Volume 234, number 2, 245–250 FEB 06005 July 1988

Review Letter

Finger proteins and DNA-specific recognition: distinct patterns of conserved amino acids suggest different evolutionary modes

François Payre and Alain Vincent

Centre de Recherche de Biochimie et Génétique Cellulaires du CNRS, 118, route de Narbonne, 31062 Toulouse Cedex, France

Received 26 April 1988

Finger proteins, the first example of which was *Xenopus* TFIIIA, share Zn^{2+} finger-like folded domains capable of binding to nucleic acids. A large number of this type of protein have been characterised from diverse organisms, indicating a wide evolutionary spread of the DNA-binding fingers. At least two classes of finger proteins may be distinguished. Class I proteins contain variable numbers of the tandemly repeating TFIIIA-like finger motif, (Y/F-X-C-X_{2-4}-C-X_3-F-X_5-L-X_2-H-X_3-H). Class II finger proteins display a single (C-X_2-C-X_{13}-C-X_2-C) motif and a facultative second putative finger. The relation between the structure of finger proteins and their recognised DNA sequences is discussed.

Finger protein; DNA binding; TFIIIA; Zn^{2+} chelation; Transcription activator

1. INTRODUCTION

The general idea that regulation of gene expression is mediated by interaction of specific proteins with DNA sequences located near gene transcriptional starts came from genetic and molecular studies on prokaryotic regulatory proteins. X-ray diffraction studies of prokaryotic repressor-DNA cocrystals [1] have demonstrated a DNA-binding protein domain composed of two α-helices ('2' and '3') connected by a β-turn, with the helix '3' lying in the major groove of the DNA double helix [1,2]. The α-helix/β-turn/α-helix motif is also found in eukaryotic proteins such as developmental gene products containing a homeo-box, or yeast mating-type proteins [3].

Although this motif was the first recognisable structural element shared by transcription factors, molecular analysis of *Xenopus* TFIIIA and its interaction with the internal control region (ICR) of the 5 S genes led Miller et al. [4] to propose an alternate DNA binding protein structure, designated as 'finger'. The finger is a 29–30 amino acid motif (Y/F-X-C-X_4-C-X_{12}-H-X_3-H-X_6-7) forming a small independent structural domain, made of a loop centered around a zinc atom coordinated by two cysteine and two histidine residues located at invariant positions [8]. The model of Klug and Rhodes [5] postulates that the tip of each loop (C-X_3-Phe-X_5-Leu-X_2-H) is in direct contact with the DNA and determines the DNA binding specificity. The present review will compare TFIIIA and other finger proteins with regard to this model.

Correspondence address: A. Vincent, Centre de Recherche de Biochimie et Génétique cellulaires du CNRS, 118 route de Narbonne, 31062 Toulouse Cedex, France

Abbreviations: TFIIIA, transcription factor A of polymerase III; ICR, internal control region; PrKC, protein kinase C; C, cysteine; F, phenylalanine; H, histidine; L, leucine; Y, tyrosine; X, any amino acid; GR, glucocorticoid receptor; ER, oestrogen receptor; PR, progesterone receptor; MR, mineralocorticoid receptor; UAS, upstream activation site

Published by Elsevier Science Publishers B.V. (Biomedical Division)
00145793/88/$3.50

2. C-X_{2-4}-C-X_{12}-H-X_3-H-X_7 FINGER PROTEINS; A NEW CLASS OF DNA BINDING PROTEINS

The model for finger protein-DNA interaction (review [5]) was based on footprinting identification of the 5 S ICR nucleotides in direct contact with a TFIIIA molecule. These contacts are distributed in nine distinct patches, all located in the major groove of the DNA molecule and each separated by five nucleotides – a distance equivalent to half a double helix period. Fairall et al. [7] proposed that TFIIIA lies on one side of the DNA double helix, with the nine successive fingers making contact with DNA (fig.1).

