Recognition of DNA by Cys_2,His_2 Zinc Fingers

RACHEL E. KLEVIT

THE "ZINC FINGER" PROTEIN MOTIF WAS SO NAMED BE-cause of the tandemly repeating pattern observed in the amino acid sequence of the transcription factor TFIIIA (*1, 2*). According to the original hypothesis, each 30-residue sequence of the type shown in Fig. 1B is an independently folded unit that binds a zinc ion and is responsible for sequence-specific DNA binding. Two-dimensional nuclear magnetic resonance analysis of peptides with sequences that conform to the consensus sequence have confirmed that the Cys_2,His_2 zinc finger motif is indeed a structural domain (*3–6*).

Hundreds of genes encoding Cys_2,His_2 zinc finger domains have been isolated, identifying them as a major class of eukaryotic DNA-binding proteins. The way in which zinc fingers recognize specific DNA sequences in various systems is therefore of great interest. The structure of the complex formed between a three-finger fragment of the protein Zif268 and its consensus DNA binding site has been described by Pavletich and Pabo, providing an atomic level view of DNA recognition by Cys_2,His_2 zinc fingers (*7*).

The sequence-specific contacts observed in the Zif268-DNA complex have characteristic simplicity and modularity. Two amino acid side chains in each finger interact with two guanines in a three-base subsite, yielding the patterns shown in Fig. 1A. In the structure, the interacting residues X, Y, Z are each separated by two residues (Fig. 1B), with X immediately preceding an α helix and Y and Z lying on the outer face of the helix (*7*).

Virtually the same pattern of contacts was inferred from change-of-specificity mutagenesis experiments on the Cys_2,His_2 protein Krox-20 (*8*). Whether those two examples are representative of Cys_2,His_2 zinc finger–DNA recognition remains to be determined. Both Zif268 and Krox-20 make contacts only with guanines of the DNA and use only arginine and histidine side chains; when any of positions X, Y, or Z is not an arginine or histidine, a contact is not observed. There is a one-to-one correspondence between the three base pairs in a finger binding site and the three DNA-binding positions in a finger, in an antiparallel orientation (provided that the DNA sequence is read 5′ to 3′, the direction of the finger binding is COOH to NH_2).

There are seven potential arrangements of guanines in a triplet if one or three contacts per finger are allowed as well as the two contacts per finger used by Zif268. By combining different fingers, many guanine-containing DNA sequences can be recognized. If bases other than guanines were used as specific H-bond contacts (*9*), the range of possible sequences increases.

With only one Cys_2,His_2 zinc finger–DNA complex structure available, it is not possible to ascertain directly the general applicability of this scheme. An indirect test is whether it will predict the binding sites for other zinc finger proteins. The sequences of the three fingers of the switch protein SWI5 reduced to their X, Y, and Z positions along with their predicted DNA triplets are:

Positions	XYZ XYZ XYZ
Amino acids	RNS RDR RAV
Bind	GAn GnG Gnn--5′

The binding site, as identified by deoxyribonuclease I (DNase I) footprint analysis is 3′-G-A-G-G-T-C-G-T-A-A-T-A-T-C-G-T-A-C-G-A-C-C-5′ which contains the predicted nine-base sequence (underlined, with predicted contacts in boldface), although the predicted Z-contact in finger 2 is missing in the natural site (*10*). A similar treatment of the yeast transcription factor ADR1 yields:

Positions	XYZ XYZ
Amino acids	RHR RLR
Bind	GGG GnG--5′

The natural site for ADR1 is 3′-A-G-A-G-G-T-T-G-A-A-T (*11*). The predicted six-base sequence is not found within the 11-base binding site, but two sequences, G-A-G-G-T-T and A-G-G-T-T-G, both give two of the three predicted contacts for finger 1 and one of two predicted contacts for finger 2. This analysis suggests that the scheme may be useful in predicting Cys_2,His_2 zinc finger DNA binding sites that can be tested experimentally.

