The bi-loop, a new general four-stranded DNA motif

(cyclic oligonucleotide/DNA structure/quadruplex/recombination/strand exchange)

STEPHEN A. SALISBURY*, SARAH E. WILSON*, HAROLD R. POWELL*, OLGA KENNARD*†, PAOLO LUBINI‡,
GEORGE M. SHELDRICK‡, NURIA ESCAJA§, ELMOSTAFA ALAZZOUZI§, ANNA GRANDAS§, AND ENRIQUE PEDROSO§

*Cambridge Crystallographic Data Centre, 12 Union Road, Cambridge CB2 1EZ, United Kingdom; ‡Institut für Anorganische Chemie der Universität Göttingen,
Tammannstrasse 4, D-37077 Göttingen, Germany; and §Departament de Química Orgànica, Universitat de Barcelona, Martí i Franquès 1-11, 08028 Barcelona,
Spain

ABSTRACT The crystal structure of the cyclic octanucle-
otide d<pATTCATTC> contains two independent molecules
that form a novel quadruplex by means of intermolecular
Watson–Crick A·T pairs and base stacking. A virtually iden-
tical quadruplex composed of G·C pairs was found by earlier
x-ray analysis of the linear heptamer d(GCATGCT), when the
DNA was looped in the crystal. The close correspondence
between these two structures of markedly dissimilar oligonu-
cleotides suggests that they are both examples of a previously
unrecognized motif. Their nucleotide sequences have little in
common except for two separated 5′-purine-pyrimidine
dinucleotides forming the quadruplex, and by implication
these so-called "bi-loops" could occur widely in natural DNA.
Such structures provide a mechanism for noncovalent linking
of polynucleotides *in vivo*. Their capacity to associate by base
stacking, demonstrated in the crystal structure of d(GCAT-
GCT), creates a compact molecular framework made up of
four DNA chains within which strand exchange could take
place.

Recognition between DNA molecules can occur in structural
motifs other than the classical double helix, where Watson–
Crick base pairing dominates. In model systems, triplexes are
formed when a third strand is accommodated sequence-
specifically within the major groove of a duplex (1–3), while
self-association of dC_n and dG_n oligonucleotides produces
four-stranded assemblies characterized respectively as the
i-motif (4–6) and G-tetrad (7–9). The relevance of these
structures to biological situations is inferred both from their
utility in providing mechanistic explanations (10–13) and by
the occurrence of susceptible sequences in natural DNA. The
crystalline state provides a ready means of characterizing such
structures at the level of atomic detail and, because of inter-
actions between adjacent molecules in the tightly packed
lattice, it may also stabilize conformations that in natural DNA
require the presence of proteins or applied torsional stress.

The backbone of the cyclic octanucleotide
d<pATTCATTC> is made up of chemically identical sugar-
phosphate units and is constrained to adopt looped confor-
mations. The organization of the bases and of rotatable bonds
throughout the molecule is difficult to predict but relevant to
understanding the deformation of DNA *in vivo*, for example,
by remotely acting forces.

MATERIALS AND METHODS

Synthesis and Crystallization. The octanucleotide was as-
sembled using phosphoramidite chemistry and a polystyrene-
derived matrix. Attachment of the 3′-terminus via a novel
linker based on 3-chloro-4-hydroxyphenylacetic acid allowed
final cyclization to take place simultaneously with deprotec-
tion and cleavage from the solid matrix. The product was of
>90% purity (A_{260}) and was purified to homogeneity by
standard chromatographic techniques. Full details of the
synthetic method will be published elsewhere. Prisms up to
≈0.3 mm across were produced after three weeks from a
solution initially containing sodium cacodylate (8.8 mM; pH
7.0), barium acetate (13.2 mM), and polyethylene glycol 4000
(14.7% wt/vol), equilibrated against polyethylene glycol 4000
(30% wt/vol) by vapor diffusion in sitting drops.

