Kopka, M. L., Yoon, C., Goodsell, D., Pjura, P., & Dickerson, R. E. (1985b) *J. Mol. Biol. 183*, 553–563.
Lown, J. W., & Krowicki, K. (1985) *J. Org. Chem. 50*, 3775–3779.
Lown, J. W., Krowicki, K., Balzarini, J., & De Clercq, E. (1986a) *J. Med. Chem. 29*, 1210–1214.
Lown, J. W., Krowicki, K., Bhat, U. G., Skorobogaty, A., Ward, B., & Dabrowiak, J. C. (1986b) *Biochemistry 25*, 7408–7416.
Matteucci, M. D., & Caruthers, M. H. (1981) *J. Am. Chem. Soc. 103*, 3185–3191.
Narendra, N., Ginell, S. L., Russu, I. M., & Berman, H. B. (1991) *Biochemistry* (following paper in this issue).
Nelson, H. C. M., Finch, J. T., Luisi, B. F., & Klug, A. (1987) *Nature 330*, 221–226.
Privé, G. G., Heinemann, U., Chandrasegaran, S., Kan, L. S., Kopka, M. L., & Dickerson, R. E. (1987) *Science 238*, 498–504.
Privé, G. G., Yanagi, K., & Dickerson, R. E. (1990) *J. Mol. Biol. 216*, (in press).
Pullman, B. (1983) *J. Biomol. Struct. Dyn. 1*, 773–794.
Sinha, N. D., Biernat, J., McNanus, J., & Koster, H. (1984) *Nucleic Acids Res. 12*, 4539–4557.
Wade, W. S., & Dervan, P. B. (1987) *J. Am. Chem. Soc. 109*, 1574–1575.
Westhof, E., Dumas, P., & Moras, D. (1985) *J. Mol. Biol. 184*, 119–145.
Wing, R. M., Drew, H. R., Takano, T., Broka, C., Tanaka, S., Itakura, K., & Dickerson, R. E. (1980) *Nature 287*, 755–758.
Yanagi, K., Privé, G. G., & Dickerson, R. E. (1990) *J. Mol. Biol. 216*, (in press).
Yoon, C., Privé, G. G., Goodsell, D. S., & Dickerson, R. E. (1988) *Proc. Natl. Acad. Sci. U.S.A. 85*, 6332–6336.
Zimmer, C. (1975) *Prog. Nucleic Acid Res. Mol. Biol. 15*, 285–318.

Crystal and Molecular Structure of a DNA Fragment: d(CGTGAATTCACG)[†,‡]

Narendra Narayana,[§,∥] Stephan L. Ginell,[§] Irina M. Russu,[⊥] and Helen M. Berman*[,§]
Department of Chemistry, Rutgers University, Piscataway, New Jersey 08855, and Department of Molecular Biology and Biochemistry, Wesleyan University, Middletown, Connecticut 06457

Received September 11, 1990; Revised Manuscript Received January 2, 1991

ABSTRACT: The crystal structure of the dodecanucleotide d(CGTGAATTCACG) has been determined to a resolution of 2.7 Å and refined to an *R* factor of 17.0% for 1532 reflections. The sequence crystallizes as a B-form double helix, with Watson–Crick base pairing. This sequence contains the *Eco*RI restriction endonuclease recognition site, GAATTC, and is flanked by CGT on the 5′-end and ACG on the 3′-end, in contrast to the CGC on the 5′-end and GCG on the 3′-end in the parent dodecamer d-(CGCGAATTCGCG). A comparison with the isomorphous parent compound shows that any changes in the structure induced by the change in the sequence in the flanking region are highly localized. The global conformation of the duplex is conserved. The overall bend in the helix is 10°. The average helical twist values for the present and the parent structures are 36.5° and 36.4°, respectively, corresponding to 10 base pairs per turn. The buckle at the substituted sites are significantly different from those seen at the corresponding positions in the parent dodecamer. Step 2 (GpT) is underwound with respect to the parent structure (27° vs 36°) and step 3 (TpG) is overwound (34° vs 27°). There is a spine of hydration in the narrow minor groove. The N3 atom of adenine on the substituted A10 and A22 bases are involved in the formation of hydrogen bonds with other duplexes or with water; the N3 atom of guanine on G10 and G22 bases in the parent structure does not form hydrogen bonds.

The *Eco*RI restriction endonuclease specifically recognizes the double-stranded DNA sequence 5′-GAATTC-3′ and cleaves it at the GpA bonds. The catalytic rates of the enzyme can be modulated by the sequences flanking the *Eco*RI recognition site. This flanking-sequence effect was first recognized by Halford et al. (1980). These authors studied the reaction of *Eco*RI endonuclease on λ phage DNA, which contains five *Eco*RI sites, numbered srI 1 to srI 5 (Daniels et al., 1983). The reaction rate at site srI 5 (5′...-TGAATTCA-...3′) was found to be enhanced relative to that at site srI 2 (5′...-TGAATTCT...3′). Moreover, the cleavage rates at these two sites exhibited markedly different dependencies on ionic strength and pH. In contrast, the cleavage rate at site srI 4 (5′...-AGAATTCT...3′) was the same as that at site srI 5. These results suggest that the orientation of the AT base pairs adjacent to the *Eco*RI site affects the recognition of this site by the protein. Specifically, when the orientation of the AT base pairs flanking the *Eco*RI site preserves the palindromic character of the site, the cleavage rates are increased (e.g., srI 4 and srI 5 on λ DNA). Otherwise, the cleavage rates are lower (e.g., srI 2 on λ phage).

