Vol. 168, No. 1, 1990 BIOCHEMICAL AND BIOPHYSICAL RESEARCH COMMUNICATIONS
April 16, 1990 Pages 254-260

A DNA-BINDING HOMEODOMAIN IN HISTONE H1

Riitta-Maaria Mannermaa and Jouko Oikarinen

Collagen Research Unit, Biocenter and Department of Medical Biochemistry, University of Oulu, Kajaanintie 52A, SF-90220 Oulu, Finland

Received February 23, 1990

The structure of the globular domain of chicken histone H1 was compared here with that of the DNA-binding homeodomain in the Drosophila Antp protein, and they were observed to display considerable similarity. Both of them consist of three or four α-helices separated by well-defined turns. Charged residues in the aminoterminal end of α3 are therefore suggested to be responsible for sequence-specific recognition of DNA by the histone. In addition, α2 of H1, with a short leucine zipper in it, may be capable of protein-protein interaction in a similar manner to the other homeodomains. © 1990 Academic Press, Inc.

Homeotic genes, the sequential expression of which is a crucial requirement for adequate body development in eukaryotes, code for transcriptional regulators primarily located in the nucleus (1-3). Their function has to date been elucidated in most detail in Drosophila and mouse, but the presence of these ubiquitous proteins has also been confirmed in man, where they may similarly control embryonal development.

The well-conserved homeodomain (HD) in a number of homeotic proteins, which is presumably responsible for the sequence-specific recognition of DNA and consists of about 70 amino acid residues, is in most cases located in the aminoterminal parts of the molecules and folds into three or four α-helices separated by well-defined turns (4,5). A helix-turn-helix structure comparable with those in a number of bacterial DNA-binding proteins is formed by α2 and α3, the aminoterminal end of α3 probably being responsible for the specific interaction in the major groove of the DNA secondary structure (6). On the other hand, α2 of the HD has in some cases been shown to be capable specific protein-protein interaction with other regulatory proteins (7).

The DNA in eukaryotes is packed into nucleosomes in which two turns of DNA are twined around a core structure that consists of pairs of histones H2A, H2B, H3 and H4 (8,9), while H1 binds the nucleosome externally to the hinge region where the DNA strands enter and leave the complex (10). It is thought to interact with all three DNA strands it is

0006-291X/90 $1.50

surrounded by. Interaction of H1 with H3, which is just opposite it on the bottom of the nucleosome, can also be demonstrated by cross-linking experiments, and may thus serve to seal the structure (11).

Polypeptide chains of various H1 subtypes comprise an N tail of 35 residues, a C tail of approximately 110 to 120 residues and mostly α-helical central globular domain of 70 residues, these accounting for the total length of 220 residues (12). The globular domain of H1 is composed of three α-helices as deduced from crystallographic data, and its α-carbon backbone is superimposable on the DNA-binding domain of the bacterial cyclic AMP-receptor protein (CRP), in which DNA recognition takes places through an α-turn-α structure (13).

Aggregation of nucleosomes to higher order solenoid structures appears to be a crucial step in the repression of certain areas of the chromatin in differentiated cells, and this process may be triggered by H1 (14), which is thought to act as a eukaryotic repressor (15,16). We have demonstrated recently by means of H1 isolated from rat liver by DNA affinity chromatography (17) and a histone H1.03-*lacZ* fusion protein produced in *E. coli* that H1 binds to the 5'-TTGGCAnnnTGCCAA-3' motif (unpublished results). This kind of interaction has also been confirmed using H1 isolated by conventional PCA extraction (18), the results thus suggesting that interaction of H1 with the specific DNA sequence may be a determinant of nucleosome positioning (19). These results are, however, contradictory to the earlier proposals that H1 interacts in a non-sequence-specific manner with a DNA minor groove (20,21).

Primary, secondary and tertiary structures of the H1 subtypes are compared here with those of the DNA-binding HD, and H1 is observed to contain an HD-like structure, presumably conferring it the capability for sequence-specific interaction with DNA.

MATERIALS AND METHODS

The computer programs described by Pustell and Kafatos (22-24) were used for DNA and protein sequence analysis, and the GeneBank Genetic Sequence Data Bank to search for the various sequences. The Desktop Molecular Modeller Program from Oxford University Press was used to visualize 3D structures of various regions of the H1 globular domain.