A challenging test for the scheme is the nine-finger Cys_2,His_2 protein TFIIIA. A 27-base site would be predicted, whereas methylation protection experiments showed protection over 50 bases (*11*). If the scheme is applied to TFIIIA and compared to its binding site, the predicted sequences can be found in clusters of two and three finger units:

Finger	1	2	3	4	5	6	7	8	9
Positions	XYZ	XYZ	XYZ	XYZ	XYZ	XYZ	XYZ	XYZ	XYZ
Amino acids	KKA	SHR	TNK	KQV	LRR	KTK	HYD	TNS	MSR
Bind	GGn	nGG	nAG	GAn	nGG	GnG	Gnn	nAn	nnG--5′

	90		80		70		60		50
3′-AGAG**GGTAGG**TTCAT**GATTGG**TCC**GGG**CTGG**GACGAACCG**AAGGCTCTAG									
Predicted	GGnnGG		nAG	GAnnGG		GnG		GnnnAnnnG	
Finger	1	2	3	4	5	6		7	8 9
His spacing	X_3	X_3	X_4	X_3	X_3	X_4		X_3	X_4 X_3
Linker length	(7)	(7)	(8)	(7)	(4)	(5)		(8)	(7)

This analysis leads to the hypothesis that TFIIIA binds to a long sequence by means of subsets of contiguous fingers that bind in the manner described for Zif268, with single fingers (fingers 3 and 6) spanning the gaps between the subsites. Several experiments support this model. Methylation protection showed that the guanines be-

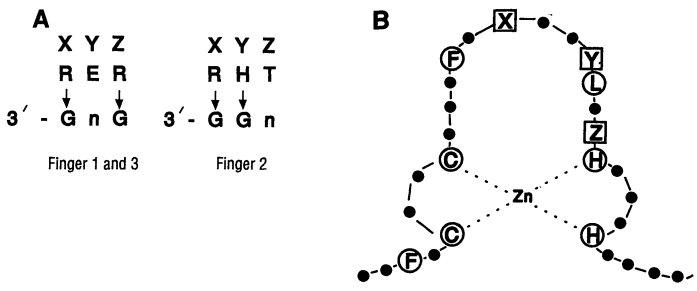

Fig. 1. (**A**) Pattern of contacts between side chain and bases in the Zif268-DNA complex. (**B**) The Cys_2,His_2 zinc finger motif. Circled residues are conserved; DNA-contacting residues, X, Y, and Z, are boxed.

Department of Biochemistry SJ-70, University of Washington, Seattle, WA 98195.

(*continued on page 1393*)

