendre cette durée ce qui permettrait de déposer de brevets sur des structures d'importance pharmaceutique. Toutefois un consensus international a été atteint pour publiciser les protéines qui sont à l'étude pour éviter les redondances.*
- *Les risques de recherche à deux vitesses sont énormes, les organismes de recherches comme les universités n'auront jamais les moyens de développer les robots qu'utilisent des sociétés de biotechnologie.*

## **I - INTRODUCTION**

### **II - CENTRES DE RECHERCHE ACADEMIQUE.**

- 1 - Institute of Proteomics. Harvard University, Massachusetts
- 2 - Laboratory of Proteomic Mass Spectrometry. The Scripps Research Institute, Californie
- 3 - Protein Structure Initiative. Lawrence Berkeley Laboratory, Californie.

### **III- SOCIETES DE BIOTECHNOLOGIES.**

- 1 - Proteome/Incyte Genomics. Beverly, Massachusetts.
- 2 - Structural GenomiX. La Jolla, Californie
- 3 - GeneFormatics. La Jolla, Californie
- 4 - Syrrx. La Jolla, Californie
- 5 - Zyomyx. Fremont, Californie.
- 6 - Guava Technologies. Burlingame, Californie
- 7 - SurroMed. Mountain View, Californie
- 8 - CIPHERGEN. Fremont, Californie.
- 9 - Abgenix. Fremont, Californie
- 10 - Signature BioScience. Hayward, Californie

### **IV - ANALYSE ET CONCLUSIONS.**

- 1 - Les avancées technologiques
- 2 - La création d'un modèle économique
- 3 - Le futur légal des données.
- 4 - Conclusions

## I - INTRODUCTION

L'achèvement du séquençage du génome humain et l'analyse du génome de nombreux organismes modèles et microbiens ont ouvert la voie à une nouvelle discipline, la protéomique. La protéomique est l'étude de l'ensemble des protéines, de leurs localisations, modifications, interactions et finalement de leurs fonctions. Dans son sens le plus large, cette définition inclut aussi la génomique structurale, domaine qui analyse la structure tridimensionnelle (3D) des protéines et qui permet d'avoir une meilleure compréhension de leurs fonctions moléculaires.

La protéomique revêt une importance capitale pour la communauté biomédicale: les pathologies sont souvent le résultat de dysfonctionnements des protéines et ces dernières sont aussi des cibles pour les actions thérapeutiques. La protéomique augmentera donc notre compréhension des mécanismes moléculaires des maladies et améliorera le diagnostic et les thérapies. Elle constitue une nouvelle vague d'innovations dans le processus de la découverte des médicaments qui combinée à la première vague (la génomique) et à celle qui est déjà amorcée (*molecular design* ou conception assistée de médicaments) va continuer à modifier la façon dont les nouvelles thérapies médicamenteuses sont imaginées.

Plusieurs initiatives récentes qui ont pris naissance à la fin de l'année 2000 et au cours de l'année 2001 illustrent l'importance de ce domaine de recherche aux Etats-Unis:

- La création du *Institute of Proteomics* à l'Université de Harvard (MA) pour définir les fonctions de toutes les protéines et de créer une base de données accessible à l'ensemble de la communauté scientifique. L'institut a formé un consortium d'entités publiques et privées pour financer cette initiative (budget évalué à 100 millions de dollars).
- Le développement des outils technologiques par plusieurs sociétés de biotechnologie pour analyser le protéome. On peut citer par exemple les puces à protéines qui permettent d'identifier des marqueurs biologiques associés à des pathologies. Le marché des puces à protéines évalué à 45 millions de dollars en 2000 pourrait exploser à 500 millions dans les 5 prochaines années.
- La *Protein Structure Initiative* (PSI) du *National Institute of Health* (NIH), a permis le financement de 7 centres de recherche pilote sur 5 ans pour un montant total de 150 millions de dollars. Ces centres répartis sur l'ensemble du territoire sont des consortiums qui regroupent en moyenne 3 à 8 instituts de recherche et qui développent et testent les techniques d'analyse des protéines<sup>8</sup>.
- *The Human Proteome Organization* (HUPO) a été créé en février 2001 et regroupe les experts des différents instituts de recherche mondiaux en protéomique.
- Les premiers résultats spectaculaires sont apparus : *CuraGen* (New Haven, CT) a cartographié toutes les interactions protéiques d'un genome entier (*Saccharomyces cerevisiae*)<sup>9</sup> ; les réalisations de *Hybrigenics* (France)
- *Celera Genomics* (Rockville, MD) a levé 855 millions de dollars pour un programme de génomique fonctionnelle et de protéomique<sup>10</sup> et *Perkin Elmer Biosciences* (Foster City, CA) a créé le *Proteomic Research Center* à Framingham (MA)<sup>11</sup>

---

<sup>8</sup> <http://www.nigms.nih.gov/funding/psi.html>

<sup>9</sup> Nature, Février 2000

<sup>10</sup> Mars 2000

<sup>11</sup> Mars 2000

- Celera, Compaq et le Sandia National Laboratory (Albuquerque, NM) ont signé un accord<sup>12</sup> pour construire un nouvel ordinateur capable d'effectuer 100 trillions d'opérations par seconde dans le but d'étudier la fonction, la structure et l'interaction des protéines chez les humains<sup>13</sup>

La protéomique aura sans aucun doute, des retombées commerciales importantes. Le modèle de développement, largement utilisé au cours des années précédentes par le secteur de la génomique, semble pouvoir être repris pour la protéomique avec le développement d'outils technologiques d'analyse du protéome et/ou la création de bases de données.

Du fait de l'importance de ce domaine de recherche, la Mission pour la Science et la Technologie aux Etats-Unis a organisé une mission d'étude pour des experts du CNRS, de l'INSERM, du Ministère de la Recherche et de la société *Hybrigenics*<sup>14</sup> en octobre 2001. En se basant sur des exemples pris dans les régions de Boston, de San Francisco et de San Diego<sup>15</sup>, nous avons analysé les initiatives conduites pour étudier le protéome à la fois dans un cadre académique et dans le contexte plus large du développement de ce secteur économique.

En s'appuyant sur des rencontres avec des professeurs d'université, des Présidents, des *Chief Executive Officer (CEO)*, *Chief Scientific Officer (CSO)* et *Chief Technological Officer (CTO)* de sociétés de biotechnologie, nous présentons, dans un premier temps, les centres académiques pour ensuite décrire quelques sociétés de biotechnologie. Nous terminons par une analyse de ce nouveau secteur de l'industrie des biotechnologies et les conclusions que ce voyage d'étude a apportées à la lumière de la situation française.

## II - CENTRES DE RECHERCHE ACADEMIQUE.

### 1 – Institute of Proteomics, Harvard University, Massachusetts<sup>16</sup>

*Joshua LaBaer, M.D., Ph.D. Director,*<sup>17</sup>

*Steven Lloyd, Associate Director of Finance and Operations*

*Plusieurs chercheurs, dont Gerald Marsischky, étaient aussi présents.*

Le principal objectif de ce nouvel institut est la création d'une base de donnée de gènes humains sous forme d'ADNc<sup>18</sup> clonés dans un vecteur multifonction de type Gateway permettant l'expression de l'ADNc<sup>19</sup> dans différents systèmes (*E. coli*, *Baculovirus*) pour déterminer la structure, la fonction et les partenaires moléculaires de la protéine. La construction de ces banques de clones permet de produire l'ensemble des protéines d'un organisme, ceci sous forme naturelle,

---

<sup>12</sup> Janvier 2001

<sup>13</sup> Projet de 4 ans

<sup>14</sup> Pierre Legrain, Vice President R&D a malheureusement annulé son voyage

<sup>15</sup> Seules des contraintes de calendrier nous ont forcé à nous cantonner à ces régions. Des initiatives tout aussi intéressantes et originales ont lieu à l'*University of Michigan* (Sam Hanash) ou à *The Institute for System Biology* (Leroy Hood et Ruedi Aebersold à Seattle).

<sup>16</sup> <http://www.hip.harvard.edu>

<sup>17</sup> Department of BCMP. Harvard Medical School 250 Longwood Ave. Boston, MA 02115. Tel : 617-432-2659. Fax : 617-432-5117

<sup>18</sup> Pas d'investissement dans la recherche et le séquençage de nouveaux ADNc mais une utilisation des banques disponibles.

<sup>19</sup> Ils amplifient les séquences correspondantes par RT-PCR, le premier brin d'ADN étant produit dans le cas de la banque humaine à partir d'ARNm extraits de cellules de patients. Les amorces pour la RT-PCR sont définies par traitement informatique des données de séquences de banques de cDNA (NCBI et NCI collections de cDNA). Une version exprimant la protéine du start codon au stop codon est produite, ainsi qu'une version où le stop codon est éliminé et des séquences codant poly-His ou GST sont rajoutées en 3'.

ou avec un tag His ou GST et mettre ces banques à disposition des instituts académiques (gratuit) et industriels (payant). Le principe est de produire un clone pour chaque ARNm potentiel d'un organisme donné ou du moins chaque gène, lorsqu'il y a trop de variants d'épissage possibles. Ils utilisent le système Flex (*Full Length Expressed Gene*) assez versatile pour permettre les différents types d'expression souhaités, programme qui est décrit sur leur site internet<sup>20</sup>. Une technique rapide de clonage dans un vecteur par recombinaison d'ADN dans les bactéries est utilisée<sup>21</sup>. Le problème technique non résolu concerne les ARNm de grande taille pour lesquels l'amplification par PCR peut poser un problème. Ils ont déjà produit tous les clones en ce qui concerne la banque de *S. cerevisiae* (G. Marsischky, Responsable) et ont largement avancé sur les banques de clones de drosophile, *Plasmodium falciparum* et *Pseudomonas*. Ils ont produit 1.800 clones humains. Ils ont environ actuellement 16.000 clones décrits sur leur site<sup>22</sup>.

Leur objectif est de monter un consortium et d'assurer une distribution des clones qui pourrait être réalisée par plusieurs sociétés privées. Ils seraient très intéressés à ce que des laboratoires français participent. Leur objectif est de disposer et de distribuer environ 100.000 clones d'ici trois ans, caractérisés chacun par un code barre. Selon eux, ils ont une bonne proportion de protéines exprimées correctement (de meilleurs résultats étant obtenus avec le tag His qu'avec le tag GST). L'automatisation des procédés leur permet de tester l'expression d'un très grand nombre de clones en même temps. Sur le plan thématique, ils sont directement intéressés par :

- la fonction de gènes codant des protéines inconnues chez la levure et *Pseudomonas*,
- le cancer et les maladies du pancréas<sup>23</sup>, et en particulier les gènes qui permettent à une cellule de traverser les membranes.

