

Les domaines de liaison de l'ADN des facteurs de transcription eucaryotes

Pierre-Olivier Angrand

Les facteurs de transcription eucaryotes se lient à l'ADN grâce à des domaines indépendants particuliers. Ces facteurs de transcription peuvent être groupés en plusieurs classes, définies par la structure tridimensionnelle de leur domaine de liaison à l'ADN. En dépit de différences structurales importantes, ces différentes classes de facteurs de transcription utilisent fréquemment une hélice α comme motif protéique de reconnaissance de l'ADN. Cette hélice s'ajuste dans le grand sillon de l'ADN et, par l'intermédiaire de ponts salins et de liaisons hydrogènes, interagit spécifiquement avec les paires de bases.

Code à une lettre des acides aminés essentiels :

A : Ala ; C : Cys ; D : Asp ;
 E : Glu ; F : Phe ; G : Gly ;
 H : His ; I : Ile ; K : Lys ;
 L : Leu ; M : Met ; N : Asn ;
 P : Pro ; Q : Gln ; R : Arg ;
 S : Ser ; T : Thr ; V : Val ;
 W : Trp ; Y : Tyr ;
 X : n'importe quel acide aminé

Les premiers facteurs de transcription (voir *m/s* n° 8, vol. 3, p. 487 et n° 9, vol. 3, p. 546) dont l'anatomie fonctionnelle a été décrite furent les protéines GAL4 et GCN4 de la levure *S. cerevisiae*. Les résultats primordiaux issus de ces études établissent la distinction entre un domaine de liaison à l'ADN et un domaine d'activation de la transcription ; les 147 résidus N-terminaux de GAL4 (881 acides aminés) sont suffisants pour sa fixation de manière spécifique à l'ADN (sur la séquence UAS_G), mais ne permettent pas d'activer la transcription. Si l'on remplace ces 147 résidus par le domaine de fixation à l'ADN du répresseur bactérien LexA, la protéine hybride active la transcription d'un gène de levure dont l'UAS_G est remplacée par l'opérateur de LexA [1]. Des résultats identiques sont obtenus avec

le facteur GCN4. Dans une variante de cette expérience, la protéine GCN4 dans son ensemble est fusionnée avec le domaine de liaison à l'ADN de la protéine LexA. L'hybride alors formé, qui porte deux domaines différents de fixation à l'ADN, est capable d'activer la transcription, aussi bien à partir d'un opérateur LexA que d'un site de reconnaissance à GCN4 [2]. L'ensemble de ces observations indique que les domaines d'activation et de liaison à l'ADN sont non seulement distincts, mais également remarquablement indépendants.

Le clonage et l'analyse des ADNc de plusieurs facteurs de transcription chez la levure, la drosophile, les plantes et les vertébrés montrent que beaucoup sont organisés en modules structuraux, et plusieurs domaines fonctionnels ont été décrits : le domaine de liaison à l'ADN, le

ADRESSE ET TIRÉS À PART

P.-O. Angrand : attaché temporaire d'enseignement et de recherche à l'université de Paris XI. Unité de génétique de la différenciation, département de biologie moléculaire, Institut Pasteur, 25, rue du Docteur-Roux, 75724 Paris Cedex 15, France.

RÉFÉRENCES

1. Brent R, Ptashne M. A eukaryotic transcriptional activator bearing the DNA specificity of a prokaryotic repressor. *Cell* 1985 ; 43 : 729-36.
2. Hope IA, Struhl K. Functional dissection of a eukaryotic transcriptional activator protein, GCN4 of yeast. *Cell* 1986 ; 46 : 885-94.
3. Mitchell PJ, Tjian R. Transcriptional regulation in mammalian cells by sequence-specific DNA binding proteins. *Science* 1989 ; 245 : 371-8.
4. Frankel AD, Kim PS. Modular structure of transcription factors : implication for gene regulation. *Cell* 1991 ; 65 : 717-9.
5. Harrison SC. A structural taxonomy of DNA-binding domains. *Nature* 1991 ; 353 : 715-9.
6. Anderson WF, Ohlendorf DH, Takeda Y, Matthews BW. Structure of the cro repressor from bacteriophage λ and its interaction with DNA. *Nature* 1981 ; 290 : 754-8.
7. McKay DB, Steitz TA. Structure of catabolite gene activator protein at 2.9 Å resolution suggests binding to left-handed B-DNA. *Nature* 1981 ; 290 : 744-9.
8. Pabo CO, Lewis M. The operator-binding domain of λ repressor : structure and DNA recognition. *Nature* 1982 ; 298 : 443-7.
9. Harrison SC, Aggarwal AK. DNA recognition by proteins with the helix-turn-helix motif. *Annu Rev Biochem* 1990 ; 59 : 933-69.
10. Gehring WJ. Homeo boxes in the study of development. *Science* 1987 ; 236 : 1245-52.
11. Scott MP, Tamkun JW, Hartzell III GW. The structure and function of the homeodomain. *Biochim Biophys Acta* 1989 ; 989 : 25-48.
12. Laughton A. DNA binding specificity of homeodomains. *Biochemistry* 1991 ; 30 : 11357-67.
13. Affolter M, Schier A, Gehring WJ. Homeodomain proteins and the regulation of gene expression. *Curr Opin Cell Biol* 1990 ; 2 : 485-95.
14. Qian YQ, Billeter M, Otting G, et al. The structure of the *Antennapedia* homeodomain determined by NMR spectroscopy in solution : comparison with prokaryotic repressors. *Cell* 1989 ; 59 : 573-80.
15. Kissinger CR, Liu B, Martin-Blanco E, et al. Crystal structure of an engrailed homeodomain-DNA complex at 2.8 Å resolution : a framework for understanding homeodomain-DNA interactions. *Cell* 1990 ; 63 : 579-90.

domaine d'activation de la transcription et, dans de nombreux cas, un domaine de dimérisation ou de liaison à un ligand [3, 4]. Dans cet article, nous discuterons la structure des domaines de liaison à l'ADN rencontrés chez les facteurs de transcription eucaryotes (pour une revue [5]).

Structure en hélice-coude-hélice et homéoprotéines

Au début des années 1980, la cristallographie révélait la structure tridimensionnelle du domaine de liaison à l'ADN des premières protéines régulatrices : le répresseur λ Cro [6], le répresseur λ CI [7] et la protéine CAP d'*E. coli* [8]. Ces protéines prokaryotiques adoptent, dans leur domaine de liaison à l'ADN, une structure voisine en motif nommé hélice-coude-hélice (*helix-turn-helix*) parce que constitué de deux hélices α , de 8-10 acides aminés chacune, séparées par 3 acides aminés formant un coude ([5] et voir *figure 1A*). L'hélice en position C-terminale, dite de reconnaissance, s'ajuste dans le grand sillon de l'ADN grâce au soutien de l'hélice N-terminale ([5, 9] et voir *m/s n° 7, vol. 3, p. 428*).

Lorsque les premiers gènes homéotiques et de segmentation furent clonés et caractérisés chez la drosophile, plusieurs d'entre eux (*ftz, Antp, Ubx*) révélèrent posséder un domaine d'environ 180 pb (60 acides aminés) très similaire (80-90 % d'acides aminés en commun), et baptisé domaine homéo [10]. Aujourd'hui, plus d'une soixantaine de gènes régulateurs qui possèdent un domaine homéo sont connus chez la drosophile. Ces protéines sont regroupées dans une superfamille ramifiée, dont les membres les plus éloignés ne présentent guère plus de 20 % de résidus identiques [11]. Chaque branche de cette famille d'homéoprotéines est conservée au cours de l'évolution, de la drosophile aux vertébrés (protéines Hox des mammifères, Xhox du xénope) et l'existence des homéoprotéines n'est pas limitée aux organismes segmentés puisqu'on les trouve également chez la levure, chez l'oursin de mer, chez le nématode *C. elegans* et même chez les plantes [11, 12].

Rapidement, les similitudes de séquences et les prédictions de structures secondaires ont suggéré que le domaine homéo renfermait un motif hélice-coude-hélice analogue au domaine de liaison à l'ADN des activateurs et répresseurs prokaryotes [12]. Il est maintenant établi que les homéoprotéines se lient à l'ADN pour contrôler la transcription [13]. De plus, la détermination de la structure tridimensionnelle des domaines homéo des protéines *Antennapedia* (*Antp* [14]), *Engrailed* (*En* [15]) et *MAT α 2* [16] démontre que l'homologie entre le domaine homéo et le motif hélice-coude-hélice s'étend à la structure tertiaire ; les hélices 2 et 3 du domaine homéo se superposent globalement aux hélices de soutien et de reconnaissance des protéines hélice-coude-hélice des prokaryotes (*figure 1A*).

La résolution des interactions ADN-protéine à domaine homéo, par cristallographie aux rayons X (*En* [15], *MAT α 2* [16]) et par RMN (*Antp* [17]), montre que l'hélice de reconnaissance (l'hélice 3) du domaine homéo se positionne dans le grand sillon de l'ADN et interagit avec certaines bases du site de liaison (*figure 1B*). Un bras N-terminal (du côté de l'hélice 1, voir *figure 1A*) se fixe dans le petit sillon de l'ADN et interagit avec les bases adjacentes à celles reconnues par l'hélice 3 (*figure 1B*).

Les homéoprotéines se lient généralement à des séquences cibles dont le cœur est constitué du motif ^{5'}TAAT^{3'} [12, 18]. Les résultats des études génétiques [19, 20] et structurales [15, 17] montrent que le résidu 9 de l'hélice de reconnaissance (résidu 50 du domaine homéo) interagit spécifiquement avec les groupements fonctionnels d'une base située en 3' du cœur ^{5'}TAAT^{3'} (*figure 1B*). Cette interaction est particulièrement importante car elle permet à un domaine homéo donné de distinguer son site de liaison spécifique sur l'ADN [12, 18] (voir *figure 1C*).

