

Des protéines en 3D

Par Gilbert Deléage et Christophe Geourjon*

Connaître la séquence des protéines constitue une première étape indispensable, mais insuffisante, pour comprendre leur rôle dans l'organisme. Afin d'avancer dans cette voie, encore faut-il déterminer leur structure dans l'espace. La bio-informatique a un rôle à jouer, en proposant une pléthore de méthodes.

* Pôle rhône-alpin de bio-informatique, Institut de biologie et chimie des protéines, IBCP-CNRS UMR 5086, 7, passage du Vercors, 69367 Lyon cedex 07. E-mail : g.deleage@ibcp.fr c.geourjon@ibcp.fr

Les astérisques renvoient au glossaire p. 71

Depuis la publication, le 26 juin dernier, de la première version de la séquence complète du génome humain, il est théoriquement possible d'écrire la séquence – nature et ordre des acides aminés – de presque toutes les protéines humaines. Mais cette simple séquence, dite aussi structure primaire, ne suffit pas, loin s'en faut, pour décrire une protéine, comprendre son rôle biologique et ses interactions avec ses voisines ou ses substrats. En effet, ces longues chaînes d'acides aminés se replient sur elles-mêmes et adoptent des conformations très spécifiques dans l'espace. Or c'est cette conformation, ou « structure 3D », qui détermine les propriétés et la fonction d'une protéine. Il est donc indispensable de la définir très précisément – avec une résolution atomique – afin, par exemple, d'en tirer d'éventuelles applications thérapeutiques. La séquence d'une protéine ne permet pas, dans l'état actuel des connaissances, de prédire sa structure 3D (voir l'encadré 1). Il faut la déterminer expérimentalement au moyen de méthodes physiques, telles que la cristallographie ou la résonance magnétique nucléaire (RMN) (voir *Biofutur* [1997] 162, [1999] 189 et 193, le *Technoscope*). Un problème se pose dès lors : si le nombre

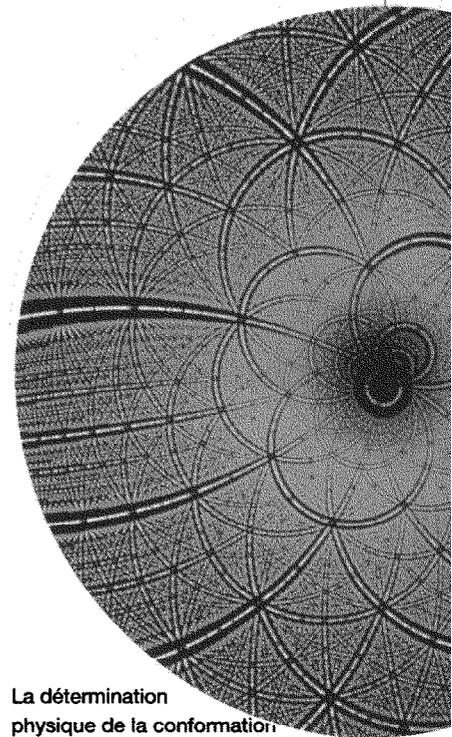
exact de protéines de l'organisme humain est encore sujet à controverse, on sait au moins qu'il s'élève à plusieurs dizaines de milliers (voir l'article de B. Jordan, p. 38). Étant donné le coût et les limites des méthodes physiques, il est impensable d'explorer expérimentalement la structure de chacune d'entre elles. La bio-informatique, nouvelle discipline en plein développement (voir les articles p. 44 et 76) peut apporter des solutions indirectes, sinon en déterminant totalement la conformation des protéines, tout au moins en réduisant l'ampleur de la tâche, et ce à plusieurs niveaux.

> La bio-informatique à la rescousse

Tout d'abord, elle rationalise le choix des protéines à étudier, lesquelles enrichiront les bases de données (séquences, structures et propriétés biologiques). Ensuite, elle permet de s'approcher de leur structure réelle par différentes voies : en comparant les séquences de protéines assurant des fonctions voisines dans différents organismes, en classant les grands types de repliements, en prédisant et même en modélisant une structure

La détermination physique de la conformation des protéines – ici diffraction sur un cristal de Rubisco – reste extrêmement coûteuse.

