

# Assemblage des complexes ribonucléoprotéiques

Nicolas Hugo, Philippe Bouvet



► Les interactions ARN-protéines interviennent dans de nombreuses fonctions. Ces interactions se font le plus souvent au sein d'assemblages macromoléculaires complexes. Des développements méthodologiques récents ont permis des avancées importantes dans la compréhension de la spécificité de ces interactions. L'objectif de cet article est d'exposer les différentes techniques d'étude des complexes ARN-protéines ainsi que les dernières données concernant l'utilisation thérapeutique des ARN. ◀



École Normale Supérieure de Lyon, Cnrs UMR 5665, 46, allée d'Italie, 69364 Lyon Cedex 07, France. pbouvet@ens-lyon.fr

Pendant de nombreuses années, les ARN ont été considérés comme de simples intermédiaires entre l'ADN qui contient les gènes (l'information génétique) et les protéines. On sait maintenant que les ARN possèdent de nombreuses autres fonctions et interviennent à différents niveaux de la régulation de l'expression génique. Ces fonctions dépendent de l'établissement d'interactions ARN-protéines, qui gouvernent la formation des complexes ribonucléoprotéiques (complexes RNP). Ces complexes très dynamiques jouent un rôle majeur dans la transcription, dans les mécanismes de régulation post-transcriptionnelle (épissage, stabilité des ARN, routage et compartimentation des ARN), ainsi que lors de la traduction. En outre, la mise en évidence récente de différentes classes de petits ARN, ARNs (small nuclear RNA), ARNsno (small nucleolar RNA), microARN, ouvre de plus la possibilité de nouvelles fonctions encore insoupçonnées. On sait maintenant qu'un dérèglement de l'expression ou de la maturation de certains ARN peut être à l'origine de certaines maladies, puisque environ 10 % des mutations impliquées dans des maladies génétiques affectent différents aspects du métabolisme des ARN. Grâce à une meilleure connaissance des mécanismes de régulation de l'expression génique, les ARN peuvent aussi être considérés comme des cibles et des outils thérapeutiques permettant de moduler l'expression

génique (aptamères, ribozymes, interférence ARN...).

Cet article a pour but de présenter les principales méthodologies d'étude ainsi que les dernières techniques qui peuvent être utilisées pour élucider les mécanismes de reconnaissance ARN-protéines. Plusieurs exemples d'utilisation des ARN en tant qu'outils thérapeutiques seront aussi exposés (→). Les principaux domaines protéiques d'interaction avec les ARN sont présentés dans le *Tableau 1*. Une revue contenant une liste exhaustive de ces domaines a été récemment publiée [1].

## Étude des interactions ARN-protéines

### Identification des séquences d'ARN liées aux protéines

Une des premières étapes dans l'étude de l'assemblage de complexes ribonucléoprotéiques consiste à identifier les différents composants de ce complexe. Bien souvent, on ne connaît que partiellement les partenaires. Par exemple, on peut avoir identifié une séquence importante pour la régulation de la stabilité d'un ARN, mais on ne connaît pas la protéine qui reconnaît spécifiquement cette séquence, ou

(→) m/s 1999, n° 5, p. 677

bien, on connaît la protéine, mais pas la séquence d'ARN qui reconnaît cette protéine. Dans ce dernier cas, la technique SELEX (*systematic evolution of ligands by exponential enrichment*) [2] permet d'isoler des séquences d'ARN qui interagissent avec une molécule « appât ». On réalise une banque contenant des séquences provenant soit d'ARN cellulaires (SELEX génomique) [3], soit de fragments d'un ARN naturel (SERF: *selection of random RNA fragment*) [4], soit d'ADN synthétiques codant pour une région d'ARN aléatoire de longueur variable que l'on crible contre la protéine « appât ». Les séquences retenues sont isolées puis amplifiées par PCR après transcription inverse. Plusieurs étapes d'isolement et d'amplification en condition de stringence croissante permettent d'obtenir les séquences d'ARN les plus affines pour la protéine « appât ». Les ARN

sélectionnés sont ensuite séquencés et alignés pour isoler la (ou les) séquence(s)/structure(s) consensus reconnue(s) par la protéine « appât ». Cette technique peut être facilement automatisée pour effectuer un criblage à haut débit [5].

Le plus souvent, une protéine recombinante est utilisée pour effectuer cette sélection. Néanmoins, si la production de la protéine recombinante est impossible et que la purification de la protéine native peut être effectuée, on peut utiliser la protéine native comme appât. C'est ce que nous avons fait pour réaliser le SELEX avec la nucléoline, une protéine présente dans le nucléole et qui présente quatre domaines de liaison aux ARN du type RRM (*RNA recognition motif*) [6]. Cette étude a permis d'isoler deux séquences d'ARN reconnues par la nucléoline. L'une de ces séquences

DOMAINES	CARACTÉRISTIQUES	STRUCTURES	PROTÉINES TYPES	ARN CIBLES	REMARQUES	RÉFÉRENCES
RRM ( <i>RNA recognition motif</i> ) RBD ( <i>RNA binding domain</i> )	80 à 90 acides aminés 2 motifs très conservés: RNP1 (R/K-G-F/Y-G/A-F/Y-V-X-F/Y) et le motif RNP-2 (L/I-FY-V/I-G/K-G/N-L)	4 feuillets $\beta$ + 2 hélices $\alpha$ Topologie: $\beta\alpha\beta\beta\alpha\beta$	Protéines du complexe d'épissage (U1A et sxl, 2 RRM), <i>poly(A) binding protein</i> impliquée dans la polyadénylation des ARNm; nucléoline (4 RRM, protéine jouant un rôle dans la biogenèse des ribosomes)	Simple brin, tiges boucles, boucles internes	Domaine de liaison aux ARN le plus représenté. Il est retrouvé dans des protéines qui interviennent dans toutes les étapes de régulation post-transcriptionnelles. 1 à 5 domaines par protéine	[43-45]
KH ( <i>hnRNP K homologue</i> )	60 à 70 acides aminés 1 motif très conservé: V-I-G-X-X-G-X-X-I	Topologie de type $\beta\alpha\beta\beta\alpha$	hnRNP K, Nova, facteur d'épissage SF1/BBP, protéine ribosomique S3, Sam 68, FMRP...	Simple brin	Présent dans plus d'une centaine de protéines. Le domaine KH peut dimériser 1 à 15 domaines par protéine	[46-48]
RGG ( <i>arginine and glycine rich domain</i> ) encore appelé GAR	Domaines riches en arginine, glycine et acides aminés aromatiques. Taille du domaine assez variable, plusieurs répétitions du motif RGGF/Y	Structure non déterminée, mais des études spectroscopiques mettent en évidence des spirales $\beta$	hnRNP U, hnRNP A1, protéines nucléolaires impliquées dans la biogenèse des ribosomes: GAR-1, fibrillarine, nucléoline, protéine FMR...	Structures en <i>G-quartet</i> , Spécificité d'interaction avec les acides nucléiques encore mal connue	Présent dans des protéines qui possèdent d'autres domaines d'interaction avec les ARN. Cible de modification post-traductionnelle (méthylation des arginines). C'est aussi un domaine d'interaction protéine-protéine	[49-52]
Ds-RBD ( <i>double strand RNA binding domain</i> )	Environ 70 acides aminés	Topologie de type $\alpha\beta\beta\beta\alpha$	RNase III, protéine Dicer, Staufen...	ARN partiellement ou complètement double brins	C'est aussi un domaine d'interaction protéine-protéine. 1 à 5 domaines par protéine	[53-55]
ARM ( <i>arginine rich motif</i> )	10 à 30 acides aminés riches en arginine	Les motifs ARM peuvent adopter différentes structures (hélice $\alpha$ , tige boucle $\beta$ )	Rev, Tat, $\lambda$ N peptide...	Sillon majeur d'un ARN double brin	De nombreux appariements non canoniques sont induits par l'interaction du peptide ARM	[56-58]

