

DIEU NE JOUE PAS AUX DÉS, IL PRÉFÈRE LE LEGO

Pierre Chardin

Notre génome porte l'information pour fabriquer près de 100 000 protéines différentes (80 000 selon une estimation récente).

Le séquençage systématique de centaines de milliers de fragments d'ADNc humains (EST pour *expressed sequence tags*) fait que l'on dispose maintenant de séquences partielles pour une majorité de ces protéines. La quasi totalité devrait être connue d'ici la fin du millénaire. Ce travail impressionnant permet d'avoir une idée de plus en plus précise de la diversité des protéines codées par notre génome. Actuellement plus de la moitié de ces fragments séquencés « au hasard » présentent des analogies avec des protéines déjà connues, et il apparaît que ces 100 000 protéines se répartissent en un relativement petit nombre (quelques milliers) de grandes familles, ayant chacune quelques dizaines de membres. Comme toujours en biologie il y a des exceptions, protéines « uniques en leur genre » ou familles de milliers de protéines, comme celle des récepteurs à sept domaines transmembranaires qui permettent à nos organes sensoriels de percevoir des informations extrêmement diverses et subtiles [1]. Mais cette répartition en familles aide beaucoup lors de la découverte d'une nouvelle protéine car, lorsqu'elle appartient à une famille déjà caractérisée, on peut avoir une première idée de sa fonction.

On observe aussi que beaucoup de protéines, et presque toutes celles de grande taille, sont constituées de

« domaines », c'est-à-dire de modules structuraux que l'on retrouve dans d'autres protéines, et qui présentent en général des analogies de séquence. Par exemple les protéines Vav sont constituées d'une mosaïque de domaines appelés CH*, Dbl*, SH2*, SH3*, PH*, et de motifs structuraux comme les *leucine zipper*, « doigt de zinc », ou « hélice/boucle/hélice » [2]. Dans la protéine IRS-1 on retrouve un « domaine PH » associé à un « domaine PTB* », alors que dans la protéine Shc, un « domaine PTB » [3] est associé à un « domaine SH2 ». On peut immédiatement avoir une idée de la fonction de chaque domaine, si elle a déjà été comprise pour une autre protéine. A l'inverse, quand un nouveau domaine est mis en évidence, on recherche un point commun entre toutes les protéines où il est trouvé, ce qui aide à imaginer quelle pourrait être sa fonction. C'est le cas par exemple du domaine PH que l'on trouve dans de très nombreuses protéines ayant des cibles membranaires [4, 5].

De plus, ces domaines sont remarquablement conservés au cours de l'évolution ; presque tous ceux qu'on connaît chez les mammifères sont aussi présents chez la drosophile ou le nématode, et beaucoup se retrouvent même chez les plantes et les

ADRESSE

P. Chardin : directeur de recherche à l'Inserm. Institut de pharmacologie moléculaire et cellulaire, Cnrs UPR 0411, 660, route des Lucioles, Sophia Antipolis, 06560 Valbonne, France.

* CH : calponin homology ; Dbl : conservé dans Dbl ; SH2, SH3 : Src homology 2 and 3 ; PH : pleckstrin homology ; PTB : phosphotyrosine binding.

levures. Les techniques génétiques, particulièrement puissantes pour ces organismes, ont aidé à élucider la fonction de certains de ces modules. Cette organisation en « domaines » est, en outre, une parfaite illustration du « bricolage moléculaire » dont parle François Jacob [6], que l'on peut aussi comparer à un gigantesque jeu de Lego. Ainsi l'évolution puise dans un répertoire limité de « domaines » ayant chacun une fonction simple, et les juxtapose pour faire émerger, par assemblage, des protéines de plus grande taille, à fonctions complexes. Cette récupération d'objets existants pour faire du neuf est sans doute la voie la moins coûteuse en termes d'évolution.

On estime que d'ici l'an 2000, la structure tridimensionnelle sera connue pour au moins une protéine de chacune des grandes familles de domaines à l'exception des protéines membranaires. L'accumulation de ces données permettra de découvrir de nouvelles analogies structurales, en l'absence d'analogies de séquences détectables, et suggèrera des analogies de fonction qui passent encore inaperçues.

