

COMPRENDRE LES ROUA



GES DU VIVANT

Le complexe AmtB-GlnK de la bactérie «Escherichia coli»



L'Institut suisse de bio-informatique a été fondé il y a dix ans. Il compte 20 groupes actifs dans les sciences de la vie et le développement d'outils informatiques capables d'étudier gènes et protéines

L'ISB possède Swiss-Prot, la banque de protéines la plus complète du monde. Le groupe genevois qui la gère vient de publier le catalogue de toutes les protéines humaines connues

Trois autres équipes de l'ISB sont basées à Genève. Elles sont actives dans l'étude du protéome et de l'évolution des génomes ainsi que dans l'analyse de séquences d'ADN et de protéines

Dossier réalisé par Anton Vos
Photos: Olivier Vogelsang

SWISS-PROT PUBLIE LE CATALOGUE DES PROTÉINES HUMAINES

L'Institut suisse de bioinformatique a été créé il y a dix ans pour assurer, entre autres, la pérennité de la base de données Swiss-Prot. Celle-ci regroupe toutes les protéines connues appartenant à 11 000 espèces, dont l'être humain qui en possède 60 000

Après une décennie d'existence, l'Institut suisse de bio-informatique (ISB) rassemble une vingtaine de groupes de recherche répartis entre Genève, Lausanne, Berne, Bâle et Zurich. Le plus grand d'entre eux, le groupe Swiss-Prot basé à Genève, gère et développe la banque de protéines du même nom qui est devenue un outil indispensable – et unique – pour des centaines de milliers de chercheurs du monde entier. Le 2 septembre dernier, les chercheurs de cette équipe ont rendu public, lors d'un congrès à Sienne en Italie, le catalogue le plus complet à ce jour de toutes les protéines humaines connues – le protéome humain. Entretien avec Amos Bairoch, professeur adjoint au Département de biologie structurale et bio-informatique et coresponsable du groupe Swiss-Prot de l'ISB.

AMOS BAIROCH
professeur adjoint au Département de biologie structurale et bio-informatique, Faculté de médecine

Quel est le point commun à tous les groupes de l'ISB?

Amos Bairoch: Tous allient, à des degrés divers, la recherche fondamentale dans les sciences de la vie et le développement d'outils informatiques leur permettant d'analyser le génome ou le protéome. C'est ce qu'on appelle la bio-informatique. Paradoxalement, les activités du groupe Swiss-Prot que je co-dirige sont plus proches du métier d'éditeur puisque notre tâche consiste à publier le catalogue de protéines le plus complet qui existe.

Vous êtes le créateur de Swiss-Prot. Qu'est-ce qui vous a amené à collectionner les protéines?

Au cours de mon diplôme en biochimie, je me suis lancé dans un travail d'analyse et de

séquençage de protéines à l'aide d'un spectromètre de masse. Comme il s'agissait d'une technique nouvelle pour l'époque, il a fallu commander l'appareil. Cependant, lorsqu'il est arrivé à Genève, il est immédiatement tombé

en panne. L'ordinateur qui l'accompagnait, lui, fonctionnait. J'ai alors légèrement changé mon sujet de diplôme afin de développer un programme informatique capable d'analyser des séquences de protéines que le spectromètre



UE

de masse était censé produire un jour. J'ai poursuivi dans la même voie durant ma thèse, entre 1983 et 1989. Cela a abouti à un programme d'analyse de séquences de protéines et d'acides nucléiques, appelé PC/Gene (ce logiciel a connu



«Je n'ai pas pu abandonner le projet de Swiss-Prot et, vingt-deux ans après, je suis toujours attelé à cette tâche»

un franc succès et a été commercialisé par une firme américaine jusqu'en 1993). Seulement, tout au long de la fabrication du logiciel, il a bien fallu que je le teste avec de véritables séquences de protéines. C'est pourquoi, dès mon diplôme, je me suis tourné vers la base de données de protéines américaine PIR (Protein Information Resource), l'héritière de l'*Atlas of Protein Sequence and Structure*, dont l'existence remonte à 1965. J'ai eu accès aux premières données informatisées de PIR en 1984. Je les ai intégrées dans mon programme PC/Gene, mais de nombreux problèmes sont apparus. Il manquait des informations sur les protéines permettant entre autres de faire le lien entre elles et les gènes correspondants. J'ai envoyé des lettres aux responsables pour demander d'y remédier. Je n'ai reçu aucune réponse. Quand je les ai finalement rencontrés au cours d'un congrès scientifique, ils m'ont répondu qu'ils n'avaient pas le temps de s'occuper de ce genre de choses.

Qu'avez-vous fait ?

J'ai cherché à créer une version améliorée de PIR. Les utilisateurs PC/Gene ont apprécié et

m'ont demandé s'ils pouvaient consulter cette nouvelle banque de protéines indépendamment du logiciel. C'est ainsi qu'est né Swiss-Prot, en 1986. La mise à jour continue de cette base de données a vite demandé beaucoup de travail – et il fallait que je termine ma thèse. Je me suis alors tourné vers le Laboratoire européen de biologie moléculaire (EMBL) à Heidelberg, qui gérait déjà une base de données d'acides nucléiques, c'est-à-dire de l'ADN. Ils ont accepté de distribuer Swiss-Prot et d'engager une personne pour s'en occuper. Mais même ainsi, cela n'a pas suffi pour absorber tout le travail. Je n'ai donc pas pu abandonner le projet et, vingt-deux ans après, je suis toujours attelé à cette tâche. Quand j'ai fini ma thèse, Swiss-Prot contenait déjà 10 000 protéines, soit plus du double de la première version de 1986.

Où avez-vous trouvé les premiers fonds pour entretenir Swiss-Prot ?

Au début, la vente du logiciel PC/Gene m'a permis d'engager deux personnes. Le Fonds national pour la recherche scientifique m'en a payé deux autres durant cinq ans. Mais, je n'ai pas pu renouveler cette dernière source de revenus. En 1995, notre pays n'avait pas encore d'accords avec l'Union européenne en matière de recherche scientifique. Toute aide m'a été refusée par ce canal aussi. En 1996, nous avons lancé un cri d'alarme sur le Web, annonçant notre disparition dans les quinze jours si l'on n'obtenait pas de financement. Swiss-Prot étant devenu très populaire dans la communauté scientifique, nous avons reçu une avalanche de lettres de soutien. Des revues comme *Science* et *Nature* ont publié des articles à notre sujet.

Comment avez-vous résolu le problème ?

Cette agitation a réveillé les dirigeants ►

locaux, dont Guy-Olivier Segond, alors conseiller d'Etat genevois en charge de l'Action sociale et de la santé. Cet ardent défenseur des nouvelles technologies nous a permis de rester à flot le temps de trouver une solution. Après beaucoup de cogitations et quelques mesures transitoires, il a été décidé de créer un institut de recherche directement subventionné par la Confédération (à l'instar du Paul Scherrer Institut). C'est devenu l'ISB, fondé en 1998, qui réunissait alors cinq groupes, dont le mien, chargé de s'occuper de Swiss-Prot. Cette stratégie nous a assuré la moitié de nos revenus. L'autre moitié est venue de l'industrie, pour laquelle la consultation de Swiss-Prot est devenue payante – la gratuité étant garantie pour les laboratoires académiques.

Est-ce le cas encore aujourd'hui ?

