

# Les régulations par les petits ARN

François Dautry

FRE 2937, Institut André Lwoff, Villejuif

Auteur correspondant : François Dautry, [fdautry@vjf.cnrs.fr](mailto:fdautry@vjf.cnrs.fr)

Reçu le 5 Septembre 2007

**Résumé** – L'ARN a pendant longtemps été cantonné à un rôle d'intermédiaire passif entre l'ADN porteur de l'information génétique et les protéines qui, dans la plupart des cas, mettent en œuvre cette information. Pourtant les molécules d'ADN et d'ARN peuvent s'hybrider ce qui ouvre la possibilité de nombreuses interactions entre les gènes et leurs transcrits et, en particulier, de rétrocontrôles. Au cours des dix dernières années un ensemble de travaux a mis en évidence que de telles régulations existaient effectivement chez les eucaryotes même si le schéma n'est pas exactement celui que l'on aurait pu prédire. Des ARN jouent ainsi un rôle essentiel dans la régulation de la traduction et de la dégradation des ARN messagers dans le cadre de l'interférence par l'ARN et des phénomènes apparentés. Deux éléments distinguent ces ARN régulateurs : leur taille (une vingtaine de nucléotides) et le fait qu'à un moment de leur genèse ils ont été des ARN double brin. Le champ des activités biologiques de ces petits ARN est toutefois beaucoup plus vaste que celui des ARN messagers et s'étend aussi à l'organisation du génome à travers le contrôle de la compaction de la chromatine, les réarrangements de gènes ou l'organisation spatiale du noyau. Ainsi des petits ARN interviennent-ils sur tous les niveaux de l'expression génique et sont-ils au centre des régulations cellulaires chez les eucaryotes.

**Abstract** – Regulations through small RNAs.

For a long time RNA molecules have been viewed as simple intermediates between DNA and proteins as conveyed by the name "messenger RNA". However, the similarity between RNA and DNA creates multiple opportunities for regulatory interactions between genes and their transcripts. Over the last ten years a large body of studies in different eukaryotes has shown that indeed RNA molecules play major roles in the control of gene expression. The first pathway to be analyzed has been the control of translation and degradation of messenger RNA by RNA interference and the related processes. This has led to the realization that regulatory RNA molecules have two specific features : a size around twenty nucleotides and the fact that at one point they have been double stranded. The field of action of small regulatory RNA is however much broader and also covers several levels of genomic organisation such as chromatin compaction, gene rearrangement and spatial organisation of the nucleus. Thus small regulatory RNA are involved at all levels of gene expression and are central to cellular regulations.

---

## Introduction

De par la similarité de leur structure chimique, l'ADN et l'ARN partagent la même capacité de codage de l'information génétique. Les molécules d'ARN, du fait de leur flexibilité, peuvent de plus générer une grande

variété de structures dont certaines sont capables de catalyser des réactions chimiques. Il est donc couramment proposé que les organismes vivants se soient initialement développés à partir de molécules d'ARN qui pouvaient à la fois être le support de l'information génétique et catalyser les principales réactions

nécessaires au développement de l'organisme. Pourtant, dans la très grande majorité des organismes vivants actuels, l'information génétique est stockée sous forme d'ADN tandis que, en dehors de quelques réactions essentielles comme la liaison peptidique ou l'épissage des ARN, la catalyse est assurée par des protéines. Ainsi, mis à part quelques ARN structuraux, l'ARN se trouve apparemment cantonné au rôle de vecteur de l'information génétique entre le génome et les protéines qu'assurent les ARN messagers. Une autre conséquence de cette spécialisation est que l'ADN est presque toujours sous forme de molécule double brin alors que l'ARN est au contraire sous forme simple brin.

