



Exploration des interactions protéiques à l'échelle génomique

Pierre Colas

P. Colas: Équipe de Brian B. Rudkin, Laboratoire de biologie moléculaire et cellulaire, UMR 5665, École normale supérieure de Lyon, 46, allée d'Italie, 69364 Lyon Cedex 07, France.

► L'étude fonctionnelle des milliers de nouvelles protéines découvertes grâce aux projets de séquençage de génomes constitue un défi scientifique et technologique majeur. À plus d'un titre, l'exploration exhaustive des interactions protéiques apparaît comme l'une des approches les plus puissantes de la génomique fonctionnelle. Cette revue décrit les différentes stratégies expérimentales envisageables reposant sur la technique du double-hybride. Elle aborde les problèmes et les limitations inhérents à cette technique ainsi que les résultats attendus et la difficulté de leur exploitation. Enfin, des méthodes d'exploration alternatives sont présentées. ◀

La progression fulgurante de la génomique a donné naissance à une nouvelle discipline communément appelée génomique fonctionnelle, dont le but est d'attribuer une fonction – ne serait-ce qu'approximative – aux milliers de protéines identifiées par les divers projets de séquençage. Parmi les différentes approches développées ou envisagées en génomique fonctionnelle, l'exploration des interactions protéiques à grande échelle semble des plus prometteuses, pour des raisons aussi bien fondamentales que techniques. En premier lieu, puisque l'immense majorité des protéines exercent leur fonction en interagissant avec d'autres protéines, l'une des façons les plus directes d'appréhender la fonction d'une protéine consiste à identifier ses partenaires d'interaction. En second lieu, la méthodologie permettant la mise en évidence d'interactions protéiques est désormais suffisamment fiable pour être applicable à l'échelle génomique. Enfin, cette méthodologie se prête volontiers à l'automatisation, condition essentielle à une application à grande échelle. A la lumière de quelques travaux précurseurs qui ont établi la faisabilité d'une telle entre-

prise, cet article tente de présenter les différentes approches expérimentales pouvant être envisagées, les problèmes que posera inévitablement leur mise en œuvre à grande échelle, mais aussi leurs promesses et leurs limitations.

Criblages « double-hybride »

Apparue voici 10 ans [1], la technique du double-hybride est rapidement devenue l'une des méthodes de choix pour la mise en évidence d'interactions entre protéines, et son utilisation est désormais répandue dans de nombreux laboratoires. Bien que d'autres méthodes puissent être envisagées à plus long terme, le double-hybride est aujourd'hui la seule technique applicable à l'échelle génomique. Les principes de cette technique et ses différentes déclinaisons ont fait l'objet de nombreuses revues auxquelles le lecteur non averti est invité à se reporter [2-4]. Un rappel très succinct est proposé dans la *figure 1*.

En 1996, le laboratoire de Stanley Fields publia les résultats de la première exploration exhaustive des interactions protéiques au sein d'un orga-