Only a few amino acids are invariant in each finger. These are precisely interspersed with variable amino acids and probably provide a framework of tertiary folding; no stretches of contiguous invariant amino acids, as is the case for the helix/turn/helix motif [2], are seen in TFIIIA finger tips. Fingers appear to be polarised structures since all contain an N-terminal pair of cysteine and a C-terminal pair of histidine residues. This polarity is also seen in the asymmetric distribution of the invariant hydrophobic amino acids, Tyr, Leu and Phe at positions 1, 11 and 17, respectively (fig.2) and correlates with the absence of symmetry in the TFIIIA binding site.

Several genes with predicted protein products showing TFIIIA-like fingers have been isolated. They include *Drosophila* serendipity (sry)-β and sry-δ [9], Krüppel (Kr) [10], hunchback (Hb) [11], snail [12] and terminus (Ter) [34]; yeast ADRI [13], and SWI5 [5]; human Sp1 [20] and sex-determining region of the chromosome [39]. Additional finger protein genes have been cloned by low stringency hybridization with *Drosophila* finger related probes [14–16,40]. The selective DNA-binding properties of some Zn^{2+} finger proteins other than TFIIIA have been established. Direct evidence exists that ADRI is a gene-specific positive transcription factor [17]. Sp1 was first characterised as enhancing transcription at the late SV40 promoter by binding a GGG CCG hexanucleotide (the 'GC box' found in upstream control regions of several other genes [18]). In addition, the *Drosophila* Krüppel [19] and sry-δ (Payre et al., in preparation) proteins bind DNA specifically but the recognition sequences are not yet known. Finally TFIIIA, Sp1 and SWI5 have been demonstrated to require Zn^{2+} for binding to DNA [5,20].

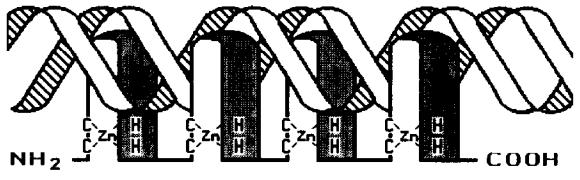

Fig.1. Schematic representation of a model for the interaction of fingers with the DNA double helix (adapted from [7]). The part of each finger which is predicted to form an α-helix is shaded [33,41].

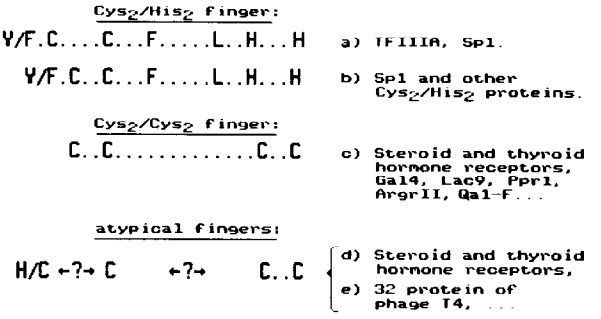

Fig.2. Patterns of conserved amino acids in the various putative Zn^{2+} fingers and corresponding classes of finger proteins; the consensus tyrosine (Y) residue in Cys_2/His_2 fingers is often replaced by a phenylalanine. References: a, [4,20]; b, [9–16,20,39,40]; c, [22–26]; d, [22]; e, [31].

Multi-finger proteins analysed so far present a large variation in the number and position of fingers or groups of fingers. It therefore seems unlikely that, within a given protein, all fingers are functionally equivalent [21]. The TFIIIA-like finger appears as a modular structure allowing a wide combination of solutions for rapid protein evolution and specificity in DNA recognition.

3. A VARIANT (C-X_2-C-X_{13}-C-X_2-C) FINGER

A variant DNA binding motif reminiscent of the TFIIIA (Cys_2/His_2) finger has also been observed. This motif, designated as Cys_2/Cys_2, has the symmetrical consensus structure C-X_2-C-X_{13}-C-X_2-C (fig.2), and is found in various transcription activators, including receptors for steroid and thyroid hormones, and vitamins A and D (review

[22]), and factors from unicellular eukaryotes: Gal4, Ppr1, ArgrII in *S. cerevisiae*, Lac9 in *Kluyveromyces lactis*, and Qa 1-F in *Neurospora crassa* [23–25]. These finger proteins show some functional analogies: acting selectively on the transcription of specific gene(s) they can mediate transitions between two physiological states and control metabolism in response to a hormonal stimulus or to changes in nutrient source. Moreover, their function is itself mediated by their interaction with one or several specific ligands.