Les domaines conservés associés aux domaines homéo

Certaines homéoprotéines possèdent d'autres motifs très conservés qui se

révèlent être impliqués dans la reconnaissance de séquences d'ADN situées au voisinage des séquences reconnues par le domaine homéo.

Le domaine Prd

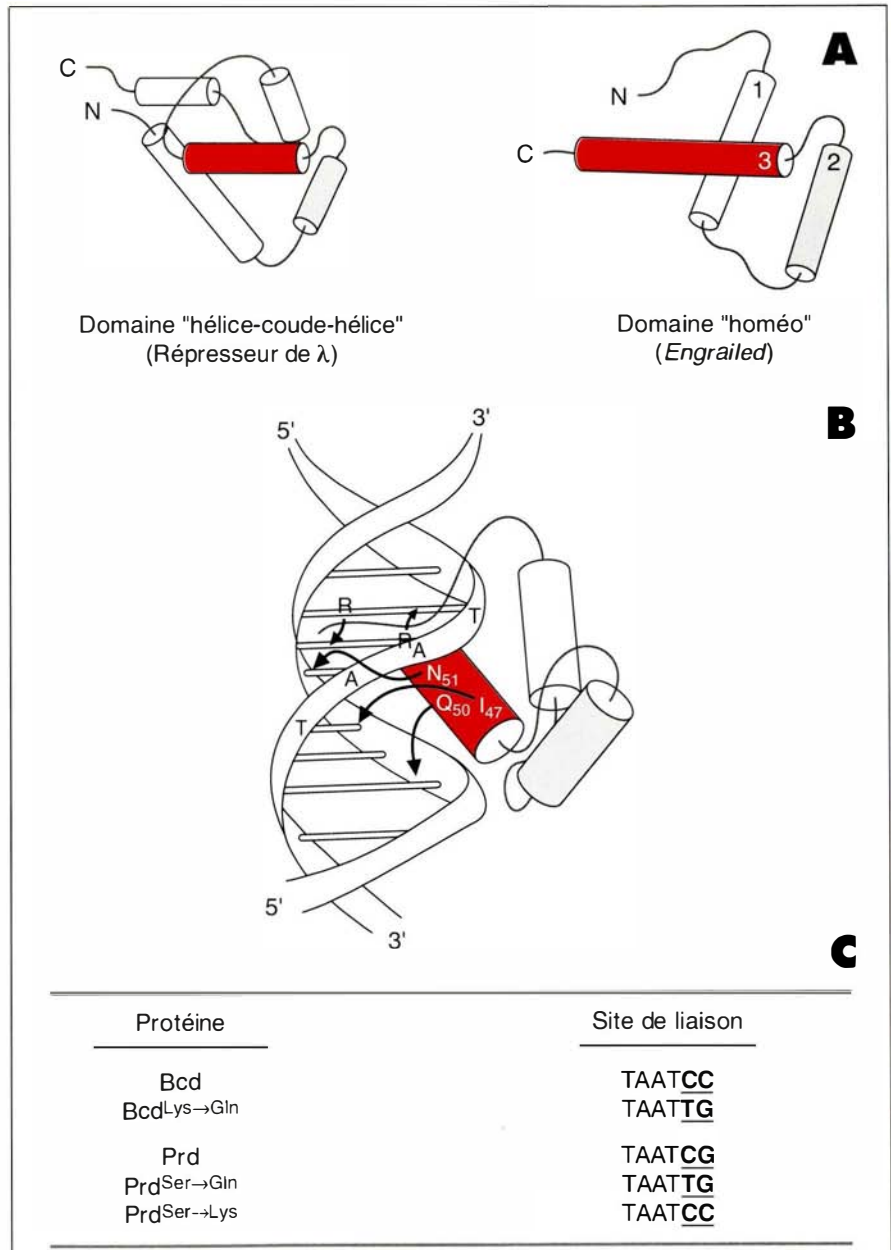
Dans l'homéoprotéine Paired (Prd), une région de 128 acides aminés constitue le domaine Prd [21]; ce domaine très conservé est également présent dans d'autres homéoprotéines de drosophile telles que celles du *locus gooseberry*, Gsb-p et Gsb-d, ou dans des protéines régulatrices de mammifère. Le domaine Prd agit de façon indépendante ou par coopération intramoléculaire pour augmenter la spécificité de l'homéoprotéine à l'ADN [22].

Le domaine POU

Le domaine POU est un motif de 135 acides aminés initialement identifié chez les facteurs de transcription de mammifère Pit-1, Oct-1, Oct-2 et dans la protéine Unc-86 qui est impliquée dans la différenciation des neurones adultes et de plusieurs lignages neuroblastiques chez *C. elegans* ([23] et voir *m/s* n° 3, vol. 5, p. 172). Ce domaine est constitué de deux régions : un sous-domaine spécifique (POU_s), et dans la moitié C-terminale un domaine homéo (POU_h). Celui-ci définit une nouvelle famille, éloignée de toutes celles repérés jusque-là (au mieux 21 résidus identiques sur 60 et, fait unique pour les homéoprotéines, l'acide aminé 9 de l'hélice de reconnaissance est une cystéine). Le domaine homéo divergent de Pit-1 ou Oct-2 se lie aux séquences du type 5'TAAT^{3'}, précédées par un motif 5'ATGC^{3'}, qui est reconnu par le sous-domaine POU_s adjacent [24, 25]. Outre ce rôle dans la reconnaissance de l'ADN, le domaine POU semble impliqué dans d'autres fonctions biologiques telles que la formation d'homodimères ou d'hétérodimères entre les membres de la famille des protéines à domaine POU, l'interaction avec d'autres protéines comme la protéine activatrice VP-16 du HSV, ou l'activation de la transcription.

La répétition LIM

Dans certaines protéines, le domaine homéo se trouve associé à un motif



RÉFÉRENCES

16. Wolberger C, Vershon AK, Liu B, *et al.* Crystal structure of a MAT α 2 homeodomain-operator complex suggests a general model for homeodomain-DNA interactions. *Cell* 1991 ; 67 : 517-28.
17. Otting G, Qian YQ, Billeter M, *et al.* Protein-DNA contacts in the structure of a homeodomain-DNA complex determined by nuclear magnetic resonance spectroscopy in solution. *EMBO J* 1990 ; 9 : 3085-92.
18. Treisman J, Harris E, Wilson D, Desplan C. The homeodomain : a new face for the helix-turn-helix ? *BioEssays* 1992 ; 14 : 145-50.
19. Hanes SD, Brent R. DNA specificity of the bicoid activator protein is determined by homeodomain recognition helix residue 9. *Cell* 1989 ; 57 : 1275-83.
20. Treisman J, Gönczy P, Vashishtha M, *et al.* A single amino acid can determine the DNA binding specificity of homeodomain proteins. *Cell* 1989 ; 59 : 553-62.
21. Bopp D, Burri M, Baumgartner S, *et al.* Conservation of a large protein domain in the segmentation gene *paired* and in functionally related genes of *Drosophila*. *Cell* 1986 ; 47 : 1033-40.
22. Treisman J, Harris E, Desplan C. The *paired* box encodes a second DNA-binding domain in the *Paired* homeo domain protein. *Genes Dev* 1991 ; 5 : 594-604.
23. Herr W, Sturm RA, Clerc RG, *et al.* The POU domain : a large conserved region in the mammalian *pit-1*, *oct-1*, *oct-2*, and *Caenorhabditis elegans unc-86* gene products. *Genes Dev* 1988 ; 2 : 1513-6.
24. Kristic TM, Sharp PA. Interactions of the Oct-1 POU subdomains with specific DNA sequences and with the HSV α -transactivator protein. *Genes Dev* 1990 ; 4 : 2383-96.
25. Verrijzer CP, Alkema MJ, van Weperen WW, *et al.* The DNA binding specificity of the bipartite POU domain and its subdomains. *EMBO J* 1992 ; 11 : 4993-5003.
26. Freyd G, Kim SK, Horvitz HR. Novel cysteine-rich motif and homeodomain in the product of the *Caenorhabditis elegans* cell lineage gene *lin-11*. *Nature* 1990 ; 344 : 876-9.
27. Karlsson O, Thor S, Norberg T, *et al.* Insulin gene enhancer binding protein Isl-1 is a member of a novel class of proteins containing both a homeo- and a Cys-His domain. *Nature* 1990 ; 344 : 879-82.
28. 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.
- riche en cystéine/histidine, le domaine LIM (pour *Lin-11*, *Isl-1*, *Mec-3* ; les trois premiers membres de la famille). Le gène *lin-11* est impliqué dans le développement de la vulve hermaphrodite chez *C. elegans*. *Isl-1* est une protéine qui se fixe sur le *enhancer* du gène I de l'insuline chez le rat. Le gène *mec-3* a aussi été décrit chez *C. elegans* ; il est impliqué dans la différenciation des neurones mécanosensoriels [26, 27]). La structure du domaine LIM est la suivante :
- $$C-X_2-C-X_{17-19}-H-X_2-C-X_2-C-X_2-C-X_{7-11}-(C)-X_8-C$$
- Ce motif constitué de résidus cystéine séparés par deux acides aminés (C-X₂-C) est souvent associé aux protéines capables de fixer des ions métalliques. Sur cette base, il est probable que les répétitions LIM soient impliquées dans des coordinations métalliques. Cependant, le domaine LIM diffère très nettement des motifs consensus de fixation des métaux déjà décrits pour les protéines en doigt à zinc (Zn) de la classe TFIIIA ou de la famille des récepteurs aux hormones stéroïdes (voir plus loin). La fonction des répétitions LIM n'est pas encore connue, mais il est probable que cette structure, par des interactions avec l'ADN, puisse augmenter la spécificité de liaison à l'ADN. Alternativement, le domaine LIM pourrait interagir avec d'autres protéines de l'appareil de transcription ou du cytosquelette.

