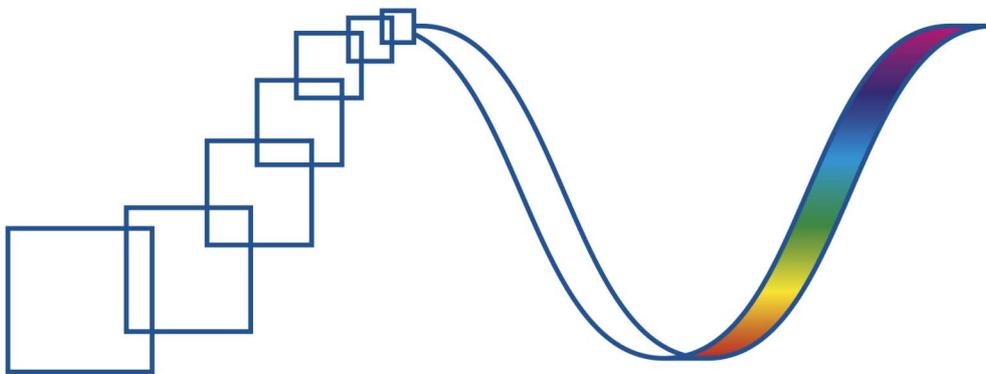


Dossier de presse



PROGRAMME DÉCRYPTHON

Conférence de presse

Paris, le 15 mars 2005

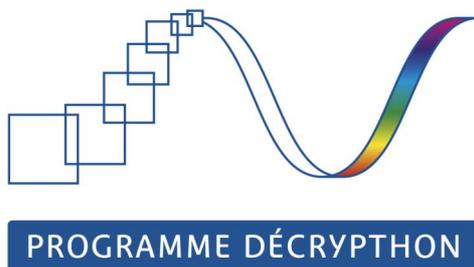
CNRS

3, rue Michel-Ange – 75016 Paris

Programme Décrypthon

Sommaire

Communiqué de presse	3
<hr/>	
Une puissance de calcul au service de la génomique et protéomique	5
Le concept et les objectifs	5
Rappel historique	5
La collaboration de trois partenaires fondateurs	6
<hr/>	
Une ambition scientifique	7
Les enjeux de la Bio-informatique	7
Deux projets pilotes déjà impliqués	7
Projet Poch et Deléage (Illkirch-Lyon)	
Projet Branlant, Leclerc et Guermeur (Nancy)	
Des recherches pluridisciplinaires	8
Les projets scientifiques futurs	8
Témoignage d'un chercheur impliqué dans le Programme Décrypthon	9
<hr/>	
Un enjeu technologique	10
Le Grid Computing, l'innovation technologique au service de la recherche	10
De la première opération Décrypthon : un projet à base de grille sur internet...	10
... au Programme Décrypthon : une étape supplémentaire avec la grille de serveurs	10
<hr/>	
Décrypthon Entreprise : à calcul partagé, financement partagé	12
Un financement innovant	12
Les besoins de financement	12
Les modalités de participation	12
<hr/>	
Les partenaires fondateurs	14
AFM – Objectif : guérison	15
IBM - Une expertise technologique appliquée aux sciences du vivant	17
CNRS - une expertise scientifique dans le domaine des sciences de la vie et des sciences et technologies de l'information et de la communication	19



Accélérer la recherche en génomique et protéomique grâce aux technologies de grid computing

Parce que l'innovation technologique, notamment informatique, peut répondre à de nouveaux besoins de la recherche en génétique, trois acteurs majeurs dans leur domaine, l'**AFM**, association de malades, le **CNRS**, organisme de recherche, et **IBM**, société d'informatique, lancent le **Programme Décrypthon, une plate-forme de Grid Computing**. Objectif : accélérer la recherche en génomique et en protéomique pour progresser plus rapidement dans la compréhension des maladies génétiques, notamment des maladies neuromusculaires.

Lancée lors du Téléthon 2001, une première opération Décrypthon avait permis la réalisation, en quelques semaines, d'une première base de données de toutes les protéines du monde vivant grâce à 75 000 ordinateurs individuels mis en réseau. Aujourd'hui, le Programme Décrypthon permet à la recherche d'aller encore plus loin et encore plus vite dans la compréhension du vivant.

La plate-forme de *Grid Computing*

Le Programme Décrypthon est une plate-forme technologique de pointe reposant à nouveau sur le Grid Computing. Cette plate-forme sera composée de deux grilles spécifiques pour générer la puissance de calcul nécessaire au traitement de projets de recherche très complexes. Le recours à l'une ou l'autre des grilles, voire à la combinaison des deux, s'effectuera en fonction des spécificités des programmes de recherche sélectionnés.

- **La grille dite « universitaire »** est constituée des supercalculateurs des centres universitaires de Bordeaux 1, Lille 1 (USTL) et Paris Pierre et Marie Curie (Paris 6 Jussieu), auxquels seront ajoutés des serveurs à base de technologies Power 4+ et Power 5 de dernière génération offerts par IBM. Un serveur central, fédérant l'ensemble de ces ressources, sera également offert par IBM à l'université d'Orsay (Paris Sud) qui en assurera l'exploitation. L'ensemble de ces ressources sera connecté par le réseau à haut débit RENATER (Réseau National de Télécommunications pour la Technologie, l'Enseignement et la Recherche) qui relie les établissements français d'enseignement supérieur et de recherche. La puissance de calcul initiale de cette grille universitaire sera de 298 Gflops¹ auxquels pourront s'ajouter, si nécessaire, une partie des 473 Gflops déjà présents dans les trois universités.
- **Une grille dite « d'internautes »** qui pourra être activée dans un deuxième temps en fonction des projets scientifiques à venir. Elle accueillera alors les ordinateurs individuels. (fonctionnant sous Windows dans un premier temps et plus tard sous MacOS ou Linux) d'internautes souhaitant participer au Programme Décrypthon en mettant à disposition des scientifiques des « cycles de calcul » non utilisés par leur machine. La puissance de cette grille d'internautes variera en fonction du nombre d'internautes mobilisés, sachant que chaque ordinateur a une puissance théorique moyenne d'environ 1 Gflops.