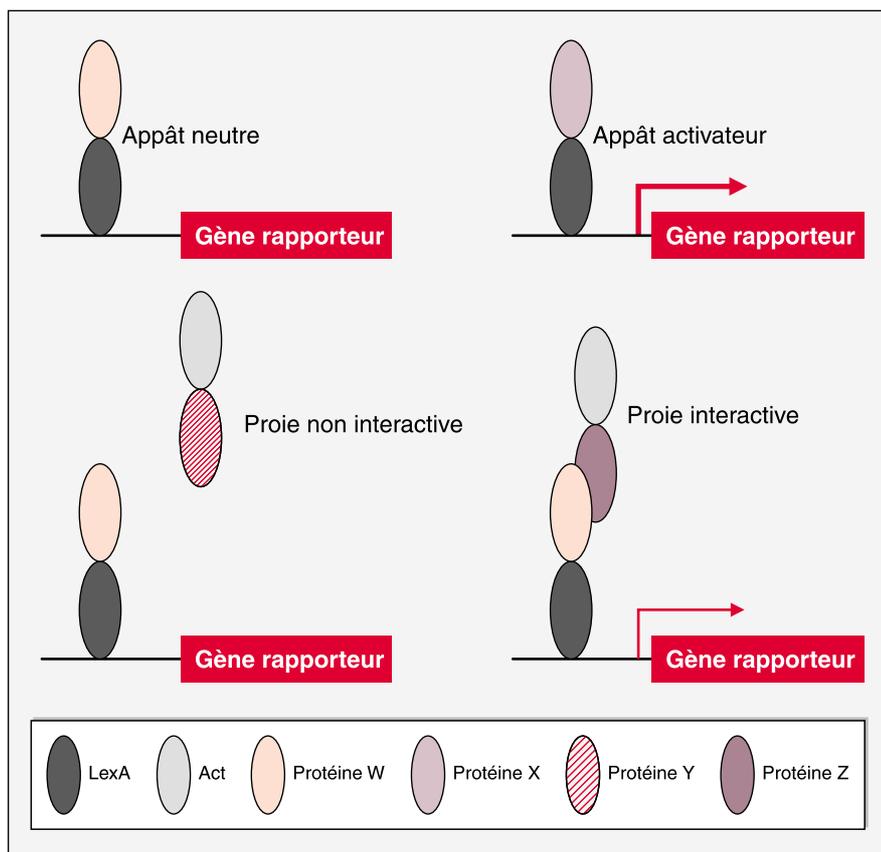


Figure 1. **Principe élémentaire de la technique double-hybride.** Cette technique repose sur deux propriétés fondamentales des facteurs de transcription. En premier lieu, ces facteurs sont modulaires et réunissent un domaine de liaison à l'ADN et un domaine activateur, interchangeable entre différents facteurs. En second lieu, la reconstitution d'un facteur de transcription actif peut être assurée par le rapprochement topographique entre ces deux domaines. Ainsi, l'interaction entre une chimère constituée du domaine de liaison à l'ADN (ici LexA) fusionné à un appât et une chimère constituée du domaine activateur (ici Act) fusionné à une proie reconstitue le facteur de transcription actif. Ce dernier peut alors transcrire des gènes rapporteurs tels que LEU2 et lacZ, conférant aux levures respectivement la capacité de pousser sur milieu dépourvu de leucine, et une couleur bleue sur milieu contenant du X-Gal. L'un des problèmes de la technique du double-hybride réside dans le fait que de nombreuses protéines fusionnées au domaine de liaison à l'ADN activent spontanément la transcription des gènes rapporteurs, indépendamment de toute interaction avec une proie.

nisme, le bactériophage T7 [5]. Deux banques génomiques, l'une d'appâts et l'autre de proies ont été construites et introduites dans deux souches de levure de sexe opposé. Les deux souches furent ensuite conjuguées afin d'isoler les individus diploïdes présentant un phénotype d'interaction et d'identifier les paires de protéines interactives. Vingt-cinq interactions purent ainsi être identifiées, parmi lesquelles 9 impliquent des protéines dont la fonction est inconnue et 11 concernent des interactions intramolé-

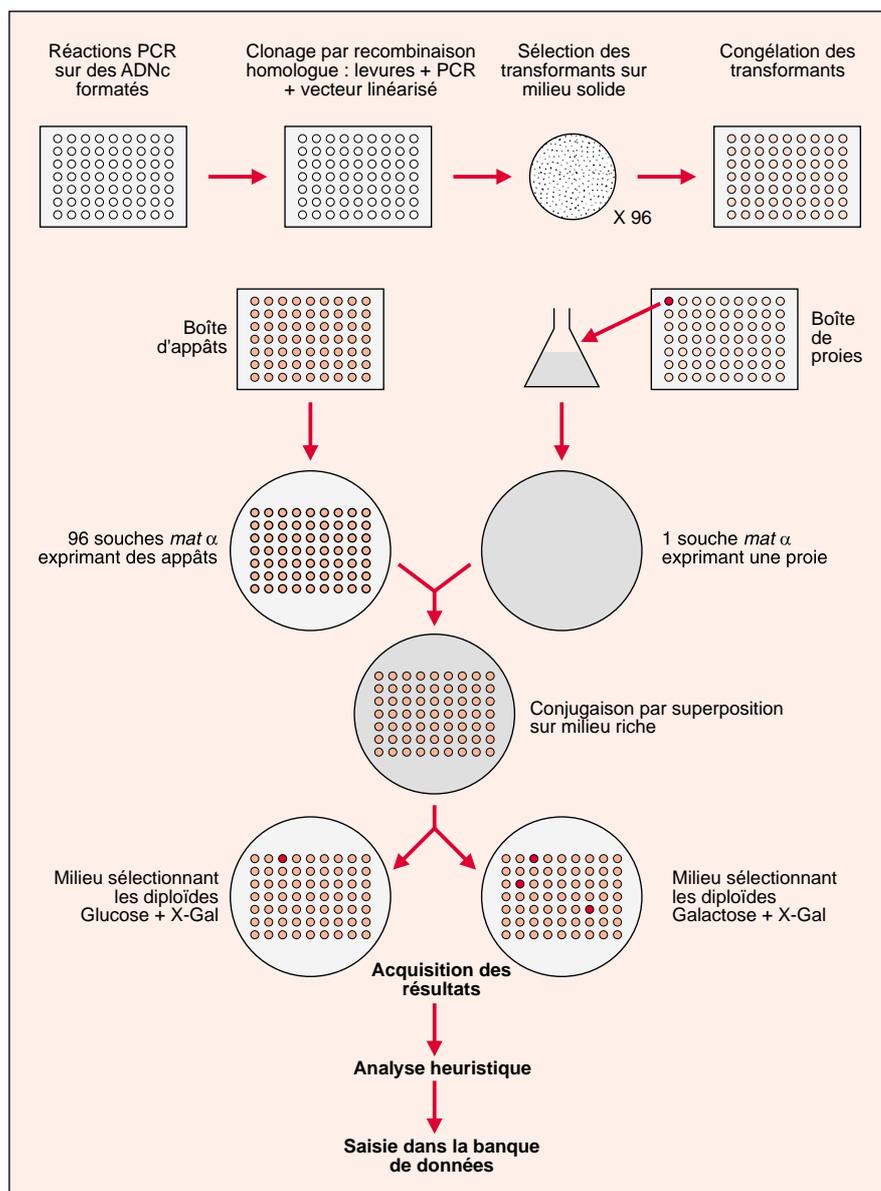
culaires dues à la fragmentation aléatoire des régions codantes intervenue lors de la construction des banques. Il faut cependant noter que parmi les 4 interactions détectées auparavant par des techniques biochimiques, seules 2 furent détectées par ce criblage. Bien que ce travail fasse date dans l'histoire encore jeune de la génomique fonctionnelle, force est de souligner que le protéome du bactériophage T7 ne comprend que 55 protéines.