Tableau I. Principaux domaines d'interaction avec les ARN.

appelée NRE (*nucleolin recognition element*) est retrouvée sur l'ensemble de la séquence de l'ARNr précurseur [7] tandis que l'autre, appelée ECM (*evolutionary conserved motif*), est présente uniquement dans la région 5' externe de l'ARN pré-ribosomique [8]. L'analyse par délétion de protéines recombinantes a permis de montrer que les deux premiers domaines RRM de la protéine (RRM 1-2) étaient requis pour l'interaction avec la séquence NRE. Un nouveau SELEX effectué avec la protéine recombinante RRM1-2 a permis d'obtenir la même séquence consensus que celui réalisé avec la protéine entière.

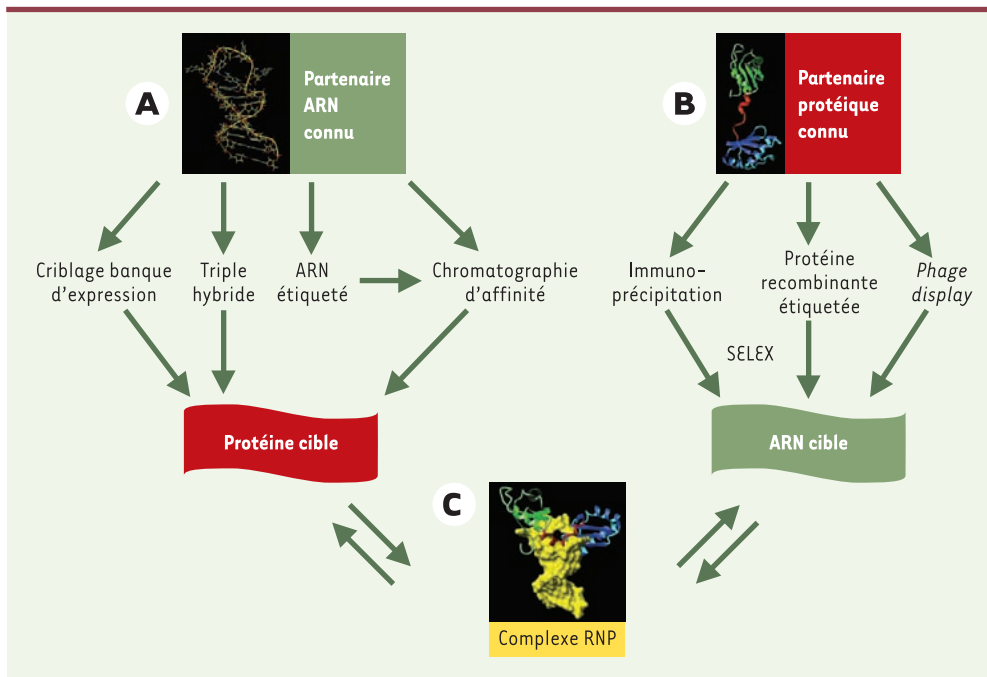
Lorsque la séquence d'ARN reconnue par la protéine d'intérêt est déjà connue, une partie de cette séquence peut être rendue aléatoire et utilisée comme banque de départ pour sélectionner les séquences les plus affines pour la protéine. Cette stratégie a été utilisée notamment pour la protéine du complexe d'épissage U1A qui s'associe avec une séquence tige-boucle du petit ARN U1. Une mutagenèse aléatoire a été réalisée sur la boucle de cet ARN de

mettre en évidence les nucléotides importants pour conserver l'interaction avec la protéine U1A [9]. Ces études sont complémentaires des analyses phylogénétiques pour prédire les nucléotides importants dans l'interaction.

La vérification de l'efficacité de la sélection passe souvent par la mesure de l'affinité de la protéine pour la séquence sélectionnée. Le plus souvent, des expériences de retard sur gel sont réalisées afin de déterminer les constantes d'association et de dissociation. La technique de résonance plasmonique de surface (RPS), qui permet l'analyse des interactions ARN-protéines en temps réel, peut aussi être utilisée pour déterminer les paramètres cinétiques des interactions [10]. Grâce à la grande sensibilité de cette méthode, des interactions de faible affinité peuvent être caractérisées.

