Proc. Natl. Acad. Sci. USA
Vol. 89, pp. 11109–11110, December 1992

Commentary

Sp1 and the subfamily of zinc finger proteins with guanine-rich binding sites

Jeremy M. Berg

Department of Biophysics and Biophysical Chemistry, The Johns Hopkins University School of Medicine, Baltimore, MD 21205

The Cys_2His_2 zinc finger proteins are a class of DNA binding proteins that contain sequences of the form (Tyr,Phe)-Xaa-Cys-$Xaa_{2,4}$-Cys-Xaa_3-Phe-Xaa_5-Leu-Xaa_2-His-Xaa_{3-5}-His, usually in tandem arrays (1–4). Each of these sequences binds a zinc(II) ion to form a structural domain termed a zinc finger. Structural studies by NMR (5–11) and x-ray crystallography (12) have revealed that these domains adopt very similar structures that consist of a β hairpin followed by a helix. Yet, because of variations of certain key amino acids from one zinc finger to the next, each domain makes its own unique contribution to DNA binding affinity and specificity.

The most well-understood members of this class of proteins are a subset that bind to relatively guanine-rich binding sites. This subset includes Sp1 (13), the Zif268/NGFI-A/Krox-20,24/Egr1,2/Wilm tumor family (14–19) and yeast ADR1 (20). Kriwacki *et al.* (21) report studies of the DNA binding domain of Sp1 that consists of three zinc finger domains. They demonstrated three major points. (*i*) They showed that a 92-amino acid peptide corresponding to the zinc finger region of Sp1 with very little flanking sequence specifically bound DNA containing sites that approximated the Sp1 binding site 5′-GGGGCGGGGC-3′. (*ii*) They demonstrated the ability of this peptide to distinguish between different DNA sequences in a manner reminiscent of that of intact Sp1. (*iii*) They showed that the 5′-GGGGCG region of the binding site contributed more to the overall binding affinity than does the GGGC-3′ region.

These observations can be interpreted in light of other reports. Several groups have demonstrated that fragments of Sp1 that included the zinc finger domains showed DNA binding properties very similar to those of the intact protein (22–26). The smaller fragment studied by Kriwacki *et al.* (21) is the smallest polypeptide studied to date beginning only 7 residues before the first Cys residue and extending only 3 amino acids past the last His residue.

The crystallographic studies of the three zinc finger domains of Zif268 bound to a cognate oligonucleotide have provided the most detailed structural information to date concerning zinc finger protein–DNA recognition (12). Three amino acids per domain play the most direct role in determining site preference. These are indicated as $\mathbf{Xaa^A}$, $\mathbf{Xaa^B}$, and $\mathbf{Xaa^C}$ in the sequence below:

(Tyr,Phe)-Xaa-Cys-$Xaa_{2,4}$-Cys-Xaa_3-

Phe-Xaa-$\mathbf{Xaa^A}$-Xaa-Xaa-$\mathbf{Xaa^B}$-

Leu-Xaa-$\mathbf{Xaa^C}$-His-Xaa_{3-5}-His.

The first and third fingers of Zif268 have $\mathbf{Xaa^A}$ = Arg, $\mathbf{Xaa^B}$ = Glu, $\mathbf{Xaa^C}$ = Arg. Each of these binds the triplet GCG with the two Arg residues forming a pair of hydrogen bonds with guanine and the Glu interacting indirectly with cytosine. Note that the Arg in $\mathbf{Xaa^A}$ interacts with the guanine at the 3′ end of the triplet. The second domain of Zif268 has $\mathbf{Xaa^A}$ = Arg, $\mathbf{Xaa^B}$ = His, $\mathbf{Xaa^C}$ = Thr. This finger interacts with the sequence 5′-TGG-3′ in the cocrystal structure via two hydrogen bonds from the Arg to the 3′ guanine and one hydrogen bond between the His and N7 of the central guanine.