3D compatible avec différents paramètres constatés ou imposés. Les protéines naturelles sont constituées d'enchaînements de 50 à 30 000 acides aminés (la moyenne est d'environ 400), ou résidus, qui peuvent être de 20 types différents. On comprend dès lors pourquoi chacune d'elles est unique : pour une modeste chaîne de 100 résidus, il existe théoriquement 20^{100} séquences possibles. Qui plus est, chaque acide aminé est lié à ses voisins par deux liaisons covalentes, dites peptidiques, autour desquelles il peut pivoter librement, d'où un nombre astronomique de conformations possibles dans l'espace. Il existe heureusement quelques formations périodiques particulières que l'on retrouve fréquemment le long des chaînes : les hélices α , les brins β et les coudes. Elles constituent la structure secondaire des protéines. Finalement, la molécule entière adopte une conformation tridimensionnelle thermodynamiquement stable, dite structure tertiaire, due à un ensemble de liaisons interatomiques non covalentes, comme les liaisons hydrogène ou les ponts salins (voir la figure p. 68). De plus, des liaisons



covalentes particulières, appelées ponts disulfure, peuvent contribuer à stabiliser la molécule. Bien que chacune de ces liaisons soit énergétiquement faible, leur ensemble suffit à stabiliser la conformation de la protéine dans les conditions physiologiques. Outre leur importance fonctionnelle, les structures secondaires et tertiaires présentent un intérêt cognitif majeur pour le biologiste, car elles ont été mieux conservées que les séquences au cours de l'évolution (1). Enfin, certaines protéines possèdent un niveau supplémentaire d'organisation, la structure quaternaire, qui consiste en l'assemblage de plusieurs sous-unités protéiques (*voir plus loin*).

Si la prédiction *ab initio* de la structure tertiaire semble encore hors de portée, il existe de nombreuses méthodes, encore imparfaites, de prédiction de l'emplacement des éléments de structure secondaire (hélices α , brins β). D'un point de vue bioinformatique, on peut considérer cette opération comme le passage d'une séquence utilisant un alphabet

à 20 lettres (les acides aminés) à une séquence ayant un alphabet à trois lettres (les structures secondaires : H pour hélice, E pour brin étendu, R pour le reste). Pour cela, il faut employer des méthodes statistiques, dont les premières datent de 1978. Toutes reposent sur la connaissance des structures 3D d'un échantillon alors restreint – une trentaine – de protéines « modèles » servant de base de comparaison. L'une utilise les « tables d'occurrence », qui reposent sur les proportions observées de chacun des 20 acides aminés dans un état structurel donné (2). Une autre, appelée théorie de l'information, a introduit un nouveau concept en prenant en compte l'influence des résidus voisins sur la conformation d'un acide aminé donné (3).

Depuis, grâce aux progrès des méthodes physiques expérimentales, le nombre de conformations connues de protéines de séquences significativement différentes – moins de 25 % d'acides aminés identiques entre deux séquences – a atteint 1 500. Chaque jour, six nouvelles structures 3D sont déposées dans la *Protein data bank*, une banque de données internationale gérée par le Research Collaboratory for Structural Bioinformatics (4). Si cet échantillon n'est pas nécessairement représentatif de l'ensemble des protéines, il constitue néanmoins une base suffisante pour mettre en œuvre des méthodes prévisionnelles plus

sophistiquées que les précédentes : méthodes des « plus proches voisins » (*Nearest neighbours method**), d'auto-optimisation, d'apprentissage par réseaux de neurones, ou de chaînes cachées de Markov* (5, 6, 7). En règle générale, ces méthodes permettent de prédire correctement les structures secondaires dans 72 % des cas environ, au lieu de 33 % par tirage aléatoire d'une conformation parmi les trois considérées. Il semble difficile d'améliorer ce résultat, ne serait-ce que parce que l'influence de la structure tertiaire sur la secondaire n'est pas prise en compte. Ces méthodes sont disponibles pour les biologistes sur le serveur Internet que nous avons installé (8, 9, 10).

Quant à la structure tertiaire, si l'on ne peut pas encore la prévoir directement à partir de la séquence, il existe plusieurs manières de l'approcher. La modélisation moléculaire consiste à construire un modèle structural en comparant la séquence de la molécule étudiée avec celle des protéines de conformation connue disponibles dans les banques de données. En effet, deux protéines ayant des séquences similaires possèdent certainement des structures tridimensionnelles proches, notamment au niveau de leur chaîne principale : elles partagent le même repliement, ou « *fold* ». On estime que, dans l'univers des protéines, il existe au plus quelques milliers de repliements différents (11). La première

(1) C. Chothia, A.M. Lesk (1986) *EMBO J.* 5, 823-826.

(2) P.Y. Chou, G.D. Fasman (1978) *Adv. Enzymol.* 47, 45-148.

(3) J. Garnier *et al.* (1978) *J. Mol. Biol.* 120, 97-120.

