Vol. 131, No. 1, 1985 BIOCHEMICAL AND BIOPHYSICAL RESEARCH COMMUNICATIONS
August 30, 1985 Pages 269-276

CONFORMATIONAL MICROHETEROGENIETY IN A DNA DOUBLE HELIX: STRUCTURE OF RESTRICTION ENDONUCLEASE Bam Hl RECOGNITION SITE

M.H. Sarma, M.M. Dhingra*, G. Gupta and R.H. Sarma

Institute of Biomolecular Stereodynamics
State University of New York at Albany
Albany, New York 12222

Received July 10, 1985

Structural studies using 500 MHz ^{1}H NMR spectroscopy on Bam Hl recognition site d(GGATCC)$_2$ in solution at 19^o is reported. The resonances from the sugar ring and base protons have been assigned from the 2D-COSY and NOESY spectra. Analyses of the NOESY cross-peaks between the base protons H8/H6 and sugar protons H2'/H2", H3' reveal that the nucleotide units G2, A3 and C6 adopt (C3'-endo, χ = 200^o-220^o) conformation while Gl, T4 and C5 exhibit (C2'-endo, χ = 240^o-260^o) conformation. NMR data clearly suggest that the two strands of d(GGATCC)$_2$ are conformationally equivalent and there is a structural two-fold between the two A-T pairs. The above information and the NOESY data are used to generate a structural model of d(GGATCC)$_2$. The important features are: (i) Gl-G2 stack, the site of cleaveage, shows an alternation in sugar pucker i.e. C2'-endo, C3'-endo as in a B-A junction, (ii) G2-A3 stack adopts a mini A-DNA, both the sugars being C3'-endo, (iii) A3-T4 stack, the site of two-fold, displays an A-B junction with alternation in sugar pucker as C3'-endo, C2'-endo, (iv) T4-C5 stack adopts a mini B-DNA both the sugars being C2'-endo and (v) C5-C6 stack exhibits a B-A junction with C2'-endo, C3'-endo sugar puckers. Thus, our studies demonstrate that conformational microheterogeniety with a structural two fold, is present in the Bam Hl recognition site.

The self-complementary hexanucleotide duplex d(GGATCC)$_2$ is the DNA recognition site of the restriction endonuclease Bam Hl(1,2). The enzyme cleaves the P-03' bonds between two Gs on either strand as shown below:

5' Gl↓G2-A3-T4-C5-C6 3' ↓ :sites of cleavage
3' C6-C5-T4-A3-G2↓Gl 5' ● :two-fold symmetry in the base-sequence

Chemical studies have shown that the presence of the exact two-fold symmetry in the DNA base-sequence is very essential for the recognition process of the enzyme (3). Therefore, in order to explain the structural basis of Bam Hl recognition, one has to examine whether there is a structural two-fold, in the DNA fragment coincident with the two-fold, in the base-sequence as postulated for the DNA recognition of CRO, CAP and CI proteins (4-6). In addition, one has to examine if there is a

* Permanent address: Tata Institute of Fundamental Research, Hami Bhabha Road, Bombay, 400005, India.

0006-291X/85 $1.50

special structural feature in the G1-G2 sequence which might explain why it is the only site of enzyme cleavage (1,2). With these goals, we undertook the structural studies of $d(GGATCC)_2$ in solution using 2D NMR spectroscopy at 500 MHz.

Materials and Methods

The Bam H1 fragment $d(GGATCC)_2$ was synthesized from blocked dinucleotides using solution phase phosphotriester techniques (for details see ref. 7). For all 500 MHz NMR measurements we used 12 mM of $d(GGATCC)_2$ in 100 mM NaCl (pH=6.8). TSP was used as internal standard for chemical shift measurements. Regular 1D NMR and 1D difference NOE spectra of $d(GGATCC)_2$ in water were collected using a time shared long pulse sequence with a notch filter of 3750 Hz. For COSY and NOESY experiments, 512 experiments were performed in the t_1 direction with a spectral width of 5000 Hz and a data set of 1K in the t_2 dimension and 124 FIDs were collected with a delay of 1 sec. Both in COSY and NOESY experiments the residual HDO peak was suppressed by presaturation. The mixing time for NOESY was 350 msec.