Non. Ce système, encouragé par les autorités suisses, a convenu à tout le monde, sauf au gouvernement des Etats-Unis. Pour faire pression, ces derniers ont débloqué des fonds pour créer une base de données concurrente, dont l'accès aurait été gratuit pour les entreprises. Ils auraient pu facilement exploiter toutes nos données, puisqu'elles appartiennent au domaine public. Nous n'avions donc pas le choix. Nous avons participé au concours d'attribution du fonds et nous l'avons remporté. La banque de protéines s'appelle depuis 2000 UniProt et est un consortium entre Swiss-Prot, l'EBI (l'Institut européen de bioinformatique, appartenant à l'EMBL) et le PIR (dont les données ont été intégrées dans Swiss-Prot). Ce qui est curieux, c'est que l'argent que nous prélevions sur les firmes privées – qui en avaient les moyens – provient maintenant des contribuables américains, au travers du budget du National Institute of Health (NIH).

Cet arrangement vous convient-il ?

Pas vraiment. Aujourd'hui, avec la faiblesse du dollar face à l'euro et les coupes budgétaires subies par le NIH, on doit s'attendre à une baisse de nos subventions américaines, qui se montent en tout à 2,5 millions de dollars par année. Le comble est que l'idée de vendre les services de Swiss-Prot aux entreprises a refait surface dernièrement.

Vous venez de publier, le 2 septembre dernier, la première version actualisée du protéome humain. De quoi s'agit-il ?

Swiss-Prot est une banque de protéines dont chaque fiche est vérifiée manuellement. Ce contrôle ne concerne pas seulement la séquence en acides aminés de la protéine, mais aussi toutes les informations qui la concernent: la fonction de la protéine, où et quand elle est exprimée ou détruite, quelles maladies lui sont éventuellement associées, etc. Ces fiches signalétiques ne seront jamais définitives puisqu'on publie chaque jour de nouvelles recherches concernant les protéines. De tous les organismes représentés dans Swiss-Prot (animaux, plantes, champignons, microbes, unicellulaires, etc.), le plus important pour nous est bien sûr l'être humain. Son génome est connu depuis 2001, ce qui a permis d'identifier la plupart des gènes. On estime aujourd'hui que l'homme possède près de 21 000 gènes produisant des protéines. Parmi eux, beaucoup n'ont pas encore été étudiés par les chercheurs: ils sont considérés comme des

gènes puisqu'ils en ont l'apparence, mais là où les protéines qu'ils produisent n'ont pas encore été détectées. Le but de notre publication est donc de fournir, en l'état actuel des connaissances, la description la plus précise possible de toutes les protéines humaines connues. Il en manque, puisqu'il existe certainement quelques gènes qui n'ont pas encore été identifiés. D'autres sont probablement de trop, puisque leur existence n'est que théorique.

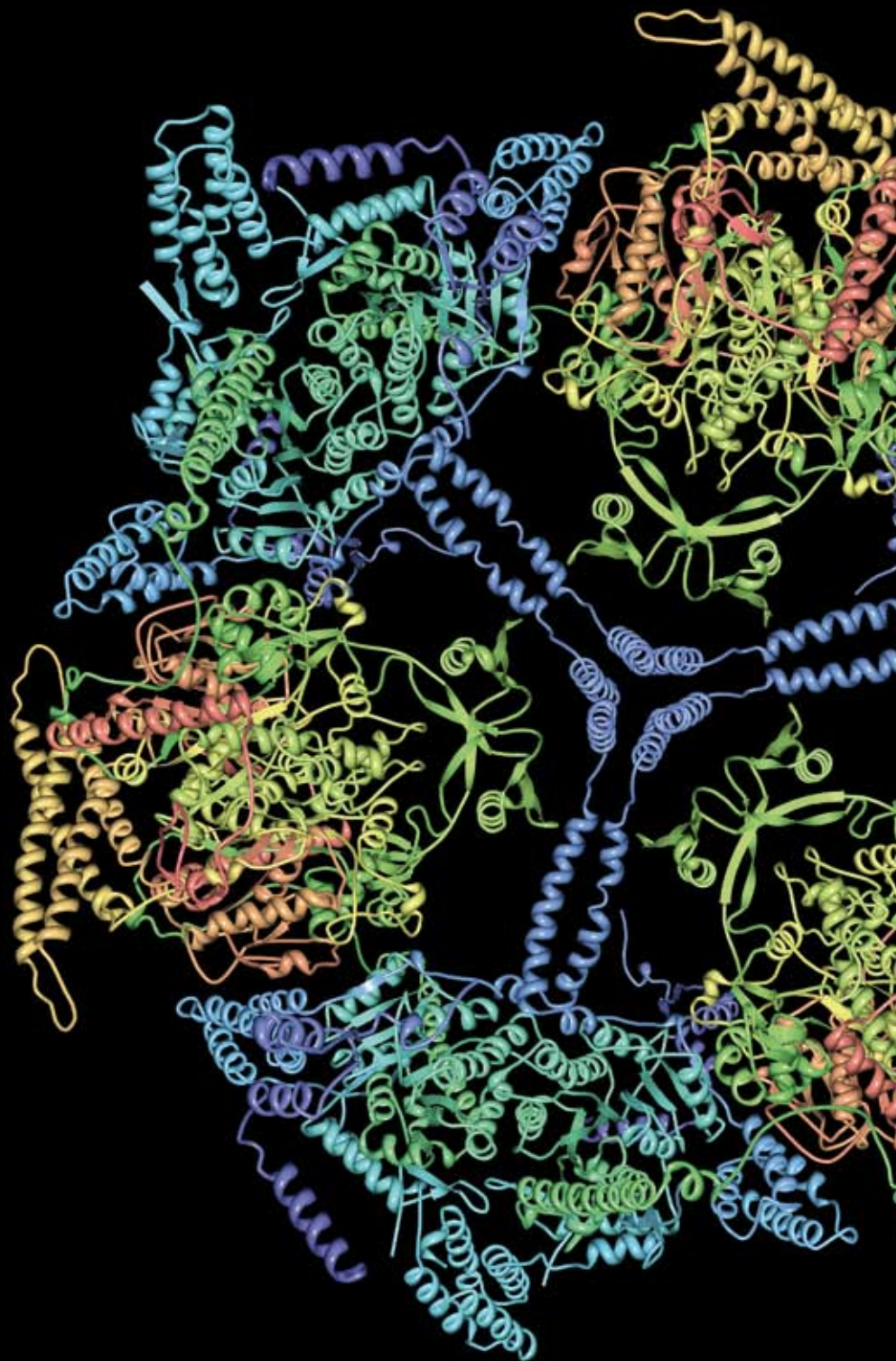
Nous possédons 21 000 gènes, mais combien de protéines différentes produisent-ils ?

Un gène peut en effet produire plusieurs protéines différentes selon qu'il utilise toute sa séquence ou qu'il en néglige certaines portions. En général, toutes les molécules issues du même gène se ressemblent énormément. Dans Swiss-Prot, chaque fiche correspond donc à un gène et répertorie toutes les protéines qui lui sont associées. La nature étant faite d'exceptions, il arrive cependant que le même gène produise des protéines très différentes les unes des autres. Dans ce cas, nous créons plusieurs entrées distinctes. En bref, j'estime le nombre de protéines produites par notre génome à 60 000.

Faudra-t-il du temps pour trouver les protéines manquantes ?

Oui, la tâche sera longue. L'ADN est présent dans toutes les cellules de l'organisme, mais les protéines, elles, ne sont produites que dans les tissus où elles sont nécessaires et parfois même à certains moments de la journée. La protéine qui n'est exprimée que dans un petit nombre de cellules, à très petite dose et pendant quelques heures par journée ne sera pas facile à trouver. De plus, l'identification par spectrométrie de masse est encore très laborieuse. On peut se contenter de l'ARN messager. Mais il n'est jamais sûr à 100% que cette étape intermédiaire entre le gène et la protéine aboutisse bel et bien à la fabrication d'une protéine. L'incertitude grandit encore si l'on se contente des gènes qui peuvent être, sans que cela soit évident à voir, des pseudo-gènes, c'est-à-dire des gènes inactifs. On peut parfois résoudre ce problème en réalisant des comparaisons entre espèces. En effet, si un gène humain inconnu se retrouve presque inchangé chez un animal aussi éloigné du ►

«On estime
aujourd'hui
que l'être humain
possède près
de 21 000 gènes
produisant 60 000
protéines.»