(4) voir www.rcsb.org/pdb

(5) J.M. Levin *et al.* (1986) *FESB Lett.* 205, 303-308.

(6) C. Geourjon, G. Deléage (1994) *Protein Eng.* 7, 157-164.

(7) B. Rost, C. Sander (1993) *J. Mol. Biol.* 232, 584-599.

(8) C. Combet *et al.* (2000) *TIBS* 29(12), 147-150.

(9) voir <http://pbil.ibcp.fr/NPSA>

(10) C. Blanchet *et al.* (2000) *Bioinformatics* 16, 286-287.

(11) C. Zhang, C. Delisi (1998) *J. Mol. Biol.* 284, 1301-1305.

1 - Du « gène bleu » au « Seti » protéique

Des expériences classiques de dénaturation – traitements chimiques ou physiques modérés qui induisent la perte de la conformation sans altérer la structure primaire – ont démontré que la conformation « native » d'une protéine dépend à la fois de sa séquence et du milieu dans lequel elle est solubilisée. Il semble que, dans la plupart des cas, aucun autre paramètre n'intervienne... si ce n'est l'interaction avec d'autres protéines. C'est donc bien la séquence des acides aminés qui contient l'information nécessaire au repliement de la protéine dans sa conformation biologiquement active. Ce repliement représente le meilleur compromis entre l'enfouissement des résidus d'acides aminés hydrophobes (alanine, leucine, isoleucine, proline et valine), puisque la plupart des milieux organiques sont aqueux, et les possibilités de rotation autour des liaisons chimiques. Par ailleurs, les protéines dénaturées ne mettent que quelques millisecondes à quelques jours pour retrouver leur structure tridimensionnelle originelle. Puisque le temps de transition entre deux conformations, déterminé par spectroscopie, est de l'ordre de 10^{-13} seconde, il est évident que la protéine n'acquiert pas successivement tous les états possibles dans sa recherche de la meilleure stabilité thermodynamique, mais qu'elle suit un repliement hiérarchisé et chronologique. Les mécanismes assurant ce « choix » restent malheureusement inconnus, donc irréproductibles. Pour déduire la

conformation d'une protéine de sa seule séquence, les biologistes en sont réduits à explorer systématiquement l'univers des conformations possibles, à la recherche du meilleur compromis. Une tâche gigantesque... pour laquelle deux types de projets existent. Le premier est dû à la société IBM, qui s'est lancée dans la construction d'un ordinateur d'un million de processeurs, appelé « Blue Gene », théoriquement capable de simuler *ab initio* le repliement d'une protéine à partir de sa séquence (1, 2). Cette machine, qui atteindra le petaflop (10^{15} opérations/seconde), devra tout de même fonctionner une année entière pour calculer le repliement d'une seule protéine ! Le second, baptisé *fold@home*, est un projet planétaire du groupe de Vijav Pande, à l'université Stanford de Californie (États-Unis) (3). Il s'agit de solliciter les internautes afin de distribuer les calculs sur leur machine personnelle, dont les capacités sont rarement utilisées dans leur totalité. L'opération est largement inspirée du projet *seti@home* (recherche de signaux provenant d'une éventuelle intelligence extraterrestre), qui a réussi à mobiliser 500 000 internautes et à obtenir en 18 mois une puissance de calcul équivalente à plus de 400 000 années d'un PC ! ■

(1) R.F. Service (1999) *Science* 286, 2250.