The three zinc finger domains (shown, in order, fingers 1–3) of Sp1 have potential contact residues

$\mathbf{Xaa^A}$ = Lys, $\mathbf{Xaa^B}$ = His, $\mathbf{Xaa^C}$ = Ala

$\mathbf{Xaa^A}$ = Arg, $\mathbf{Xaa^B}$ = Glu, $\mathbf{Xaa^C}$ = Arg

$\mathbf{Xaa^A}$ = Arg, $\mathbf{Xaa^B}$ = His, $\mathbf{Xaa^C}$ = Lys.

Assuming that the corresponding contacts from the Zif268 structure apply to Sp1, the binding site can be predicted to be 5′-NGG-GCG-NGN-3′, where the 5′-NGG triplet is contacted by the third finger, the central GCG is contacted by the second finger, and the NGN triplet is contacted by the first finger. The His residues in position $\mathbf{Xaa^B}$ are likely to bind well to adenine as well as guanine since the observed hydrogen bond involved N7. This has been directly demonstrated for an Sp1 variant with $\mathbf{Xaa^B}$ = His in the second domain (27). Thus, the predicted site can be written 5′-N(G/A)G-GCG-N(G/A)N-3′. This clearly is a good approximation to the Sp1 consensus (23, 28). Moreover, it accounts for the many aspects of the variations in the known Sp1 binding sites as well as the decreased affinity of the Sp1 peptide for two mutant sites reported by Kriwacki *et al.* (21). Finally, this analysis provides a rationalization for the asymmetry of the contributions to overall binding affinity noted by Kriwacki *et al.* since the first domain (which contacts the 3′ end of the binding site) has only $\mathbf{Xaa^B}$ = His predicted to make a directly precedented contact. While it is likely that $\mathbf{Xaa^A}$ = Lys also hydrogen bonds to a base, $\mathbf{Xaa^C}$ = Ala is obviously incapable of such an interaction.

This observation raises one of the remaining puzzles about Sp1–DNA interactions. If $\mathbf{Xaa^C}$ = Ala does not interact directly with the DNA, why is the underlined guanine in 5′-GGG-GCG-GGG so well conserved in the known Sp1 binding sites? This position is a thymine in several of the known high-affinity sites but no direct studies have been reported that bear on the effects of mutations at this position on Sp1 binding affinity. It is possible that changes in this site do not affect affinity significantly or that guanine is preferred but is recognized by a more indirect mechanism.

These studies illustrate one approach to investigating specific protein–DNA interactions—namely, examination of the ability of a given protein to distinguish between different binding sites. An alternative approach is to investigate the effects of changes in the protein sequence on DNA binding affinity and specificity. Change-of-specificity mutants (in which DNA binding ability is not simply disrupted but changed to a different preferred DNA sequence) are particularly useful. The first such mutants were reported by Nardelli *et al.* (29) for Krox-20, a close homologue of Zif268, with three zinc finger domains and the same contact residues. The wild-type protein was shown to bind 5′-GCG-GGG-GCG-3′ but not 5′-GCG-GCG-GCG-3′. In the second domain, $\mathbf{Xaa^B}$ was changed from His to Glu and $\mathbf{Xaa^C}$ was changed from Thr to Arg. The double mutant showed the reverse specificity for the two DNA sequences above. Obviously, the two mutations put the same contact residues in the second domain that were already found in the first and third domains. Thus, binding to 5′-GCG-GCG-GCG-3′ is

not surprising but it is important to realize that these experiments were published prior to the report of the Zif268 cocrystal structure. Other mutants studied resulted in reduction in the selectivity of DNA binding.