La technique SELEX permet de sélectionner des ARN qui ont une grande affinité pour la cible protéique mais qui ne correspondent pas forcément à des séquences naturelles. En effet, un des problèmes majeurs de cette technique est que la sélection se fait complètement *in vitro* et il faut

ensuite démontrer que les cibles ARN isolées à partir d'une banque d'ARN aléatoires synthétiques ont une réalité biologique. Des expériences complémentaires sont donc nécessaires pour valider les cibles sélectionnées. Dans l'exemple de la nucléoline, on savait que la protéine interagissait avec l'ARN pré-ribosomique naissant. Une recherche par homologie de séquences et différentes délétions ont permis de valider les séquences SELEX [6-8]. Un autre exemple d'approches complémentaires pour identifier les cibles ARN reconnues par une protéine concerne la protéine FMR (*fragile X mental retardation*) qui est associée au syndrome de fragilité lié au chromosome X, forme commune de retard mental, dont la fréquence est de 1/4000 chez les hommes et 1/8000 chez les femmes. La protéine FMR possède deux domaines KH (*K homology*



**Figure 1. Stratégies d'études des complexes ribonucléoprotéiques.** Le plus souvent, un des deux partenaires du complexe est connu (l'ARN ou la protéine). **A.** La structure de l'ARN est spécifiquement reconnue par les deux premiers domaines de liaison à l'ARN de la nucléoline. **B.** Structure des deux premiers domaines de liaison de la nucléoline à l'ARN. **C.** Structure du complexe formé par ces deux domaines avec leur cible ARN. Si la séquence d'ARN est identifiée, on peut alors utiliser différentes techniques (chromatographie d'affinité, triple hybride, criblage d'une banque d'expression) pour identifier la protéine cible. Si la protéine du complexe ribonucléoprotéique (RNP) est connue, des expériences d'immunoprécipitation peuvent fournir des sondes pour cribler des micropuces portant des séquences de différents gènes ou être utilisées dans des expériences de SELEX (*systematic evolution of ligands by exponential enrichment*) afin d'identifier les séquences ARN cibles. Les complexes RNP peuvent aussi être purifiés par des techniques biochimiques, puis les protéines du complexe peuvent être identifiées grâce aux techniques de protéomique.

domain) et un domaine RGG (*arginine and glycine rich domain*). Une expérience de SELEX, réalisée en utilisant une protéine FMR recombinante comme appât pour cribler une banque d'ARN de 96 nucléotides (dont 52 résidus centraux aléatoires), a permis de sélectionner six séquences d'ARN différentes comportant toutes des structures appelées *G-quartet* [11]. L'interrogation de bases de données a ensuite permis de caractériser 31 ARNm susceptibles, sur la base de leur homologie avec la séquence consensus, de se lier à la protéine FMR. Sur douze de ces séquences testées pour leur interaction avec FMR, seulement six avaient une affinité significative pour la protéine.

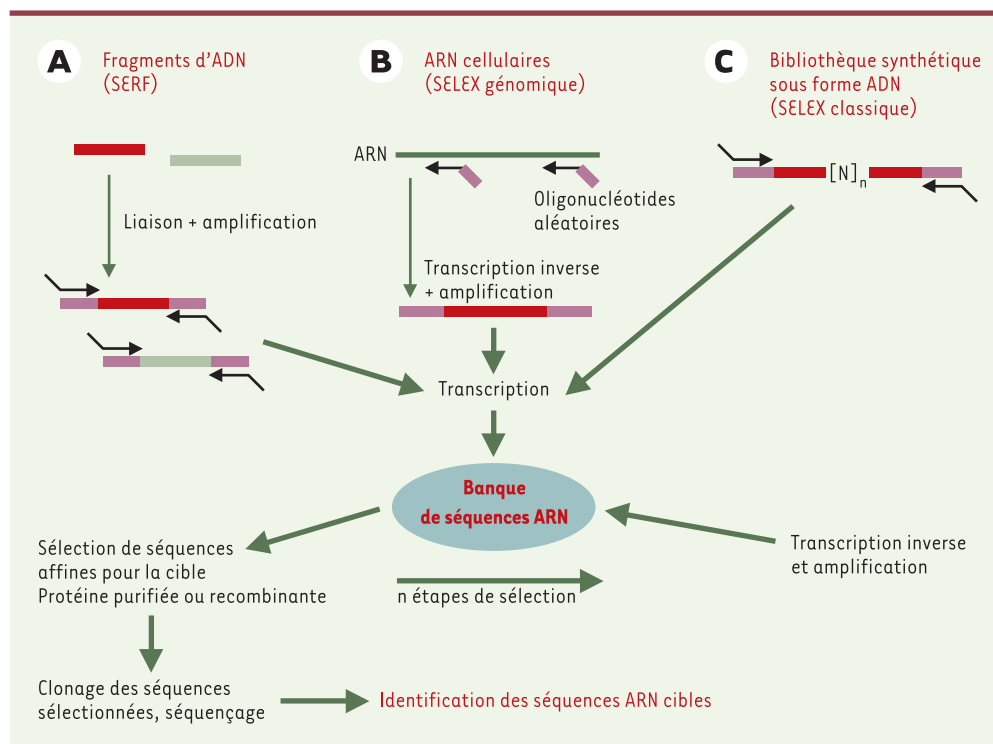
La ribonomique [12] a également été utilisée pour identifier les cibles cellulaires de la protéine FMR. Cette technique permet l'identification et l'analyse d'une fraction d'ARN associés fonctionnellement ou structurellement. À partir d'expériences d'immunoprécipitation de complexes ribonucléoprotéiques associés à la protéine FMR, les ARN présents dans ces complexes ont été isolés et ont servi de sondes pour cribler plus de 25 000 séquences présentes sur des puces à ADN [13]. Le criblage de ces puces avec les ARN immunoprécipités par la protéine FMR a permis l'identification de 432 gènes dont l'ARN est présent dans les ribonucléoparticules contenant la FMR. Dans une autre expérience, 251 ARNm montrent un profil de polyribosomes différent en l'absence de FMR, suggérant qu'un dérèglement de la traduction de ces ARNm peut être l'une des causes du syndrome de fragilité lié au chromosome X. Il est aussi remarquable que près de 70 % de ces ARN possèdent des séquences qui permettent la formation de structures de type *G-quartet*, confirmant ainsi les résultats de l'expérience de SELEX réalisée avec la protéine FMR. La protéine FMR interagit aussi avec son propre ARN dans une région de 100 nucléotides également susceptible de former une structure en *G-*

*quartet* [14]. La comparaison des ARNm, isolés par ces différentes approches, a permis de sélectionner huit ARNm cibles, dont le dérèglement pourrait être à l'origine des retards mentaux observés chez les patients atteints du syndrome de l'X fragile.

Les ARN purifiés à partir d'immunoprécipitation de complexes ribonucléoprotéiques issus soit de deux types cellulaires, soit d'un même type cellulaire soumis à différentes conditions, peuvent aussi servir à la réalisation d'ADN complémentaires. Ces derniers pourront ensuite être utilisés dans les techniques classiques de banques soustractives et d'analyse d'expression différentielle (*differential display*) afin d'identifier les ARN immunoprécipités et d'analyser la dynamique de leur association au sein des complexes ribonucléoprotéiques [15].