(2) www.research.ibm.com/compsci.bluegene.html

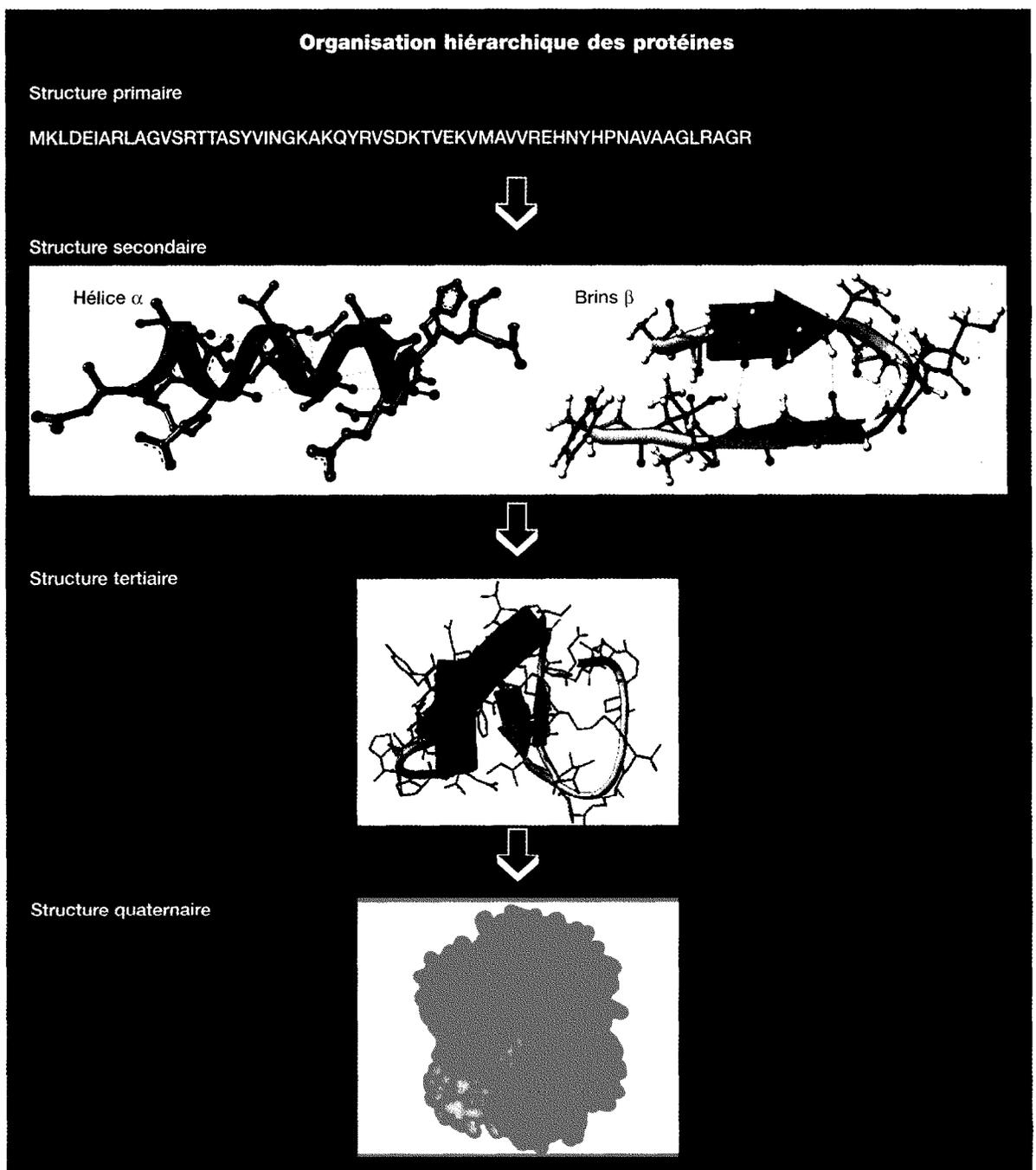
(3) www.stanford.edu/group/pandegroup/Cosm/index.html

*** étape consiste donc à estimer la similarité des deux séquences, ce qui implique de déterminer par le calcul un alignement qui maximise le nombre d'acides aminés identiques entre les deux. Pour cela, trois types d'opérations *in silico* sont autorisés : substitution ou insertion/délétion d'acides aminés et glissement de parties de la séquence. Lorsqu'une protéine répertoriée partage plus de 30 % d'acides aminés, après alignement, avec celle que l'on veut étudier, sa structure tridimensionnelle peut servir d'« empreinte » pour construire un modèle de la seconde. On peut dès lors utiliser soit une méthode dite

substitutive, soit une méthode géométrique pour cette construction.

Les méthodes substitutives consistent à échanger les chaînes latérales – les parties des résidus non engagées dans les liaisons peptidiques, qui diffèrent d'un acide aminé à l'autre – de l'empreinte par celles de la séquence à modéliser, en conservant le plus possible leur orientation. Cette substitution se fait sur ordinateur en utilisant au cas par cas un logiciel graphique, ou automatiquement, grâce à des bases de données de rotamères (ensemble des orientations possibles pour une chaîne latérale). Si certains atomes se rapprochent trop, du fait

de l'encombrement stérique des chaînes remplaçantes, les chaînes latérales sont réorientées « manuellement », de façon à minimiser ces mauvais contacts. Enfin, on détermine l'énergie potentielle de la molécule selon les principes de la mécanique moléculaire, qui considère notamment chaque liaison de deux atomes comme un ressort oscillant autour d'une position d'équilibre favorable. Cette énergie est ensuite minimisée sur l'ensemble de la molécule, grâce à différentes méthodes mathématiques (méthode de plus grande pente, gradient conjugué, méthode de Newton-Raphson). Bien que très utilisée, et



De la séquence des acides aminés (structure primaire) au repliement dans l'espace (structure tertiaire) – voire, dans certains cas, à l'association de plusieurs monomères, dite structure quaternaire –, la conformation tridimensionnelle des protéines comporte plusieurs niveaux d'organisation. Elle est responsable de leur fonction.