A more extensive series of change-of-specificity mutants of Sp1 itself has been developed by Desjarlais and Berg (25–27). These involved changes in the second zinc finger domain. The first of these changed the binding specificity from 5′-GGG-GCG-GGG-3′ to 5′-GGG-GCT-GGG-3′ via mutation of Xaa^A from Arg to Gln and Xaa^B from Glu to Asp as well as one additional change in a residue that interacts with the Xaa^A side chain (25, 26). The fact that three amino acid sequence changes were required to change the preferred binding site in a single position illustrates the necessity to consider the entire recognition helix; only certain sets of contact residues are mutually compatible. An additional mutant was produced by changing Xaa^B from Asp to Asn in the above context. This mutant binds to 5′-GGG-GAT-GGG-3′. A single change of Xaa^B from Glu to His in the wild-type Sp1 background bound to 5′-GGG-G(A/G)G-GGG-3′. This is the inverse of one of the Krox-20 mutants discussed above. Finally, a mutant with Xaa^B changed from Glu to Leu bound 5′-GGG-G(A/C/T)G-GGG-3′.

A final member of this protein subfamily that has been extensively studied is yeast ADR1. This protein has two zinc fingers in its DNA binding domain and it generally binds as a dimer on sites with approximate dyad symmetry (30). The two putative contact residues (shown, in order, fingers 1 and 2) are

$$Xaa^A = Arg, Xaa^B = His, Xaa^C = Arg$$

$$Xaa^A = Arg, Xaa^B = Leu, Xaa^C = Arg.$$

One of the preferred binding sites is 5′-TTG-GAG-3′. These contacts are consistent with those above except that the Arg in position Xaa^C in the second finger abuts thymine rather than guanine. Three change-of-specificity mutants were produced via an extensive study of mutations in the first zinc finger (31). Changing Xaa^A from Arg to Gln produced a protein that preferred the site 5′-TTG-GAA-3′. It is interesting to note that the preference associated with the Xaa^A = Gln differs from that observed in the Sp1 mutant above. Glutamine is notable for its ability to interact with DNA bases in different ways; in the cocrystal structure of 434 repressor bound to DNA, three Gln residues make three different types of contact with DNA (32). A second mutant involved changing Xaa^B from His to Thr. This mutant protein preferred 5′-TTG-GCG-3′. Finally, a mutant involving a change of Xaa^C from Arg to Asn preferred 5′-TTG-AAG-3′. These results on ADR1 confirm and extend the results from the Sp1 mutants.

Because of the apparent modular nature of the zinc finger domains in proteins of this class, it has been very tempting to hope that codes could be deduced that interrelate the amino acid sequences of zinc finger domains and the sequences of their preferred binding sites. This would allow prediction of binding site sequences for naturally occurring zinc finger proteins and design of zinc finger proteins with preselected DNA binding properties. As the above discussion indicates, considerable progress toward this goal has been made for this small subset of zinc finger proteins that interact with guanine-rich binding sites. However, it is important to note that many zinc finger proteins have amino acids in the potential contact positions whose interactions with DNA have not been elucidated. Furthermore, some zinc finger proteins such as *Drosophila* hunchback have been shown to interact with very (A+T)-rich binding sites (33, 34). Future studies should reveal whether similar rules can be developed for these protein–DNA interactions.