Les domaines en doigt à zinc

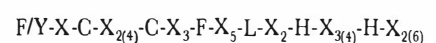
TFIIIA, une protéine de 40 kDa, interagit avec 50 pb du gène 5S, et forme également un complexe stable avec son produit, l'ARN 5S. Ces propriétés singulières, ainsi que de nombreuses autres études structurales et fonctionnelles, ont conduit Klug *et al.* à proposer un motif structural radicalement nouveau pour la reconnaissance de l'ADN ([28] et voir *m/s n° 7, vol. 3, p. 428*). Les 30 kDa de TFIIIA qui lient l'ADN sont constitués de la répétition avec variation, neuf fois en tandem, d'un petit domaine de 30 acides aminés. Chacun de ces petits domaines se replie indépendamment autour d'un atome de zinc que coordonnent deux cystéines et deux histidines, et interagit

indépendamment avec l'ADN *via* les résidus polaires et basiques présents dans la boucle formée. La protéine ne contacte l'ADN que d'un seul côté. Chaque doigt à Zn interagit avec 5 pb dans le grand sillon, contiguës aux 5 pb reconnues par le doigt suivant, si bien que les neuf doigts à Zn couvrent bien environ 50 pb. Ce modèle structural a été pour partie confirmé, mais, surtout, le doigt à Zn apparaît maintenant comme un motif universel, largement répandu parmi les protéines eucaryotes ayant de l'affinité pour l'ADN. En dépit de quelques similarités dans leur structure primaire, plusieurs classes de motifs en doigt à Zn sont décrites ([5] et *m/s n° 10, vol. 4, p. 624*). Les structures tridimensionnelles de ces motifs sont très différentes, mais tous possèdent une même fonction : permettre l'exposition d'une hélice α (de reconnaissance) dans le grand sillon de l'ADN.

La famille C₂-H₂ (classe 1)

Dans la classe des protéines à doigt C₂-H₂, les doigts à Zn sont très semblables à ceux de TFIIIA, sur le plan de la taille, de l'écartement, et des résidus invariants en plus des deux cystéines et des deux histidines ; c'est le nombre des doigts qui varie : de 2 à... 37 (!). Ces doigts C₂-H₂ sont trouvés dans de nombreux facteurs de transcription de mammifère (Sp1), de levure (ADR1, SWI5) ou dans le produit de plusieurs gènes impliqués dans le développement chez la drosophile — gènes de segmentation (*krüppel*, *hunchback*), de polarité dorso-ventrale (*snail*). Mais beaucoup d'autres gènes comportant des doigts de ce type ont été repérés chez la souris et l'homme, notamment par l'utilisation de l'hybridation croisée à l'aide d'une sonde *krüppel*.

L'analyse de plus de 200 motifs en doigt C₂-H₂ révèle la séquence consensus suivante :



La structure tridimensionnelle du motif C₂-H₂ a été établie par RMN [29, 30] et par cristallographie [31] (*figure 2B*) : les résidus 1-10 forment une épingle antiparallèle. La Cys₃ se trouve dans le premier brin β , et la Cys₆ est localisée dans la boucle

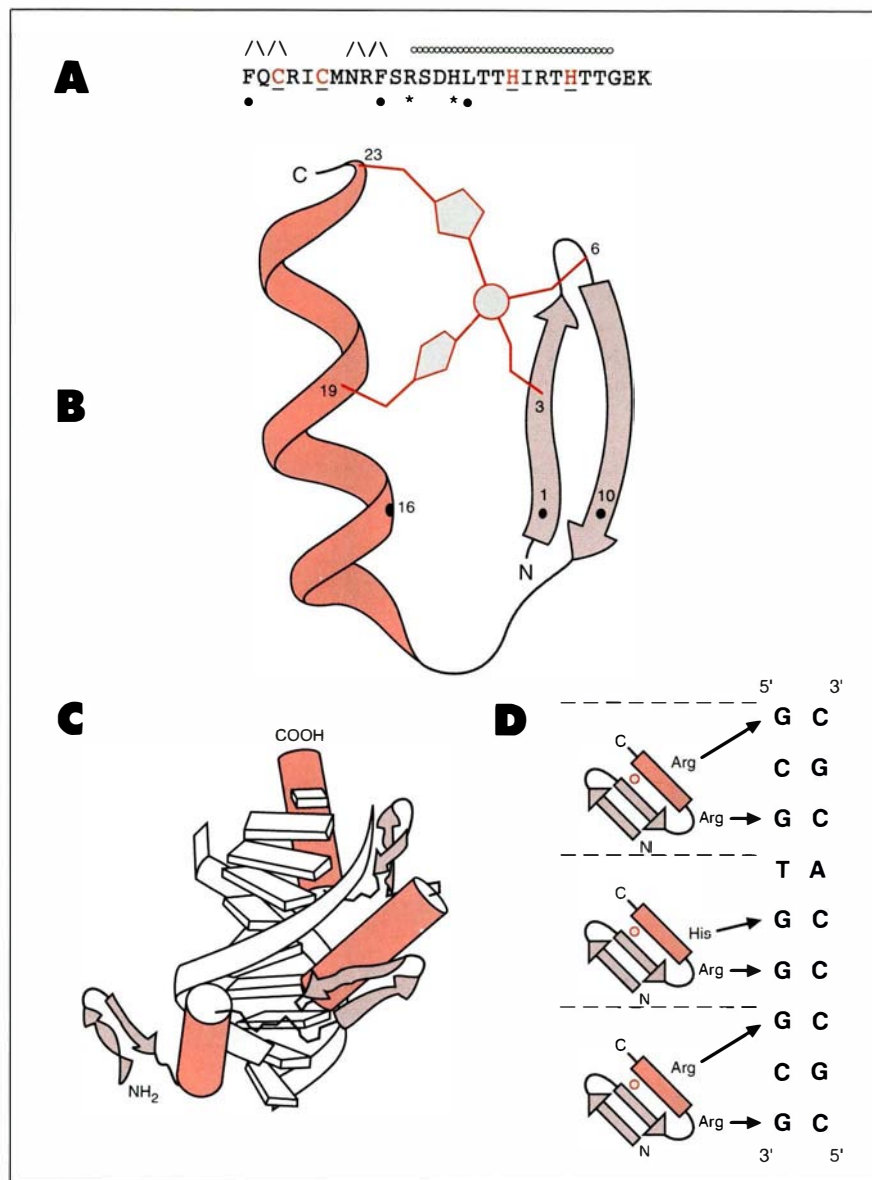


Figure 2. **Le motif de doigt à zinc de type C₂-H₂.** **A.** Séquence en acides aminés du deuxième doigt à Zn de la protéine Zif268 de xénope. Les acides aminés impliqués dans les structures en brin β sont notés \wedge ; ceux qui adoptent une structure en hélice α sont marqués par ∞ . Les résidus Cys₃, Cys₆, His₁₉, et His₂₃ responsables de la coordination au Zn sont en rouge soulignés. La position des acides aminés qui forment le cœur hydrophobe (Phe₁, Phe₁₀ et Leu₁₆), et de ceux qui interagissent avec les bases (Arg₁₂ et His₁₅) est indiquée respectivement par • et * (d'après [31]). **B.** Structure tridimensionnelle schématisée du motif de doigt à Zn de type C₂-H₂. Les résidus du cœur hydrophobe sont indiqués comme en **A**. L'hélice (de reconnaissance) est en rouge et les brins β apparaissent en bistre. **C.** Diagramme schématisé du complexe formé entre le domaine de liaison à l'ADN de la protéine Zif268 (qui possède trois doigts à Zn de type C₂-H₂) et sa séquence cible (tiré de [31]). Les hélices de reconnaissance de chaque doigt C₂-H₂ s'enfoncent dans le grand sillon de l'ADN, et le motif s'enroule autour de la double hélice. **D.** Résumé en projection, des interactions entre Zif268 et les bases de l'ADN. Tous les doigts C₂-H₂ interagissent avec le même brin d'ADN, et l'unité élémentaire de reconnaissance par un doigt est un groupe de trois bases (adapté de [31]).

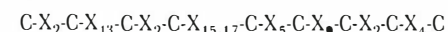
m/s n° 6-7 vol. 9, juin-juillet 93

située entre les deux brins β . Cette épingle β est suivie par une hélice (résidus 12-24) d'environ trois tours et demi, dans laquelle se trouvent les deux autres ligands du Zn : His₁₉ et His₂₃. L'hélice est maintenue contre l'épingle β en une structure miniglobulaire à la fois par la coordination au Zn des résidus Cys et His, et par des interactions hydrophobes (entre les résidus Tyr/Phe₁, Phe₁₀ et Leu₁₆). Enfin, en accord avec les études par mutagenèse dirigée [32, 34], les chaînes latérales des acides aminés basiques et polaires de l'hélice, orientées vers l'extérieur de la structure, sont responsables de la spécificité de liaison à l'ADN, dans le grand sillon [31]. Les interactions se font alors avec 3 pb (et non 5 pb, comme proposé par Klug *et al.*), et la protéine s'enroule autour de l'ADN, de sorte que les 3 pb contigus sont reconnus par l'hélice du doigt suivant ([31] et voir figures 2C et 2D).

Ainsi, le motif structural de base ne reconnaît que 3 pb, mais il est répété n fois avec des variantes au sein du même polypeptide, permettant la reconnaissance de motifs de longueurs variées et pas forcément symétriques, car un monomère peut suffire à l'interaction (c'est le cas de TFIIIA et de Sp1). Pour TFIIIA, la présence de neuf doigts C₂-H₂ est sans doute essentielle à sa stabilité sur le gène 5S lors du passage de l'ARN pol III. En revanche, les facteurs de transcription de l'ARN pol II comme Sp1 se fixent plutôt en amont des gènes dans des régions sans nucléosome, si bien que deux ou trois doigts C₂-H₂ peuvent suffire pour une fixation stable et spécifique à l'ADN.