La thérapie génique consiste à implanter des gènes codant des protéines médicaments directement dans des cellules ou un organe malades. Les protéines seront donc synthétisées sur le site même des lésions. Les vecteurs transportant ces gènes peuvent être, par exemple, des virus non pathogènes modifiés de manière à se diriger spécifiquement vers la cible choisie. Pour ce faire, une méthode consiste à insérer un peptide (une courte séquence d'acides aminés) spécifique d'un récepteur ou d'un type cellulaire de la cible dans la protéine de la capsid virale. En l'occurrence, il s'agissait d'introduire dans les vecteurs viraux un peptide de 14 acides aminés (dit L14) qui se lie spécifiquement à une intégrine* de la surface des globules blancs (1). Cette méthode soulève deux difficultés majeures : l'insertion du peptide ne doit pas altérer l'assemblage du virus, et le peptide doit être exposé à la surface du virus.

En collaboration avec l'équipe de Michael Hallek, de l'université de Munich (Allemagne), nous avons déterminé, dans la capsid virale (virus adéno-associé, AAV, humain de type 2), les sites d'insertion les plus favorables, à l'aide d'une analyse de séquence, de prédictions de structure et de modélisation moléculaire.

Nous avons ensuite obtenu des vecteurs mutants porteurs du peptide, et le tropisme de l'un de ces virus a été effectivement modifié. L'adénovirus est devenu capable d'infecter spécifiquement une lignée de globules blancs dont les intégrines ne se lient pas normalement à lui (2, 3). ■

(1) voir www.lmb.uni-muenchen.de/groups/hallek/Hallek.html#AAV

(2) A. Girod *et al.* (1999) *Nat. Med.* 5, 1052-1056.

(3) Brevet international n° WOEP9904288, 21 juin 1999, Medigene (1999) M. Hallek, M. Ried, G. Deléage & A. Girod.

avec succès, cette méthode ne permet aucune hypothèse quant à la structure de la protéine au niveau des insertions et délétions d'acides aminés (qui différencient la protéine étudiée de l'empreinte). Pour ces régions, il faut recourir à des algorithmes particuliers de recherche de boucles dans les banques de données, puis faire un choix parmi les candidats possibles.

> Ne pas violer les contraintes...

Plus récentes, les méthodes géométriques de construction d'un modèle de conformation consistent à mesurer des distances sur la protéine « empreinte », puis à les appliquer à la protéine à modéliser grâce à des algorithmes (*distance geometry**) de reconstruction à partir de ce jeu de distances. Après optimisation énergétique par mécanique et dynamique moléculaires, on obtient un ensemble de solutions, ou faisceau de structures. Ces solutions doivent satisfaire au mieux à la fois les données de la chimie (liaisons covalentes, angles de valence, rayons de van de Waals) et les contraintes de distance déduites de l'empreinte. La convergence du faisceau et l'absence de violation des contraintes (l'écart entre la contrainte introduite et la distance mesurée sur le modèle ne doit pas excéder 0,05 nm) sont des critères de qualité. Cette approche présente l'avantage de ne pas faire d'*a priori* sur les insertions et délétions. De plus, la superposition de toutes les solutions générées permet d'appréhender visuellement la définition atteinte par les modèles, notamment au niveau des boucles.

Actuellement, les performances de ces méthodes dépendent directement du taux de similarité entre la séquence de la protéine à modéliser et celle de

l'empreinte. Dès lors, comment améliorer la qualité des modèles lorsque aucune homologie n'est détectable par alignement de séquences ? C'est l'un des enjeux de la bio-informatique. Si les banques de données ne recèlent aucune séquence similaire à celle de la protéine à étudier, on peut recourir à une méthode encore en cours de développement : le repliement inverse, ou *threading*. Elle consiste à estimer la compatibilité entre la séquence de la protéine et un ensemble représentatif des repliements connus. Il faut pour cela disposer d'un échantillon de toutes les unités de repliements (*folds*) possibles, puis essayer d'y appliquer successivement la séquence, en cherchant la solution qui répond au mieux à une contrainte donnée. En général, il s'agit d'optimiser la répartition spatiale des acides aminés hydrophiles et hydrophobes. Même si l'on obtient des résultats encourageants, la description des contraintes reste très grossière, et il faut prendre en compte d'autres facteurs, tels que le rapport surface/volume, le rayon de giration, etc. Entendons-nous bien : les méthodes de repliement inverse ne donnent pas la solution du problème, mais proposent un classement par ordre décroissant des meilleurs candidats. Il n'est donc pas possible pour l'instant de trouver la bonne conformation, mais plutôt de distinguer les solutions possibles des solutions incompatibles. Cela représente tout de même une réduction d'un facteur 100 du nombre de structures à étudier, et fait de cette approche une étape préliminaire indispensable avant d'imaginer d'autres méthodologies.

