The Helix-Turn-Helix DNA Binding Motif

Richard G. Brennan and Brian W. Matthews
From the Institute of Molecular Biology and Department of Physics, University of Oregon, Eugene, Oregon 97403

One of the most striking findings to emerge from the determinations of the structures of Cro, CAP, and λ-repressor (1–3) was that these three sequence-specific DNA-binding proteins had in common a substructure now named the helix-turn-helix motif. The apparent importance of this motif was suggested not only by its striking similarity to the respective structures (4,5) but also from amino acid sequence comparisons (6–11) which indicated that a similar motif might occur in other DNA-binding proteins.

It was proposed (1, 3, 12–14) that the helix-turn-helix motif interacts with operator DNA in a manner similar to that shown in Fig. 1 for Cro. For Cro, the recognition of a specific sequence on DNA was thought to be achieved by a network of hydrogen bonds and other contacts between the side chains of the protein and the parts of the base pairs exposed within the grooves of the DNA. The second helix of the helix-turn-helix motif, which was thought to bind within the major groove of the DNA, was presumed to be especially important in recognition.

The purpose of this report is not to review, in detail, the structures of the repressor-operator complexes. Rather, we will first discuss the use of amino acid sequence comparisons to locate putative helix-turn-helix motifs. Then we will review the structural correspondence between known helix-turn-helix units. Finally, the role of the helix-turn-helix in recognition of sequence-specific binding sites on DNA will be evaluated in light of the recently defined repressor-operator complexes.

Sequence Comparisons

There has been considerable interest in the use of amino acid sequences to identify putative helix-turn-helix motifs in proteins of unknown structure. Predictions that have been confirmed by subsequent structure determination include the helix-turn-helix units of λ-repressor (3, 6), trp repressor (9, 23, 24), 434 Cro protein, and 434 repressor (6, 7, 19, 22, 25, 26). Prediction of a helix-turn-helix in lac repressor (7, 8) has been verified by two-dimensional NMR spectroscopy (27), and predictions for the biotin repressor of E. coli (28) are strongly supported by genetic analysis (29).

These successful predictions give confidence in the methods on which they were based. They also provide a representative set of examples against which new predictions can be assessed.

In comparing amino acid sequences, especially those that have marginal correspondence, it is essential that defined statistical procedures be used. Also it is very desirable that the significance of a proposed matching of sequences be assessed by controls carried out with unrelated or shuffled sequences (e.g. see Refs. 8, 30, and 31). We will focus our discussion on the method employed in our laboratory to search for putative helix-turn-helix units and, as listed above, found to be effective in a number of cases.

A new amino acid sequence is compared with a “master set” of prealigned sequences, 22 amino acids in length, taken from proteins known or assumed to contain helix-turn-helix units (9, 31). Our "master set" consists of 10 proteins (Fig. 2). All but three of these (P22 Cro, P22 repressor cII, and λ cII protein) are now confirmed as containing a helix-turn-helix motif (see above). An amino acid sequence of interest is compared with the master set by evaluating, in turn, the correspondence between every possible 22-amino acid segment and the master set of sequences. For largely historical reasons it is convenient to use a system of scoring in which sequences with the best correspondence with the master set have the lowest scores. The score for each 22-amino acid segment is determined by summing the total number of amino acids in the master set that differ with the segment being tested. The score so obtained is normalized by dividing by 220 (i.e. 22 amino acids × 10 sequences in the master set) to give the average amino acid change per codon (AAC). Scores obtained from comparing many 22-amino acid segments with the master set are pooled to provide statistical estimates of the significance of an individual AAC score (9, 31).

Segments with the lowest scores (AAC values) are the most likely candidates as helix-turn-helix motifs. It is not possible to draw a sharp dividing line between those proteins that do have a helix-turn-helix and those that probably do not. If, however, a 22-amino acid segment is found with an AAC score of 0.80 or less, it can be regarded as a strong candidate for a helix-turn-helix motif. Such sequences are shown in Fig. 2.1

This procedure successfully predicts all known examples of helix-turn-helix motifs in sequence-specific DNA-binding proteins. In addition it does not misidentify any 22-amino acid segments that are known to be helix-turn-helix units.

An amino acid sequence that corresponds to a putative helix-turn-helix motif can be given greater credence if it is compatible with known stereochemistry (9). Inspection of the structure of Cro suggested that a helix-turn-helix unit would probably have the following characteristics, with residue numbering as in Fig. 2 (9). (i) Residue 9 should be a glycine (having a conformation rare for nonglycines). (ii) Residues 4 and 15 should not be charged (these residues being buried close to DNA). (iii) The residues of the turn are always small.

