

Les « doigts à zinc », éléments de reconnaissance de l'ADN

Les « protéines dactyles » dont les doigts sont stabilisés par des liaisons de coordination entre un atome de zinc et soit deux histidines et deux cystéines, soit quatre cystéines constituent une famille de molécules ayant la propriété d'interagir avec des séquences particulières d'ADN. Cette famille peut elle-même être subdivisée en plusieurs sous-groupes définis par la participation — ou non — d'histidine à la formation du doigt, par le nombre de doigts et par l'interaction symétrique ou asymétrique avec l'ADN double brin. Nombre des protéines dactyles sont des modulateurs de la transcription des gènes.

Nicole Helbecque
Jean-Pierre Hénichart

Un des objectifs de la biologie moléculaire consiste à élucider précisément les mécanismes de régulation de la transcription des gènes. Cette recherche passe par l'identification de facteurs de transcription capables de reconnaître des séquences spécifiques d'ADN et de s'y fixer (*m/s* n° 7, vol. 3, p. 428). L'étude par analyse cristallographique aux rayons X de l'interaction du répresseur et de la protéine Cro du phage lambda avec un ADN en double brin a permis de mettre en évidence un motif « hélice-tour-hélice » (*figure 1*), qui se retrouve dans différentes protéines de phages, de levures ou de bactéries ainsi que dans des protéines comprenant des séquences homéoboxes chez des eucaryotes [1]. La particularité des sites d'ADN participant à ce type de contact est d'être « palindromiques », c'est-à-dire d'avoir une symétrie par rapport à un point et de comporter ainsi des séquences identiques sur les deux brins. Plus récemment, Miller, McLachlan et Klug [2] ont montré l'existence de séquences répétitives pour le facteur de transcription TFIIIA chez un amphibien, le xénope. Ils ont proposé un repliement de chacune de ces séquences

autour d'un atome de zinc pour former des domaines structuraux séparés, les « doigts à zinc ». Par la suite, des séquences protéiques similaires ont été mises en évidence pour de nombreuses protéines capables de se lier à l'ADN [3] parmi lesquelles des facteurs de transcription impliqués dans la régulation des ARN polymérases II et III.

Étude du facteur de transcription TFIIIA

Le facteur de transcription TFIIIA est une protéine de 344 résidus qui existe sous forme de monomère de poids moléculaire 38 600 daltons. Elle a un rayon de Stokes* de 34 Å, un coefficient de sédimentation de 2,8 S** et est fortement asymétrique [4]. Elle se trouve associée à de l'ARN 5S au sein d'un complexe ribonucléoprotéique 7S qui est sa forme de stockage et qui contient de sept à onze atomes de zinc par molécule. Par protéolyse de la protéine TFIIIA on obtient des fragments

ADRESSE

N. Helbecque : Ingénieur de recherche. J.-P. Hénichart : Directeur de recherche à l'Inserm. Inserm U. 16, place de Verdun, 59045 Lille Cedex, France.

* Rayon de Stokes : rayon de la sphère d'hydratation d'une protéine.

** S : unités Svedberg, unités de centrifugation.

d'environ 3 000 daltons. La détermination de la séquence complète de TFIIIA réalisée à partir d'ADN cloné a montré que l'enchaînement des résidus 13 à 276 est constitué de neuf séquences similaires de 30 acides aminés chacune, comportant une paire de cystéines (Cys) et une paire d'histidines (His) à des positions invariantes suivant la séquence consensus : (Tyr Phe) X Cys X₂₋₄ Cys X₃ Phe X₅ Leu X₂ His X₃ His X₂₋₆, où X représente n'importe quel acide aminé.

En présence de zinc, ces segments forment des complexes, les « doigts à zinc », impliquant les deux Cys et les deux His et un atome de zinc tétraédriquement coordonné.

