Vol. 157, No. 1, 1988 BIOCHEMICAL AND BIOPHYSICAL RESEARCH COMMUNICATIONS
November 30, 1988 Pages 100-105

MODELING OF THE HOMOEO DOMAIN SUGGESTS SIMILAR STRUCTURE TO REPRESSORS

Panagiotis A. Tsonis[1,2], Vasili Carperos[2], and Teruna Shiahaan[2]

[1]VA Medical Center (111E)
Indiana University
Indianapolis, IN 46202

[2]La Jolla Cancer Research Foundation
10901 N. Torrey Pines Road
La Jolla, CA 92037

Received October 11, 1988

Summary The sequences of the homoeo domain containing the three α- helices were modeled based on secondary structure prediction, sequence homologies and coordinates of known helix-turn-helix motif containing DNA-binding proteins. The model reveals very similar three dimensional structure to repressors and suggest binding to DNA with its helix 3.

The homoeo box is a 60 amino acid region initially found in the products of some homoeotic genes in Drosophila and subsequently in most organisms (1-3). The degree of sequence conservation among homoeo boxes is astonishing. In some cases, homoeo boxes from genes isolated from fly and humans share more than 90% identity (4), Such a conservation during evolution comes to suggest that the homoeo box has a unique function in all animals. When the homoeo box sequences were first revealed it was found that they share sequence similarities with the helix-turn-helix motif of some DNA binding regulatory proteins, i.e. λ-cro, λ-repressor. This with the additional observation that the products of homoeo-box containing genes are localized in the nucleus (5) come to suggest that the homoeo box is in fact a regulatory domain that binds to DNA. Such a property has been shown for the product of the engrailed gene (6). Secondary structure prediction and sequence similarities show that the C-terminus of the homoeo box contains the helix-turn-helix motif similar to that found in DNA-binding proteins (7,8). The helices are :RIDIANAL; and ERQIKIWFQ. This motif is

0006-291X/88 $1.50

actually very conserved among all homoeo boxes. However, a close inspection of homoeo boxes reveals that sequences in the N-terminus, namely, **LELEKEFH** is also conserved in most homoeo boxes (9). In addition, this octapeptide is predicted to form an α-helix. Interestingly, the sequence **LELEKE** has been found in many proteins that are known to form -helices (9). Such an observation along with the homology of the helix-turn-helix motif of DNA-binding proteins to sequence in the C-terminus of the homoeo boxes suggests that there are three potential α-helices in the homoeo domain: helix 1: **LELEKEFH**; helix 2: **RIDIANAL**; and helix 3: **ERQIKIWFQ** (See Figure 1). Based on the above observations we attempted to model the three-dimensional structure of the homoeo domain in order to further understand its possible relationship to DNA-binding proteins and its interaction with DNA.

RESULTS AND DISCUSSION

For the modeling, we considered the sequences presented in Figure 1 from amino acid 16 where the first helix starts to amino acid 52 where the third helix ends. From the 16-52 amino acids, 16-23, 33-40, and 44-52 were considered α-helices (see above and boxed in Figure 1). The first amino acid, Leu 16, and the last, Glu 52, were considered neutral in order to avoid possible