Biochemical evidence indicates that the contacts of the protein with bases outside the recognition site play an important role in binding and cleavage. Strong ethylation interference has been observed for the two phosphate groups upstream from the recognition site (i.e., ...pNpGAATTC...) (Lu et al., 1981; Lesser et al., 1990). The patterns of ethy-

[†] This work was supported by grants from the NIH (GM 21589) for H.M.B. and the NSF (88-17589) for I.M.R.
* Address correspondence to this author.
[‡] The coordinates as well as the structure factors were deposited in the Protein Data Bank and the Nucleic Acid Database.
[§] Rutgers University.
[∥] Present address: Department of Molecular Biophysics and Biochemistry, Yale University, New Haven, CT 06511.
[⊥] Wesleyan University.

0006-2960/91/0430-4449$02.50/0

lation interference at these sites have been shown to depend on the DNA base sequence (Lesser et al., 1990). These results strongly suggest that sequence-specific conformational features of the DNA are important for recognition by *Eco*RI endonuclease.

In order to understand the role of DNA conformation in the flanking-sequence effect of *Eco*RI endonuclease, the structure of DNA molecules containing the *Eco*RI site and various flanking sequences must be characterized at the atomic level. This is currently done by X-ray diffraction, NMR, and molecular dynamics techniques. The first DNA molecule studied is the dodecamer 5′-CGTGAATTCACG-3′, which provides a model for site srI 5 in λ phage DNA. We report here the crystal structure of this dodecamer, which in conjunction with the forthcoming refined structure of the DNA–*Eco*RI endonuclease complex (Kim et al., 1990) will help elucidate the molecular basis of the flanking-sequence effects.

Materials and Methods

DNA Synthesis and Crystallization. The DNA dodecamer was synthesized by the solid-support phosphoramidite method (Matteuci & Caruthers, 1980, 1981). The crystal was grown from a hanging drop containing 1.22 mM DNA dodecamer, 30 mM sodium cacodylate (pH 7.0), 15 mM magnesium chloride, 2.0 mM spermine, and 10% (v/v) 2-methyl-2,4-pentanediol (MPD) by using vapor diffusion against a 22% MPD reservoir at 4 °C. One crystal, 0.70 × 0.15 × 0.08 mm^3, grown over a period of 1 week was used for data collection.

Data Collection and Processing. The crystal was flash frozen and mounted by a method similar to that reported by Hope (1988). It was transferred from the mother liquor to an oil drop (50% paratone-N and 50% mineral oil). The precipitate and adhering mother liquor were removed by a filter paper wick and by moving it around in the oil. The crystal was scooped from the oil drop with a glass fiber and quickly placed on a goniometer in the path of a cold nitrogen stream (−90 °C) that was momentarily blocked off. The unit-cell dimensions of the crystal are a = 24.78, b = 40.85, and c = 65.67 Å in the space group $P2_12_12_1$. A Nicolet multiwire area detector on a Rigaku RU-200 rotating anode generator was used. Data were collected with four different settings by the rotation–oscillation method. Frame scanning, unit-cell determination, data reduction, and scaling were performed with the program package XENGEN version 1.3 (Howard et al., 1987). The unweighted merging R value was 8.75% with 2075 reflections [$I > 2\sigma(I)$] to a resolution of 2.1 Å, which represents 56% of the possible reflections at that resolution. The F_o's are based upon the simply summed intensities. As discussed in the next section, the refinements were done with 1532 reflections [$I > 2\sigma(I)$], corresponding to 76% of the reflections expected between 10.0- and 2.7-Å resolution.

Structure Solution and Refinement. A V_m value (Matthews, 1968) of 2.3 Å^3/Da indicates the presence of two dodecamer strands in the asymmetric unit and a DNA content of about 50% by weight in the unit cell. The overall similarity of these unit-cell dimensions with the published dimensions of d(CGCGAATTCGCG) (Wing et al., 1980) suggest an isomorphous structure. Accordingly, the coordinates of that dodecamer, with appropriate changes in the sequence, were used as a starting model in the refinement. Rigid-body refinement was initiated by using the program COREL S (Sussman et al., 1977). The model was refined against the experimental data, first as a duplex and subsequently by breaking it into 12 base-paired dinucleosides and 22 phosphate groups. The R factor converged to 26.0% for 6–3-Å resolution, with a correlation coefficient of 0.85. At this stage, refinement was

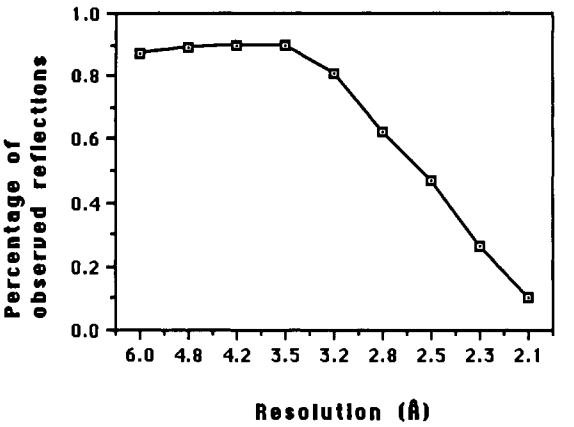

FIGURE 1: Plot showing the variation of the percentage of observed reflections [$I > 2\sigma(I)$] with the resolution (Å).

