



# DNA mimicry by proteins and the control of enzymatic activity on DNA

David T.F. Dryden

School of Chemistry, The University of Edinburgh, The King's Buildings, Edinburgh EH9 3JJ, UK

**Cells are unable to perform any function on their DNA in the absence of proteins, and it is of vital importance that these proteins only perform their function at appropriate times during the cell cycle. Thus, DNA-binding proteins are always controlled by a wide range of other factors, primarily other proteins. These controlling factors usually block access of the protein to the DNA, often operating by simple competitive inhibition. However, it has recently been demonstrated that DNA-binding proteins can be controlled by the direct binding of the control protein to the DNA-binding site on the first protein. The structures of these control proteins have revealed that they mimic the structure and electrostatics of DNA. This review highlights the roles of DNA mimics in the control of DNA-binding proteins, suggests other possible candidate proteins using DNA mimicry, and puts forward a range of potential uses of DNA mimics.**

## Introduction

DNA requires interaction with sequence-specific enzymes for any useful function. Such enzymes recognize specific target sequences, specific structures or specific chemical modifications in the DNA molecule and catalyze reactions such as DNA replication, transcription, recombination, repair or, in prokaryotes, restriction. All of these interactions between the enzymes and DNA have to be carefully controlled, and for many years the paradigm for control has been competition between the enzyme and a control protein for the same specific target on the DNA molecule. Textbook examples of this are the control of the *lac* operon [1], the lysis or lysogeny switch of phage lambda [2] and the control of gene expression by DNA methylation in eukaryotes [3]. Binding of the enzyme or the control protein precludes DNA binding by the competitor in a direct and obvious manner, as shown in Figure 1.

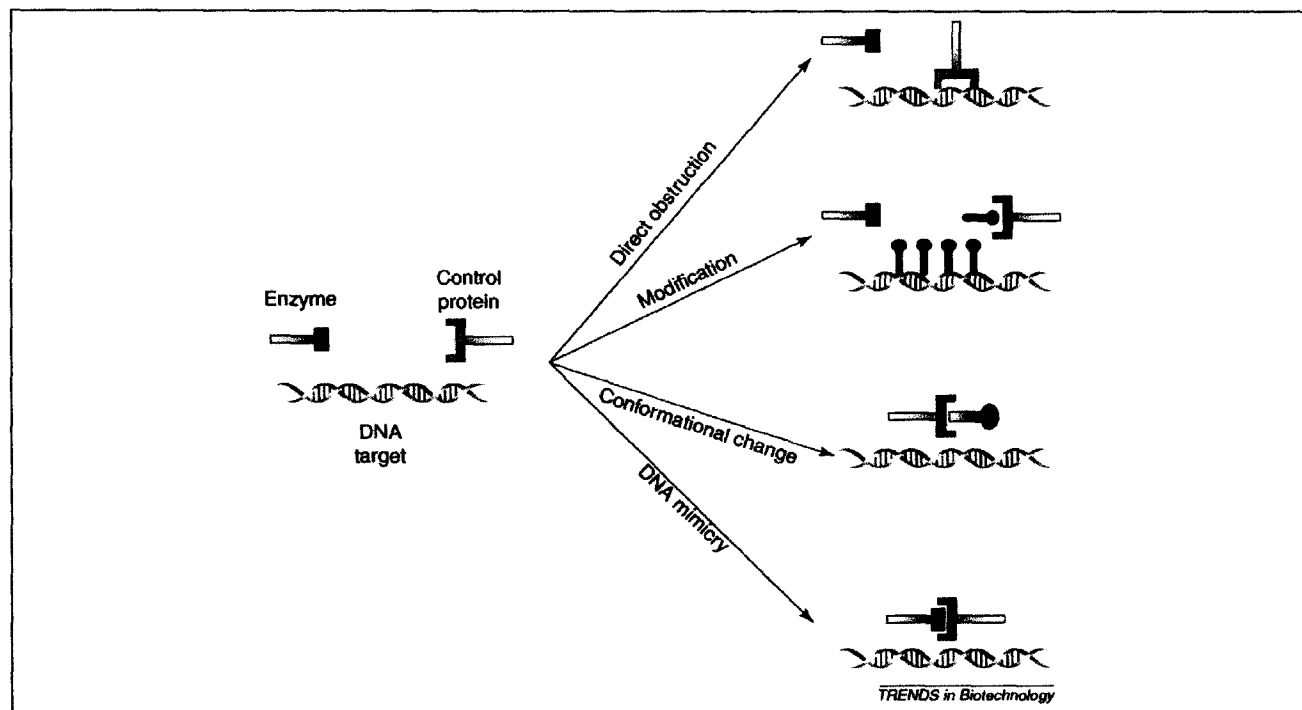
However, another mechanism does exist, namely an interaction between the enzyme and the control protein that results in the blocking of the interaction between the enzyme and DNA (Figure 1). This prevention of DNA binding can occur in either an indirect or a direct fashion. An indirect conformational change in the DNA-binding site can be induced by the interaction of the control protein with the enzyme (e.g. anti-sigma factors sequester bacterial sigma factors and prevent assembly of a functional RNA polymerase [4,5]). This review, however, will only consider the simplest and most direct mechanism for preventing DNA binding, namely the direct binding of the control