Le Programme Décrypthon utilisera la suite logicielle *Grid MP* de la société américaine United Devices, dans le cadre de la planification, l'exécution et le suivi des programmes informatiques de chaque projet.

¹ La puissance d'un ordinateur s'exprime en nombre d'opérations « flottantes » par seconde, c'est-à-dire en nombre d'opérations élémentaires (addition, soustraction...) sur des nombres décimaux à virgule flottante. L'abréviation est le « flops » pour *FLOating-Point operation Per Second*. Le Giga Flops (Gflops) correspond à 1 milliard de flops.

L'enjeu scientifique : la révolution protéomique

Le déchiffrement complet du génome humain, en 2003-2004, a montré que les liens biologiques qui lient les éventuelles modifications du génome et les maladies qui en découlent, sont plus complexes. L'approfondissement des connaissances dans le domaine de la protéomique constitue donc un axe prioritaire de la recherche en génétique à l'aube de ce troisième millénaire.

Pendant 3 ans, la plate-forme du Programme Décryphon accueillera une dizaine de projets de recherche ayant pour but, notamment, de mieux comprendre le fonctionnement et le rôle des protéines, de prédire leur fonction, de progresser dans la connaissance de leur structure en trois dimensions, ou de croiser les données du génome avec celles du protéome.

Le Programme Décryphon est déjà une réalité : **deux projets pilotes** ont été sélectionnés auprès de la communauté scientifique française et étrangère.

- « **les mutations structurales avec leurs conséquences sur le phénotype des pathologies humaines** ». Un projet coordonné par Olivier Poch (Illkirch, IGBMC / CNRS / Inserm) et Gilbert Deléage (Lyon, IBCP / CNRS).
L'objectif est double : mettre au point une grille d'analyse descriptive des protéines avec des mutations connues, et élaborer un outil prédictif pour faciliter la compréhension de ces mutations dans les maladies humaines. Ce projet s'inscrit dans le contexte fondamental de la compréhension des mécanismes qui contrôlent la fonction des protéines. La façon dont une protéine se replie dans l'espace détermine ses interactions avec l'environnement cellulaire, y compris la manière dont elle va interagir avec les molécules thérapeutiques. Son démarrage est prévu en mai 2005.
- « **Développement et utilisation d'approches informatiques pour l'analyse des liens existant entre les défauts d'épissage et les maladies génétiques** ». Un projet piloté par Christiane Branlant, Fabrice Leclerc (Nancy, Laboratoire de Maturation des ARN et Enzymologie Moléculaire / CNRS), et Yann Guermeur (Nancy, LORIA / CNRS / INP Lorraine / INRIA).
Il s'agit de l'analyse des liens existant entre les défauts d'épissage (processus de fabrication de la protéine à partir de l'ADN) et les maladies génétiques. En apportant des données fondamentales pour la compréhension des maladies génétiques, cette étude devrait permettre le développement de thérapies basées sur la réparation de l'épissage impliquant, par exemple, le saut d'exon. Son démarrage est prévu à l'automne 2005.

Un appel d'offres scientifique a été lancé fin 2004 afin de sélectionner de nouveaux projets.

Les entreprises sollicitées pour « prendre part »

Le Programme Décryphon s'appuie sur un financement innovant, reposant sur un principe clair : **à calcul informatique partagé, financement partagé**. L'AFM organise donc une opération de collecte spécifique baptisée « **Décryphon Entreprise** ». Toutes les entreprises sont ainsi invitées à souscrire au Programme Décryphon en achetant une ou plusieurs parts de ce projet :

1 part du Programme Décryphon = 1000 Euros.

Le coût prévisionnel du Programme Décryphon est estimé à 1,8 million d'euros par an sur 3 ans, dont 40 % sont assurés par des contributions en nature d'IBM, du CNRS et de l'AFM, et 60 % seront assurés grâce aux souscriptions des entreprises.

L'adhésion des entreprises à l'objectif et aux ambitions du Programme Décryphon leur permet de devenir actrices d'un programme scientifique innovant et, plus globalement, de s'inscrire dans une démarche de soutien à la recherche et à l'innovation en santé. En achetant des parts, elles deviennent des Entreprises Partenaires du Programme Décryphon.

Pour souscrire, s'informer et suivre le Programme Décryphon, rendez-vous sur www.decryphon.fr

Contacts Presse :

AFM : Estelle Assaf, Delphine Carvalho - 01 69 47 28 28

CNRS : Martine Hasler - 01 44 96 46 35

IBM : Véronique Barone - 01 49 05 73 35

Le concept et les objectifs

Le Programme Décryphon consiste à mettre en place une plate-forme technologique de pointe reposant sur le Grid Computing pour apporter à des équipes de chercheurs des moyens de calcul très importants leur permettant d'accélérer leurs travaux.

Son objectif est clair : accélérer la recherche en génomique et en protéomique pour progresser plus rapidement dans la compréhension des maladies génétiques, notamment des maladies neuromusculaires.

Pendant 3 ans, la plateforme de Grid computing du Programme Décryphon accueillera une dizaine de projets de recherche dans le but, notamment, de mieux comprendre le fonctionnement et le rôle des protéines, de prédire leur fonction, de progresser dans la connaissance de leur structure en trois dimensions, de croiser les données du génome avec celles du protéome.

Aujourd'hui, le Programme Décryphon est déjà une réalité : deux projets pilotes ont été sélectionnés avec le CNRS auprès de la communauté scientifique française et étrangère :

- Une équipe dirigée par Olivier Poch et Gilbert Deléage (Illkirch : CNRS/IGBMC Institut de Génétique et de Biologie Moléculaire et Cellulaire - Lyon : CNRS/IBCP Institut de Biologie et de Chimie des Protéines) pour la mise au point d'une grille d'analyse descriptive des protéines avec des mutations connues et d'un outil prédictif pour faciliter l'analyse des conséquences de ces mutations dans les maladies humaines. Le démarrage est prévu en mai 2005.
- Une équipe dirigée par Christiane Branlant, Fabrice Leclerc et Yann Guermeur (Nancy : Laboratoire de maturation des ARNs et enzymologie moléculaire et LORIA Laboratoire lorrain de Recherche en Informatique et ses Applications – CNRS/INRIA) pour l'analyse des liens existant entre les défauts d'épissage (processus de fabrication de la protéine à partir de l'ADN) et les maladies génétiques. Le démarrage est prévu à l'automne 2005.