continued according to the Konnert and Hendrickson (Hendrickson & Konnert, 1981) restrained least-squares refinement procedure with NUCLSQ (Westhof et al., 1985). All observed data [$I > 2\sigma(I)$] between 5 and 2.1 Å were used. Several rounds of refinement and $2F_o - F_c$ maps were made with the PROTEIN (Steigemann, 1985) package and displayed on an Evans and Sutherland PS390 using the program FRODO (Jones, 1978). Atoms were moved in the refined model to fit the electron density. It was found that the number of bad distances (distances deviating from ideality by more than 2σ) and the rms deviation for bond lengths and bond angles were high and the quality of the electron-density map was poor. Specifically, the map contained an uneven density and was peaky in character. The number of observable reflections above the 2σ level represented only ~49% (2075/4250) of the total number of reflections to 2.1 Å, ~70% (1806/2565) to 2.5 Å, and ~76% (1582/2062) to 2.7 Å (Figure 1). Trials of refinement and inspection of electron-density maps were done at several resolutions, and it was decided to use data up to a limit of 2.7 Å. These data behaved well in the refinement process and yielded high-quality $2F_o - F_c$ maps. Fourier maps were made with each base pair removed, and the model was fitted into the resultant electron density. The electron densities were in general seen at the 2σ level. Some of the sugar–phosphate linkages had to be fit at a lower contour level. The R factor converged at 23.0% for 6–2.7 Å. The geometry of the structure was monitored during the refinement. The conformational angles fluctuated considerably. On the other hand, the bases maintained stable geometrical values as judged by analyses of the base morphology (Figure 2). To check if the starting model was biasing the final structure, the structure was rerefined by using fiber B-DNA coordinates (Arnott et al., 1976). The same base geometries were obtained as had been obtained in the first set of refinements.

Both $F_o - F_c$ and $2F_o - F_c$ maps were used to locate solvent molecules. Only peaks that were well shaped, had a minimum value of 2.5 standard deviations, and were within 2.3–3.3 Å of possible hydrogen-bonding partners were accepted as possible solvent molecules. The structure is heavily hydrated with 85 solvent positions, all assigned as water oxygen atoms with full occupancy. It was not possible to identify sodium, magnesium, or spermine ions. The solvent molecules were included in the final stages of refinement. The relative averaged minimal, maximal, and mean crystallographic B values are 4, 8, and 6 Å^2 for the bases, 7, 10, and 8 Å^2 for the deoxyribose sugars, and 7, 12, and 9 Å^2 for the phosphate groups. These values are indicative of the relative motion of the sugar–phosphate backbone compared to the bases. A similar trend

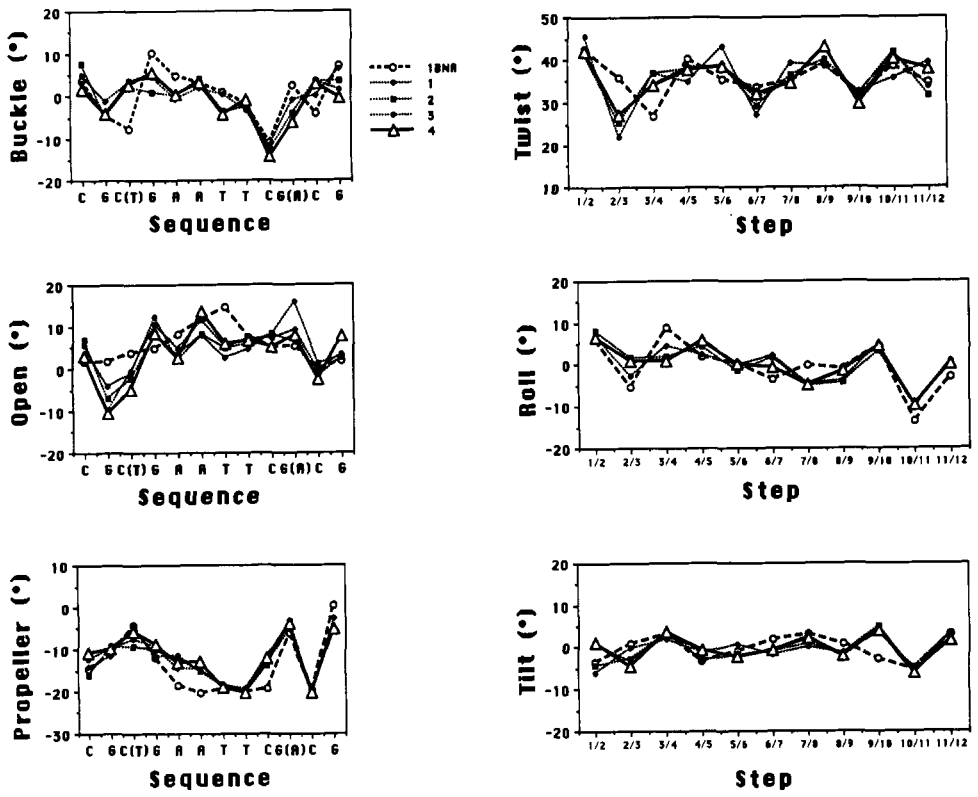

FIGURE 2: Plots showing the variation in the base-pair geometry at various stages of the structure refinement are compared with d-(CGCGAATTCGCG) (Drew et al., 1981). 1BNA represents the parent dodecamer and GTG represents the present structure. The other conditions are as follows: (1) The refinement stage before addition of solvent molecules with an R factor of 23.0% for 6–2.7 Å. (2) This represents a stage with 46 solvent molecules and an R factor of 20.0% for 6–2.7 Å. (3) An intermediate stage with 68 water molecules and an R factor of 18.0% for 10–2.7 Å. (4) The final set of values with 85 water oxygens and an R factor of 17.0% for 10–2.7 Å.

for the B values is also seen in the parent structure.

The details of the refinement are shown in Table I. Refinement restraints were applied to covalent bond lengths and bond angles, chirality of sugar atoms, planarity of base atoms, nonbonded repulsive contacts, and hydrogen-bond distances. The deoxyribose rings in the structure were fit to the electron density and allowed to refine. Soft pseudorotation restraints were imposed only on the deoxyribose belonging to residue 24. Weak hydrogen-bonding restraints were used for the Watson–Crick base pairs. Loose restraints were imposed on some of the sugar–phosphate backbone torsion angles about the P–O5′ and C5′–C4′ bonds. Soft restraints were imposed on the individual isotropic temperature factors.