Au-delà de ce rôle de vecteur de l'information génétique pour l'ARN, la présence dans une même cellule, mais pas nécessairement dans le même compartiment, de molécules d'ADN et d'ARN crée de nombreuses possibilités de régulations sur la base d'une complémentarité de séquence. Ainsi on peut facilement envisager tant des interactions entre ARN, par exemple dans le cytoplasme, que des rétrocontrôles des ARN messagers vers les gènes. De fait, en fonction de la longueur de la région d'hybridation il leur est possible d'être soit spécifique d'un gène, soit au contraire d'agir sur une famille de gènes. Pour un génome de la taille de celui des mammifères, une séquence de 16 nucléotides est en principe suffisante pour assurer une reconnaissance unique, tandis qu'une séquence plus courte peut être trouvée sur de nombreux gènes. Il est donc tentant de vouloir faire jouer un rôle central aux ARN dans la régulation de l'expression génique. Mais une difficulté préalable a longtemps retardé les approches expérimentales, celle de l'identification d'ARN régulateurs parmi la grande quantité d'ARN messagers ou structuraux présents dans une cellule. La réponse apportée par les travaux sur l'interférence par l'ARN est remarquablement simple et repose sur deux critères : d'une part la nature double brin, et d'autre part la taille (une vingtaine de nucléotides). D'où l'importance des petits ARN régulateurs qui font l'objet de la plupart des présentations de cette journée. En ce qui concerne l'aspect double brin un résumé extrêmement simple de la situation pourrait être : ARN simple brin = ARN messager ou structural, ARN double brin = ARN régulateur. Si cette séparation est beaucoup trop schématique pour être universelle, elle n'en a pas moins un pouvoir explicatif très important. L'une des complications vient de ce que, à partir d'un ARN double brin, il ne persiste dans la cellule qu'un petit fragment simple brin qui sert à guider une machinerie enzymatique. Ainsi, *in fine*, les ARN régulateurs ne sont-ils plus distinguables sur la base de leur nature double ou simple brin mais de leur taille. De plus les études sur l'interférence par l'ARN ont mis en évidence l'existence de deux modes d'inter-

action entre le petit ARN guide et les cibles : un mode spécifique (plus de 17 nucléotides), et un mode relâché (une dizaine de nucléotides) suggérant qu'ils peuvent agir soit sur une cible unique soit sur une famille de cibles.

Indépendamment des études sur l'interférence et les petits ARN, de nombreuses observations suggèrent un rôle central des ARN dans l'organisation du noyau et dans la structuration de la chromatine entre euchromatine (régions exprimées) et hétérochromatine (non exprimées). Ces observations soulèvent de nombreuses questions et en particulier un paradoxe apparent : comment peut-on faire jouer un rôle à des ARN dans l'organisation de régions qui ne sont pas transcrites ? Une partie de la réponse est fournie par l'observation qu'en plus des gènes classiquement identifiés, l'ensemble du génome est transcrit mais à un niveau beaucoup plus faible. En outre, la plupart de ces transcrits sont non codants ce qui renforce leur démarcation d'avec les ARN messagers. Cette transcription en dehors des « gènes classiques » a été initialement interprétée comme définissant la limite de la spécificité de la transcription mais tend à être vue comme participant à une interaction permanente entre l'ADN et son environnement. Pour une part ces mécanismes sont très proches de ceux de l'interférence par l'ARN et mettent en œuvre des petits ARN guides. Il est cependant encore trop tôt pour dire si tous les phénomènes d'organisation de la chromatine reposent sur ce concept d'ARN guide. Ainsi, alors que des ARN non codants sont impliqués dans l'inactivation du chromosome X et l'empreinte parentale, il n'est pas encore clair si ces mécanismes sont apparentés à ceux des régulations par les petits ARN.

Une particularité des travaux sur les fonctions régulatrices des ARN vient de ce qu'ils se sont développés grâce à des études menées en parallèle sur l'ensemble des organismes eucaryotes. Il s'est ainsi dégagé une vision de ces régulations qui transcende les clivages habituels entre espèces. Néanmoins, quelle que soit l'importance que l'on accorde aux petits ARN, il ne faut pas perdre de vue qu'au moins un organisme eucaryote, *S. cerevisiae*, est dépourvu de ces régulations. Ainsi, les petits ARN doivent ils être vus non comme une composante indispensable du fonctionnement des eucaryotes mais plutôt comme l'un des moyens d'atteindre le niveau de sophistication dans la régulation de l'expression génique dont les organismes complexes ont besoin. De plus, il convient d'aborder avec prudence l'apparente similarité des régulations par les ARN qui ont été observées chez les différents eucaryotes car, dans le détail, il existe d'importantes différences dans la mise en œuvre de ces régulations. Un des aspects les plus spectaculaires de ces différences est l'existence chez les plantes, les champignons, des levures et les nématodes, de po-

lymères ARN dépendant de l'ARN. Ces enzymes, tout comme les réplases virales, permettent de transformer des ARN simple brin en double brin et donc de générer des molécules régulatrices d'ARN. Ainsi, non seulement, la frontière entre ARN messagers et ARN régulateurs double brin peut elle être franchie, mais, en plus, l'amplification des ARN régulateurs rend possible leur transmission au cours des divisions cellulaires et de la reproduction. Chez ces organismes les ARN régulateurs peuvent donc être le support d'informations épigénétiques en complément du génotype porté par l'ADN.