X-Ray Analysis. Suitable crystals were mounted within a
thin film of perfluoropolyether model RS3000 (Riedel-de
Haën, Seelze, Germany) in rayon loops and flash frozen in a
stream of nitrogen maintained at 120 K (Oxford Cryosystems,
Oxford). X-ray intensity data ($\lambda = 0.934$ Å) were recorded on
a phosphor image plate system (MAR Research, Hamburg,
Germany) at station X11 of DORIS II at Deutsches Elek-
tronen Synchrotron (Hamburg, Germany). Two separate data
sets of consecutive oscillation images each covering 120° were
collected with short and long exposure times to maximize the
dynamic range of detection. These were respectively of 40 and
120 exposures using 180- and 300-mm diameter image plates
at 91.5 and 95.5 mm from the crystal. Reflection data were
extracted and merged using Denzo/Scalepack (14). The
merged dataset contained 19,971 unique reflections with
$R_{merge1} = 0.067$ and $R_{merge2} = 0.057$; $\chi^2 = 1.011$, 100%
complete to 1.1 Å. Crystal data: P2$_1$2$_1$2$_1$; $a = 22.8$ Å, $b = 27.6$
Å, $c = 55.3$ Å. A single heavy atom, Ba, was located by
inspection of the Patterson function and the high quality and
resolution of the data allowed subsequent expansion (15) to
include about half the nucleic acid in the asymmetric unit. The
remainder together with water molecules, one sodium ion and
11 barium ions in locations of partial occupancy were added
during the course of refinement with SHELXL-93 (16), finally
adding hydrogens to the nucleic acid at geometrically deter-
mined positions. Atoms other than H were added according to
chemical criteria as well as the presence of positive peaks in
electron density maps ($F_o - F_c$ and $2F_o - F_c$). Water molecules
in the vicinity of disordered barium ions were added tenta-
tively. Disorder in residues T13 and T17 (see text) was not
incorporated into the model. The least squares refinement
(against F_o^2) proceeded in conjugate gradient mode until close
to convergence and then finally with two cycles as a full matrix
giving $R_1 = 0.183$ (w$R_2 = 0.513$) for 12,662 reflections 9.0–1.1
Å ($F_o > 4\sigma F_o$); $R_1 = 0.0192$ for all 14,810 unique data recorded
in this resolution range. The asymmetric unit contained two

molecules of d<pATTCATTC>, 105 water molecules treated as oxygen atoms, one Na^+ and 11 Ba^{2+} ions of various occupancies in the range 1.0–0.183. The DNA atoms alone were modeled as anisotropic.