The final R factor is 17.0% with a correlation coefficient of 0.94 for 1532 [$I > 2\sigma(I)$] observations between 10 and 2.7 Å. The final rms shifts and deviations for atomic positions, bond lengths, and bond angles are 0.065 Å, 0.013 Å, and 1.4°, respectively.

Results and Discussion

Overall Structure. The asymmetric unit consists of two chemically equivalent self-complementary dodecanucleotide strands forming an antiparallel duplex with 12 Watson–Crick base pairs. The helix is of the B-DNA type. We have numbered the bases C1 through G12 in the 5′ to 3′ direction on one strand and C13 through G24 in the 5′ to 3′ direction on the other strand. The solvent molecules are labeled W1 to W85. The replacement of C by T at the third position and G by A at the tenth position is shown schematically below:

d(C-G-C-G-A-A-T-T-C-G-C-G) Parent dodecamer (Wing et al., 1980)

↓ ↓

d(C-G-T-G-A-A-T-T-C-A-C-G) This structure

The sequence contains two GTG/CAC triplets and is the first B-form structure having such a GTG/CAC stretch. This analysis shows that the overall geometry has not been affected

Table I: Crystallographic and Stereochemical Refinement Parameters[a]

resolution range	10–2.7 Å
number of reflections [$I > 2\sigma(I)$]	1532
temperature of study	−90.0 °C
final R factor	17.0%
correlation coefficient	0.935
distances > 2σ	79
σ on F_o map (e/Å^3)	0.33
highest $F_o - F_c$ peak (e/Å^3)	0.23
lowest $F_o - F_c$ peak (e/Å^3)	−0.28
σ on $F_o - F_c$ map(e/Å^3)	0.07
sugar–base bond distances	0.013/0.020 Å
sugar–base bond angle distances	0.029/0.030 Å
phosphate bond distances	0.031/0.030 Å
phosphate angle and H-bond distances	0.038/0.040 Å
planar groups	0.008/0.015 Å
chiral volumes	0.175/0.15 Å^3
single torsion contacts	0.159/0.40 Å
multiple torsion contacts	0.253/0.40 Å
possible hydrogen bonds	0.278/0.40 Å
isotropic thermal factors	
sugar–base bonds	1.019/3.5 Å^2
sugar–base angles	1.362/3.5 Å^2
phosphate bonds	2.889/5.0 Å^2
phosphate angles, H-bonds	1.944/10.0 Å^2
weighting scheme applied to the structure factors[b]	[1/(SIGAPP)^2]
AFSIG	13.5
BFSIG	−100.0

[a] The R factor is defined as $\sum|F_o - F_c|/\sum F_o$, and the correlation coefficient is $\sum[(F_o - \langle F_o\rangle)(F_c - \langle F_c\rangle)]/[\sum(F_o - \langle F_o\rangle)^2\sum(F_c - \langle F_c\rangle)^2]^{1/2}$. For stereochemical parameters the left number gives the rms deviation from ideality and the right number is the σ value used in the refinement. The weight applied on the corresponding restraint is the inverse square of σ. [b] SIGAPP = AFSIG + BFSIG·(STHOL − 0.1666667).

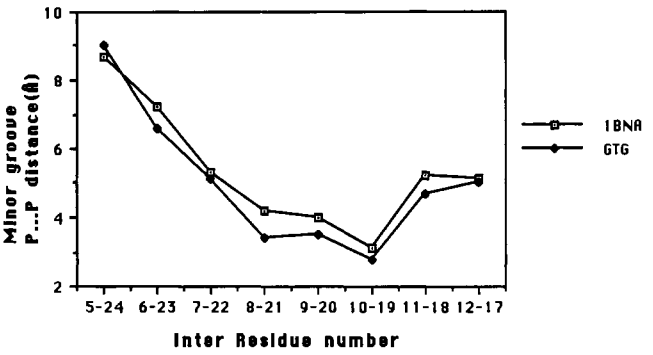

FIGURE 3: Plot of the minor-groove width (Å) in this structure (GTG) as compared to parent dodecamer (1BNA).

by the substitutions. The total rms deviation between the two structures is 0.69 Å, with the average decreasing from the phosphates to the sugars and the bases (0.81, 0.61, and 0.51 Å, respectively). The average unit helical twist for the present and parent structures are 36.5° and 36.4°, respectively, corresponding to 10 base pairs per turn of the duplex. The average axial rise per residue is 3.3 Å as in the parent structure. The structure displays a characteristic narrow minor groove and a wide major groove, both of approximately similar depth, as seen in the parent structure. The bend angle between the first four and last four base pairs is 10° as compared with the value of 12° in the parent dodecamer.

Base-Pair Morphology. There are variations in the local geometry of the bases. Figure 2 shows a comparison of the values (Babcock et al., 1989, 1990) of twist, tilt, and roll for the 11 base-pair steps and buckle, open, and propeller twist for the 12 base pairs in this structure with those of the parent structure. As pointed out before, the bases were distinctly defined in the Fourier maps and did not move significantly during the refinement. In this sense, the base geometry may be considered more reliable than the torsion angles and thus worthy of some comment.