1 Tables giving the AAC scores for these and other DNA-binding proteins, as well as statistical estimates of significance and a full set of references, are available on request.

FIG. 1. Model for the presumed complex of a dimer of Cro repressor with DNA (1, 12). For clarity the side chains of the residues within the helix-turn-helix units have been removed, and only the a-carbon atoms are shown as small white spheres. The helix-turn-helix on the right of the figure is most obvious. The first helix extends from right to left; the second helix (the "recognition helix") is seen end-on within the major groove of the DNA. For the remainder of the protein acidic residues are drawn red, basic residues blue, hydrophilic noncharged residues are green, and hydrophobic residues are shown gray (original figure prepared by Dr. D. H. Ohlendorf, University of Oregon and Du Pont).
FIG. 2. Amino acid segments that agree best with a master set of 10 helix-turn-helix motifs. The score (AAC value) gives the correspondence between each segment and the master set (lower values indicate better agreement). The number to the left of each segment is the amino acid sequence number of the first residue in the segment. The numbering at the top of the figure follows the convention of Pabo and Sauer (17). Amino acids that are most strongly conserved (Ala or Gly in position 5, Gly in position 9, and Ile or Val in position 15) are underlined.

Master set of sequences

| Protein | Sequence | Score |
|---------|----------|-------|
| λ Cro   | 14-FGQTKT4KDLCVYQSAJNKAIH | 0.70  |
| 434 Cro | 16-MTQTELATKAGVKQQLIEAILH | 0.66  |
| P22 Cro | 11-GTQRAVAKLIGISDAAYSQWKE | 0.67  |
| λ Repressor | 31-LSPEGAVSDKMCGQSGVCALFP | 0.70  |
| 434 Rep. | 16-NQAEQAQVGVTQQLSIQEQLEN | 0.65  |
| P22 Rep.(C7) | 19-IKTQAAALGKYQYSNYASQWHER | 0.67  |
| λ cII | 26-LGCTEKTEAKAVGVQDKSQJRSWRK | 0.75  |
| lacR | 4-VTLTDAEYAGVSTQTXSVRN | 0.69  |
| CAP | 167-ITRQIEGQIVGCSRETITCRCSL | 0.70  |
| trp R | 66-MSQRELKNCAGCAIATICTCSN | 0.71  |

Other DNA-binding proteins

| Protein | Sequence | Score |
|---------|----------|-------|
| cl P22 | 23-KCQRKVAEACINRESQISRWMK | 0.71  |
| CalK | 7-AITKEVRALGQEVATYSRVIY | 0.73  |
| TetR | 23-ILTTRKLAGEQLVEPQTLWVHK | 0.75  |
| TetR | 23-ILTTRKLAGEQLVEPQTLWVHK | 0.75  |
| DeoK | 22-LHRLDAALLGVSMTETBRDIN | 0.75  |
| o43 | 330-KTLCEYVEFVGTREIRQIEA | 0.75  |
| λ Rep. | 434-LSRQELAQKGVTTQQSLEQLEN | 0.65  |
| P22 Rep.(C7) | 19-IKTQAAALGKYQYSNYASQWHER | 0.67  |
| λ cII | 26-LGCTEKTEAKAVGVQDKSQJRSWRK | 0.75  |
| lacR | 4-VTLTDAEYAGVSTQTXSVRN | 0.69  |
| CAP | 167-ITRQIEGQIVGCSRETITCRCSL | 0.70  |
| trp R | 66-MSQRELKNCAGCAIATICTCSN | 0.71  |

(iii) Residues 3-8 and 15-20 should not be proline (being within α-helices). (iv) Residue 5 should not be a β-branched residue (being wedged between the two helices). There are a number of proteins that have been proposed to include potential helix-turn-helix motifs but agree poorly with the master set of reference sequences used here. Examples include homeo boxes, histones, some transcriptional activator proteins, resolvases and inversion proteins, and the AIDS protein "tat." The poor scores associated with these proteins are due to the low matches of the motif to the reference sequences used here. (34)
Minireview: Helix-Turn-Helix DNA Binding Motif

Structural Comparisons

The initial recognition of the significance of the helix-turn-helix motif derived as much from structural comparisons as from considerations of amino acid sequences. In particular, CAP and Cro were seen to contain a similar helix-turn-helix substructure in their presumed DNA binding regions even though the rest of these protein structures had little in common (4). 24 a-carbon atoms (residues 13–36) including the helix-turn-helix of Cro could be superimposed with a root mean square discrepancy of 1.1 Å on the corresponding 24 a-carbons (165–189) of CAP (4). The significance of this result was enhanced by the finding that the structure ofpressor also contained a helix-turn-helix substructure (5). In this case 23 a-carbons (residues 31–53) could be superimposed on residues 14–36 of Cro with a root mean square discrepancy of only 0.7 Å. A systematic search through all protein structures in the Brookhaven Data Bank failed to find any 22 a-carbon segment that corresponded to the helix-turn-helix as seen in Cro, a-repressor, and CAP (4, 35) (Fig. 3).