Afin de sélectionner de nouveaux projets qui pourraient bénéficier de la plate-forme Décryphon, un nouvel appel d'offres scientifique a été lancé fin 2004 auprès de la communauté scientifique française et étrangère.

Rappel historique

Après avoir réalisé, entre 1992 et 1996, les premières cartes du génome humain dans son laboratoire Généthon, l'AFM se lance en 2001 dans la cartographie du protéome. C'est ainsi qu'elle s'associe à IBM pour l'opération Décryphon lors du Téléthon 2001. Ce véritable défi scientifique, technologique et humain est alors relevé avec brio : 75 000 internautes mobilisés, des milliards de calculs complexes effectués en moins de 2 mois, 550 000 protéines du monde vivant (animal, végétal, humain) cartographiées.

En effet, ce premier projet Décryphon s'est appuyé sur la mobilisation de plus de 75 000 internautes qui ont accepté de « mettre à disposition » la puissance inutilisée de leur PC pour effectuer les milliards de calculs complexes nécessaires. Grâce à cette mobilisation unique en France, ces calculs ont pu être effectués en moins de deux mois, alors qu'il aurait fallu plus de 1170 années pour les réaliser à l'aide d'un seul ordinateur ! Chaque ordinateur a contribué à hauteur d'environ 133 heures, soit plus de 10 millions d'heures de calcul au total. 21 serveurs IBM ont hébergé l'ensemble des solutions et des données pendant toute la durée de l'opération.

La cartographie du protéome ainsi générée par ce projet a été mise à la disposition de la communauté scientifique (www.infobiogen.fr). Il s'agit d'une véritable bibliothèque de comparaison des protéines des différentes espèces du monde vivant (animal, végétal, humain). Elle contient près de 2,2 millions de fichiers répartis en 17 000 répertoires, soit l'équivalent d'une bibliothèque de plus de 25 000 volumes.

Aujourd'hui, la recherche en sciences de la vie a de plus en plus besoin de puissances de calcul pour avancer : **chaque année, le volume de données en biologie est multiplié par deux**. C'est ainsi qu'est née la volonté de poursuivre le recours aux technologies de grille (ou Grid Computing) dans le cadre de programmes de recherche en protéomique ou en génomique.

La collaboration de trois partenaires fondateurs

Le Programme Décryphon repose sur un partenariat original. Il rassemble une association de malades, l'AFM, un organisme de recherche, le CNRS et une société, IBM : trois entités aux missions, domaines de compétences et stratégies distinctes, mais qui se rejoignent dans la recherche de l'innovation et du progrès.

L'AFM coordonne l'appel à projets auprès de la communauté scientifique et contribue au financement des projets sous plusieurs formes : financement direct aux équipes, financement d'éventuels services associés. Le Programme Décryphon s'inscrit dans la stratégie scientifique de l'AFM, organisatrice du Téléthon chaque année depuis 1987 et acteur majeur de la recherche. Cette stratégie vise un objectif : guérir les maladies neuromusculaires, des maladies rares pour la plupart d'origine génétique.

IBM apporte son expertise technologique dans les technologies de grille (ou Grid Computing). Dans le cadre du programme IBM Shared University Research, IBM France dote 4 universités de supercalculateurs de dernière génération. (cf la partie Un enjeu technologique)

Le CNRS assure le pilotage du programme scientifique. Il apporte également une expertise scientifique et technologique sur le portage des projets sur la grille et sur la mise en place des moyens de calcul.

En complément de ces trois partenaires fondateurs, le programme s'appuie sur la participation de 4 universités (Bordeaux, Lille, Paris Pierre et Marie Curie et Orsay) où des supercalculateurs ont été installés par IBM, ainsi que sur le réseau à très haut débit RENATER (Réseau National de Télécommunications pour l'Enseignement et la Recherche), sur lequel sera connecté l'ensemble de ces ressources. Enfin, le Programme Décryphon exploitera la suite logicielle *Grid MP* de la société américaine United Devices.

Les enjeux de la Bio-informatique

Établir la carte du génome humain, mais aussi celle des génomes de diverses espèces animales et végétales a constitué un des enjeux majeurs de la recherche en génétique des deux dernières décennies. Cet objectif, partagé par l'ensemble de la communauté scientifique, allait conduire au cours des années 1992 à 1996 à la réalisation des premières cartes du génome humain au sein du laboratoire Généthon de l'AFM.

Le déchiffrement complet du génome humain, en 2003-2004, a montré que les liens biologiques qui lient les éventuelles modifications du génome et les maladies qui en découlent, sont plus complexes. La connaissance des mécanismes d'expression ou de non expression des gènes et la caractérisation des produits qui en découlent, c'est à dire les protéines, se révèle donc d'un intérêt majeur. Il s'agit à la fois de comprendre les mécanismes fondamentaux du fonctionnement cellulaire, mais aussi à terme, de développer des méthodes qui pourraient être utilisées dans le domaine de la clinique humaine. Ainsi, l'élaboration de cartographies du protéome constitue un axe prioritaire de la recherche en génétique à l'aube de ce troisième millénaire.

Il s'agit de mettre en œuvre de nouvelles méthodes permettant de croiser les informations en provenance à la fois de la génomique et de la protéomique. Le traitement de flux d'informations impose d'innover dans le domaine de la bio-informatique et de développer de nouveaux modes de calcul permettant la prise en compte simultanée de nombreux paramètres. Le premier projet Décryphon lancé en 2001 en a montré l'intérêt.

La plate-forme mise en place dans le cadre du Programme Décryphon aujourd'hui, répond à cet enjeu encore plus pressant depuis quelques années, avec l'explosion des connaissances dans les domaines des sciences de la vie.