Le programme de distribution n'a pas encore réellement commencé : l'institut recherche encore les financements estimés à 100 millions de dollars<sup>24</sup>. Le budget de l'institut, qui teste la validité des stratégies qui seront mises en place dans FLEX est d'environ 1 million de dollar par an. Ce qui représente le coût le plus élevé est le séquençage des clones. Ils ont au départ débuté avec de l'argent de la Harvard Medical School, puis ils ont obtenu un soutien du NIH. En parallèle, ils ont obtenu un contrat avec une compagnie privée.

Le programme développé par l'*Institute for Proteomics* est une ressource très importante pour disposer d'ADNc complets dans une forme facilement exprimable, et pour disposer des différents variants d'épissage pour chaque gène. Ce programme présente certaines similarités avec le programme sur les ADNc humains et de souris en cours au Centre National de Séquençage à Evry, mais ils semblent nettement en retard. Des possibilités d'échanges sur leur façon de procéder et celles d'Evry, en particulier sur les modes de distribution des clones, seraient bénéfiques. Au niveau CNRS, il ne semble pas y avoir d'initiative similaire en ce qui concerne *S. cerevisiae*, *P. falciparum* et *Pseudomonas* et beaucoup de laboratoires pourraient être intéressés. Pour la levure, une banque permettant d'exprimer les différentes ORF est déjà commercialisée mais c'est très cher et le vecteur de clonage offre sans doute beaucoup moins de possibilités au niveau choix du système d'expression. L'institut est très ouvert à des collaborations, ce qui est un moyen de valider l'effort qu'ils ont fait pour *Pseudomonas* et *P. falsiparum* : le nombre de laboratoires concernés en France est assez limité. En particulier, pour *P. falsiparum*, H. Vial peut se charger de contacter ses

---

<sup>20</sup> <http://kotel.med.harvard.edu/FLEX/logon.do>

<sup>21</sup> la recombinaison est faite en 1 heure, ce qui permet de réaliser rapidement un grand nombre de clonages par des systèmes robotisés traitant un grand nombre d'échantillons à la fois.

<sup>22</sup> <http://kotel.med.harvard.edu/FLEX/System/demo>

<sup>23</sup> Ils développent en parallèle des travaux de traitement des données de la littérature pour établir des liens entre gènes et maladies.

<sup>24</sup> le coût correspondrait seulement aux frais de distribution

collègues au travers du GDR parasitologie. Pour *S. cerevisiae*, le nombre de laboratoires est plus élevé. Evry pourrait après établissement de collaborations individuelles, être un centre distributeur de leur clones. Une autre possibilité est l'ensemble CGM, Janin... qui s'implique dans la génomique structurale de la levure. Même si certains laboratoires (Curie, A Nicolas) se sont équipés de la banque d'ORF de levure, il serait opportun de voir ce qu'il pourrait être fait avec La Baer dans ce domaine.

**2 - Laboratory of Proteomic Mass Spectrometry. The Scripps Research Institute, Californie<sup>25</sup>**  
*Professor John Yates III<sup>26</sup>, Principal Investigator*  
*Laurence Florens, Research Scientist<sup>27</sup>*

Le Professeur Yates est un des pionniers dans l'analyse de mélanges protéiques complexes par spectrométrie de masse et l'objectif de ce groupe est de développer des technologies d'analyse du protéome qui évite l'utilisation de gel 2D. Le travail réalisé actuellement sur le protéome de *Plasmodium* par une post-doc française, Laurence Florens<sup>28</sup>, a très bien illustré la stratégie mise en œuvre de manière systématique et depuis plusieurs années par son groupe<sup>29</sup> et la complexité du problème<sup>30</sup>.

Toutes les analyses sont basées sur le même principe. Les cellules sont lysées et à la fois la fraction soluble et la fraction insoluble<sup>31</sup> sont analysées. La séparation des protéines est réalisée en utilisant la nanoHPLC<sup>32</sup> directement couplée au spectromètre de masse de type ESI-ion trap, suivant le principe de « tandem mass spectrometry » (MS-MS) qui permet d'avoir directement la séquence des peptides. Les résultats concluants<sup>33</sup> illustrent les nouvelles approches dans le protéome à moyen et haut débit. En particulier, l'importance de la gestion et de l'analyse des données sont bien

---

<sup>25</sup> <http://fields.scripps.edu>

<sup>26</sup> Department of Cell Biology, SR11. 10550 North Torrey Pines Road. LaJolla, CA 92130. Tél: 858-784-8862. Fax: 858-784-8883  
[jyates@scripps.edu](mailto: jyates@scripps.edu)

<sup>27</sup> [florens@scripps.edu](mailto: florens@scripps.edu)

<sup>28</sup> thèse chez M. Bruschi à Marseille

<sup>29</sup> Link *et al.*, 1999, Direct analysis of protein complexes using mass spectrometry. *Nat. Biotechnol.* 17, 676–682.

<sup>30</sup> L'analyse du parasite est réalisée à plusieurs phases du cycle biologique, dans deux types d'hôtes (anophèle et homme). Les protéines analysées proviennent donc soit du *Plasmodium*, soit de l'anophèle, soit de l'homme et doivent donc être discriminées, ceci afin de développer des inhibiteurs. Quatre stades du développement sont étudiés : Sporozoïtes, Trophozoïtes, Mérozoïtes, Gamétocytes

<sup>31</sup> Dans le cas de la fraction insoluble produite lors de la lyse des cellules, une solubilisation est réalisée par traitement à l'acide formique et au Bromure de cyanogène, avant hydrolyse par des endopeptidases, suivie de l'analyse LC-MS-MS comme pour la fraction soluble.

<sup>32</sup> Les peptides résultants de l'hydrolyse par des protéases sont purifiés par nanochromatographie liquide dans des microcapillaires, qu'ils remplissent eux même avec la ou les phases chromatographiques souhaitées. Avant remplissage, le capillaire est effilé à une extrémité, afin d'être placé directement dans le spectromètre de masse après remplissage. Les peptides arrivant à l'extrémité de la pointe vont être ionisés.

<sup>33</sup> La séquence du génome de *P. falciparum* est bientôt terminée et ils ont accès aux données actuelles pour leurs analyses de peptides. Par cette approche, ils ont pu identifier 645 protéines exprimées au stade sporozoïtes, 640 protéines exprimées au stade trophozoïtes, 599 protéines exprimées au stade mérozoïtes et 99 exprimées au stade gamétocytes. Le matériel biologique leur est fourni par un laboratoire de biologie qui dissèque les glandes salivaires des moustiques. La séparation entre matériel protéique de l'hôte et du parasite n'est pas parfaite, ils ont beaucoup de contamination par des protéines de l'Anophèle qu'ils ne peuvent pas jusque là identifier du fait que la séquence génomique est incomplète et non annotée. Ils ont bien sûr marqué un grand intérêt pour les données qui seront produites par le séquençage complet du génome de l'Anophèle. En dépit de l'état partiel des connaissances sur ce génome. Ils ont tout de même pu déjà tirer des conclusions. Moins de 20 % des protéines identifiées au stade sporozoïtes sont retrouvées dans le stade où le parasite se trouve dans le sang des patients infectés. Dans le cas de cette forme on a bien sûr des contaminations par des protéines humains. Les formes sporozoïtes, gamétocytes et mérozoïtes diffèrent par l'expression de protéines de surface variantes. Une fois dans les cellules sanguines, le parasite sécrète des protéines qui s'associent pour former un complexe lui permettant d'adhérer aux cellules endothéliales (Knob associated proteins). Une étude plus particulière est menée sur ces protéines, afin de rechercher des inhibiteurs de l'adhésion. Cette approche leur a permis de caractériser au total 1.400 protéines.

apparus : leur système d'exploitation des données repose sur un ensemble de 30 PC. Le laboratoire a développé plusieurs logiciels (SEQUEST, DFCA, DTASelect, Contrast...) permettant l'identification automatisée des peptides à partir de leur profil en spectrométrie de masse et la manipulation de quantités importantes de résultats. Un logiciel permet également l'assemblage des peptides séquencés au sein d'un gène.

Le laboratoire de John Yates a ainsi analysé une cinquantaine de complexes chez la levure<sup>34</sup>. Un programme similaire avait été proposé à Bordeaux dans le cadre des demandes de Génopole. La compétition avec le laboratoire de Yates risque d'être dure à assumer en France, si on ne met pas en place des plateaux techniques aussi performants que celui qu'il maîtrise. Leurs projets s'intègrent dans les préoccupations de l'institut, notamment sur la biologie cellulaire (organelles de la synthèse protéique et sur le *Plasmodium* et ses mécanismes d'adhésion à la cellule endothéliale) :

- *P. falciparum* (différents stades) en vue de générer un vaccin de type ADN
- *S. cerevisiae* : étude de complexes
- Protéome humain
- Lymphome
- Cancer du sein
- Anthrax depuis très récemment

Selon eux, l'approche LC-MS-MS est amenée à remplacer complètement les approches basées sur l'électrophorèse 2D suivie d'analyses par spectrométrie de masse. Un des reproches fait jusque là au système LC-MS-MS par rapport aux systèmes basés sur l'électrophorèse 2D était la difficulté de réaliser des quantifications. Il est possible de réaliser des marquages par différentes approches : incorporation de  $O^{18}$  lors de l'hydrolyse par les protéases, marquage par formation de méthylesters, marquage N-terminal ou encore marquage *in vivo* par emploi de sulfate d'ammonium  $^{15}N$  (beaucoup utilisé pour les applications levure). La méthode de marquage ICAT commercialisée par IBI est trop chère. Les appareils de spectrométrie de masse de type « Ion Traps » ne sont pas trop chers. John Yates dispose d'un ensemble de 6 Ion Trap Thermo finnigan fonctionnels et va recevoir un Q-Tof.

C'est un laboratoire de haute technicité, qui a cherché les moyens les plus économiques pour avoir une grande efficacité en analyse du protéome. Ce qui frappe, c'est sûrement la modestie en terme d'appareillage couplée à une utilisation optimisée des équipements et à un grand savoir-faire qui permet aux chercheurs de réaliser eux-mêmes leur colonne de nanoHPLC. L'accent a été plus mis sur la bioinformatique et sur l'interprétation des résultats que sur la course aux équipements, tout cela animé par une densité impressionnante de chercheurs post-doctoraux (environ 15). Les chercheurs sont biologistes, physico-chimistes ou informaticiens. Tout est orchestré comme cela pourrait être fait dans une biotech<sup>35</sup>, avec un grand souci de productivité. En terme plus pratiques, il faut souligner l'extrême concentration des moyens humains et matériels dans un espace très limité. La pièce où se trouvent les 6 spectromètres de masse avec les appareils de nanoLC est très petite, tout comme le laboratoire, qui renferme paillasses, bureaux et réseaux de PC.

---

<sup>34</sup> Ils travaillent avec une équipe de biologistes qui a « taggué » environ 200 protéines de levures et qui recherche par l'intermédiaire de ce tag, tous leurs partenaires.