RESULTS AND DISCUSSION

Oligonucleotide Structure. The asymmetric unit (Fig. 1) is a dimer formed by two molecules of the cyclic DNA octamer. The molecules are related by two perpendicular noncrystallographic dyad axes and have closely similar conformations (rms deviation = 1.18 Å or 1.15 Å, for the two chemically equivalent superpositions). A third dyad relates the two sequence-equivalent halves of both octamers (rms deviation = 0.83 Å). The d(pATTC) tetramers that make up the chemically unique part of the molecule have similar conformations but their environments in the crystal and their interactions with neighboring molecules differ. Remarkably, the conformational similarity extends also to the cytidine residues that participate in interactions of two distinct types between asymmetric units. The dimer unit itself is stabilized by a combina-
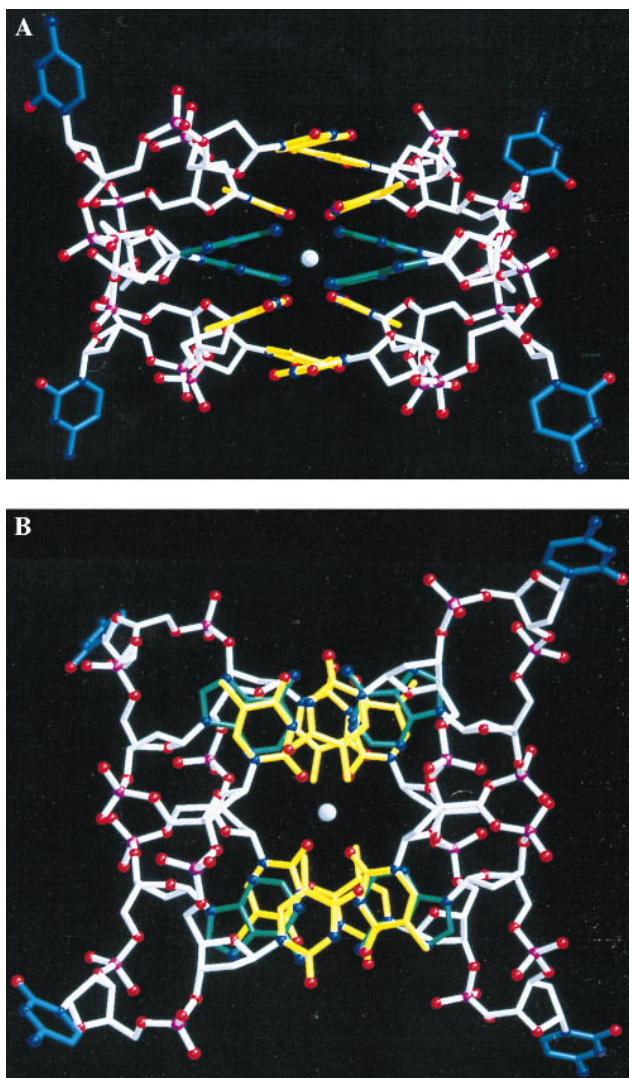

FIG. 1. The two cyclic octanucleotides of which the asymmetric unit is composed, showing (*A*) the relative orientations of base planes and (*B*) a perpendicular view illustrating the stacking of bases. The central sodium ion is included but water molecules and the numerous barium ions located in the analysis are omitted for clarity. Bases are colored: adenine, green; thymine, yellow; cytosine, blue; and the backbone, white (19, 20).

tion of effects that could apply to other sequences, supporting the suggestion that the structure is an example of a more generally accessible motif.

A dominant feature of the structure is the base pairing of A and T between the two molecules that make up the dimer (Fig. 2). The individual base pairs are planar and display little of the propeller twist and buckle observed in the crystal structures of DNA duplexes (17, 18). The four A·T pairs are arranged into two stacks, each of which is capped at either end by a thymine base. There is extensive intermolecular overlap between these unpaired bases and the thymines of the adjacent base pairs. Base pairs in the two stacks face each other via their minor grooves to produce two tiers of a quadruplex. Within each layer, however, the base pairs are not coplanar but have a mutual inclination of $\approx32°$.

The center of the quadruplex is occupied by a sodium ion coordinated to the O^2 atoms of the four base-paired thymines. Together with two water molecules these form an octahedral arrangement, with a uniform Na. . .O distance of 2.8 Å. Coordination is expected to reduce lateral electrostatic repulsion within the quadruplex arising from juxtaposition of the four carbonyl groups, but interactions within the sugar-phosphate backbone are also likely to contribute to the stability of the arrangement. The loops are a consequence of the covalent linkage of the oligomer, but there are many hydrophobic contacts across the component cyclic backbones. Phosphate groups are brought into close proximity: the closest approaches are between O2P of T2 and T6 (4.8 Å) and of the equivalent pair T12 and T16 (4.7 Å).(Residues are numbered A1, T2, etc., for molecule 1 and A11, A12, etc., for molecule 2.)

These phosphate groups do not make direct interactions with cations, but repulsive effects are moderated by nearby centers of positive charge assigned as single barium ions, each distributed between three principal locations. Disordered cations are common in oligonucleotide crystals but because of the large atomic number of barium and the high resolution of the present structure many partially occupied sites could be identified. Only one barium ion position was found to be fully occupied, directly coordinated to O2P of A15 (2.88 Å). The chemically equivalent site in the region of O2P of A11 refined to an occupancy of 0.56.