Deux projets pilotes déjà impliqués

« les mutations structurales avec leurs conséquences sur le phénotype des pathologies humaines »

Un projet de Olivier Poch et Gilbert Deléage à Illkirch (Institut de Génétique et de Biologie Moléculaire et Cellulaire – CNRS / Université Louis Pasteur, Strasbourg 1 / INSERM) et Lyon (Institut de Biologie et Chimie des Protéines – CNRS / Université de Lyon 1)

Le but est double :

- établir une grille d'analyse descriptive des protéines avec des mutations connues
- fournir un outil prédictif pour faciliter l'analyse des conséquences de ces mutations dans les maladies humaines.

Ce projet s'inscrit dans le contexte fondamental de la compréhension des mécanismes qui contrôlent la fonction des protéines cellulaires.

La structure tridimensionnelle d'une protéine est unique et spécifique. La connaissance de cette structure 3D est une étape indispensable et essentielle pour déterminer précisément la fonction de la protéine en cause et son mécanisme d'action. La façon dont une protéine se replie dans l'espace détermine ses interactions avec l'environnement cellulaire, y compris la manière dont elle va interagir avec les molécules thérapeutiques.

Plus précisément, pour atteindre cet objectif, les informations évolutives et structurales d'une protéine mutée seront analysées et intégrées dans des modèles pour établir et quantifier les facteurs d'impact structural des mutations afin de les comparer aux phénotypes observés dans des pathologies humaines bien connues, telles que les myopathies ou certaines maladies du développement. Lorsque ces données seront bien établies, les approches et les outils développés dans le cadre de ce projet seront mis à disposition de l'ensemble de la communauté scientifique afin de permettre la prédiction des impacts structuraux d'un grand nombre de mutations et de maladies humaines. Cette démarche de modélisation et de simulation sera mise en œuvre au travers d'un portail web faisant appel aux plus récentes

technologies informatiques de calcul massif permettant d'assurer l'analyse et la prédiction des impacts structuraux, l'intégration rapide et fiable de grandes quantités d'informations hétérogènes et l'accès aux résultats générés.

« Développement et utilisation d'approches informatiques pour l'analyse des liens existant entre les défauts d'épissage et les maladies génétiques »

Un projet de Christiane Branlant, Fabrice Leclerc et Yann Guermeur à Nancy (Laboratoire de Maturation des ARN et Enzymologie Moléculaire - CNRS / Université de Nancy 1 et Laboratoire Lorrain de Recherche en Informatique et ses applications - CNRS / Universités de Nancy 1 et 2 / INP Lorraine / INRIA).

Pour synthétiser une protéine, le gène (présent dans la molécule d'ADN) délivre dans la cellule un code de fabrication dans le cadre d'un processus qui a été baptisé : « épissage ». Au cours de cette étape, la cellule assemble bout à bout les éléments composant le code originel qui sont appelés exons. Le même gène peut donner lieu à plusieurs codes de fabrication différents, par suite d'un épissage dit « alternatif ». C'est ce processus qui permet à un seul gène de produire plusieurs protéines à la fois. L'analyse des mutations dans les gènes chez des personnes atteintes de maladies génétiques et les conséquences de l'épissage devraient apporter des données fondamentales pour la compréhension des maladies génétiques. Cette étude devrait permettre le développement de thérapies basées sur la réparation de l'épissage, impliquant, par exemple, le saut d'exon.

Des recherches pluridisciplinaires

Le Programme Décryphon se traduit également par le développement de recherches pluridisciplinaires qui impliquent le département des Sciences de la vie (SDV) et le département des Sciences et technologies de l'information et de la communication (STIC) du CNRS. Ces recherches concernent le traitement des séquences, des structures et l'algorithmique associée, la construction de phylogénies et la modélisation de phénomènes biologiques. Elles sont intégrées dans l'une des priorités du CNRS intitulée « modélisation du vivant ». 9 postes de chercheurs ont été mis à l'affichage pour le concours 2005 dans la thématique « modélisation des systèmes biologiques ». Le département STIC a contribué à la mise en place du Programme Décryphon depuis sa création, à la fois dans le domaine de la bio-informatique et dans le domaine du calcul intensif et des grilles. Il apporte de plus un soutien financier important à ce programme sous la forme de bourses de thèse et de post doctorants.

Les projets scientifiques futurs

Un appel d'offres annuel a été mis en place pour sélectionner 2 à 3 nouveaux projets scientifiques chaque année sur une durée de trois ans. Pour l'année 2006, l'AFM a lancé un appel d'offres fin novembre 2004. Les projets soumis dans ce cadre seront expertisés et évalués par le comité scientifique du Programme Décryphon, dans le courant du mois d'avril 2005.

Témoignage d'un chercheur impliqué dans le programme Décryphon :

" Un outil pour aller plus loin, plus vite "

Olivier POCH,

Biologiste, directeur de recherche au sein de l'Institut de Génétique et de Biologie Moléculaire et Cellulaire de Strasbourg (CNRS / Université Louis Pasteur, Strasbourg 1 / Inserm)

" Le projet scientifique que je coordonne avec Gilbert Deléage à Lyon repose sur le rapport entre les protéines, les mutations génétiques et les maladies humaines. Nous étudions plus précisément la structure en trois dimensions des protéines. Il s'agit de manipuler des données très complexes, regroupant plusieurs paramètres. La plate-forme de *grid computing* qui est mise en place dans le cadre du Programme Décryphon répond donc tout à fait à nos besoins : la mise en grille de supercalculateurs nous apporte la puissance de calcul nécessaire à ces analyses complexes, et nous bénéficions d'une synergie précieuse entre les mondes de la biologie, de la médecine et de l'informatique. La recherche en biologie est aujourd'hui submergée par des milliers de données intéressantes. Nous, chercheurs, avons besoin de nouveaux outils, comme l'informatique, pour que tout cela acquière un sens. Le Programme Décryphon nous permettra d'aller plus loin, plus vite. Et nous savons qu'au bout, il y a des malades qui attendent. »

Le Grid Computing, l'innovation technologique au service de la recherche

Les technologies dites de « grille », ou *Grid Computing*, permettent de mettre en partage de façon sécurisée les données et les programmes de multiples ordinateurs, qu'ils soient de bureau, personnels ou supercalculateurs. Ces ressources sont mises en réseau et partagées grâce à des solutions logicielles dédiées. Elles peuvent ainsi générer, à un instant donné, un système virtuel doté d'une puissance gigantesque de calcul et une capacité de stockage en rapport pour mener à bien des projets scientifiques ou techniques requérant une grande quantité de cycles de traitement ou l'accès à de gros volumes de données.