Dans les méthodes exposées jusqu'ici, le choix de l'empreinte repose sur la connaissance de la séquence et la prise en compte de certaines contraintes chimiques, physiques ou moléculaires.

Il existe une tout autre approche, reposant cette fois sur les données biologiques expérimentales. En effet, deux protéines partageant une même fonction biologique doivent avoir des éléments de structure proches. Sur la base d'études expérimentales comparatives (mutagenèse dirigée, activités enzymatiques, interactions moléculaires), le biologiste émet des hypothèses sur la structure tridimensionnelle et bâtit un modèle initial qui sera confronté sans cesse à l'expérimentation. Si cette approche, dite de modélisation moléculaire par analogie, donne de bons résultats pour certains types de repliement, elle reste d'usage marginal et ne s'applique que très difficilement dans le contexte de la génomique structurale*. En effet, au cours de l'analyse de la totalité des protéines issues d'un génome, on se trouve forcément en présence de nombreuses séquences de protéines, dites « orphelines », ne possédant aucune homologie, biologique ou moléculaire, avec des protéines connues, y compris chez d'autres organismes.

> Utiliser les données biologiques

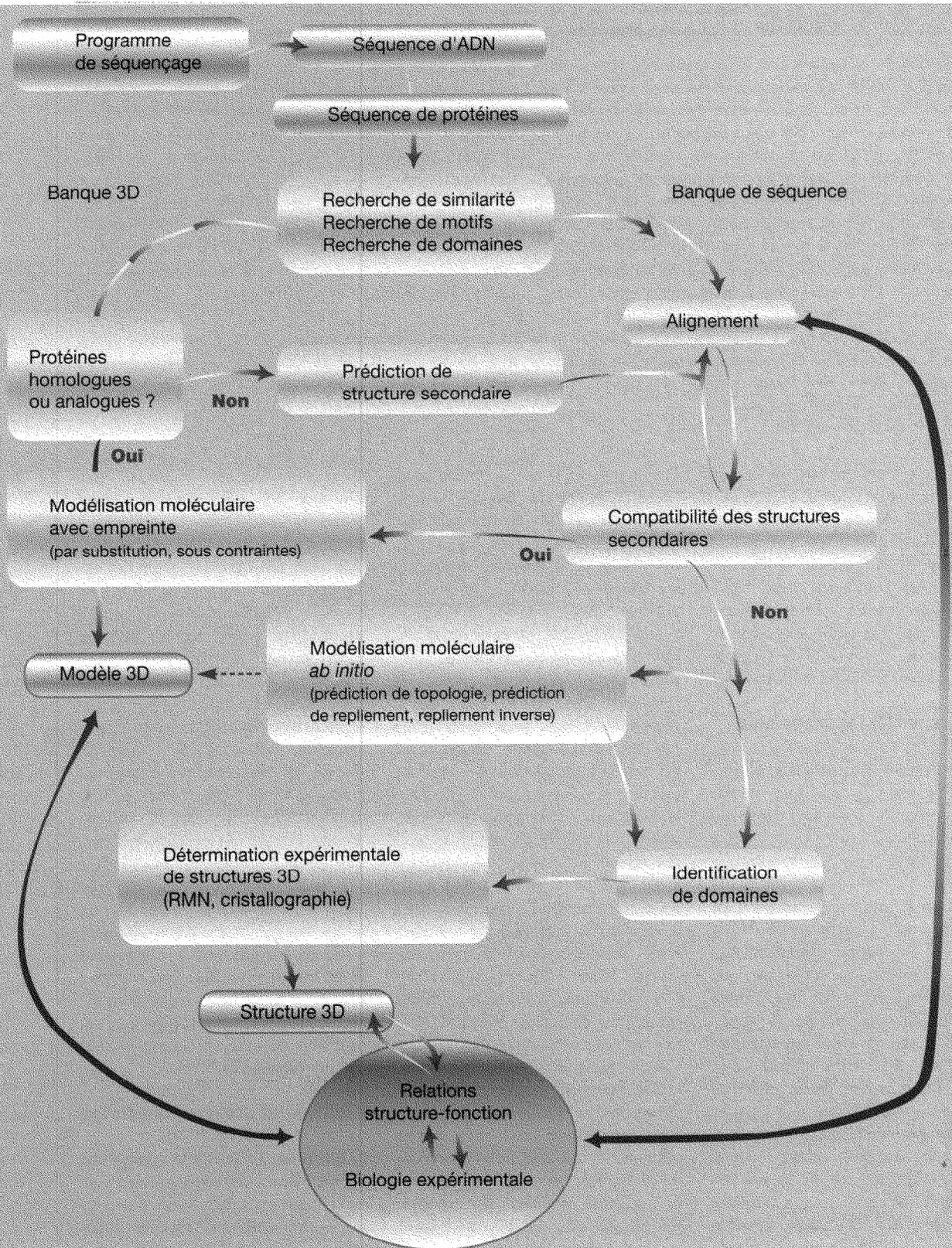
Enfin, pour ajouter à la complexité du problème, certaines protéines se présentent sous la forme d'assemblages moléculaires, appelés oligomères selon la terminologie de Jacques Monod et Jean-Pierre Changeux, qui associent de façon spécifique plusieurs molécules protéiques, identiques ou non, dites monomères ou sous-unités. Ces édifices constituent la structure quaternaire. L'association des monomères repose sur une propriété générale des protéines : la reconnaissance moléculaire. Cette dernière intervient tant au cours de l'interaction des protéines avec leur substrat qu'entre les

