

C-1 XP2 system and make serious computational demands. We are therefore adapting the method to use transputers so that the resources needed will not be beyond the scope of typical research laboratories. The ability to compute redox potentials should permit the rational design of bioreductive agents with a specified difference in free energy between oxidized and reduced forms. Furthermore, it now appears possible to calculate redox potentials of systems for which experimental determination is not possible.

This work was conducted pursuant to a contract with the National Foundation for Cancer Research.

Received 18 February; accepted 12 May 1988.

- Adams, G. E. & Stratford, I. J., *Biochem. Pharmac.* **35**, 71-76 (1986).
- Reynolds, C. A., Richards, W. G. & Goodford, P., *Anti-Cancer Drug Design* **1**, 291-295 (1987).
- Reynolds, C. A., Richards, W. G. & Goodford, P., *J. chem. Soc.* 551-556 (1988).
- Burridge, J. M., Quarendon, P., Reynolds, C. A. & Goodford, P. J., *J. molec. Graphics* **5**, 165-166 (1987).
- Cieplak, P., Singh, U. C. & Kollman, P. A., *Int. J. quant. Chem. QBS14*, 65-74 (1987).
- Wong, C. F. & McCammon, J. A., *J. Am. chem. Soc.* **108**, 3830-3832 (1986).
- Singh, U. C., Brown, F. K., Bash, P. A. & Kollman, P. A., *J. Am. chem. Soc.* **109**, 1607-1614 (1987).
- Bash, P. A., Singh, U. C., Brown, F. K., Langridge, R. & Kollman, P. A., *Science* **235**, 574-575 (1987).

- Bash, P. A., Singh, U. C., Langridge, R. & Kollman, P. A., *Science* **236**, 564-568 (1987).
- Rao, S. N., Singh, U. C., Bash, P. A. & Kollman, P. A., *Nature* **328**, 551-554 (1987).
- Kennedy, K. A., Teicher, B. A., Rockwell, S. & Sartorelli, A. C., *Biochem. Pharmac.* **29**, 1-8 (1980).
- Zwanzig, R. W., *J. chem. Phys.* **22**, 1420-1426 (1954).
- Postma, J. P. M., Berendsen, H. J. & Haak, J. R., *Faraday Symp. Chem.* **17**, 55-67 (1982).
- Tembe, B. L. & McCammon, J. A., *Computers Chem.* **8**, 281-283 (1984).
- Lybrand, T. P., McCammon, J. A. & Wipff, G., *Proc. natn. Acad. Sci. U.S.A.* **83**, 833-835 (1986).
- Jorgensen, W. L. & Ravimohan, C., *J. chem. Phys.* **83**, 3050-3054 (1985).
- Singh, U. C., Weiner, P. K., Caldwell, J. W. & Kollman, P. A., *AMBER Version 3.0* (Dept. Pharmaceutical Chemistry, Univ. San Francisco, 1986).
- Roothaan, C. C., *J. Rev. mod. Phys.* **23**, 69-89 (1951).
- Møller, C. & Plesset, M. S., *Phys. Rev.* **46**, 618-622 (1934).
- Dewar, M. J. S. & Stewart, J. J. P., *AMPAC, Quantum Chemistry Program Exchange Bulletin* **6**, 24 (1986).
- Clark, W. M., *Oxidation-Reduction Potentials of Organic Systems* (Balliere, Tindall & Cox, London, 1960).
- Jorgensen, W. L., Chandrasekhar, J., Madura, J., Impey, R. W. & Klein, M. L., *J. chem. Phys.* **79**, 926-935 (1983).
- van Gunsteren, W. F. & Berendsen, H. J. C., *Molec. Phys.* **34**, 1311-1327 (1977).
- Weiner, S. J., *J. Am. chem. Soc.* **106**, 765-784 (1984).
- Weiner, S. J., Kollman, P. A., Nguyen, D. T. & Case, D. A., *J. comput., Chem.* **7**, 230-252 (1986).
- Singh, U. C. & Kollman, P. A., *J. comput. Chem.* **5**, 129-145 (1984).
- Frisch, M., *Gaussian 82 Revision H Version* (Carnegie-Mellon University, Pittsburgh, 1985).
- Amos, R. D., *CADPAC 3.0* (University of Cambridge, 1986).
- Binkley, J. S., Pople, J. A. & Hehre, W. J., *J. Am. chem. Soc.* **102**, 939-947 (1980).
- Hariharan, P. C. & Pople, J. A., *Theor. chim. Acta* **28**, 213-222 (1973).