15. [5-55; 14-38], designated N* because of its native-like structure as determined by x-ray crystallography of a recombinant model [C. Eigenbrot, M. Randal, A. A. Kossiakoff, *Protein Eng.* **3**, 591 (1990)], does not oxidize readily to N because the remaining thiols, on cysteines 30 and 51, are buried and thus inaccessible to oxidizing agent (*2, 11*). N* coeluted with native BPTI by IEC (see Fig. 3) and was recognized after NMR studies of refolded BPTI detected a contaminating species [D. J. States *et al.*, *Nature* **286**, 630 (1980)].
16. T. E. Creighton, *J. Mol. Biol.* **87**, 579, 603 (1974).
17. _____, *Methods Enzymol.* **107**, 305 (1984).
18. J. S. Weissman and P. S. Kim, unpublished data.
19. T. E. Creighton, *Biophys. Chem.* **31**, 155 (1988).
20. T.-Y. Lin and P. S. Kim, *Biochemistry* **28**, 5282 (1989).
21. T. E. Creighton, *J. Mol. Biol.* **96**, 777 (1975).
22. Y. Goto and K. Hamaguchi, *ibid.* **146**, 321 (1981).
23. G. H. Snyder *et al.*, *Biochemistry* **20**, 6509 (1981).
24. D. M. Rothwarf and H. A. Scheraga, *J. Am. Chem. Soc.* **113**, 6293 (1991).
25. J. S. Weissman and P. S. Kim, unpublished data. For an example, see the zero time point of Fig. 6A.
26. The rate-limiting step is taken to be the step with the greatest free energy difference between its transition state and the lowest free energy intermediate preceding it [S. Yagisawa, *Biochem. J.* **263**, 985 (1989)].
27. T. E. Creighton, *Biochem. J.* **270**, 1 (1990).
28. J. S. Weissman, Z.-Y. Peng, P. S. Kim, unpublished results. Assignments for native BPTI are given by G. Wagner and K. Wüthrich, *J. Mol. Biol.* **155**, 347 (1982).
29. T. E. Creighton, *J. Mol. Biol.* **113**, 313 (1977).
30. L. F. McCoy, E. S. Rowe, K.-P. Wong, *Biochemistry* **19**, 4738 (1980).
31. T. Kiefhaber, H.-P. Grunert, U. Hahn, F. X. Schmid, *Proteins*, in press.
32. D. P. Goldenberg and T. E. Creighton, *Biopolymers* **24**, 167 (1985).
33. T. G. Oas and P. S. Kim, *Nature* **336**, 42 (1988).
34. D. J. States *et al*, *J. Mol. Biol.* **195**, 731 (1987).
35. For reviews, see: J. E. Rothman, *Cell* **59**, 591 (1989); F. X. Schmid, *Curr. Opinion Struct. Biol.* **1**, 36 (1991).
36. P. S. Kim and R. L. Baldwin, *Annu. Rev. Biochem.* **59**, 631 (1990).
37. T. E. Creighton and P. S. Kim, *Curr. Opinion Struct. Biol.* **1**, 3 (1991).
38. F. X. Schmid and R. L. Baldwin, *J. Mol. Biol.* **135**, 199 (1979); J. B. Udgaonkar and R. L. Baldwin, *Nature* **335**, 694 (1988); H. Roder *et al.*, *ibid.*, p. 700; M. Bycroft *et al.*, *ibid.* **346**, 488 (1990).
39. J. Baum, C. M. Dobson, P. A. Evans, C. Hanley, *Biochemistry* **28**; 7 (1989); F. M. Hughson, P. E. Wright, R. L. Baldwin, *Science* **249**, 1544 (1990); M.-F. Jeng, S. W. Englander, G. A. Elöve, A. J. Wand, H. Roder, *Biochemistry* **29**, 10433 (1990).
40. C. M. Dobson, *Curr. Opinion Struct. Biol.* **1**, 22 (1991).
41. For example, in the subdomain model (*11, 33, 36*) protein folding is viewed as a hierarchical condensation process [G. D. Rose, *J. Mol. Biol.* **134**, 447 (1979)] involving cooperatively folded subdomains of native structure, and the predominant source of cooperativity is native tertiary interactions.
42. A BPTI folding mixture (12 ml) was split into three portions of 4 ml each (*10*). At 5 minutes, folding was quenched by the addition of 1 ml of 0.125, 0.5, or 2 M iodoacetate stock solution (*43*), or by the addition of 50, 200, or 800 μl of formic acid (*44*).
43. Iodoacetate quenching is achieved by the addition of one-fourth volume of a stock (5×) solution of sodium iodoacetate in 250 mM tris-HCl, pH 6.8. After a period inversely proportional to the final concentration of iodoacetate used (1 minute for 400 mM) the protein is desalted on a PD-10 column equilibrated in 10 mM HCl. For the HPLC separations, buffer A is 0.1 percent trifluoroacetic acid (TFA) in water, and buffer B is 0.1 percent TFA, 90 percent acetonitrile, and 10 percent water. The gradient used is 0 minutes, 90 percent A; 15 minutes, 75 percent A; 75 minutes, 72 percent A; 135 minutes, 70 percent A; 155 minutes, 69 percent A.
44. Acid quenching is achieved by the addition of 1/20th volume of 88 percent formic acid, to give a final pH of ~2. When a delay between acid quenching and HPLC analysis was necessary because multiple time points were taken, the experiment was repeated with the order of chromatography altered to demonstrate that rearrangement was not occurring while the mix awaited separation. The gradient used for the HPLC separations is 0 minutes, 90 percent A; 15 minutes, 75 percent A; 35 minutes, 73 percent A; 50 minutes, 72 percent A; 140 minutes, 69 percent A.
45. Acid-quenched intermediates were purified by HPLC and lyophilized. Rearrangement was initiated by the addition of degassed folding buffer. The experiment was carried out in a water bath at 25°C in an anaerobic chamber (Coy Laboratory Products, Ann Arbor, MI). The single disulfides were allowed to rearrange for 2.5 minutes at pH 8.7 or 45 minutes at pH 7.3.
46. The [30-51; 14-38] intermediate was allowed to rearrange in folding buffer containing urea. The appearance of N* and N^{SH}_{SH} was monitored as a function of time and the rate of the sum of these intermediates was fit to an exponential in order to determine the unimolecular rate constant for rearrangement. In the presence of urea, a significant population of intermediates appears transiently, and the rate of formation of N* and N^{SH}_{SH} deviates from a simple exponential. Nevertheless, the data demonstrate that the rate of formation of N* and N^{SH}_{SH} increases substantially upon addition of denaturant.
47. Samples were prepared for NMR (Bruker 500 MHz spectrometer) by adding lyophilized protein to Milli-Q water 10 percent D_2O containing 20 μM EDTA and the pH was adjusted with NaOH. HPLC analysis indicated that less than 10 percent of the [30-51; 14-38] had rearranged or oxidized during the data collection.
48. J. S. Weissman and P. S. Kim, in preparation. The rate of oxidation with cyclic oxidizing agents like DTT is proportional to the rate of the intramolecular process that brings the cysteine thiols together. This makes it possible to compare steps in folding that involve intermolecular disulfide bond formation to those that involve only intramolecular thiol-disulfide exchange (*17*).
49. We thank J. P. Staley for helpful discussions; R. L. Baldwin and D. P. Goldenberg for helpful comments on earlier drafts of this manuscript; and Z.-Y. Peng for assistance with NMR measurements. Supported by the Karl Taylor Compton and NSF predoctoral fellowship programs (J.S.W.) and by grants from the NIH (GM 41307 and RR 05927), the Pew Memorial Trust, and the Rita Allen Foundation.