Assignments of the protons in $d(GGATCC)_2$

The assignments of the protons were carried out at 19°C at which the DNA exists in the duplex state as demonstrated by observing Watson-Crick paired NH protons of G and T of $d(GGATCC)_2$ in 90% H_2O + 10% D_2O. The non-exchangeable protons of $d(GGATCC)_2$ were assigned in the following three steps.

1. Sugar Protons H1', H2'/H2", H3'

Figure 1 shows a COSY spectrum of $d(GGATCC)_2$ in which the scalar couplings between H1' and H2'/H2" belonging to the same sugar manifest as cross-peaks. By observing the cross-peaks between H1' and H2'/H2", we were able to group them into six sets belonging to six different sugar residues. Then by observing the cross-peaks between H2'/H2" and H3' as shown in Figure 2, we identified the spin system comprising H1', H2'/H2", H3' belonging to different sugar units.

2. Base-Protons

H2 of A3 was located at 7.94 ppm by performing a 1-D NOE experiment in 90% H_2O + 10% D_2O in which Watson-Crick paired NH proton of T4 was irradiated. There are two strong NOESY cross-peaks between Me of T4 and the base-protons within 8.50-7.00 ppm. The stronger cross-peak at higher field (7.24 ppm) gives the location of H6 of T4 while the other at a lower field (8.27 ppm) is H8 of A3. Strong NOESY cross-peaks are observed between H6 and H5 of a cytosine residue - thus, two sets (H6, H5) belonging to C5 and C6 are identified (Figure 3). The set with H6 at 7.61 and H5 at 5.70 ppm belongs to C5 because H6 at 7.61 ppm shows a NOESY cross-peak with neighbouring H6 of T4. The other set with H6 at 7.67 and H5 at 5.77 ppm belongs to C6. The signals at 7.84 and 7.87, then, should belong to H8s of G1 and G2. Therefore, by combination of 1D NOE and 2D NOESY experiments on $d(GGATCC)_2$, we are able to sequentially assign the base-protons except for H8s of G1 and G2 (which would be discussed later in the text).

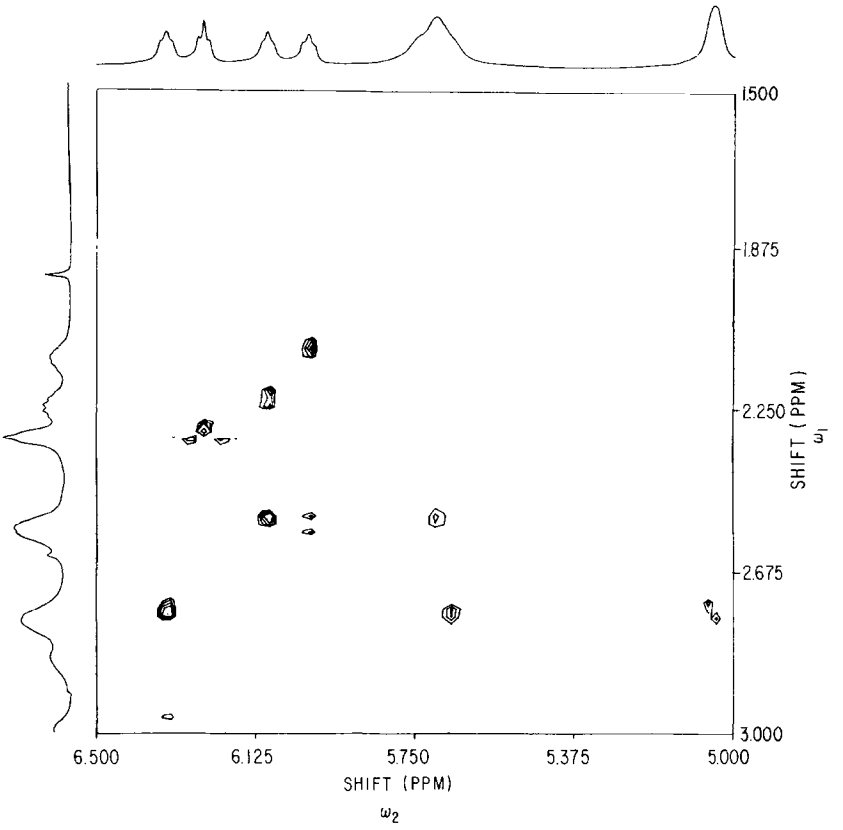

Figure 1. 500 MHz ^{1}H COSY spectrum of d(GGATCC)$_2$ at 19°C showing the cross-peaks between the sugar protons H1' and H2'/H2".