## L'interférence par l'ARN

La découverte de l'interférence par l'ARN en 1998 (Fire *et al.*, 1998) marque une étape décisive dans l'étude des régulations par les ARN. L'article initial apporte au moins deux éléments clés. Tout d'abord, c'est la mise en évidence d'un rôle spécifique des ARN double brin, en l'occurrence de la capacité de ces ARN à induire une suppression post-transcriptionnelle de l'expression génétique basée sur une homologie de séquence. C'est aussi la démonstration que, chez *C. elegans*, cette suppression est transmissible sur une ou deux générations avant que le phénotype lié au génotype ne soit restauré. Ainsi, dès le départ, une dimension épigénétique très forte a été associée à ces régulations.

L'étude de l'interférence chez *C. elegans* a bénéficié de circonstances favorables qui ne se retrouvent pas nécessairement chez d'autres espèces. Ainsi la capacité de *C. elegans* à assimiler les ARN double brin par voie alimentaire n'est pas partagée par d'autres espèces de nématodes, mais elle s'est avérée très utile en permettant une mise en œuvre particulièrement aisée de l'interférence<sup>1</sup>. De même, la propagation des ARN double brin de cellule en cellule permet une extinction (*silencing*) systémique qui est partagée avec les plantes mais pas la drosophile et les vertébrés (au moins jusqu'à présent).

À partir de l'observation initiale, l'analyse du mécanisme a progressé très rapidement en s'appuyant sur les études réalisées chez les plantes et les champignons. Deux messages essentiels s'en sont dégagés. D'une part, les ARN double brin de grande taille sont pris en charge par une enzyme de la famille des RNase III, l'éminceuse (*dicer*) qui les transforme en petites molécules d'une vingtaine de nucléotides. Les

petits ARN interférents ainsi générés sont les effecteurs de l'interférence et peuvent être utilisés à la place des grandes molécules de l'expérience initiale. D'autre part, l'analyse du mécanisme de l'interférence a permis de dégager un paradigme pour les régulations par les petits ARN. Dans sa forme la plus simple le complexe effecteur RISC (*RNA Induced Silencing Complex*) comprend une protéine et un petit ARN guide qui correspond à l'un des brins d'un petit ARN interférent. C'est la présence de l'ARN guide qui va conférer à cette protéine une spécificité de séquence en permettant une reconnaissance d'un ARN cible par complémentarité de séquence. Ainsi, par exemple, la protéine Argonaute 2 (chez la drosophile et les mammifères) associée à un petit ARN constitue une nucléase spécifique de séquence (Hammond *et al.*, 2000; Rivas *et al.*, 2005). Si la complémentarité entre l'ARN guide et l'ARN cible est parfaite ou presque parfaite, Argonaute 2 peut alors couper l'ARN messager et ainsi empêcher l'expression du gène correspondant. Le schéma proposé est d'une remarquable élégance et peut en principe s'appliquer à de nombreuses activités biologiques. Si l'appellation d'interférence par l'ARN désignait initialement l'activité endonucléase spécifique de séquence, la très grande similarité avec les mécanismes de régulations de la traduction et, dans une moindre mesure, de compaction de la chromatine font qu'en fonction des contextes cette appellation peut aussi être utilisée pour désigner d'autres mécanismes de régulation par des petits ARN.

## Les sources d'ARN double brin

Dans les expériences initiales de Fire et Mello (Fire *et al.*, 1998), l'ARN double brin est un artéfact de laboratoire introduit par l'expérimentateur. Plusieurs sources d'ARN double brin peuvent *a priori* être envisagées en fonction des organismes. Tout d'abord, mis à part les rétrovirus, tout virus dont le génome est constitué d'ARN génère au cours de sa réplication de l'ARN double brin. De fait, chez les plantes où la majorité des virus ont un génome ARN, l'interférence par l'ARN est un important mécanisme antiviral. La pression de sélection est suffisamment forte pour que de nombreux virus aient un gène inhibiteur de l'interférence dans leur génome. Par ailleurs, les éléments répétés présents dans le génome des eucaryotes sont aussi susceptibles de générer des ARN double brin soit du fait de leurs promoteurs internes soit du fait de leur transcription à partir de promoteurs cellulaires voisins. Une implication de l'interférence dans la répression de la transposition a d'ailleurs été mise en évidence chez des nombreux organismes (Soifer, 2006). Par ailleurs, les études chez *S. pombe* ont permis