<sup>35</sup> Par ailleurs, John Yates anime une compagnie privé qui travaille pour Syngenta.

John Yates est ouvert aux collaborations avec des biologistes français si ses moyens humains le lui permettent : une solution est d'envoyer le chercheur qui ferait lui même les analyses. Clairement, son système est parfaitement adapté à l'analyse de complexes protéiques et il tend à l'adapter pour une utilisation sur la cellule entière. Le tout est soit de fractionner en plusieurs « pools » les protéines cellulaires, soit de faire un très grand nombre d'analyses LC-MS-MS successives pour réussir à identifier au moins un peptide de chaque protéine exprimée. Pour la quantification des protéines exprimées dans une cellule, c'est sans doute encore assez peu précis, lorsque l'analyse est faite sur une cellule entière, par contre cela doit être tout à fait satisfaisant pour l'analyse de complexe. De quoi remettre en question les analyses 2D beaucoup pratiquées en France !

### **3 - Protein Structure Initiative, Lawrence Berkeley Laboratory, Californie.**

*Professor Sung Hou Kim<sup>36</sup>*

Le laboratoire du Professeur Kim constitue l'un des exemples de la *Protein Structure Initiative*<sup>37</sup> (PSI) du National Institute of General Medical Sciences (NIGMS) du NIH. Ce programme est focalisé sur un nombre limité de centres (7 plus 2), avec un coût considérable (5 millions de dollars par an et par centre). C'est en fait un programme test : il s'agit d'évaluer la faisabilité du programme, son coût (humain et matériel) et sa productivité potentielle. Si chaque centre a tout loisir de mettre en œuvre la stratégie de son choix, il y a eu une répartition globale des programmes de recherche (par organisme ou par famille de protéines), en essayant de minimiser (mais sans interdire) les redondances possibles. Un programme sur *Arabidopsis* vient d'être financé cette année. Les entreprises comme *Structural GenomiX* et *Syrrx* participent au programme. Les aventures de Steve Burley, PI qui manageait le programme à Rockefeller University et qui a rejoint Structural Genomix en janvier 2002, témoignent d'un pragmatisme<sup>38</sup> par rapport aux entreprises privées assez impensable chez nous !

On peut se poser la question de savoir si cette initiative académique aura les moyens de rivaliser avec les initiatives privées. On se doute que toutes les structures d'intérêt économique seront réalisées par le secteur privé. On voit bien l'importance que le secteur académique reste dans la course<sup>39</sup> pour ne pas laisser complètement le champ libre aux initiatives privées qui ne publieront pas les structures, ou partiellement, ou trop tard. L'initiative académique européenne devrait être coordonnée avec l'américaine. Est-ce le cas ? Il serait intéressant de voir comment les programmes de génomique structurale français (Génopoles) pourraient être associés à ce programme.

## **III- SOCIETES DE BIOTECHNOLOGIES**

### **1 - Proteome. Beverly, Massachusetts<sup>40</sup>.**

*Peter E. Hodges, Ph.D. Research Group<sup>41</sup>*

---

<sup>36</sup> ALS. Calvin Lab, UC Berkeley. Room 220. Tél : 510.486-4333. [shkim@lbl.gov](mailto:shkim@lbl.gov)

<sup>37</sup> <http://www.nigms.nih.gov/funding/psi.html>

<sup>38</sup> Steve Burley restera PI du New York Structural Genomics Consortium

<sup>39</sup> Le DoE (Department of Energy) considère une initiative d'un montant de 70 millions de dollars pour créer un centre de protéomique au Pacific Northwest National Laboratory (PNNL) à Richland (WA). Ce centre serait ouvert grandement aux universités et aux instituts de recherche pour répondre à différentes problématiques.

<sup>40</sup> <http://www.proteome.com>

<sup>41</sup> 100 Cummings Center, Suite 345M. Beverly, MA 01915. Tel : 978-922-1643. Fax : 978-922-3971. [ph@proteome.com](mailto:ph@proteome.com)

Jim Garell, l'un des fondateurs de *Proteome*, a été aussi le fondateur de la *Quest Protein Database* au *Cold Spring Harbor Laboratory* (NY). Son équipe a été une des premières à développer les logiciels d'analyse et d'utilisation des gels 2D à grande échelle, comme le premier programme vendu par *Pharmacia*. A sa fondation en 1995, *Proteome* avait pour objectifs :

- la production de données en utilisant la technique des gels 2D,
- la création de bases de données bien annotées en nettoyant les données publiques ceci en commençant par le protéome de la levure *S. cerevisiae*, puis en étendant à *C. elegans* en 1998, *S. pombe*, d'autres levures pathogènes, *Mycoplasma* et maintenant les protéomes humains, de rats et de souris.

C'est ce deuxième objectif qui est devenu prépondérant et la compagnie a fermé son laboratoire « humide » depuis environ un an<sup>42</sup>. Ces bases de données associent des informations de génomique fonctionnelle, d'expression, de publications, de recherche de motifs par des algorithmes, et des données phénotypiques (chez les souris mutantes). Ces bases de données permettent donc de disposer de tout un faisceau d'informations provenant de différentes sources sur des résultats obtenus par les techniques de post-génomique (analyse du transcriptome et du protéome). Ces résultats désignent des gènes dont on connaît souvent peu de choses, et qui sont en quantité trop importantes pour être tous analysés et exploités. C'est un outil d'analyse et de tri de cibles potentielles, dont on a pu constater l'importance dans toutes les sociétés visitées. La philosophie de la compagnie était de donner un accès libre aux données pour les institutions académiques et de faire payer les compagnies privées (Philosophie similaire à celle de GENBIO défendue par Amos Bairoch dans le cadre de l'institut de Bioinformatique Suisse).

En décembre 2000, *Proteome* a été racheté par *Incyte*, compagnie qui considère que les laboratoires publics constituent aussi un marché. La remise en question de cette philosophie a probablement conduit les deux fondateurs à démissionner en septembre 2001. Actuellement, *Proteome* a 100 employés essentiellement des biologistes ayant un PhD qui fouillent la littérature pour retirer des informations traduites sous forme d'annotations. Le problème est que la collecte d'informations porte essentiellement sur les résumés des articles, d'où un risque de caractère superficiel des données, pour un spécialiste. Néanmoins, les banques de données qu'ils ont ainsi construites présentent un intérêt pour celui qui par analyse globale aboutit à une protéine, dont il ne sait rien, et sur laquelle il veut rapidement avoir une vision de ce qui a déjà été fait. C'est bien sûr intéressant pour l'industrie pharmaceutique, mais cela peut aussi être utile en milieu académique. Jusque là, les données sur les organismes modèles sont librement accessibles sur leur site<sup>43</sup>. Leurs données sur les autres champignons pathogènes ne sont accessibles que moyennant financement. L'accès à l'analyse du protéome humain, qu'ils ont commencée en 2000, nécessitera une souscription de la part des organismes académiques. Bien sûr, l'accès à toutes leurs bases de données est payante pour toute institution faisant des profits.

La banque de données sur *S. cerevisiae* porte sur 6280 protéines et repose sur 19717 références. La banque de données du protéome humain (human PSD) porte sur 22396 protéines et décrit 135834 propriétés. Une banque spécifique a été réalisée sur les récepteurs couplés aux protéines G (GPCR-PD). Elle n'est pas accessible librement aux académiques. Elle porte sur les données homme, rat, souris. Elle porte sur 740 récepteurs couplés aux protéines G (3000 protéines fonctionnellement liées à ces récepteurs, 9000 références). Y sont décrits en particulier, les ligands,

---

<sup>42</sup> D'une manière générale, l'électrophorèse 2D que nous avons beaucoup développée en Europe est remplacée aux USA par des approches spectrométrie de masse.

<sup>43</sup> YPD yeast (*S. cerevisiae*) Pombe PD (*S.pombe*) Worm PD (*C.elegans*) MycoPath PD (*Mycoplasma*)

les régulations, les variants connus... Pour qu'une information soit entrée dans la banque, elle doit être collectée par un scientifique, puis vérifiée successivement par deux autres scientifiques. Les données sont continuellement mises à jour.

Dans ce type de banque se pose le problème du choix du mode de présentation des données. Leur choix a été de le faire gène par gène et non pas par protéine (on sait bien que par épissage alternatif, un gène donné peut coder plusieurs protéines différentes). La fiche concernant un gène donné procure une indication sur les variants connus d'épissage alternatif, mais la banque n'est sans doute pas exhaustive dans ce domaine (les données sur l'épissage alternatif ne sont pas accessibles). Le type d'information contenue dans une fiche est : rôle dans l'organisme, rôle dans la cellule, propriétés biochimiques, localisation dans l'organisme ou la cellule, implication dans des processus biologiques... Lorsque la structure 3D de la protéine n'est pas connue, les domaines de la protéine sont prédits en faisant une recherche par Blast à partir des structures 3D connues de la PDB. Dans le cas des protéines de souris, les données sur les phénotypes de 1500 mutants ont été introduites. Au niveau expression, les données acquises avaient déjà été introduites avant le rachat par *Incyte*. En ce qui concerne les protéines de fonction non connues, sur la base de comparaisons avec d'autres protéines connues, des expériences sont proposées pour vérifier leurs fonctions potentielles. Il n'y a pas de recherche possible au sein de la banque humaine par mot-clef correspondant à une pathologie donnée, mais pour chaque gène, les variants connus pour générer des maladies génétiques sont indiqués. Les modifications post-traductionnelles connues ne sont indiquées que pour la banque de récepteurs couplés aux protéines G. Rien n'est encore fait sur les autres protéines. Une autre lacune est que la structure 3D de la protéine n'est pas visualisable dans leurs fiches. En fait, ils ont beaucoup de biologistes et peu d'informaticiens, cela explique les imperfections dans la présentation des données. Ils n'ont pas réalisé de banque sur la drosophile, car il en existait une ailleurs. Apparemment, il n'y a aucune initiative sur *Arabidopsis*. Tout ce qui est fait vise l'industrie pharmaceutique. Jusque là les plantes ne les intéresse pas. Il reste une place pour les académiques dans ce domaine.

La fusion avec *Incyte* va aussi permettre d'accroître l'exploitation des données sur l'expression des gènes. L'objectif est maintenant de relier les données sur les protéines aux données d'expression obtenues à partir des microarrays. En particulier, pour une protéine donnée, on aura l'information sur toutes les protéines qui présentent un profil d'expression similaire. Pour cela une connection pourra être faite avec Spotfire. *Incyte* conservera ses propres banques de données<sup>44</sup>. Pour avoir le package des informations, il faudra souscrire à chacune des deux sociétés et un lien entre les deux ensembles de banques de données pourra être établi. Ils sont ouverts à l'établissement de souscriptions regroupées pour des instituts (10 à 15 laboratoires) typiquement nos IFR. L'interlocuteur est Kathy O'Neill<sup>45</sup>.