3. Connectivity Between the Base and Sugar Protons

At first, the connectivities between the base-protons H8/H6 and H1's were obtained from the NOESY spectrum shown in Figure 3. In a duplex structure, the base proton of the 5'-terminal nucleotide (in the present case H8 of G1) is only close in space with H1' of its own sugar and is expected to give a single cross-peak along the H1'-axis (see Figure 3). Similarly, H1' of the 3'-terminal nucleotide (in the present case H1' of C6) is only close to its own base proton (i.e. H6 of C6) and should show a single cross-peak along the H8/H6 axis (see Figure 3). Once H1's belonging to terminal nucleotides i.e. G1 and C6, are assigned the connectivities between the H1's and the base-protons of internal nucleotides are established based on the fact that H8/H6 of internal nucleotide (i.e. G2, A3, T4, C5) could be close to its own sugar H1' and the H1' of the 5'-neighbour. All the connectivities involving base-protons H8/H6 and sugar protons H1'are shown in Figure 3. Thus by combining the results of COSY and NOSEY (Figures 1-3) we are able to assign the base and the sugar ring protons H1', H2', H2" and H3' of six nucleotidyl units.

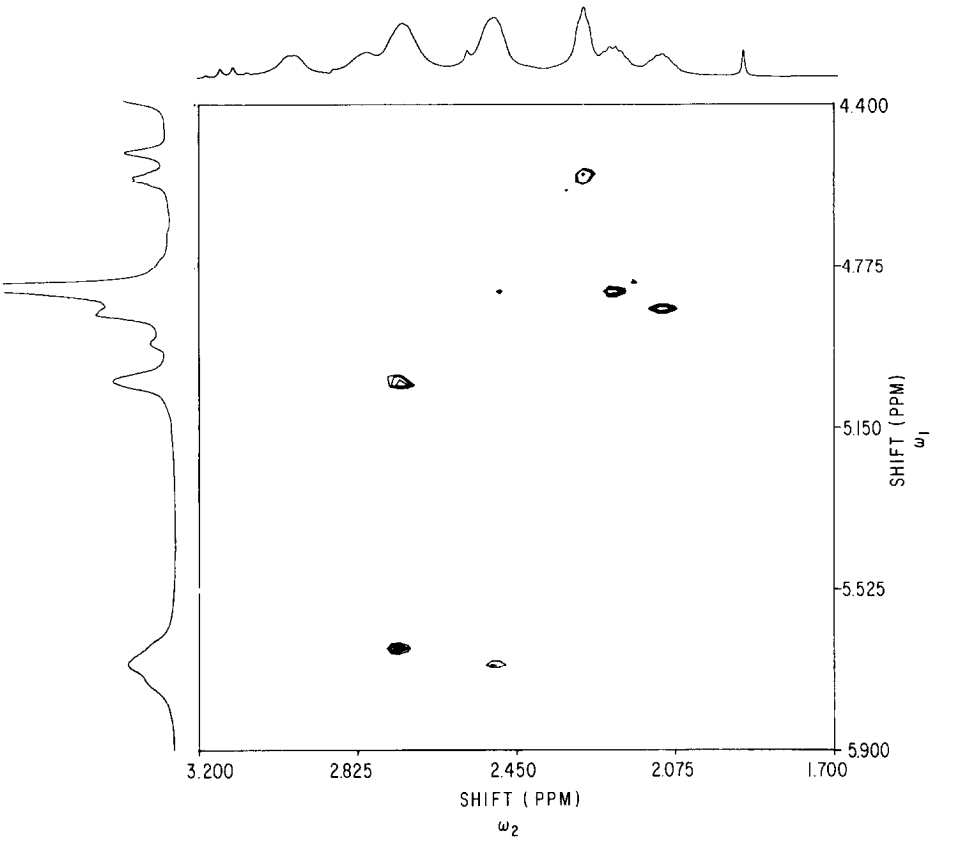

Figure 2. 500 MHz ^{1}H COSY spectrum showing the cross-peaks between the sugar protons H3' and H2'/H2".