The three-dimensional structure of a DNA duplex containing looped-out bases

Leemor Joshua-Tor*, Dov Rabinovich*, Håkon Hoppe†, Felix Frolow*, Ettore Appella‡, Joel L. Sussman*§

* Department of Structural Chemistry, The Weizmann Institute of Science, Rehovot 76100, Israel
 † Laboratory of Cell Biology, National Cancer Institute, National Institutes of Health Bethesda, Maryland 20205, USA

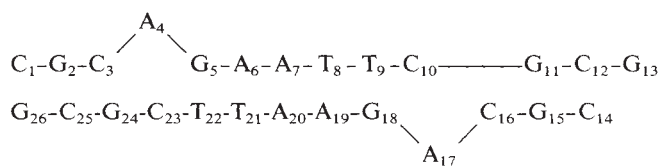
Unpaired bases in DNA have been assigned a possible role in the mechanism of frameshift mutagenesis in sequences with repeated base pairs¹. They also occur in quasipalindromic DNA sequences, which have been implicated in mutagenesis where there are no repeated base pairs, through the formation of single-stranded hairpin loops^{2,3}. The conformation of unpaired bases in DNA has been the subject of numerous thermodynamic as well as high resolution NMR (nuclear magnetic resonance) studies (reviewed in ref. 4). The NMR studies in solution⁵ have shown that the duplex of the tridecamer DNA fragment d(CGCAGAATTCGCG) remains intact, and that the unpaired adenosines are stacked into the duplex. Having crystallized this oligonucleotide and determined its structure, we find its conformation in the crystal is close to that of a B-DNA duplex, with the two additional adenosines looped out from the double helix and causing little disruption of the rest of the structure.

The introduction of an addition base in an otherwise fully complementary double helix could, in principle, produce several alternative structures as shown schematically in Fig. 1 for the tridecamer. These include either a misalignment of the bases in the vicinity of the additional base (Fig. 1a), or accommodation of the extra base without drastic effect on the other base pairs by a looping-out of the extra base away from the duplex (Fig. 1b), or its insertion into the double helix (Fig. 1c). A two-dimensional NMR study⁶ of a very similar sequence d(CGCAGAGCTCGCG) which differs only in the two middle base pairs, strongly suggests that the additional adenosines in both sequences are stacked into the duplex.

Details of the synthesis by the phosphotriester solid-phase method, purification and preliminary crystallization conditions of the duplex:

† Permanent address: Department of Chemistry, University of California, Davis, California 95616, USA.

§ To whom correspondence should be addressed.



have been reported⁷. Shock-cooling to $\sim -150^\circ\text{C}$ (ref. 8) resulted in virtually no radiation-induced damage and thus extended the lifetime of the crystals in the X-ray beam almost indefinitely. X-ray data were collected on a Rigaku AFC5-R rotating anode diffractometer operated at 15 kW. The space group is C2 with unit-cell dimensions $a = 78.48$, $b = 42.84$, $c = 25.16$, $\beta = 99.36$, and one DNA duplex per asymmetric unit. We observed no significant changes in the unit-cell dimensions when the crystal was cooled from 4 to -150°C and virtually no changes during data collection. The structure was solved using ULTIMA⁹ and refined to an R -factor of 15% and a correlation coefficient of 0.95.

Single-crystal X-ray analysis reveals a strikingly different conformation for the unpaired adenosines of the tridecamer in the solid state, as compared to the one derived from solution NMR studies. The overall structure is close to that of B-DNA¹⁰ but the two additional adenosines loop out from the duplex and are not inserted as observed in solution. This looping out is accomplished with minimal disruption of the overall conformation of the helix (Fig. 2) with a mean helix rotation (for the 12 base pairs) of 35° and rise per base pair of 3.38 \AA , as calculated by the program HELIX (written by John Rosenberg)^{11,12}. Figure 3 shows a difference Fourier map in which the atoms belonging to one of the extra adenosines (A17) were omitted from the calculated structure factors. The difference map for the other extra adenosine (A4) is of similar quality. The two adenosines