Peu après, le laboratoire de Pierre Legrain proposa une approche simi-

laire dans laquelle une souche exprimant un appât donné est conjuguée à une souche contenant une banque génomique de proies [6]. Les protéines interactives isolées sont alors utilisées comme appâts dans des criblages itératifs, et des mécanismes moléculaires de régulation peuvent ainsi être explorés au fil d'une « promenade protéomique ». Appliquée initialement à l'étude du mécanisme de régulation de l'épissage chez *Saccharomyces cerevisiae*, cette approche est à présent appliquée à des protéomes de petite taille tels que celui du virus de l'hépatite C, de *Helicobacter pylori* (<http://www.hybrigenics.fr/>), ou encore de *Bacillus subtilis*.

Parallèlement à ces travaux, Russ Finley développa dans le laboratoire de Roger Brent une stratégie différente fondée sur l'utilisation d'une collection de souches de levures formatée exprimant des appâts d'identité connue, et pouvant être conjuguées à une souche exprimant une proie d'identité également connue [2]. Cette collection d'appâts fut assemblée grâce aux dons de nombreux scientifiques ayant utilisé le système double-hybride construit dans le laboratoire de Roger Brent. Récemment, cette collection comptait environ 800 protéines différentes, issues de divers organismes et participant à de nombreux mécanismes de régulation. Dans la première application publiée, un criblage de cette collection réalisé en utilisant comme proie la protéine de drosophile Roughex permit de révéler une interaction avec la cycline E et de préciser ainsi la fonction de ce régulateur important dans le développement de l'œil [7]. Cette collection restreinte et hétéroclite a permis à ses utilisateurs de découvrir d'autres interactions fort intéressantes et informatrices (figure 2). Dans un effort plus ample mais concentré sur une espèce, le laboratoire de Stanley Fields amplifia par PCR les 6000 gènes codant pour les protéines de *Saccharomyces cerevisiae*, de manière à ce que les produits d'amplification puissent être clonés rapidement dans différents vecteurs, notamment de double-hybride, par recombinaison homologue chez la levure [8]. Plus récemment, le groupe GeneNet du laboratoire Clontech amplifia plus de 250 000 EST (*expressed sequence tag*) humains par PCR en utilisant des amorces permettant

Figure 2. **Principe d'un criblage «double-hybride» d'une collection d'appâts formatée.** Des amplifications d'une collection formatée d'ADNc [9] sont effectuées par PCR en une ou deux étapes [8] à l'aide d'amorces présentant une identité de séquence avec les vecteurs d'appât ou de proie. Les levures sont alors transformées avec ces produits d'amplification en présence du vecteur d'appât ou de proie linéarisé. Les transformants sont sélectionnés sur milieu solide, puis récupérés et re-formatés afin de constituer un stock congelé. Les boîtes d'appâts sont inoculées à partir du stock congelé, tandis que les boîtes de proies sont inoculées à partir d'une culture liquide elle-même inoculée à partir d'une souche congelée. Les souches d'appâts et la souche de proie sont conjuguées par réplique sur milieu riche, et les conjugants diploïdes sont alors obtenus par réplique sur milieu de sélection, contenant deux sources de carbone différentes et du X-Gal afin de révéler l'activité de la β -galactosidase codée par le gène rapporteur lacZ. Dans le système double-hybride développé dans le laboratoire de R. Brent [2], l'expression des proies est placée sous le contrôle d'un promoteur inductible par le galactose et réprimé par le glucose (GALp), tandis que l'expression des appâts est placée sous le contrôle d'un promoteur constitutif (ADHp). Cette configuration permet de distinguer facilement les phénotypes d'interaction imputables à une activation spontanée de la transcription par certains appâts (tels que celui présenté sur la première ligne de la matrice) des signaux positifs pouvant correspondre à des interactions entre appâts et proies (tels que les deux autres signaux présentés).