### Identification des protéines liées aux ARN

Au cours des études sur la régulation post-transcriptionnelle, on commence bien souvent par déterminer la



**Figure 2. Identification des séquences ARN reconnues par une protéine cible.** La technique du SELEX (*systematic evolution of ligands by exponential enrichment*) a été initialement développée avec une banque d'ARN synthétisée à partir d'oligonucléotides aléatoires (C). L'inconvénient de cette approche est d'isoler des cibles ARN non naturelles. Cette technique a été adaptée pour isoler directement des séquences naturelles. Le SELEX génomique (B) permet la construction d'une bibliothèque d'ARN cellulaire (transcription inverse avec des oligonucléotides de séquences aléatoires) alors que la technique de SERF (*selection of random RNA fragment*) (A) permet de transcrire des séquences d'ARN à partir de fragments d'ADN d'intérêt. Dans les trois cas (A, B, C), une fois la bibliothèque d'ARN synthétisée, la sélection peut se faire avec la protéine cible par les approches classiques (immunoprécipitation, retard sur gel, rétention des complexes sur filtres, etc.). Lorsque les ARN cibles auront été sélectionnés et identifiés, il restera à démontrer qu'ils ont une signification biologique.

séquence d'ARN qui est importante pour la régulation. On recherche ensuite les facteurs protéiques qui interagissent spécifiquement avec ces séquences. Grâce au développement de systèmes de transcription *in vitro* et à la synthèse chimique d'ARN modifiés (rendus par exemple résistants aux ribonucléases par incorporation de 2'-O-méthyl ribonucléotides), on peut très facilement synthétiser de grandes quantités d'ARN. Ces ARN peuvent alors être utilisés pour réaliser des expériences de chromatographie d'affinité pour les protéines. La séquence d'ARN cible peut aussi être fusionnée à une autre séquence qui représente un site d'interaction pour une protéine connue. On peut citer l'exemple du site de reconnaissance de la protéine d'enveloppe du phage MS2 ou celui d'un aptamère de la streptomycine (*StreptoTag*) [16]. L'ARN hybride peut être alors soit exprimé *in vivo* pour récupérer les complexes nucléoprotéiques *via* une chromatographie d'affinité comportant la protéine MS2 ou la streptomycine immobilisée, soit synthétisé *in vitro* puis incubé ensuite avec des extraits cellulaires pour récupérer la protéine qui se fixe sur la séquence cible. Il existe maintenant de nombreux exemples dans lesquels des chromatographies d'affinité ont été utilisées pour purifier soit un complexe ARN-protéine simple, soit une ribonucléoparticule plus complexe. On citera l'identification de la protéine EDEN-BP (*embryonic deadenylation element-binding protein*) qui interagit avec une courte séquence dans la région non traduite (3'UTR) de certains ARNm qui sont désadénylés au cours de la fécondation de l'ovocyte de xénope [17], et l'isolement de petites particules ribonucléoprotéiques impliquées dans l'épissage. Les interactions non spécifiques, avec la cible ARN, de nombreuses hnRNP (*heterogeneous nuclear ribonucleoproteins*) abondamment présentes dans les extraits représentent une difficulté majeure de ces expériences.

Une autre technique très élégante d'identification d'une protéine interagissant avec une séquence d'ARN est le système du triple hybride [18]. Cette technique, dérivée du système double hybride développé dans la levure, consiste à construire un ARN hybride dont une partie est reconnue par une protéine couplée à un domaine de fixation à l'ADN et la seconde partie est l'ARN d'intérêt reconnu par une protéine cible fusionnée au domaine d'activation de la transcription. Bien que peu utilisée encore, cette technique a notamment permis d'isoler la protéine HBP (*histone binding protein*) qui se lie avec une structure en tige-boucle dans la région 3'UTR des ARN des histones [19]. Cette protéine est impliquée dans la maturation de l'extrémité 3' de l'ARN. On peut noter que, dans ce dernier exemple, des techniques biochimiques de purification n'avaient pas permis d'isoler cette protéine en quantité suffisante pour pouvoir l'identifier.

Le criblage d'une banque d'expression avec une sonde ribonucléique correspondant à la séquence d'ARN d'intérêt (par la technique de *Northwestern*) devrait permettre *a priori* l'identification de la protéine qui reconnaît cette séquence. Bien que cette stratégie, qui semble assez directe, ait été utilisée avec succès pour cloner plusieurs protéines qui interagissent avec l'ADN, il n'existe que très peu d'exemples de clonage de protéines affines pour l'ARN [20]. Le problème de la renaturation des protéines et la présence de RNases sont sans doute des freins importants à l'utilisation de cette technique.

### Analyse fine des interactions

Au cours de ces dernières années, de nombreux systèmes génétiques ont vu le jour, permettant d'identifier les résidus nucléotidiques ou protéiques importants pour l'assemblage de particules ribonucléoprotéiques. Un système utilise l'activité d'anti-termination transcriptionnelle de la protéine N du bactériophage  $\lambda$  [21], et a été utilisé principalement pour l'identification de peptides qui se lient au motif riche en arginine. Un autre test est fondé sur la régulation de la traduction dans la bactérie *E. coli* [22]. Chez cette bactérie, l'initiation de la traduction s'effectue par le recrutement du ribosome sur une région localisée en général quelques nucléotides en amont du codon d'initiation, appelée séquence de Shine-Dalgarno (S/D). Le ribosome couvre alors une région d'environ 40 nucléotides. L'introduction de la cible ARN en amont de la boîte S/D ne modifie pas l'interaction du ribosome et donc l'expression du gène rapporteur en aval (gène *LacZ* par exemple). En revanche, l'expression dans *E. coli* de la protéine qui reconnaît la séquence cible sur l'ARN peut empêcher stériquement l'interaction du ribosome et donc la traduction du gène rapporteur. Grâce à ce système, et après une mutagenèse aléatoire de la protéine, des liaisons spécifiques entre la protéine Rev (*regulation of expression of viral proteins*) et sa cible RRE (*Rev response element*), ou entre les RRM 1-2 de la nucléoline et sa cible NRE [23], ont pu être déterminées. Ces liaisons ont été ensuite confirmées par des analyses structurales des deux complexes. Contrairement aux tests génétiques précédents qui ont lieu *in vivo* - donc dans un environnement complexe - la technique du *phage display* permet *in vitro* d'isoler aussi de nouvelles cibles d'ARN et de déterminer les acides aminés impliqués dans la reconnaissance spécifique et la force de l'interaction. Le principe de cette technique est de lier physiquement une protéine à sa séquence codante grâce à un phage qui expose à sa surface une protéine codée par un plasmide qu'il contient (protéine en fusion avec la protéine d'attachement du bactériophage M13, par exemple). On peut ainsi introduire diverses mutations dans la protéine à tester et réaliser une banque de bacté-