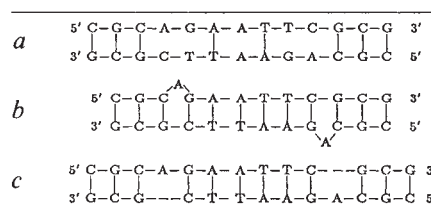
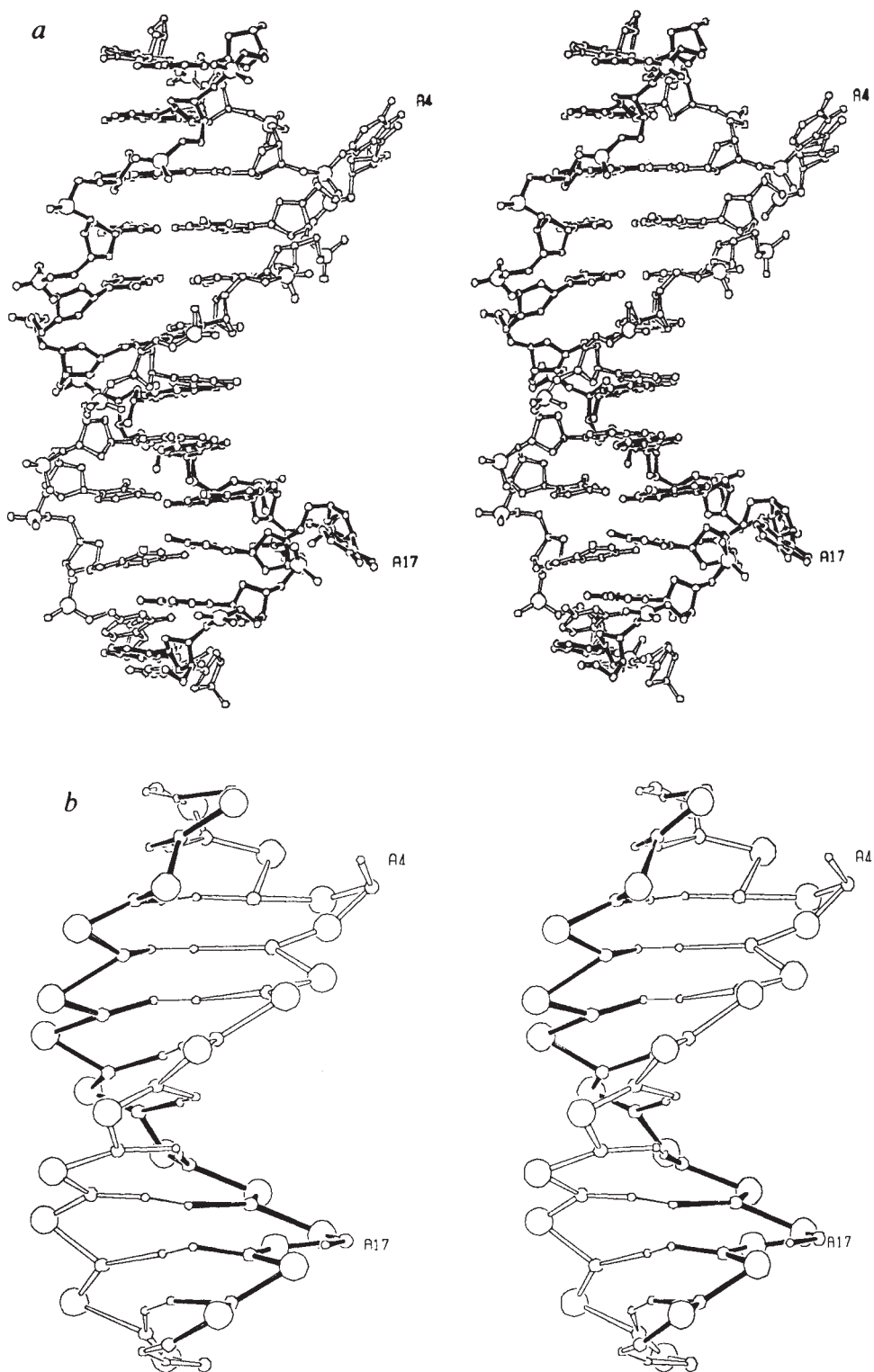


Fig. 1 Schematic representations of possible secondary structures for a tridecamer of the sequence d(CGCAGAATTCGCG)₂. a, An intact duplex with five misaligned base pairs. b, A modified duplex with the unpaired adenosines bulging out. c, A modified duplex with the unpaired adenosines stacked in the helix.