The finger is determinant for DNA binding and recognition of target sequences. Genetic experiments with Gal4 have shown that the integrity of the finger and Zn^{2+} are required for DNA binding [29,30]. Transfection experiments [27,28] with mutagenised steroid hormone receptors defined the finger domain as the DNA binding domain [27,28]. Only one Cys_2-X_{13}-Cys_2 motif is present per protein. Hormone receptors display an additional putative 'hybrid' finger of the general structure C-X_5-C-X_{12}-C-X_4-C (or, alternatively H-X_{3-4}-C-X_{15}-C-X_2-C) (fig.2). A putative hybrid finger is also present in the product of the gene 32 from bacteriophage T4 which binds aspecifically single strand DNA. Again, the Zn^{2+} (one atom per protein) is implicated in the DNA binding process [31]. Because the C-X_2-C-X_{13}-C-X_2-C finger was conserved from yeast to human, one can argue that this finger structure is essential for binding DNA. The respective contributions of the finger itself and finger adjacent sequences to the binding specificity remain a matter of debate [25,32]. A duplicated Cys_2/Cys_2 finger motif is present in the predicted sequence of the protein kinases C from *Drosophila* and other eukaryotes [26]. This is perhaps surprising since protein kinases are not thought to be DNA binding proteins.

4. TYPE I AND TYPE II FINGERS CORRELATE WITH TWO CLASSES OF DNA BINDING PROTEINS

In spite of initial confusion, Cys_2/His_2 and Cys_2/Cys_2 fingers are not structurally equivalent. They correlate with two classes of DNA-binding proteins (designated here as class I and II, respectively) and two types of recognition sites. First, reiteration of fingers is restricted to class I proteins. While known class I proteins contain from one (Ter) [34] to 37 fingers (Xfin) [14], class II proteins display no more than two non-equivalent potential fingers (with the possible exception of PrKc). Second, an attempt to convert a Cys_2/Cys_2 into a Cys_2/His_2 finger by site directed mutagenesis was unsuccessful. The replacement of the C-terminal cysteines by histidines in the finger of the human oestrogen receptor lead to its inactivation [32]. It is not known whether this is due to the shorter spacing of the histidines – two amino-acid residues as opposed to a minimal three in all described Cys_2/His_2 fingers (fig.2) – or to inadequate folding of the finger tip, or both. Indeed, a prediction of partial α-helical secondary structure of some Cys_2/His_2 fingers (including the histidine residues) has been reported [33,41]. Because some proteins display a Cys_2-X_n-Cys_2 motif without a priori binding to DNA [23], target recognition by Cys_2/Cys_2 fingers could require to adopt specific conformations contributed by elements in the finger itself or in adjacent sequences.

A strong protein conservation of the Cys_2/Cys_2 finger sequence among proteins displaying homologous functions is observed. Within the steroid hormone receptor superfamily, the finger region is far better conserved than any other domain of the proteins, including the hormone binding region [22]; the finger domains of receptors binding the same hormone in different species are frequently identical (fig.3) and those of receptors binding different steroid hormones in the same species remain similar [35,36]. A cross-species relationship between primary structure and function has also been reported for yeast transcription activators; Lac9 from *K. lactis* and Gal4 from *S. cerevisiae* which are functionally homologous (Lac9 can complement a Gal4 mutation), and bind to homologous DNA sequences, present highly homologous DNA binding domains [37]. In contrast, the primary sequence of each Cys_2/His_2 domain reported up to now is unique (although several fingers within a given protein can be very similar [43]). Striking examples are the products of the Krh [16], mrk1, mkr2 [15], krox [43] and xfin [14] genes, all isolated via hybridization with a *Drosophila* Krüppel DNA probe ([10], fig.3). The only motif shared by these proteins and Krüppel is a 7 amino acid link separating adjacent fingers (the H/C link). The conservation of this 7 amino acid