<sup>1</sup> Au laboratoire les nématodes sont élevés sur des boîtes de pétri avec un tapis bactérien. Il suffit donc d'utiliser une souche bactérienne qui exprime un ARN double brin pour induire une interférence chez le nématode qui les consomme.

d'établir que le maintien de l'hétérochromatine constitutive péri-centromérique nécessitait une transcription des séquences répétées qui les constituent. Dans des cellules mutantes pour l'éminceuse on peut détecter la présence de transcrits sens et antisens correspondants aux motifs répétés caractéristiques de ces régions.

Au-delà de ces ARN double brin issus de l'appariement de deux molécules, les ARN simple brin et en particulier les ARN messagers constituent une abondante source d'ARN « presque double brin » du fait de l'importance de leurs structures secondaires. Si la frontière entre ARN structurés et substrats des enzymes de la famille des RNases III est encore à préciser, il est d'ores et déjà apparent que chez la plupart des organismes ce sont bel et bien des ARN issus de structures en tige boucle qui constituent la principale source de petits ARN guides. Ces ARN guides issus de transcrits simple brin Pol II ou Pol III ont été baptisés micro ARN (voir plus bas). Enfin, pour les organismes chez lesquels il existe une polymérase ARN dépendant de l'ARN, tout substrat de cette activité peut être converti en ARN double brin. On comprend aisément que si ce processus peut être très utile, par exemple dans le cadre de mécanisme de contrôle de qualité des ARN messagers ou d'une réponse de type immunitaire à des ARN étrangers, il est essentiel que cette conversion soit très contrôlée pour éviter une inhibition de l'expression d'ARN messagers « authentiques » (l'équivalent d'une réaction auto-immune).

## La cosuppression chez les plantes

Avant même la découverte de l'interférence par l'ARN, d'autres travaux avaient suggéré l'existence de régulations spécifiques de séquences qui n'entraient pas dans les cadres habituels. Ainsi, la co-suppression avait été décrite chez les plantes ou l'extinction (*quelling*) chez le champignon *Neurospora crassa*. La co-suppression désigne le fait que dans des expériences de transgénèse l'introduction de copies supplémentaires d'un gène peut induire une suppression complète ou partielle de l'expression des gènes résidents. Sans être systématique, ce phénomène avait pu être suffisamment étudié pour démontrer qu'il s'agissait d'un mécanisme spécifique de séquence. A la lumière de l'interférence par l'ARN, il a pu rapidement être montré que dans de nombreux cas la co-suppression reflétait l'expression d'ARN double brin liés à la transcription des nombreuses copies du transgène, intégrées dans le génome lors de la transgénèse. Si le lien entre interférence et co-suppression est indubitable, les différences entre organismes font qu'il faut se garder d'être trop catégorique. Ainsi, il existe clairement chez les plantes la possibilité qu'un ARN possédant des caractéristiques « anormales » soit converti en

ARN double brin et ainsi déclenche une inhibition de l'expression. C'est probablement la situation chez certaines plantes transgéniques chez lesquelles on ne détecte pas de production significative d'ARN double brin. Dans ce cas un transcrit issu du transgène, détecté comme « anormal » et converti en un ARN double brin par la RNA polymérase RNA dépendante, serait immédiatement pris en charge par la machinerie de co-suppression et donc ne s'accumulerait pas dans la cellule. C'est aussi à la suite de la conversion d'un ARN messenger en ARN double brin que l'on peut observer une extension du *silencing* aux séquences situées en amont et en aval de la région initialement ciblée (Baulcombe, 2004).