Fig. 2 Stereo diagrams of the tridecamer $d(\text{CGCAGAATTCGCG})_2$ structure showing the extra adenosines bulging out. *a*, Ball and stick model with residues 1-13 in white and residues 14-26 in black. *b*, A schematic representation of the structure in the same view where the sugar-phosphate backbone is represented by phosphorous atoms (large spheres) linked to C1' atoms of the sugars (medium-sized spheres), and the bases by links from the C1' atoms to the N1 atoms in purines or to the N3 atoms in pyrimidines (small spheres). The base-pair interactions are drawn as thin lines. The overall conformation is close to that of B-DNA. The structure was solved via the molecular replacement technique, using the program ULTIMA⁹. A B-DNA dodecamer duplex (with the two extra adenosines omitted) was used as a search model. Initial rigid-body refinement led to an *R*-factor of 38% for the 25-6 Å data. The structure was refined in stages by the constrained restrained least-squares (CORELS) program¹⁹, first as a rigid-body, then as separate base pairs, and finally as individual phosphate groups and nucleosides, using only group thermal parameters²⁰. After the *R*-factor reached ~25% for 10-3 Å data, a difference Fourier map was calculated based on this refined dodecamer which showed clearly the presence of electron density at sites where the two omitted nucleotides might be. The backbones of these two additional residues, consisting of the phosphates and sugars, were initially fitted on a Vector General computer graphics display with the program FRODO.²¹ After additional cycles of refinement and difference Fourier maps, the two adenine bases could be seen. Further refinement with the addition of 50 solvent molecules, using the Hendrickson-Konnert refinement program²² as modified by Westhof for nucleic acids²³, led to a current *R*-factor of 15.0% for 10.0-2.8 Å data for F_{obs} greater than 2σ (r.m.s. deviation of bond length from ideality is 0.01 Å). A detailed description of the structure solution will be given elsewhere.



bulge out from the double helix in quite different ways (Fig. 2). A4 is in the *anti*-conformation, extending out from the backbone and protruding into the minor groove. It appears to be stabilized in part by hydrogen bonding through a water molecule to an oxygen atom of phosphate 12 of a symmetrically-related tridecamer ($\frac{1}{2} - x, \frac{1}{2} + y, -z$). A17 is in a *syn*-conformation, lying along the backbone with its adenine base close to phosphate 16. The base pairs flanking the extra adenosines are stacked on top of each other.

Indications of the existence of other looped-out structures come from solution NMR studies of related sequences, including $d(\text{CAAACAAAG}) \cdot d(\text{CTTTTTTG})$, which contains an unpaired cytosine¹³, and $d(\text{CGCTGAGCTCGCG})_2$, which contains unpaired thymines⁴. In the case of the tridecamer $d(\text{CGCAGAATTCGCG})_2$ two conformations are evidently possible: one with the unpaired adenosines stacked into the duplex as found in solution, and one with the looped-out base as observed here in the crystal. Preliminary refinement of the

Fig. 3 A difference Fourier ($F_{\text{obs}} - F_{\text{calc}}$) map superimposed on a stereo drawing of the region of one of the extra adenosine nucleotides (A17). The entire A17 nucleotide was omitted in the computation of the calculated structure factors. Its density can be seen clearly in the map (lowest contour level at 2σ).

