Zipperless bZips and Zipped Homeodomains

By Masashi SUZUKI^{†)} and Rafael GIRALDO^{*)}

MRC Laboratory of Molecular Biology, Hills Road, Cambridge, CB2 2QH, U.K.

(Communicated by Setsuro EBASHI, M. J. A., Jan. 12, 1995)

Abstract : DNA recognition of zipperless basic-zipper (bZip) proteins and zipped homeodomains (HDs) are discussed in the light of the structure and the DNA-binding specificity of ordinary bZip and HD proteins. Proteins, Mastermind, Skin-1 and Srg-1, have twelve residue sequences which are very similar to that of the basic domain but lack the “zipper” sequences, which follows the basic domain in bZip proteins. It is predicted here that these proteins and Zeste and RAP1 have a domain which is composed of three α -helices as in an HD and the proteins bind to the DNA by a recognition helix of the HD-bZip type.

Key words : Zeste; Skin-1; Srg-1; RAP1; mastermind.

Introduction. Transcription factors of the homeodomain (HD) family and those of basic-zipper (bZip) family have been characterised extensively in their biological functions, structures and DNA-binding specificities.^{1),2)} A homeodomain is a globular domain and is composed of three α -helices (Fig. 1b); its third helix binds to the DNA. A basic-zipper, on the contrary, is a single, continuous α -helix (Fig. 1c); its N-terminal “basic” part binds to the DNA, while the C-terminal “zipper” interacts with the “zipper” of another protein for dimerisation. It was therefore surprising that the same type of recognition helices was identified in HD as well as in bZip proteins.³⁾

A helical segment used for DNA-recognition by the two families (a probe helix) is composed of twelve residues.^{3)–5)} Residues at position 2, 7, 9, 11 and 12 bind to DNA phosphates thereby fix the binding geometry (position 2 is occupied by a hydrophilic residue and positions, 7, 9, 11 and 12 by Lys or Arg) (Fig. 2a). Residues at positions 1, 4, 5 and 8 contact the DNA bases. Position 1 is occupied by one of four residues,⁴⁾ Val, Ile, Asn and His, or by Thr. In a few examples such as Hox11 the position is occupied by Thr, which is the largest among small residues⁴⁾ (see Fig. 3 for further discussion).

After the first analysis of the two families was

published,³⁾ some more interesting information have been published on the two families. In particular, Sessa *et al.*⁶⁾ have reported a possible structure of the HDzip proteins, Athb-1 and 2. The proposed structure is a combination of a HD and a bZip. The two proteins have sequences of normal HDs. But the HD sequences are followed by zipper-like sequences in which Leu and other hydrophobic residues are repeating with the periodicity of 3.5 residues (Fig. 2a). Therefore, in the two proteins the recognition helices and the zippers are combined essentially in the same way as in a bZip. Indeed, the HDzip proteins fold into a dimer and bind to DNA sequences which have two direct binding sites of ordinary HDs related by a two fold symmetry⁶⁾ (Fig. 1d). The HDzip proteins may give a straightforward insight into the similarity of recognition helices of the HD and bZip families.

Another interesting observation is that, as it will be described in this paper, some proteins have a recognition helix which is essentially the same as those found in bZip proteins but are clearly missing the zipper⁷⁾ and that therefore these proteins cannot fold into the bZip fold, although their DNA-binding mode is expected to be the same as that of bZips. We here discuss a possible three dimensional structure of these zipperless bZips in comparison with the bZip, HD and HDzip proteins, and describe further the DNA-recognition framework shared by the bZip and HD transcription factor families.

Zipperless bZips. The majority of basic-zipper

^{*)} Current address: Departamento de Microbiología Molecular. Centro de Investigaciones Biológicas (CIB-CSIC). c/Velázquez, 144. 28006 Madrid, Spain.

^{†)} Correspondence to: M. Suzuki.