*Proteome* offre un outil intéressant pour nos laboratoires académiques. Tant que *Incyte* ne change pas la règle du jeu, il est possible d'accéder aux données sur les organismes modèles. Pour le protéome humain, il faudra estimer à partir de l'évaluation des modèles, si cela vaut la peine de mettre en œuvre le système de souscriptions regroupées qu'ils proposent, cela en tenant compte du coût estimé. Leur mode de fonctionnement par extraction de données de la littérature est très proche de celui de l'Institut de l'Information Scientifique et Technique (INIST), à la différence qu'ils sont focalisés sur un aspect donné de la littérature. Une différence importante est que l'INIST utilise aussi les données sur les brevets. Nous avons sans doute avec l'INIST, les masses de connaissances dans les laboratoires de biologie et les grands laboratoires d'informatique français pour la mise en

<sup>44</sup> « EST sequencing information » et « Data base microarrays »

<sup>45</sup> [Keo@proteome.com](mailto:Keo@proteome.com)

œuvre, des possibilités d'avoir des initiatives complémentaires dans le domaine, soit sous forme de start-up sur des thèmes plus spécialisés, soit au travers de la filiale commerciale de l'INIST. L'absence de banque dans le domaine des plantes correspond sans doute à un marché faible, mais en appui sur l'INIST, où le personnel est employé par le CNRS, on peut peut-être envisager la réalisation d'une banque de données sur les plantes en liaison avec le programme Génoplante. En ce qui concerne les pathogènes, nous n'avons pas su quels organismes étaient couverts. Il faut bien sûr dans tout ça tenir compte de ce qui se passe à Genève autour de la *Swiss Prot*. L'INIST représente une possibilité dans le domaine que nous n'exploitons pas assez au CNRS.

En ce qui concerne *Proteome*, il est vraisemblable que des changements importants sont à attendre à la suite du rachat par *Incyte*. Il faudra suivre l'évolution sur un à deux ans. Ce rachat traduit bien l'importance montante du protéome et sa nécessaire association avec le transcriptome. L'histoire de cette société montre l'extrême proximité entre le secteur concurrentiel et le secteur académique. A partir de données publiques, il est possible de générer de nouvelles informations qui éventuellement peuvent donner lieu à des brevets ou à de nouvelles connaissances à accès limité. Doit-on réguler cet accès à l'information et obliger les compagnies, si elles utilisent des données du secteur académique, à permettre l'accès de leurs bases de données au secteur académique ?

## 2 - Structural GenomiX. La Jolla, Californie<sup>46</sup>

*Tim Harris, PhD, President & CEO*<sup>47</sup>

*Eric de La Fortelle, PhD, MBA, Director, Business Development*<sup>48,49</sup>

Une vision très complète et très claire de la stratégie de *Structural GenomiX* pour la découverte de nouveaux médicaments a été donnée. C'est une stratégie très intégrée qui est mise en place, dont on connaît déjà quelques exemples, mais qui devrait se généraliser dans les années qui viennent avec l'apparition des nouvelles cibles et la disponibilité des structures 3D. Selon les spécialistes, la structure 3D est un outil de sélection des bonnes voies pour la conception des médicaments par les chimistes. *Structural GenomiX* a pour objectif principal la détermination à haut débit (5000 structures en 5 ans) des structures 3D de protéines par radiocristallographie et la modélisation moléculaire. Partant de la connaissance du génome, des outils bioinformatiques<sup>50</sup> sont mis en œuvre pour réaliser une annotation fonctionnelle du génome et identifier des cibles potentielles. Dans un second temps, où les étapes de biologie moléculaire et de biochimie sont automatisées et robotisées, les gènes correspondants sont exprimés dans des systèmes hétérologues<sup>51</sup> et les protéines recombinantes purifiées en vue de leur cristallisation. Les méthodologies ont pour but :

- d'améliorer la purification des protéines,
- d'automatiser la cristallisation,
- d'automatiser autant que possible, la détermination des structures,

<sup>46</sup> <http://www.stromix.com>

<sup>47</sup> 10505 Roselle Street, San Diego, CA 92121. (858) 558 4850. [tim@stromix.com](mailto:tim@stromix.com)

<sup>48</sup> [eric@stromix.com](mailto:eric@stromix.com)

<sup>49</sup> Français centralien, qui après une thèse débutée au LURE avec R. Fourme et finie avec Gérard Bricogne en Angleterre, a réalisé un post-doct de plusieurs années au MRC à Cambridge avec G. Bricogne, avant de décider de tenter l'aventure des Biotech. Pour cela, il a suivi une formation d'un an dans une école de commerce, en France, avant d'être embauché par *Structural GenomiX* à San Diego. Philippe Chambon est un des capitaux-risqueur de cette société.

<sup>50</sup> Obtenus grâce au rachat de *Prospect Genomics* (Belmont, CA) en avril 2001. Cette société est spécialisée dans les bases de données de structures 3D de protéines, la modélisation 3D, et la recherche de « leads » *in silico*.

<sup>51</sup> L'expression est faite soit chez *E. coli*, soit par le système baculovirus ou encore par emploi du Sindiki Forest virus pour les protéines membranaires. La levure *Pichia* s'est avérée peu intéressante pour les protéines membranaires contrairement à ce qui était pensé, il y a encore peu de temps. L'expérience négative d'autres utilisateurs, leur a évité d'essayer de mettre ce système en place comme l'ont fait certains. La plupart de leurs protéines bactériennes ont été produites chez *E. Coli*.

- de réaliser du criblage virtuel, une fois la structure connue.

Les cristaux de protéines sont ensuite analysés sur la ligne de lumière que possède *Structural GenomiX* au synchrotron d'Argonne<sup>52</sup> (Illinois), 8 millions de dollars avec la possibilité d'en construire 2 autres si nécessaire et l'installation d'une équipe sur place à Chicago. C'est une stratégie classique<sup>53</sup>, mais qui présente chez *Structural GenomiX* un certain nombre de caractéristiques. D'un point de vue économique, ils se focalisent sur des familles de protéines qui intéressent l'industrie, à savoir les récepteurs nucléaires, les kinases et les protéines bactériennes susceptibles d'être des cibles pour de nouveaux antibiotiques. Les kinases sont retrouvées comme centre d'intérêt de beaucoup de Biotech, mais ce n'est pas le cas pour les récepteurs nucléaires. Nous n'avons pas vu non plus beaucoup de recherches sur les bactéries pathogènes à part l'Anthrax. Leur grande stratégie est d'aborder les protéines par famille, de déterminer quand cela n'est pas fait, la structure 3D de différents membres de la famille et de modéliser la structure 3D des autres membres, ceci afin que les chimistes puissent orienter leur choix dans les criblages chimiques d'inhibiteurs ou d'activateurs potentiels (aide aux chimistes pour prendre les décisions au niveau du travail sur les « leads »<sup>54</sup>). Un certain nombre de structures ont déjà été déterminées par *Structural GenomiX*<sup>55</sup> : leur première structure 3D a été établie en mars 2000, ils en sont à 60, avec une résolution moyenne de 2 Å. Dans le cadre d'un partenariat avec la *Cystic Fibrosis Foundation*, un programme spécifique est mis en œuvre pour l'analyse structurale du récepteur à 12 domaine transmembranaire impliqué dans la mucoviscidose (gène CFTR). Ils vont faire la structure domaine par domaine. Ce projet est établi sur 4 ans.

Pour chaque protéine, dont une structure est souhaitée, tous les orthologues connus chez les bactéries ou les eucaryotes sont recherchés, y compris chez *Arabidopsis*. Les différents gènes orthologues sont clonés et les capacités de cristallisation de leurs produits d'expression sont testées. 100 à 200 gènes voire jusqu'à 300 sont testés pour une protéine donnée. Si la structure 3D n'est pas obtenue pour l'espèce qui les intéresse mais pour une autre, celle souhaitée sera modélisée à partir de la structure établie. Toutes les opérations depuis le clonage, la culture, l'extraction et la cristallisation sont robotisées. Un système a été mis au point pour supprimer les électrophorèses sur gel dans les tests d'expression. Une étape qu'ils considèrent importante est la mise au point des conditions de lyse pour éviter les problèmes d'insolubilité. Les cristallisations sont réalisées dans des plaques à 96 puits. Là encore, la mise au point d'un robot va leur permettre un examen automatique de 10.000 plaques par jour. Dans un premier temps, pour chaque protéine, 100 conditions sont testées, ceci à deux températures différentes. Si rien n'est obtenu au bout de 3 mois, la condition est abandonnée. L'affinement des conditions de cristallisation est aussi automatisée. Toute la partie robotique est réalisée avec une petite société spécialisée. Nous avons vu fonctionner le robot. Nous n'avons rien vu de la partie production et purification de protéines recombinantes. L'analyse automatique des cristaux en formation nécessite des dispositifs d'analyse d'images sophistiqués (le volume dans lequel se forment les cristaux n'étant pas totalement miniaturisé, la mise au point de la caméra sur le cristal n'est pas immédiate, d'où le choix de faire plusieurs clichés sur le même puit).

---

<sup>52</sup> Argonne a plusieurs autres lignes ouvertes par des consortiums ou des groupements pharmaceutiques pour la recherche en biologie structurale : IMCA construite par un consortium de 12 compagnies pharmaceutiques et chimiques, DND fondée par Dupont et Northwestern University, BioCars fondée par University of Chicago, et en cours de construction : SER par un consortium de 20 instituts de recherche conduits par University of Georgia, NE par un consortium d'instituts du Nord-est, incluant Columbia, Cornell, Harvard, Memorial Sloan Kettering...

<sup>53</sup> Conduite aussi chez *Emerald BioStructures* (Bainbridge Island, WA) qui travaille sur l'automatisation du screening de plusieurs conditions de cristallisation en parallèle.

<sup>54</sup> A noter que Tim Harris a des contacts avec Glaxo et aussi avec Philippe Froegel.

<sup>55</sup> A son actif cette société a résolu la structure de LyxS une enzyme critique pour la virulence de nombreuses bactéries. Elle constitue une nouvelle cible intéressante pour la conception de nouveaux antibiotiques.

En ce qui concerne la partie bioinformatique, il faut signaler un effort important sur le développement d'outils (accords avec *Compaq* pour la fourniture de matériel) pour la modélisation et l'analyse de domaines structuraux, l'étude structurale de complexes, le criblage virtuel de molécules, etc... Une base de données de domaines structuraux a été créée (ModBase<sup>56</sup>). *Structural GenomiX* a des accords avec *Rockefeller University* d'un côté et avec *Celera* de l'autre. Le partenariat avec *Celera* est stratégiquement très intéressant pour les 2 sociétés. *Celera* propose ainsi d'intégrer la base de données de structure de protéines modélisées par ordinateur, représentant 58% des protéines disponibles. Pour le chercheur, c'est un accès direct à des informations utiles même si la valeur de ces modèles est sujette à caution. Ainsi, ils ont fourni leur banque ModBase de 300.000 modèles de structures 3D de protéines à *Celera* contre l'obtention d'ADNc qui les intéresse. Cette banque de 300.000 modèles réalisée à partir de la PDB est du domaine public<sup>57</sup>, mais si une firme s'en sert pour découvrir un médicament, elle doit payer une redevance à *Structural GenomiX*, ce que *Celera* n'aura pas à faire avec cet accord.