De tels criblages ont été effectués de façon manuelle par les différents membres du laboratoire de R. Brent (R. Finley, A. Reymond, A. Mendelsohn, K. Robzyk, A. Saunders, B. Cohen, P. Colas) à l'aide de la collection d'appâts construits par les différents utilisateurs du système double-hybride développé au laboratoire. La plupart des résultats bruts peuvent être consultés sur : <http://www.molsci.org/JeffsDatabasePages/front.html>. La procédure expérimentale proposée ici devrait permettre un grand degré d'automatisation.

de cloner par recombinaison homologue les trois phases de lecture de ces EST dans un vecteur de proie [9]. Les transformants contenant les proies peuvent alors être formatés en plaques de 96 puits. Ces travaux ouvrent la voie vers la constitution de collections d'appâts formatées, couvrant des protéomes complets et pouvant être criblées selon une approche présentée dans la figure 2. Le Tableau 1 résume les avantages et les inconvé-

nients respectifs des approches faisant appel à des banques et à des collections formatées.

Problèmes inhérents à la méthodologie

Tout biologiste ayant eu la joie ou l'infortune d'utiliser la technique du double-hybride a été confronté au fléau que constituent les résultats faussement positifs. De nombreuses

revues ont tenté de définir la nature de ces faux positifs, d'en cerner la causalité complexe et multiple, et de proposer des solutions permettant de les distinguer des véritables interactions [2-4]. Bien que n'étant nullement rédhibitoire dans la perspective d'analyses à grande échelle, ce problème devra toutefois être abordé avec la plus grande compétence. Ainsi, l'un des défis majeurs posés par la mise en œuvre de criblages

Tableau I

CRIBLAGES « DOUBLE-HYBRIDE FAISANT APPEL À DES BANQUES OU À DES COLLECTIONS FORMATÉES

	Construction	Normalisation	Identité des protéines	Faux négatifs Faux positifs	Automatisation
Banques	Techniques classiques bien éprouvées	<ul style="list-style-type: none"> • Excellente pour les banques génomiques • Problématique pour les banques d'ADNc 	À déterminer par séquençage ou hybridations sur puces à ADN	Moins de faux négatifs avec des banques génomiques ou d'ADNc à amorçage aléatoire. Plus de faux positifs ?	Peu aisée jusqu'à l'obtention de signaux positifs
Collections formatées	Recombinaison homologue. Risques d'erreurs dues aux réactions de PCR	Parfaite	Connue	Plus de faux négatifs avec des ORF complètes	Quasi intégrale depuis la construction des vecteurs jusqu'à la saisie des résultats

double-hybride à l'échelle génomique semble bien d'élaborer une heuristique performante qui permettrait d'attribuer un indice de confiance à chaque interaction observée. Une telle heuristique pourra faire largement appel aux développements les plus sophistiqués de la technologie de l'information voire de l'intelligence artificielle, puisqu'elle devra intégrer et pondérer de nombreux paramètres dont voici un bref aperçu. L'utilisation de la collection formatée du laboratoire Brent a révélé que près de 10% des protéines exprimées en appâts activent spontanément la transcription (figure 1). Une approche de génomique fonctionnelle n'étant pas acceptable si elle exclut 10% d'un protéome, un bon compromis consisterait à n'exclure que les activateurs les plus forts et à considérer avec beaucoup de circonspection les activateurs plus faibles produisant des phénotypes d'interaction plus marqués en présence de certaines proies. Des analyses statistiques devraient permettre d'identifier les appâts et les proies suspects s'ils sont trop souvent impliqués dans un phénotype d'interaction, et il est donc prévisible que l'accumulation progressive de grands volumes de résultats révèle l'identité de ces protéines suspectes avec un indice de confiance croissant. Par exemple, certaines proies produisent un phénotype d'interaction plus marqué avec la grande majorité des appâts activant spontanément la transcription, et ces signaux peuvent

alors être écartés sans trop d'états d'âme. L'heuristique définie devra également incorporer des données issues d'autres approches de génomique fonctionnelle, telles que les analyses globales de l'expression génique [10] ou, mieux encore, de l'expression protéique [11]. Ainsi, un signal positif observé avec un couple de protéines dont la co-expression n'est confirmée par aucune de ces analyses pourra être considéré comme suspect, mais pourra parfois suggérer l'existence d'une interaction réelle entre des protéines très similaires. L'analyse comparée des phénotypes d'interaction obtenus entre une protéine et les différents membres d'une famille protéique doit être systématiquement menée, et l'observation d'interactions différentielles permettra souvent d'attribuer un bon indice de confiance aux résultats positifs. Par ailleurs, il est particulièrement utile de déterminer chaque fois que possible si les interactions observées avec les protéines « sauvages » sont également observées avec certains variants alléliques à l'origine de phénotypes bien caractérisés [2, 12]. La génétique et l'analyse des interactions protéiques constituent en effet deux approches naturellement complémentaires, et un excellent indice de confiance peut généralement être attribué aux interactions qui ne sont plus observées notamment dans le cas de certaines mutations de type « perte de fonction ». Afin de tenter de valider