Conformations of the individual nucleotides

Two favored nucleotide conformations, in general, are: (i) C2'-endo, χ = 240°-260° and (ii) C3'-endo, χ = 200°-220° (8,9). Conformation (i) is present in the right-handed B-DNA while the conformation (ii) is present in the right-handed A-DNA. These two conformations are distinguishable in terms of the distances between the base proton H8/H6 and sugar protons H2'/H2" and H3'. For (C2'-endo, χ = 240°-260°) conformation, the intranucleotide H8/H6---H3' distance is beyond 4.5 Å while the intranucleotide distances are: H8/H6---H2' = 2.0-2.4Å, H8/H6---H2" = 4Å (8). Thus for such a nucleotidyl geometry a strong NOESY cross-peak is expected between H8/H6 and H2' of its own sugar but no cross-peak between H8/H6 and H3' of its own sugar. For C3'-endo, χ=200°-220° conformation, the intranucleotide H8/H6---H3' distance is 2.7-3.1Å; while the intranucleotide distance H8/H6---H2' is 3.8-4.3Å, hence for such a nucleotidyl geometry, a NOESY cross-peak between H8/H6 and H3' of its own sugar should be present while the cross-peak between H8/H6 and H2' of its own sugar (if present) should be very weak (8).

Figure 4 shows the 1D projections indicating the NOESY cross-peaks between the base-protons H8/H6 and sugar-protons H1', H2'/H2", H3'. Figure 4 reveals that for

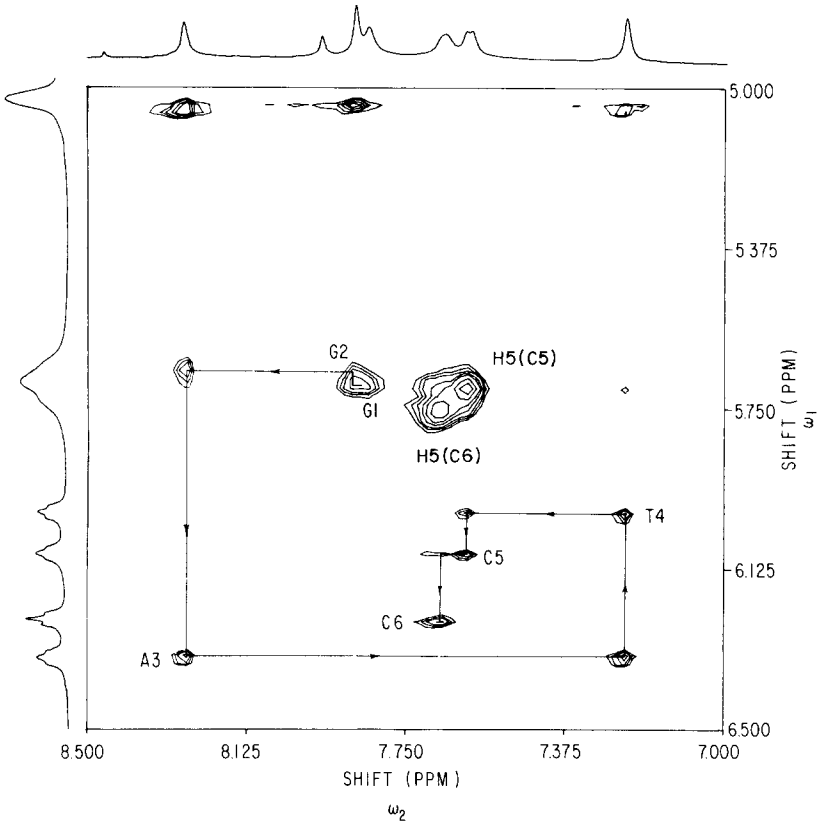

Figure 3. 500 MHz ^{1}H NOESY spectrum showing sequential connectivities between the base protons H8/H6 and the sugar protons H1'. The cross-peaks between H6 and H5 of cytosines are also indicated.