protein to the DNA-binding site of the enzyme, resulting in the physical occlusion of the DNA-binding site. This would require a high binding affinity between the two proteins so that the protein-protein interaction can compete out binding of DNA by the enzyme. One way to ensure such a strong interaction is to make the control protein 'mimic' the natural substrate for the enzyme, namely the DNA molecule to which the enzyme would normally bind. This would imply mimicry by the control protein of the shape and electrostatics of the DNA target normally recognized by the enzyme.

Exactly this sort of molecular mimicry is now known to target many of the enzyme functions occurring upon DNA, including transcription, repair and restriction (Table 1). Of further interest is the recent observation that the DNA mimic HI1450, encoded by *Haemophilus influenzae*, binds to histone-like proteins, and homologous sequences are found in many proteobacteria [16,17]. Perhaps HI1450 and other DNA mimics have a role in controlling chromosomal structure in prokaryotes (and perhaps even in eukaryotes)? Two dramatic examples of proteins mimicking substantial lengths of double-stranded DNA are considered in Box 1.

Mimicry has long been exploited by the pharmaceutical industry, and with the discovery of protein mimics of DNA it seems that nature might be suggesting a new path to follow. Structural aspects of DNA mimicry by proteins have been the subject of several recent reviews [26–30], and the reader is directed to these for precise details of the structures of the mimics. However, common characteristics of these mimics are mimicry of the shape adopted by the DNA when bound to the target enzyme (e.g. the *ocr* protein of phage T7 mimics the bent DNA recognized by type I restriction enzymes [11,12] and dTAF<sub>II</sub>230 mimics the unwound and bent DNA of the TATA-box [19]) and correspondence in the location of carboxylates on the surface of the protein with the location of phosphates on the surface of the DNA molecule. The protein structures used by the DNA mimics cover all known secondary structure types – alpha helices, beta strands and peptide loops – thus potentially making bioinformatics approaches to detecting mimics rather difficult. In the absence of an atomic structure, suggestions of DNA mimicry can be derived from an abundance of negatively charged residues, relatively small size (<180 aa) and physical evidence of direct blockage of the DNA binding site on the target protein. It is highly probable that many examples of DNA mimicry lie unappreciated in the literature and these could be re-examined using structural methods. Putnam and Tainer [27] suggest

Corresponding author: Dryden, D.T.F. (david.dryden@ed.ac.uk).



**Figure 1.** Mechanisms for control of a site-specific DNA-binding enzyme. The enzyme needs to recognize and bind to a specific DNA target sequence to perform its function. This function is controlled by another protein. The control can be obtained by direct obstruction of the binding site by the control protein, by chemical modification of the binding site by the control protein, by an indirect conformational change in the enzyme induced by binding of the control protein to change the binding site conformation on the enzyme or, lastly, by direct blocking of the DNA binding site on the enzyme by a DNA mimic.

that the Arn protein of phage T4 and the gp2 and gp5.9 proteins of phage T7 might be DNA mimics. The author would also like to suggest that a range of phage antirepressor proteins from phages P1, P4, P22, N15, 186 and WΦ [31–40] might also function by mimicking DNA structure. All these systems are worthy of further study.

### Mimics for technology

#### Current uses and challenges with known mimics

Nature has clearly evolved a range of proteins capable of resembling different DNA structures and of controlling different aspects of protein–DNA interactions (Table 1). During the analysis of these proteins, several technical issues and areas for further research have become apparent.