toutes les interactions observées par criblages « double-hybride » du protéome du ver *Caenorhabditis elegans*, Marc Vidal propose d'isoler pour chaque signal obtenu des allèles mutants perdant leur capacité d'interaction, en utilisant une approche de mutagenèse aléatoire et des gènes rapporteurs contre-sélectionnables*. Dans cette approche, ces allèles mutants sont introduits dans des vers portant une mutation de type « perte de fonction » sur le gène correspondant, et une éventuelle absence de complémentation permet alors de valider le signal « double-hybride » observé [4]. En disposant des indices de confiance obtenus grâce à cette analyse heuristique, le scientifique pourra s'engager résolument dans l'étude approfondie des interactions impliquant les protéines auxquelles il s'intéresse.

Un autre problème important concerne les résultats faux négatifs dont la fréquence est difficile à estimer. Différentes raisons bien connues des utilisateurs de la technique du double-hybride peuvent expliquer les faux négatifs. Il apparaît que les interactions impliquant des fragments de protéines produits de façon aléatoire lors de la construction des banques génomiques sont parfois détectées

* Ces gènes rapporteurs codent pour des protéines qui tuent les levures en présence de composés chimiques appropriés. Leur utilisation permet donc la sélection de mutations ou de molécules qui inhibent l'interaction entre l'appât et la proie.

beaucoup plus facilement que lorsque les protéines sont exprimées dans leur intégralité. L'utilisation systématique de domaines protéiques pourrait ainsi diminuer grandement la fréquence des signaux faussement négatifs. Toutefois, cette stratégie semble plus difficile à appliquer à des analyses exhaustives de grands protéomes.

Promesses et limitations

Les expériences de double-hybride à l'échelle génomique mettront en évidence un très grand nombre d'interactions entre protéines, et apporteront ainsi des informations fort utiles sur la fonction de la grande majorité des protéines impliquées dans ces interactions [2]. De nouveaux mécanismes moléculaires de régulation, ou tout au moins de nouveaux liens entre des mécanismes connus seront révélés. En outre, l'analyse combinée des séquences protéiques et des interactions différentielles impliquant soit des protéines d'une même famille fonctionnelle soit différentes formes alléliques d'une même protéine permettront de cerner les bases structurales de nombreuses interactions. Par ailleurs, l'élaboration de grandes cartes d'interactions protéiques devrait permettre d'inventorier de nombreux profils d'interactions, puis de rechercher exhaustivement les groupes de protéines présentant un profil d'interactions donné. Au-delà de la simple observation des interactions binaires, ce type d'analyse pourrait se révéler utile dans la prédiction de la fonction régulatrice d'un groupe de protéines qui interagissent (figure 3). La réalisation de l'ensemble de ces promesses dépendra étroitement du développement de banques de données aussi performantes que possible, disposant des liens nécessaires avec d'autres banques, et présentant des interfaces graphiques conviviales permettant une visualisation optimale des résultats (voir pour exemple <http://www.molsci.org/JeffsDatabasePages/front.html>). Les bénéfices attendus sont donc immenses, aussi bien pour la recherche académique que pour la recherche industrielle qui s'intéresse vivement aux criblages « double-hybride » à grande échelle avec l'espoir d'identifier de nouvelles cibles thérapeutiques. En

