No limits on restriction

The structure of the endonuclease R.PvuII in the absence and presence of its recognition site indicates how the interaction with DNA might proceed.

THE bacterial restriction-modification (R-M) proteins did not evolve just for the convenience of molecular biologists and the profit of biotechnology companies, but to promote the survival of bacteria by helping to shape their primitive 'immune' systems. The sheer diversity of the sequence-specific type-II endonucleases that form one arm of the R-M system is also fascinating to students of biomolecular architecture, providing as they surely will a wealth of information on the nature of the interface between protein and DNA. Now two new structures are reported, one in this month's *Nature Structural Biology*¹ and the other to appear shortly in the *EMBO Journal*², of the PvuII endonuclease (R.PvuII) of the bacterium *Proteus vulgaris*. In the first¹ no DNA is present, but in the second the nuclease is complexed with its cognate DNA sequence (5'-CAGCTG-3')². The structures suggest a scenario for the initial steps of the protein-DNA interaction, as well as providing some clues about the general mechanism of action of the type-II restriction endonucleases.

Type-II bacterial R-M systems consist of a pair of proteins, both of which bind specifically to the same short, 4–8-base-pair recognition sequence. One of the pair is an endonuclease which, when bound to DNA in the presence of Mg²⁺, cleaves both strands of the DNA, cutting the duplex clean in two. The other is a DNA methyltransferase that adds a methyl group onto one of the bases in the recognition sequence. Methylation prevents binding of the recognition sequence by the endonuclease and hence the cleavage reaction. In this way the host bacterium's DNA is protected from destruction while foreign, unmethylated DNA — from viruses and other species of bacteria — is open to attack.

Over 2,000 restriction endonucleases, clocking up around 200 different DNA specificities between them, have been

isolated to date. The lack of homology between the 60 or so known gene sequences hints at the structural diversity of this class of enzymes. R.EcoRI and R.EcoRV show little similarity in either sequence, structure or mechanism of binding to DNA^{3,4}. On the other hand, R.BamHI does share substantial homology with R.EcoRI (refs 5, 6).

The structure of the unbound form of R.PvuII reveals that two monomers are non-covalently linked through a dimerization subdomain to form a U-shaped dimer¹; like many other restriction enzymes, R.PvuII is only active as a homodimer. The monomers are joined at the base of the 'U' by two helix-loop-helix structures — one from each monomer — which together form a pseudo three-helix bundle: an extensive hydrophobic core and a number of salt bridges and hydrogen bonds lock the bundle and the dimer together.

The shape of the dimer immediately suggests a possible binding site for DNA, neatly cradled in the cleft of the 'U'. Comparison of the R.PvuII structure with that of R.EcoRV bound to DNA³ further strengthens this suspicion. R.PvuII shares some structural and topological features with R.EcoRV, a restriction enzyme that, like R.PvuII, slices through the DNA duplex, leaving blunt rather than 5' or 3' overhanging ends. Elements of secondary structure that correspond to the DNA-binding elements of R.EcoRV can readily be identified in the structure of R.PvuII. The region that is similar to the 'R-turn' in R.EcoRV, which is involved in making base-specific contacts with DNA in this latter enzyme, falls across two antiparallel β -strands which line the inside of the cleft in apo-R.PvuII. Indeed, the co-crystal structure² reveals that this two-stranded sheet interacts with the major groove of the DNA in a fashion reminiscent of that seen for the MetJ and Arc repressors⁷. Similarly, the loop between the first and second helices of the dimerization subdomain is shown to be the equivalent of the R.EcoRV DNA-binding 'Q-turn', and again is situated strategically at the base of the cleft. This loop is seen to make backbone and base-specific contacts in the R.PvuII-DNA complex².

The residues and structures involved in catalysis in R.EcoRV have their counterparts in R.PvuII. But, unlike the DNA-binding domain, here the comparison can be extended to the catalytic residues of

R.EcoRI and R.BamHI (refs 5, 6) as well, for although these enzymes might interact with DNA in different ways, they all catalyse the cleavage of an identical DNA backbone to yield 5'-phosphate and 3'-hydroxyl groups. Whether this similarity is the result of convergent or divergent evolution is difficult to say — indeed the evolutionary relationship of the group as a whole is still a matter for debate⁵.

Although there are notable similarities between R.PvuII and R.EcoRV, there are also some important differences between the two structures. The cleft of the 'U' in the unbound form of R.PvuII is much wider than it is in the structure of both free and bound R.EcoRV and so could not contact DNA as extensively with both subunits as in R.EcoRV. Athanasiadis and colleagues¹ suggest that the initial stages of the protein-DNA interaction may occur through the equivalent of the R.EcoRV Q-turn. In the unbound form, this loop in the two monomers would be the only DNA-binding motif capable of making symmetry-related contacts with the DNA recognition site. The potential flexibility of the loops connecting the various subdomains would then allow the enzyme to grasp the DNA more tightly. Indeed, the co-crystal shows a dramatic conformation change in R.PvuII: the mouth of the cleft closes up, wrapping around the DNA and bringing the tips of the two arms of the 'U' into contact, a movement of some 25 Å. The DNA, on the other hand, does not deviate significantly from B-form, indicating that in this case distortion of the double helix is not required for recognition².

We can look forward to the next clutch of structures furnished by the restriction endonucleases, which no doubt will continue to provide a pertinent illustration of the variability and unpredictability of the protein-DNA interface.

Guy Riddihough

Guy Riddihough is Editor of Nature Structural Biology.

Also in this month's *Nature Structural Biology*: a copper-iron-sulphur cluster in a ferredoxin; apomyoglobin as a molten globule; haem binding in a bacterioferretin; the unique structure of *Serratia* endonuclease; GroEL fails to recognize a molten globule; the general base in ras p21; and entropic effects of disulphide bonds on protein stability.

1. Athanasiadis, A. *et al. Nature struct. Biol.* **1**, 469–475 (1994).
2. Cheng, X., Balendiran, K., Schildkraut, I. & Anderson, J.E. *EMBO J.* (in the press).
3. Kim, Y., Grable, J.C., Love, R., Greene, P.J. & Rosenberg, J.M. *Science* **249**, 1307–1309 (1990).
4. Winkler, F.K. *et al. EMBO J.* **12**, 1781 (1993).
5. Newman, M., Strzelecka, T., Dörner, L.F., Schildkraut, I. & Aggarwal, A.K. *Nature* **368**, 660–664 (1994).
6. Newman, M., Strzelecka, T., Dörner, L.F., Schildkraut, I. & Aggarwal, A.K. *Structure* **2**, 439–452 (1994).
7. Riddihough, G. *Nature* **368**, 82 (1994).