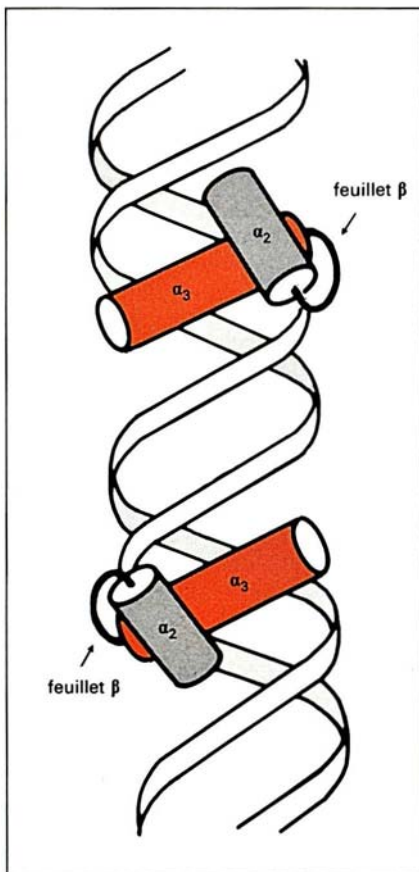


Figure 1. **Représentation schématique du motif « hélice-tour-hélice » d'interaction de l'ADN avec le complexe récepteur-opérateur du phage lambda.** L'hélice α_3 se loge dans le grand sillon tandis que l'hélice α_2 vient en contact avec le squelette de l'ADN. Sur le schéma, les hélices sont représentées par des cylindres. (D'après [1]).

m/s n° 10 vol. 4, décembre 88

L'initiation de la transcription du gène de l'ARN 5S par l'ARN polymérase III nécessite la formation d'un complexe stable de transcription entre la séquence d'ADN concernée et trois facteurs de transcription au moins, TFIIIA, B et C [2]. La première étape de ce processus d'initiation consiste en une interaction spécifique entre la protéine TFIIIA et la séquence d'ADN qui contrôle la transcription et qui comporte environ 50 nucléotides. Ce complexe est ensuite stabilisé par la fixation des facteurs B et C.

Le fait que le complexe TFIIIA-Zn ait plusieurs doigts joue un rôle important dans sa liaison à l'ADN. Il a été montré récemment que les deux doigts les plus proches de l'extrémité N-terminale de la protéine sont en contact avec chaque nucléotide du grand sillon sur un tour d'ADN tandis que les cinquième et sixième doigts se lient à l'ADN dans le petit sillon, les deux derniers doigts étant en contact pour leur part avec chaque nucléotide du grand sillon sur un tour d'ADN.

Les sites nucléotidiques protégés par TFIIIA lors de sa liaison à l'ADN ont été étudiés par *foot printing*, technique d'« empreinte » laissée par la protéine fixée à l'ADN lorsque ce dernier est digéré par les nucléases (la région de l'ADN liée à TFIIIA est alors protégée contre la digestion) [5], ce qui a permis de déterminer une séquence répétitive d'ADN protégée par TFIIIA, XCXGGXPYXGGPu où X représente n'importe quelle base, Pu symbolise une base purique et Py une base pyrimidinique.

Protéines susceptibles de former des doigts à zinc

Suite à cette étude relative au facteur de transcription du xénope, des enchaînements du type Cys X₂₋₄ Cys X₃ Phe X₅ Leu X₂ His X₃ His X₅ susceptibles de former des doigts à zinc [6] ont été mis en évidence dans de nombreuses autres protéines (Tableau I) qui jouent un rôle de régulation de la croissance ou de la différenciation cellulaire. C'est le cas, par exemple, des produits des gènes Kruppel et Serendipity β et δ chez la drosophile, du gène *ADR-1* chez la levure, des gènes *mkr₁* et *mkr₂* expri-

més principalement dans le cerveau chez la souris [7], du gène *NGF-1A* qui est exprimé dans le cerveau, les poumons et les ganglions cervicaux supérieurs chez le rat, et du gène *Egr-1* qui est exprimé surtout dans le cœur, le cerveau et les poumons chez les mammifères. C'est le cas également de la protéine codée par le gène *Sp-1* humain, isolée primitivement de cellules HeLa, puis retrouvée dans de nombreux autres tissus (cellules T, tissu placentaire) et qui se lie à des motifs riches en GC [8]. C'est le cas enfin du produit du gène *Krox 20* de souris qui pourrait jouer un rôle au niveau de la prolifération cellulaire. Ces protéines contiennent un minimum de deux doigts séparés par une chaîne de sept à huit résidus d'acides aminés de séquence très conservée.