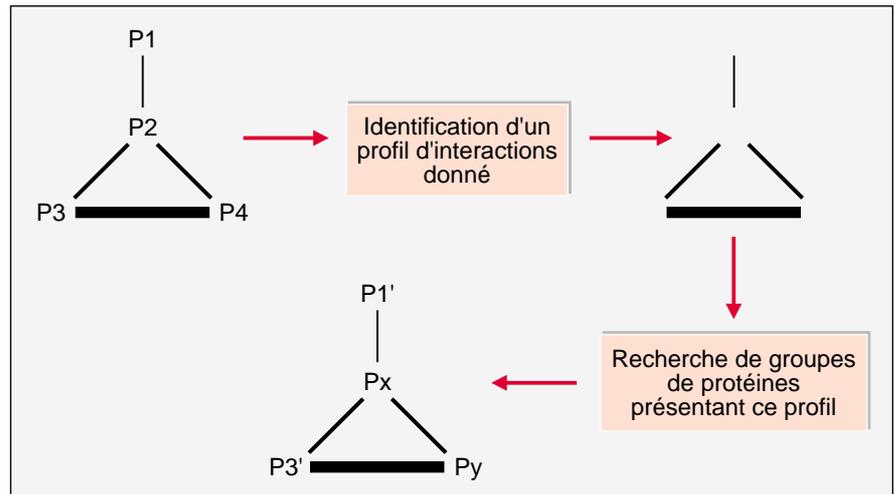


Figure 3. **Identification et recherche de profils d'interactions dans une banque de donnée.** Larry Lok a développé un programme permettant de rechercher les groupes de protéines présentant un profil d'interactions défini (<http://genome.molsci.org/~larrylok/CtdFormServer/CtdFormServer.html>). Dans l'exemple proposé, 4 protéines P1, 2, 3, 4, exerçant des fonctions connues, interagissent avec des affinités proportionnelles à l'épaisseur des traits. Le programme permet d'identifier un autre groupe de 4 protéines présentant un profil d'interactions identique. Si P1' et P3' appartiennent respectivement aux mêmes familles fonctionnelles que P1 et P3, et si les fonctions de Px et Py sont inconnues, la similarité du profil d'interactions pourrait alors fournir un indice sur les fonctions de Px et Py.

outre, il est important de souligner que le coût d'une exploration exhaustive des interactions au sein du protéome humain, paramètre essentiel de la génomique fonctionnelle, est estimé à 50 millions de dollars, somme très modique en regard du milliard de dollars que pourrait coûter un programme d'inactivation systématique de tous les gènes murins! [13]. Cependant, il convient de ne pas oublier les nombreuses limitations de la technique du double-hybride qui, appliquée à l'échelle génomique, laissera dans l'ombre un grand nombre d'interactions importantes. Ainsi, les interactions dépendant de modifications post-traductionnelles ne se produisant pas chez la levure ne pourront pas être détectées. Il est certes possible d'exprimer, par exemple, des tyrosine kinases afin de phosphoryler l'appât et de détecter des interactions impliquant des résidus phospho-tyrosine [14], mais la mise en œuvre systématique d'un tel artifice est peu réaliste. En outre, les interactions impliquant des protéines activant fortement la transcription de façon spontanée, ou des protéines ne

puvant être localisées dans le noyau ne pourront pas être détectées. D'autres systèmes « double-hybride » ont été développés, dans la levure ou dans d'autres cellules, dans lesquels une interaction protéique induit la reconstitution d'une activité biochimique facilement sélectionnable [3, 4, 15]. L'expérience collective concernant ces différents systèmes est à ce jour insuffisante pour évaluer leur applicabilité à l'échelle génomique. L'une des limitations les plus fondamentales des diverses techniques de double-hybride réside dans le fait que seules les interactions binaires sont analysées. Il est certes possible de co-exprimer une troisième protéine sous forme native [16], ou sous forme d'un deuxième appât [17] et d'appréhender ainsi des interactions d'ordre supérieur, mais l'applicabilité de ces systèmes plus complexes à l'échelle génomique est là encore improbable. Toutefois, il faut noter que la mise en œuvre de criblages « double-hybride » qui utilisent en parallèle des appâts constitués de différents membres d'un complexe protéique et une banque de proies exprimant des fragments aléa-

toires de protéines peut permettre dans certains cas de cartographier les différentes régions d'interaction et de modéliser l'architecture du complexe multiprotéique [18]

Autres méthodes d'exploration

La spectrométrie de masse appliquée aux protéines connaît depuis peu des progrès spectaculaires qui laissent présager une utilisation croissante de cette technique dans l'identification des interactions protéiques [19]. La connaissance complète du génome d'un organisme permet l'identification par MALDI-TOF (*matrix-assisted laser desorption/ionization time-of-flight*) de protéines ou parfois des produits de leur digestion protéolytique, sur la base de leur masse. Ainsi, la spectrométrie de masse a-t-elle permis d'élucider la composition de certains complexes protéiques tels que l'épissagéome humain [20] ou l'APC (*anaphase promoting complex*) de levure [21]. L'énorme avantage de la spectrométrie de masse sur les techniques de double-hybride est qu'elle permet d'identifier les complexes protéiques d'ordre supérieur tels qu'ils peuvent exister dans la cellule. Son inconvénient majeur réside dans la nécessité de purifier ces complexes par différentes méthodes physiques auxquelles les interactions les plus faibles ne résistent pas toujours. De plus, contrairement aux techniques de double-hybride, cette approche ne permet pas d'appréhender les interactions protéiques individuelles qui président à la formation du complexe, et ne donne aucune indication sur les affinités de ces interactions. Il apparaît donc que les techniques de double-hybride et la spectrométrie de masse constituent des approches plus complémentaires que concurrentes dans l'exploration exhaustive des interactions protéiques.