La famille C₄ (classe 2)

La deuxième classe de doigt à Zn correspond à un domaine d'environ 70 acides aminés présents dans les récepteurs nucléaires aux hormones stéroïdes (et aux molécules apparentées) (*m/s* n° 3, vol. 3, p. 172 ; n° 3, vol. 4, p. 196 ; n° 2, vol. 8, p. 156). Ce motif, capable de se lier aux boîtes HRE (*hormone responsive elements*), présente la structure suivante :



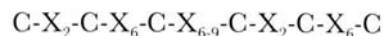
Dans ce cas, deux atomes de Zn sont chélatés, chacun par quatre résidus cystéine, ce qui définit en fait deux

RÉFÉRENCES

29. Lee MS, Gippert GP, Soman KV, *et al.* Three-dimensional solution structure of a single zinc finger DNA-binding domain. *Science* 1989 ; 245 : 635-7.
30. Klevit RE, Herriott JR, Horvath SJ. Solution structure of a zinc finger domain of yeast ADR1. *Proteins* 1990 ; 7 : 215-26.
31. Pavletich NP, Pabo CO. Zinc finger-DNA recognition : crystal structure of Zif268-DNA complex at 2.1 Å. *Science* 1991 ; 252 : 809-17.
32. Thukral SK, Morrison ML, Young ET. Alanine scanning site-directed mutagenesis of the zinc fingers of transcription factor ADR1 : residues that contact DNA and that transactivate. *Proc Natl Acad Sci USA* 1991 ; 88 : 9188-92.
33. Thukral SK, Morrison ML, Young ET. Mutations in the zinc fingers of ADR1 that change the specificity of DNA binding and transactivation. *Mol Cell Biol* 1992 ; 12 : 2784-92.
34. Nardelli J, Gibson TJ, Vesque C, Charnay P. Base sequence discrimination by zinc-finger DNA-binding domains. *Nature* 1991 ; 349 : 175-8.
35. Laudet V, Hänni C, Coll J, *et al.* Evolution of the nuclear receptor gene superfamily. *EMBO J* 1992 ; 11 : 1003-13.
36. Härd T, Kellenbach E, Boelens R, *et al.* Solution structure of the glucocorticoid receptor DNA-binding domain. *Science* 1990 ; 249 : 157-60.
37. Luisi BF, Xu WX, Otwinowski Z, *et al.* Crystallographic analysis of the interaction of the glucocorticoid receptor with DNA. *Nature* 1991 ; 352 : 497-505.
38. Schwabe JWR, Neuhaus D, Rhodes D. Solution structure of the DNA-binding domain of the oestrogen receptor. *Nature* 1990 ; 348 : 458-61.
39. Hollenberg SM, Evans RM. Multiple and cooperative *trans*-activation domains of the human glucocorticoid receptor. *Cell* 1988 ; 55 : 899-906.
40. Danielsen M, Hinck L, Ringold GM. Two amino acids within the knuckle of the first zinc finger specify DNA response element activation by the glucocorticoid receptor. *Cell* 1989 ; 57 : 1131-8.
41. Mader S, Kumar V, de Verneuil H, Chambon P. Three amino acids of the oestrogen receptor are essential to its ability to distinguish an oestrogen from a glucocorticoid-responsive element. *Nature* 1989 ; 338 : 271-4.
42. Umesono K, Evans RM. Determinants of target gene specificity for steroid/thyroid hormone receptors. *Cell* 1989 ; 57 : 1139-46.
- doigts à Zn organisés en deux modules dans le motif. Une cinquantaine de protéines à doigt de type C₄ sont décrites chez les mammifères, le poulet, le xénope ou la drosophile [35] ; toutes possèdent ce motif organisé en deux modules de doigts C₄ (*figure 3A*).
- Les déterminations de la structure tridimensionnelle des doigts C₄ du récepteur des glucocorticoïdes (GR [36, 37]) et du récepteur des œstrogènes (ER [38]) révèlent une conformation du domaine radicalement différente de celle du doigt C₂-H₂. Chacun des deux modules du motif C₄ adopte une conformation globalement similaire, constitué d'une boucle suivie d'une hélice α amphipatique de 11-13 résidus. Chaque ion Zn est lié par deux cystéines situées au début de la boucle, et par les deux autres localisées dans la partie N-terminale de l'hélice α, formant ainsi une structure en double boucle Zn-hélice [5]. Les deux hélices (des modules 1 et 2) s'organisent selon une structure en X, compacte, maintenue par des interactions hydrophobes (une dizaine de résidus sont impliqués dans la formation du cœur hydrophobe qui maintient la structure globale [37, 38]) (*figure 3B*). Conformément aux études par mutagenèse dirigée [39-42], l'hélice α du module 1 est responsable de la liaison à l'ADN [37]. Les résidus de l'hélice, tournés vers l'extérieur du motif, contactent l'ADN dans le grand sillon (*figure 3C*).
- Deux déterminants sont responsables de la spécificité de l'interaction ADN-protéine à doigt C₄ : (1) le premier passe par une interaction spécifique entre l'hélice α du module 1 et le HRE ; (2) le deuxième passe par une interaction protéine-protéine spécifique qui se traduit par une interaction ADN-protéine spécifique dépendante de l'espacement des deux demi-sites du HRE.
- Trois acides aminés de la région N-terminale de l'hélice α du module 1 (domaine P, voir *figures 3A et 3B*) sont responsables de la spécificité de reconnaissance de différents HRE ; la substitution de ces trois résidus dans le récepteur des œstrogènes (ER) (Glu, Gly et Ala) par les résidus correspondants du récepteur des glucocorticoïdes (GR) (Gly, Ser et Val) produit une protéine incapable de reconnaître les ERE (éléments de réponse aux œstrogènes - ⁵AGGTCAXXXTGACCT³), mais capable de se lier aux GRE (éléments de réponse aux glucocorticoïdes - ⁵AGAACAAXXXTGTCT³) [41]. De même, la substitution de ces mêmes acides aminés modifie la spécificité d'interaction à l'ADN du GR et du T₃R (récepteur des hormones thyroïdiennes) [40, 42].
- Un second groupe de résidus (les cinq acides aminés d'une région appelé domaine D, voir *figures 3A et 3C*), localisés dans la boucle Zn du module 2, est impliqué dans la discrimination de l'espacement entre les deux demi-sites des HRE. Le remplacement du domaine D du ER par celui du T₃R rend le ER capable de reconnaître la séquence T₃REp (T₃REp = ERE sans les trois bases centrales du motif, voir *figure 3D*) [42, 43]. Les études biochimiques révèlent que les récepteurs des hormones stéroïdes se lient sous forme de dimères sur les HRE palindromiques et qu'un des déterminants de la dimérisation se trouve associé au doigt C₄ [44]. Comme le montrent les analyses de la structure tridimensionnelle des GR et ER, le domaine D (qui forme une boucle à la surface de la molécule, se trouve à 15-20 Å de l'hélice de reconnaissance du module 1) est responsable de l'interaction entre les deux monomères ; les hélices α des modules 1 de chaque monomère se positionnent alors dans les sillons adjacents de l'ADN, séparés par 34 Å [37, 38] (*figure 3C*).

La famille C₆ (classe 3)

Tous les membres de la famille des protéines à doigt de type C₆ sont des activateurs de transcription chez la levure (GAL4, HAP1, LEU3, PPR1...). Parmi ces protéines, GAL4 est le facteur le mieux caractérisé ; il se fixe à l'ADN sous forme dimérique sur une séquence palindromique. Le domaine de liaison à l'ADN est localisé dans la région N-terminale de la protéine GAL-4 : les résidus 1-74 sont suffisants à la reconnaissance de l'ADN [45] (*figure 4A*, p. 728). Ce domaine possède un motif dont la structure est la suivante :