Une autre classe de « doigts à zinc » pour laquelle les histidines ont été remplacées par des cystéines a également été caractérisée. La séquence consensus correspondante est Cys X₂ Cys X₁₃ Cys X₂ Cys (a) ou encore Cys X₂ Cys X₆ Cys X₆ Cys X₂ Cys (b) (Tableau I, p. 627). Cette classe de doigts est retrouvée dans diverses protéines telles que les produits du gène *E1A* d'adénovirus (a), qui jouent un rôle crucial de régulation positive de la transcription, des produits des gènes *GAL-4*, *PPR-1* et *ARGR-2* de *Saccharomyces cerevisiae* (b), du gène *LAC-9* de *Kluyveromyces lactis* (b) et du gène *qa-1f* de *Neurospora crassa* (b). Ces protéines d'eucaryotes unicellulaires ont en commun une séquence consensus (code à une lettre)* du type :

	KK	D	K
A CDXCR Pho	Pho KC	XXXXPXCX	C Pho
	RR	S	R

où Pho représente un acide aminé de type hydrophobe et X représente n'importe quel type d'acide aminé. A cette classe (a) se rattachent également les récepteurs des hormones stéroïdiennes [9] pour lesquels la séquence de la partie qui se lie à l'ADN (région C) est fortement conservée : il y a environ 90 % d'homologie pour le récepteur d'une même hormone dans différentes

* A = Ala ; C = Cys ; D = Asp ; K = Lys ; R = Arg ; S = Ser ; P = Pro.

RÉFÉRENCES

1. Pabo CO, Sauer RT. Protein-DNA recognition. *Ann Rev Biochem* 1984 ; 53 : 293-321.
2. Miller J, McLachlan AD, Klug A. Repetitive zinc-binding domains in the protein transcription factor IIIA from *Xenopus oocytes*. *EMBO J* 1985 ; 4 : 1609-14.
3. Evans RM, Hollenberg SM. Zinc fingers : gilt by association. *Cell* 1988 ; 52 : 1-3.
4. Bieker JJ, Roeder RG. Physical properties and DNA-binding stoichiometry of a 5S gene-specific transcription factor. *J Biol Chem* 1984 ; 259 : 6158-64.
5. Fairall L, Rhodes D, Klug A. Mapping of the sites of protection on a 5S RNA gene by the *Xenopus* transcription factor III A. *J Mol Biol* 1986 ; 192 : 577-91.
6. Vincent A. TF III A and homologous genes. The « finger » proteins. *Nucleic Acids Res* 1986 ; 14 : 4385-91.
7. Chowdhury K, Deutsch U, Gruss P. A multigene family encoding several « finger » structures is present and differentially active in mammalian genomes. *Cell* 1987 ; 48 : 771-8.
8. Kadonaga JT, Carner KR, Masiarz FR, Tjian R. Isolation of a cDNA encoding transcription factor Sp1 and functional analysis of the DNA binding domain. *Cell* 1987 ; 51 : 1079-90.
9. Gehring U. Steroid hormone receptors : biochemistry, genetics, and molecular biology. *TIBS* 1987 ; 12 : 399-402.
10. Berg JM. Proposed structure for the zinc-binding domains from transcription factor III A and related proteins. *Proc Natl Acad Sci USA* 1988 ; 85 : 99-102.