## Les régulation de la traduction et les micro ARN

Si *C. elegans* a joué un rôle clé dans la mise en évidence de l'interférence par l'ARN, c'est aussi chez le nématode qu'avait été observée l'existence d'une régulation de la traduction par des petits ARN. En effet, certains gènes impliqués dans le contrôle temporel du développement comme *let-7* et *lin-4* n'ont pour seul produit détectable dans la cellule qu'un petit ARN non codant d'une vingtaine de nucléotides. La similarité entre ces petits ARN et les ARN guides de l'interférence par l'ARN est évidente, mais elle est encore plus frappante lorsque l'on considère la genèse de *lin-4* et *let-7*. En effet l'ARN de 21 nt est le produit d'une maturation d'un transcrit plus long (d'environ 70 nt dans ce cas) et dans lequel cette séquence est en grande partie appariée au sein d'une tige boucle. La démonstration que l'éminceuse est aussi responsable de cette maturation place ces petits ARN sur un pied d'égalité avec les produits de maturation des grands ARN double brin. De fait, ces petits ARN servent de guide dans des complexes très proches de ceux identifiés pour les ARN interférents et dont un constituant essentiel est une protéine argonaute. Dans le cas de *lin-4* et de *let-7* l'inhibition de l'expression a principalement lieu au niveau de la traduction tandis que dans d'autres cas une déstabilisation de l'ARN messenger peut jouer un rôle important (voir par exemple le cas de la dégradation des ARN messagers maternels chez le poisson zèbre (Giraldez *et al.*, 2006)). L'organisation au sein du cytoplasme de la régulation de la traduction est abordée dans la présentation de D. Weil et illustre les liens étroits entre la régulation de la traduction et de la dégradation.

Si pendant longtemps il a été pensé que les petits ARN régulateurs de la traduction étaient une spécificité de *C. elegans*, le lien avec l'interférence a permis de montrer qu'en fait de très nombreux

ARN cellulaires étaient maturés en petits ARN guides désignés par l'appellation de micro ARN. A l'heure actuelle, 474 micro ARN ont été identifiés chez l'homme et les estimations sur le nombre réel varient entre 800 et plusieurs dizaines de milliers. Les micro ARN (ou leurs précurseurs) constituent donc peut-être la classe la plus abondante de gènes régulateurs, devant celle des kinases. La grande difficulté est de prédire leurs gènes cibles, car contrairement au cas de l'interférence « classique » la reconnaissance a lieu par un appariement imparfait entre l'ARN guide et la cible. Les estimations les plus réservées prédisent que l'expression de 30 % des gènes est contrôlée par des micro ARN. I. Nagibneva présente un exemple d'analyse de la contribution des micro ARN au contrôle de la différenciation musculaire. Les virus se sont appropriés les micro ARN tout comme les autres niveaux de régulation cellulaire, et peuvent ainsi modifier leur expression ou celle de gènes cellulaires par l'expression de micro ARN.

## L'extinction (*silencing*) transcriptionnelle

Chez les plantes, contrairement aux premiers travaux chez *C. elegans*, il est possible d'induire une co-suppression au niveau transcriptionnel; le moyen le plus simple de le mettre en évidence étant de faire exprimer un ARN double brin dont la séquence correspond à une région promotrice et non pas à un ARN messenger. Ainsi il est possible d'induire une compaction de la chromatine de manière spécifique de séquence. Une autre observation importante chez les plantes est que, même lorsque l'ARN db est homologue à des régions transcrites et que l'on induit une extinction post-transcriptionnelle, il existe aussi une action au niveau de la chromatine sous forme de méthylation de l'ADN. La séparation entre mécanisme transcriptionnel et post-transcriptionnel est donc probablement en partie artificielle. D'ailleurs, chez *C. elegans* une analyse plus approfondie de l'approche initiale de Fire et Mello a permis de mettre en évidence qu'à la suite d'un silencing par coupure de l'ARN messenger on peut observer, au moins pour certains gènes, une extinction à long terme<sup>2</sup> due à une inhibition de la transcription (Vastenhouw *et al.*, 2006). Mais c'est chez la levure *S. pombe* que les études les plus complètes ont pu être réalisées en combinant des approches génétiques et biochimiques. Ces travaux, présentés par A. Verdell, ont permis de décrire les principaux acteurs des régulations transcriptionnelles et d'illustrer tout à la fois les similarités et les différences avec l'interférence par l'ARN. Ainsi, l'unique protéine argonaute de *S. pombe* associée à un petit ARN est

au cœur du processus d'inhibition transcriptionnelle et post-transcriptionnelle chez cet organisme. Toutefois, l'inhibition transcriptionnelle nécessite en plus l'intervention d'un complexe ayant une activité de polymérisation de l'ARN et des enzymes de modification des histones (Zaratiegui *et al.*, 2007).