Un nombre croissant de techniques élégantes et sophistiquées sont constamment mises au point afin d'étudier les interactions protéiques [15]. La plupart de ces techniques semblent aujourd'hui trop laborieuses pour être appliquées à grande échelle, mais nul ne saurait affirmer qu'aucune d'entre elles n'intéressera la génomique fonctionnelle du XXI^e siècle. Très récemment, la mise

au point de puces à ADN double brin permettant de détecter et de quantifier les interactions entre ADN et protéines a introduit de façon magistrale une perspective excitante d'exploration de ces interactions à l'échelle génomique [22]. Cette avancée majeure préfigure peut-être la mise au point plus lointaine de puces à plus court terme l'élaboration de puces réunissant à l'ensemble des protéines capables de lier l'ADN et permettant d'étudier les interactions entre ces protéines et leurs partenaires.

Enfin, il est probable que la connaissance complète des génomes ainsi que les progrès continus de l'analyse informatique permettent de développer des méthodes de prédiction d'interaction protéique *in silico*, comme l'ont illustré récemment les travaux dirigés par David Eisenberg [23]. L'approche repose sur le fait que des paires de protéines interactives sont parfois réunies au sein d'une même chaîne polypeptidique dans un autre organisme. Des recherches d'homologie de séquence entre les 4 290 protéines d'*E. coli* et des protéines d'autres organismes ont révélé 6 809 interactions potentielles chez cette bactérie, dont certaines ont déjà été démontrées expérimentalement. Les auteurs ont conçu trois tests d'analyse informatique afin d'attribuer des indices de confiance à leurs prédictions, mais d'autres filtres devront être développés. Bien que cette méthode puisse en effet prédire des interactions physiques entre deux protéines, il semble qu'elle soit en fait plus efficace dans la prédiction d'interactions fonctionnelles, ce qui ne la rend pas inintéressante pour autant...

Conclusions

L'identification exhaustive des interactions protéiques à l'aide de criblages «double-hybride» ou d'autres méthodes semble constituer un objectif réalisable à moyen terme. Toutefois, l'analyse et l'exploitation de l'immense volume de résultats attendus constituent un grand défi lancé à la technologie de l'information. Les banques de données d'interactions protéiques promettent de fournir des indices utiles, parfois

déterminants, sur la fonction d'un très grand nombre de protéines en dressant «l'anatomie» des réseaux moléculaires de régulation. Si cette connaissance anatomique est essentielle, le but ultime de la biologie moderne est bien d'accéder à la «physiologie» des mécanismes régulateurs, d'appréhender leur fonctionnement dynamique dans le contexte intégré et complexe de la cellule et de l'organisme sain et malade [24]. Certaines approches originales développées récemment et utilisant directement les découvertes d'interactions protéiques, permettent de manipuler ces interactions et d'accéder ainsi à un haut niveau de résolution dans l'étude des mécanismes de régulation [25, 26]. De telles approches ouvrent la voie vers une exploitation imaginative des résultats promis par la génomique fonctionnelle, et sera sans doute l'un des défis les plus difficiles que la biologie du siècle prochain devra relever ■

Remerciements

Je remercie vivement le Dr Pierre Legrain et le Dr Philippe Noirot de leurs commentaires éclairés sur le manuscrit et des informations qu'ils m'ont communiqué. Je suis très reconnaissant au Dr Roger Brent et à mes anciens collègues au sein de son laboratoire pour tous les moments scientifiques excitants que nous avons partagés, notamment au cours de nos nombreuses explorations d'interactions protéiques.

RÉFÉRENCES

1. Fields S, Song O. A novel genetic system to detect protein-protein interactions. *Nature* 1989; 340: 245-6.
2. Brent R, Finley RLJ. Understanding gene and allele function with two-hybrid methods. *Annu Rev Genet* 1997; 31: 663-704.
3. Colas P, Brent R. The impact of two-hybrid and related methods on biotechnology. *Trends Biotechnol* 1998; 16: 355-63.
4. Vidal M, Legrain P. Yeast forward and reverse 'n'-hybrid systems. *Nucleic Acids Res* 1999; 27: 919-29.
5. Bartel PL, Roecklein JA, SenGupta D, Fields S. A protein linkage map of *Escherichia coli* bacteriophage T7. *Nat Genet* 1996; 12: 72-7.
6. Fromont-Racine M, Rain JC, Legrain P. Towards a functional analysis of the yeast genome through exhaustive two-hybrid screens. *Nat Genet* 1997; 16: 277-81.