First, several of the mimic structures have been solved in complex with their target protein [6–9,19]. This

obviously simplifies their comparison with structures of the DNA bound to the target protein but it also suggests that mimics could be used instead of DNA for co-crystallization experiments. For example, two structures of the sequence specificity subunits of type I restriction enzymes have recently been solved [41,42]; however, the DNA target sequences recognized by these two specificity subunits are unknown and only simple models of the protein–DNA interface could be constructed. It is possible that the addition of the ocr protein would enable a co-crystal structure to be obtained to better define the putative DNA-binding interface on the specificity subunits.

Second, Hoffman and colleagues, of Epicentre Biotechnologies (Type I inhibitor improves transformation efficiencies by blocking type I restriction and modification systems *in vivo*. 2002 *EPICENTRE Forum* 9, 8;

**Table 1.** Proteins most likely to function by DNA mimicry and their biological role.

DNA mimic or putative mimic	Targeted enzyme or protein	Targeted DNA function	References and notes
Uracil glycosylase inhibitor	Uracil glycosylase	Repair	[6–9]
Ocr	Type I restriction enzymes	Restriction	[10–12]
ArdA, ArdB	Type I restriction enzymes and some type II restriction enzymes		[13–15]
Hi1450	Histone-like proteins	DNA packaging	[16,17]
MfpA	DNA gyrase	Topology	[18]
dTAF <sub>II</sub> 230	RNA polymerase	Transcription	[19]
Ocr	RNA polymerase		[20]
DinI	RecA	Recombination	[21–24] DinI was suggested to be a mimic of single-stranded DNA but evidence is contradictory.
p53	Replication protein A	Single-strand binding	[25] Mimicry by a fragment of p53 blocks the single-stranded DNA binding activity of RPA.

### Box 1. Examples of DNA mimicry.

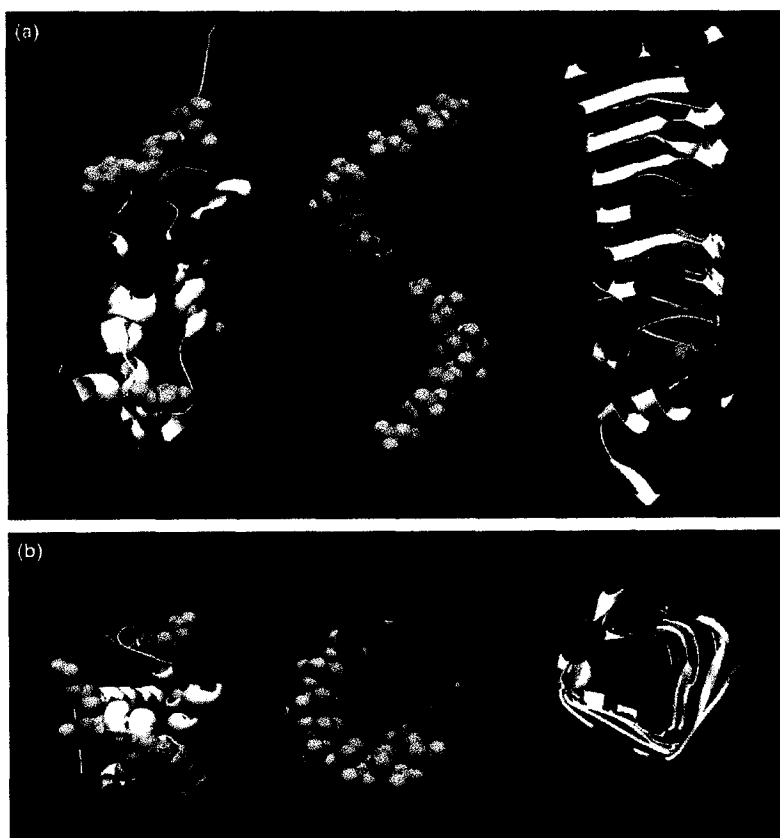
#### Ocr: an inhibitor of type I DNA restriction and modification (R/M) enzymes.