Now that the helix-turn-helix has been observed in a number of other DNA-binding proteins, it is possible to extend the initial structural comparisons. The correspondences between the helix-turn-helix unit of Cro (i) and those seen in CAP (2), a-repressor (3), trp repressor (24), phase 434 Cro (25), and phase 434 headpiece (26) is given in Table I. In making these comparisons Cro was chosen as the "reference" structure, in part because it was used for the initial comparisons, but also because it agrees better with most of the other helix-turn-helix motifs than they do among themselves. The backbone segment that is most precisely conserved in all six of the above proteins consists of 21 a-carbon atoms (Table I). Outside this region, the individual structures may differ substantially.

There are three other proteins that have been reported to contain substructures similar to the helix-turn-helix motif. The first of these is "DNA-binding protein II," a nonspecific DNA-binding protein from Bacillus steaothermophilus (36), the second is cytochrome c peroxidase, and the third is the ribosomal protein L7/L12 (Table I). The latter two helix-turn-helix structures were found in a recent search through the Brookhaven Data Bank by Richardson and Richardson (57). Although the helix-turn-helix units in these three proteins correspond reasonably well with that of Cro, none agrees as well as the helix-turn-helix units in the proteins that are known to bind DNA sequence specifically (Fig. 3). Nevertheless, it is intriguing that two of the three examples are proteins that interact with nucleic acids. The amino acid sequences within these three helix-turn-helix units do not obviously correspond with the master set of 10 sequences. However, the amino acid sequences of several homologs of the non-specific DNA-binding protein have been determined and one of these, Rhizobium meliloti (DBP II (Rm)) yields a score of 0.80 (Fig. 2). Furthermore, this score is for the set of 22 amino acids that correspond to the observed helix-turn-helix in the B. steaothermophilus enzyme. The significance, if any, of this correspondence is a matter for debate.

Role of the Helix-Turn-Helix in Recognition

For Cro protein the "recognition helix" (i.e. the second helix of the helix-turn-helix motif) protrudes from the surface of the protein and matches in shape the major groove of the DNA (Fig. 1). This complementarity in shape can facilitate binding of the protein to both sequence-specific and nonspecific sites on the DNA. As shown in Table II, the backbone conformations of the helix-turn-helix motifs in other sequence-specific DNA-binding proteins are very similar to that seen in Cro. Notwithstanding this overall similarity, the helix-turn-helix motifs do not interact with the DNA in exactly the same way in each repressor-operator complex (19–23). This had been anticipated from early structural comparisons of Cro and a-repressor (5, 13), although "helix-swap" experiments (38, 39) suggested similar rather than dissimilar geometries of binding. It can now be seen directly in the different repressor-operator complexes that the respective helix-turn-helix motifs adopt approximately similar but distinctly different binding geometries with respect to the DNA operator (16–23).

The helix-turn-helix does not function as a fixed "reading head" that is always aligned in the same way relative to the DNA. Rather, recognition is achieved by a combination of factors. Specific hydrogen bonds and van der Waals contacts between the side chains of the protein and the parts of the base pairs exposed within the grooves of the DNA clearly play a major role. Appropriately placed water molecules that bridge between the protein and the DNA may also participate in the recognition of a specific sequence. The DNA may also adapt to enhance contacts within the grooves as well as interactions with the phosphate backbone.

The role of the helix-turn-helix appears to be to provide a rigid underlying framework that supports the recognition surface of the protein but not to define how that recognition surface is aligned with the DNA.