Organization of the dimers within the crystal is less symmetric than the intrinsic symmetry of the units themselves could allow. Intermolecular interactions at each end of the two stacks of bases are of two types. At one, stacking is continued
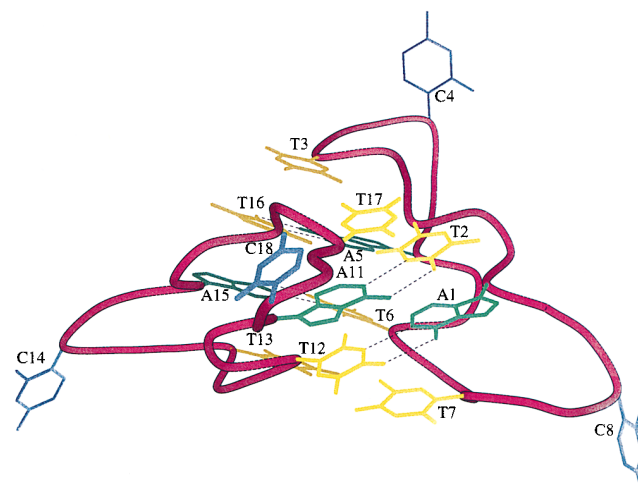

FIG. 2. Schematic representation (19) in which the sugar-phosphate backbone atoms of the dimer are replaced by ribbons. Interactions in the farther stack, where bases are indicated by darker colors, are the same as in the nearer.

between two units by the insertion of cytosines of two other symmetry-related dimers and extends across four asymmetric units in the sequence T3-C14-C18-T7, while at the other, the faces of T13 and T17 are exposed to solvent. The two stacked thymines are in only one of the independent octamer molecules in the asymmetric unit, but each is part of a different base stack (Fig. 2). Stacking between dimer units is thus continuous in the crystal but only one column of bases is involved at each junction, producing an alternating pattern. The remaining cytosines (C4 and C8) do not take part in stacking between asymmetric units but adjoin regions occupied by partially ordered water molecules. One face of each heterocyclic ring is close to the methyl group respectively of T2 and T6, while N^4 atoms form electrostatic contacts of ≈ 2.7 Å with O1P of A1 and A5, respectively.

The Bi-Loop as a General Motif. A quadruplex of the type found in this structure was observed previously in the crystal structure of d(GCATGCT) (21), when the heptamer molecules were folded upon themselves and associated into dimers through formation of G·C pairs. In that case the columns of bases, which were equivalent by crystallographic symmetry, were capped at one end by thymine and at the other by adenine. The central thymine of the sequence extended from the loops in an analogous manner to the cytosines of the present structure and was involved in stacking between dimers. Superposition of the two structures in Fig. 3 illustrates their similarity: the rms difference between the positions of common atoms of the central four base pairs of the two structures is only 0.71 or 0.72 Å, depending on orientation. Interactions that in each individual structure might seem to favour the bi-loop motif are not present in both. This again suggests that the motif may be a more general property of DNA.

In the heptamer structure, potentially disruptive effects due to juxtaposition of O^2 atoms of cytosine were moderated through H-bonds to GN^2 across the quadruplex, rather than by coordination to a cation as observed in the present case for thymine O^2. In both structures the close approach of sugar-phosphate backbones results in extensive hydrophobic contacts but also short interphosphate distances. The shortest distance between phosphate oxygen atoms in the structure of d(GCAT-GCT) is only 3.7 Å (compared with 4.8 Å in the cyclic octamer), but the greater proximity in this case is likely to be due to coordination of a Mg^{2+} ion between two such pairs of atoms in symmetry-related dimers. There is no direct coordination of DNA atoms to barium ions in the cyclic octamer structure. The short phosphate-phosphate distances are considerably less than are encountered in double helix structures, although some phosphate oxygens in G-tetrad and i-motif crystal structures are 5–5.5 Å apart, and in these also backbone contacts resemble those in bi-loop structures.