(12) voir <http://www.structuralgenomics.org/main.html>

(13) voir <http://s2f.carb.nist.gov/>

(14) voir <http://proteome.bnl.gov>

(15) A. Sali (1998) *Nat. Struc. Biol.* 5, 1029-1032.



© RÉALISÉ PAR E. SIMONIN D'APRÈS G. DELEAGE, C. GEOURJON

Un chercheur connaissant la séquence d'une protéine et désirant déterminer sa conformation dans l'espace dispose de différentes méthodes, qu'il choisira selon les renseignements complémentaires (existence de protéines homologues, données biologiques fonctionnelles,...) dont il dispose.

oligomères. Mettant en jeu des paramètres de type géométrique (complémentarité des surfaces en regard), dynamique (adaptation des surfaces moléculaires) et chimique (propriétés des groupes chimiques portés sur ces surfaces et capacité à engager des liaisons faibles), la reconnaissance

moléculaire permet une interaction très spécifique, très efficace, mais aussi réversible, entre les partenaires. Pour les besoins de la recherche pharmaceutique, les scientifiques développent des algorithmes d'amarrage moléculaire, ou *docking*, qui permettent de comprendre et mimer les

mécanismes de reconnaissance entre des molécules chimiques appelées ligands, souvent de petite taille, et des protéines. Il est cependant irréaliste de vouloir appliquer ce type de méthode à des molécules protéiques interagissant pour former un oligomère, du fait du nombre immense de

possibilités et de l'absence d'une théorie exacte. Il est donc encore impossible de prévoir l'éventuelle structure quaternaire d'une protéine. Pour comprendre les interactions spécifiques, il faut d'abord déterminer avec une grande précision (résolution inférieure à 2 Å) la conformation des molécules impliquées, par cristallographie ou RMN. Il s'agit ensuite d'identifier l'ensemble des liaisons faibles mises en jeu. Pour cela, la bio-informatique dispose d'outils comme la phylogénie structurale, qui consiste à comparer la conformation de protéines ayant des fonctions biologiques identiques dans différentes espèces. Cette comparaison permet parfois de comprendre comment les structures ont évolué dans l'histoire des espèces. À titre d'exemple, les protéines des organismes extrémophiles (vivant dans des environnements extrêmes en termes de température, de pH, de salinité, ou autre) peuvent apporter des enseignements sur la conservation de ces liaisons faibles. Nous l'avons vu, avec la connaissance des génomes complets, et en particulier de celui de l'homme, un nombre considérable de nouvelles protéines

est brusquement dévoilé. Dès lors, se pose la question de la détermination expérimentale de leur conformation : c'est tout l'objet de la génomique structurale. Des pays comme les États-Unis ou l'Allemagne se sont déjà engagés dans cette démarche (12, 13, 14). En France, une réflexion nationale est actuellement menée, afin de définir rapidement le cahier des charges d'une approche combinant biologie et bio-informatique structurales. Dans cette perspective, le problème à résoudre est le choix des premières protéines à étudier, en fonction de leur intérêt biologique et structural. Les contingences économiques doivent également être prises en compte, étant donné le coût de la détermination expérimentale de la structure d'une protéine : environ 200 000 dollars actuellement. La bio-informatique assumera un rôle essentiel dans le processus de décision, sur deux points. Tout d'abord, en déduisant les fonctions probables des protéines à partir de leur séquence, elle aide à identifier les molécules d'intérêt. Ensuite, elle est capable de prédire si une protéine présente une structure originale