L'ISB et Swiss-Prot en quelques chiffres

► L'Institut suisse de bioinformatique (ISB) est un institut académique comptant une vingtaine de groupes, soit environ 250 personnes réparties entre Bâle, Berne, Genève, Lausanne et Zurich. Il coordonne la recherche et la formation dans le domaine de la bio-informatique en Suisse et assure un certain nombre de services à la communauté scientifique.

► Une des équipes genevoises, le groupe Swiss-Prot, s'occupe de la mise à jour de la base de données de protéines UniProtKB/Swiss-Prot (www.uniprot.org). Elle regroupe actuellement plus de 390 000 fiches, validées manuellement une à une, correspondant à autant de gènes. Ces derniers proviennent d'environ 11 000 espèces différentes, de l'homme aux bactéries en passant par le mammouth et les plantes.

► Chaque fiche contient des informations détaillées sur le nom et la fonction des protéines produites par le gène en question, sur la classification de l'espèce dont celui-ci provient, sur sa séquence, etc. Cette base de données permet aux chercheurs d'identifier des protéines trouvées dans des échantillons biologiques, de découvrir des ressemblances avec des protéines d'autres espèces, etc.

► Le groupe Swiss-Prot fait partie d'un consortium, appelé UniProt, regroupant également l'European Bioinformatics Institute (basé au Royaume-Uni) et la Protein Information Resource (basé aux Etats-Unis).

► Les chercheurs de l'ISB participent à l'effort international d'identification des gènes et des protéines du monde vivant, que ce soit pour l'être humain, les bactéries, les plantes, les champignons ou encore les virus. Leurs recherches s'orientent également vers la phylogénie, à savoir la science qui établit les liens de parenté entre les espèces. Un autre domaine est le développement de techniques permettant d'exploiter la littérature scientifique (l'information ne vient pas toute seule), d'outils servant à analyser les séquences et les structures des protéines, de programme de modélisation en trois dimensions, etc.

► L'ISB possède une école doctorale et propose plusieurs maîtrises universitaires en bio-informatique, dont une en protéomique et bio-informatique à Genève.

point de vue de l'évolution que la souris, alors il y a beaucoup de chance qu'il soit toujours actif et produise des protéines.

Qui utilise Swiss-Prot?

La grande majorité des utilisateurs sont des biologistes, des biochimistes et des médecins qui exploitent la base de données dans le cadre de leurs recherches fondamentales ou cliniques. Les bio-informaticiens sont également nombreux à utiliser Swiss-Prot comme terrain d'essai pour leurs logiciels, ainsi que pour valider ou infirmer certaines hypothèses. De plus en plus de chimistes s'y intéressent aussi, notamment ceux qui cherchent à développer, à partir de protéines existantes, des molécules aux propriétés spécifiques, comme des enzymes capables de dégrader des hydrocarbures ou autres produits polluants. Idem pour les experts en nanotechnologies qui trouvent dans la banque des pièces détachées pour leurs minuscules machines. Swiss-Prot attire aussi des professeurs et des enseignants en biologie. Et puis, il y a ceux que l'on n'attendait pas: les avocats. Ceux-ci sont en général intéressés de savoir à quelle date certaines informations ont été publiées pour la première fois. Cela concerne souvent des litiges ou des procès au sujet de brevets.

Certains génomes décryptés appartiennent à des individus identifiables, comme Craig Venter, l'ex-patron de la firme Celera Genomics. Ils comportent des petites variations personnelles qui peuvent se retrouver au niveau des protéines. Comment gérez-vous ce genre de données qui relèvent de la sphère privée?

Contrairement aux banques de données ADN, qui collectionnent le code génétique entier, Swiss-Prot n'est que partiellement concernée par ce débat. En effet, la plupart des variations entre individus sont situées dans les zones dites «non codantes» de l'ADN. Seul 1% d'entre elles touche des gènes et concerne potentiellement la fabrication des protéines. La majorité de ces modifications sont toutefois neutres et n'affectent pas, en fin de compte, la séquence d'acides aminés. Cela dit, celles qui ont bel et bien un effet sur les protéines, bien que rares, sont évidem-

ment aussi celles qui concernent la recherche médicale et, partant, la sphère privée du donneur. Craig Venter a décidé de rendre publique l'intégralité de son génome. C'est son choix. Seuls les spécialistes peuvent reconnaître les données spécifiques qui ont été tirées de son ADN, mais c'est possible. Toute la question est de savoir ce que l'on va faire avec les génomes des autres individus, de plus en plus nombreux, dont le génome sera décrypté. Nous attendons que les débats éthiques parviennent à un consensus pour savoir si les variantes personnelles que l'on pourra détecter seront enregistrées anonymement ou pas, à l'aide d'un nom de code reconnaissable uniquement par la personne en question et/ou son médecin, etc. Mais il faut faire vite. Dans quelques années, n'importe qui pourra décrypter son génome en quelques heures et pour seulement 100 francs.

Vraiment?

Il existe actuellement à Genève une machine qui lit 3 milliards de «lettres» du code génétique en deux ou trois semaines pour un coût avoisinant les 100 000 francs. La nouvelle génération de ces appareils, déjà sur le marché, fait de même en une semaine et pour 10 000 francs. Il existe même un prototype de machine basé sur une technologie différente, capable de lire l'ADN en direct et qui promet de décrypter un génome humain entier en une heure.

Quel intérêt un individu peut-il avoir à décrypter son génome et à connaître son protéome?

Le génome et le protéome d'une personne peuvent fournir des informations sur d'éventuelles prédispositions à développer des maladies ou des allergies à certaines substances. Là aussi intervient une question éthique. Faut-il informer un patient que son profil protéomique lui confère un risque augmenté pour telle ou telle affection? La règle de base est de répondre à cette question par l'affirmative seulement si la personne peut agir contre cette menace (traitement préventif, changement de mode d'alimentation, etc.). Si la maladie en question est incurable, mieux vaut ne rien dire.

«Dans quelques années, n'importe qui pourra décrypter son génome en quelques heures et pour seulement 100 francs»

Votre banque de protéines ne va-t-elle pas exploser si tout le monde y dépose les siennes?

Non. En fait, tout le monde possède en gros les mêmes protéines. A l'échelle de Swiss-Prot, les variations internes à l'espèce humaine sont très peu nombreuses et facilement intégrables. Ce qui fait considérablement grossir notre base de données c'est l'arrivée de génomes de nouvelles espèces. Bientôt, les chercheurs décrypteront l'ADN entier de toutes les formes de vie qu'ils connaissent. Les zoos feront de même avec leurs propres animaux. Dans vingt ans, Swiss-Prot comptera peut-être une centaine de milliards de protéines.

Quel est l'intérêt de cataloguer tous les génomes et protéomes de la vie sur Terre?