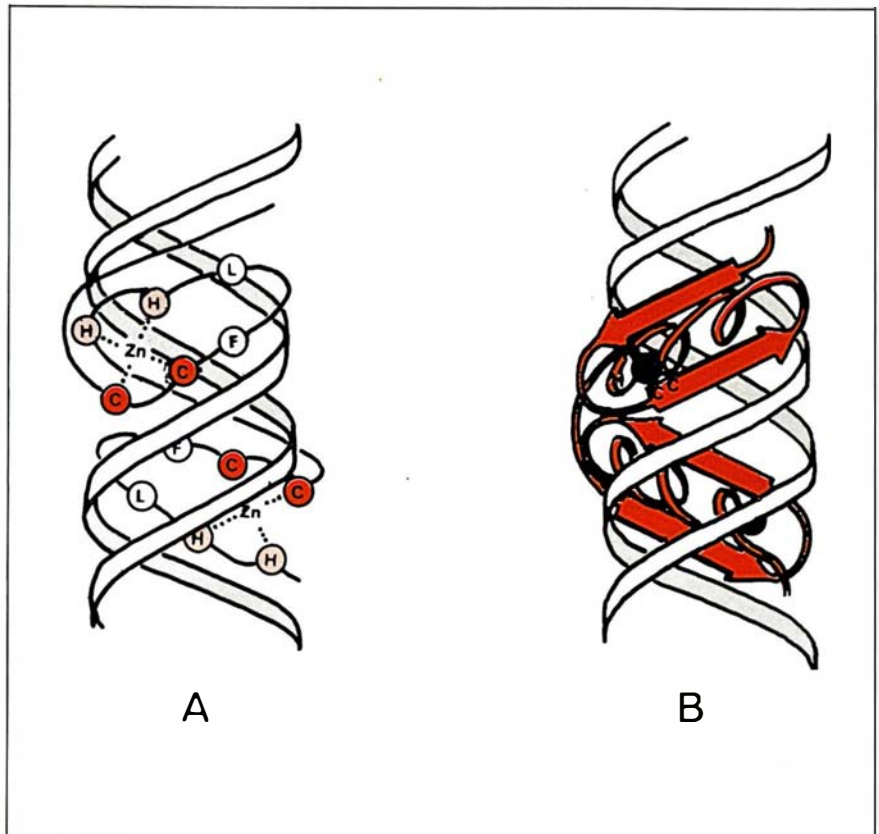


Figure 2. Structures probables d'un «doigt à zinc» du type TFIIIA et schémas d'interaction avec l'ADN. A. Sont indiqués les acides aminés invariants du facteur TFIIIA ainsi que les acides aminés impliqués dans la liaison à l'atome de zinc. B. L'atome de zinc est symbolisé par un rond noir et les paires de cystéines et d'histidines formant la structure en «doigt» sont indiquées respectivement par C (cystéines) et par H (histidines). (D'après [10]).

espèces, et 50 à 60 % d'homologie pour différents récepteurs dans la même espèce. Tous ces récepteurs contiennent deux séquences successives riches en cystéine, la première répondant à la séquence consensus (code à une lettre)* :

S	S
Pho CXVCXDXA	G XH Pho XXX CXXCXXFF
T	T

Il est à noter qu'une séquence du même type a été mise en évidence d'une part dans le cas du récepteur de l'acide rétinoïque chez l'homme (qui joue un rôle dans la différenciation cellulaire), d'autre part dans le cas de la protéine oncogène du virus d'érythroblastose aviaire (*v-erb A*) et de sa contrepartie cellulaire (*c-erb A*).

* V = Val ; T = Thr ; G = Gly ; H = His ; F = Phe.

Cependant il faut remarquer que si cette famille des récepteurs stéroïdiens se compose de protéines ayant deux séquences successives riches en cystéine, seule la première de ces séquences répond à la formule consensus Cys X₂ Cys X₆ Cys X₆ Cys X₂ Cys.

Enfin d'autres séquences consensus ont été mises en évidence dans le cas de protéines de rétrovirus ou de bactériophages ainsi que pour quelques synthétases d'ARN de transfert. Ces séquences sont répertoriées dans le Tableau I.

Interaction de ces protéines avec l'ADN

Depuis la mise en évidence par Miller [2] de la structure protéique en doigt comme médiateur de l'interaction avec l'ADN au cours de la trans-

cription, on a identifié par séquençage de clones d'ADN de nombreuses autres protéines possédant l'une ou l'autre des séquences consensus citées précédemment, à savoir Cys X₂₋₁, Cys X₃, Phe X₅, Leu X₂, His X₃, His ou Cys X₂, Cys X₁₃, Cys X₂, Cys et Cys X₂, Cys X₆, Cys X₆, Cys X₂, Cys. Pour pouvoir les classifier en tant que « doigts à zinc », il faut encore s'as-

surer que ces protéines nécessitent la présence de ce métal pour interagir avec l'ADN. Ceci n'a été démontré que pour quelques protéines outre le facteur de transcription TFIIIA du xénope, à savoir les protéines *SPI* et *GAL-4*. Par ailleurs, on a de très fortes présomptions en ce qui concerne la protéine du gène *Uvr A* d'*E. coli* ainsi que le récepteur des

estrogènes humain.

Il est à noter que les deux principales classes de protéines citées précédemment semblent différentes quant à leur mode d'interaction avec l'ADN. Dans le cas des protéines possédant plusieurs « doigts à zinc » (de 2 à 13) séparés par de courtes séquences de sept à huit acides aminés (*figure 2A*), groupe auquel appartient le facteur