riophages codant pour la protéine de fusion. L'ARN cible est synthétisé *in vitro*, immobilisé sur une colonne portant un ARN complémentaire d'une partie de l'ARN cible, puis mis en contact avec les phages portant la ribonucléoprotéine à leur surface. Les phages portant des protéines qui reconnaissent l'ARN cible sont alors retenus. Cette technique a été affinée en utilisant le phage T7 en combinaison avec une souche d'*E. coli* déficiente en ribonucléase et utilisée pour l'identification de nouvelles ribonucléoprotéines [24].

(→) m/s  
2001, n° 5,  
p. 609

### Protéomique et analyse des complexes ribonucléoprotéiques

Le développement de nouvelles techniques de spectrométrie de masse et la mise au point d'étiquettes (Tag) permettent maintenant de déterminer la composition des particules ribonucléoprotéiques complexes (→).

Certaines particules ribonucléoprotéiques forment de très grosses structures qui peuvent être isolées par des techniques biochimiques. Il est cependant difficile d'assurer que le complexe ainsi isolé n'a pas été altéré ou contaminé au cours de l'isolement. Par exemple, l'analyse du protéome du nucléole, qui peut être considéré comme une grosse particule ribonucléoprotéique puisqu'il contient un assemblage complexe de plusieurs ARN et de nombreuses protéines, a été réalisé par deux groupes différents [25, 26]. Certes de nombreuses protéines sont communes aux deux analyses, mais près d'un tiers des protéines d'une étude n'ont pas été retrouvées dans l'autre. Ces différences pourraient être dues en partie à des techniques de purification différentes des nucléoles. Un autre exemple est donné par la petite particule ribonucléoprotéique nucléolaire U3 (snoRNP U3) impliquée dans la maturation des ARN ribosomiques dans le nucléole. La composition de cette snoRNP, identifiée il y a plus de 30 ans, a fait l'objet de nombreuses études. Mais ce n'est que très récemment [27], grâce à l'utilisation de protéines étiquetées, que la composition de cette snoRNP a pu être déterminée. Cette technique permet en général, en une seule étape, d'enrichir très fortement le complexe étudié. Afin d'améliorer cette purification, un certain nombre de double - voire de triple - étiquettes [28] ont été développées. La plus utilisée est sans doute la double étiquette Tap Tag (*tandem affinity purification*) développée par l'équipe de B. Séraphin [29]. Cette méthode semble particulièrement bien adaptée pour la purification, dans des conditions natives, de particules ribonucléoprotéiques à partir de la levure ou de lignées cellulaires transfectées de façon stable avec la protéine de fusion. Dans l'exemple de l'isolement de la ribonucléoprotéine U3 [27], deux étiquettes étaient portées par deux protéines différentes (protéines qui étaient déjà connues comme faisant partie du complexe). Par

(→) m/s  
2002, n° 3,  
p. 274

(→) m/s  
2001, n° 6-7,  
p. 771

cette technique, 28 protéines ont été identifiées dans la particule snoRNP U3, dont 17 nouvelles protéines. Trois autres protéines co-purifiées avec la particule étaient des contaminants (→).

### ARN et complexes ribonucléoprotéiques : applications thérapeutiques

Comme nous l'avons vu précédemment, les ARN sont sujets à de nombreuses régulations post-transcriptionnelles (→). La disponibilité des séquences génomiques rend possible l'identification de mutations responsables de maladies génétiques. Une compilation de ces mutations peut être consultée sur le site *The Human Gene Mutation Database* (<http://archive.uwcm.ac.uk/uwcm/mg/hgmd0.html>). Près de 60 % de ces mutations concernent les régions codantes et affectent directement les structures et les fonctions des protéines. Un nombre croissant de maladies génétiques impliquent des mutations (environ 10 % des mutations identifiées) affectant différents aspects du métabolisme des ARN [30]. On peut citer une mutation dans la structure IRE (*iron responsive element*) présente à l'extrémité 5' non traduite de l'ARNm codant pour la ferritine, protéine importante pour l'homéostasie du fer. Cette mutation empêche l'interaction de la protéine IRP (*iron responsive protein*) et provoque un dérèglement de la traduction de l'ARNm codant pour la ferritine, responsable d'un syndrome associant une hyperferritinémie et une cataracte. Un autre exemple vient d'être récemment donné dans la dystrophie myogénique de type 1. Dans cette maladie, l'hyper-excitabilité des muscles serait en rapport avec un dérèglement de l'épissage alternatif de l'ARNm CIC-1 (codant pour un canal chlore spécifique des muscles) [31]. Il n'est pas étonnant avec ces données que les ARN soient considérés de plus en plus comme des cibles pharmacologiques importantes. Une meilleure compréhension de l'assemblage et de la structure des complexes ribonucléoprotéiques permettront le développement de nouveaux cribles pour l'identification de molécules thérapeutiques [32]. Ainsi, l'interaction de certains antibiotiques avec l'ARN ribosomique procaryote est connue depuis longtemps. La résolution à l'échelle atomique de la structure de complexes antibiotique-ARN ribosomiques permet maintenant de rationaliser la mise au point de nouveaux antibiotiques [33].

L'utilisation des ARN comme molécules thérapeutiques est limitée *in vivo* par leur extrême sensibilité aux nucléases. Plusieurs types de modifications peuvent être apportées pour les protéger. Ainsi, des ARN modifiés par l'ajout de molécules de haut poids moléculaire comme le poly-éthylène glycol ou bien insérés dans des liposomes ont une durée de vie plus longue (de plusieurs heures à plusieurs



(→) m/s  
2002, n° 6-7,  
p. 665

(→) m/s  
2000, n° 12,  
p. 1430

jours) dans des fluides biologiques comme le sang. De même, des analogues nucléotidiques (remplacement du groupement 2'-hydroxyl du ribose par la fluorine) ou des molécules synthétiques de type N3'-P5' désoxyphosphoramidate ont parfois une affinité similaire aux ARN de départ mais présentent une résistance accrue aux nucléases.