Aujourd'hui, la recherche, en particulier dans le domaine médical, a de plus en plus recours aux technologies de Grid Computing car les besoins en calcul et en stockage sont considérables. Tous les ans, le volume de données en biologie est en effet multiplié par deux. Le Grid offre la possibilité d'accélérer la découverte de nouveaux traitements. Il permet de réaliser des prévisions plus précises et de réduire le temps de recherche et de développement de médicaments.

A titre d'exemple, le projet *Smallpox Research Grid*, parrainé par l'université d'Oxford, IBM et la société United Devices, a permis aux scientifiques d'analyser des millions de molécules afin d'identifier lesquelles pourraient entrer dans la fabrication de médicaments contre la variole. En moins de six mois, 45 traitements antivarioliques potentiels ont été décelés.

De la première opération Décrypthon : un projet à base de grille sur internet...

En 2001, l'AFM s'associe à IBM, précurseur dans la technologie de *Grid Computing*, pour la première opération Décrypthon. L'objectif : réaliser la cartographie du protéome en faisant appel aux technologies de grille sur internet. Le projet a ainsi permis la comparaison exhaustive de 500 000 protéines du monde vivant.

Grâce à la mobilisation de plus de 75 000 internautes qui ont accepté de mettre à disposition la puissance inutilisée de leurs PC, des milliards de calculs complexes ont pu être effectués en moins de deux mois, alors qu'il aurait fallu plus de 1170 années pour les réaliser à l'aide d'un seul ordinateur de l'époque !

... au Programme Décrypthon : une étape supplémentaire avec la grille de serveurs

L'AFM et IBM sont de nouveau partenaires, avec le CNRS, dans un projet plus large de Grid Computing, le Programme Décrypthon.

La grille développée pour le Programme Décrypthon permettra de mettre en commun les puissances de calculs ou les capacités de stockage disponibles auprès de supercalculateurs ou de simples ordinateurs individuels.

La plate-forme de calcul du Programme Décrypthon sera dotée de deux grilles spécifiques :

- Une grille dite « universitaire » constituée des supercalculateurs des centres universitaires de Bordeaux 1, Lille 1 (USTL) et Pierre et Marie Curie (Paris 6 Jussieu), auxquels seront ajoutés des serveurs à base de technologies Power 4+ et Power 5 de dernière génération offerts par IBM. Un serveur central, fédérant l'ensemble de ces ressources, sera également offert par IBM à l'université d'Orsay (Paris Sud) qui en assurera l'exploitation. L'ensemble de ces ressources sera connecté par le réseau à haut débit RENATER (Réseau National de Télécommunications pour la Technologie, l'Enseignement et la Recherche) qui relie les établissements français d'enseignement supérieur et de recherche.

La puissance de calcul initiale de cette grille universitaire sera de 298 Gflops² auxquels pourront s'ajouter, si nécessaire, une partie des 473 Gflops déjà présents dans les trois universités.

- Une grille dite « d'internautes » qui pourra être activée dans un deuxième temps en fonction des projets scientifiques à venir. Elle accueillera alors les ordinateurs individuels (fonctionnant sous Windows dans un premier temps et plus tard sous MacOS ou Linux) d'internautes souhaitant participer au Programme Décryphon en mettant à disposition des scientifiques des « cycles de calcul » non utilisés par leurs machines.

La puissance de cette grille d'internautes variera en fonction du nombre d'internautes mobilisés, sachant que chaque ordinateur a une puissance théorique moyenne d'environ 1 Gflops.

Le Programme Décryphon utilisera pour les deux grilles (universitaire et d'internautes) la suite logicielle *Grid MP* de la société américaine United Devices, partenaire d'IBM et leader mondial pour les solutions de grille de calcul. *Grid MP* pourra piloter l'intégralité des ressources des grilles en optimisant la planification, l'exécution et le suivi des programmes informatiques de chaque projet scientifique utilisant la puissance de la grille. Cette solution permettra aussi de prendre en charge l'intégration de nouvelles ressources (supercalculateurs, stockage, ordinateurs individuels, etc.) fournies par d'autres universités, centres de calculs nationaux ou communautés d'internautes.

Les particularités des deux types de grilles :

- La grille universitaire permet de résoudre des calculs en parallèle sur de longues durées, ces calculs pouvant être fortement couplés entre eux, c'est-à-dire ayant la capacité de partager et d'échanger des données communes de taille importante à une fréquence élevée. En effet, cette grille est constituée d'un réseau à haut débit, disponible 24h/24, et dotée de machines puissantes (plusieurs Gigas octets de mémoire vive par poste).

- La grille d'internautes est constituée d'ordinateurs « volatiles » - chaque internaute se connecte et se déconnecte d'Internet à son rythme, allume et éteint son ordinateur à sa convenance - connectés par des liaisons à plus faible débit. Les calculs soumis à cette grille requièrent donc d'être faiblement couplés, sur une durée de traitement limitée et nécessitant l'accès à peu de données. En revanche, le grand nombre d'ordinateurs disponibles permet de distribuer massivement ces traitements.

En fonction des spécificités des programmes de recherche sélectionnés, on aura recours à l'une ou l'autre des grilles ou aux deux. En effet, tous les projets scientifiques ne permettent pas, pour leurs calculs, d'être découpés en petits morceaux. À titre d'exemple, pour les deux projets pilotes menés en 2005, seule la grille universitaire sera utilisée. D'autres projets sélectionnés à l'avenir pourraient, en revanche, requérir un couplage des deux grilles.

² La puissance d'un ordinateur s'exprime en nombre d'opérations « flottantes » par seconde, c'est-à-dire en nombre d'opérations élémentaires (addition, soustraction...) sur des nombres décimaux à virgule flottante. L'abréviation est le « flops » pour FLOating-Point operation Per Second. Le Giga Flops (Gflops) correspond à 1 milliard de flops.