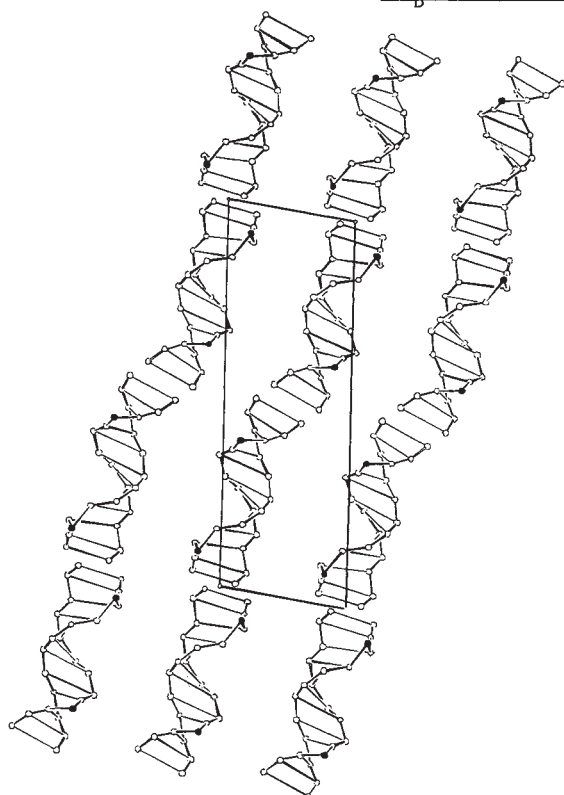
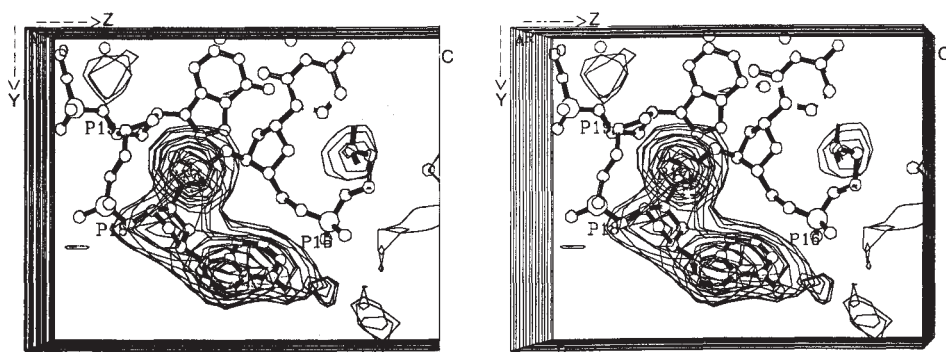


Fig. 4 A packing diagram of the DNA tridecamer structure showing one layer of molecules as viewed down the crystallographic b axis. Each strand is represented by lines connecting the $C1'$ atoms, and the base pairs are drawn as straight lines connecting the $C1'$ atoms on opposite strands. The two extra adenosine nucleotides are shown as filled circles. The packing in the crystal mimics a continuous DNA polymer with stacking interactions between adjacent duplexes. The rotation between the juxtaposed base pairs of neighboring duplexes differs from that of regular B-DNA, however, (-85° across the twofold axis at the origin and 39° across the twofold axis at $\frac{1}{2}, 0, \frac{1}{2}$).

crystal structure of a related 15-mer $d(\text{CGCGAAATTT-ACGCG})_2$ shows that, as in the tridecamer, the unpaired bases loop-out from the duplex¹⁴, although the packing scheme is different, as was the temperature of data collection. As for the tridecamer, however, two-dimensional NMR solution studies of the 15-mer indicate that these extra purines are inserted¹⁵. These studies clearly demonstrate the flexibility of the DNA molecule and suggest that even weak interactions, with other DNA molecules or proteins, could easily shift the conformation between two very different states. The salt-induced B-to-Z transition^{16,17} is another example of such a change.

The manner in which the tridecamer packs in this crystal simulates a long DNA polymer as the molecules stack lengthwise, although in alternating directions (Fig. 4). If the helix were disrupted, for example stretched or bent as in the structures derived by distance geometry calculations on two-dimensional

NMR data of $d(\text{CGCAGAGCTCGCG})_2$ (ref. 6) and by energy minimization for the tridecamer¹⁸, a B-DNA 'polymer-like' packing arrangement would probably not be possible and the resemblance to a long DNA molecule would be lost. In solution, where oligonucleotide molecules are well separated by intervening solvent, a DNA fragment of this size could accommodate an extra nucleotide either looped-out or inserted into the duplex. In fact, the duplex $d(\text{CGCTGAGCTCGCG})_2$ exhibits both conformations in solution, as observed by NMR studies at different temperatures⁴.