Les ribozymes et les ARN antisens ont fait l'objet de nombreuses études afin de les utiliser dans la régulation de l'expression génique [34, 35]. L'interaction de l'ARN antisens avec sa cible forme un duplex qui est supposé conduire à la dégradation de l'ARN cible ou à une inhibition de sa traduction. Cependant, pour observer ces effets, l'ARN antisens doit être présent en large excès. En revanche, les ribozymes peuvent être recyclés après avoir clivé leur ARN cible. Il existe de nombreux exemples de contrôle de l'expression génique par des ribozymes. Le développement d'un nouveau type de ribozyme, le « maxizyme » (*minimized, active, X-shaped [fonctioning as a dimer], and intelligent [allosterically controllable] ribozyme*) a été utilisé pour inactiver l'ARN chimérique Bcr-Abl issu d'une translocation chromosomique dans la leucémie myéloïde chronique (LMC). Cette molécule est active *in vitro* et *in vivo*. En effet, des souris auxquelles on injecte des cellules de LMC transduites par un vecteur rétroviral exprimant le maxizyme ne sont pas affectées alors que les souris témoins ayant reçu des cellules de LMC transduites par un vecteur témoin (n'exprimant pas le maxizyme) meurent de leucémie diffuse [36]. Cela suggère une utilisation prometteuse du maxizyme pour la thérapie génique (→).

Les ARN peuvent aussi être utilisés comme molécules inhibitrices de l'activité de certaines protéines. Un exemple classique concerne l'inhibition de la réplication du génome du virus VIH dans des cellules CD4 par l'expression d'un petit ARN qui mime la structure de l'élément tige-boucle TAR de l'ARN du virus [37]. Grâce à la stratégie du SELEX décrite plus haut, il est possible d'identifier de petites molécules d'ARN, appelées aptamères, qui possèdent une forte affinité pour une protéine donnée, ceci n'étant pas exclusivement réservé aux ribonucléoprotéines. Ainsi, une recherche par SELEX d'aptamères ARN fixant fortement la séquence TAR de l'ARN a permis d'isoler plusieurs molécules qui inhibent spécifiquement l'activité transcriptionnelle dépendante de la protéine Tat *in vitro* [38]. Plusieurs ARN aptamères, inhibiteurs de l'adhérence cellulaire dépendant de la sélectine ont pu être identifiés. D'autres molécules inhibant l'interaction du trypanosome (*T. cruzi*) avec ses cellules cibles ont aussi été caractérisées [39]. Certaines d'entre elles sont capables, à la concentration de 1 µM, de diminuer de 50 à 70 % l'invasion cellulaire par *T. cruzi*. Le premier aptamère ARN à usage thérapeutique en cours d'essai est une molécule d'ARN sélectionnée contre le facteur de croissance VEGF-165 (*vascular endothelial*

*growth factor 165*). Cet aptamère modifié semble très stable, puisqu'il peut être retrouvé 28 jours après avoir été injecté dans l'humeur vitreuse [40] et possède des propriétés thérapeutiques intéressantes. Ainsi, il est capable d'inhiber la néovascularisation dans les tests d'angiogénèse. Plus récemment, il a été montré que des oligoribonucléotides sélectionnés avec le facteur de coagulation IXa agissent comme de puissants anticoagulants [41]. Cet effet peut être inhibé par un oligonucléotide antisens (antidote) complémentaire de l'aptamère.

La découverte de l'efficacité du mécanisme de l'interférence ARN (initialement mis en évidence chez le nématode *Caenorhabditis elegans*) dans des cellules humaines, ouvre de nouvelles voies thérapeutiques (→). L'interférence ARN utilise un court fragment d'ARN double brin (en général 21-23 nucléotides) qui s'apparie par complémentarité de bases avec sa cible ARN, provoquant sa destruction. Cette stratégie, récemment employée pour inhiber l'infection de cellules par le VIH [42], sera probablement très utilisée pour essayer de contrôler l'expression de gènes impliqués dans différentes maladies.

Ces quelques exemples montrent la grande variété de stratégies fondées sur l'ARN qui peuvent être utilisées dans la régulation de l'expression génique. De nombreuses entreprises de biotechnologie développent de nouvelles molécules thérapeutiques en s'appuyant sur les mécanismes moléculaires de régulations post-transcriptionnelles et sur les données structurales de plus en plus nombreuses concernant les complexes ribonucléoprotéiques. Ainsi, l'identification et la caractérisation de nouveaux complexes ribonucléoprotéiques permettront l'identification de nouvelles cibles pharmacologiques. ♦

## REMERCIEMENTS

Les travaux provenant de notre équipe sont soutenus par le Cnrs (équipe ATIP), la Fondation pour la Recherche Médicale (FRM) et le ministère de la Recherche (ACI nanosciences).

## SUMMARY

### Assembly of ribonucleoparticle complexes

RNA-proteins interactions are involved in numerous cellular functions. These interactions are found in most cases within complex macromolecular assemblies. The recent development of tools and techniques to study RNA-protein complexes has significantly increased our knowledge in the nature of these specific interactions. The aim of this article is to present the different techniques used to study RNA-protein complexes, as well as recent data concerning the application of RNA as therapeutic molecules. ♦