Actuellement, la société a 130 employés et un capital de 85 millions de dollars. Environ 1/3 du personnel travaille dans le domaine de la modélisation, du docking et du design d'inhibiteurs. Leur objectif une fois qu'ils auront assez de structures 3D établies est de rechercher des molécules chimiques d'intérêt pharmaceutique par tests de cristallisation, qui, si la technique est assez automatisée, peut être une approche rentable de screening. Pour cela, ils ont acheté des bibliothèques de molécules chimiques et ils ont établi un partenariat avec plusieurs sociétés pharmaceutiques spécialisées dans la recherche de « hits ». Lorsqu'ils établissent une structure 3D, ils essaient d'en tirer un maximum d'informations sur le plan de la fonction et du mécanisme d'action. Ils vont s'équiper en RMN uniquement pour la détermination de la structure des molécules chimiques ; pour la 3D des molécules biologiques, ils restent focalisés sur l'approche RX et ils n'envisagent pas non plus d'utiliser la RMN pour le criblage de molécules chimiques potentiellement intéressantes.

Clairement, et malgré une impression de solidité et de compétence, *Structural GenomiX* paraît plus en phase initiale que *Syrrx*. Au niveau cristallisation par exemple, *Syrrx* est plus avancé et compétent en robotisation. La force de *Structural GenomiX* est sans doute de ne pas dépendre d'une autre société et de pouvoir établir plus librement des accords et des contrats avec beaucoup d'entreprises. Leur force est aussi de coupler aspects modélisation et cristallisation et d'avoir un fort potentiel en modélisation.

A noter que l'approche de génomique structurale basée sur le test de nombreuses espèces est aussi celle du consortium universitaire<sup>58</sup> qu'on a visité. C'est sans nul doute la meilleure approche, pour avoir toutes les chances d'établir la structure 3D d'une protéine donnée ou de la protéine d'une famille donnée. Les mutations naturelles distinguant les espèces sont assez larges pour qu'on ait des chances de trouver dans une collection de 100 à 200 protéines de la même famille, une protéine qui a de bonnes propriétés de cristallisation. C'est la stratégie opposée de ce qui est fait en France ou dans les programmes de génomique structurale, un organisme unique est étudié et on essaye de cristalliser les protéines inconnues qui veulent bien cristalliser. L'approche

---

<sup>56</sup> ModBase a été développée en milieu universitaire à *Rockefeller University* et une licence exclusive a été donnée à *Structural Genomix*.

<sup>57</sup> Lorsqu'il y a risque de publication de la structure par un laboratoire académique, ils publient, ce qui leur permet de se faire connaître scientifiquement.

<sup>58</sup> ils font partie du consortium BioXtal, E. de la Fortelle a fait sa thèse en même temps que Guillaume Lhermite qui a initié ce consortium.

américaine semble nettement plus efficace. Bien sûr, elle implique de la grande échelle pour clonage, production et cristallisation.

### 3 – GeneFormatics. La Jolla, Californie<sup>59</sup>

Steve Betz, PhD, Senior Manager, Protein Production<sup>60</sup>

Brian Hoffman, Manager, Research & Genome Analysis

Cette société a été fondée en 1998 et a levé 41 millions de dollars<sup>61</sup> sur le principe de la découverte des nouvelles fonctions des protéines à partir de la connaissance de leurs séquences. Sur cette base, la compagnie a créé une plate-forme logicielle DIAMOND (*Diverse Integrated Automated Method of Novel Discovery*) combinant différents logiciels<sup>62</sup> pour virtuellement partir du gène d'une protéine inconnue, essayer de prédire sa structure et en déduire sa fonction. Partant là encore des données génomiques, des modèles structuraux et des sites fonctionnels sont établis. La société tire sa force d'une méthode de modélisation de la structure et de la fonction des protéines basée sur un algorithme développé au *Scripps Research Institute* par Jacqueline Fetrow et Jeffrey Sklonick<sup>63,64</sup>. Cet algorithme aligne la séquence sur la structure la plus ressemblante dans les bases de données. Le site actif est identifié dans l'alignement séquence-structure grâce à un descripteur en 3 dimensions. L'annotation fonctionnelle des motifs structuraux (sites catalytiques, sites de fixation de métaux, de nucléotides...) est réalisée à l'aide de descripteurs structuraux (*fuzzy functional forms* ou FFF).

Société essentiellement dédiée à la modélisation des structures *in silico*<sup>65</sup>, *GeneFormatics* vient de fusionner avec une compagnie spécialisée dans l'utilisation de la RMN pour l'obtention de structures de protéines<sup>66</sup>, dans le but de mesurer la « flexibilité » de ces protéines et analyser leurs interactions avec de petites molécules. La seule originalité dans la stratégie de *GeneFormatics* est l'utilisation de la RMN, qui ne nécessite pas d'étape de cristallisation mais dont le champ d'application ne concerne que les petites protéines solubles. Les cibles potentielles pour de nouveaux médicaments sont les mêmes que partout ailleurs (kinases en particulier). Ils utilisent le système *In vitro gene* pour la production des protéines. Ils ont acquis un appareil permettant de réaliser 2300 cultures de 1 ml. Ils disposent d'un robot permettant de réaliser les PCR pour les clonages, l'isolement des protéines qui sont analysées directement par spectrométrie de masse pour éviter les gels (plaque MALDI). Le marquage des protéines pour la RMN doit être réalisé soit lors des cultures bactériennes, soit *in vitro*. Selon eux, en une semaine, ils peuvent établir la structure 3D d'un domaine protéique par RMN. En parallèle à la RMN, sur le plan expérimental, ils développent

---

<sup>59</sup> <http://www.geneformatics.com>

<sup>60</sup> 5830 Oberlin Drive, Suite 200. San Diego, CA 92121-3754. Tél : 858-882-5927. [Stevebetz@geneformatics.com](mailto:Stevebetz@geneformatics.com)

<sup>61</sup> 62 personnes dont 30 PhD.

<sup>62</sup> Prospector™, Sapphire™ pour prédire les structures 3D. FFF™ pour la recherche automatique de sites fonctionnels potentiels (motifs kinase, phosphatase, lipase, protéase, site de fixation de cofacteurs, sites de liaison d'ions métalliques). C'est associé à une banque de données de motifs fonctionnels où des liens sont faits avec la littérature. Logiciel pour la recherche de tête de série (leads) : Touchstone™.

<sup>63</sup> Skolnick & Fetrow, 2000, From genes to protein structure and function : novel applications of computational approaches in the genomic era. *Trends Biotechnol.* 18, 34-39

<sup>64</sup> Jeffrey Sklonick est un bioinformaticien impliqué dans l'approche dite du THREADING ou de l'enfilage

<sup>65</sup> approche conduite aussi par des sociétés telles que *Structural Bioinformatics* (San Diego, CA) *Prospect Genomics* (Belmont, CA) ou *Myriad Proteomics* (Salt Lake City, UT), un consortium regroupant *Myriad Genetic*, *Oracle* et *Hitachi* d'un montant de 185 millions de dollars.

<sup>66</sup> tout comme *Structure-Function Genomics* (Princeton, NJ) et *Triad Therapeutics* (San Diego, CA)

des tests des propriétés enzymatiques prédites *in silico* avec un appareillage Beckmann d'électrophorèse capillaire et un système multi-spectrométrie.

*GeneFormatics* a donné l'impression d'une entreprise virtuelle à qui il a été demandé de s'investir dans une approche plus pratique pour l'identification de cibles. Il est clair que nos interlocuteurs n'avaient qu'une vision théorique de la biologie et des approches expérimentales qu'ils devaient apprendre à maîtriser. En l'absence d'informations plus concrètes, il est impossible d'avoir une idée précise de ce que la compagnie est capable de vendre, même si elle prétend être la seule compagnie de génomique structurale ayant des clients. Des accords avec diverses entreprises, *Celera* en particulier, montrent cependant que *GeneFormatics* a su intéresser divers groupes aux informations structurales qu'elle est en mesure d'apporter à travers son FFF. Malgré tout, la partie biologique de la compagnie a semblé complètement déconnectée de la partie *in silico*, impression confirmée par la visite qui a montré des locaux pratiquement vides et sans activité significative de recherche en biologie. Brucker a cependant investi dans leur entreprise, pour leur permettre de s'équiper en spectrométrie de masse, RX et RMN. Lorsque nous avons visité les installations, il y avait un gros étalement d'appareillage, mais pas le sentiment qu'il y avait beaucoup d'activité. Il serait intéressant de savoir ce qu'est devenu cette entreprise dans un an. De toutes les entreprises développant des stratégies de génomique structurale, *GeneFormatics* est apparu la plus fragile.

#### 4 – Syrrx. La Jolla, Californie<sup>67</sup>

Wendell Wierenga, PhD, CEO<sup>68,69</sup>

Kenneth Goodwill, PhD, Associate Director, Protein Chemistry<sup>70</sup>

Cette compagnie est une « spin-off » du *Genomic Institute of the Novartis Research Foundation*<sup>71</sup> (GNF) dont l'objectif est la détermination de structures 3D de protéines d'intérêt pharmaceutique. Cette compagnie en 18 mois a réussi une levée de fonds 80 millions de dollars et a actuellement un personnel de 95 personnes. *Syrrx* est une compagnie de tout premier plan (et probablement le leader) dans le domaine de la génomique structurale. La visite des installations a démontré sans aucune ambiguïté que le haut débit existait réellement dans ce domaine. Si la compagnie met en avant une stratégie expérimentale de découverte de nouveaux médicaments à partir d'une approche structurale finalement assez proche de celle que *Structural GenomiX*, l'avance technologique (protégée par des brevets importants dans la miniaturisation des dispositifs de cristallisation) semble assez confortable. Elle peut se définir de la manière suivante : (a) la production de protéines se fait effectivement à haut débit, depuis la culture des bactéries jusqu'à la purification des protéines recombinantes<sup>72</sup> ; (b) leurs robots sont tels qu'ils peuvent obtenir 2 mg de protéine recombinante par chaque condition de culture, répartis de manière entièrement automatique dans 12 x 96 puits par un robot, permettant ainsi de définir près de 1200 conditions différentes (en fait 2 x 600, car la cristallisation est réalisée à deux températures différentes<sup>73</sup> dans deux pièces

---

<sup>67</sup> <http://www.syrrx.com>

<sup>68</sup> 10450 Science Center Drive, Suite 100. San Diego, CA 92121. 858.622-8528. [Wendell.wierenga@syrrx.com](mailto:Wendell.wierenga@syrrx.com)

<sup>69</sup> qui avait travaillé un moment chez Juvenal à Paris.