Un financement innovant

Pour remplir ses objectifs pendant les trois prochaines années, le Programme Décryphon s'appuie sur un financement innovant. L'AFM organise une opération de collecte spécifique : **Décryphon Entreprise**. Autour d'une communauté de valeurs qui lie les trois partenaires fondateurs, les petites, moyennes et grandes entreprises sont invitées à relever le défi et à participer à ce grand programme scientifique et technologique au service du combat contre la maladie.

Le Programme Décryphon reposant sur le **calcul informatique partagé**, Décryphon Entreprise repose sur l'idée d'un **financement partagé** : les entreprises sont invitées à souscrire au Programme Décryphon en achetant des parts de ce projet.

Au-delà de cette souscription financière, leur adhésion à l'objectif et aux ambitions du Programme Décryphon leur permet de devenir actrices d'un programme scientifique innovant, et plus globalement, de s'inscrire dans une démarche de soutien à la recherche et à l'innovation en santé. Les entreprises partenaires seront d'ailleurs tenues régulièrement informées du déroulement du programme.

Les besoins de financements

L'AFM se comporte comme un investisseur intransigeant pour lequel les dividendes s'apprécient en progrès scientifiques et médicaux concrets. Les dons sont efficacement utilisés et les partenaires s'engagent à rendre compte fidèlement de leur utilisation.

Sur une base d'une dizaine de projets accueillis sur trois ans, le coût prévisionnel du Programme Décryphon est estimé à 1,8 million d'euros par an sur 3 ans :

- 40 % sont assurés par des contributions en nature d'IBM, du CNRS et de l'AFM à travers l'apport de serveurs, d'expertise et de compétences
- 60 % seront apportés par les souscriptions des entreprises : 1/3 serviront à soutenir directement les travaux des chercheurs et 2/3 les coûts supplémentaires de pilotage, mise en œuvre, sécurisation et suivi de la plate-forme technologique ; ainsi que les coûts pour faciliter l'accès aux résultats (portail...).

Les modalités de participation

Pour participer au Programme Décryphon, les entreprises achètent une ou plusieurs parts du projet. Elles deviennent ainsi des Entreprises Partenaires du Programme Décryphon.

1 part du Programme Décryphon = 1000 Euros

Pour devenir Entreprise Partenaire, les entreprises devront compléter le « Bon de souscription Entreprise Partenaire » qu'elles trouveront sur le site internet du Programme Decryphon, www.decrython.fr, sur lequel elles trouveront également toutes les informations scientifiques et technologiques relatives au Programme.

Des mesures fiscales avantageuses

L'ensemble des versements au titre du mécénat permet aux Entreprises Partenaires de bénéficier d'une réduction d'impôt sur le revenu ou les sociétés, de 60% du montant de la souscription pris dans la limite de cinq pour mille du chiffre d'affaire hors taxe. Au-delà de 5 pour mille ou en cas d'exercice déficitaire, un report est possible sur les 5 exercices suivants.

Par exemple : pour une contribution à hauteur de 10 000 Euros au Programme Décryphon, la réduction d'impôt est de 6 000 Euros, soit une dépense réelle de 4 000 Euros seulement.

Les partenaires fondateurs



Objectif : guérison

Une association de malades et de parents de malades

Créée en 1958 par des malades et leur famille, l'AFM se bat pour guérir les maladies neuromusculaires et réduire le handicap qu'elles engendrent. Ces maladies, évolutives et lourdement invalidantes, pour la plupart d'origine génétique et toutes rares, ont longtemps été oubliées de la science et de la médecine et non prises en compte par le système de santé.

L'AFM s'est donc fixé deux missions :

- guérir les maladies neuromusculaires
- aider les malades au quotidien.

En 1976, elle est reconnue d'utilité publique. En 1987, pour se donner les moyens d'atteindre son objectif, elle crée le Téléthon qui, chaque année, fait appel à la générosité du public durant le premier week-end de décembre.

Une association moteur et accélérateur de la recherche

Depuis le 1^{er} Téléthon, l'AFM a choisi d'intervenir au-delà du champ des maladies neuromusculaires. Parce que celles-ci sont des maladies minoritaires, l'AFM mène une politique d'intérêt général pour une plus grande efficacité. Elle intervient donc sur des problématiques plus globales :

- *Parce que les maladies neuromusculaires touchent le muscle et le neurone, l'AFM s'attache à faire exister la myologie, science et médecine du muscle.*
- *Parce que les maladies neuromusculaires sont d'origine génétique, l'AFM développe des outils d'intérêt général, des moyens et des savoirs pour accélérer la compréhension des maladies génétiques et ouvrir le chemin aux thérapies issues de la connaissance des gènes.*
- *Parce que les maladies neuromusculaires font partie des maladies rares, négligées par le système de santé publique, l'AFM a développé une stratégie globale pour la reconnaissance de ces malades exclus de la vie économique.*

Au-delà des quelque 400 programmes de recherches financés chaque année, l'AFM a également choisi de se doter et de développer des structures de recherche directement dédiées à ses axes de progrès.

- **Généthon** : *Fer de lance des thérapeutiques innovantes.*

Créé et financé par l'AFM depuis 1991, Généthon est un laboratoire unique au monde, un accélérateur de la recherche génétique capable de se réorganiser pour s'adapter à chaque nouvelle étape sur la voie de la guérison. Après les cartes du génome, après avoir fourni des vecteurs de thérapie génique à la communauté scientifique, Généthon explore les nouvelles pistes thérapeutiques (thérapie cellulaire et génique). Il prépare notamment plusieurs essais thérapeutiques de phase I, sur l'Homme, qui concerneront des maladies neuromusculaires et une maladie du système immunitaire.

Parmi ces essais :

La piste du saut d'exon, une technique innovante de thérapie génique

Fin 2004, une équipe de Généthon obtenait des résultats prometteurs en utilisant une technique innovante de thérapie génique, le saut d'exon. Elle a en effet réussi à réparer les muscles de souris atteintes de myopathie de Duchenne. Un essai sur l'Homme est prévu pour 2007 et un important programme d'identification des maladies candidates à la technique du saut d'exon a démarré.