## Les ARN comme mode d'emploi du génome

Les généticiens des plantes ont introduit à partir des années 1950 le concept de paramutation et, avec lui, l'idée que l'expression des gènes peut être influencée non seulement par l'environnement génétique mais aussi par l'expression du gène lui-même à la génération précédente. Une formulation imagée de ces concepts est qu'en plus du génome il en existe un mode d'emploi qui peut aussi être transmis aux générations suivantes. Si l'on est prêt à suivre cette idée, les ARN messagers ou des fragments issus de ceux-ci pourraient remplir ce rôle de mode d'emploi en conférant une spécificité de séquence. Dans le contexte très spécifique du réarrangement du génome des ciliés c'est effectivement ce qui se passe, les ARN exprimés à la génération *n* servant de guides pour le réarrangement du génome somatique à la génération *n+1*. On peut envisager une forme beaucoup plus générale de ce mode d'emploi, du type « les gènes exprimés à la génération *n* » doivent être exprimés à la génération *n+1* car ils ont été utiles au développement de l'organisme. De nouveau, dans le cas des plantes, ceci serait l'illustration directe de l'idée que d'une génération à la suivante les plantes sont le plus souvent dans le même environnement et qu'une jeune pousse exprimant les mêmes voies métaboliques que la génération précédente aura un avantage de croissance. En même temps, une programmation trop rigide priverait la plante de toute capacité d'adaptation et limiterait l'intérêt de la dissémination des graines. Il est donc essentiel que ce mode d'emploi ne soit pas aussi strictement défini que le génome lui-même. A ce stade de nos connaissances, il n'est pas possible de faire un lien direct avec les données sur les régulations par les petits ARN puisque celles-ci sont essentiellement répressives, on peut cependant construire assez facilement des boucles autoentretenuës en utilisant l'inhibition d'un inhibiteur (voir par exemple Fazi *et al.*, 2005). Ce concept de mode d'emploi du génome est donc encore spéculatif, en dehors du cas des ciliés. L'observation par l'équipe de M. Rassoulzadegan (Rassoulzadegan *et al.*, 2006) d'une paramutation chez la souris constitue donc un développement important dans notre vision des mécanismes épigénétiques chez les animaux. La démonstration que des phénotypes peuvent

<sup>2</sup> L'extinction a pu être suivie pendant 70 générations.

être induits par l'introduction d'ARN et tout particulièrement de micro ARN dans l'œuf fécondé renforce le lien avec les régulations par les petits ARN mais le détail des mécanismes en cause reste à explorer.

## Références

- Baulcombe, D. (2004). RNA silencing in plants. *Nature* 431, 356-363.
- Fazi, F., Rosa, A., Fatica, A., Gelmetti, V., De Marchis, M.L., Nervi, C., and Bozzoni, I. (2005). A minicircuitry comprised of microRNA-223 and transcription factors NFI-A and C/EBPalpha regulates human granulopoiesis. *Cell* 123, 819-831.
- Fire, A., Xu, S., Montgomery, M.K., Kostas, S.A., Driver, S.E., and Mello, C.C. (1998). Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*. *Nature* 391, 806-811.
- Giraldez, A.J., Mishima, Y., Rihel, J., Grocock, R.J., Van Dongen, S., Inoue, K., Enright, A.J., and Schier, A.F. (2006). Zebrafish MiR-430 promotes deadenylation and clearance of maternal mRNAs. *Science* 312, 75-79.
- Hammond, S.M., Bernstein, E., Beach, D., and Hannon, G.J. (2000). An RNA-directed nuclease mediates post-transcriptional gene silencing in *Drosophila* cells. *Nature* 404, 293-296.
- Rassoulzadegan, M., Grandjean, V., Gounon, P., Vincent, S., Gillot, I., and Cuzin, F. (2006). RNA-mediated non-mendelian inheritance of an epigenetic change in the mouse. *Nature* 441, 469-474.
- Rivas, F.V., Tolia, N.H., Song, J.J., Aragon, J.P., Liu, J., Hannon, G.J., and Joshua-Tor, L. (2005). Purified Argonaute2 and an siRNA form recombinant human RISC. *Nat. Struct. Mol. Biol.* 12, 340-349.
- Soifer, H.S. (2006). Do Small RNAs Interfere With LINE-1? *Journal of Biomed. & Biotechnol.* 2006, 29049.
- Vastenhouw, N.L., Brunschwig, K., Okihara, K.L., Muller, F., Tijsterman, M., and Plasterk, R.H. (2006). Gene expression : long-term gene silencing by RNAi. *Nature* 442, 882.
- Zaratiegui, M., Irvine, D.V., and Martienssen, R.A. (2007). Noncoding RNAs and gene silencing. *Cell* 128, 763-776.