Tableau I
CLASSIFICATION DES DOIGTS A ZINC

Origine	Nom de la séquence ou de la protéine	Nombre de doigts	Rôle de la protéine
Séquence type Cys X₂₋₄ Cys X₃ Phe X₅ Leu X₂ His X₂₋₄ His			
Xénope	TFIIIA	9	Facteur de transcription
Drosophile	Serendipity β, δ	5(β) ; 6 + 1 (δ)	Gène de transcription chez l'embryon
	Kruppel	4	Gène de segmentation
	Hunchback	4 + 2	Gène de segmentation
Souris	mkr ₁ , mkr ₂	7(1) ; 9(2)	Contrôle de l'expression génétique pendant le développement
	Krox 20	3	Contrôle de la transcription
Levure	ADR1	2	Active la transcription du gène ADH2
Mammifères	Sp 1	3	Facteur de transcription
	NGF-IA	3	Stimule la différenciation neuronale
	Egr-1	3	Contrôle la transcription
	TDF	13	Détermination du sexe
Séquence type Cys X₂ Cys X₆ Cys X₆ Cys X₂ Cys			
Eucaryotes unicellulaires	GAL-4	1	Métabolisme du galactose
	PPR-1	1	Régulation de la biosynthèse de la pyrimidine
	ARGR-2	1	Métabolisme de l'arginine
	LAC-9	1	Métabolisme du galactose
	qa-1f	1	Active la transcription de gènes impliqués dans l'utilisation de l'acide quinique
Séquence type Cys X₂ Cys X₁₃ Cys X₂ Cys			
Adénovirus	E1A	1	Stimulation de la transcription
Mammifères	Récepteurs stéroïdiens	1	Médiateurs de l'action des hormones stéroïdiennes, contrôle de la transcription
Séquence type Cys X₂ Cys X₄ His X₄ Cys			
Rétrovirus	gène gag	1	Formation du <i>core</i> protéique des virions
Séquence type Cys X₃ His X₅ Cys X₂ Cys			
Bactériophage	gène 32	1	Déstabilise l'hélice
Séquence type Cys X₂ Cys X₉ Cys X₂ Cys ou Cys X₂ Cys X₆ His X₂ His			
Bactéries	ARNt synthétases	1	Métabolisme des acides nucléiques

TFIIIA, la protection est asymétrique. On peut envisager un mode d'interaction avec l'ADN tel que la protéine s'enroule en quelque sorte autour de la double hélice d'ADN, les « doigts » successifs suivant l'hélice formée par le grand sillon, ce qui amène ces « doigts » à faire « face » (figure 2A, p. 626). Cependant il a été proposé (dans un premier temps) que ces « doigts » possèdent une structure en hélice α à leur extrémité C-terminale. Par la suite un modèle plus élaboré a été imaginé, à savoir une structure du type « feuillet β -coude β -hélice α » (figure 2B, p. 626). Cela conduit à un autre modèle d'interaction avec l'ADN dans lequel l'hélice α et le coude β seraient situés dans le grand sillon tandis que le feuillet β serait plus éloigné de l'axe de l'hélice. La protéine formerait ainsi une sorte de berceau dans lequel se logerait l'ADN, ce qui amène ici aussi les « doigts » à faire « face » (figure 2B). En fait des études très récentes par *footprinting* ont permis de montrer l'interaction des

doigts successifs avec les bases nucléiques d'abord dans le petit sillon, puis dans le grand sillon ; la protéine fait donc bien face à sa cible lors de son interaction avec celle-ci. En revanche, la protection s'exerce sur les deux brins de l'ADN en ce qui concerne les protéines à doigt unique du type Cys X₂ Cys X₁₃ Cys X₂ Cys (famille des récepteurs stéroïdiens). Il faut remarquer que ces enchaînements ont une forte tendance à adopter une conformation en hélice α à leur extrémité N-terminale.

Le mécanisme d'interaction de ces protéines avec l'ADN est totalement inconnu à ce jour bien que l'on sache que la liaison à l'hormone soit un préalable pour la spécificité d'interaction du récepteur avec l'acide nucléique en cause.

Toutes ces remarques amènent à penser que les deux classes majeures de protéines à doigts interagissent différemment avec l'ADN, bien qu'elles jouent un rôle à un niveau ou à un autre dans la régulation de la transcription ■

Summary

Until the early 80's the only motif described for protein-DNA interactions was the so-called « helix-turn-helix » motif. Recent studies on *Xenopus* transcription factor III A have led to propose another structure for recognition of nucleic acids by proteins, the « zinc finger ». This structure consists of a peptidic chain of approximately thirty amino acids, containing invariable cysteal and/or histidyl residues able to complex one zinc atom. It has been found in proteins from different origins, many of them playing a role in the regulation of transcription.

TIRÉS A PART

N. Helbecque.