## RÉFÉRENCES

1. Anantharaman V, Koonin EV, Aravind L. Comparative genomics and evolution of proteins involved in RNA metabolism. *Nucleic Acids Res* 2002; 30: 1427-64.
2. Tuerk C, Gold L. Systematic evolution of ligands by exponential enrichment: RNA ligands to bacteriophage T4 DNA polymerase. *Science* 1990; 249: 505-10.
3. Singer BS, Shtatland T, Brown D, Gold L. Libraries for genomic SELEX. *Nucleic Acids Res* 1997; 25: 781-6.
4. Stelzl U, Nierhaus KH. SERF: *in vitro* election of random RNA fragments to identify protein binding sites within large RNAs. *Methods* 2001; 25: 351-7.
5. Drolet DW, Jenison RD, Smith DE, Pratt D, Hicke BJ. A high throughput platform for systematic evolution of ligands by exponential enrichment (SELEX). *Comb Chem High Throughput Screen* 1999; 2: 271-8.
6. Ghisolfi-Nieto L, Joseph G, Puvion-Dutilleul F, Amalric F, Bouvet P. Nucleolin is a sequence-specific RNA-binding protein: characterization of targets on pre-ribosomal RNA. *J Mol Biol* 1996; 260: 34-53.
7. Serin G, Joseph G, Faucher C, et al. Localization of nucleolin binding sites on human and mouse pre-ribosomal RNA. *Biochimie* 1996; 78: 530-8.
8. Ginisty H, Serin G, Ghisolfi-Nieto L, et al. Interaction of nucleolin with an evolutionarily conserved pre-ribosomal RNA sequence is required for the assembly of the primary processing complex. *J Biol Chem* 2000; 275: 18845-50.
9. Tsai DE, Harper DS, Keene JD. U1-snRNP-A protein selects a ten nucleotide consensus sequence from a degenerate RNA pool presented in various structural contexts. *Nucleic Acids Res* 1991; 19: 4931-6.
10. Katsamba PS, Park S, Laird-Offringa IA. Kinetic studies of RNA-protein interactions using surface plasmon resonance. *Methods* 2002; 26: 95-104.
11. Darnell JC, Jensen KB, Jin P, Brown V, Warren ST, Darnell RB. Fragile X mental retardation protein targets G quartet mRNAs important for neuronal function. *Cell* 2001; 107: 489-99.
12. Tenenbaum SA, Lager PJ, Carson CC, Keene JD. Ribonomics: identifying mRNA subsets in mRNP complexes using antibodies to RNA-binding proteins and genomic arrays. *Methods* 2002; 26: 191-8.
13. Brown V, Jin P, Ceman S, et al. Microarray identification of FMRP-associated brain mRNAs and altered mRNA translational profiles in fragile X syndrome. *Cell* 2001; 107: 477-87.
14. Schaeffer C, Bardoni B, Mandel JL, Ehresmann B, Ehresmann C, Moine H. The fragile X mental retardation protein binds specifically to its mRNA via a purine quartet motif. *EMBO J* 2001; 20: 4803-13.
15. Rodgers ND, Jiao X, Kiledjian M. Identifying mRNAs bound by RNA-binding proteins using affinity purification and differential display. *Methods* 2002; 26: 115-22.
16. Bachler M, Schroeder R, von Ahsen U. StreptoTag: a novel method for the isolation of RNA-binding proteins. *RNA* 1999; 5: 1509-16.
17. Paillard L, Omilli F, Legagneux V, Bassez T, Maniey D, Osborne HB. EDEN and EDEN-BP, a *cis* element and an associated factor that mediate sequence-specific mRNA deadenylation in *Xenopus* embryos. *EMBO J* 1998; 17: 278-87.
18. SenGupta DJ, Zhang B, Kraemer B, Pochart P, Fields S, Wickens M. A three-hybrid system to detect RNA-protein interactions *in vivo*. *Proc Natl Acad Sci USA* 1996; 93: 8496-501.
19. Martin F, Schaller A, Eglite S, Schumperli D, Muller B. The gene for histone RNA hairpin binding protein is located on human chromosome 4 and encodes a novel type of RNA binding protein. *EMBO J* 1997; 16: 769-78.
20. Sagesser R, Martinez E, Tsagris M, Tabler M. Detection and isolation of RNA-binding proteins by RNA-ligand screening of a cDNA expression library. *Nucleic Acids Res* 1997; 25: 3816-22.
21. Harada K, Martin SS, Frankel AD. Selection of RNA-binding peptides *in vivo*. *Nature* 1996; 380: 175-9.
22. Jain C, Belasco JG. A structural model for the HIV-1 Rev-RRE complex deduced from altered-specificity rev variants isolated by a rapid genetic strategy. *Cell* 1996; 87: 115-25.
23. Bouvet P, Jain C, Belasco JG, Amalric F, Erard M. RNA

## GLOSSAIRE

### Aptamère

Petite molécule oligonucléotidique ou peptidique qui possède une grande affinité et une grande spécificité pour une molécule cible (une protéine par exemple).

### G-quartet (G-quadruplex)

Structure ARN ou ADN retrouvée dans des séquences simple-brins riches en guanine.

### Northwestern

Révélation de protéines séparées sur gel de polyacrylamide à l'aide d'une séquence ARN marquée.

### Phage display

Expression d'une banque de protéines à la surface de phages.

### Protéomique

Analyse des protéines d'un complexe par combinaison de techniques de séparation et d'identification des protéines

### Ribonomique

Ensemble de techniques qui permettent l'analyse de séquences d'ARN structurellement ou fonctionnellement associées.

### Ribozyme

Séquences d'acides nucléiques qui possèdent des propriétés catalytiques. Les ribozymes sont à l'origine des « maxizymes » qui regroupent deux ribozymes dont l'un porte l'activité de clivage et l'autre la fonction de reconnaissance du substrat. Le clivage a lieu uniquement si le substrat est reconnu.

### RPS

Résonance plasmonique de surface. Technologie utilisée par l'appareil « Biacore » pour suivre en temps réel les interactions entre molécules.

### SELEX

*Systematic evolution of ligand by exponential enrichment*. Isolement au sein d'une banque d'acides nucléiques aléatoires des séquences qui présentent une affinité pour une molécule cible.

### Triple hybride

Reconstitution d'un complexe de transcription actif par interaction entre un ARN chimérique, dont une partie est un fragment ARN reconnu par une protéine couplée à un domaine de fixation à l'ADN, et la seconde partie est un ARN X reconnu par une protéine cible fusionnée au domaine d'activation de la transcription.