## Glossaire

### Argonautes (protéines)

Les protéines argonautes sont au cœur de toutes les régulations par les petits ARN décrites jusqu'à présent. Initialement caractérisées à travers des mutations qui affectent le développement des plantes et de la drosophile, les protéines argonautes comprennent deux domaines caractéristiques : PAZ et Piwi. Il existe un nombre variable de gènes argonautes en fonction des organismes; de 1 chez *S. pombe* à 27 chez *C.*

*elegans*. Elles sont actuellement regroupées en trois sous-familles : argonautes, Piwi et le 3<sup>ème</sup> groupe. Les protéines argonautes sont le point de départ du paradigme actuel sur les régulations par les petits ARN. En effet, au moins *in vitro*, il suffit d'une protéine argonaute (Ago2 chez les mammifères) et d'un petit ARN d'une vingtaine de nucléotides pour constituer une nucléase spécifique de séquence. Le petit ARN se lie au domaine PAZ tandis que le domaine Piwi porte l'activité endonucléase qui ne coupe que l'ARN passager, laissant intact le brin guide et permettant ainsi une activité catalytique.

### (petits) ARN interférents (siRNA)

Lors de l'introduction dans des cellules de molécules de grandes tailles d'ARN double brin, celles-ci sont rapidement découpées par l'éminceuse en des fragments d'une vingtaine de nucléotides, les petits ARN interférents. Ces molécules peuvent être introduites directement pour induire une interférence. Chez les mammifères cette stratégie permet d'éviter d'induire une réponse de type interféron puisque celle-ci n'est efficacement induite que par des ARN double brin de plus d'une trentaine de nucléotides. De plus les petits ARN interférents générés par l'éminceuse ont deux nucléotides libres en 3' de chaque brin, ce qui leur permet d'échapper à la détection par la réponse immunitaire innée induite par Rig-1.

### Drosha

Certaines tiges boucles présentes sur des ARN cellulaires (transcrits Pol II ou Pol III) peuvent être reconnues par une enzyme de la famille des RNase III drosha, présente dans le noyau de la plupart des eucaryotes. Celle-ci excise la tige boucle qui est alors transportée dans le cytoplasme par l'exportine 5. Dans le cytoplasme, l'éminceuse excise la boucle et génère ainsi une tige fortement structurée qui, comme les ARN interférents possède deux nucléotides libres en 3' de chaque brin du fait que les deux coupures ont été faites par des enzymes de la famille des RNase III, et peut être incorporée dans un complexe RISC. C'est ainsi que sont produits les micro ARN présents chez de nombreux eucaryotes, mais pour l'instant pas encore décrits chez la levure *S. pombe*.

### Eminceuse (dicer)

Lors de l'introduction de molécules d'ARN double brin de grande taille chez la plupart des eucaryotes celles-ci sont découpées en fragments d'une vingtaine de nucléotides par une enzyme de la famille des RNase

III : l'éminceuse. Les RNase III sont des RNases spécifiques des ARN double brin et qui génèrent des extrémités 3' protubérantes de 2 nucléotides. Il existe un gène d'éminceuse chez les mammifères et le poisson zèbre qui est essentiel pour le développement précoce, deux chez la drosophile et quatre chez les plantes.

### Micro ARN

Initialement l'interférence par l'ARN a été observée avec des molécules d'ARN double brin soit exogènes, soit produites dans le cadre d'une infection virale, soit résultant de l'expression des nombreuses copies d'un transgène. Pourtant la forme la plus courante d'une réponse de type interférence est probablement due à des régions structurées en tige boucle sur des molécules simple brin transcrites par l'ARN polymérase II ou III. Après une maturation par Drosha et par l'éminceuse, ces molécules sont transformées en des équivalents de petits ARN interférents, les micro ARN. Si le nombre exact de micro ARN possibles chez un organisme n'est pas connu, il est cependant probable que ces molécules représentent la classe la plus abondante de molécules régulatrices. Les micro ARN sont incorporés dans des complexes de type RISC avec en général un très fort biais en faveur d'un seul des deux brins qui sert donc de guide. Chez les mammifères, la reconnaissance des ARN cibles se fait par un appariement imparfait qui privilégie la moitié 5' de l'ARN guide (la région d'amorçage); en dépit d'importants progrès dans les algorithmes, la prédiction des cibles est encore délicate.