<sup>70</sup> [ken.goodwill@syrrx.com](mailto:ken.goodwill@syrrx.com)

<sup>71</sup> Novartis a créé, il y a 3 ans un institut à San Diego sur la génomique et la bioingénierie, ils font aussi du design d'équipement, en particulier d'automates (200 personnes). *Syrrx* bénéficie de tout ce potentiel de compétences d'où l'avance par rapport à *GenomiX*.

<sup>72</sup> Cultures à haute densité d'*E. coli* (souche optimisée protéase<sup>-</sup>), 96 clones cultivés en même temps de façon à toujours travailler avec des robots utilisant 96 échantillons en même temps. La purification sur colonne d'affinité est robotisée, elle prend 4 heures. Ils atteignent ainsi une pureté d'environ 75 %. Les purifications ultérieures sont réalisées sur FPLC.

<sup>73</sup> à 4°C, et à température ambiante.

dédiées); (c) le suivi de la formation des cristaux est entièrement automatisé<sup>74</sup> (même si l'analyse finale des images est faite manuellement, car plus rapide). Ils ont commencé par s'intéresser à une bactérie thermophile *Thermus thermotoga* (cristallisation de 200 ORF) pour mettre le système de radiocristallographie en route au niveau bactérien (ils s'investissent beaucoup sur les mécanismes de résistance aux B-lactamase). Au niveau des eucaryotes, ils investissent sur *C. elegans*. La différence avec l'approche mise en œuvre chez *Structural GenomiX* est la miniaturisation de l'étape de formation de cristaux, qui simplifie l'analyse d'image automatique (l'objet à regarder est toujours très petit, sa taille est de l'ordre de grandeur de la profondeur de champ de l'objectif, la mise au point est simplifiée). Les cristaux obtenus sont montés à San Diego, envoyés au LBL de Berkley où ils disposent aussi de robots<sup>75</sup>, puis sont analysés avec la ligne de lumière que Syrrx y possède<sup>76,77</sup>. En une semaine, ils analysent environ 200 cristaux (plusieurs cristaux sont analysés pour une même protéine). Ils considèrent qu'ils ont en moyenne un résultat positif sur 10 % de ces cristaux. Une fois les structures établies, ils font du screening virtuel par informatique de ligands potentiels, ceci à partir d'une énorme banque de molécules dont ils disposent (2,7 millions de molécules). Ils travaillent dans ce cadre avec une société de San Diego dédiée au « docking » de molécules (*Molsoft*). En parallèle, ils développent leurs propres algorithmes (8 personnes, c'est clairement moins que chez *Structural GenomiX*). Ils ont aussi une équipe de Chimie. Au niveau humain, comme la plupart des sociétés visitées, ils s'intéressent aux kinases. Ils ont déjà exprimé 150 kinases de différentes espèces, en général, par le système baculovirus. *C. elegans* qui présente beaucoup d'enzymes qu'on retrouve chez l'homme est intéressant dans ce cadre.

*Syrrx* a des liens étroits avec *Cubist Pharmaceuticals* (design d'antibiotiques et d'antibactériens), *Molsoft* (criblage virtuel). Les liens restent très forts avec Novartis, ce qui représente une grande force pour eux. Ils font partie de la *Protein Structure Initiative*<sup>78</sup> subventionné par le NIH. Ils ont en charge une partie des cristallisations, ce qui s'adapte parfaitement bien aux compétences qu'ils ont développées en matière de robotisation et de miniaturisation des procédés de cristallisation.

Manifestement, ils n'ont pas décrit réellement les protéines sur lesquelles ils investissaient à fond car ils sont sous contrat avec d'autres sociétés pharmaceutiques. *Thermotoga* choisi pour mettre en route leur système, a pu leur servir pour les recherches de cibles potentielles de nouveaux antibiotiques chez les bactéries. L'utilisation de *C. elegans* pour des systèmes conservés chez l'homme est sans doute plus d'actualité dans leurs projets. C'est la société qui nous a le plus impressionnés. Comparée à *Structural GenomiX*, l'industrialisation de la production de protéines et la robotisation pour la production de cristaux sont très en avance. *Syrrx* a de réelles capacités de production de nombreuses structures et montre la faisabilité de ces stratégies à haut débit. Là, c'était concret : on a vu les employés à l'œuvre devant les machines. C'était tard dans l'après-midi mais

---

<sup>74</sup> Les premiers tests effectués visent à estimer le degré de solubilité de la protéine. Si elle est satisfaisante, ils lancent les tests de cristallisation qui sont miniaturisés, 15 nl par essai (10 plaques de 96 puits par protéine). Tout a été conçu pour qu'il n'y ait pas d'évaporation. Un film ferme les puits de manière hermétique à la fin. Lorsque la protéine s'agrège et qu'ils ont des problèmes de cristallisation, leur stratégie est de produire une large gamme de mutants, par mutagenèse aléatoire du gène et de rechercher des mutants qui cristallisent, d'où la nécessité de tout faire à grande échelle.

<sup>75</sup> Ce robot installé en collaboratin avec l'équipe de Thomas Earnest 5Advance Light Source au LBL) devrait permettre de multiplier par 10 les structures qui pourront être analysées par une ligne de lumière, pour en avoir 1000 par jour.

<sup>76</sup> Le prix de 2.4 millions de dollars pour la ligne de lumière de l'ALS a été partagée entre Syrrx et le GNF pour obtenir 75% du temps d'utilisation. Les autres 25% sont réservés au public.

<sup>77</sup> Lorsque nous étions là-bas, ils allaient envoyer du personnel à l'ESRF à Grenoble car le synchrotron d'Argonne était arrêté pour 6 semaines.

<sup>78</sup> Dans la cadre du Scripps Research Institute

c'était encore en plein activité, contrairement à *GeneFormatics*, et même *Structural GenomiX* où l'activité était nettement moins débordante. La cadence est impressionnante.

### **5 – Zyomyx, Fremont, Californie.**<sup>79</sup>

*Jonathan Forman, PhD, Manager of DNA Microarray Technology*<sup>80</sup>

*David Wilson, PhD, Research Scientist*<sup>81</sup>

Cette compagnie est à l'interface des sciences de l'ingénieur et de la biologie. Leur objectif est de fabriquer des puces à protéines afin de pouvoir mieux analyser le protéome cellulaire (interaction, biomarqueurs<sup>82</sup>...). Cette compagnie est relativement récente, sans site Web (en construction) et il est donc difficile d'en cerner la stratégie exacte. Nous n'avons eu qu'une présentation rapide, et donc superficielle, des nombreux développements technologiques qui sont actuellement réalisés par *Zyomyx*. Ils concernent la détection (fluorescence, résonance plasmonique, ellipsométrie, etc...), les microsystèmes, les surfaces et les matériaux, la microfluidique, etc... La fabrication des puces à protéines se fait à partir de la mise au point de procédés concernant la chimie de surface, les technologies de distribution parallèle et des microfluides. Ils ont notamment mis au point des surfaces spéciales pour ces biopuces ainsi que des traitements chimiques devant permettre une meilleure fixation des sondes ADN et une augmentation très significative du signal. *Zyomyx* veut mettre au point une puce à protéines qui devrait présenter des caractéristiques intéressantes (10000 plots par cm<sup>2</sup>) pour détecter et quantifier les protéines grâce à des collaborations avec des sociétés disposant d'anticorps en nombre suffisant pour arriver à la taille d'une puce à protéine. *Zyomyx* se concentre donc sur la technologie même des puces et développe des alliances avec des entreprises produisant des anticorps (*BD Biosciences, Cambridge Antibody Technology et Dyax Corporation*). Ces alliances sont complétées avec des sociétés disposant de techniques de détection, en particulier *Axon* qui fabrique et commercialise un scanner de microarray très populaire qu'ils adapteront à la collecte de données et l'analyse des biopuces, ou *MDS Proteomics* impliquée dans la spectrométrie de masse.

La protéomique est un domaine qui nécessite beaucoup d'innovation technologique. *Zyomyx* développe des outils pour permettre la protéomique à haut débit, pour miniaturiser les dispositifs d'analyse des protéines et pour permettre l'analyse multiparallèle. Si l'ensemble de la stratégie présentée semble tout azimut, on peut se poser la question de la réalité de certains projets. En effet, les puces à plots présentées ne semblent pas être faciles à utiliser avec des systèmes de détection comme la résonance plasmonique. L'intérêt du dispositif pour des puces à ADN a été souligné.

### **6 - Guava Technologies, Burlingame, Californie**<sup>83</sup>

*Philippe Goix, PhD, MBA, Founder & President*<sup>84</sup>

*Jack Giroux, Executive Vice President of Commercial Operations*<sup>85</sup>

---

<sup>79</sup> <http://www.zyomyx.com>

<sup>80</sup> 26101 Research Road. Fremont, CA 94538.510.266.7734. [jforman@zyomyx.com](mailto:jforman@zyomyx.com)

<sup>81</sup> [dwilson@zyomyx.com](mailto:dwilson@zyomyx.com)

<sup>82</sup> Un biomarqueur est une molécule ou macromolécule biologique (ou un ensemble de molécules ou macromolécules biologiques) associée à un état physiologique ou pathologique ou au suivi d'un tel état.

<sup>83</sup> <http://www.guavatechnologies.com>

<sup>84</sup> 863C Mitten Road. Burlingame, CA 94010-1303. Tel : 650.552-0706. [pgoix@guavatechnologies.com](mailto:pgoix@guavatechnologies.com)

<sup>85</sup> [jgiroux@guavatechnologies.com](mailto:jgiroux@guavatechnologies.com)

La visite de cette entreprise, un peu en marge des préoccupations directes de la mission d'étude, s'est révélée très intéressante. Cette compagnie, fondée par un français, Philippe Goix, a créé un cytofluorimètre personnel, astucieux, de petite taille, facilement utilisable, qui peut être considéré comme « l'apple » des cytofluorimètres. Ce cytomètre de flux de paillasse avec un laser d'excitation et deux voies de détection par deux photomultiplicateurs est capable d'analyser de très petits nombres de cellules avec une très bonne sensibilité. Le produit actuellement commercialisé, est très convaincant. Il repose sur une idée simple, l'exploitation de systèmes miniaturisés pour la circulation des fluides et l'adéquation à un besoin. En même temps, *Guava Technologies* développe des méthodes d'analyses (viabilité cellulaire, expression de protéines, apoptose). Cette offre globale est originale et devrait intéresser de nombreux laboratoires. Son prix (50 000\$) est toutefois un peu élevé pour le marché français, même s'il se situe bien en dessous des prix des trieurs et des gros cytomètres de flux. Actuellement, sans concurrent, on peut imaginer que son prix va baisser et que cet appareil devrait trouver une niche commerciale. La sortie d'un modèle permettant de travailler sur des plaques 96 puits devrait aussi positionner cet appareil comme un outil de criblage à haut débit dans le domaine des essais cellulaires.