1. Miller, J., McLachlan, A. D. & Klug, A. (1985) *EMBO J.* **4**, 1609–1614.
2. Klug, A. & Rhodes, D. (1987) *Trends Biochem. Sci.* **12**, 464–470.
3. Evans, R. M. & Hollenberg, S. M. (1988) *Cell* **52**, 1–3.
4. Berg, J. M. (1990) *Annu. Rev. Biophys. Biophys. Chem.* **19**, 405–421.
5. Lee, M. S., Gippert, G. P., Soman, K. V., Case, D. A. & Wright, P. E. (1989) *Science* **245**, 635–637.
6. Párraga, G., Horvath, S. J., Eisen, A., Taylor, W. E., Hood, L., Young, E. T. & Klevit, R. E. (1988) *Science* **241**, 1489–1492.
7. Neuhaus, D., Nakaseko, Y., Nagai, K. & Klug, A. (1990) *FEBS Lett.* **262**, 179–184.
8. Kochoyan, M., Keutmann, H. T. & Weiss, M. A. (1991) *Biochemistry* **30**, 9396–9402.
9. Omichinski, J. G., Clore, G. M., Appella, E., Sakaguchi, K. & Gronenborn, A. M. (1990) *Biochemistry* **29**, 9324–9334.
10. Omichinski, J. G., Clore, G. M., Robien, M., Sakaguchi, K., Appella, E. & Gronenborn, A. M. (1992) *Biochemistry* **31**, 3907–3917.
11. Krizek, B. A., Amann, B. T., Kilfoil, V. J., Merkle, D. L. & Berg, J. M. (1991) *J. Am. Chem. Soc.* **113**, 4518–4523.
12. Pavletich, N. P. & Pabo, C. O. (1991) *Science* **252**, 809–817.
13. Kadonaga, J. T., Carner, C., Masiarz, F. R. & Tjian, R. T. (1987) *Cell* **51**, 1079–1090.
14. Christy, B. A., Lau, L. F. & Nathans, D. (1988) *Proc. Natl. Acad. Sci. USA* **85**, 7857–7861.
15. Milbrandt, J. (1987) *Science* **238**, 797–799.
16. Chavrier, P., Vesque, C., Galliot, B., Vigneron, M., Dollé, P., Duboule, D. & Charnay, P. (1990) *EMBO J.* **9**, 1209–1218.
17. Joseph, L. J. (1988) *Proc. Natl. Acad. Sci. USA* **85**, 7164–7168.
18. Gessler, M., Poustka, A., Cavenee, W., Neve, R. L., Orkin, S. H. & Bruns, G. A. P. (1990) *Nature (London)* **343**, 774–778.
19. Haber, D. A., Buckler, A. J., Glaser, T., Call, K. M., Pelletier, J., Sohn, R. L., Douglass, E. C. & Housman, D. E. (1990) *Cell* **61**, 1257–1269.
20. Hartshorne, T. A., Blumberg, H. & Young, E. T. (1986) *Nature (London)* **320**, 281–287.
21. Kriwacki, R. W., Schultz, S. C., Steitz, T. A. & Caradonna, J. P. (1992) *Proc. Natl. Acad. Sci. USA* **89**, 9759–9763.
22. Courey, A. J. & Tjian, R. (1988) *Cell* **55**, 887–898.
23. Thiesen, H.-J. & Bach, C. (1990) *Nucleic Acids Res.* **18**, 3203–3209.
24. Kuwahara, J. & Coleman, J. E. (1990) *Biochemistry* **29**, 8627–8631.
25. Desjarlais, J. R. & Berg, J. M. (1992) *Proteins: Struct. Funct. Genet.* **12**, 101–104.
26. Desjarlais, J. R. & Berg, J. M. (1992) *Proteins: Struct. Funct. Genet.* **13**, 272.
27. Desjarlais, J. R. & Berg, J. M. (1992) *Proc. Natl. Acad. Sci. USA* **89**, 7345–7349.
28. Kadonaga, J. T., Jones, K. A. & Tjian, R. (1986) *Trends Biochem. Sci.* **11**, 20–23.
29. Nardelli, J., Gibson, T. J., Vasque, C. & Charnay, P. (1991) *Nature (London)* **349**, 175–178.
30. Eisen, A., Taylor, W. E., Blumberg, H. & Young, E. T. (1988) *Mol. Cell. Biol.* **8**, 4552–4556.
31. Thukrel, S. K., Morrison, M. L. & Young, E. T. (1992) *Mol. Cell. Biol.* **12**, 2784–2792.
32. Aggarwal, A. K., Rodgers, D. W., Drottar, M., Ptashne, M. & Harrison, S. C. (1988) *Science* **242**, 899–907.
33. Stanojevic, D., Hoey, T. & Levine, M. (1989) *Nature (London)* **341**, 331–335.
34. Treisman, J. & Desplan, C. (1989) *Nature (London)* **341**, 335–337.