Le fait de connaître tous ces génomes permet de retracer l'évolution des espèces et d'étudier en détail l'arbre de la vie. En revanche, posséder la séquence de toutes les protéines du monde vivant ne servira probablement à rien dans la grande majorité des cas. De temps en temps, il se peut cependant que l'on découvre une protéine dont les caractéristiques peuvent être utiles en médecine, en biotechnologie ou pour l'industrie agroalimentaire. ■

DE CHROMOSOME EN CHROMOSOME

L'ensemble du code génétique de l'être humain est divisé en 24 morceaux (l'un d'entre eux, le Y, ne se trouve que chez les hommes) et compacté dans des structures appelées chromosomes. Chacun contient un certain nombre de gènes producteurs de protéines

Pour ses 10 ans, l'Institut suisse de bioinformatique a mis sur pied l'exposition *Chromosome Walk, au fil du génome humain*. Il s'agit d'un balade passant par 23 postes consacrés aux chromosomes humains. Elle permet l'immersion dans le monde des gènes, des protéines et de la bio-informatique. Le texte ci-dessous en est inspiré.

le chromosome 2 est le résultat de la fusion de deux plus petits. Chez l'animal le plus proche de l'homme, le chimpanzé, il existe encore en deux parties (2A et 2B), tout comme chez le gorille et l'orang-outang.

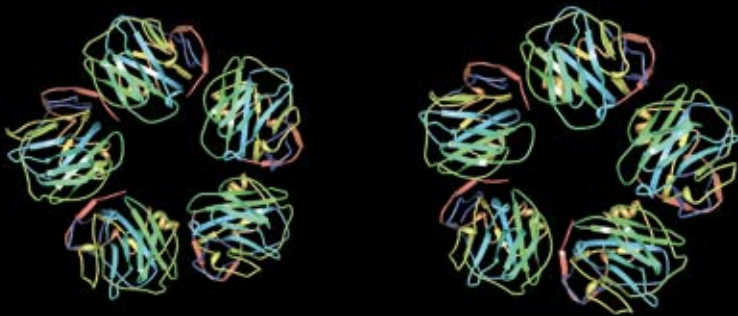
Le chromosome 3 contient plus de 1000 gènes. Chez les animaux et les plantes, la majo-

ou pas. Une approche purement automatique ne peut cependant éviter certaines erreurs comme de confondre certains pseudogènes (des anciens gènes devenus obsolètes et inopérants) avec des gènes. C'est pourquoi il est indispensable de procéder à une vérification «manuelle».

Le chromosome 5 contient plus de 900 gènes dont une partie est encore mal connue. En effet, sur les 21 000 gènes recensés dans le génome humain, un peu plus de la moitié seulement est «bien connue» dans le sens qu'elles correspondent à des protéines caractérisées par les scientifiques. Pour de nombreux autres, seul l'ARN messager a été mesuré. Une troisième catégorie est composée des gènes considérés comme tels car ils ressemblent à des gènes actifs et étudiés chez d'autres animaux (homologues). Cependant, il est possible qu'un gène soit actif chez la souris, par exemple, tandis que son homologue humain soit inopérant. Finalement, les gènes restants ne le sont pour l'instant que par pure prédiction.

Le chromosome 6 contient environ 1150 gènes. Comme tous les gènes, ils présentent ce qu'on appelle des variations de nucléotides ou single nucleotide polymorphism (SNP). Ce sont des petits changements (une lettre du code A, G, T, C qui est remplacée par une autre) qui s'observent entre les individus sans que cela ait des conséquences pathologiques. Un projet est en cours pour séquencer le génome d'un millier de volontaires humains pour pouvoir répertorier précisément toutes ces variations.

Le chromosome 7 contient le gène CFTR qui, s'il subit une mutation, provoque l'apparition de la mucoviscidose. Il existe un certain nombre de maladies génétiques qui sont ►



La protéine humaine c-reactive, dont le gène se trouve sur le chromosome 1

Le chromosome 1 est le plus long des chromosomes humains et le dernier à avoir été totalement séquencé. Déroulé, l'ADN qu'il contient mesure 8 centimètres et compte environ 2200 gènes. Une copie du génome humain entier, soit les 23 chromosomes déroulés et mis bout à bout, mesure 1 mètre. Cela signifie que chaque cellule de notre corps (qui possède deux copies de chaque chromosome) ne contient pas moins de 2 mètres d'ADN. Le nombre total des gènes humains est estimé à 21 000.

Le chromosome 2 contient le gène correspondant à la plus grosse protéine humaine connue, la titine (34 350 acides aminés), qui est un des composants du muscle. Il est admis que

rité des gènes ne sont pas d'un seul tenant. Ils sont éclatés en plusieurs morceaux séparés par des plages qui ne semblent servir à rien. Chacun de ces gènes permet la fabrication de plusieurs protéines différentes, selon que toutes ses parties sont mobilisées ou seulement une combinaison d'entre elles. Chez la mouche, un seul gène (le Dscam) est capable de produire 38 000 protéines différentes par ce mécanisme, c'est-à-dire près de 3 fois plus que le nombre de gènes total de cette espèce d'insecte.

Le chromosome 4 contient plus de 800 gènes. Les scientifiques utilisent des algorithmes pour déterminer quelles portions de l'ADN ont le plus de chances de correspondre à des gènes



dues à l'altération d'un seul gène (Maladie de Huntington, Achondroplasie...). En revanche, des affections comme le diabète, la maladie d'Alzheimer ou les maladies cardiovasculaires sont dites multifactorielles. Plusieurs gènes différents peuvent être impliqués ainsi que des facteurs environnementaux.

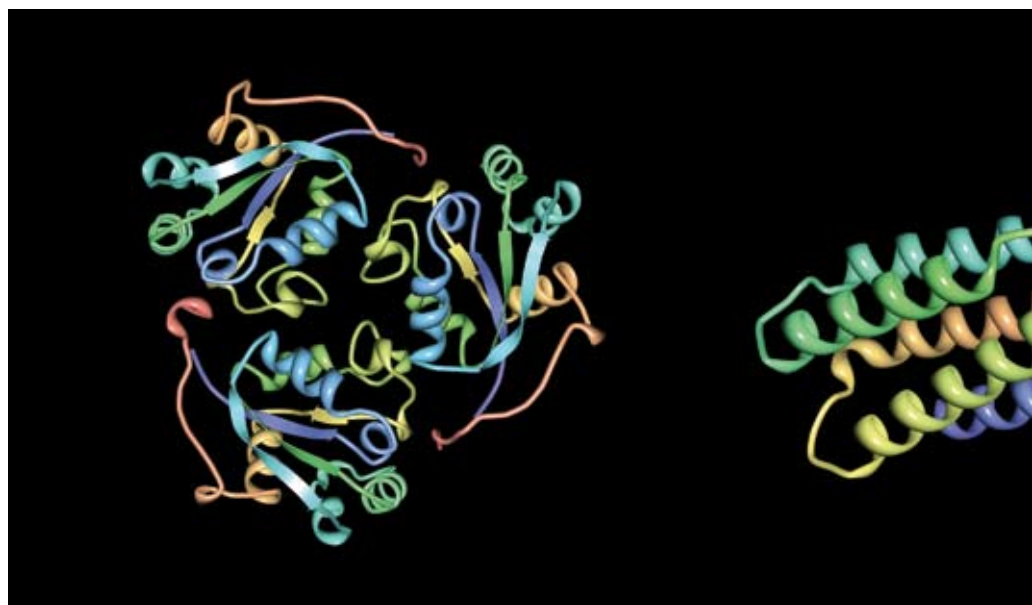
Le chromosome 8 contient un vestige de gène impliqué dans la fabrication de la vitamine C. L'être humain, au cours de son évolution, a en effet abandonné cette fonction probablement en raison d'un apport suffisant de vitamine C par l'alimentation. Du coup ce gène est devenu superflu et n'a plus été soumis à la pression de la sélection naturelle. Il a commencé à accumuler des mutations jusqu'à devenir inopérant. C'est ce qu'on appelle un pseudogène.