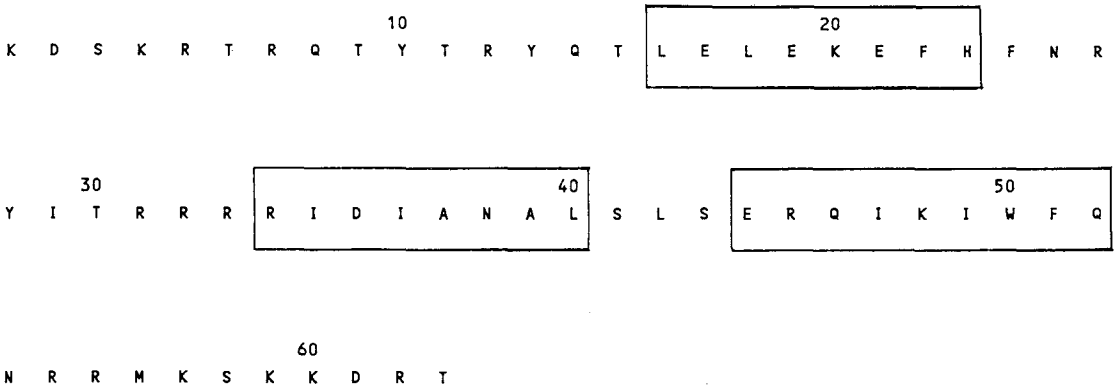

Fig. 1 Amino aid sequences of the homoeo domain (taken form the ftz gene [7]) used in the modeling. Sequences 16-52 used in the model presented in Fig. 2. The three helices are boxed.

circularization of the domain during minimization. Secondary structure predictions show the amino acids in between, to be in turns. A β-I turn was introduced at the Tyr 27. Sequences 33-52 were modeled using the coordinates from a protein containing the helix-turn-helix motif, the catabolite gene activator protein (CAP) (Brookhaven Data Bank file 2GAP) (10). The modeling was carried out in an Evans-Sutherland PS-390 computer with **INSIGHT** program and the energy minimization was done using MicroVax II computer equipped with the DISCOVER program (BIOSYM technologies). A stereo view of the predicted three-dimensional structure of the homoeo domain is shown in Figure 2. The backbone reveals the three characteristic helices, helix 1 being on the bottom of the model. A close look reveals that the three helices form a structure that is very similar to the one found in repressors (11-14). The above proposed structure can actually account for the mode of interaction between homoeo domain and DNA. As well established by crystallography, the helix-turn-helix structure of λ-repressor or λ-cro or other DNA-binding protein containing the helix-turn-helix motif is interacting with DNA in a fashion depicted in Figure 3. Helix 2 and helix 3 are embedded in the major groove of DNA; Helix 3 being the one to make contact with bases in the major groove. Helix 1 is placed in the model in similar fashion to helix 1 of the repressor and other DNA-binding proteins. The

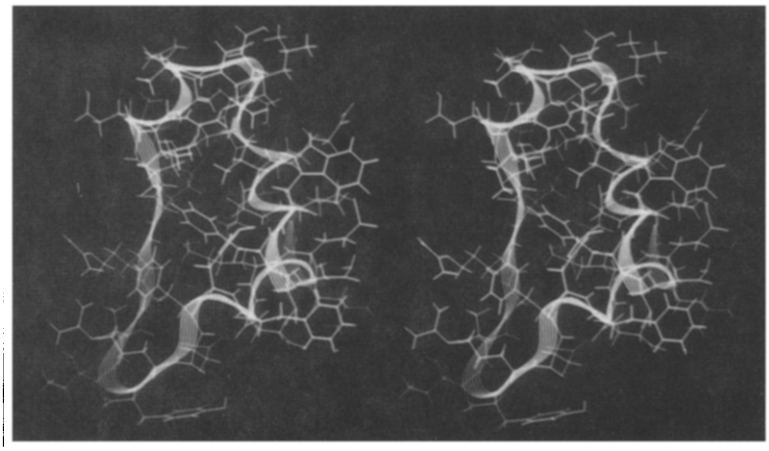

Fig. 2 Stereo view of the three dimensional structure of the homoeo domain predicted from our studies.

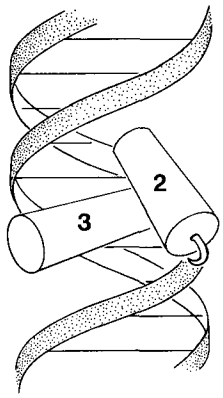

Fig. 3 A sketch of the λ-repressor depicted from ref 12 and its interaction with DNA.

above model is showing that the homoeo box could interact with DNA in a fashion similar to that of DNA-binding proteins. However, some other points warrant discussion. Most of the DNA-binding proteins interact with DNA as dimers or tetramers (10, 11, 12, 13). Such a case of possible dimers of homoeo boxes binding to two major grooves is not known. There is no indication of helix-helix interaction of two homoeo domains of hydrophobic kind. In fact, the whole homoeo box domain is highly hydrophilic, and a careful observation of the extending arm at the C-terminus of the homoeo box (amino acid 54-63) shows that it contains charged amino acids (6 out of 10 are Arg or Lys). The same is valid for the first 10 amino acids in the N-terminus of the homoeo box. In fact, the peptides containing the first 10 amino acids in the N-terminus and the last 10 amino acids in the C-terminus are very similar (15):