The two dinucleotides that make up the core of the bi-loop are essentially fragments of B-DNA (23), with a RMS deviation of 0.71 Å. It emphasizes the stability of this form of the double helix and is remarkable since not only do the environments of the A-T dinucleotides differ greatly from that of other B-DNA crystals but the residues also form part of a cyclic backbone. The bi-loop motif itself represents part of the conformational surface of DNA and may be accessible in biological situations such as in a complex with protein or when torsional stress is applied remotely to a region of double-stranded polynucleotide.

It is tempting to speculate that we have observed in this crystal structure a DNA motif of general occurrence in nature. Such a result could be compared with the unexpected discovery of Z-DNA, which was first characterized in crystals of d(CG)$_2$ (24) and d(CG)$_3$ (25), but only later connected with biological functions (26). The minimal requirement for forming a bi-loop, as suggested by these crystallographic results, is a pair of sequences of the type . . . RYNYRYN. . . (R, purine; Y, pyrimidine; N, any nucleotide), where the RY dinucleotides are complementary, but it may also be open to other sequences.

The two heptanucleotides of the type described above, in which RY is either AT or GC, contain only four unique residues and are consequently expected to occur very commonly in nature. Much lower statistical probabilities are, however, associated with multiple repeats of these sequences and their occurrence in natural DNA at frequencies higher than in a random polynucleotide of comparable length could be taken as indicating some specialized function. For example, 11,866 sequences of the type (-ATNNATN-(N)n-)$_5$, where n is 2–20, were identified in the primates section of the EMBL DNA sequence database (27), $\approx 1,000$-fold more than expected in a comparable random sequence of 5.3×10^7 nucleotides. Poly d(AT) and other repetitive sequences that make up significant proportions of genomic DNA are, in addition, effectively multiple repeats of these heptamers.

Implications for Recognition Between DNA Molecules. The quadruplex within the bi-loop is composed of two dinucleotide fragments of double-stranded B-DNA. One of these fragments can be extended by model building into an infinite double helix with the result that the other 2-bp fragment is accommodated within its minor groove with little distortion. Extension of both fragments, however, leads to serious steric clash between backbones so that the crystal structure cannot be used as a prototype of Holliday junction models in the way, for example, that has been suggested by crystal packing of oligonucleotide duplexes (28, 29). The quadruplex structure is, moreover, not consistent with the topology required of such four-way recombination intermediates.

In an alternative view, the six residues ATTCAT of each of cyclic octamer molecule in the present structure make up a loop within a single strand of DNA. Interaction of two such loops in this motif could apply to longer sequences, with the emerging chains either at the same or at opposite ends of the quadruplex. Homologous regions of double stranded DNA
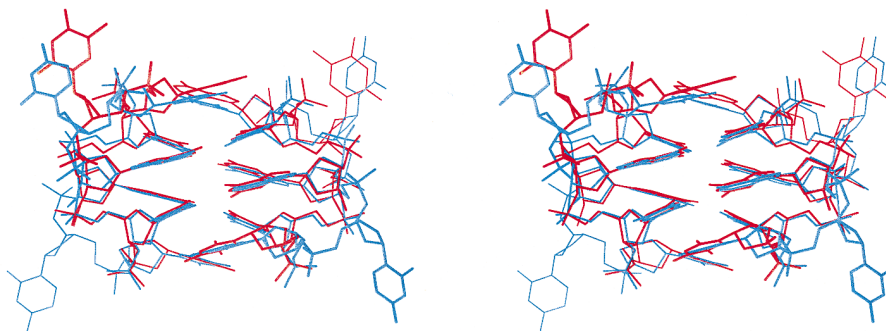