The dimeric ocr protein, encoded by bacteriophage T7, is produced in huge amounts immediately after infection of *E.coli* by the phage. Ocr overwhelms the host type I restriction enzyme [46,47]. The type I restriction enzymes are large, multifunctional enzymes and were the first restriction enzymes to be detected and purified [48,49]. They are apparently present in at least 50% of prokaryotes and display numerous different sequence specificities [50]. Where it has been tested, ocr knocks out all type I restriction enzymes irrespective of their DNA target specificity, implying that ocr targets not sequence specificity but the general DNA binding function of the restriction enzyme [11]. Crystallography [11] revealed that ocr mimicked the bent DNA substrate recognized by the type I enzyme [12] and that the electrostatic mimicry of phosphates by carboxyls was confined to regions of DNA bordering the DNA sequence [10], thus explaining why ocr works against type I enzymes that recognize different DNA sequences. Figure 1 shows the extensive phosphate mimicry on a monomer of ocr.

#### MfpA: a *Mycobacterium tuberculosis* protein that binds to DNA gyrase.

The spread of antibiotic resistance through bacterial populations is of great concern worldwide. The quinolone drugs are of

considerable use because they irreversibly inhibit bacterial DNA gyrase, an essential enzyme in bacteria [44,45]. However, quinolone tolerance is now occurring clinically and in some cases appears to be related to the acquisition by horizontal transfer of the *qnr* gene. The Qnr protein appears to be a member of the pentapeptide-repeat family of proteins [30]. These proteins can form highly elongated structures approximately 2 to 3 nanometres in diameter, an almost perfect framework on which to construct a DNA mimic. The recent structure of MfpA, a *qnr* gene product encoded by *Mycobacterium tuberculosis*, shows a dimeric structure >10 nm in length with a diameter of 2nm at the dimer interface, expanding to 2.7 nm at the extreme ends. The surface of MfpA is decorated with enough carboxyl groups to act as a mimic of double-stranded DNA [18] (Figure 1). MfpA binds to DNA gyrase and rescues it from irreversible inhibition by quinolone drugs and it was immediately suggested that it would fit well into a DNA binding site on the target enzyme, DNA gyrase. In their structural model of MfpA docked onto DNA gyrase, Blanchard and colleagues [18,30] have identified two faces of this molecule that are most likely to interact with the DNA binding site of the enzyme. Binding of MfpA to the gyrase maintains the enzyme in an active form until required, effectively reducing both the binding of the quinolone drug to the gyrase and its clinical efficacy.



**Figure 1.** A structural comparison of the location of the acidic residues (space-filled) on a monomer of the ocr protein (left) with phosphate groups on a simple B-form DNA molecule (centre with phosphates space-filled in red and yellow for each strand) and with the acidic residues (space-filled) on a monomer of the MfpA protein (right). (a) The ocr monomer is orientated with its N-terminus (and dimer interface) at the bottom and the C-terminus at the top. The pattern of the two chains of phosphates on DNA is clearly mimicked by the acidic residues on the ocr monomer, which have been coloured appropriately. The MfpA monomer is shown with its N-terminus at the top and its C-terminus (and dimer interface) at the bottom. The acidic residues suggested in [30] as mimicking the phosphates on DNA which interact with DNA gyrase, are coloured red. Other possible DNA-mimicking acidic residues on the back of the monomer are coloured in pale red. Although the pattern of acidic residues across MfpA roughly follows the path of the phosphates on DNA, the mimicry of B-form DNA by MfpA is not as obvious as with the ocr protein. This might suggest that MfpA mimics a distorted DNA molecule rather than B-form DNA. (b) Views of the molecules from above. It is apparent that the proteins have approximately the same diameter as B-form DNA.

http://www.epibio.com/forum.asp), recognized that the highly negative charge of the ocr dimer would enable it to be used in experiments on bacterial transformation by electroporation. Because most bacteria contain resident type I restriction systems, which impose a strong barrier to transformation, it was reasoned that the simultaneous electroporation of bacteria with ocr and DNA could increase the efficiency of transformation. This appears to be the case, and Epicentre now market ocr as an additive for use in bacterial transformation.

Third, how do we recognize mimics from sequence information? The recognition of DNA mimics presents a challenge for bioinformatics because their sequences and structures are so divergent but it also provides an opportunity to analyze the spread of these genes through bacteria and mobile genetic elements. For instance, the Ard proteins seem to be particularly widespread on conjugative plasmids, transposable elements and bacterial genomes [15,43]. The distribution of DNA mimics, both actual and putative, on mobile genetic elements is of considerable importance, as indicated by the spread of Qnr quinolone-resistance proteins such as the MfpA DNA mimic in clinical environments [44,45]. The pharmaceutical industry should take action to study these antibiotic resistance mechanisms and their potential for control.