The three-dimensional structure reported here supports the slip-mispairing model proposed by Streisinger¹. It implies that certain DNA molecules can easily accommodate additional looped-out nucleotides in one strand, with minimal disruption of the overall duplex structure. This may be a clue to the observed tendency of certain DNA molecules to undergo frame-shift mutations. It is conceivable that the enzymes involved in DNA replication may be insensitive to a looped-out base in the vicinity of the replication fork, as the integrity of the local three-dimensional structure of the duplex is maintained. It will be especially intriguing to correlate the observed frequencies of frame-shift mutations in various DNA sequences with possible differences in the way the additional nucleotides are accommodated by either insertion or looping out from the double helix, and how the cellular repair mechanisms recognize and cope with these types of aberrant structures.

We thank Mark Saper, Zippora Shakked, Miriam Hirshberg, Boaz Shaanan, Peter Stern, David Davies, Zvi Livneh and Eric Westhof for help at various stages of this work. This work was supported by grants from the US Army Research Office, the US-Israel Binational Science Foundation, the Kimmelman Center for Biomolecular Structure and Assembly and the Minerva Foundation, Munich, Germany.

Received 15 April; accepted 1 June 1988.

- Streisinger, G. *et al. Cold Spring Harbor Symp. quant. Biol.* **31**, 77-84 (1966).
- Ripley, L. S. *Proc. natn. Acad. Sci. U.S.A.* **79**, 4128-4132 (1982).
- De Boer, J. G. & Ripley, L. S. *Proc. natn. Acad. Sci. U.S.A.* **81**, 5528-5531 (1984).
- Patel, D. J., Shapiro, L. & Hare, D. Q. *Rev. Biophys.* **20**, 35-112 (1987).
- Patel, D. J. *et al. Biochemistry* **21**, 445-451 (1982).
- Hare, D., Shapiro, L. & Patel, D. J. *Biochemistry* **25**, 7456-7464 (1986).
- Saper, M. A. *et al. J. molec. Biol.* **188**, 111-113 (1986).
- Hope, H. *Acta crystallogr.* **B44**, 22-26 (1988).
- Rabinovich, D. & Shakked, Z. *Acta Crystallogr.* **A40**, 195-200 (1984).
- Wing, R. *et al. Nature* **287**, 755-758 (1987).
- Dickerson, R. E., Kopka, M. L. & Pjura, P. in *Biological Macromolecules and Assemblies: Nucleic Acids and Interactive Proteins* (eds Jurnak, F. A. & McPherson, A.) **2**, 38-126 (Wiley, New York, 1984).
- Fratini, A. V., Kopka, M. L., Drew, H. R. & Dickerson, R. E. *J. biol. Chem.* **257**, 14686-14707 (1982).
- Morden, K. M., Chu, Y. G., Martin, F. H. & Tinoco, I. Jr *Biochemistry* **22**, 5557-5563 (1983).
- Miller, M., Harrison, R., Wlodawer, A., Appella, E. & Sussman, J. L. *Nature* **334**, 85-86 (1988).
- Roy, S., Sklenar, V., Appella, E. & Cohen, J. *Bipolymers* **26**, 2041-2052 (1987).
- Pohl, F. M. & Jovin, T. M. *J. molec. Biol.* **67**, 375-396 (1972).
- Wang, A. H.-J. *et al. Nature* **282**, 680-686 (1979).
- Hirshberg, M., Sharon, R. & Sussman, J. L. *J. biomolec. struct. Dynam.* **5**, 965-979 (1988).
- Sussman, J. L., Holbrook, S. R., Church, G. M. & Kim, S.-H. *Acta crystallogr.* **33**, 800-804 (1977).
- Sussman, J. L., Holbrook, S. R., Warrant, R. W., Church, G. M. & Kim, S.-H. *J. molec. Biol.* **123**, 607-630 (1978).
- Jones, T. A. *J. appl. Crystallogr.* **11**, 268-272 (1978).
- Hendrickson, W. A. & Konnet, J. H. in *Biomolecular Structures, Conformation, Function and Evolution* (eds Srinivasan, R., Subramanian, E. & Yathindra, N.), **1**, 43-57 (Pergamon, Oxford, 1981).
- Westhof, E., Dumas, P. & Moras, D. *J. molec. Biol.* **184**, 119-145 (1985).