nucleotides G2, A3 and C6, there are distinct NOEs at H3' from H8/H6 which suggest that these nucleotides adopt (C3'-endo, χ=200°-220°) conformation. However, for nucleotides G1 and C5 there are no noticeable NOE at H3's from H8/H6 which indicate that they adopt (C2'-endo, χ=240°-260°) conformation; this point is further confirmed by the fact that there is strong NOE at H2' of G1 and C5 from H8 and H6 respectively (Figure 4). For the nucleotide T4 there is a residual NOE at H3' from H6 which we believe is an effect of spin-diffusion from H2' of T4- a strong site of NOE (Figure 4). In view of the fact that a strong NOE at H2' of T4 from H6 of T4 is only consistent with ((C2'-endo, χ=240°-260°) conformation, we rule out the possibility of (C3'-endo, χ=200°-220°) conformation for T4. Having obtained the individual nucleotide geometries of six residues, molecular models are constructed such that (i) Watson-Crick base pairing with acceptable hydrogen bond geometries is present all through the structure, (ii) there are no steric compressions in the structure and (iii) interproton distances between two neighboring nucleotides are also consistent with the NOE data.

Figure 5 shows skeletal and space filling views of a model projected along the helix-axis. In this model the structural two-fold exactly coincides with the

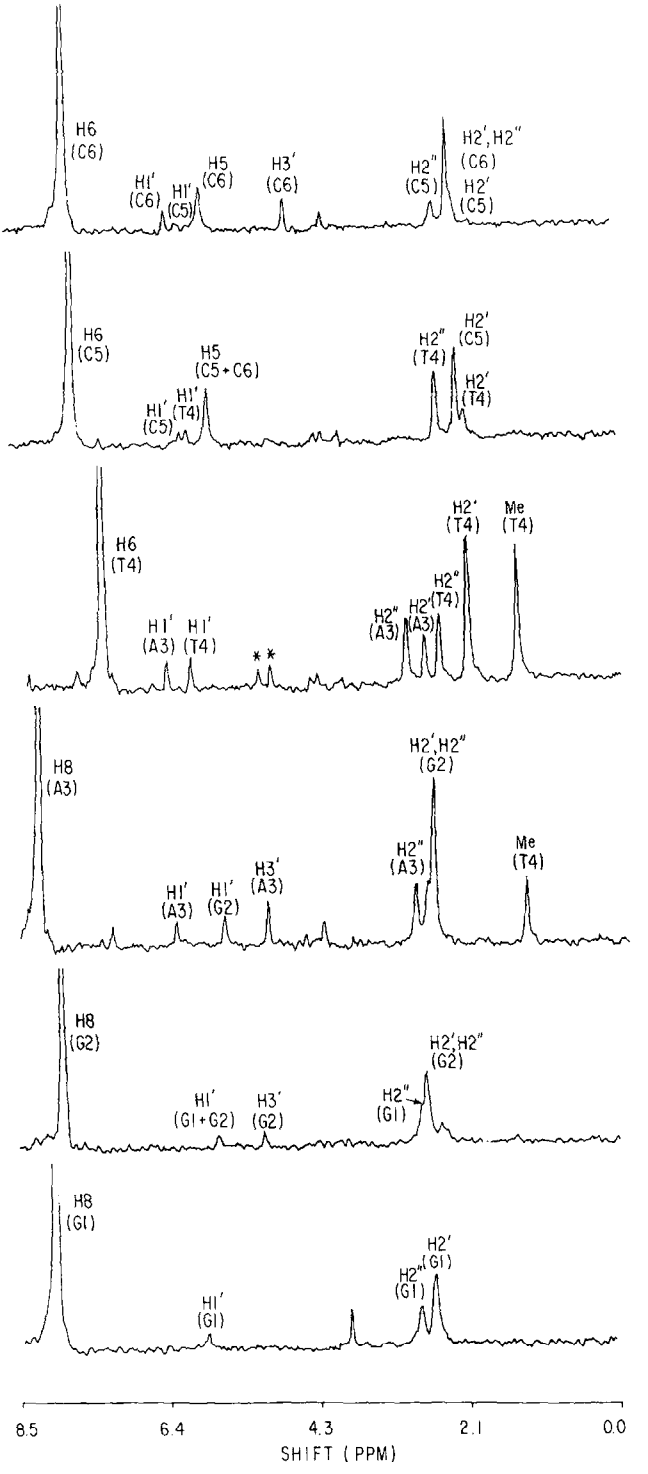

Figure 4. 1D projections from 2D-NOESY data showing the NOEs from the base H8/H6 to sugar protons (The asterisks in T4 projection indicates site of diffusion (see text)).