The structure exhibits the characteristic negative propeller twist at all the base pairs. As was seen in the parent structure (Drew et al., 1981), the central AT-rich region has a higher propeller twist (mean = −16.3°) than the ends of the duplex (mean = −9.5°). There is no significant change in the propeller twist at the substituted sites. The graph of the values of "open" in the parent structure differs from this one although there are no significant changes with respect to any single base pair. The values for buckle between the T3 and A22 bases (2.8°) and the A10 and T15 bases (−6.0°) are significantly different from those seen at the corresponding positions in the parent structure (−7.6°, 2.4°). Although the global helical twist is the same as that in the parent structure, there are differences at the local twists. The GpT at step 2 is unwound with respect to the parent structure (27° vs 36°) and the TpG at step 3 is overwound (34° vs 27°). The rise per residue (not illustrated) at step 3 is 3.4 Å and at step 10 is 3.2 Å, which are quite different from that seen in the parent structure, whose corresponding values are 3.0 and 3.7 Å. The values for the tilt and roll angles are similar in both structures.

Sugar–Phosphate Backbone. The P–P distances on both strands range from 5.8 to 7.3 Å with an average of 6.7 Å. In the parent structure the range is from 6.2 to 7.1 Å with an average of 6.7 Å (Dickerson et al., 1983). The measurements provide an estimate of groove width, which is defined as the distance between phosphorus atoms less 5.8 Å to allow for the van der Waals radii of the phosphate groups. In this structure the width of the minor groove varies from 9.0 Å [P(A5)–P(G24)] to 2.8 Å [P(A10)–P(T19)] (Figure 3). There is a conspicuous narrowing of the minor groove at the central AATT region, a feature also seen in the parent structure. The width of the major groove varies from 11 Å [P(A5)–P(A17)] to 12.4 Å [P(G4)–P(A18)]. These values are in good agreement with related B-DNA structures (Dickerson et al., 1983). There is a wider variation of the individual conformation angles (Figure 4) than was observed in the parent structure, although the *average* values are in close agreement with those seen in the parent structure (Drew et al., 1981). The glycosidic torsion angle, χ, is in the usual anti conformation on all the bases. The values for the torsion angle C5′–C4′–C3′–O3′ (δ) associated with the conformation of the deoxyribose ring ranges from 92.5° (C3′ endo) to 163.5° (C2′ endo). The pseudorotation (Altona & Sundaralingam, 1972) parameters for deoxyribose sugars have a wide range of values, with most of the phase values in the 110°–200° domain. The torsion angles α and γ on residue 3 (the site at which C is replaced by T) are −102° and 101°, respectively, and deviate considerably from the corresponding values in parent dodecamer (−63° and 59°, respectively). However, because of the relatively low resolution of the data and the fact that during the course of refinement the values of these angles fluctuated considerably, it is difficult to draw definitive conclusions about any single torsion angle.

Crystal Packing. The crystal packing as illustrated in Figure 5 involves the overlapping ends of the minor grooves of duplexes related by the 2_1 screw along the *c* axis. As in the parent structure, the N2 and N3 atoms of residue 2 form

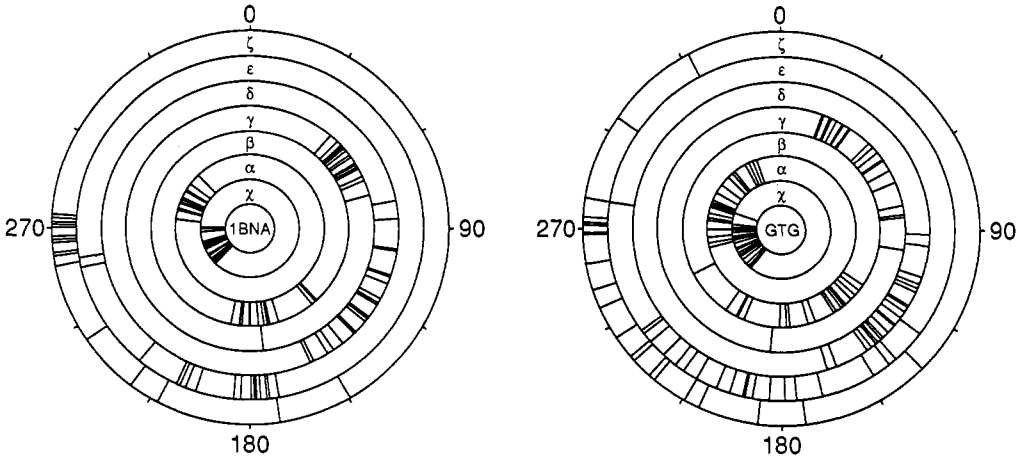

FIGURE 4: Conformational rings (Srinivasan & Olson, 1980) showing the distribution of sugar–phosphate backbone and glycosidic torsion angles (°) in parent (left) and present (right) structures.

FIGURE 5: Schematic diagram showing the direct intermolecular contacts. The backbone is represented by connecting P atoms by a smooth curve and the bases by solid horizontal lines. For clarity the main molecule is represented by bold lines and the symmetry-related molecules by shaded and thin lines. Hydrogen bonds are shown as arrows.

hydrogen bonds with the N3 and N2 atoms, respectively, of residue 12. Similar contacts are found between residues 14 and 24. The terminal O3′ atom on G12 donates a hydrogen bond to N3 of A22 of a symmetry-related duplex. In the parent molecule, in which there is a guanine at position 22, the hydrogen bond is with the N2 atom. In that case the N2 most likely serves as the donor. This is the greatest difference in the packing between the two otherwise isomorphous structures and is directly related to the substitution of bases in the flanking region. The O3′ of G24 is involved in a hydrogen bond with N3 of G16 in a symmetry-related molecule. This contact is the same in the crystal structure of the parent molecule.