The complexes involving X-repressor headpiece, 434 Cro, and 434 repressor headpiece all support the expectation that these proteins recognize their specific binding sites on the DNA by direct contacts with the edges of the base pairs exposed within the grooves of the DNA (19–22). In the complex of trp repressor, however, there are no

![FIG. 3. Histogram showing the result of a search through the Brookhaven Data Bank for helix-turn-helix motifs. The inset shows the superposition of the helix-turn-helix motifs as initially seen in Cro and CAP (4). The horizontal axis is the root-mean-square discrepancy between the 24 a-carbon atoms in the helix-turn-helix of Cro and all possible 24 a-carbon segments for the known protein structures. Also shown are the root mean square discrepancies between the helix-turn-helix of Cro and the eight other proteins listed in Table I. Note that the histogram is taken from the original search (4, 34) and was carried out using the 24 a-carbon atoms 13–36 of Cro repressor. In contrast, the individual values 1–8 shown in the figure were calculated for 21 a-carbon segment 16–36. As is clear from the figure and from Table I, this 21 a-carbon segment is well conserved in all known repressor and activator crystal structures. Beyond the 21 a-carbons, however, the protein backbones may differ substantially in the different structures.](image-url)
direct contacts between the protein and the base pairs that have been shown in to be important in recognition. Sigler and colleagues (23) argue that trp repressor provides an example of indirect readout. In other words, they suggest that trp operator is recognized indirectly through sequence-specific changes induced in the geometry of the phospho backbone that, in turn, permit the formation of a tight complex with the protein. This rationalization for the specific binding of trp repressor poses several questions. In particular, the observed distortions in the structure of the trp operator are not large, and it is not obvious why different DNA sequences could not assume conformations similar to that adopted by the trp operator. The crystals of the trp repressor-operator complex were grown from 50 mM NaCl, 11 mM CaCl₂, 10 mM cadaverine and 35% methylpentanediol (23). Such slow salt, high alcohol conditions would be expected to increase the affinity of nonspecific DNA binding. Indeed, von Hippel and co-workers (40) have shown that in low salt the affinity of lac repressor for nonspecific sites on DNA can approach, if not exceed, its affinity for the lac operator. Furthermore, they have also shown that the nonspecific binding of lac repressor is enhanced by the presence of glycerol (41). If trp repressor behaves similarly to lac repressor, it raises the possibility, first suggested to us by von Hippel, that the interhelix angle is not invariant. In this scenario, the evolution of DNA-protein complexes would be heavily constrained where the angle between the two a-helices. On the other hand, the recent evidence (42-44) suggests that the helix-turn-helix unit of CAP has a glycine and a proline at position 1 and have suggested that this "invariant" contact helps maintain the angle between the two a-helices. It seems unlikely that the helix-turn-helix unit has a conformation that is energetically favored during the folding process. Inspection of observed helix-turn-helix motifs shows that there can be sequences that bind to DNA with different alignments. There is also no obvious structural reason for the conservation of the helix-turn-helix motif. Because structures corresponding to the helix-turn-helix motif are not found in protein structures in general (Fig. 3), it seems unlikely that the helix-turn-helix unit has a conformation that is energetically favored during the folding process. Inspection of observed helix-turn-helix motifs shows that there can be sequences that bind to DNA with different alignments.