- *l'Institut de Myologie : Fer de lance de la science et de la médecine du muscle.*

Créé par l'AFM en 1997, l'Institut de Myologie est implanté au sein du groupe hospitalier de la Pitié-Salpêtrière à Paris, en collaboration avec l'AP-HP et l'Inserm. Ce centre international d'expertise se consacre à toutes les questions relatives au muscle : ses pathologies, ses accidents, son vieillissement. Trois grandes missions structurent son organisation :

- La recherche fondamentale et clinique (quatre laboratoires et seize études cliniques en cours ou prévues),
- La prise en charge médicale (une consultation pluridisciplinaire sur les maladies neuromusculaires)
- L'enseignement de la myologie (trois cycles de formation et un centre de documentation à la disposition des chercheurs, des médecins, des étudiants mais aussi des malades)

Avec sa ténacité et la confiance du public renouvelée chaque année lors du Téléthon, l'AFM a déjà accompli ce qui paraissait jusqu'ici impossible en avançant à pas de géant vers la guérison. Avec son laboratoire Généthon, elle a publié les premières cartes du génome humain et a soutenu la découverte des gènes responsables de plus de 750 maladies. Elle a développé de nouveaux outils et démontré la pertinence des nouvelles thérapies issues de la connaissance des gènes, avec des premières percées encourageantes. Aujourd'hui, grâce à l'action majeure de l'AFM, les thérapies génique et cellulaire sont entrées dans l'ère des essais sur l'Homme.

Le Programme Décryphon

C'est en toute logique que l'AFM a porté la première opération Décryphon, lancée lors du Téléthon 2001. Les résultats ont illustré l'intérêt du grid computing dans l'accélération de l'acquisition d'informations et de connaissances, nécessaires aux progrès de la recherche et de la science.

Aujourd'hui, les moyens mis en œuvre grâce à la plate-forme de calcul du Programme Décryphon permettent à la recherche d'aller encore plus loin et encore plus vite dans la compréhension du vivant. Parce qu'au quotidien, le temps presse toujours pour les malades, le Programme Décryphon vient étoffer les moyens que se donne l'AFM.

Une expertise dans le domaine des Sciences de la vie et des Sciences et technologies de l'information et de la communication

Riche de ses huit départements scientifiques et de ses deux instituts, le CNRS est le plus important des établissements publics de recherche sur le territoire national. 11 600 chercheurs permanents, 14 400 ingénieurs, techniciens et administratifs participent conjointement à la recherche fondamentale dans l'ensemble des disciplines scientifiques.

Le département des Sciences de la Vie avec 290 laboratoires et 5900 agents (3150 chercheurs et 2750 ingénieurs, techniciens et administratifs) est situé à la convergence de toutes les disciplines. La recherche dans le domaine du vivant, outre son intérêt majeur dans la progression des connaissances, recouvre d'énormes enjeux sociaux et économiques. La compréhension des mécanismes impliqués dans l'expression des grandes pathologies au niveau moléculaire, cellulaire et de l'organisme entier, l'évaluation de l'influence des facteurs génétiques et de ceux liés à l'environnement sont indispensables pour développer, à terme, des thérapies efficaces. L'étude de l'évolution des génomes, chez les plantes, les animaux et chez l'homme, ses conséquences sur le maintien de la biodiversité, l'étude des interactions des gènes avec l'environnement qui influencent directement le développement de certaines maladies, sont, parmi tant d'autres secteurs explorés, des axes majeurs de recherche du Département des Sciences de la Vie. La structure même du CNRS avec ses possibilités d'approches pluri et interdisciplinaires, confère au Département des Sciences de la Vie une capacité de décrypter le vivant dans une approche globale : une situation qui demeure unique au niveau international. Les possibilités d'approches aux interfaces sont nombreuses impliquant tour à tour ou en synergie des collaborations étroites des laboratoires du département des Sciences de la vie avec ceux du domaine de la Chimie, des Sciences de l'Univers, des Sciences Physiques et Mathématiques, des Sciences Physiques pour l'Ingénieur et des Sciences de l'homme et de la Société, ou encore comme c'est le cas dans le cadre du Programme Décryphon avec le Département des Sciences et des Technologies de l'Information et de la Communication.

Le département des Sciences et Technologies de l'Information et de la Communication met son expertise dans le domaine des systèmes intelligents au service de l'homme et de la société. Son impact est majeur pour la progression des systèmes d'information et de communication, dans les domaines de l'assistance, mais aussi dans ceux liés à la connaissance : apprentissage, didactique... Son influence s'exerce également dans le domaine de la culture en générant de nouvelles voies pour le stockage, la comparaison et le croisement de diverses données relatives à un même objet de recherche. Si la robotique constitue un des domaines majeurs de l'activité du Département des Sciences et Technologies de l'Information et de la Communication, la mise au point de nouvelles méthodes permettant le traitement d'un grand nombre de données, recueillies dans des domaines scientifiques variés : physique statistique, communications téléphoniques, transmission d'informations par satellite, robotique à distance, gestion de données financières, maintenance industrielle... constitue également une des voies privilégiées d'investigation. La production d'outils efficaces pour appréhender le fonctionnement de systèmes complexes est à la base des activités de ce département.

Décrypter les génomes, établir les cartographies du protéome impliquent, comme nous l'avons vu précédemment, une étroite collaboration entre biologistes et informaticiens. Le Programme Décryphon est sans aucun doute, à la fois dans ses objectifs fondamentaux et appliqués, mais aussi dans sa démarche de mise en œuvre, un exemple clé montrant toute l'utilité et l'efficacité d'une collaboration

étroite entre chercheurs, biologistes et bioinformaticiens (CNRS), spécialistes de l'ingénierie informatique et bio-informatique (IBM), et association de malades (AFM).



Une expertise technologique appliquée aux Sciences du vivant

IBM investit dans une structure dédiée aux Sciences du vivant

La course internationale au décryptage du génome humain à l'aube du 3ème millénaire a fait découvrir au grand public les enjeux scientifiques mais aussi technologiques et sociétaux qui se profilent pour les années à venir autour des Sciences du vivant.