24 April 1990; accepted 9 August 1990

(*continued from page 1367*)

tween bases 60 and 70 were not well protected by TFIIIA, a result consistent with finger 6 behaving differently compared to the other fingers. Hydroxyl radical footprinting of finger deletion mutants of TFIIIA is consistent with the placement of fingers 1-2 and 8-9 as was shown (*12*).

The less regular mode of binding hypothesized for TFIIIA relative to the Zif268 structure may arise from irregularities in some of the TFIIIA zinc fingers. Fingers 3, 6, and 8 have His-X_4-His spacings between the histidine ligands instead of the more prevalent His-X_3-His spacing. The His-X_4-His fingers have more conformational flexibility than the His-X_3-His fingers (*13, 14*). Also, the linker length between fingers may influence the precise mode of binding (number of residues between the last His and first Cys in neighboring fingers). In the proposed model, the three subsites are spanned by His-X_4-His fingers with irregular linker lengths. These variations on the more regular His-X_3-His domain may serve to increase and diversify the binding sites available to multifinger proteins.

How general, then, is the recognition pattern of a single protein-DNA complex? Structures of the helix-turn-helix motif suggest that if there is a pattern, it will be more complex than the scheme outlined here. However, a scheme for making predictions that can be tested experimentally, both by structural analysis and by site-directed mutagenesis, should yield some insights into the nature of this class of zinc finger–DNA recognition.

REFERENCES AND NOTES

1. J. Miller, A. D. McLachlan, A. Klug, *EMBO J.* **4**, 1609 (1985).
2. R. S. Brown, C. Sander, P. Argos, *FEBS Lett.* **186**, 271 (1985).
3. G. Párraga *et al.*, *Science* **241**, 1489 (1988).
4. M. S. Lee, G. P. Gippert, K. V. Soman, D. A. Case, P. E. Wright, *ibid.* **245**, 635 (1989).
5. R. E. Klevit, J. R. Herriott, S. J. Horvath, *Proteins: Struct. Funct. Genet.* **7**, 215 (1990).
6. J. G. Omichinski, G. M. Clore, E. Appella, K. Sakaguchi, A. M. Gronenborn, *Biochemistry* **29**, 9324 (1990).
7. N. P. Pavletich and C. O. Pabo, *Science* **252**, 809 (1991).
8. J. Nardelli, T. J. Gibson, C. Vesque, P. Charnay, *Nature* **349**, 175 (1991).
9. N. C. Seeman, J. M. Rosenberg, A. Rich, *Proc. Natl. Acad. Sci. U.S.A.* **73**, 804 (1976).
10. K. Nagai, Y. Nakaseko, K. Nasmyth, D. Rhodes, *Nature* **332**, 284 (1988).
11. S. Thukral, A. Eisen, E. T. Young, *Mol. Cell. Biol.* **11**, 1566 (1991).
12. L. Fairall, D. Rhodes, A. J. Klug, *J. Mol. Biol.* **192**, 577 (1986).
13. R. X. Xu, S. J. Horvath, R. E. Klevit, *Biochemistry* **30**, 3365 (1991).
14. M. Kochoyan *et al.*, *ibid.*, p. 3371.