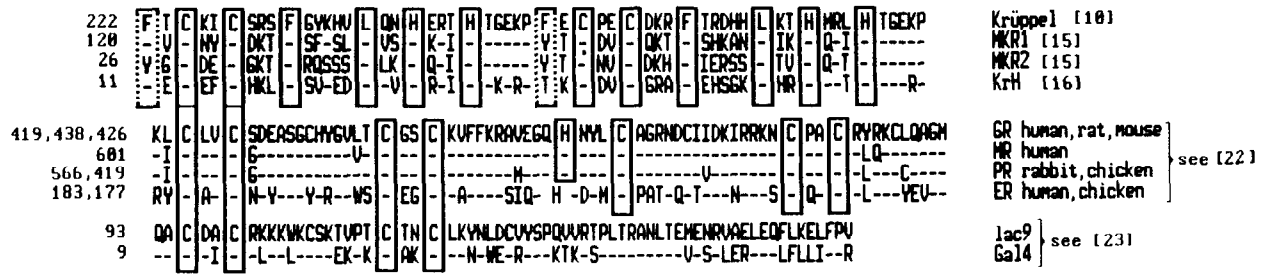

Fig.3. Patterns of conserved amino acids in the DNA binding domain of some class I and class II finger proteins. Consensus amino acids are boxed. The complete sequence is indicated only for one protein in each group (Krüppel, h GR lac 9); the amino acids which are identical at the same positions in other proteins of the group are represented by hyphens. Positions of the NH_2 terminal amino acids are indicated. Only the two NH_2 proximal fingers of class I proteins were arbitrarily taken for this comparison.

stretch in a subset of finger proteins, independent of the rest of the sequence, suggests that it could be important for folding and/or interaction of these proteins with DNA but it is not essential for DNA binding specificity.

A potential model for studying whether there is concerted evolution of Cys_2/His_2 finger proteins and their DNA binding specificity is the *Drosophila* sry-β and -δ, genes which probably arose from a duplication of an ancestral gene already containing 6 fingers [9]. Firstly, the degree of sequence similarity between individual fingers found at homologous positions in the duplicated region is variable. Secondly, an additional finger is found in the carboxy terminal portion of the sry-δ protein. Finally, limited sequence similarity between these two proteins is observed outside the finger domain [9,21].

No sequence conservation between Cys_2/His_2 and Cys_2/Cys_2 fingers has been found. Whether they evolved from a single primordial finger remains an issue to be solved. It is likely, however, that functional speciation of class I and class II finger proteins involved separate mechanisms. Class I proteins evolved through reiteration of a structural backbone motif F/Y-X-C-X_{2-4}-C-X_3-F-X_5-L-X_2-H-X_3-H with rapid sequence divergence at other positions in the fingers and no indication for a conservation of the adjacent regions. Class II proteins evolved through diversification of the primary sequence of a single C-X_2-C-X_{13}-C-X_2-C finger and adjacent regions with coevolution of this DNA binding domain with the DNA target, and other domains of the protein (such as hormone or effector protein(s) binding domains).

5. HOW FAR DOES THE TFIIIA MODEL APPLY TO OTHER FINGER PROTEINS?

The structural diversity of finger proteins raises the question of the general validity of the TFIIIA model of nucleic acid binding. Notwithstanding the structural difference between Cys_2/His_2 and Cys_2/Cys_2 Zn^{2+} fingers and its relation to the nature of recognised DNA sequences (asymmetric versus (partly) palindromic) we will consider here essentially class I proteins.

The postulate that each finger binds to half a DNA turn and that a periodicity of guanine residues in the binding site correlates with the interaction of contiguous fingers came from studies of the TFIIIA 5 S-ICR complex and examination of other gene regulatory sequences [6]. However, Sp1 which displays three fingers for a 10 bp long consensus binding site [18] provides an example in which the number of 5.5 bp periodicities in the recognition site does not correlate with the number of fingers.