In addition to the interactions involving minor-groove atoms and the terminal O3′ atoms, there is a hydrogen-bonding interaction between the terminal O5′(C1) and O5′(T7) and the O1P(T8) belonging to a neighboring symmetry-related duplex. This is correlated with a backbone torsion angle involving O5′(C1) of 29.9° as compared with 174.2° in the parent structure. It is possible that the poor diffraction of these crystals compared to the parent dodecamer may be related in some way to these minor differences in crystal packing.

It should be noted that the substituted base pairs T3-A22 and A10-T15 are located in one of the interfaces between symmetry-related molecules in which the ends of the duplexes form hydrogen bonds with other molecules. Furthermore, one difference in the hydrogen-bonding pattern is due directly to the substitution of G to A. These hydrogen-bonding differences that involve base atoms could affect the base geometries. Hence, the differences seen in the twist patterns (Figure 2) are more likely to be a consequence of these packing differences than to an intrinsic sequence effect.

Hydration. Although the resolution of the data is lower than that used in the room-temperature study of the parent dodecamer, it was possible to locate 85 solvent molecules of which 50% are similar to those seen in the parent structure. The fact that so many were located may be a consequence of the lower temperature used in this data collection. The majority of the waters belong to the primary shell, and only a few were seen in the upper level. There are 13 water molecules in the minor groove. A total of 26 water oxygens are involved in the hydrogen-bonding scheme on the floor of the major groove. The remaining 46 waters are involved in the interactions with the sugar–phosphate backbone with an average of 2 per phosphate group.

Of the 13 water molecules found in the minor groove, 9 have a similar pattern as in the parent dodecamer. Interestingly, water molecules are seen beyond the spine observed in the parent structure (Drew & Dickerson, 1981) extending into the end part of the duplex (Figure 6). These water molecules are involved in minor-groove interactions not seen in the parent: W2 interacts with N3 of the substituted base (A10), as well as with three other atoms. W9 is hydrogen bonded to O1′(A17), O3′(G24), and W4.

In the major groove, N7 and N6 of adenine, N7 and O6 of guanine, O4 of thymine, and N4 of cytosine are solvated. The majority of the waters in the major groove belong to the primary shell, and only a few were seen in the upper level. Intrabase water bridges are observed between N7 and O6 of residue G4 and between N7 and N6 of residue A5. A water molecule (W38) is trapped between the methyl group and a phosphate oxygen O1P at residue T15. Similarly, waters W47 and W85 are involved in methyl–water–phosphate-oxygen interaction at residues T7 and T20, respectively. Such interactions were also observed in the parent structure (Drew & Dickerson, 1981).

Most of the phosphate anionic oxygens are involved in hydrogen bonds and are individually hydrated as seen in the structure of the parent sequence at 16 K and in the 9-bromo derivative d(CGCGAATT^{Br}CGCG) at 280 K (Kopka et al., 1983) (MPD7). Some solvent molecules are involved in bidentate intraphosphate interactions between a free phosphate oxygen and a phosphodiester oxygen O5′ that were also seen in MPD7 structure. Water molecules at residues T3 and T8 interact with two free phosphate oxygens of the same phosphate group, a feature seen in the structure of d(CGCGAATTAGCG) (Hunter et al., 1986). The phosphate O3′ and O5′ atoms are hydrated sparsely.

Conclusions

We report the molecular and crystal structure of a DNA fragment that contains the *Eco*RI restriction site but with different flanking sequences than the d(CGCGAATTCGCG) sequence whose structure was determined previously. Most elements of the crystal structure are very similar to those of the parent molecule, including the overall helix geometry, the crystal packing, and the hydration structure.

The main differences observed are in the local base geometries and in some possibly important elements of the hydrogen bonding. The substituted bases are near the ends of the duplex that are involved in multiple hydrogen-bonding interactions. There is a significant difference in the hydrogen-bonding pattern involving the substituted A22 because this base does not have a 2 amino group and therefore cannot form the same hydrogen bond as in the native structure that contains a guanine at this site. Whether or not this and other more subtle