#### Potential uses of new DNA mimics

There is much basic research still to be performed in this area but several technological applications of the DNA mimics already present themselves. To reach their full potential, these areas require new mimics: either naturally occurring mimics or laboratory 'designer mimics' constructed by protein design to target any desired protein-DNA interaction. It can be envisaged that DNA mimics could be used in the following, increasingly speculative, areas:

- DNA-binding proteins and enzymes are often difficult to crystallize for structural analysis. It is also often the case that for a protein that recognizes a specific DNA target sequence the actual target sequence is unknown, thus precluding co-crystallization studies. Clearly, when crystallizing DNA-binding proteins it might be possible to use DNA mimics instead of actual DNA molecules in the preparation of co-crystals to aid elucidation of the DNA-binding site on the target protein. Such a use would necessarily rely upon the non-sequence-specific recognition functions of the target protein to enable the use of 'generic' DNA mimics. These generic mimics could imitate, for example, the structure and electrostatics of straight or bent DNA in either its A-form or B-form. Such generic mimics need to be constructed.
- Generic DNA mimics could, of course, be incorporated into a whole gamut of labeling experiments for detection of target DNA-binding proteins and their isolation using affinity methods. An early example of this, although done in reverse, was the demonstration that *E. coli* RNA polymerase, when covalently attached to a chromatography matrix to form an affinity column, could specifically bind and isolate the ocr DNA mimic protein

from an extract of bacteriophage  $\lambda$  [18]. Although the major role of the ocr protein is to knock out DNA restriction, it is sufficiently close in structure to the DNA bound by RNA polymerase to mimic this second protein-DNA interaction. DNA mimics could replace DNA in affinity chromatography, pull-down assays or micro-arrays. A use for fluorescently labeled and targeted mimics in cell imaging could even be envisaged.

- The potential for using mimics in microbiological analyses, as demonstrated by Hoffman *et al.* for the ocr protein, could possibly be extended to use the generic mimics mentioned above. Such mimics might be able to target not only bacterial restriction systems but also other systems such as replication, repair and drug resistance. Thus, diagnostic or even therapeutic uses of mimics in clinical settings might eventually become possible.

#### Conclusion

The authors of the MfpA paper [18] concluded that '[the] structure appears robust enough to allow for surface amino acid substitutions that could tailor specificity and could provide a platform for the rational design of proteins that specifically target DNA-binding proteins of known structure.' This statement could clearly include all of the known DNA mimics and such a use should be investigated in parallel with a search for further examples in nature and the design of synthetic mimics.

#### Acknowledgements

I would like to thank The Leverhulme Trust for the first grant supporting my work on ocr, performed in close collaboration with Malcolm Walkinshaw (University of Edinburgh), and the BBSRC and Darwin Trust for subsequent support. I also gratefully acknowledge the great amount of experimental work performed by my colleagues on ocr, particularly Malcolm Walkinshaw, Shane Sturrock, Paul Taylor, Costel Atanasiu and Jason Blackstock. I am also grateful to David Ratner (Amherst College) for information on his experiments, Les M. Hoffman (Epicentre Biotechnologies, Madison, WI) for communication of results before publication, Noreen Murray and Garry Blakely (University of Edinburgh) for many useful discussions and William Studier (Brookhaven National Laboratory) for the supply of materials.

#### References

- 1 Muller-Hill, B. (1996) *The lac operon: a short history of a genetic paradigm*, Walter de Gruyter
- 2 Ptashne, M. (2004) *A Genetic Switch: Phage Lambda Revisited* (3rd edition), Cold Spring Harbor Laboratory Press
- 3 Klose, R.J. and Bird, A.P. (2006) Genomic DNA methylation: the mark and its mediators. *Trends Biochem. Sci.* 31, 89-97
- 4 Hughes, K.T. and Mathee, K. (1998) The anti-sigma factors. *Annu. Rev. Microbiol.* 52, 231-286
- 5 Campbell, E.A. *et al.* (2003) Crystal structure of *Escherichia coli* sigmaE with the cytoplasmic domain of its anti-sigma RseA. *Mol. Cell* 11, 1067-1078
- 6 Mol, C.D. *et al.* (1995) Crystal structure of human uracil-DNA glycosylase in complex with a protein inhibitor: protein mimicry of DNA. *Cell* 82, 701-708
- 7 Savva, R. and Pearl, L.H. (1995) Nucleotide mimicry in the crystal structure of the uracil-DNA glycosylase-uracil glycosylase inhibitor protein complex. *Nat. Struct. Biol.* 2, 752-757
- 8 Sanderson, R.J. and Mosbaugh, D.W. (1996) Identification of specific carboxyl groups on uracil-DNA glycosylase inhibitor protein that are required for activity. *J. Biol. Chem.* 271, 29170-29181
- 9 Putnam, C.D. *et al.* (1999) Protein mimicry of DNA from crystal structures of the uracil-DNA glycosylase inhibitor protein and its