RESULTS AND DISCUSSION

The primary and secondary structures of an HD from the *Antennapedia* protein (1,5), which is regarded as a model for this unique functional unit, were compared with those of H1 (12,13). The primary structure of the globular domain of the histone may be aligned with that of the HD, the two sequences displaying 12% identity and 34% similarity when the

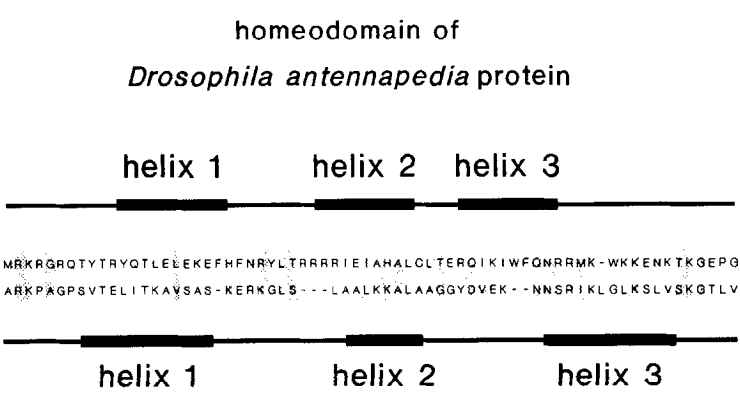

Fig.1. Alignment of the primary structure of the Antennapedia homeodomain with that of the globular domain of histone H1. Single letter codes for amino acids are used, and gaps introduced for optimal alignment. Residues identical or similar to those in the homeodomain are hatched, the amino acids being grouped (V, I, M, L), (P, A, G, S, T), (D, E, N, Q), (K, R, H) and (F, Y, W). Positions of the three α-helices in H5, representing H1 here, and the homeodomain secondary structures are depicted.

amino acids were grouped as generally approved (Fig. 1). The degree of primary structure homology is not convincing enough to substantiate similar function of the two domains, however, although it should be noted that the homologous area comprises principally the whole globular domain and only that in H1.

The secondary structure of H1 is known on the basis of that of an avian erythrocyte-specific H1 variant called H5 (Fig. 1). The globular domain of H5 consists of three or four α-helices (21). α2 and α3 lie almost antiparallel and are held together by α1, as deduced from crystallographic data (Fig. 2). By analogy to CRP, the α-turn-α structure between α2 and α3 may be deduced from DNA recognition. The model HD also folds into three or four α-helices separated by a well-defined turn between α2 and α3 (Fig. 1). The similarity of the secondary structures becomes illustrative, when the primary structures of the globular domain of H1.01 and Antp HD are aligned as described above. The helices in the two polypeptide chains are also aligned, the putative sites for interaction with DNA then being superimposed in the two. The tertiary structure of the HD may also be superimposed on the H1/H5 globular domain in a similar manner to that of the CRP DNA-binding domain.

It should be noted that the residues assumed to be conserved in various HDs (1) are not conserved in the H1 subtypes. Hydropathy

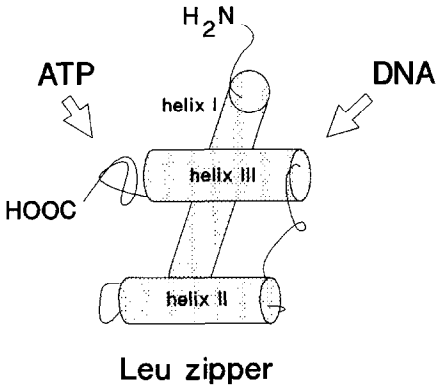

Fig.2. A schematic representation of the histone H1 tertiary structure. Helices are depicted by the cylinders, and the putative sites for sequence-specific interaction with DNA and recognition of ATP by the arrows. A short leucine zipper in α2 that may capable of protein-protein interaction is also depicted.

analysis revealed, however, that the H1.01 globular domain and Antp HD share similar charge distribution along their α-helices, with the exception of α3 (Fig. 3). They may thus easily fold to similar higher order structures, as has already been confirmed by the crystallographic experiments, but the analysis suggests that the two domains may even be capable of similar interactions with other molecules such as DNA.