Le plus souvent, la connaissance de la structure tridimensionnelle d'une protéine permet de modéliser la structure de la chaîne principale pour les autres protéines de cette famille, mais pas la position des chaînes latérales. C'est donc une information très utile pour prédire les acides aminés essentiels, les modifier par mutagenèse, et étudier en quoi ils modifient le fonctionnement de la protéine, mais elle n'est pas suffisante pour prédire le mécanisme précis d'interaction avec un ligand et éventuellement concevoir de nouveaux agents pharmacologiques permettant de bloquer cette interaction. Les progrès parallèles de la biologie structurale, et des programmes de modélisation qui s'affineront en se confrontant aux données accumulées, devraient nous permettre de connaître (ou de prédire avec une bonne précision) la structure de ces domaines pour de très nombreuses protéines, dans les dix ans qui viennent. Cette information aidera vraisemblablement à concevoir de nouveaux médicaments modifiant

spécifiquement l'activité de certaines de ces protéines, lorsqu'elles sont impliquées dans des maladies.

La connaissance de la structure ne donne pas immédiatement accès à la fonction ; par exemple, la structure des protéines Ras ou des domaines PH a été déterminée avant que l'on ne comprenne leur fonction, mais en permettant une mutagenèse intelligente, elle facilite la compréhension de cette fonction.

Par ailleurs, la progression spectaculaire des techniques de « double-hybride » [7] et des « aptamères* » [8] permet d'établir des listes de partenaires potentiels ou de cibles peptidiques ou nucléiques d'une protéine. L'accumulation rapide de ces informations et leur mise en commun dans des banques à libre accès définira un réseau d'interactions possibles entre protéines. Il restera ensuite à déterminer leurs localisations intracellulaires pour montrer que les protéines en question ont l'occasion de se rencontrer, ou à obtenir par la génétique la confirmation de leur interaction fonctionnelle.

Comme on le verra dans ce numéro, l'étude des domaines protéiques isolés est en bonne voie, mais il est beaucoup plus difficile d'avoir accès à la structure de protéines complètes avec de multiples domaines.

La tyrosine kinase src, premier oncogène découvert, a focalisé l'attention des meilleures équipes depuis vingt ans, ses principaux domaines ont été définis et leur structure déterminée, alors qu'en parallèle s'accumulaient les données fonctionnelles montrant l'importance des réarrangements de ces différents domaines lors de l'activation. Mais ce n'est que très récemment que la structure de la protéine presque complète a pu être déterminée au prix de gros efforts, donnant enfin une image cohérente qui explique toutes ces observations [9].

* Aptamères : petites molécules d'acides nucléiques, ARN ou ADN, fabriquées par PCR et sélection in vivo, pour leur aptitude à se lier spécifiquement à une macromolécule donnée, ou pour une activité biologique particulière.

S'il existe une unité remarquable de ces familles de domaines dont les fonctions sont pour l'essentiel conservées au cours de l'évolution, la nature a rivalisé d'imagination pour combiner ces différents domaines, et inventer de nouvelles façons de modifier leurs interactions pour moduler l'activité des protéines dans lesquelles ils se trouvent. Il reste donc un travail considérable pour comprendre le positionnement de ces différents domaines dans les protéines de grande taille et leurs réarrangements qui jouent bien souvent un rôle régulateur essentiel ■

RÉFÉRENCES

1. Bockaert J. Récepteurs couplés aux protéines G : des concepts à la structure. *Med Sci* 1996 ; 12 (suppl n°10) : 79-85.
2. Romero F, Dusanter-Fourt, Fischer S. La protéine Vav, une mosaïque de domaine de signalisation. *Med Sci* 1997 ; 13 : 629-38.
3. Borg JP, Fournier E, Birnbaum D, Margolis B. PTB, un domaine d'interaction protéine-protéine important dans le « jeu de dominos » de la transmission du signal. *Med Sci* 1997 ; 13 : 647-56.
4. Boivin P, Lecomte MC. Les domaines homologues de la pleckstrine. *Med Sci* 1997 ; 13 : 639-46.
5. Chardin P. Rôle du domaine PH dans le recrutement membranaire des facteurs d'échange pour les petites protéines G. *Med Sci* 1997 ; 13 : 731-3.
6. Jacob F. Le jeu des possibles. Paris : Fayard, 1981.
7. Plessis A, Camonis JH. Le système double-hybride, mode d'emploi. *Med Sci* 1994 ; 10 (suppl n° 6-7) : I-IX.
8. Colas P, *et al.* Genetic election of peptide aptamers that recognize and inhibit Cdk2. *Nature* 1996 ; 380 : 548-50.
9. Xu W, Harrison SC, Eck MJ. Three-dimensional structure of the tyrosine kinase c-Src. *Nature* 1997 ; 385 : 595-602.

TIRÉS À PART

P. Chardin.