## 7 – SurroMed, Mountain View, Californie<sup>86</sup>

*Michael Natan, CTO<sup>87</sup>*

Cette compagnie développe un ensemble de technologies très originales pour réaliser des analyses biologiques multi-parallèles, à haut débit et à moindre coût. La stratégie de Surromed consiste à trouver des biomarqueurs de maladie. Deux technologies ont été principalement développées :

- le Surroscan permettant d'identifier plus de 200 types cellulaires (marqués ou non avec des anticorps fluorescents) sur des microvolumes de sang ou d'autres fluides biologiques et qui remplace les techniques de cytométrie en flux,
- des nanoparticules ou un système de nano-code-barre<sup>88</sup> afin de multiplexer les analyses sanguines<sup>89</sup> (Elisa, Metabolite, Proteome) et qui permet d'étiqueter la cible de son choix grâce à une liaison de l'étiquette avec le réactif de son choix (anticorps, ligand). Ce type de structure combine l'avantage des puces à protéines (dispositifs de fixation de protéines type puce CIPHERGEN) pour l'analyse protéomique et la simplicité d'utilisation des billes magnétiques dans les divers essais biologiques.

La miniaturisation de ces technologies est en cours, ce qui devrait permettre leurs utilisations pour le phénotypage des animaux de laboratoires, les souris en particulier.

*Surromed* ne vend pas directement sa technologie. Elle signe des partenariats avec des hôpitaux ou des institutions lui permettant d'avoir accès à des cohortes de malades. Les exemples présentés (Asthme, maladies psychiatriques, diabète, arthrite, maladies auto-immunes ...) sont convaincants. Un accent est mis sur les études longitudinales (pour l'identification de marqueurs de progression de la maladie) et sur les études pharmacologiques (répondeurs/non répondeurs, profil

<sup>86</sup> <http://www.surromed.com>

<sup>87</sup> 2375 Garcia Avenue, Mountain View, CA 94043-1104. [mnatan@surromed.com](mailto:mnatan@surromed.com). Tel: 650-230-1600. fax: 650-230-1927

<sup>88</sup> Ces petites barres, formées d'empilements de couches d'or et d'argent dont la fabrication est d'une simplicité remarquable, peuvent être utilisées dans une très grande diversité d'applications. Ces nanobarres peuvent être fonctionnalisées avec des protéines, des acides nucléiques, de petites molécules. Lorsqu'elles sont fonctionnalisées avec des molécules permettant la fixation de petites acides, basiques, hydrophobes, etc... elles peuvent être utilisées dans des applications de protéomique quantitative. Si des anticorps sont fixés à la surface, elles peuvent être utilisées dans l'analyse quantitative d'antigènes dans des fluides biologiques. La diversité des structures susceptibles d'être obtenues (à l'image des codes barre) permet le tri des nanobarres et ainsi l'identification et la quantification en parallèle des diverses molécules qui ont servi à les fonctionnaliser.

<sup>89</sup> Ce système permet également de mesurer des ARN.

pharmacodynamique). Ces patients sont phénotypés, c'est à dire analysés sur le plan clinique, le plus précisément possible. Parallèlement, des études biologiques très larges, qui visent à intégrer l'ensemble des paramètres biologiques mesurables (cellulaires, protéomiques, métaboliques, etc...), des paramètres cliniques obtenus lors du traitement d'un malade, ainsi que toute autre information disponible (expression génique, etc...) sont entreprises. *Surromed* mesure des grandes séries de paramètres hormonaux et métaboliques classiques et recherche également à mettre en évidence des molécules de petit poids moléculaire (l'orgeomics) par spectrométrie de masse après séparation des protéines par HPLC. Ils font aussi de la nanoextraction en phase solide avant la spectrométrie de masse. La stratégie de *Surromed* est une sorte d'épidémiologie biologique à grande échelle, faisant curieusement l'impasse sur les SNPs<sup>90</sup>, mais centrée sur la biologie cellulaire et la protéomique. L'analyse bioinformatique des résultats (datamining par différentes méthodes) est annoncée comme un point fort.

On peut se demander comment *Surromed* va trouver de nouveaux biomarqueurs. La biologie à grande échelle, l'exploitation poussée des résultats par du datamining très performant, l'utilisation de nouveaux outils comme la cytométrie de flux en routine ou la spectrométrie de masse sont sans doute des atouts. Une telle stratégie globalisante est très certainement intéressante, mais on peut questionner la véritable valeur de l'information qui sera retirée et l'utilisation qui pourra en être faite. En effet, la compréhension de la relation « génotype + environnement = phénotype » (qui est mise en avant) implique la maîtrise d'un ensemble considérable de paramètres que l'on commence seulement à étudier chez les bactéries... De là à en faire un système d'aide au diagnostique, il y a un grand pas qu'il semble encore bien difficile à franchir. La visite des laboratoires a permis de voir non seulement la réalité du travail expérimental, mais aussi la simplicité remarquable des technologies mises en jeu.

## **8 – CIPHERGEN, Fremont, Californie.<sup>91</sup>**

*William Rich, President & CEO<sup>92</sup>*

*Veronique Riethuisen, Director of Business Development<sup>93</sup>*

*Egisto Boschetti, PhD, Director of R&D<sup>94</sup>*

*Matthew Hogan, Vice President and CFO<sup>95</sup>*

Cette compagnie apparaît comme une compagnie assise (par comparaison aux autres). Elle a développé une technologie d'analyse du protéome de bas niveau permettant d'être utilisée en clinique et de découvrir de nouveaux biomarqueurs. Cette technologie associe des puces à protéines de moindre qualité que chez *Zyomyx* et un MALDI-TOF peu performant. L'ensemble constitue toutefois un système extrêmement facile à utiliser par un chercheur peu familier avec la spectrométrie de masse. Cette technologie rencontre un succès important en recherche clinique et mitigé dans le secteur académique. Une stratégie alternative pour le secteur académique semble se mettre en place. Le procédé consiste à adsorber les protéines dans un premier temps sur une puce à protéines sur 8 supports différents (surface phase inverse, anioniques, cationiques, affinité pour des métaux, etc...) permettant l'adsorption des protéines selon des caractéristiques différentes et dans le but de fractionner des échantillons biologiques pour leur analyse par spectrométrie de masse. Des

<sup>90</sup> *Surromed* considère qu'il n'y a pas de gènes candidats valables pour l'instant

<sup>91</sup> <http://www.ciphergen.com>

<sup>92</sup> 6611 Dumbarton Circle. Fremont, CA 94555. Tel: 510-505-2100. Fax: 510-505-2101. [werich@ciphergen.com](mailto:werich@ciphergen.com)

<sup>93</sup> [vriethuisen@ciphergen.com](mailto:vriethuisen@ciphergen.com)

<sup>94</sup> [egisto.boschetti@biosepra.com](mailto:egisto.boschetti@biosepra.com)

<sup>95</sup> [mhogan@ciphergen.com](mailto:mhogan@ciphergen.com)

logiciels sont fournis pour l'analyse des profils obtenus, avec des licences annuelles. La conception est simple et est très bien adaptée à la protéomique clinique dont l'objectif est d'identifier des marqueurs d'un état pathologique donné, dans un tissu donné. Il serait sûrement intéressant de pouvoir évaluer cet outil dans plusieurs pathologies. La limitation principale tient au poids moléculaire des protéines qui doit être inférieur à 30 kDa. Ils considèrent leur outil comme complémentaire du gel 2D. Seulement 20 microlitres de serum sont nécessaires. Les développements actuels visent d'une part à permettre l'augmentation significative du nombre d'analyses (puces 96 puits) et d'autre part à coupler les puces à protéines à l'analyse par MS/MS.

*Ciphergen* est la seule entreprise visitée qui commercialise effectivement un appareil ou une technologie. Par rapport à l'esprit de la mission, *Ciphergen* est un peu à part. D'une part nous avons eu droit à une présentation très stratégique par l'ensemble de l'équipe de direction. Cette approche très marketing a peut être empêché d'approfondir certains aspects de la stratégie de l'entreprise qui n'ont pas paru évidents (l'acquisition de *BioSeptra* et la volonté d'affirmer que l'appareil peut être une aide à la plurification de protéines à grande échelle). D'autre part, *Ciphergen* propose un produit simple vendu clé en main à une clientèle bien ciblée (cliniciens). Bien qu'étant un outil de protéomique, l'appareil vendu par *Ciphergen* n'est pas en mesure de rivaliser avec un Maldi-Tof, par contre il répond à certaines questions posées par les préparations d'échantillons en milieu clinique pour l'analyse protéomique. Il ne semble pas que l'objectif de l'entreprise soit d'investir massivement dans le type de recherche que fait *SurroMed* ; pourtant les démarches Nanobarcode et puces à protéines sont très semblables, même si la première semble être susceptible d'applications bien plus larges.

#### 9 – **Abgenix, Fremont, Californie**<sup>96</sup>

*Greg Landes, Ph.D. – VP, Product Discovery*<sup>97</sup>

*Bruce Keyt, Ph.D. - VP, Preclinical Research*<sup>98</sup>

*Mark Nocerini – Director, Assay Development*<sup>99</sup>

*Abgenix* est une entreprise qui produit des anticorps pour le traitement de maladies, en particulier le cancer, et les problèmes liés aux maladies autoimmunes, les transplantations, etc... Les implications thérapeutiques des anticorps humanisés produits par cette société sont d'un intérêt médical considérable. Le cancer, les maladies infectieuses pour lesquels on ne dispose pas d'antibiotiques efficaces sont des cibles majeures pour ces anticorps. Ces anticorps sont produits par des souris ayant acquis une grande partie des gènes des chaînes lourdes et légères des immunoglobulines humaines. Le développement de la technologie des *XenoMouse* pour produire des anticorps humains chez la souris a paru tout à fait remarquable et susceptible d'applications majeures à moyen terme. Par ailleurs, l'analyse physiologique de telles souris devrait pouvoir faire l'objet de programme de recherche en collaboration avec le réseau des Génopoles (à Strasbourg et/ou Lyon).

Lors de cette visite, nous avons été frappés par la volonté des responsables de ne donner qu'un minimum d'informations. Par exemple, la réponse à une question sur la réalité de la stratégie protéomique qu'ils affichent est restée vague, or ils ont des relations avec Mathias Mann, un

---

<sup>96</sup> <http://www.abgenix.com>

<sup>97</sup> 6701 Kaiser Drive, M/S 11, Fremont, CA 94555. Tel: 510- 608-6509. Fax: 510-608-4606. [landes\\_g@abgenix.com](mailto:landes_g@abgenix.com)

<sup>98</sup> [keyt\\_b@abgenix.com](mailto:keyt_b@abgenix.com)

<sup>99</sup> [nocerini\\_m@abgenix.com](mailto:nocerini_m@abgenix.com)

spécialiste mondialement connu de la protéomique. Ils doivent donc bien avoir des programmes précis dans ce domaine !