Fig. 1. Different protein foldings which can adopt a recognition helix of the HD-bZip type (a probe helix). Those of E2 (a), a homeodomain (b), a bZip (c), an HDzip (d) and a bHLH (e). DNA-recognition helices are shown hatched.

proteins have a recognition helix of the sequence, NhyXAAXBXB², where hy is a hydrophilic residue, B is Lys or Arg, N is Asn and A is Ala. Two putative transcription factors of *C. elegans*, Skin-1 and Srg-1,⁷ have such twelve residue sequences only at their very C-terminal ends and thus in these proteins a zipper simply cannot be made to follow the preceding recognition helix:

Skn-1 **NKVAARTCRQRR**TDRHDKMSHYI<end
Srg-1 **NKAAARTCRERRR**SKPY<end.

Another putative transcription factor of *Drosophila*, Mastermind,⁸ has a twelve residue sequence followed by Gln and His residues after an insertion of Pro residues. Again this domain does not possess a zipper of the leucine zipper type;

Mast **NKRAAKKTDKKLP**DPSQQHQQQQHQQQQHQQH.

What kind of structure do these putative DNA-binding domains fold into? The HDzip structure, even if its zipper part is removed, could still bind to the DNA as an ordinary HD. Therefore a quite obvious candidate for the zipperless bZip structure is an HD-like fold—i.e. a globular structure composed of three α -helices, in which the third is the DNA-recognition helix. Since for this structure no residue is necessary to be placed C-terminal to the recognition helix, it may be appropriate even for Srg-1, which has only five residues there.

The sequences of Skin-1, Srg-1 and Mastermind

are compared with that of the standard HD (Fig. 2b). Some features of these sequences indeed suggest a HD-like structure. First, these proteins have three stretches which are free from Pro and Gly, the residues which might break an α -helix. Second, each of the three stretches is about fifteen residues long, the size similar to those of the three helices in a homeodomain. And third, most importantly, each domain has hydrophobic residues on the a and d phases (see below).

An α -helix has seven phases, a–g, and the a and d phases are positioned on the same side of an α -helix (Fig. 3a). In HDs hydrophobic residues are found on the two phases in the three helices and the hydrophobic residues on one helix pack against those on the other helices. Hydrophobic residues are found on the two phases in the three stretches of the proteins Skin-1, Srg-1 and Mastermind. This suggests that the three stretches are indeed α -helical and pack against each other as in the HD. The idea is further supported by the fact that the a-d phase of the first helix does not match that of the second and that of the second helix does not match that of the third. Therefore these helices do not fold into a single continuous helix but are likely to have the two breaks.

The first residues in the twelve residue segments are positioned differently in the third helices (Fig. 2b). In Skin-1 it starts N-terminal to the two hydrophobic

Fig. 2. Comparison of the sequences of transcription factors of the HD-bZip type. (a) Comparison of the sequences of the recognition helices (basic) and the zippers. Some important residues are shown in bold characters—aa1, which is occupied by one of the four residues, Val, Ile, His, Asn (except in Hox11—see Fig. 3), aa7, 9, 11, 12, which are occupied by Lys or Arg, and hydrophobic residues in the zippers. Positions 1, 4, 5, 8, which are predicted to contact the DNA bases, are shown underlined. (b) Comparison of the sequences of DNA-binding domains which are predicted to have three α -helices (H1–H3). The standard HD sequence¹⁾ is shown for comparison. The hydrophobic residues at the a and d phases are shown in bold characters. The twelve residue segments of the probe helix are shown underlined.

positions in the third helix (on the b phase), in the HD it starts between the two hydrophobic positions (on the f phase), and in Mastermind it starts C-terminal to the two (on the f phase). This again coincides with the three helix packing, because the b and f phases are positioned on the surface opposite to the a and d phases and thus the residues which are expected to contact the DNA bases are likely to be exposed and placed opposite from the protein core (Fig. 3a).

Similarity of the sequence of another factor,

Zeste, to those of HDs¹⁾ as well as to those of bZips⁷⁾ have been pointed out. Thus Zeste might be another protein which can bridge the two families. Indeed, the Zeste DBD appears to be another example of the three helix packing (Fig. 2b). It has a twelve residue sequence, IQKFWLNSRLRK (see Fig. 4j for its DNA-binding specificity).