Le chromosome 9 compte environ 140 millions de paires de base, c'est-à-dire les nucléotides qui constituent l'ADN et dont il existe quatre types connus sous les appellations A, G, T et C. Au total, notre génome compte 3 milliards de ces lettres. Seuls 3 à 5% d'entre elles composent des gènes. La grande majorité restante a une fonction inconnue, pour autant qu'elle en ait une.

Le chromosome 10 contient le gène CYP2C19 qui code pour une protéine active dans le foie. Cette dernière contribue à dégrader les médicaments que l'on absorbe. Une petite modification de ce gène peut exercer une influence sur la vitesse de cette «digestion». Connaître les variations individuelles de la séquence de CYP2C19 permettrait de prescrire des doses de médicaments adaptées à chaque patient.

Le chromosome 11 contient environ 1400 gènes. Comme les autres chromosomes, il contient des trous, c'est-à-dire des régions qui n'ont pas pu être séquencées. Ces dernières sont composées de séquences qui se répètent un grand nombre de fois. Cette particularité empêche son décryptage à l'aide des techniques actuelles. Il existe environ 400 de ces trous dans le génome humain, surtout dans les régions appelées centromères, c'est-à-dire celles qui se situent au croisement des deux barres du X que représente chaque chromosome.

Le chromosome 12 contient un des quatre groupes de gènes HOX: les HOXC. Les gènes HOX sont parfois surnommés gènes architectes. Ils contribuent au développement des embryons en faisant en sorte que les cellules se spécialisent de manière correcte en fonction



de l'endroit où elles se trouvent dans l'organisme. C'est notamment grâce aux gènes HOX que l'avant-bras précède le bras, que celui-ci se prolonge par la main puis les doigts.

Le chromosome 13 contient plusieurs centaines de gènes. L'étude de l'ADN a révolutionné la classification des espèces. En effet, connaissant le taux de mutation dans les gènes, il est possible d'estimer la distance évolutive qui sépare deux espèces différentes. Ce genre de calculs permet de redessiner l'arbre de l'évolution, parfois avec plus de précision que la paléontologie. C'est ainsi qu'il est apparu que l'hippopotame est plus proche de la baleine que du cochon, contrairement à ce que l'on a déduit de l'analyse du squelette et des dents.

Le chromosome 14 compte un certain nombre de gènes inconnus. Il est possible de se faire une idée de la fonction des protéines qui leur sont associées en étudiant leur séquence. On observe en effet que les protéines sont souvent formées de «domaines» (longs de plusieurs centaines d'acides aminés) bien précis que l'on peut retrouver tels quels dans d'autres protéines. Plusieurs milliers d'entre eux ont été identifiés. En comparant les domaines d'une nouvelle protéine avec ceux des protéines bien connues, il est ainsi possible, dans certains cas, de lui prédire une fonction. C'est ce qui s'appelle dresser un portrait-robot.

Le chromosome 15 contient un des gènes qui déterminent la couleur des yeux. Une étude

sur des familles danoises a montré qu'une seule petite mutation dans celui-ci permet de passer des yeux bruns aux yeux bleus.

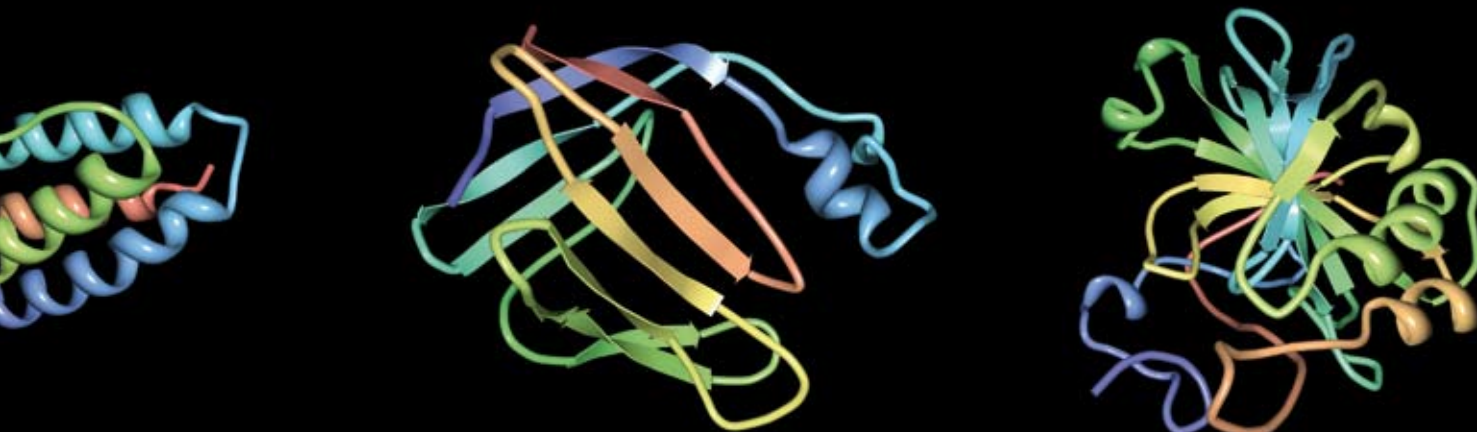
Le chromosome 16 contient des vestiges de gènes qui ressemblent beaucoup à ceux qui, chez la souris, correspondent à des récepteurs de phéromones. Plusieurs études ont montré que l'être humain est toujours sensible à une forme de communication olfactive inconsciente. Mais on ignore encore comment.

Le chromosome 17 contient un gène correspondant à un type de collagène, c'est-à-dire une protéine fibreuse, très résistante à la traction. C'est une telle protéine qui a été isolée dans les restes d'un tyrannosaure vieux de plus de 68 millions d'années. Bien que cette étude américaine demande encore des confirmations indépendantes, elle confirme sur des bases moléculaires que les dinosaures et les oiseaux sont très proches du point de vue de l'évolution.

Le chromosome 18 contient le gène BCL2 qui correspond à une protéine responsable de la régulation de l'apoptose (mort programmée des cellules). Ce gène, lorsqu'il est muté, est impliqué dans un grand nombre de cancers, notamment la leucémie, le mélanome, le cancer du sein et de la prostate.

Le chromosome 19 contient un peu plus de 2 centimètres d'ADN et environ 1500 gènes. Pour près de la moitié des gènes humains, les cher-

Les protéines «nucleoside diphosphate kinase A» (dont le gène se trouve sur le chromosome 17), l'apolipoporphin (non humaine), la «cellular retinoic acid binding protein-type» (chromosome 1) et la «Carbonic anhydrase 2» (chromosome 8).



cheurs n'ont pas encore confirmé l'existence d'une protéine correspondante. Un projet international, doté d'un budget de 11 milliards de dollars, est sur le point de démarrer pour tenter de trouver ces molécules et de déterminer à quoi elles pourraient bien servir.

Le chromosome 20 contient le gène de la protéine prion dont une variante est responsable de la maladie de Creutzfeld-Jakob. Des protéines homologues chez d'autres espèces sont également responsables des maladies de la vache folle (l'encéphalopathie spongiforme bovine qui est transmissible à l'homme par l'alimentation) et de la tremblante du mouton.

Le chromosome 21 est le seul à avoir été en partie décrypté par une équipe genevoise, celle de Stylianos Antonarakis, professeur au Département de médecine génétique et développement. Il arrive chez certaines personnes que ce chromosome existe en trois exemplaires au lieu de deux, ce qui provoque la trisomie 21. Il contient également le gène de la protéine précurseur de l'amyloïde (APP), à l'origine de la formation des plaques amyloïdes que l'on retrouve dans le cerveau des patients atteints de la maladie d'Alzheimer.