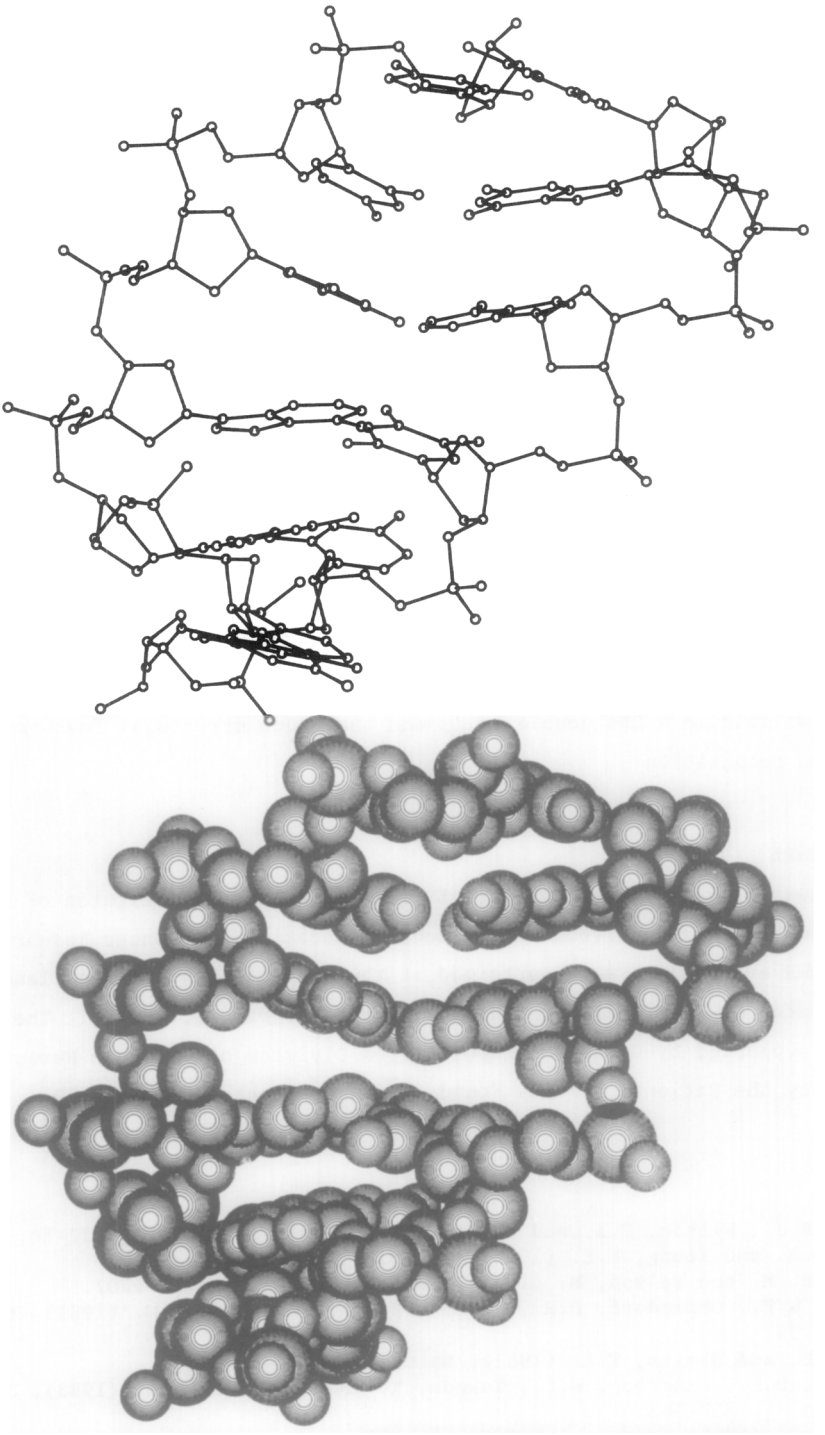

Figure 5. Skeltal and space-filling model of d(GGATCC)$_2$ projected along the helix-axis. Notice that there is a structural two-fold in the model between two A=T pairs, coincident with the two-fold in the base-sequence. The G1-G2 stack-the site of cleavage-is conformationally different from the rest of the molecule. In the space-filling model, for clarity, the sizes used for the van der Waal radii are : 1.2 Å for O, 1.4 Å for N, 1.5 Å for C and 1.6 for P.

two-fold in the base-sequence between two A-T pairs. The important features of the structure are as follows: (i) The G1-G2 stack, the site of cleavage, shows alternation of sugar pucker i.e C2'-endo, C3'endo as in a B-A junction; (ii) the G2-A3 stack forms a mini A-DNA with both sugars in C3'-endo pucker; (iii) the A3-T4 stack, the site of two-fold, exhibits an A-B junction with alternation of sugar pucker as C3'endo, C2'-endo; (iv) the T4-C5 stack adopts a mini B-DNA with C2'-endo pucker for both the sugars; (v) the C5-C6 stack exhibits a B-A junction with C2'-endo, C3'endo sugar puckers. In the resulting DNA double helix, the sugar residues of each Watson-Crick pair, have different puckers viz. in the pair G1≡C6, the sugar attached to G1 has C2'-endo pucker while that of C6 has C3'-endo pucker and similarly for other pairs G2≡C5, A3=T4 etc. Thus, in this model, conformational differences of the nucleotides across a base-pair provide additional structural specificities for the recognition process. The most significant feature of this structure is the fact that the G1-G2, the site of cleavage, is conformationally different from the rest