- complex with *Escherichia coli* uracil-DNA glycosylase. *J. Mol. Biol.* 287, 331–346
- 10 Atanasiu, C. *et al.* (2002) Interaction of the *ocr* gene 0.3 protein of bacteriophage T7 with EcoKI restriction/modification enzyme. *Nucleic Acids Res.* 30, 3936–3944
  - 11 Walkinshaw, M.D. *et al.* (2002) Structure of Ocr from bacteriophage T7, a protein that mimics B-form DNA. *Mol. Cell* 9, 187–194
  - 12 Su, T.-J. *et al.* (2005) DNA bending by M.EcoKI methyltransferase is coupled to nucleotide flipping. *Nucleic Acids Res.* 33, 3235–3244
  - 13 Belogurov, A.A. and Delver, E.P. (1995) A motif conserved among the type I restriction-modification enzymes and antirestriction proteins: a possible basis for mechanism of action of plasmid-encoded antirestriction functions. *Nucleic Acids Res.* 23, 785–787
  - 14 Zavil'gel'skii, G.B. (2000) Antirestriction. *Mol. Biol. (Mosk.)* 34, 854–862
  - 15 Wilkins, B.M. (2002) Plasmid promiscuity: meeting the challenge of DNA immigration control. *Environ. Microbiol.* 4, 495–500
  - 16 Parsons, L.M. *et al.* (2004) Solution structure of the highly acidic protein HI1450 from *Haemophilus influenzae*, a putative double-stranded DNA mimic. *Proteins* 54, 375–383
  - 17 Parsons, L.M. *et al.* (2005) HU-alpha binds to the putative double-stranded DNA mimic HI1450 from *Haemophilus influenzae*. *Protein Sci.* 14, 1684–1687
  - 18 Hegde, S.S. *et al.* (2005) A fluoroquinolone resistance protein from *Mycobacterium tuberculosis* that mimics DNA. *Science* 308, 1480–1483
  - 19 Liu, D. *et al.* (1998) Solution structure of a TBP-TAF(II)230 complex: protein mimicry of the minor groove surface of the TATA box unwound by TBP. *Cell* 94, 573–583
  - 20 Ratner, D. (1974) The interaction of bacterial and phage proteins with immobilized *Escherichia coli* RNA polymerase. *J. Mol. Biol.* 88, 373–383
  - 21 Ramirez, B.E. *et al.* (2000) Solution structure of DinI provides insight into its mode of RecA inactivation. *Protein Sci.* 9, 2161–2169
  - 22 Voloshin, O.N. *et al.* (2001) A model for the abrogation of the SOS response by an SOS protein: a negatively charged helix in DinI mimics DNA in its interaction with RecA. *Genes Dev.* 15, 415–427
  - 23 Yoshimasu, M. *et al.* (2003) An NMR study on the interaction of *Escherichia coli* DinI with RecA-ssDNA complexes. *Nucleic Acids Res.* 31, 1735–1743
  - 24 Lusetti, S.L. *et al.* (2004) The DinI protein stabilizes RecA protein filaments. *J. Biol. Chem.* 279, 30037–30046
  - 25 Bochkareva, E. *et al.* (2005) Single-stranded DNA mimicry in the p53 transactivation domain interaction with replication protein A. *Proc. Natl. Acad. Sci. U. S. A.* 102, 15412–15417
  - 26 Goodsell, D.S. (2005) Recognition in action: DNA mimicry. *J. Mol. Recognit.* 18, 427–430
  - 27 Putnam, C.D. and Tainer, J.A. (2005) Protein mimicry of DNA and pathway regulation. *DNA Repair* 4, 1410–1420
  - 28 Tock, M.R. and Dryden, D.T.F. (2005) The biology of restriction and anti-restriction. *Curr. Opin. Microbiol.* 8, 466–472
  - 29 Dryden, D.T.F. and Tock, M.R. (2006) DNA mimicry by proteins. *Biochem. Soc. Trans.* 34, 317–319
  - 30 Vetting, M.W. *et al.* (2006) Pentapeptide repeat proteins. *Biochemistry* 45, 1–10