Le chromosome 22 est, en 1999, le premier à avoir été entièrement décrypté par le *Human Genome Project*, un consortium public. Dès 1998, une compétition est engagée entre cette dernière et une société privée, Celera Genomics, dont l'objectif est de commerciali-

ser ses découvertes. En 2001, les deux groupes publient en même temps une première ébauche du génome humain. Celera Genomics avouera plus tard avoir utilisé des données du consortium public pour avancer plus vite. Le séquençage total prend officiellement fin en 2003 et les résultats sont depuis gratuitement accessibles sur Internet. Au total, il aura fallu quinze ans, 3000 scientifiques et 2,7 milliards de dollars pour en venir à bout.

Le chromosome X contient environ 900 gènes. Les femmes possèdent, dans chacune de leurs cellules, deux copies de ce chromosome, tandis que les hommes n'en ont qu'une seule. C'est pourquoi il existe des maladies liées à des mutations du chromosome X (hémophilie, daltonisme...) qui sont beaucoup plus fréquentes chez les hommes, ces derniers ne pouvant pas compter sur une copie de secours du gène endommagé.

Le chromosome Y contient environ 70 gènes, dont 21 sont communs au chromosome X. L'un des gènes propres au Y est le SRY. Il détermine la différenciation sexuelle. Certaines mutations dans ce gène peuvent donner naissance à des femmes alors même qu'elles possèdent une paire de chromosomes sexuels XY. ■

L'exposition «Chromosome Walk, au fil du génome humain» se tient sur la Terre de Pregny, aux Conservatoire et Jardin botaniques de la Ville de Genève, du 1^{er} au 30 septembre et sur la place de Milan à Lausanne du 6 au 19 octobre (Renseignements: 022/379 50 50 ou sp_com@isb-sib.ch).

Il aura fallu quinze ans, 3000 scientifiques et 2,7 milliards de dollars pour décrypter tout le génome humain



LE PROFIL DU PROTÉOME

Les maladies ou les intoxications peuvent entraîner des différences dans le contenu en protéines de nos cellules ou liquides biologiques. L'équipe de Frédérique Lisacek, active dans la protéomique, contribue à étudier et comprendre ces variations parfois très fines

Une cellule malade et une cellule saine ne possèdent pas le même contenu en protéines. Le dysfonctionnement dont souffre la première va engendrer un déséquilibre dans sa production de molécules par rapport à la seconde. C'est dans l'étude de ces différences qu'est impliqué le Groupe informatique protéomique (PIG), une des quatre équipes de l'Institut suisse de bioinformatique (ISB) basées à Ge-

nève. Toute la difficulté de ce qu'on appelle l'étude du protéome (contenu en protéines d'une cellule, d'un liquide biologique...) est de séparer les molécules les unes des autres, de les identifier, d'en déterminer la quantité relative et de réaliser des comparaisons de la façon la plus précise possible. L'objectif est à la mesure de l'effort: en identifiant ces petites variations entre les situations saines et patho-

logiques, il est possible de trouver des protéines impliquées dans des maladies, dans la réponse à une intoxication ou à un médicament, etc. Autant d'indications pavant la voie à une médecine du futur offrant un traitement personnalisé pour chaque patient.

Bien que les protéines représentent leur fonds de commerce, les chercheurs du PIG, dirigé par Frédérique Lisacek, n'en voient

FRÉDÉRIQUE
LISACEK
Faculté de
médecine

passer aucune dans leur laboratoire. Leur travail consiste en effet principalement à offrir (et utiliser) les outils informatiques nécessaires aux différentes étapes d'une étude en protéomique, surtout celles d'identification et de caractérisation des protéines.

MELANIE AU TOP 10

«Un de nos logiciels les plus utilisés par la communauté scientifique est Melanie, explique Frédérique Lisacek. Il a été créé par mon prédécesseur à la tête de ce groupe, le professeur Ron Appel, actuellement directeur de l'ISB. Il permet de comparer des images d'électrophorèse bidimensionnelle (une technique qui sépare les protéines, prises dans un gel visqueux, selon leur charge électrique et

leur masse moléculaire). Il est commercialisé depuis les années 1990 et nous continuons à l'actualiser et l'améliorer.»

En plus de fournir et développer ce genre de logiciels, le PIG se charge de la consultation et de l'extraction d'informations de la multitude de bases de données relatives aux protéines et aux peptides. Car il est crucial, mais de loin pas aisé, de se frayer efficacement un chemin parmi les millions de fiches techniques répertoriant toutes les molécules du monde vivant connues à ce jour. Cela demande de la méthode.

C'est d'ailleurs le sujet d'un article récent publié par Frédérique Lisacek et ses collègues dans la revue *Proteomics Clinical Applications* du mois d'août 2007. Les auteurs y répertorient les principales bases de données et outils de recherche dédiés à la protéomique appliquée aux fluides biologiques (sang, urine...). Le tout accompagné de conseils sur l'utilisation des uns et des autres afin d'obtenir des interprétations fiables de résultats expérimentaux.

En bref, le PIG dispose d'une boîte à outils bien garnie pour une bonne recherche en protéomique. Une boîte à outils consultable sur le site ExPASy.org, que l'équipe de Frédérique Lisacek cogère avec le groupe Swiss-Prot de l'ISB. On y trouve l'accès à de nombreuses bases de données présentes sur Internet et une liste spectaculaire de logiciels d'analyse de données biologiques.

POSITION CHARNIÈRE

«Nous participons à de nombreux projets de recherche suisses et européens, note Frédérique Lisacek. Nous jouons généralement le rôle d'interface entre, d'une part, les biologistes, qui souhaitent accroître leurs connaissances des mécanismes de la vie et, de l'autre, les mathématiciens, physiciens et informaticiens qui leur fournissent les outils pour y parvenir. J'adore cette position charnière. Même si, par expérience, je sais que les personnes venant de

ces différentes disciplines peuvent mettre une année avant de se comprendre.»

La toxico-protéomique est un sujet phare du PIG, traité en collaboration avec l'équipe de Denis Hochstrasser professeur au Département de biologie structurale et bioinformatique. L'idée consiste à mesurer de quelle manière un produit toxique ou un médicament agit sur le contenu en protéines des cellules ou fluides biologiques.

«Nous cherchons de très petites variations dans la structure des protéines, précise Frédérique Lisacek. Comme la phosphorylation ou la glycosylation de certaines protéines, c'est-à-dire l'ajout d'un groupe phosphate ou d'un sucre sur la molécule. Ce sont là des processus normaux et courants. Mais ces variations peuvent parfois traduire une situation toxique ou pathologique sans que l'on sache ni comment ni pourquoi. Par exemple, il a déjà été observé que certaines protéines subissent de telles modifications chez les patients atteints d'un cancer et subissant une chimiothérapie, un traitement qui sauve des vies, mais dont la composante toxique demeure importante. Pour mieux comprendre l'effet toxique sur les protéines, il faut pousser l'investigation plus loin en cherchant notamment des combinaisons de modifications caractéristiques de la toxicité.»

L'équipe genevoise s'intéresse également à l'influence spécifique des sucres dans l'apparition de certaines maladies, notamment l'arthrose (dans le cadre d'un projet européen), et le diabète en collaboration avec le Groupe de recherche de protéomique biomédicale, dirigé par Jean-Charles Sanchez maître d'enseignement et de recherche à la Faculté de médecine.

Grâce à ces recherches, on espère parvenir un jour à identifier quels profils protéomiques sont associés à des maladies ou à des sensibilités particulières face à l'administration de substances toxiques et/ou médicamenteuses. «Il serait intéressant de déterminer si un patient fait partie d'une proportion de la population qui ne tolère pas tel ou tel traitement, par exemple, explique Frédérique Lisacek. Cela permettrait de lui proposer un médicament alternatif. En d'autres termes, l'idée est de participer au développement d'une médecine personnalisée, dans laquelle on pourrait prescrire le bon dosage du bon médicament à chaque patient.» ■

«Les personnes venant de ces différentes disciplines (biologie et informatique) peuvent mettre une année avant de se comprendre.»