of the molecule. Therefore, it is not unlikely that such a structural peculiarity of the G1-G2 stack might explain why the restriction enzyme Bam H1 very specifically cleaves at G1-G2 site. Thus our studies demonstrate for the first time the presence of distinct conformational microheterogeniety with a structural two fold in a DNA double helix and that such structural features may play a key role in recognition.

Acknowledgement

This research is supported by a grant from the National Institutes of Health (GM29787) and by a contract from the National Foundation for Cancer Research. The high field NMR experiments were performed at the NMR Facility for Biomolecular Research located at the F. Bitter National Magnet Laboratory,. M.I.T. The NMR Facility is supported by Grant No. RR00995 from Division of Research Resources of the NIH and by the National Science Foundation under Contract No. C-670.

References

1. Roberts, R.J., Wilson, G.A. and Young, F.E. (1977), Nature, 265, 82-84.
2. Wilson, G.A. and Young, F.E. (1975), J. Mol. Biol., 97, 123-125.
3. McClelland, M. and Nelson, M. (1985), Nucl. Acid. Res., r201-r207.
4. Anderson, W.F., Ohlendorf, D.H., Kakede, Y. and Matthews, B.W. (1981), Nature, 290, 754-759.
5. McKay, D.B. and Steitz, T.A. (1981), Nature, 290, 744-749.
6. Ohlendorf, D.H., Anderson, W.f., Takeda, Y. and Matthews, B.W. (1983), J. Biomol. Str. & Dyn. 1, 553-563.
7. Sarma et al. (1985), manuscript in preparation.
8. Gupta, G., Sarma, M.H. and Sarma, R.H. (1985), Int. J. Quantum. Chem., in press.
9. Note that IUB-IUPAC $\chi=240^{o}$ is equal to 60^{o} (i.e $180^{o}+60^{o}=240^{o}$) according to the old nomenclature.