## **10 - Signature BioScience, Hayward, Californie<sup>100</sup>**

*Dr. John Hefti, MD, PhD, Founder & CTO<sup>101</sup>*

Cette entreprise permet une analyse globale des composés protéiques de la cellule et est fondée à partir de brevets de John Hefti sur l'utilisation des micro-ondes et des ondes radio pour analyser les modifications de l'environnement immédiat des protéines induites par leurs changements de conformation lorsqu'elles fonctionnent ou qu'elles fixent un ligand (technologie MCS pour Multipole Coupling Spectroscopy). Cette technique permet donc en particulier de suivre les effets de drogues sur la cellule, ce qui en fait un outil intéressant pour le screening biologique. Apparemment les profils seraient plus ou moins spécifiques de l'activation d'une voie fonctionnelle donnée.

Il est assez remarquable que John Hefti, physicien, cherche ensuite à développer sa technologie pour mettre au point de nouvelles méthodes de criblage afin d'identifier de nouvelles cibles et de nouveaux médicaments. Il a en effet développé une plate-forme (WaveScreen) qui permet d'appliquer la technologie MCS au criblage de molécules sur cellules. La réponse analysée est alors la réponse intégrée lors de l'addition de la drogue. La technologie développée peut être utilisée pour de nombreuses applications (identification de mismatch lors d'hybridations d'acides nucléiques, etc...).

## **IV - ANALYSE ET CONCLUSIONS**

Nul ne doute que le développement de nouveaux ordinateurs puissants, de robots et d'autres outils « high-tech » va permettre aux chercheurs d'obtenir la structure de centaines de protéines dans le temps qu'il fallait auparavant pour en obtenir une. La protéomique, tout comme la génomique, représente l'émergence d'un nouveau moyen de faire de la recherche qui ne dépend plus de l'expérimentation de modèles spécifiques sur le comportement cellulaire, mais d'un flot de données qu'il faut produire, comparer, analyser et exploiter. Ce sont les défis de l'industrialisation et de la modification profonde du concept de la recherche en biologie (les données précèdent l'hypothèse) que les initiatives décrites dans ce document tentent de relever. L'importance de la protéomique dans les applications cliniques et l'enjeu des nouveaux biomarqueurs pour la physiopathologie mais surtout le suivi et le traitement des maladies chroniques sont évidents. On peut se demander s'il n'y a pas un déplacement, vers cette approche protéomique, de l'intérêt des laboratoires pharmaceutiques en complément de l'approche génétique dans les maladies complexes. En effet, jusqu'à présent, peu de cibles sont venues de la mise en évidence de gènes de prédisposition dans les maladies complexes (diabète, asthme, maladies neurodégénératives ou cardiovasculaires).

Concrètement, la protéomique doit répondre à 3 enjeux majeurs : les avancées technologiques, la création d'un modèle économique et le futur légal des données.

### **1 - Les avancées technologiques**

Le développement de l'automatisation et de la robotisation des unités de productions va permettre un changement d'échelle très net dans l'analyse des protéines et marque le début de

<sup>100</sup> <http://www.signaturebio.com>

<sup>101</sup> 21124 Cabot Blvd., Hayward, CA 94545. Tel : 510.576-2334. [Jhefti@signaturebio.com](mailto:Jhefti@signaturebio.com)

l'industrialisation de ce domaine de recherche (analyse en parallèle de plusieurs centaines de protéines). Que cela soit dans les laboratoires académiques ou dans les sociétés privées, les moyens déployés sont apparus considérables et très inégaux par rapport à notre pays.

Le programme *Protein Structure Initiative* du NIH est beaucoup plus ambitieux que ce qui existe chez nous. Le financement du programme de Génomique structurale des Génopoles est très loin de celui d'un seul des centres américains. Une initiative similaire, la *Proteomics Initiative*, est en cours de création par le *National Heart, Lung and Blood Institute* (NHLBI) l'un des instituts du NIH. Cette initiative prévoit de financer 10 centres de recherche pilote pour développer des techniques innovantes d'analyse du protéome et de les appliquer à des problèmes biologiques d'importance pour comprendre les maladies cardiaques et pulmonaires. Ces centres seraient financés pendant 7 ans pour un total de 100 millions de dollars<sup>102</sup>

Les entreprises impliquées dans la protéomique font du développement technologique (*Zyomyx*, *SurroMed*). Certaines apportent des solutions originales, mais il est à craindre que les coûts des produits soient alors très élevés et ne soient pas disponibles pour la recherche académique (c'est déjà le cas pour les systèmes commercialisés de type ICAT). Les laboratoires académiques font aussi du développement technologique (John Yates). Chacun développe des outils bioinformatiques ... La génomique structurale à haut débit existe. *Syrrx* la pratique... Les start-ups prennent en compte la multidisciplinarité de manière plus importante qu'en Europe et en France (cf. l'automatisation chez *Syrrx* et la comparaison de ces entreprises avec des entreprises françaises ou européennes comme *BioXtal*). L'interface académique-start up est plus important aux Etats-Unis qu'en Europe, bien que la différence s'amenuise. L'université française a du mal à suivre cette évolution actuellement.

Deux technologies semblent avoir le vent en poupe pour découvrir les nouveaux biomarqueurs en clinique: la cytométrie de flux (pour l'identification des cellules impliquées) et la spectrométrie de masse (pour le petit protéome). Leur simplification et leur arrivée « sur la paillasse » est une source de résultats nouveaux et de grands progrès dans ces technologies vont se faire prochainement. Les puces protéiques ne devraient pas tarder à faire leur apparition vu les alliances entre les spécialistes des supports et les producteurs d'anticorps. Le transcriptome souffre de la non-disponibilité de matériel cellulaire pertinent chez les patients.

## 2 - La création d'un modèle économique

La vitalité entrepreneuriale des Etats-Unis a été évidente lors de cette mission. La prise de risque des financiers (mesurée par le montant levé par les entreprises) ne semble pas corrélée au contenu scientifique des projets présentés. Il semble que les banques d'investissement ne sont pas associées à des experts pour analyser les projets. Nous avons été très surpris des sommes considérables rassemblées pour créer des entreprises sur des bases scientifiques parfois très discutables (exemple *GeneFormatics*).

Les initiatives publiques conduites par le gouvernement américain (*Protein Structure Initiative*) et les autres pays (*Structural Genomics Consortium* en Angleterre, initiatives au Canada, au Japon et en Allemagne) ne semblent pas inquiéter les sociétés de biotechnologie qui misent sur la génomique structurale pour développer un *business model*. Ces sociétés pensent que la vitesse qui

---

<sup>102</sup> <http://www.nhlbi.nih.gov/meetings/proteomics.htm>

résulte de l'automatisation de chaque étape attirera la clientèle des groupes pharmaceutiques à la recherche de structures de protéines.

Les bases de données de motifs structuraux sont très certainement des valeurs ajoutées importantes. Mais, il n'est pas clair pourquoi les entreprises qui se construisent sur la valeur de ces bases de données cèdent en même temps des droits d'utilisation à d'autres sociétés, *Celera* par exemple, et perdent alors (nous semble-t-il) une partie de leur avantage initial (exemple *GeneFormatics*, *Structural GenomiX*).

Toutes les entreprises ont pour objectif le criblage de molécules sur cellules, parfois sur des protéines. C'est en particulier surprenant lorsque ce ne sont pas des compagnies pharmaceutiques (exemple *Signature BioScience*). Les cibles potentielles choisies sont alors très semblables, et ce quelque soit la compagnie.

### 3 - Le futur légal des données.

La disponibilité des données structurales du génome n'est pas encore résolue comme dans beaucoup de cas similaires où des bases de données sont en jeu. Si tout le monde s'accorde dans le milieu académique pour dire que les données devront être publiques, il y a une demande européenne et japonaise pour étendre cette durée ce qui permettrait de déposer de brevets sur des structures d'importance pharmaceutique<sup>103</sup>. Toutefois un consensus international a été atteint pour publiciser les protéines qui sont à l'étude et pour éviter les redondances. Dans le même temps, les PSI du NIH sont soumis à des conditions différentes, puisque le NIGMS pousse les centres à identifier de nouvelles protéines à un taux de 200 par an en 2006.

Le secteur académique a aussi un rôle majeur à jouer dans la protéomique appliquée aux maladies. Il est essentiel, pour des raisons éthiques et industrielles, que les biomarqueurs ne deviennent pas la propriété de quelques sociétés. Des dotations d'équipement adaptés devraient être coordonnées sur les sites de recherche clinique sous l'égide des spécialistes en protéomique et des études cliniques devraient être lancées sur ce domaine.

Enfin, les brevets issus de la protéomique peuvent certainement fragiliser les brevets issus de la génomique. En effet, un gène peut donner lieu à différentes protéines, et ces différentes protéines peuvent avoir des fonctions et des intérêts biomédicaux différents de ceux protégés par les brevets accordés au gène. Cela donnera lieu à toute une nouvelle série de brevets protégeant les découvertes issues de la protéomique. Cet aspect n'a pas échappé aux sociétés de génomique qui ont lancé leurs propres programmes de protéomique (Incyte, Celera, Human Genome Science)

### 4 – Conclusions :

En conclusion, cette mission fort instructive a permis plusieurs constatations, qui à coté des commentaires plus spécifiques indiqués plus haut peuvent se résumer ici. Les risques de recherche à deux vitesses sont énormes. Les organismes de recherche comme les universités n'auront jamais les moyens de développer les robots qu'utilise *Syrrx*. Si cela semble être le cas aux Etats-Unis, ce sera encore plus vrai chez nous ! Que faire ? Sous-traiter la production de cristaux à des compagnies ? Laisser à la recherche universitaire les structures les moins intéressantes en terme d'applications ? Faudra-t-il refaire dans les laboratoires académiques les structures déjà résolues par les entreprises privées ?

---

<sup>103</sup> Meeting de Airlie House, Warrenton, VA. 4-6 avril 2001

Des rencontres avec les français de San Diego et San Francisco (post-docs et entrepreneurs) ont eu lieu en marge de cette mission d'étude. Si ces rencontres ont révélé la vitalité de cette communauté, il est aussi apparu une véritable méconnaissance de l'évolution importante du système français et européen. Il faudrait pouvoir organiser des séjours de 1 semaine en France pour certains post-docs afin qu'ils se mettent au courant et ramènent l'information dans leur communauté. Beaucoup de chercheurs ont acquis des compétences qui seraient très intéressantes à valoriser, même si les chances d'un retour direct en France sont limitées du fait de raisons personnelles. On peut au moins envisager de mettre en place des collaborations avec des laboratoires français. L'idée serait de développer des collaborations avec des post-docs français ayant acquis une maturité scientifique dans le but de puiser dans ce vivier de compétences.