The DNA-binding protein, RAP1, may be another example of such three helix-packings (Fig. 2b). RAP1 is a yeast transcription factor as well as a telomere

Fig. 3. The a-g phase of residues in an α -helix (a) and possible contacting of a Thr residue to the T and C bases (b, c). (a) In an α -helix the a and d phases are on the same side, while the b and f phases on the opposite side. Therefore, when the a and d phases are occupied by hydrophobic residues for packing against the rest of the protein, the b and f phases are exposed to the solvent and thus can be used for contacting the DNA bases. (b) and (c) T Thr residue can contact the T or C bases by hydrogen bonding and at the same time by hydrophobic interaction.⁴⁾ Such contacts coincide with the DNA-binding specificity of Hox11.¹³⁾

binding protein.⁹⁾ The DNA-binding domain of a RAP1 monomer seems to be further composed of two repeats and the each repeat has a twelve residue segment starting with His (Fig. 2b). Indeed some residues in the twelve residue segments have been identified as important for DNA-binding (David Shore, personal communication). The DNA-binding specificity of RAP1 has been experimentally identified and as a

Fig. 4. Base-residue contacts predicted for some transcription factors of the HD-bZip type (the probe helix family). (a) shows a key to the diagrams in (b)–(l). It is essentially a sketch of the DNA major groove; the two DNA strands (top—the W strand and bottom—the C strand, C1 pairs with W1, C2-W2, C3-W3) and the recognition helix (the centre) in which four positions, 1, 4, 5 and 8, face the DNA bases. The base-residue contacts found in (b)–(l) are accumulated (note that the same contacting pattern is shared by the proteins). In (b)–(l) specific contacts⁴⁾ are indicated by circles. The predictions are made according to the rules.^{4),5)} References for the binding specificity of the individual factors are Guazzi *et al.*¹⁴⁾ for TTF1 (thyroid nuclear factor 1) in (b), Goutte and Johnson¹²⁾ for Maf1 in (c), Ades and Sauer¹⁵⁾ for a mutant HD (QA50) in (d), Sessa *et al.*⁶⁾ for Athb in (e), Dear *et al.*¹⁶⁾ for Hox11 in (f), Morinaga *et al.*¹⁷⁾ for the first HD of ATBF1 (α -fetoprotein enhancer-binding factor 1) in (g), Neufeld *et al.*¹⁸⁾ for CDP (CCAAT-displacement protein) in (h), Kataoka¹⁹⁾ *et al.* for Maf in (i), Biggin *et al.*²⁰⁾ for Zeste in (j) and Gilson *et al.*⁹⁾ and Graham and Chambers²¹⁾ for two repeats of RAP1 in (k, l). See Fig. 3 for the Thr-T/C contact in (f).

whole it is consistent with the binding specificity expected from the recognition pattern of the bZip-HD type, although some details might be different since distortion of the DNA upon binding to RAP1 has been detected.⁹⁾

A transcription factor, E2, has a recognition helix of the HD-bZip type³⁾ (a probe helix) but folds into a

Fig. 5. Different ways of combining two recognition helices of the HD-bZip (probe helix) type. Those in a homodimer of Mat α 2 (a), a Mat α 1-Mat α 2 heterodimer (b), a E2 homodimer (c), a bZip dimer on the CRE type site and a bHLH dimer (d), and a bZip dimer on the AP1 type site and a HDZip dimer (e). Each diagram is a sketch of the DNA major groove, which are drawn in the same way as in Fig. 4(a). The three basepairs bound by each recognition helix are drawn hatched.

β -barrel.¹⁰⁾ Recognition helices which are closely related to the HD-bZip type are found in a bundle of four helices⁴⁾ (Fig. 1e) in a dimer of the basic-helix-loop-helix-zipper family.¹¹⁾ Thus, it seems now clear that a recognition helix of the HD-bZip type (a probe helix) can be incorporated into several different folds.