### P bodies ou corps GW 182

Chez tous les eucaryotes (y compris la levure *S. cerevisiae*) les principaux composants de la voie de dégradation 5' des ARN s'accumulent dans des structures cytoplasmiques, les *P bodies* ou corps GW 182. On y trouve en particulier les enzymes de décoiffage (dcp1 et dcp2), l'unique exoribonucléase 5' cytoplasmique Xrn1, mais aussi les protéines argonautes et d'autres protéines régulatrices de la dégradation ou de la traduction des ARN messagers. En fonction du type cellulaire et de la physiologie on peut observer de quelques corps GW 182 à plus d'une dizaine. Il est probable que la plupart de ces constituants sont aussi présents de manière diffuse dans le cytoplasme. La présence dans les corps GW 182 des constituants des complexes RISC (protéines argonautes, petits ARN) suggère que l'interférence par l'ARN pourrait avoir lieu dans les mêmes sites que la dégradation « classique » des ARN messagers.

### Piwi (domaine protéique)

Toutes les protéines argonautes des eucaryotes comprennent quatre domaines (N-terminal, PAZ, central et Piwi). La structure tridimensionnelle de la protéine argonaute d'une archaebactérie *P. furiosus* a permis de constater que le domaine Piwi avait une structure proche de celle de la RNase H, ce qui a conduit à l'identification des protéines argonautes comme des trancheuses.

### Piwi (sous famille des protéines argonautes)

Les protéines argonautes de la sous famille Piwi sont principalement exprimées dans les cellules des lignées germinales chez les mammifères. Il en existe quatre chez l'homme et trois chez la souris qui sont associées à des petits ARN plus longs que les ARN interférents classiques (26 à 29 nucléotides). Ces petits ARN ne sont pas générés par l'éminceuse mais par une action répétitive des protéines Piwi sur certains transcrits.

### Trancheuse (*slicer*)

Dans l'interférence par l'ARN au sens strict le petit ARN guide confère une spécificité de séquence à une nucléase. C'est en 2004 que la protéine qui porte cette activité nucléase a été identifiée et appelée trancheuse. Chez les mammifères il s'agit de la protéine Ago2. L'activité nucléase est portée par le domaine Piwi de la protéine argonaute et trois acides aminés jouent un rôle clé dans la catalyse (Asp/Asp/Asp avec la possibilité d'avoir Glu, His ou Lys pour le 3<sup>ème</sup> acide aminé). Puisque l'activité nucléase porte sur un substrat double brin on peut comprendre que seules les protéines argonautes associées à un ARN guide soient actives – et spécifiques de séquence. Parmi les protéines argonautes certaines ont les acides aminés nécessaires pour l'activité nucléase mais ne sont pas pour autant toutes actives dans des tests *in vitro* (hsAgo2 est active mais pas hsAgo3 qui pour l'instant reste une trancheuse silencieuse). Il existe aussi un nombre variable en fonction des organismes de protéines argonautes qui n'ont pas les acides aminés nécessaires pour la catalyse, par exemple hsAgo1 et hsAgo4 ou la majorité des protéines argonautes du 3<sup>ème</sup> groupe chez *C. elegans*. Les fonctions biologiques de ces trancheuses émoussées n'ont pas encore été identifiées. Certaines protéines de la famille Piwi ont les acides aminés nécessaires à la catalyse et sont des trancheuses actives *in vitro*.

### **RISC (*RNA Induced Silencing Complex*)**

C'est le complexe effecteur de l'interférence qui comprend un petit ARN guide et l'ensemble des protéines nécessaire à son activité biologique. Dans le cadre de l'interférence au sens strict, le complexe a une activité nucléase spécifique de séquence qui peut être reconstituée *in vitro* avec une protéine argonaute (la trancheuse, Ago2 chez les mammifères) et un petit ARN interférent. Il est probable que *in vivo* le complexe comprend d'autres protéines, et une forme composée de trois protéines (l'éminceuse, TRBP et Ago2) a été identifiée chez les mammifères. La nature des complexes RISC impliqués dans le silencing par les micro ARN n'est toujours pas clairement identifiée; est-ce le même complexe

que celui qui porte l'activité nucléase ou diffère-t-il par la nature de la protéine argonaute et /ou d'autres constituants ?

### **RITS (*RNA induced Initiation of Transcriptional Silencing*)**

Lorsque des petits ARN induisent un silencing au niveau de la transcription par compaction de la chromatine, ils sont associés à un complexe distinct de celui des régulations post-transcriptionnelles. Un tel complexe a été caractérisé chez *S. pombe* et comprend en plus de l'unique protéine argonaute de cet organisme (Ago1) les protéines Tas3 et chip1.