– différente des molécules déjà répertoriées – qu'il sera intéressant de résoudre expérimentalement. Elle pourra alors constituer une empreinte pour la modélisation d'autres protéines de la même famille. On estime ainsi que la connaissance de 10 000 structures originales suffirait à modéliser la quasi-totalité des protéines connues (15).

Pour en savoir plus

• *Nat. Struct. Biol.* (2000) vol. 7, numéro spécial, 927-994. <http://structbio.nature.com/structuralgenomics/>

Glossaire

Chaîne de Markov : progression aléatoire en estimant la probabilité de transition à chaque étape.

Distance geometry : reconstruction d'un objet à partir d'un ensemble de contraintes géométriques (distances et angles).

Génomique structurale : détermination de la structure 3D des protéines localisées à partir des génomes.

Intégrine : protéine membranaire servant de récepteur d'adhérence.

Liaison covalente : liaison chimique de forte énergie, donc stable, reposant sur la mise en commun d'électrons entre les atomes impliqués.

Liaison faible : différents types de liaisons chimiques de basse énergie, faisant intervenir des forces comme, par exemple, l'attraction électrostatique.

Nearest neighbours method : utilisation de la similarité entre sous-séquences de protéine pour prédire la structure.

Pont disulfure : liaison covalente entre groupements sulfhydryles portés par des acides aminés de type cystéine.



CENTRE NATIONAL
DE LA RECHERCHE
SCIENTIFIQUE

CNRSFormation
au service de l'Entreprise

"La génomique et ses conséquences pour l'innovation dans les industries du vivant"

du 7 au 9 mars 2001 - ESPCI - PARIS (75)

Pierre-Gilles DE GENNES (Directeur de l'ESPCI) La bio-ingénierie : une nouvelle science pour l'ingénieur.

Michel MORANGE (PARIS VII) La biologie moléculaire : histoire d'une révolution moléculaire dans les sciences du vivant.

Claude BENICOURT (ENS-Cachan) Quelques définitions et précisions terminologiques sur les méthodologies du génie génétique.

Jacques HAEICH (Directeur du Programme Génome au Ministère de la Recherche) Données historiques et techniques sur la mise en oeuvre des programmes de séquençage des génomes.

Francis QUETIER (Université d'Evry) Apport du séquençage et prospective dans le domaine végétal.

Jean ROSSIER (ESPCI) L'après-génome ou comment utiliser les résultats.

Jean-Michel CLAVERIE (CNRS) Acquisition, mise en forme et traitement des résultats. Problèmes d'interprétation. Naissance de la bioinformatique.

Techniques combinatoires en biologie et en chimie :

- Les "puces" à ADN. Principe, élaboration et utilisation.

- Applications des biopuces.

- La spectrométrie de masse.

- Les biocapteurs.

- La chimie combinatoire.

Pierre LEGRAIN (HYBRIGENICS) Secteurs d'utilisation des nouvelles technologies induites par la connaissance des génomes.

Prospective dans les différents domaines considérés.

Marc GIGET (CNAM) Des concepts aux produits industriels : la longue marche de l'innovation.

Maurice CASSIER (CNRS-CERMES) "Nouvelle économie" et nouveaux modes de production dans le domaine de la génomique et de la post-génomique.

Bernard CALVINO (ESPCI) Bioéthique et réglementation : réalités ou fantasmes ?

Présentations de quelques entreprises innovantes dans le domaine de la génomique et de la post-génomique.

Public : cadres de la R&D et décideurs des industries utilisant les procédés du vivant : biotechnologies, industries pharmaceutiques et agroalimentaires
Programme complet des conférences et inscriptions : CNRSFormation <http://www.cnrs-gif.fr/cnrsformation/Incubio.html>

CNRSFormation Bât. 31 Avenue de la Terrasse 91198 Gif-sur-Yvette Cedex Tél. : 01 69 82 44 55 Fax : 01 69 82 44 89