Residues in the N-terminal part of α3 may be suggested in the case of H1 to be involved in recognition of DNA bases on the bottom of the major groove. The well-conserved Asn^{77} and Ser^{78} residues in the case of chicken H1.01 project to the side of the helix facing the DNA groove

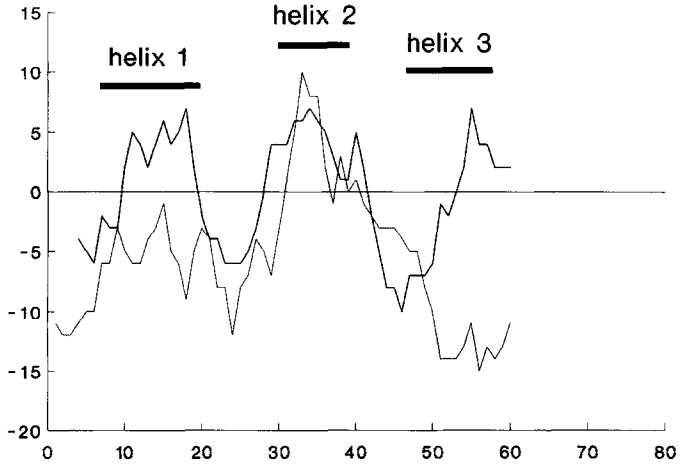

Fig.3. Hydropathy analysis. Hydropathy plot of the homeodomain-like structure in H1 (thick line) and that of the Antennapedia homeodomain (thin line). Positions of α-helices in H1 are depicted by horizontal lines.

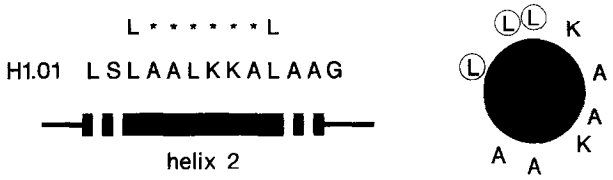

Fig.4. A leucine zipper in histone H1. Alignment of the primary structure of α2 in the globular domain of chicken H1.01 with the consensus for leucine zippers. The helical structure of H5 is shown below, and the helical wheel on the right, to demonstrate the projection of hydrophobic residues to one side of the helix.

when a model of the region from Gly^{69} to Lys^{81} is constructed by a computer using energy-minimization and coordinates from the 3D structure of H5 (data not shown). On the other hand, a leucine zipper may be identified in α2 (Fig. 4) (25). When α2 is added to the computer model, Leu 59 and 66 project to the opposite side with respect to the residues thought to interact with DNA, and the leucine zipper is obviously capable of interacting with other proteins even when the H1 molecule is bound to DNA. These results thus suggest that the globular domain of H1/H5 is an HD that is capable of interacting both with DNA and proteins, as demonstrated previously in the case of the other HDs.

A region in H1/H5 displays homology with nucleotide-binding sites in various enzymes capable of using nucleotide substrates, such as protein kinases and G_α-proteins (26). This region resides in the hinge between the globular domain and C tail, and may therefore be of importance in modulating DNA binding. Actually, we have demonstrated very recently that nucleotides bind to H1 and that low concentrations of ATP stimulate binding of H1 to its recognition motif on DNA (unpublished results).

When thinking of the function of the leucine zipper in H1, the most interesting possibility is that it may mediate nucleosome aggregation. When the angles between α2 and α3 are taken into consideration, the H1 molecules dimerized through the N tails and bound to successive nucleosomes may easily be staggered through α2 in the latter of the two monomers in an upstream H1 dimer and the former in a downstream H1 dimer, and this would result in the formation of a continuous 30 nm solenoid structure with approximately six nucleosomes per turn, the primary constituent of repressed heterochromatin.

A novel DNA-binding unit, called an SPKK-repeat, has very recently been identified in histones, a primary interaction taking place with an exceptionally narrow minor groove in an A+T rich DNA region (27-29). This kind of interaction may occur between the N-terminal ends of H2A and H2B and the C tail of H1/H5 and DNA. These results suggest, there-

fore, that the C tail of H1 may also be involved in the H1-DNA interaction, in addition to the globular domain.

Sequence-specific interaction of the H1/H5 globular domain and interaction of the SPKK-repeats in the C tail with A+T rich regions may together be responsible for the binding of H1 to the nucleosome hinge region. It may easily be envisaged that the C tails of H1 serve in this model to gather the entering and leaving DNA strands and thereby seal the chromatosome. In addition, interaction of H1 with the core nucleosome may occur. These interactions of H1 with DNA and proteins may thus serve as driving forces for generation of the irregular but constant spacing of nucleosomes observed in vivo and tight packaging of the DNA into nucleosomes (30), while ordered aggregation of H1 to homopolymers through the leucine zipper in α2 may be responsible for the aggregation of the nucleosomes and consequent H1-directed repression of eukaryotic chromatin activity (14,31).