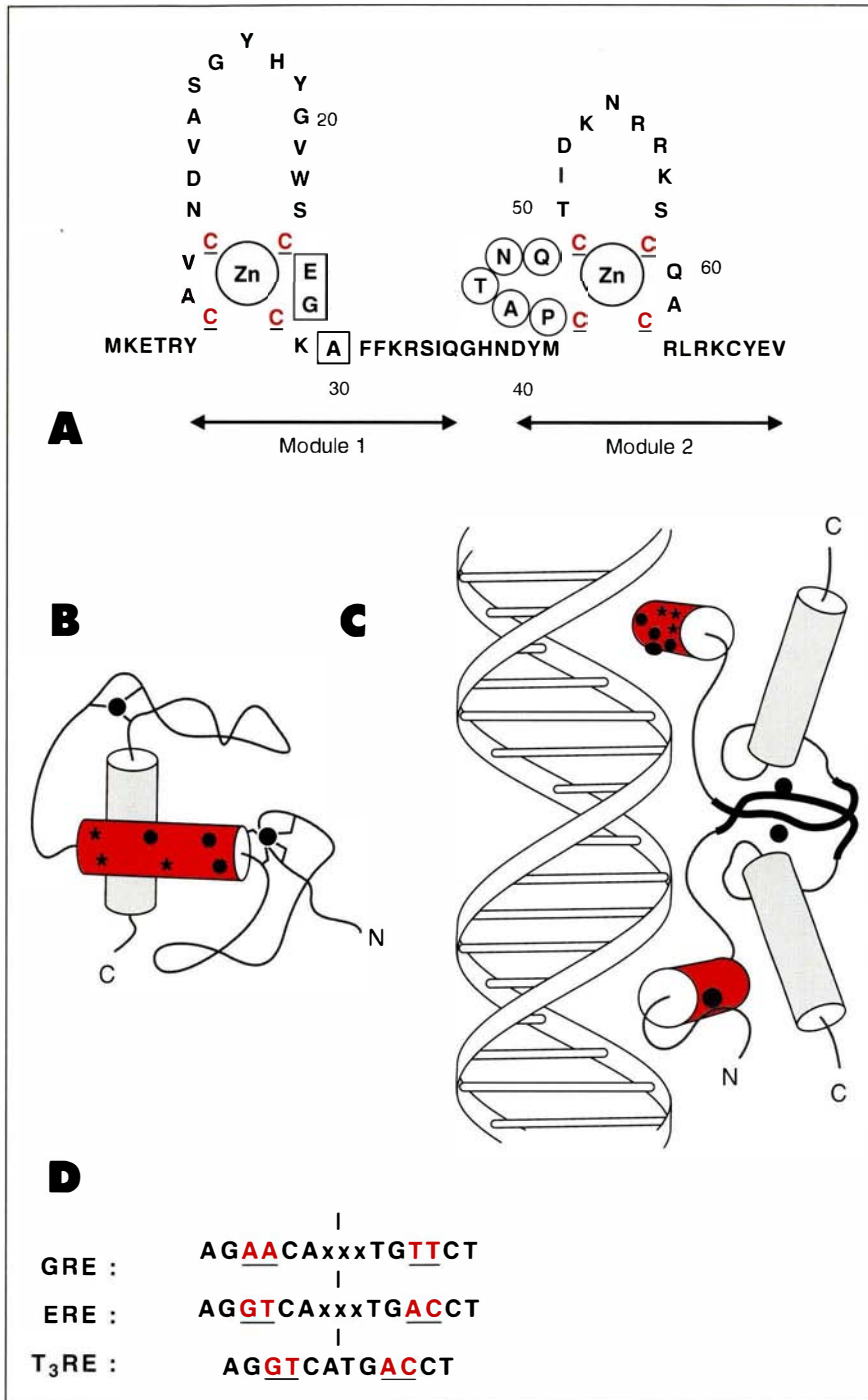


Figure 3. **Le motif de doigt à zinc de type C₄.** **A.** Séquence en acides aminés des 70 premiers résidus du domaine de liaison à l'ADN du récepteur des œstrogènes (ER). Les résidus Cys responsables de la coordination au Zn sont soulignés. Les acides aminés des domaines P et D sont en caractères gras, respectivement encadrés et cerclés. Le domaine de liaison à l'ADN en doigt à Zn de type C₄ s'organise en deux modules. **B.** Représentation schématique de la structure tridimensionnelle du domaine de liaison à l'ADN du ER, en double boucle-Zn-hélice (d'après [38]). L'hélice de reconnaissance du premier module (en rouge) et l'hélice de soutien du deuxième module (en gris) adoptent une structure en X maintenue par des interactions hydrophobes. Les trois acides aminés du domaine P responsables de la discrimination entre les différents éléments de réponse aux hormones (HRE) (Glu₂₅, Gly₂₆ et Ala₂₉) sont indiqués par •. Sur la même face de l'hélice, les résidus basiques Lys₂₈, Lys₃₂ et Arg₃₃, conservés dans les membres de cette classe de facteurs de transcription sont indiqués par *. Dans le cas du récepteur des glucocorticoïdes (GR), il a été montré que deux de ces résidus (Lys₂₈ et Arg₃₃) sont responsables des interactions spécifiques avec les bases du grand sillon de l'ADN [37]. **C.** Modèle de l'interaction entre deux monomères de récepteur et le site HRE palindromique. Les hélices de reconnaissance des modules 1 se positionnent dans les deux grands sillons d'ADN adjacents, séparés par 34 Å (d'après [37, 38]). Le domaine D, à l'interface entre chaque monomère, est en trait épais. **D.** Les éléments palindromiques de réponse aux hormones (HRE). Les éléments de réponse aux glucocorticoïdes (GRE), aux œstrogènes (ERE) et aux hormones thyroïdiennes (T₃RE) sont montrés. Le GRE et l'ERE diffèrent par seulement deux paires de bases dans chaque demi-site. En revanche, le T₃RE diffère de l'ERE par l'espacement des deux demi-sites.

Ce motif contient six cystéines qui se lient à deux ions Zn selon une structure en *cluster* analogue à celle rencontrée dans les métallothionéines. Récemment, la structure tridimensionnelle du doigt de type C₆ de GAL4 a été établie par RMN [46, 47] et par cristallographie du com-

plexe doigt C₆ de GAL4-ADN [48]. Le polypeptide est organisé en deux petites hélices α (résidus 10-18 et 27-35) reliées par une boucle. Ces deux hélices adoptent une structure globulaire maintenue par les interactions entre les six cystéines et les deux ions Zn (figure 4B). Les chaînes

latérales des acides aminés situés dans la région C-terminale de la première hélice (Lys₁₇ et Lys₁₈) sont responsables de la reconnaissance spécifique du triplet ⁵CCG³ dans le grand sillon ([48] et voir figures 4B et 4C).

En outre, la dimérisation, par simple effet de positionnement du doigt à Zn

RÉFÉRENCES

43. Hirst MA, Hinck L, Danielsen M, Ringold GM. Discrimination of DNA response elements for thyroid hormone and estrogen is dependent on dimerization of receptor DNA binding domains. *Proc Natl Acad Sci USA* 1992 ; 89 : 5527-31.
44. Tsai SY, Carlstedt-Duke J, Weigel NL, et al. Molecular interactions of steroid hormone receptor with its enhancer element : evidence for receptor dimer formation. *Cell* 1988 ; 55 : 361-9.
45. Carey M, Kakidani H, Leatherwood J, et al. An amino-terminal fragment of GAL4 binds DNA as a dimer. *J Mol Biol* 1989 ; 209 : 423-32.
46. Baleja JD, Marmorstein R, Harrison SC, Wagner G. Solution structure of the DNA-binding domain of Cd₂-GAL4 from *S. cerevisiae*. *Nature* 1992 ; 356 : 450-3.
47. Kraulis PJ, Raine ARC, Gadhavi PL, Lauc ED. Structure of the DNA-binding domain of zinc GAL4. *Nature* 1990 ; 356 : 448-50.
48. Marmorstein R, Carey M, Ptashne M, Harrison SC. DNA recognition by GAL4 : structure of a protein-DNA complex. *Nature* 1992 ; 356 : 408-14.
49. Corton JC, Johnston SA. Altering DNA-binding specificity of GAL4 requires sequences adjacent to the zinc finger. *Nature* 1989 ; 340 : 724-7.
50. Witte MM, Dickson RC. The C6 zinc finger and adjacent amino acids determine DNA-binding specificity and affinity in the yeast activator proteins LAC9 and PPR1. *Mol Cell Biol* 1990 ; 10 : 5128-37.
51. Orkin SH. Globin gene regulation and switching : circa 1990. *Cell* 1990 ; 63 : 665-72.
52. Green LM, Berg JM. A retroviral Cys-Xaa₂-Cys-Xaa₄-His-Xaa₄-Cys peptide binds metal ions : spectroscopic studies and a proposed three-dimensional structure. *Proc Natl Acad Sci USA* 1989 ; 86 : 4047-51.
53. Summers MF, South TL, Kim B, Hare DR. High-resolution structure of an HIV zinc fingerlike domain via a new NMR-based distance geometry approach. *Biochemistry* 1990 ; 29 : 329-40.
54. Landschulz WH, Johnson PF, McKnight SL. The leucine zipper : a hypothetical structure common to a new class of DNA binding proteins. *Science* 1988 ; 240 : 1759-64.
55. Vison CR, Sigler PB, McKnight SL. Scissors-grip model for DNA recognition by a family of leucine zipper proteins. *Science* 1989 ; 246 : 911-6.

de type C₆, est également impliquée dans la spécificité de reconnaissance à l'ADN ; la protéine GAL4 se fixe à la séquence palindromique UAS_G (5'CGGN₁₁CCG3'), alors qu'une autre protéine, appartenant à la même classe de facteurs de transcrip-

tion à doigt à Zn de type C₆, PPR1, reconnaît la séquence UAS_I (5'CGGN₆CCG3'). Le remplacement de la région de liaison (*linker*, voir figure 4C) entre le doigt à Zn et le domaine de dimérisation de GAL4, par celui de PPR1 est responsable du