REFERENCES

1. Anderson, W. F., Ohlendorf, D. H., Takeda, Y., and Matthews, B. W. (1981) Nature 290, 754-758
2. McKay, D. B., and Steitz, T. A. (1991) Nature 349, 744-749
3. Pabo, C. O., and Lewis, M. (1983) Nature 306, 440-447
4. Steitz, T. A., Ohlendorf, D. H., McKay, D. B., Anderson, W. F., and Matthews, B. W. (1983) Proc. Natl. Acad. Sci. U. S. A. 79, 3097-3100
5. Ohlendorf, D. H., Anderson, W. F., Pabo, C. O., and Matthews, B. W. (1983) J. Mol. Biol. 169, 757-769
6. Anderson, W. F., Takeda, Y., Ohlendorf, D. H., and Matthews, B. W. (1982) J. Mol. Biol. 159, 145-153
7. Sauer, R. T., Yocum, R. R., Doolittle, R. F., Lewis, M., and Pabo, C. O. (1986) Nature 323, 467-471
8. Matthews, R. W., Ohlendorf, D. H., Anderson, W. F., and Takeda, Y. (1989) Proc. Natl. Acad. Sci. U. S. A. 79, 1428-1432
9. Ohlendorf, D. H., Anderson, W. F., and Matthews, B. W. (1983) J. Mol. Biol. 19, 109-114
10. Gejo-Sanzeactive, B., and Cossart, P. (1982) EMBO J. 6, 591-595
11. Weber, J. T., McKay, D. B., and Steitz, T. A. (1982) Nature 301, 5105-5109
12. Ohlendorf, D. H., Anderson, W. F., Fisher, R. G., Takeda, Y., and Matthews, B. W. (1983) Nature 308, 718-720
13. Lewis, M., Jeffrey, A., Wang, J., Ladner, R., Ptashne, M., and Pabo, C. O. (1982) Cold Spring Harbor Symp. Quant. Biol. 47, 435-440
14. Weber, J. T., and Steitz, T. A. (1984) Proc. Natl. Acad. Sci. U. S. A. 81, 3097-3097
15. Ohlendorf, D. H., and Matthews, B. W. (1983) Annu. Rev. Biophys. Bioeng. 12, 293-294
16. Takeda, Y., Ohlendorf, D. H., Anderson, W. F., and Matthews, B. W. (1983) Science 221, 1020-1029
17. Pabo, C. O., and Sauer, R. T. (1984) Annu. Rev. Biochem. 53, 293-321
18. Schleier, R. (1985) Science 221, 1182-1187
19. Anderson, J. E., Ptashne, M., and Harasien, S. C. (1987) Nature 326, 846-851
20. Aggarwal, A. K., Rodgers, D., Drottta, M., Ptashne, M., and Harrison, S. C. (1988) Science 242, 899-907
21. Jordan, S. K., and Pabo, C. O. (1988) Science 242, 883-889
22. Wolberger, C., Dong, J., and Harrison, S. C. (1988) Nature 335, 789-791
23. Otwinski, R. W., Schevitz, R. W., Zhang, R.-G., Lawmon, C. L., Joachimiak, A., Marmostein, R. Q., Luisi, B. F., and Sigler, P. B. (1988) Nature 335, 321-324
24. Schleier, R. W., Ortwinski, Z., Joachimiak, A., Lawmon, C. L., and Sigler, P. B. (1988) Nature 317, 782-786
25. Mondragon, A., Wolberger, C., and Harrison, S. C. (1986) J. Mol. Biol., in press
26. Mondragon, A., Subbiah, S., Almo, S. C., Drottor, M., and Harrison, S. C. (1988) J. Mol. Biol., in press
27. Zoidnerweg, E. R. P., Kajstein, R., and Wuthrich, K. (1983) Proc. Natl. Acad. Sci. U. S. A. 80, 5837-5841
28. Howard, P. K., Shaw, J., and Otsoke, A. J. (1986) Gene (Amst.) 35, 221-331
29. Buoncristiani, M. R., Howard, P. K., and Otsoke, A. J. (1986) Gene (Amst.) 44, 250-261
30. Doolittle, R. F. (1981) Science 214, 149-159
31. Bloomer, R. G., Weaver, L. H., and Matthews, B. W. (1986) Chem. Scien. 249, 251-256
32. Dodd, I. B., and Egan, J. B. (1987) J. Mol. Biol. 194, 557-564
33. White, S. W. (1987) Protein Eng. 1, 375-376
34. Hochschuld, A., Irwin, N., and Ptashne, M. (1983) Cell 32, 319-325
35. Matthews, B. W., Ohlendorf, D. H., Anderson, W. F., Fisher, R. G., and Takeda, Y. (1983) Cold Spring Harbor Symp. Quant. Biol. 47, 427-433
36. Tuskka, L., Apeltt, K., Dijk, J., White, S. W., and Wilson, K. S. (1984) Nature 310, 576-580
37. Richardson, J. S., and Richardson, D. C. (1988) Proteins: Structure, Function and Genetics, in press
38. Hochschuld, A., Dohmen, J. III, and Ptashne, M. (1986) Cell 47, 807-816
39. Wharton, R. P., and Ptashne, M. (1996) Trends Biochem. Sci. 11, 71-73
40. Wang, A. C., Revzin, A., Butler, A. P., and von Hippel, P. H. (1977) Nucleic Acids Res. 4, 1579-1596
41. Revzin, A., and von Hippel, P. H. (1977) Biochemistry 16, 4769-4776
42. Beales, R., Scheid, R. M., van Boom, J. H., and Kajstein, R. (1987) J. Biol. 193, 213-225
43. Lehning, N., Sartorius, J., Niemöller, M., Goenger, G., Wicken-Bergmann, B. v., and Muller-Hill, B. (1987) EMBO J. 6, 3145-3153
44. Lehning, N., Sartorius, J., Oehler, S., Wicken-Bergmann, B. v., and Muller-Hill, B. (1988) Proc. Natl. Acad. Sci. U. S. A. 85, 7947-7951

Acknowledgments—We thank Dr. Peter von Hippel for helpful discussions on sequence-specific and sequence-nonspecific DNA-protein interaction. We are also grateful to Drs. C. O. Pabo, S. C. Harrison, P. B. Sigler, and T. A. Steitz for providing coordinates, and Drs. J. S. Richardson and D. C. Richardson for a copy of their manuscript in advance of publication.