Fig. 3. Stereoview of the superposition (LSQKAB, ref. 22) of the asymmetric unit of the present structure (blue) and an equivalent portion of the crystal structure of d(GCATGCT) (red).
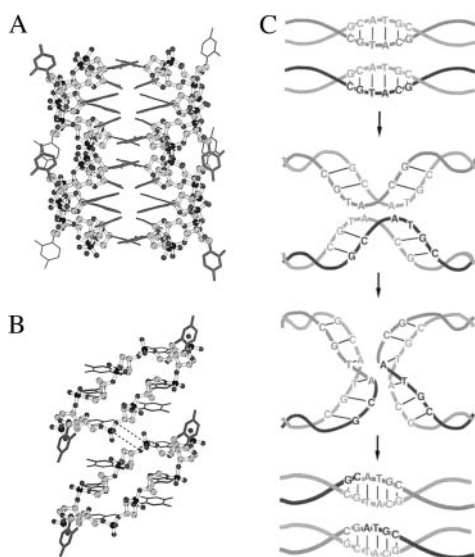
FIG. 4.   (*A*) Two dimers of the cyclic octamer arranged so that the thymine rings of T3 and T17 of one are in van der Waals contact with those of T7 and T13 of the other. For clarity, only backbone atoms are shown as spheres. (*B*) Side view showing only the nearer molecule of each dimer. Putative new phosphodiester bonds between O3′ of T2 and P of T3 are represented as dashed lines. (*C*) Proposed course of recombination according to this mechanism between two duplexes containing the self-complementary sequence d(GCATGC).

could thus be linked by pairs of bi-loops formed after local disruption of the duplexes in a formally, if not structurally, analogous way to "kissing complexes" between RNA molecules (30, 31).

Watson–Crick pairing within bi-loops allows recognition between DNA strands, and these structures could provide a mechanism for sequence alignment prior to genetic recombination. The backbones of the two polynucleotide chains are widely separated in the isolated dimer (Fig. 1) so that the structure does not readily present itself as a framework in which scission and religation of phosphodiester bonds could occur. An interesting possibility is raised, however, when two bi-loops approach each other so that the bases at the end of each stack are in contact (Fig. 4*A*). Steric complementarity is a consequence of the symmetry of the bi-loop structure, and this type of stacking is indeed found in crystals of d(GCAT-GCT). Atoms (P and O3′) that would undergo enzymatic transesterification between the two backbones during recombination are within 4.5–7 Å, depending on whether the stacked bases are thymines as in the present (modeled) case (Fig. 4*B*) or adenines as in the crystal structure of d(GCATGCT). This process would result in a double-stranded join staggered by two nucleotide residues and, where the two bi-loops were formed from duplexes of the same sequence, it would be preserved in the products (Fig. 4*C*). Such an outcome has parallels with some natural processes that have been described (32, 33).

The close resemblance between the crystal structures of d(GCATGCT) and d<pATTCATTC> is likely to be more than coincidence, and we believe may be evidence of a motif of possible biological relevance. By taking advantage of the conformational restraints imposed on the octamer by its cyclic backbone, this molecule and its analogs can be used to generate the bi-loop motif also in solution and would provide a basis for biochemical and immunological experiments to investigate the occurrence of such structures in living organisms.