Different ways of combining two probe helices. In Fig. 4 we summarise base-residue contacts predicted for some transcription factors of the HD-bZip type, including of those mentioned in earlier sections. These fall into the same pattern (Fig. 4a) specific to this group.^{4),5)} On the basis of this and earlier identification of three basepairs directly bound by individual recognition helices, it is possible to describe the way(s) in which two recognition helices are combined in homo- and heterodimers of the proteins (Fig. 5). For example, DNA-binding of the Mat α 1-Mat α 2 hetero-dimer and the Mat α 2 homodimer has been characterised well by foot-printings experiments.¹²⁾ By using the three basepairs bound by Mat α 2 in a crystal structure¹³⁾ and those predicted to be bound by Mat α 1 (Fig. 4c) the way(s) in which these helices are combined can be described (Figs. 5a, b).

These spacing differences in the binding sites are consequences of differences in the way(s) in which these proteins dimerise. These spacing and dimerisation differences are important for discrimination—

even if individual recognition helices are properly designed, unless they are incorporated into an appropriate dimer structure so that their separation fit to the spacing between the two direct binding sites in the target DNA sequence, the DNA-binding of the protein will not be strong.

Acknowledgements. We thank Prof. H. Katsura and Dr. D. Smoller for informing us the literatures of the zipperless bZip proteins.

References

- 1) Duboule, D. ed. (1994) Guidebook to the Homeobox Genes, Oxford University Press, New York.
- 2) Hurst, H. C. (1994) Protein Profile vol. 1, Issue 2 Transcription factors 1: bZip proteins, Academic Press, London.
- 3) Suzuki, M. (1993) EMBO J. 12, 3221–3226.
- 4) Suzuki, M. (1994) Structure 2, 317–326.
- 5) Suzuki, M., and Yagi, N. (1994) Proc. Natl. Acad. Sci. U.S.A. 91, 12357–12361.
- 6) Sessa, G., Morelli, G., and Ruberti, I. (1993) EMBO J. 12, 3507–3517.
- 7) Bowerman, B., Eaton, B. A., and Priess, J. R. (1992) Cell 68, 1061–1075.
- 8) Smoller, D., Friedel, C., Schmid, A., Bettler, D., Lam, L., and Yedvobnick, B. (1990) Genes Develop. 4, 1688–1700.
- 9) Gilson, E., Roberge, M., Giraldo, R., Rhodes, D., and

- Gasser, S. M. (1993) *J. Mol. Biol.* **231**, 293–310.
- 10) Hegde, R. S., Grossman, S. R., Laimins, L. A., and Sigler, P. B. (1992) *Nature* **359**, 505–512.
- 11) Ferré-D'Amaré, A. R., Prendergast, G. C., Ziff, E. B., and Burley, S. K. (1993) *Nature* **363**, 38–45.
- 12) Goutte, C., and Johnson, A. D. (1994) *EMBO J.* **13**, 1434–1442.
- 13) Wolberger, C., Vershon, A. K., Liu, B., Johnson, A. D., and Pabo, C. O. (1991) *Cell* **67**, 517–528.
- 14) Guazzi, S., Price, M., De Felice, M., Damante, G., Mattei, 15) Ades, S. E., and Sauer, R. T. (1994) *Biochemistry* **33**, 9187–9194.
- 16) Dear, T. N., Sanchez-Garcia, I., and Rabbitts, T. H. (1993) *Proc. Natl. Acad. Sci. U.S.A.* **90**, 4431–4435.
- 17) Morinaga, T., Yasuda, H., Hashimoto, T., Higashio, K., and Tamaoki, T. (1991) *Mol. Cell Biol.* **11**, 6041–6049.
- 18) Neufeld, F. J., Skalnik, D. G., Lievens, P. M.-J., and Orkin, S. H. (1992) *Nature Genetics* **1**, 50–55.
- 19) Kataoka, K., Noda, M., and Nishizawa, M. (1994) *Mol. Cell Biol.* **14**, 700–712.
- 20) Biggin, M. D., Bickel, S., Benson, M., Pirrotta, V., and Tjian, R. (1988) *Cell* **53**, 713–722.
- 21) Graham, I. R., and Chambers, A. (1994) *Nucleic Acids Res.* **22**, 124–130.