  - 31 Susskind, M.M. and Botstein, D. (1978) Molecular genetics of bacteriophage P22. *Microbiol. Rev.* 42, 385–413
  - 32 Prell, H.H. and Harvey, A.M. (1983) P22 antirepressor protein prevents *in vivo* recA-dependent proteolysis of P22 repressor. *Mol. Gen. Genet.* 190, 427–431
  - 33 Heinzl, T. *et al.* (1992) C1 repressor of phage P1 is inactivated by noncovalent binding of P1 Coi protein. *J. Biol. Chem.* 267, 4183–4188
  - 34 Velleman, M. *et al.* (1992) The Bof protein of bacteriophage P1 exerts its modulating function by formation of a ternary complex with operator DNA and C1 repressor. *J. Biol. Chem.* 267, 12174–12181
  - 35 Riedel, H.D. *et al.* (1993) The antirepressor of phage P1. Isolation and interaction with the C1 repressor of P1 and P7. *FEBS Lett.* 334, 165–169
  - 36 Liu, T. *et al.* (1998) The E protein of satellite phage P4 acts as an anti-repressor by binding to the C protein of helper phage P2. *Mol. Microbiol.* 30, 1041–1050
  - 37 Shearwin, K.E. *et al.* (1998) The Tum protein of coliphage 186 is an antirepressor. *J. Biol. Chem.* 273, 5708–5715
  - 38 Liu, T. and Haggard-Ljungquist, E. (1999) The transcriptional switch of bacteriophage Wφ, a P2-related but heteroimmune coliphage. *J. Virol.* 73, 9816–9826
  - 39 Ravin, N.V. *et al.* (1999) The anti-immunity system of phage-plasmid N15: identification of the antirepressor gene and its control by a small processed RNA. *Mol. Microbiol.* 34, 980–994
  - 40 Eriksson, S.K. *et al.* (2000) Interacting interfaces of the P4 antirepressor E and the P2 immunity repressor C. *Mol. Microbiol.* 36, 1148–1155
  - 41 Calisto, B.M. *et al.* (2005) Crystal structure of a putative type I restriction-modification S subunit from *Mycoplasma genitalium*. *J. Mol. Biol.* 351, 749–762
  - 42 Kim, J.S. *et al.* (2005) Crystal structure of DNA sequence specificity subunit of a type I restriction-modification enzyme and its functional implications. *Proc. Natl. Acad. Sci. U. S. A.* 102, 3248–3253
  - 43 Cerdano-Tarraga, A.M. *et al.* (2005) Extensive DNA inversions in the *B. fragilis* genome control variable gene expression. *Science* 307, 1463–1465
  - 44 Tran, J.H. and Jacoby, G.A. (2002) Mechanism of plasmid-mediated quinolone resistance. *Proc. Natl. Acad. Sci. U. S. A.* 99, 5638–5642
  - 45 Li, X.Z. (2005) Quinolone resistance in bacteria: emphasis on plasmid-mediated mechanisms. *Int. J. Antimicrob. Agents* 25, 453–463
  - 46 Dunn, J.J. *et al.* (1981) Amino acid sequence of the gene 0.3 protein of bacteriophage T7 and nucleotide sequence of its mRNA. *J. Biol. Chem.* 256, 2579–2585
  - 47 Mark, K.K. and Studier, F.W. (1981) Purification of the gene 0.3 protein of bacteriophage T7, an inhibitor of the DNA restriction system of *Escherichia coli*. *J. Biol. Chem.* 256, 2573–2578
  - 48 Murray, N.E. (2000) Type I restriction systems: sophisticated molecular machines (a legacy of Bertani and Weigle). *Microbiol. Mol. Biol. Rev.* 64, 412–434
  - 49 Loenen, W.A. (2003) Tracking EcoKI and DNA fifty years on: a golden story full of surprises. *Nucleic Acids Res.* 31, 7059–7069
  - 50 Roberts, R.J. *et al.* (2005) REBASE—restriction enzymes and DNA methyltransferases. *Nucleic Acids Res.* 33, D230–D232