1. Cooney, M., Czernuszewicz, G., Postel, E. H., Flint, S. J. & Hogan, M. E. (1988) *Science* **241,** 456–459.
2. Doan, T. L., Perrouault, L., Praseuth, D., Habhoub, N., Thuong, J.-L., Lhomme, J. & Helene, C. (1987) *Nucleic Acids Res.* **15,** 7749–7760.
3. Moser, H. E. & Dervan, P. B. (1987) *Science* **238,** 645–650.
4. Berger, I., Kang, C.-H., Fredian, A., Ratliff, R., Moyzis, R. & Rich, A. (1995) *Nat. Struct. Biol.* **2,** 416–425.
5. Gehring, K., LeRoy, J. L. & Gueron, M. (1993) *Nature (London)* **363,** 561–565.
6. Kang, C., Berger, I., Lockshin, C., Ratliff, R., Moyzis, R. & Rich, Z. (1994) *Proc. Natl. Acad. Sci. USA* 11636–11640.
7. Kettani, A., Kumar, R. A. & Patel, D. J. (1995) *J. Mol. Biol.* **254,** 638–656.
8. Kang, C., Zhang, X., Ratliff, R., Moyzis, R. & Rich, A. (1992) *Nature (London)* **356,** 126–131.
9. Laughlan, G., Murchie, A. I. H., Morgan, D. G., Moore, M. H., Moody, P. C. E., Lilley, D. M. J. & Luisi, B. (1994) *Science* **265,** 520–524.
10. Mohanty, D. & Bansal, M. (1995) *Biophys. J.* **69,** 1046–1067.
11. Blackburn, E. H. (1991) *Nature (London)* **350,** 569–573.
12. Venczel, E. A. & Sen, D. (1996) *J. Mol. Biol.* **257,** 219–224.
13. Unrau, P. & Johnson, J. R. (1995) *J. Theor. Biol.* **177,** 73–86.
14. Otwinowski, Z. (1993) in *Proceedings of the CCP4 Study Weekend: Data Collection and Processing 29–30 January 1993*, eds. Sawyer, L., Isaacs, N. & Bailey, S. (SERC Daresbury Laboratory, U.K.), pp. 56–62.
15. Sheldrick, G. M. & Gould, R. O. (1995) *Acta Crystallogr. B* **51,** 423–431.
16. Sheldrick, G. M. & Schneider, T. R. (1997) *Methods Enzymol.* **277,** in press.
17. Bernstein, F. C., Koetzle, T. F., Williams, G. J. B., Meyer, E. F., Brice, M. D., Rogers, J. B., Kennard, O., Shimanouchi, T. & Tasumi, M. (1977) *J. Mol. Biol.* **112,** 535–542.
18. Berman, H. M., Olson, W. K., Beveridge, D. L., Westbrook, J., Gelbin, A., Demeny, T., Hsieh, S., Srinivasan, A. R. & Schneider, B. (1992) *Biophys. J.* **63,** 751–759.
19. Kraulis, P. J. (1991) *J. Appl. Crystallogr.* **24,** 946–950.
20. Merritt, E. A. & Murphy, M. E. P. (1994) *Acta Crystallogr. D* **50,** 869–873.
21. Leonard, G. A., Zhang, S., Peterson, M. R., Harrop, S. J., Helliwell, J. R., Cruse, W. B. T., d'Estaintot, B. L., Kennard, O., Brown, T. & Hunter, W. N. (1995) *Structure (London)* **3,** 335–340.
22. Collaborative Computing Project Number 4 (1994) *Acta Crystallogr. D* **50,** 760–763.
23. Arnott, S. & Hukins, D. W. L. (1972) *Biochem. Biophys. Res. Commun.* **47,** 1504–1509.
24. Drew, H. R. Takano, T., Tanaka, S. Itakura, K. & Dickerson, R. E. (1980) *Nature (London)* **286,** 567–573.
25. Wang, A. H.-J., Quigley, G. J., Kolpack, F. J., Crawford, J. L., van Boom, J. H., van der Marel, G. A. & Rich, A. (1979) *Nature (London)* **282,** 680–686.
26. Herbert, A. & Rich, A. (1996) *J. Biol. Chem.* **271,** 11595–11598.
27. European Bioinformatics Institute (1996) *EMBL DNA Sequence Database* (Cambridge, U.K.).
28. Goodsell, D., Greskowiak, K. & Dickerson, R. (1995) *Biochemistry* **34,** 1022–1029.
29. Timsit, Y., Westhof, E., Fuchs, R. P. P. & Moras, D. (1989) *Nature (London)* **341,** 459–462.
30. Muriaux, D., Fossé, P. & Paoletti, J. (1996) *Biochemistry* **35,** 5075–5082.
31. Paillart, J.-C., Skrimpkin, E., Ehresmann, B., Ehresmann, C. & Marquet, R. (1996) *Proc. Natl. Acad. Sci. USA* **93,** 5572–5577.
32. Berg, D. E. & Howe, M. M. (1989) *Mobile DNA* (Am. Soc. Microbiol., Washington, DC).
33. Kucherlapati, R. & Smith, G. R. (1988) *Genetic Recombination* (Am. Soc. Microbiol., Washington, DC).