Data on yeast ADRI that specifically binds the UAS of the glucose repressible ADH2 gene [13,17] show that the two Cys_2/His_2 fingers are necessary for its function. A larger number of fingers per protein may result in an increase in length of the DNA sequences specifically recognised and a resulting increase in binding strength. This is supported by the observation that the 20 kDa proteolytic fragment of TFIIIA [38] (corresponding to the 6 amino-terminal fingers [4]) binds the 3′ end of the 5 S ICR specifically, but with a weakened binding strength [38].

The apparent discrepancy between a larger

number of fingers per protein and the small average size of *cis*-elements recognised by regulatory proteins raises another possibility. Provided that simultaneous binding of two fingers to a core sequence is sufficient for the formation of a stable complex [17], the presence of numerous fingers could result in an extended set of specific short target sites recognised per protein. Moreover, the clustering of fingers in separate groups observed in some finger proteins raises the possibility of split binding sites made of non-contiguous sequences.

The two alternatives we propose might well correspond to separate properties of different finger proteins. On the one hand, TFIIIA is essential for the transcription, by RNA polymerase III [5], of a unique set of genes (5 S). Contact between each of the 9 TFIIIA fingers (80% of the protein) and DNA might be required by the fact that TFIIIA remains bound to the transcribed ICR for many rounds of transcription. On the other hand, Sp1 enhances the transcription by polymerase II of various genes containing at least one properly positioned binding site, upstream of the transcribed region [18]. The known Sp1 binding sites present some sequence variability, especially in the nucleotides adjacent to the core GC box, reflecting different relative affinities of Sp1 for each site [18]. Here, the multifingered structure of a protein might account for such a diversity in sequence specificity and binding strength.

Another property is (so far) restricted to TFIIIA: its ability to bind specific RNAs; in addition to its role in transcription, TFIIIA is a stabilising component of 5 S RNA during oogenesis (in 7 S RNP particles) [5]. This property could be shared with some other Cys_2/His_2 finger proteins; good candidates are finger protein gene products stored in oocytes [9,11,40] or subject to post-transcriptional regulation [42].

Finally, the contribution of domains adjacent to the finger region – that are likely to be involved in interactions with other factors – to the DNA binding specificity remains to be assessed. Only then will we begin to understand better how multiple finger proteins specifically act to regulate gene expression.

Acknowledgements: This work was supported by a grant from Centre National de la Recherche Scientifique (CNRS) to A.V.; the authors wish to thank Mick Chandler for critical reading of the manuscript and Jeremy Berg, James Kadonaga and Judith Lengyel for communication of unpublished data.