L'ÉVOLUTION DES GÉNOMES

Le séquençage du génome de plusieurs espèces ouvre la porte à l'étude comparative de l'ADN et à une meilleure compréhension des mécanismes de l'évolution

Comparer les génomes afin de comprendre l'évolution des espèces, tel est le travail du Groupe évolution des génomes, membre de l'Institut suisse de bioinformatique et dirigé par Evgeny Zdobnov, professeur adjoint au Département de médecine génétique et développement. Dans cette optique, le biologiste genevois a déjà participé à l'analyse de presque tous les génomes d'insectes publiés à ce jour (plusieurs mouches et moustiques, un scarabée, l'abeille à miel, etc.) ainsi que de plusieurs vertébrés comme la souris, le rat et le poulet. Et son tableau de chasse continue à s'étoffer d'année en année: son groupe travaille actuellement sur le génome de la vache. Entretien.

Campus: Un génome, c'est-à-dire l'ensemble des gènes d'un organisme, peut-il évoluer?

Evgeny Zdobnov: Notre ADN est constamment sujet à toutes sortes de mutations, que ce soit des délétions, des transpositions, des duplications, etc. Ces modifications, lorsqu'elles se produisent, par exemple, dans une cellule de notre foie, conduisent souvent à la mort de la cellule. En revanche, si elles ont lieu dans la lignée germinale, ces altérations peuvent être transmises à la descendance. La plupart du temps, ce genre d'événements n'entraîne aucune conséquence.

Mais il arrive parfois que cela aboutisse à quelque chose d'utile. Comme un gène qui se duplique par accident et dont la nouvelle copie peut évoluer pour exercer une fonction adaptée aux besoins de l'organisme ou lui conférerait un avantage. Du coup, une telle mutation pourrait être transmise à la génération suivante et se propager ainsi petit à petit dans toute une population. Il ne s'agit pas toujours d'un avantage flagrant. Parfois un nouveau gène (ou une mutation) est conservé parce qu'il n'a pas entraîné la mort de l'organisme. Il se peut alors qu'il occupe une «niche» ou reprenne une partie de la fonction d'un ancien gène et se trans-

mette ainsi aux générations futures. C'est notamment la raison pour laquelle le génome humain, comme celui des autres espèces, est truffé de redondances. Nos gènes sont en général impliqués dans un certain nombre de fonctions différentes et une même fonction peut être exercée par plusieurs gènes distincts.

Le génome humain est-il le plus évolué?

On ne peut pas dire cela. On peut estimer que le cerveau humain est le plus complexe de tous les systèmes nerveux centraux du règne animal actuel. Mais notre génome n'est pas

plus complet, ni plus complexe que celui des autres. Entre autres, nous ne possédons pas le plus grand nombre de gènes, ni les meilleurs. Nos gènes fonctionnent exactement de la même manière que ceux d'une poule. D'ailleurs, il existe plus de différences entre les génomes de deux espèces de mouches (*Drosophila melanogaster* et *Drosophila pseudoobscura*) qu'entre ceux d'une souris et d'un être humain.

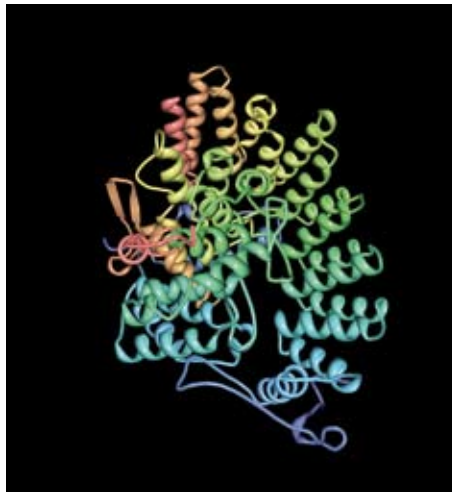
Les différences génétiques ne se traduisent donc pas forcément par des différences morphologiques...

Non. L'analyse moléculaire peut révéler beaucoup plus de différences que celles que l'on peut déduire de la forme des organismes. D'un autre côté, une minime variation génétique peut parfois entraîner un important changement morphologique.

En comparant des espèces différentes, certaines parties du génome sont plus conservées que d'autres. Pourquoi?

On constate en effet que la nature conserve, d'une espèce à l'autre et de manière assez fidèle, les gènes qui produisent des protéines, les séquences d'ADN qui produisent ce qu'on appelle les «micro-ARN» ainsi que de nombreuses zones dont on ignore encore la fonction. Cela signifie simplement que ces régions-là ne supportent pas facilement des mutations et qu'elles sont indispensables à la survie et à la reproduction des individus. Des différences apparaissent bien entendu avec le temps, mais on retrouve des gènes très semblables chez la levure et l'être humain. En revanche, tout le reste du génome accumule régulièrement des mutations aléatoires qui n'ont pas de conséquences sur sa survie et se différencie ainsi très rapidement d'une espèce à l'autre.

la protéine farnesyl-transférase dont le gène se trouve sur le chromosome 8.



«Nos gènes fonctionnent exactement de la même manière que ceux d'une poule»



EVGENY ZDOBNOV
 professeur
 adjoint au
 Département
 de médecine
 génétique et
 développe-
 ment, Faculté
 de médecine

soumis les gènes. Puis grâce à un modèle d'évolution, nous pouvons avoir une idée du temps qu'il a fallu pour obtenir les différences que nous observons entre deux génomes. Ces modèles d'évolution nécessitent toutefois des points de repère paléontologiques, à savoir des fossiles bien datés, afin d'établir une relation entre le temps et le nombre de mutations que l'on observe entre deux génomes. Et cette relation dépend aussi de facteurs comme la durée d'une génération, du nombre de descendants à chacune de ces générations, etc.

Le génome humain varie-t-il beaucoup d'un individu à l'autre?

Oui. Les chercheurs ont comparé les génomes de deux individus qui sont aujourd'hui entièrement séquencés, celui de Craig Venter, le patron de la firme Celera Genomics, et celui de James Watson, codécouvreur de la structure en double hélice de l'ADN en 1953. On a ainsi pu dénombrer entre 3 et 4 millions de petites différences. La plupart de ces variations sont neutres. Mais certaines d'entre elles expliquent les différences morphologiques ou les prédispositions à des maladies. C'est d'ailleurs pour cela que les scientifiques s'intéressent à ce champ de recherche. A tel point qu'un projet international vise à décrypter le génome entier de 1000 individus, d'origines différentes, afin d'en savoir plus sur ces petites mutations. Il faut préciser que les techniques de séquençage ont beaucoup progressé. Elles prennent de moins en moins de temps et coûtent de moins en moins cher. Rien qu'entre Genève et Lausanne, on dénombrera bientôt trois ou quatre appareils de séquençage dernier cri. Cela va nous permettre d'obtenir une quantité colossale de données. Notre frein ne sera plus le nombre de génomes, mais bien plus notre capacité à analyser ces données ainsi que la puissance informatique dont nous disposons. Nous sommes déjà les plus gourmands en la matière dans le Centre médical universitaire. ■

Que sont les micro-ARN et ces portions d'ADN mystérieuses qui ne codent pour rien?