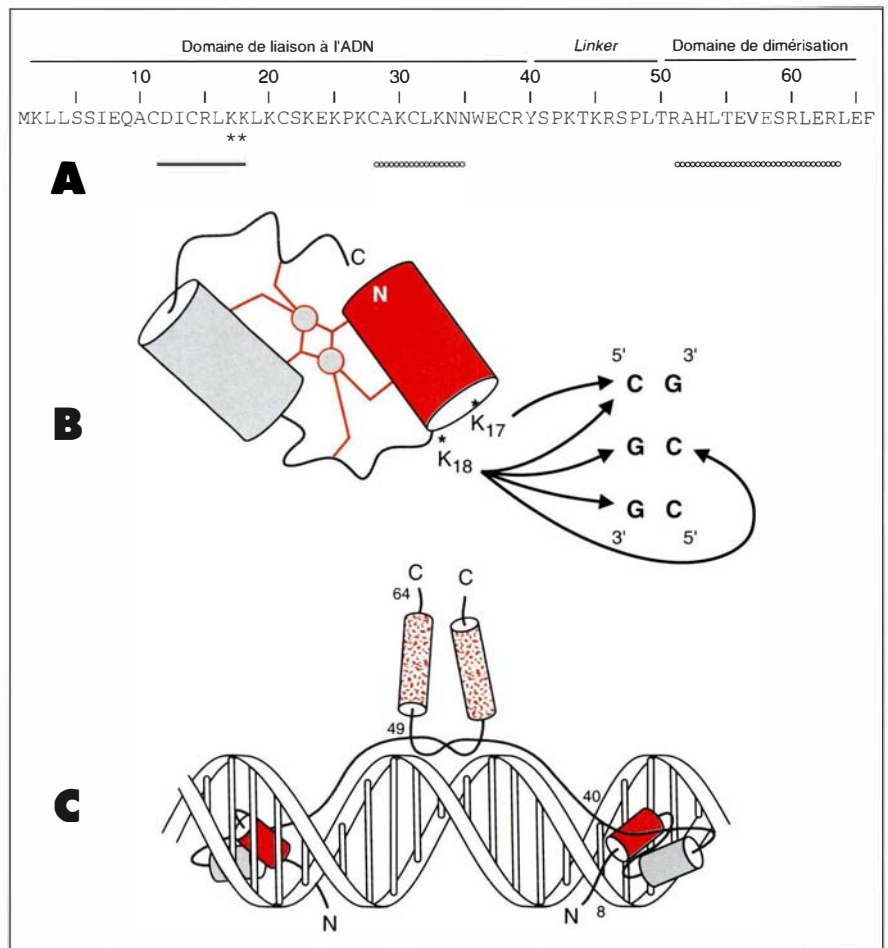


Figure 4. **Le motif de doigt à zinc de type C₆**. **A**. Séquence en acides aminés des 66 premiers résidus de la protéine GAL4 de *S. cerevisiae*. Les résidus Cys responsables de la coordination aux Zn sont soulignés. Les acides aminés qui interagissent avec les bases de l'ADN (Lys₁₇ et Lys₁₈) sont indiqués par ★. Les régions du domaine qui adoptent une structure en hélice α sont marquées par ∞. **B**. La structure tridimensionnelle du motif à doigt à Zn de type C₆ de GAL4 (résidus 1-40), est constituée de deux hélices maintenues par la coordination métallique (d'après [46, 47]). Les interactions spécifiques entre le doigt à Zn de type C₆ de GAL4 et la séquence 5'CCG3' se font par l'intermédiaire des résidus Lys₁₇ et Lys₁₈ situés dans la région C-terminale de la première hélice α (de reconnaissance, en rouge, [48]). **C**. Diagramme schématique du complexe formé entre un dimère de GAL4 et son site de liaison à l'ADN (UAS_G = 5'CGGN₁₁CCG3'). La position des acides aminés qui bordent les trois modules est indiquée : le domaine de liaison à l'ADN, résidus 8-40 ; la région de liaison (*linker*), résidus 40-49 ; le domaine de dimérisation, résidus 49-64. L'interaction ADN-protéine spécifique se fait dans le grand sillon de l'ADN. L'interaction entre les deux monomères se fait par une interaction en dorsade d'hélices (hélices marquées d'un pointillé rouge) (d'après [48]).

changement de la spécificité de liaison à l'ADN (UAS_C → UAS_L) [49, 50].

D'autres familles de doigt à Zn

Les protéines qui se fixent à la séquence GATA, comme le facteur de transcription hématopoïétique GATA-1 (*m/s n° 4, vol. 7, p. 385*), possèdent un motif de liaison à l'ADN en doigt à Zn apparemment différent des précédents et pourraient définir une nouvelle classe [51].

Enfin, une autre famille de doigt à Zn (de type C₂-H-C (C-X₂-C-X₄-H-X₄-C) [52]) trouvée dans le polypeptide GAG des rétrovirus est responsable de la liaison à l'ARN du génome viral. La structure tridimensionnelle de ce motif a été établie par RMN [53] : elle est proche de celle du doigt de type C₂-H₂, avec deux brins β antiparallèles, mais l'hélice α est remplacée par une large boucle. La nature des interactions entre le motif C₂-H-C et l'ARN est inconnue, mais remarquons que le facteur TFIIA (qui possède neuf doigts de type C₂-H₂) est également capable de se lier à l'ARN 5S.

Le domaine basique hélicoïdal

Un autre type d'élément de liaison à l'ADN a été originellement identifié par le domaine de dimérisation qui lui est associé (*leucine-zipper*, voir *m/s n° 5, vol. 8, p. 445* [54, 55]).

Les facteurs de transcription de la classe *Leu-zipper* sont caractérisés par la présence d'un domaine qui contient quatre à cinq leucines en répétition tous les sept acides aminés et qui est responsable de la dimérisation [54, 55]. Une région très basique, adjacente à la partie N-terminale de ce domaine de dimérisation, est impliquée dans la liaison de la protéine à l'ADN sur un site palindromique de 9-10 pb (*figure 5A*). La région basique de chaque monomère se lie, dans le grand sillon de l'ADN, à un demi-site. Les protéines à domaine *Leu-zipper* reconnaissent des motifs semblables, variant autour de la séquence 5'TGA³. Parallèlement à ce motif commun, un certain nombre de résidus conservés dans la région basique peuvent être identifiés. La périodicité de ces acides aminés conservés est d'environ 3,6 — soit la

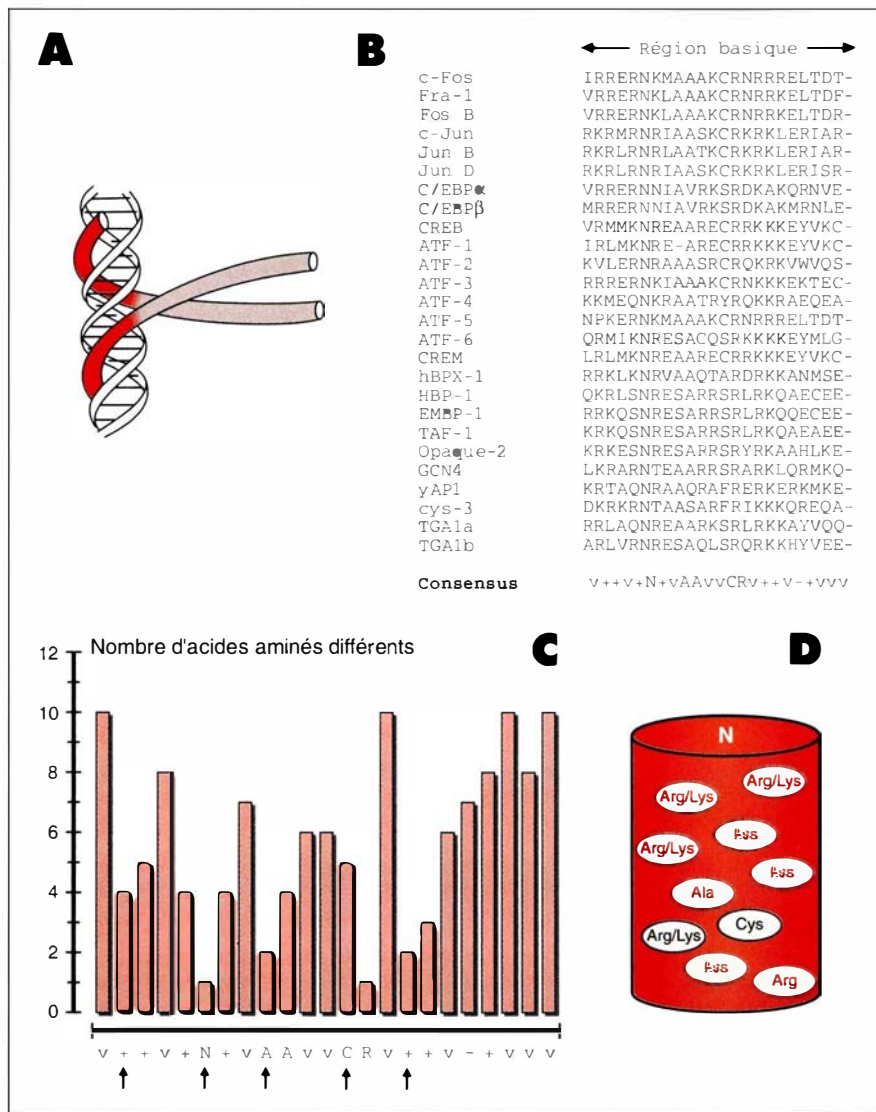


Figure 5. La région hélicoïdale basique des protéines de la classe bZIP. **A.** Modèle de l'interaction entre une protéine de la classe bZIP (sous forme dimérique) et l'ADN (selon [55]). La région basique est représentée en rouge et le domaine *Leu-zipper*, en bistre. La région hélicoïdale basique s'enroule dans le grand sillon de l'ADN. **B.** Séquence en acides aminés de la région basique de 26 protéines de la classe bZIP. La séquence consensus est indiquée : v, acide aminé variable ; +, acide aminé basique (Arg ou Lys) ; N, Asn ; A, Ala ; R, Arg ; -, acide aminé acide (Glu). **C.** Variabilité à chaque position du domaine basique. Le nombre d'acides aminés différents trouvés dans les 26 protéines bZIP, pour chaque position, est indiqué. Les flèches montrent la périodicité des acides aminés conservés, tous les 3,5 résidus (d'après [56]). **D.** Représentation graphique des résidus conservés selon une conformation en hélice α. Les résidus basiques et les autres résidus conservés (Asn - ... - Ala - Ala - ... - Cys/Ser) se trouvent sur la même face de l'hélice. (D'après [56].)