Dans le même moment, la société IBM décidait d'investir massivement dans les bio-technologies en créant, en août 2000, une division des Sciences du vivant. C'est pour mieux répondre aux besoins et attentes de ses clients qu'IBM s'est ainsi dotée d'équipes pluridisciplinaires, alliant une double expérience et compétence en informatique, mais aussi dans de nombreux domaines connexes, comme la biologie moléculaire, la génétique, la biochimie ou la recherche clinique et la médecine.

La division Sciences du vivant d'IBM soutient des actions innovantes, favorise des rapprochements et partenariats entre associations, biotechs³, industriels et monde académique. Elle travaille en outre, en collaboration avec les 8 Centres de recherche appliquée d'IBM dans le monde, qui emploient plus de 3000 chercheurs, dont 5 Prix Nobel. Ces 5 dernières années, ce sont près de 100 chercheurs spécialisés en Sciences du vivant qui ont rejoint ces Centres de recherche IBM.

IBM, accélérateur de la recherche en Sciences du vivant

Une des premières contributions d'IBM à l'enjeu de la bio-informatique a été de permettre aux chercheurs de s'affranchir des frontières géographiques et culturelles et de gagner du temps dans la manipulation de leurs données génétiques.

Le lancement, en 2001, de la plate-forme d'intégration **IBM DiscoveryLink** permet chaque jour, à des scientifiques du monde entier de relier leurs applications à des bases de données hétérogènes. Ainsi, par le biais d'Internet, un chercheur peut accéder en parallèle à la Banque de données des protéines KEGG, sur le site de Kyoto au Japon, croiser ces informations avec les données de GenBank mises à jour quasiment tous les jours aux USA, et les relier aux publications scientifiques disponibles en s'abonnant aux revues de référence en ligne, etc.

Le décryptage des séquences des génomes, qui nécessite des capacités de stockage et de calcul considérables, a également recours à la bio-informatique. En 2001, s'appuyant sur les technologies IBM du *Grid Computing*, le Décryphon a permis à la génétique de réaliser un énorme bond en avant. La cartographie du protéome ainsi générée constitue une véritable bibliothèque de comparaison des protéines des différentes espèces du monde vivant (animal, végétal, humain).

Mais le rôle de la bio-informatique ne se limite pas au séquençage. La prédiction de la structure tridimensionnelle des protéines constitue également un enjeu capital. La façon dont une protéine se replie dans l'espace détermine, en effet, ses interactions avec l'environnement cellulaire y compris la manière dont elle va interagir avec les molécules thérapeutiques. IBM a ainsi lancé depuis 5 ans, un programme de recherche de 100 millions de dollars, pour bâtir un super-ordinateur, **Blue Gene**, capable d'effectuer

³ Les biotechs sont des entreprises spécialisées dans la recherche et le développement en biotechnologies

plus d'un million de milliards d'opérations par seconde et destiné à simuler le repliement des protéines en 3D.

De la bio-informatique à l'e-Santé

Déjà, les premières applications de la bio-informatique commencent à sortir des Centres de recherche, publics ou privés. Des médicaments sont pré-testés *in silico*, par modélisation informatique, pour accélérer la mise sur le marché de nouvelles thérapeutiques. Ces traitements seront désormais ciblés, en fonction de l'analyse de profils génétiques et permettront donc de délivrer une médecine personnalisée.

Les professionnels de santé pourront aussi bénéficier demain des apports de la bio-informatique pour améliorer la prise en charge des patients, en détectant à quel traitement ils seront répondeurs. Cette approche va aussi améliorer la prévention des maladies, permettre leur suivi au long cours, à travers des hôpitaux virtuels, limitant les durées d'hospitalisation par un suivi à domicile, s'appuyant sur des Dossiers Médicaux Personnels informatisés.

Pour répondre à ces nouveaux besoins émergents, IBM a décidé en 2004, de créer un Institut de Recherche sur la Santé et d'investir 250 millions de dollars pour appliquer les possibilités de la bio-informatique aux enjeux nouveaux de la santé.

Les promesses du Grid Computing

La manipulation de l'information contenue dans le vivant atteint des volumes de données rapidement hors de portée des chercheurs, sans l'aide de la bio-informatique. Par exemple, on estime que pour stocker les séquences du génome d'un seul être humain, une capacité de 300 Tera-octets (soit trois cent fois 1000 milliards d'octets de mémoire) est nécessaire pour l'ensemble des bases. Et que dire des 1,5 millions de protéines différentes identifiées, dont le catalogage et l'identification progressive des fonctions ne cesse d'augmenter chaque jour de par le monde ?

Les technologies de grille, ou *Grid Computing*, ont rendu possibles des avancées spectaculaires telles que la cartographie du protéome en 2001. Le *Grid* permet d'attaquer d'ambitieux programmes collaboratifs, même lorsque l'on ne dispose pas sur un même site de toute la puissance de calcul, de stockage et d'applications logicielles dont on aurait besoin.

En 2004, IBM s'associe à plusieurs grandes organisations scientifiques et éducatives pour lancer un programme mondial humanitaire de grille : le World Community Grid. Ce projet consiste à répartir un logiciel de calcul sur des millions d'ordinateurs de particuliers, d'entreprises ou d'associations à travers le réseau Internet. Les données à traiter sont découpées en petits volumes et envoyées vers les ordinateurs personnels. Une fois le traitement terminé, les résultats sont renvoyés et rendus disponibles aux chercheurs. La puissance de calcul obtenue est directement mise à la disposition d'instituts de recherche dans les domaines médicaux, mais aussi pour l'amélioration des prévisions de catastrophes naturelles ou la lutte contre la faim dans le monde.

Le Programme Décryphon engagé en 2005, utilise également le Grid Computing pour mettre à la disposition des chercheurs une puissance de calcul renforcée sur des projets de recherche très complexes. La mise en parallèle des traitements, l'adaptation des algorithmes de calcul autant que le respect des contraintes de sécurité, font de cette grille un enjeu à la fois technique et humain. Il s'agit de faire collaborer des universitaires avec des industriels, des professionnels avec des bénévoles, une association de malades avec des experts académiques.

Au-delà de la prouesse technique, cette double grille du Programme Décryphon est l'emblème de la recherche de demain : une recherche ouverte sur le monde, en interaction avec lui, pluridisciplinaire dans son approche et qui apprend à travailler à l'amélioration de la vie.