Il existe, chez l'être humain, quelques centaines de gènes codant pour des morceaux très courts d'ARN. Ces derniers jouent un rôle important dans la régulation cellulaire et notamment dans la production de protéines. Quant aux segments d'ADN qui ne correspondent à aucune protéine, ni ARN, on ne connaît par leur rôle. Mais le fait qu'ils soient conservés est un indice fort qu'ils en possèdent un. On aurait pu penser qu'il s'agit de régions placées (pour

une raison ou une autre) à l'abri des mutations. Mais il a été démontré que ce n'était pas le cas; ces segments sont conservés par la sélection naturelle. Il a été démontré que certains d'entre eux sont impliqués dans la régulation de l'expression des gènes.

En analysant les différences entre les génomes, pouvez-vous calculer le temps qui sépare deux espèces de leur ancêtre commun?

Jusqu'à un certain point. Nous pouvons estimer les taux de mutation auxquels sont

DONNER DU SENS AU CODE

Le groupe Parallélisme et calcul scientifique du Département d'informatique est membre de l'Institut suisse de bioinformatique. Il développe des outils informatiques qui permettent d'extraire des informations utiles de l'ADN et des protéines



AAG CTG AAG TTC AGA CGT CTC... Comment s'y retrouver dans cet alignement de lettres qui peut compter jusqu'à 3 milliards de signes? Ce code monotone constitué d'une succession de quatre acides nucléiques (adénine, guanine, thymine et cytosine) est celui du patrimoine génétique de toutes les espèces vivantes. Et un des objectifs du groupe Parallélisme et calcul scientifique (SPC), membre de l'Institut suisse de bioinformatique,

est de développer des algorithmes capables d'extraire de l'information biologiquement utile de ces interminables brins d'ADN.

«Une de nos réalisations récentes est le programme Anâtaxis, explique Bastien Chopard, professeur adjoint au Département d'informatique et responsable du groupe SPC. Ce logiciel permet d'analyser et de comparer des portions de génomes puis de reconstituer de très grands arbres phylogénétiques, c'est-à-dire des

arbres généalogiques contenant des milliers d'espèces différentes.»

Anâtaxis est issu d'une idée originale de Gabriel Bittar, collaborateur extérieur au groupe. Il en a imaginé le concept, mais, manquant de compétences informatiques, il s'est tourné vers Bastien Chopard pour concrétiser le projet. Bernhard Pascal Sonderegger, ancien assistant au SPC, en a fait son travail de thèse qu'il a terminé en 2007.

BASTIEN CHOPARD, professeur, et Jean-Luc Falcone, du Département d'informatique, Faculté des sciences

«Les avantages d'Anâtaxis par rapport aux autres programmes de phylogénie existants sont sa robustesse, sa rapidité et sa capacité à traiter un grand nombre d'espèces à la fois, précise Bastien Chopard. Notre logiciel tient compte des inévitables erreurs et incertitudes liées au décryptage des codes génétiques. En d'autres termes, il traite des données qu'il

considère dès le départ comme pas tout à fait exactes.» De plus, il est capable de déjouer le piège de l'homoplasie: l'évolution peut en effet aboutir deux fois au même trait, dans deux espèces différentes, mais de manière indépendante, sans que les deux espèces partagent un ancêtre commun récent. Du point de vue de l'ADN, cela peut se traduire par des similarités entre deux génomes qui sont en réalité fortuites et ne correspondent pas à une quelconque parenté.

DROIT DE SORTIE

Le groupe SPC travaille également sur le développement d'un programme, très ciblé, dont le but est de pouvoir déterminer, à partir de la séquence d'une protéine, si cette dernière est destinée à sortir de la cellule qui l'a fabriquée et à circuler dans l'organisme (c'est-à-dire sécrétée) ou à demeurer à l'intérieur (dans le cytoplasme).

L'objectif n'est pas aisé à atteindre car il n'existe pas de règle absolue en la matière. Les chercheurs ont identifié des «séquences signal» qui jouent le rôle de clé et permettent à la protéine qui en dispose de passer à travers la membrane cellulaire. Mais le mécanisme d'ouverture des portes est encore largement méconnu. Il semble en tout cas très précis car Dominique Belin, professeur adjoint au Département de pathologie et d'immunologie, qui collabore avec le groupe SPC sur ce sujet, a montré qu'en modifiant un seul acide aminé dans certaines protéines habituellement non sécrétée, on pouvait leur conférer une séquence signal.

«Le rôle de pionnier de l'Université de Genève lui permet aujourd'hui de disposer d'un vivier important de bio-informaticiens»

Comme on ne connaît pas toutes ces «clés» (il en existe différentes sortes), un indice plus pertinent pour la reconnaissance d'une protéine sécrétée est la présence d'un site de clivage. Il s'agit de trois à six acides aminés par lesquels la séquence signal et le reste de la protéine sont attachés (au moment du passage à travers la membrane cellulaire, des enzymes scindent la molécule à cet endroit). Mais cela ne suffit pas non plus, puisque des chercheurs ont trouvé des exemples de protéines sécrétées qui ne disposent pas de sites de clivage.

Pour résoudre tous ces problèmes, Jean-Luc Falcone, collaborateur scientifique au sein du groupe SPC, a développé un logiciel basé sur une série de critères de décision, eux-mêmes établis à partir des connaissances actuelles en biologie moléculaire. L'algorithme est complexe, mais, au final, il permet de reconnaître une séquence signal avec un taux très élevé de réussite (98% au cours d'un essai avec le génome de la bactérie *Escherichia coli*) tout en donnant une interprétation biologique à ses choix.

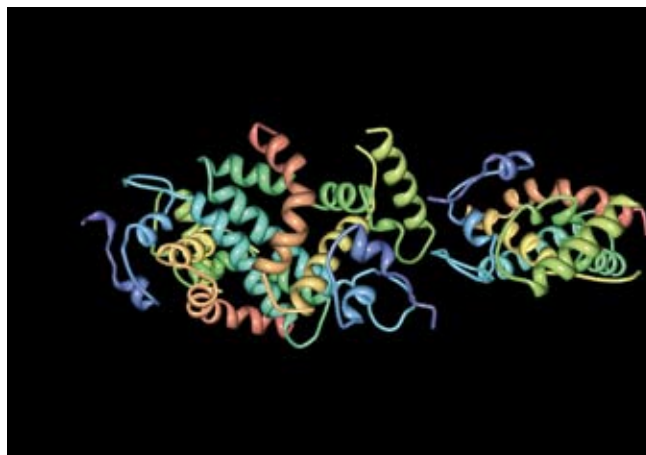
«Ces projets montrent bien l'importance qu'il y a à créer des compétences transdisciplinaires, note encore Bastien Chopard. Un pur biologiste

aura beaucoup de peine à comprendre un algorithme et à transcrire son savoir biologique en langage informatique. D'un autre côté, son expérience de sa discipline est essentielle à l'informaticien pour que son logiciel produise des résultats conformes à la réalité.»

UNE DISCIPLINE À LA MODE

L'Université de Genève a été parmi les premières à proposer une formation en bio-informatique. Un cours à option de 2 heures par semaine est introduit dès 1997 à destination des étudiants en biologie et informatique. Cela deviendra un Diplôme d'études approfondies (DEA) entre 1999 et 2004 avant de se muer finalement en une maîtrise universitaire en protéomique.

«La bio-informatique est actuellement à la mode, estime Bastien Chopard. Toutes les grandes universités offrent désormais une formation de pointe dans cette discipline. L'Université de Genève a bien sûr perdu l'exclusivité qu'elle avait dans ce domaine, mais son rôle de pionnier lui permet aujourd'hui de disposer d'un vivier important de bio-informaticiens et son implication importante dans l'Institut suisse de bioinformatique lui permet de développer une recherche de pointe.» ■



La protéine humaine «endonuclease dicer», dont le gène se trouve sur le chromosome 14