It is assumed in the above model for the H1 function that H1 binds to DNA as a dimer but no direct evidence exists to date to confirm this fact. In contrast, it is generally believed that only one H1 monomer exists per nucleosome (32). There are some hints, however, suggesting that homodimers of H1 may be formed (17,33). It remains to be elucidated in the future in more detail how the higher order aggregates and dimers of H1 are formed. In order to study the interactions of H1, we have recently introduced various deletions and point mutations to the H1-lacZ fusion protein produced in E. coli.

REFERENCES

1. Scott, M.P., Tamkun, J.W., and Hartzell, III, G.W. (1989) Biochim. Biophys. Acta **989**, 25-48.
2. Wright, C.V.E., Cho, K.W.Y., Oliver, G., and Derobertis, E.M. (1989) TIBS **14**, 52-56.
3. Levine, M., and Hoey, T. (1988) Cell **55**, 537-540.
4. Otting, G., Qian, Y.-Q., Müller, M., Affolter, M., Gehring, W., and Wüthrich, K. (1988) EMBO J. **7**, 4305-4309.
5. Qian, Y.-Q., Billeter, M., Otting, G., Muller, M., Gehring, W.J., and Wuthrich, K. (1989) Cell **59**, 573-580.
6. Brennan, R.G., and Matthews, B.W. (1989) TIBS **14**, 286-290.
7. Stern, S., Tanaka, M., and Herr, W. (1989) Nature **341**, 624-630.
8. Klug, A., Rhodes, D., Smith, J., Finch, J.T., and Thomas, J.O. (1980) Nature **287**, 509-516.
9. Morse, R.H., and Simpson, R.T. (1988) Cell **54**, 285-287.
10. Staynov, Z., and Crane-Robinson, C. (1988) EMBO J. **7**, 3685-3691.
11. Glotov, B.O., Itkes, A.V., Nikolaev, L.G., and Severin, E.S. (1978) FEBS Lett. **91**, 149-152.
12. Wells, D., and McBride, C. (1989) Nucl. Acids Res. **17**, R311-R346.
13. Clore, G.M., Gronenborn, A.M., Nilges, M., Sukumaran, D.K., and Zarbock, J. (1987) EMBO J. **6**, 1833-1842.
14. Weintraub, H. (1985) Cell **42**, 705-711.
15. Wolffe, A.P. (1989) EMBO J. **8**, 527-538.

16. Sun, J.-M., Wiaderkiewicz, R., and Ruiz-Carrillo, A. (1989) Science **245**, 68-71.
17. Ristiniemi, J., and Oikarinen, J. (1989) J. Biol. Chem. **264**, 2164-2174.
18. Sevall, J.S. (1988) Biochemistry **27**, 5038-5044.
19. Richard-Foy, H., and Hager, G.L. (1987) EMBO J. **6**, 2321-2328.
20. Crane-Robinson, C., and Ptitsyn, O.B. (1989) Protein Engineering **2**, 577-582.
21. Turnell, W.G., Satchwell, S.C., and Travers, A.A. (1988) FEBS Lett. **232**, 263-268.
22. Pustell, J., and Kafatos, F.C. (1982) Nucl. Acids Res. **10**, 51-59.
23. Pustell, J., and Kafatos, F.C. (1984) Nucl. Acids Res. **12**, 643-655.
24. Pustell, J., and kafatos, F.C. (1986) Nucl. Acids Res. **14**, 479-488.
25. Landschulz, W.H., Johnson, P.F., and McKnight, S.L. (1988) Science **240**, 1759-1764.
26. Ristiniemi, J., and Oikarinen, J. (1988) Biochem. Biophys. Res. Commun. **153**, 783-791.
27. Suzuki, M. (1989) EMBO J. **8**, 797-804.
28. Churchill, M.E.A., and Suzuki, M. (1989) EMBO J. **8**, 4189-4195.
29. Izaurralde, E., Käs, E., and Laemmli, U.K. (1989) J. Mol. Biol. **210**, 573-585.
30. Rodriquez-Campos, A., Shimamura, A., and Worcel, A. (1989) J. Mol. Biol. **209**, 135-150.
31. Shimamura, A., Sapp, M., Rodriguez-Campos, A., and Worcel, A. (1989) Mol. Cell. Biol. **9**, 5573-5584.
32. Ring, D., and Cole, R.D. (1983) J. Biol. Chem. **258**, 15361-15364.
33. Nilsson, P., Hallberg, B., Thornell, A., and Grundström, T. (1989) Nucl. Acids Res. **17**, 4061-4075.