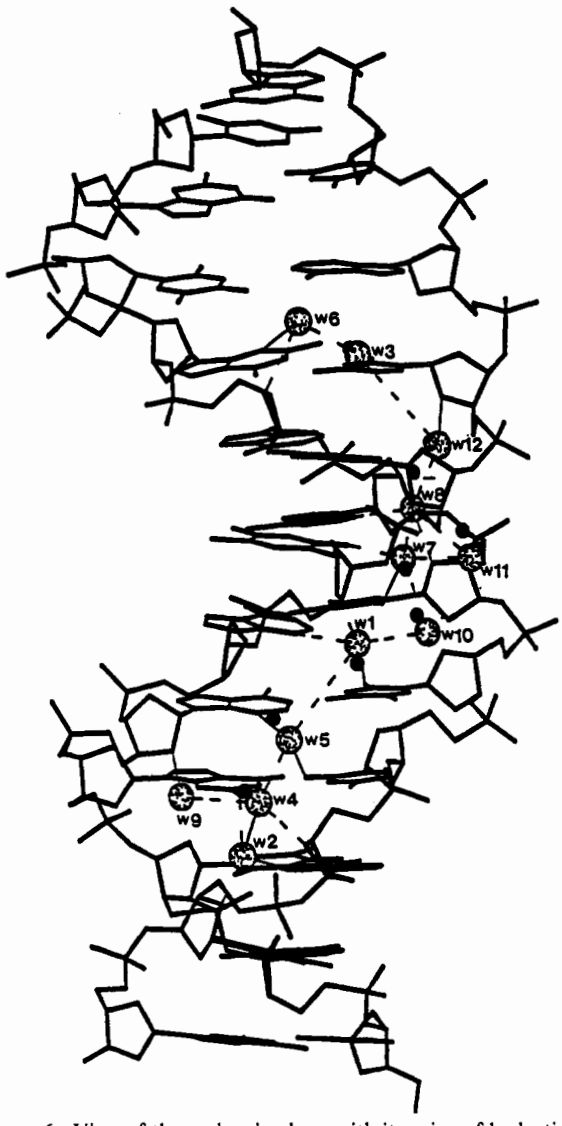

FIGURE 6: View of the molecule along with its spine of hydration in the minor groove, and for comparison the water molecules associated with the spine in the minor groove of parent dodecamer are shown. Second-shell water molecules are not drawn. The water molecules in the present structure are shown as stippled circles and those of parent dodecamer as filled circles that are half the size of the stippled circles. Only the water molecules seen in the present structure are labeled. The minor-groove interactions in this structure involving the water molecules are shown by dashed lines. Interactions involving symmetry-related atoms are not shown in the diagram. There is an extended spine of hydration in this structure as compared to the parent structure (Drew & Dickerson, 1981).

differences can be correlated to differences in the base morphologies cannot be ascertained at this time.

The results of this structural analysis have shown that because the substituted base is involved in significant intermolecular hydrogen-bonding interactions, this sequence is not a good model for studying the intrinsic effects of this particular base substitution by X-ray crystallographic methods. On the other hand, it may be possible to study these effects by either NMR or theoretical methods. Alternatively, it may be possible to embed the recognition and flanking sequences into a decamer in which hydrogen bonding will not play such an important role.

After we had determined this structure, we found that a group at UCLA had also determined the same structure although for a different reason. Preliminary comparisons have been made, and overall the structures are quite similar. The cell dimensions indicate a high degree of isomorphism. However, small differences in the base geometries and hydrogen bonding indicate that there may be some polymorphism that may in turn be related to the generally poorer quality of the crystals as compared with those of the parent structure. A collaborative effort to resolve these issues has begun.

Acknowledgments

Discussions with Richard Dickerson, Teresa Larsen, and Mary Kopka have been informative, delightful, and in the spirit of true scientific interchange. We thank Carol Afshar for her help in growing the crystals. We also thank Marla Babcock for her program to calculate the base-pair geometry and A. R. Srinivasan for his program to generate the conformational rings.

Supplementary Material Available

A table (2 pages) showing the backbone torsion angles and distances between adjacent phosphorus atoms. Ordering information is given on any current masthead page.

Registry No. d(CGTGAATTCACG), 113725-84-1; restriction endonuclease *Eco*RI, 80498-17-5.

References

Altona, C., & Sundaralingam, M. (1972) *J. Am. Chem. Soc. 94*, 8205–8212.

Arnott, S., Campbell-Smith, P. J., & Chandrasekaran, R. (1976) in *CRC Handbook of Biochemistry and Molecular Biology, Nucleic Acids* (Fasman, G. D., Ed.) Vol. 2, pp 411–422, CRC Press, Cleveland, OH.

Arnott, S., Chandrasekaran, R., Birdsall, D. L., Leslie, A. G. W., & Ratliff, R. L. (1980) *Nature (London) 283*, 743–745.

Babcock, M. S., Srinivasan, A. R., & Olson, W. K. (1989) *Book of Abstracts: Sixth Conversation in Biomolecular Stereodynamics*, pp 9–10, Albany, New York.

Babcock, M. S., Pednault, E. P. D., Srinivasan, A. R., & Olson, W. K. (1990) *A New Program for the Analysis of Nucleic Acid Structure: Local Cartesian and Helical Parameters*, Poster at the meeting of Mathematical Approaches to DNA, Jan 24–29, Santa Fe, New Mexico.

Daniels, D., Schroeder, J., Szybalski, W., Sanger, F., Coulson, A., Hong, G. D., Petersen, G., & Blattner, F. (1983) in *Lambda II* (Hendrix, R., Roberts, F., Stahl, F., & Weisberg, R., Eds.) pp 519–676, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.

Dickerson, R. E., Kopka, M. L., & Drew, H. R. (1983) *Structure and Dynamics of Nucleic Acids and Proteins* (Clementi, E., & Sarma, R. H., Eds.) pp 149–179, Adenine Press, NY.

Drew, H. R., & Dickerson, R. E. (1981) *J. Mol. Biol. 151*, 535–556.

Drew, H. R., Wing, R. M., Takano, T., Broka, C., Tanaka, S., Itakura, K., & Dickerson, R. E. (1981) *Proc. Natl. Acad. Sci. U.S.A. 78*, 2179–2183.

Halford, S. E., Johnson, N. P., & Grinsted, J. (1980) *Biochem. J. 191*, 581–592.

Hendrickson, W. A., & Konnert, J. H. (1981) in *Biomolecular Structure, Conformation, Function and Evolution* (Srinivasan, R., Ed.) Vol. 1, pp 43–57, Pergamon Press, Oxford.

Hope, H. (1988) *Acta Crystallogr. B44*, 22–26.

Howard, A. J., Gilliland, G. L., Finzel, B. C., Poulos, T. L., Ohlendorf, D. H., & Salemme, F. R. (1987) *J. Appl. Crystallogr. 20*, 383–387.

Hunter, W. N., Brown, T., & Kennard, O. (1986) *J. Biomol. Struct. Dyn. 4*, 173–191.

Jones, T. A. (1978) *J. Appl. Crystallogr. 11*, 268–272.