RÉFÉRENCES

7. Thomas BJ, Dong X, Zavitz K, *et al.* Roush downregulates G2 cyclin in G1. *Genes Dev* 1997; 11: 1289-98.
8. Hudson JR Jr, Dawson EP, Rushing KL, *et al.* The complete set of predicted genes from *Saccharomyces cerevisiae* in a readily usable form. *Genome Res* 1997; 7: 1169-73.
9. Hua SB, Luo Y, Qiu M, Chan E, Zhou H, Zhu L. Construction of a modular yeast two-hybrid cDNA library from human EST clones for the human genome protein linkage map. *Gene* 1998; 215: 143-52.
10. Brent R. Functional genomics: learning to think about gene expression data. *Curr Biol* 1999; 9: R338-41.
11. Joubert-Caron R, Caron M. Protéome et analyse protéomique: de nouveaux concepts pour de nouveaux champs d'applications biomédicales. *Med Sci* 1999; 15: 701-5.
12. Reymond A, Brent R. p16 proteins from melanoma-prone families are deficient in binding to Cdk4. *Oncogene* 1995; 11: 1173-8.
13. Fields S. The future is function. *Nat Genet* 1997; 15: 325-7.
14. Osborne, MA, Lubinus, M, Kochan, JP. Detection of protein-protein interactions dependent on post-translational modifications. In: Bartel PL, Fields S, eds. *The yeast two-hybrid system*. New York, Oxford: Oxford University Press, 1997: 233-58.
15. Mendelsohn AR, Brent R. Protein interaction methods-toward an endgame. *Science* 1999; 284: 1948-50.
16. Tirode F, Malaguti C, Romero F, Attar R, Camonis J, Egly JM. A conditionally expressed third partner stabilizes or prevents the formation of a transcriptional activator in a three-hybrid system. *J Biol Chem* 1997; 272: 22995-9.
17. Xu CW, Mendelsohn AR, Brent R. Cells that register logical relationships among proteins. *Proc Natl Acad Sci USA* 1997; 94: 12473-8.
18. Flores A, Briand JF, Gadal O, *et al.* A protein-protein interaction map of yeast RNA polymerase III. *Proc Natl Acad Sci USA* 1999; 96: 7815-20.
19. Costello CE. Bioanalytic applications of mass spectrometry. *Curr Opin Biotechnol* 1999; 10: 22-8.
20. Neubauer G, King A, Rappsilber J, *et al.* Mass spectrometry and EST-database searching allows characterization of the multi-protein spliceosome complex. *Nat Genet* 1998; 20: 46-50.
21. Zachariae W, Shevchenko A, Andrews PD, *et al.* Mass spectrometric analysis of the anaphase-promoting complex from yeast: identification of a subunit related to cullins. *Science* 1998; 279: 1216-9.
22. Bulyk ML, Gentalen E, Lockhart DJ, Church GM. Quantifying DNA-protein interactions by double-stranded DNA arrays. *Nat Biotechnol* 1999; 17: 573-7.
23. Marcotte EM, Pellegrini M, Ng H-L, Rice DW, Yeates TO, Eisenberg D. Detecting protein function and protein-protein interactions from genome sequences. *Science* 1999; 285: 751-3.
24. Colas P. Some challenges and promises of the 21st century biology. *J Mol Biol Biotech* 1999; 1: 1-3. (<http://www.jmbab.com/a1colas.htm>).
25. Colas P. Nouvelles approches dans l'étude des mécanismes moléculaires de régulation. *Hématologie* 2000 (sous presse).
26. Colas P. Combinatorial protein reagents to manipulate protein function. *Curr Opin Chem Biol* 2000; 4 (sous presse).

TIRÉS À PART

P. Colas.

Summary

Genome-scale exploration of protein interactions

Genome-scale exploration of protein interactions is one of the most potent approaches of functional genomics. Pioneering efforts have demonstrated that the two-hybrid technique could support such an endeavor. However, the problem of false positive signals remains and needs to be carefully tackled by designing a heuristic to attribute a confidence level for each interaction detected. The expected benefits of building huge databases of protein interactions include insights into the function of many novel proteins, discovery of regulatory pathways or cross-talks between known pathways, structural insights into proteins interactions and possibly analysis of protein interaction patterns. Nevertheless, this approach presents several caveats. Regular two-hybrid assays do not allow the detection of interactions depending on post-translational modifications. Furthermore, information on multi-order protein complexes cannot be accessed easily. In the near future, some of these limitations should be alleviated by a broader use of other methodologies such as mass spectrometry and micro-arrays that may be applied to genome-scale analysis.

GERDA

Groupe d'Études et de Recherches en Dermato-Allergologie
PARIS – Palais des Congrès – 5 au 7 octobre 2000

Dermato-allergo-pédiatrie – Pathologies allergiques des muqueuses

Organisation scientifique :

Dr Annik Pons-Guiraud/10, bd Maiesherbes, 75008 Paris, France.
 Tél. : 01 42 66 32 01/Fax : 01 42 66 32 13 – E-mail : Annick.Pons-guiraud@wanadoo.fr

Organisation technique :

MELTHEM/95, rue de Lourmel, 75015 Paris, France.
 Tél. : 01 44 26 16 48/Fax : 01 45 54 36 21 – E-mail : melthem@club-internet.fr