- recognition by the joint action of two nucleolin RNA-binding domains: genetic analysis and structural modeling. *EMBO J* 1997; 16: 5235-46.
24. Danner S, Belasco JG. T7 phage display: a novel genetic selection system for cloning RNA-binding proteins from cDNA libraries. *Proc Natl Acad Sci USA* 2001; 98: 12954-9.
  25. Andersen JS, Lyon CE, Fox AH, et al. Directed proteomic analysis of the human nucleolus. *Curr Biol* 2002; 12: 1-11.
  26. Scherl A, Coute Y, Deon C, et al. Functional proteomic analysis of human nucleolus. *Mol Biol Cell* 2002; 13: 4100-9.
  27. Dragon F, Gallagher JE, Compagnone-Post PA, et al. A large nucleolar U3 ribonucleoprotein required for 18S ribosomal RNA biogenesis. *Nature* 2002; 417: 967-70.
  28. Honey S, Schneider BL, Schieltz DM, Yates JR, Futcher B. A novel multiple affinity purification tag and its use in identification of proteins associated with a cyclin-CDK complex. *Nucleic Acids Res* 2001; 29: E24.
  29. Rigaut G, Shevchenko A, Rutz B, Wilm M, Mann M, Seraphin B. A generic protein purification method for protein complex characterization and proteome exploration. *Nat Biotechnol* 1999; 17: 1030-2.
  30. Mendell JT, Dietz HC. When the message goes awry: disease-producing mutations that influence mRNA content and performance. *Cell* 2001; 107: 411-4.
  31. Charlet BN, Savkur RS, Singh G, Phillips AV, Grice EA, Cooper TA. Loss of the muscle-specific chloride channel in type 1 myotonic dystrophy due to misregulated alternative splicing. *Mol Cell* 2002; 10: 45-53.
  32. Xavier KA, Eder PS, Giordano T. RNA as a drug target: methods for biophysical characterization and screening. *Trends Biotechnol* 2000; 18: 349-56.
  33. Schlunzen F, Zarivach R, Harms J, et al. A structural basis for the interaction of antibiotics with the peptidyl transferase centre in eubacteria. *Nature* 2001; 413: 814-21.
  34. Opalinska JB, Gewirtz AM. Nucleic-acid therapeutics: basic principles and recent applications. *Nat Rev Drug Discov* 2002; 1: 503-14.
  35. Sullenger BA, Gilboa E. Emerging clinical applications of RNA. *Nature* 2002; 418: 252-8.
  36. Tanabe T, Kuwabara T, Warashina M, Tani K, Taira K, Asano S. Oncogene inactivation in a mouse model. *Nature* 2000; 406: 473-4.
  37. Sullenger BA, Gallardo HF, Ungers GE, Gilboa E. Overexpression of TAR sequences renders cells resistant to human immunodeficiency virus replication. *Cell* 1990; 63: 601-8.
  38. Darfeuille F, Arzumanov A, Gryaznov S, Gait MJ, Di Primo C, Toulme JJ. Loop-loop interaction of HIV-1 TAR RNA with N3'→P5' deoxyphosphoramidate aptamers inhibits *in vitro* Tat-mediated transcription. *Proc Natl Acad Sci USA* 2002; 99: 9709-14.
  39. Ulrich H, Magdesian MH, Alves MJ, Colli W. *In vitro* selection of RNA aptamers that bind to cell adhesion receptors of *Trypanosoma cruzi* and inhibit cell invasion. *J Biol Chem* 2002; 277: 20756-62.
  40. Drolet DW, Nelson J, Tucker CE, et al. Pharmacokinetics and safety of an anti-vascular endothelial growth factor aptamer (NX1838) following injection into the vitreous humor of Rhesus monkeys. *Pharm Res* 2000; 17: 1503-10.
  41. Rusconi CP, Scardino E, Layzer J, et al. RNA aptamers as reversible antagonists of coagulation factor IXa. *Nature* 2002; 419: 90-4.
  42. Lawrence D. RNAi could hold promise in the treatment of HIV. *Lancet* 2002; 359: 2007.
  43. Birney ES, Kumar S, Krainer AR. Analysis of the RNA-recognition motif and RS and RGG domains: conservation in metazoan pre-mRNA splicing factors. *Nucleic Acids Res* 1993; 21: 5803-16.
  44. Allain FH, Gilbert DE, Bouvet P, Feigon J. Solution structure of the two N-terminal RNA-binding domains of nucleolin and NMR study of the interaction with its RNA target. *J Mol Biol* 2000; 303: 227-41.
  45. Ginisty H, Amalric F, Bouvet P. Two different combinations of RNA-binding domains determine the RNA binding specificity of nucleolin. *J Biol Chem* 2001; 276: 14338-43.
  46. De Bouille K, Verkerk AJ, Reyniers E, et al. A point mutation in the *FMR-1* gene associated with fragile X mental retardation. *Nat Genet* 1993; 3: 31-5.
  47. Darnell JC, Jensen KB, Jin P, Brown V, Warren ST, Darnell RB. Fragile X mental retardation protein targets G quartet mRNAs important for neuronal function. *Cell* 2001; 107: 489-99.
  48. Schaeffer C, Bardoni B, Mandel JL, Ehresmann B, Ehresmann C, Moine H. The fragile X mental retardation protein binds specifically to its mRNA via a purine quartet motif. *EMBO J* 2001; 20: 4803-13.
  49. Ghisolfi L, Joseph G, Amalric F, Erard M. The glycine-rich domain of nucleolin has an unusual supersecondary structure responsible for its RNA-helix-destabilizing properties. *J Biol Chem* 1992; 267: 2955-9.
  50. Birney E, Kumar S, Krainer AR. Analysis of the RNA-recognition motif and RS and RGG domains: conservation in metazoan pre-mRNA splicing factors. *Nucleic Acids Res* 1993; 21: 5803-16.
  51. Bouvet P, Diaz JJ, Kindbeiter K, Madjar JJ, Amalric F. Nucleolin interacts with several ribosomal proteins through its RGG domain. *J Biol Chem* 1998; 273: 19025-9.
  52. Hanakahi LA, Sun H, Maizels N. High affinity interactions of nucleolin with G-G-paired rDNA. *J Biol Chem* 1999; 274: 15908-12.
  53. Ryter JM, Schultz SC. Molecular basis of double-stranded RNA-protein interactions: structure of a dsRNA-binding domain complexed with dsRNA. *EMBO J* 1998; 17: 7505-13.
  54. Ramos A, Grunert S, Adams J, et al. RNA recognition by a Staufen double-stranded RNA-binding domain. *EMBO J* 2000; 19: 997-1009.
  55. Conrad C, Rauhut R. Ribonuclease III: new sense from nuisance. *Int J Biochem Cell Biol* 2002; 34: 116-29.
  56. Tan R, Frankel AD. Structural variety of arginine-rich RNA-binding peptides. *Proc Natl Acad Sci USA* 1995; 92: 5282-6.
  57. Weiss MA, Narayana N. RNA recognition by arginine-rich peptide motifs. *Biopolymers* 1998; 48: 167-80.
  58. Campisi DM, Calabro V, Frankel AD. Structure-based design of a dimeric RNA-peptide complex. *EMBO J* 2001; 20: 178-86.

TIRÉS À PART

P. Bouvet