Kim, Y., Grable, Y. C., Love, R., Greene, P. J., & Rosenberg, J. M. (1990) *Science 249*, 1307–1309.
Kopka, M. L., Fratini, A. V., Drew, H. R., & Dickerson, R. E. (1983) *J. Mol. Biol. 163*, 129–146.
Lesser, D. R., Kurpiewski, M. R., & Jen-Jacobson, L. (1990) *Science 250*, 776–786.
Lu, A.-L., Jack, W. E., & Modrich, P. (1981) *J. Biol. Chem. 256*, 13200–13206.
Matteuci, M. D., & Caruthers, M. H. (1980) *Tetrahedron Lett. 21*, 719–722.
Matteuci, M. D., & Caruthers, M. H. (1981) *J. Am. Chem. Soc. 103*, 3185–3191.
Matthews, B. W. (1968) *J. Mol. Biol. 33*, 491–497.
Srinivasan, A. R., & Olson, W. K. (1980) *Nucleic Acids Res. 8*, 2307–2329.
Steigemann, W. (1974) Ph.D. Thesis, Technical University, Munich, West Germany.
Sussman, J. L., Holbrook, S. R., Church, G. M., & Kim, S.-H. (1977) *Acta Crystallogr. A33*, 800–804.
Westhof, E., Dumas, P., & Moras, D. (1985) *J. Mol. Biol. 184*, 119–145.
Wing, R., Drew, H. R., Takano, T., Broka, C., Tanaka, S., Itakura, K., & Dickerson, R. E. (1980) *Nature (London) 287*, 755–758.

Polyamines Favor DNA Triplex Formation at Neutral pH†

Ken J. Hampel, Paul Crosson, and Jeremy S. Lee*
Department of Biochemistry, University of Saskatchewan, Saskatoon, Saskatchewan S7N 0W0, Canada
Received November 6, 1990; Revised Manuscript Received February 13, 1991

ABSTRACT: The stability of triplex DNA was investigated in the presence of the polyamines spermine and spermidine by four different techniques. First, thermal-denaturation analysis of poly[d(TC)]·poly[d(GA)] showed that at low ionic strength and pH 7, 3 μM spermine was sufficient to cause dismutation of all of the duplex to the triplex conformation. A 10-fold higher concentration of spermidine produced a similar effect. Second, the kinetics of the dismutation were measured at pH 5 in 0.2 M NaCl. The addition of 500 μM spermine increased the rate by at least 2-fold. Third, in 0.2 M NaCl, the mid-point of the duplex-to-triplex dismutation occurred at a pH of 5.8, but this was increased by nearly one pH unit in the presence of 500 μM spermine. Fourth, intermolecular triplexes can also form in plasmids that contain purine·pyrimidine inserts by the addition of a single-stranded pyrimidine. This was readily demonstrated at pH 7.2 and 25 mM ionic strength in the presence of 100 μM spermine or spermidine. In 0.2 M NaCl, however, 1 mM polyamine is required. Since, in the eucaryotic nucleus, the polyamine concentration is in the millimolar range, then appropriate purine·pyrimidine DNA sequences may favor the triplex conformation in vivo.

Triplex structures were first described over 20 years ago (Glaser & Gabbay, 1968; Morgan & Wells, 1968). In the more usual form, they consist of T-A-T and C-G-C^+ base triads in which the third pyrimidine strand winds up the major groove of an A-form helix with Hoogsteen pairing. This necessitates protonation of one of the cytosines, and the two pyrimidine strands are antiparallel. Although no X-ray crystallographic structure is yet available, recent NMR studies support this simple model (Rajagopal & Feigon, 1989; de los Santos et al., 1989). These requirements tend to restrict triplex formation to polypurine·polypyrimidine sequences (pur·pyr DNA) although some mismatches can be tolerated (Hanvey et al., 1989; Griffin & Dervan, 1989). In addition, a low pH is generally required because of the protonated cytosine, which has a pK_a of 4.5 in the free state. Thus synthetic DNAs such as poly[d(TC)]·poly[d(GA)] will dismutate to a triplex at pHs below 6 (Lee et al., 1979). Although this result seems to preclude a physiological role for triplexes, several lines of evidence are suggestive of their presence in eucaryotic cells.

First, pur·pyr tracts, some of which are over 100 base pairs in length, represent up to 1% of certain eucaryotic genomes (Hoffman-Lieberman et al., 1986; Manor et al., 1988). Thus the concentration of potential triplex-forming sequences in the nucleus is quite substantial. Second, many pur·pyr tracts are found in the 5′-flanking region of genes, and these regions are sensitive to single-strand specific nucleases both in vitro and in vivo (Schon et al., 1983; Larsen & Weintraub, 1982). By necessity, triplex formation from two duplexes requires the extrusion of a single strand (Htun & Dahlberg, 1988, 1989). Third, a triplex-specific antibody binds to certain regions of mouse, human, and polytene chromosomes (Lee et al., 1987; Burkholder et al., 1988; and unpublished observations). The binding can be abolished by the addition of competing triplex but not with *Escherichia coli* DNA, for example, in which triplexes cannot be detected (Lee et al., 1989). Although this evidence is compelling, it is not clear what factors might be responsible for stabilizing triplexes under physiological conditions.

Potential triplex stabilizers include negative supercoiling, which promotes intramolecular triplex formation (also called H-DNA) in plasmids containing pur·pyr inserts (Lyamichev et al., 1986; Wells et al., 1988; Htun & Dahlberg, 1989). However, in the normal range of superhelix densities (≈ -0.05), a pH less than 7 is still required (Htun & Dahlberg, 1988; Shimizu et al., 1989). Alternatively, triplexes containing 5-methylcytosine or phosphorothioate groups are stable at neutral pH, but there is no evidence for their presence in pur·pyr tracts (Lee et al., 1984; Povsic & Dervan, 1989; Latimer et al., 1989). Cations are also known to alter triplex

†Supported by MRC Canada by grants to J.S.L.
*Corresponding author.

0006-2960/91/0430-4455$02.50/0