62	R		D	K	K	S	K	M	R	R	54
	*		*		*				*		
1	K		D	S	K	R	T		R	Q	T	9
	*			*	*	*			*		*	
55	R	M	K	S	K	K	D		R		T	63

This could argue that (1) homoeo boxes could interact with each other to form dimers or that (2) the homoeo box folds by the interaction of its N-and C-terminus and exposes the three helices in the major groove. Further investigation and crystallographic studies could answer this question. The amino acids L18, F22, Y27, L42, W50, and F51 have been found to be conserved in all homoeo boxes sequenced to date. An inspection of the model reveals that all of these amino acids are very exposed to the outside of the modeled domain, however, their possible function is not yet known. We believe that the above considerations and the predicted three dimensional structure of the homoeo domain support the notion that the homoeo box is composed of three potential helices folded and interacting with the DNA in a similar fashion to repressors.

ACKNOWLEDGMENTS

We would like to thank Mrs. Janet Paradise Rosselle for typing the manuscript. P.A.T. is an Arthritis Investigator.

REFERENCES

1. Scott, M.P. and Weiner, A.J. (1984) Structural relationships among genes that control development: Sequence homology between the Antennapedia, Ultrabithorax and fushi tarazu loci of Drosophila. Proc. Natl. Acad. Sci. USA 81:4115-4119.

2. McGinnis, W., Garber, R.L., Wirrz, J., Kuroiwa, A., Gehring, W.J. (1984) A homologous protein-coding sequence on Drosophila homeotic genes and its conservation in other metazoans. Cell 37:403-408.

3. Holland, P.W.H., and Hogan, B.L.M. (1986) Phylogenetic distribution of antennapedia-like homeoboxes. Nature 321:251-253

4. Burglin, T.R. (1988) The yeast regulatory gene PHO2 encodes a homeobox. Cell 53:334-340.

5. Tsonis, P.A., and Adamson, E.D. (1986) Specific expression of homeobox containing genes during induced differentiation of embryonal carcinoma cells. Biochem. Biophys. Res. Comm. 137:520-527.

6. Desplan, C., Theis, J., and O'Farrell, P.H. (1985) The Drosophila developmental gene, engrailed encodes a sequence-specific DNA binding activity. Nature 318:630-635.

7. Laughon, A., Scott, M.P. (1984) Sequence of a Drosophila segmentation gene: protein structure homology with DNA-binding proteins. Nature 310:25-31.

8. Sauer, R.T., Yocum R.R., Doolittle, R.F., Lewis, M., and Pabo, C.O. (1982) Homology among DNA-binding proteins suggests use of a conserved super-secondary structure. Nature 298:447-451.

9. Tsonis, P.A., and Lambris, J.D. (1986) Similarities between a conserved sequence element of homeoboxes and other genes. FEBS Lett. 194:263-266.

10. McKay, D.B., Weber, I.T., and Steitz. T.A. (1982) Structure of a catabolite gene activator protein of a 2.9 Å resolution. J. Biol. Chem. 257:9518-9524.

11. Anderson, W.F., Ohlendorf, D.H., Takeda, Y., and Matthews, B.W. (1981) Structure of the cro repressor from bacteriophage lambda and its interaction with DNA. Nature 290:754-758.

12. Pabo, C.O., and Lewis, M. (1982) The operator-binding domain of lambda-repressor: Structure and DNA-binding. Nature 298:443-447.

13. Anderson, J.E., Ptashne, M., and Harrison, S.C. (1987) Structure of the repressor-operator complex of bacteriophage 434. Nature 326:846-852.

14. Kaplein, R., Boeleus, R., Scheek, R.M., van Gunsteren, W.F. (1988) Protein structures from NMR. Biochemistry 27:5389-5395.

15. Tsonis, P.A. Pattern formation: From theoretical models to molecular biology. In vivo (in press).