RÉFÉRENCES

56. O'Neil KT, Hoess RH, DeGrado WF. Design of DNA-binding peptides based on the leucine zipper motif. *Science* 1990 ; 249 : 774-8.
57. Neuberg M, Schuermann M, Hunter JB, Müller R. Two functionally different regions in Fos are required for the sequence-specific DNA interaction of the Fos/Jun protein complex. *Nature* 1989 ; 338 : 589-90.
58. Turner R, Tjian R. Leucine repeats and an adjacent DNA binding domain mediate the formation of functional cFos-cJun heterodimers. *Science* 1989 ; 243 : 1689-94.
59. Andrisani OM, Dixon JE. Involvement of lysine residues 289 and 291 of the cAMP-responsive element-binding protein in the recognition of the cAMP-responsive element. *J Biol Chem* 1991 ; 266 : 21444-50.
60. Tzamarias D, Pu WT, Struhl K. Mutations in the bZIP domain of yeast GCN4 that alter DNA-binding specificity. *Proc Natl Acad Sci USA* 1992 ; 89 : 2007-11.
61. Patel L, Abate C, Curran T. Altered protein conformation on DNA binding by Fos and Jun. *Nature* 1990 ; 347 : 572-5.
62. Weiss MA, Ellenberger T, Wobbe CR, et al. Folding transition in the DNA-binding domain of GCN4 on specific binding to DNA. *Nature* 1990 ; 347 : 575-8.
63. Weiss MA. Thermal unfolding studies of a leucine zipper domain and its specific DNA complex : implications for scissor's grip recognition. *Biochemistry* 1990 ; 29 : 8020-4.
64. Shuman JD, Vinson CR, McKnight SL. Evidence of changes in protease sensitivity and subunit exchange rate on DNA binding by C/EBP. *Science* 1990 ; 249 : 771-4.
65. Murre C, Schonleber McCaw P, Baltimore D. A new DNA binding and dimerization motif in immunoglobulin enhancer binding, *daughterless*, *MyoD*, and *myc* proteins. *Cell* 1989 ; 56 : 777-83.
66. Anthony-Cahill SJ, Benfield PA, Fairman R, et al. Molecular characterization of helix-loop-helix peptides. *Science* 1992 ; 255 : 979-83.
67. Vison CR, Garcia KC. Molecular model for DNA recognition by the family of basic-helix-loop-helix-zipper proteins. *New Biologist* 1992 ; 4 : 396-403.
68. Dang CV, Dolde C, Gillison ML, Kato GJ. Discrimination between related DNA sites by a single amino acid residue of Myc-related basic-helix-loop-helix proteins. *Proc Natl Acad Sci USA* 1992 ; 89 : 599-602.
69. Davis RL, Weintraub H. Acquisition of myogenic specificity by replacement of three amino acid residues from MyoD into E12. *Science* 1992 ; 256 : 1027-30.
70. Ham J, Dostatni N, Gauthier JM, Yaniv M. The papillomavirus E2 protein : a factor with many talents. *Trends Biochem Sci* 1991 ; 16 : 440-4.
71. Hedge RS, Grossman SR, Laimins LA, Sigler PB. Crystal structure at 1.7 Å of the bovine papillomavirus-1 E2 DNA-binding domain bound to its DNA target. *Nature* 1992 ; 359 : 505-12.

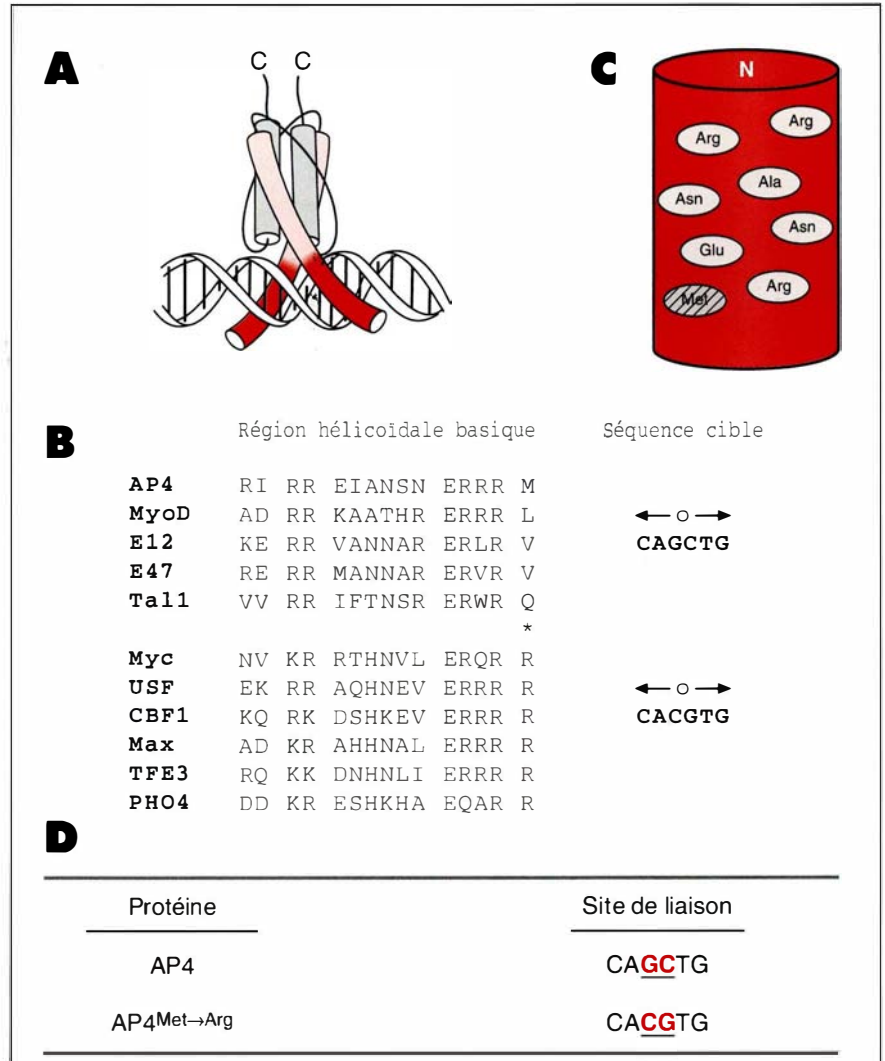


Figure 6. **La région hélicoïdale basique des protéines de la classe bHLH.** **A.** Modèle de l'interaction entre une protéine de la classe bHLH (sous forme dimérique) et l'ADN (selon [66, 67]). La région basique représentée en rouge, les première et deuxième hélices du motif HLH sont respectivement en rose et en gris. **B.** Séquence en acides aminés de la région basique de 11 protéines de la classe bHLH. Selon la nature des bases centrales du motif $5'CAXXTG3'$, ces protéines peuvent être classées en deux groupes (d'après [68]). La position de l'acide aminé responsable de la discrimination $5'CAGCTG3' \rightarrow 5'CACGTG3'$, est indiqué par *. **C.** Représentation graphique de la région basique de la protéine AP4, selon une conformation en hélice α . Les résidus basiques, les autres résidus conservés (Ala, Asn, Glu) et l'acide aminé responsable de la discrimination $5'CAGCTG3' \rightarrow 5'CACGTG3'$ (indiqué par des hachures) se trouvent sur la même face de l'hélice. **D.** Le résidu responsable de la discrimination $5'CAGCTG3' \rightarrow 5'CACGTG3'$. Le remplacement, dans AP4, de la méthionine (indiquée par * dans la partie B de la figure) par une arginine change la spécificité de liaison à l'ADN. (D'après [68].)

même périodicité que celle d'une hélice α ([56] et voir figures 5B et 5C). La structure tridimensionnelle de la région basique n'est pas définitivement établie, mais les modèles actuels postulent qu'elle adopte une structure en hélice α [55, 56]. Notons alors que les résidus conservés et basiques se trouvent tous sur la même face de cette hélice, et sont susceptibles d'interagir avec l'ADN [56] (figure 5D). De plus, des acides aminés conservés, de charge négative (Glu), et localisés à chaque extrémité de la région basique se trouvent sur la face opposée de cette hélice, créant ainsi un dipôle dont le pôle positif pointe vers l'ADN. Dans la région basique, un résidu asparagine, conservé dans toutes les protéines de la famille *Leu-zipper*, est responsable d'une flexion de l'hélice qui permet le contact des acides aminés plus distaux avec l'ADN (modèle du *scissors-grip* [55]). Ce modèle de structure hélicoïdale basique est en accord avec les études par dichroïsme circulaire [56]. De même, l'analyse par mutagenèse dirigée confirme l'importance des acides aminés conservés dans la capacité de liaison à l'ADN [57-60].

Comme d'autres domaines de liaison à l'ADN, la structure du motif de reconnaissance est ici aussi une hélice α ; mais, à l'inverse des autres motifs de liaison à l'ADN, l'hélice α n'est pas soutenue au sein d'une structure rigide (seule la dimérisation est nécessaire à la liaison à l'ADN ; deux « régions hélicoïdales basiques », liées l'une à l'autre par un pont disulfure, sont tout à fait capables de se fixer à l'ADN). Par ailleurs, l'hélice α n'existe pratiquement pas en solution, mais est stabilisée par l'interaction avec l'ADN [61-63]. Ainsi, l'absence de structure de maintien permettrait dans ce cas l'enroulement de la région basique hélicoïdale autour de l'ADN, dans le grand sillon [64].

