A happy diversity of interactions

The 1990s have seen a significant expansion of the lexicon of structural motifs in proteins known to interact with DNA, as well as an ever-increasing wealth of detail about previously characterized DNA-binding domains. And the first quarter of 1994 has not been a disappointment; papers in this[1] and other journals[2-9] emphasize further the heterogeneity of the protein-DNA interface as well as providing a glimpse of a new non-specific DNA-binding domain.

DNA must of necessity undergo a series of modifications that are central to biological form and function; replication, recombination, repair, restriction-modification (in bacteria), imprinting (in eukaryotes), transcription and packaging, for example. All these are effected by proteins that cut, clamp, kink, bind, bend, unwind, zip up, knot and unknot double-helical or, as it may be, single-stranded DNA.

Topological transformation of the double helix, a necessary requirement for many functions of DNA, is carried out by a number of different classes of topoisomerase. The recently reported crystal structure of an amino-terminal fragment of *Esherichia coli* topoisomerase I (topo I) has allowed Lima and colleagues[2] to present a model of how the enzyme may juggle DNA strands.

The protein has a roughly toroidal shape and, in that conformation, the active site is buried and inaccessible. It has been suggested that the toroid opens like a padlock, exposing the catalytic tyrosine and thus facilitating single-strand DNA (ssDNA) cleavage. The cut ends of the ssDNA associate with the two now-separated ends of the open torus, allowing passage of another DNA molecule, either single- or double-stranded, through the gap and into or out of the hole in the protein. Subsequent closure of the torus brings the cut ends

of the DNA back together. Sealing of the nick completes the topological transformation.

Interaction of topo I with DNA must be non-specific and also weak, so as to avoid formation of a stable protein-DNA complex and consequent stalling of the reaction. The average diameter of the hole in the toroid is 27.5Å, wide enough to accommodate dsDNA (~20Å in diameter), yet there are no obvious motifs that might recognize the major or minor groove. A dimer of the B protein of DNA gyrase, a type II topoisomerase, also forms a hole, 20Å in diameter but of different architecture. That too may interact with DNA during catalysis[3].

As might be expected, the electrostatic lining of the topo I and gyrase B holes is for the most part positive and similar to that of the inner surface of the processive element of DNA polymerase III holoenzyme[4], which also binds DNA non-specifically. The structure of the latter seems well suited to its role of clamping onto and racing along DNA at 750bp sec^{-1}: the diameter of the hole is 35Å, significantly larger than that of dsDNA and the 12 α-helices lining the inside of the clamp are oriented at right-angles to the direction of the grooves in the DNA, no doubt to minimize specific interaction.

Although they are not specific for particular DNA sequences, it is likely that the topoisomerases do sense local structural distortions, such as those associated with supercoiling for example, and it may therefore be possible to use such features, as well as details of the catalytic mechanism, to generate informative protein-DNA complexes. As pointed out by Kong and colleagues, the more profound lack of specificity of the processive clamp suggest that molecular dynamics simulations of solvated models of the protein with DNA may provide the most useful approach for understanding details of that interaction.

editorial

Part of the curve of the torus of topo I is formed by a predominantly β-sheet domain that shows a passing resemblance to the DNA-binding β-sheet saddle of the TATA box protein (TBP). Presumably the similarity of curvature of the sheet allows this domain partially to wrap around the DNA.

Another example of β-sheet interacting with DNA is provided by the structure[5] and mutagenesis[1] of the Arc repressor-DNA complex (discussed in the 6 March issue of *Nature*). The *Salmonella* bacteriophage Arc repressor is a member of the ribbon–helix–helix family of DNA-binding proteins, which includes the MetJ repressor. The 'recognition' domain of the dimer is a two-stranded anti-parallel β-sheet as opposed to the more familiar α-helix. Unlike the saddle of TBP, which interacts with a tremendously distorted minor groove, the β-strands of the Arc repressor sit very comfortably in the major groove of the DNA, in an analogous manner to the second α-helix of the now familiar helix–turn–helix (HTH) motif. Hin recombinase, also from *Salmonella*, recognizes its DNA recombination site using just such an HTH motif[6]. Important base-specific hydrogen bonds are mediated through water molecules wedged in the major groove between protein and DNA, and the recombinase also makes important minor groove contacts with its amino- and carboxy-terminal arms. Although superficially similar to the HTH of the homeobox, in some respects the structure is more closed similar to the HTH of the prokarotic regulators.

The structure of a second basic/helix–loop–helix/leucine zipper (b/HLH/Z) domain from the Hela cell transcription factor upstream stimulatory factor (USF) bound to DNA[7] shows many features in common with the Max-DNA complex; both form parallel, left-handed, four-helix bundles, for example. Aditionally, dimers of the USF b/HLH/Z domain can form homotetramers that simultaneously bind two distinct DNA sites — a result that has obvious implications for DNA looping and transcription.

While there has been some speculation that the zinc finger family of transcription factors may indeed have a relatively simple 'code' for DNA recognition, recent structures of zinc finger-DNA cocrystals have demonstrated considerable variability in the zinc finger-DNA interaction. Thus, any kind of simple code for recognition of DNA by zinc finger motifs as a whole is unlikely. Nonetheless, comparison of closely related DNA-binding motifs may provide a useful set of rules for designer DNA-binding domains. Two recent papers[8,9] describe the use of a phage display system to select for alterations in the specificity of the first zinc finger of Zif268 (ref 8). This kind of approach should provide both zinc fingers with novel DNA-binding specificities and insight into the parameters of the protein-DNA interaction.

The recent feast of structures continues to surprise, delight and perhaps frustrate those interested in the problem of understanding the basis of protein-DNA recognition. Knowledge of three-dimensional structure is a tangible and, more often than not, satisfying milestone in understanding; but clearly, it cannot provide all the answers. It is encouraging to hope that a recent study of the thermodynamics of sequence-specific DNA-protein binding[10] heralds a change in emphasis in the field and the beginning of the dissection of the forces that drive sequence-specific and non-specific interactions.

1. Brown, B.M., Milla, M.E., Smith, T.L. & Sauer, R.T. *Nature struct. Biol.* **1**, 164–169 (1994)
2. Lima, C.D., Wang,J.C. & Mondragon, A. *Nature* **367**, 138–146 (1994)
3. Wigley, D.B., Gideon, J.D., Dobson, E.J., Maxwell, A. & Dodson, G. *Nature* **351**, 624–629 (1991)
4. Kong, X-P., Onrust, R., O'Donnell, M. & Kuriyan, J. *Cell* **69**, 425–437 (1992)
5. Raumann, B.E., Rould, M.A., Pabo, C.O. & Sauer, R.T. *Nature* in the press
6. Feng, J-A, Johnson, R.C. & Dickerson, R.E. *Science* **263**, 348–355 (1994)
7. Ferré-D'Amaré, A.R., Pognonec, P., Roeder, R.G. & Burley, S.K. *EMBO J.* **13**, 180–189 (1994)
8. Rebar, E.J. & Pabo, C.O. *Science* **263**, 671–673 (1994)
9. Jamieson, A.C., Kim, S-H., & Wells, J.A. *Biochemistry* in the press
10. Spolar, R.S. & Record Jr., M.T. *Science* **263**, 777–784 (1994)