REFERENCES

[1] Anderson, J.E., Ptashne, M. and Harrison, S.C. (1985) Nature 316, 596–601.
[2] Pabo, C.O. and Sauer, R.T. (1984) Annu. Rev. Biochem. 53, 293–321.
[3] Gehring, W.J. (1987) Science 236, 1245–1252.
[4] Miller, J., McLachlan, A.D. and Klug, A. (1985) EMBO J. 4, 1609–1614.
[5] Klug, A. and Rhodes, D. (1987) Trends Biochem. Sci. 12, 464–469.
[6] Rhodes, D. and Klug, A. (1986) Cell 46, 123–132.
[7] Fairall, L., Rhodes, D. and Klug, A. (1986) J. Mol. Biol. 192, 577–591.
[8] Diakun, G., Fairall, L. and Klug, A. (1986) Nature 324, 698–699.
[9] Vincent, A., Colot, H.V. and Rosbash, M. (1985) J. Mol. Biol. 186, 149–166.
[10] Rosenberg, V.B., Schroder, C., Preiss, A., Kienlin, A., Coté, S., Riede, I. and Jacle, H. (1986) Nature 319, 336–339.
[11] Tautz, D., Lehmann, R., Schnurch, H., Schuh, R., Seifert, E., Kienlin, A., Jones, K. and Jackle, H. (1987) Nature 327, 383–389.
[12] Boulay, J.L., Dennefeld, C. and Alberga, A. (1987) Nature 330, 395–398.
[13] Hartshorne, T.A., Blumberg, H. and Young, E.T. (1986) Nature 320, 238–287.
[14] Ruiz i Albata, A., Perry-O'Keefe, H. and Melton, D.A. (1987) EMBO J. 6, 3065–3070.
[15] Chowdhury, K., Deutsch, V. and Gruss, P. (1987) Cell 48, 771–778.
[16] Schuh, R., Aicher, W., Gaul, V., Coté, S., Preiss, A., Maier, D., Seifert, E., Nauber, V., Schroder, C., Kemler, R. and Jacle, H. (1986) Cell 47, 1025–1032.
[17] Blumberg, H., Eisen, E., Sledziewski, A., Bader, D. and Young, E.T. (1987) Nature 328, 443–445.
[18] Kadonaga, J.T., Jones, K.A. and Tijan, R. (1986) Trend Biochem. Sci. 11, 20–23.
[19] Ollo, R. and Maniatis, T. (1987) Proc. Natl. Acad. Sci. USA 84, 5700–5704.
[20] Kadonaga, J.T., Carner, K.M., Masiarz, F.R. and Tijan, R. (1987) Cell 51, 1079–1090.
[21] Vincent, A. (1986) Nucleic Acids Res. 14, 4385–4391.
[22] Gehring, U. (1987) Trend Biochem. Sci. 12, 399–402.
[23] Johnston, M. and Dover, J. (1987) Proc. Natl. Acad. Sci. USA 84, 2401–2405.
[24] Messenguy, F., Dubois, E. and Deschamps, F. (1986) Eur. J. Biochem. 157, 77–81.
[25] Baum, J.A., Geever, R. and Giles, N.H. (1987) Moll. Cell. Biol. 7, 1256–1266.
[26] Rosenthal, A., Rhee, L., Yadegari, R., Paro, R., Ullrich, A. and Goeddel, D.V. (1987) EMBO J. 6, 433–441.
[27] Hollenberg, S.M., Giguere, V., Segui, P. and Evans, R.M. (1987) Cell 49, 39–46.
[28] Rusconi, S. and Yamamoto, K.R. (1987) EMBO J. 6, 1309–1315.

[29] Keegan, L., Gill, G. and Ptashne, M. (1986) Science 231, 699–704.
[30] Johnston, M. (1987) Nature 328, 353–355.
[31] Giedroc, D.P., Keating, K.M., Williams, K.R., Konisberg, W.H. and Coleman, J. (1986) Proc. Natl. Acad. Sci. USA 83, 8452–8456.
[32] Green, S. and Chambon, P. (1987) Nature 325, 75–78.
[33] Brown, R.S. and Argos, P. (1985) Nature 324, 215.
[34] Baldarelli, R.M., Mahoney, P.A., Salas, F., Gustavson, E., Boyer, P., Chang, M.F., Roark, M. and Lengyel, J. (1988) Dev. Biol., in press.
[35] Krust, A., Green, S., Argos, P., Kumar, V., Walter, P., Bornert, J.M. and Chambon, P. (1986) EMBO J. 5, 891–897.
[36] Jeltsch, J.M., Krozowski, Z., Quirin-Stricker, C., Gronemeyer, H., Simpson, R.J., Garnier, J.M., Krust, A., Jacob, F. and Chambon, P. (1986) Proc. Natl. Acad. Sci. USA 83, 5424–5428.
[37] Breunig, K.D. and Kuger, P. (1987) Mol. Cell. Biol. 7, 4400–4406.
[38] Smith, D.R., Jackson, T.J. and Brown, D.D. (1984) Cell 37, 645–652.
[39] Page, D.C., Mosher, R., Simpson, E.M., Fisher, E.M.C., Mardon, G., Pollack, J., McGillivray, B., De La Chapelle, A. and Brown, L.G. (1987) Cell 51, 1091–1104.
[40] Vincent, A., Kejzlarova-Lepesant, J., Segalat, L., Janicostas, C. and Lepesant, J.A. (1988) submitted.
[41] Berg, J. (1988) Proc. Natl. Acad. Sci. USA 85, 99–102.
[42] Gaul, V., Seifert, E., Schuh, R. and Jackle, H. (1987) Cell 50, 639–647.
[43] Chavrier, P., Lemaire, P., Relevant, O., Bravo, R. and Charnay, P. (1988) in press.