La région basique hélicoïdale, en tant que domaine de liaison à l'ADN, n'est pas réservée aux protéines à domaine *Leu-zipper* : une autre classe de protéines, les protéines à domaine hélice-boucle-hélice (*helix-loop-helix* [65], figure 6A), possède également, en position N-terminale adjacente à ce domaine de dimérisation, une région riche en acides aminés basiques et

très certainement de structure hélicoïdale. La structure primaire de ce domaine est différente de celle des protéines à domaine *Leu-zipper*, comme l'est le motif d'ADN reconnu par les protéines à domaine hélice-boucle-hélice. En effet, toutes les protéines de la classe bHLH (classe des

protéines qui possèdent un domaine « région basique-hélice-boucle-hélice ») reconnaissent le motif palindromique $5'CAXXTG3'$. Comme dans le cas des protéines de la classe bZIP (classe des protéines qui possèdent un domaine région basique-*Leu-zipper*), il a été montré que la région basique

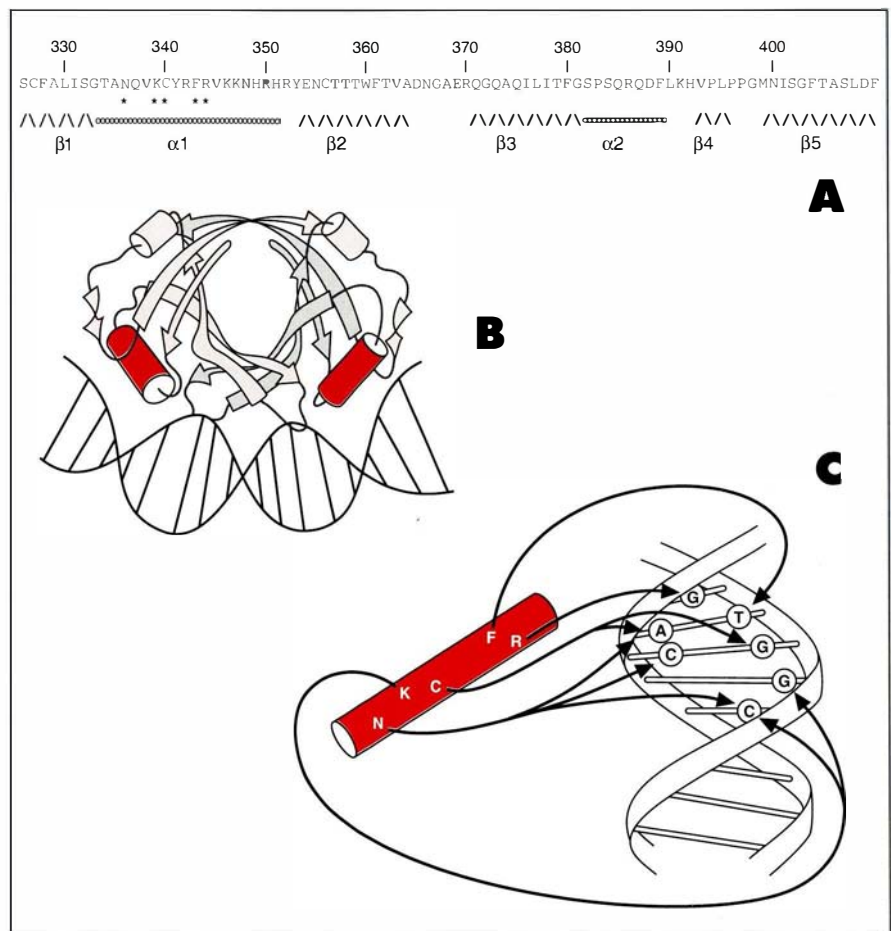


Figure 7. **Le motif en cylindre β antiparallèle de la protéine E2 du BPV.** **A.** Séquence en acides aminés du domaine de liaison à l'ADN et de dimérisation de la protéine E2 du virus bovin du papillome (BPV-1). Les acides aminés impliqués dans les structures en brin β sont notés λ ; ceux qui adoptent une structure en hélice α sont marqués par α . Les brins β et les hélices α sont numérotés. Les résidus responsables des interactions avec les bases de l'ADN (Asn₃₃₆, Lys₃₃₉, Cys₃₄₀, Phe₃₄₃ et Arg₃₄₄) sont indiqués par le code à une lettre. **B.** Diagramme schématisé du complexe formé entre un dimère du domaine de liaison à l'ADN/dimérisation de la protéine E2 du BPV et sa séquence cible (tiré de [71]). Les deux sous-unités de la protéine sont en gris et en rose, et adoptent une structure en cylindre β antiparallèle. Les hélices $\alpha 1$ (de reconnaissance — en rouge) de chaque monomère s'enfoncent dans le grand sillon de l'ADN [71]. **C.** Représentation des interactions spécifiques entre l'hélice de reconnaissance de E2 (résidus Asn₃₃₆, Lys₃₃₉, Cys₃₄₀, Phe₃₄₃ et Arg₃₄₄) et les bases du grand sillon de l'ADN (adapté de [71]).

hélicoïdale des protéines de la classe bHLH est responsable de la liaison à l'ADN [68, 69]. Par exemple, le remplacement d'un seul acide aminé dans ce domaine est responsable de la discrimination entre les motifs $5' \text{CAGCTG} 3'$ et $5' \text{CACGTG} 3'$ ([68], *figure 6*).

La protéine E2 et le cylindre β antiparallèle

Les virus du papillome constituent une famille de virus à ADN, responsables de lésions épithéliales chez les mammifères (pour revue voir [70]). Chez ces virus, les produits du gène E2 jouent un rôle central dans la régulation de la transcription et de la réplication du génome viral. La liaison à l'ADN des différentes formes de la protéine E2 se fait par l'intermédiaire d'un domaine de 85 résidus (*figure 7A*, p. 733), également suffisant pour la dimérisation de la protéine.

La détermination des interactions ADN-E2 par cristallographie aux rayons X révèle un motif original [71]. Le dimère de E2 forme un cylindre constitué de l'association de huit brins β antiparallèles ; chaque monomère de E2 participe à cette structure en cylindre β antiparallèle (*antiparallèle β -barrel*, voir *figure 7B*) par quatre brins β (les brins $\beta 1$, $\beta 2$, $\beta 3$ et $\beta 5$ de la *figure 7A*). Situés à la périphérie de ce cylindre, les résidus d'une hélice α (l'hélice $\alpha 1$ de la *figure 7A*) sont directement impliqués dans la reconnaissance des bases de l'ADN, dans le grand sillon ([71] et voir *figures 7B et 7C*).

Conclusion : unité et diversité dans la liaison à l'ADN

Les analyses fonctionnelles et structurales des domaines de liaison à l'ADN des facteurs de transcription eucaryotes révèlent que ces modules peuvent être très divers. En dépit de leurs différences structurales, ces domaines possèdent un même rôle : celui d'exposer une hélice α dans le grand sillon de l'ADN. Toutefois, chaque classe de protéines présente cette hélice de reconnaissance d'une façon qui lui est propre, au sein d'une structure bien définie.

Souvent, le domaine de liaison à l'ADN est relativement fixe (sa structure tridimensionnelle est similaire, en solution et une fois liée à l'ADN), mais dans certains cas, comme dans les protéines de la classe bZIP, ce n'est que lorsque le domaine se lie à l'ADN qu'il adopte sa structure définitive en hélice α .

En outre, un seul module de liaison à l'ADN n'est généralement pas suffisant ; les protéines des classes bZIP, bHLH, à doigt à Zn de types C_4 et C_6 , ou à cylindre β antiparallèle se fixent à l'ADN sous forme dimérique. Plus que la simple formation d'un domaine de liaison à l'ADN fonctionnel, dans certaines situations (protéines à doigts à Zn de types C_4 et C_6), la dimérisation détermine en partie la spécificité de l'interaction ADN-protéine, par un simple effet de positionnement du motif de liaison à l'ADN. Chez les protéines des classes bZIP, bHLH et à doigts à Zn de type C_4 , la capacité à faire des hétérodimères avec d'autres membres de leur classe augmente la diversité et la spécificité des interactions ADN-protéines.

La situation trouvée pour les facteurs de transcription à doigt à Zn de type C_2-H_2 est différente : ici, la structure oligomérique de la protéine est remplacée par une association covalente de modules de même type, au sein de la chaîne polypeptidique. L'association de plusieurs modules de liaison à l'ADN est une situation rencontrée dans certaines protéines à domaine homéo ; les domaines Prd et POU, mais probablement aussi les répétitions LIM, sont ainsi impliqués dans la reconnaissance de séquences d'ADN au voisinage des séquences cibles de l'homéodomaine. Enfin, d'autres protéines semblent coopérer avec les homéoprotéines afin d'augmenter ou de modifier la spécificité de liaison du domaine homéo à l'ADN. Ces différentes organisations montrent que les facteurs de transcription eucaryotes peuvent adopter des stratégies très différentes dans le but d'exposer une hélice α de reconnaissance dans le grand sillon de l'ADN. Chaque classe de protéines présente son hélice α au sein d'une structure qui lui est propre, et cela suggère que les domaines de liaison à l'ADN qui utilisent une hélice de reconnaissance

ont été réinventés plusieurs fois au cours de l'évolution ■

Summary

DNA-binding domains of eukaryotic transcription factors

Most of transcription factors recognize specific DNA sequences through small, discrete domains. In some cases, these domains can be interchanged between proteins showing that they are independently folded units. Mutational and three-dimensional structure analyses have been used to understand the means by which protein-DNA interactions are achieved. There are different kinds of DNA-binding domains. In many cases, this domain is fixed, but in others, it folds only on binding to DNA. Thus homeodomains, C_2-H_2 -class Zn fingers, steroid receptor-like C_4 -class Zn fingers, GAL4-like C_6 -class Zn fingers and E2 antiparallel β -barrel have a « recognition helix » that protrudes from a compact domain. By contrast, the basic region of bZIP-class proteins has a flexible segment that wraps around the major groove of DNA. The E2 protein, the steroid receptor-like C_4 -class Zn finger, the GAL4-like C_6 -class Zn finger, the bZIP-class and the bHLH-class of transcription factors form dimers. By opposition, most homeoproteins bind to DNA as monomers, and finally the C_2-H_2 -class Zn finger proteins form covalent concatemers. The study of DNA-binding domains of eukaryotic transcription factors shows more than one way to present an α -helix in the major groove. In evolutionary terms, this implies that these proteins have arisen from different ancestors : consequently, the DNA-binding domain must have been « invented » several times.

Remerciements

L'auteur remercie Mary Weiss pour son soutien et ses encouragements lors de la rédaction de cette revue.