Hansen, U. M., & McClure, W. R. (1980) *J. Biol. Chem. 255*, 9564–9570.
Higuchi, R., Krummel, B., & Saiki, R. K. (1988) *Nucleic Acids Res. 16*, 7351–7367.
Hinrichsen, A. I., Ortner, I. M., & Hartmann, G. R. (1985) *FEBS Lett. 193*, 199–202.
Ikeda, R. A., & Richardson, C. C. (1986) *Proc. Natl. Acad. Sci. U.S.A. 83*, 3614–3618.
Josse, J., Kaiser, A. D., & Kornberg, A. (1961) *J. Biol. Chem. 236*, 864–875.
Kahn, J. D., & Hearst, J. E. (1989) *J. Mol. Biol. 205*, 291–314.
Kammerer, W., Deuschle, U., Gentz, R., & Bujard, H. (1986) *EMBO J. 5*, 2995–3000.
Kassavetis, G. A., Zentner, P. G., & Geiduschek, E. P. (1986) *J. Biol. Chem. 261*, 14256–14265.
Kinsella, L., Hsu, C. J., Schulz, W., & Dennis, D. (1982) *Biochemistry 21*, 2719–2723.
Levin, J. R., Krummel, B., & Chamberlin, M. J. (1987) *J. Mol. Biol. 196*, 85–100.
Luse, D. S., & Jacob, G. A. (1987) *J. Biol. Chem. 262*, 14990–14997.
Maitra, U., & Hurwitz, J. (1967) *J. Biol. Chem. 242*, 4897–4907.
Maniatis, T., Fritsch, E., & Sambrook, J. (1982) *Molecular Cloning*, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.
Maxam, A. M., & Gilbert, W. (1977) *Proc. Natl. Acad. Sci. U.S.A. 74*, 560–564.
McClure, W. R. (1985) *Annu. Rev. Biochem. 54*, 171–204.
Mosig, H., Schaffner, A. R., Sieber, H., & Hartmann, G. R. (1985) *Eur. J. Biochem. 149*, 337–343.
Munson, L. M., & Reznikoff, W. S. (1981) *Biochemistry 20*, 2081–2085.
Panka, D., & Dennis, D. (1985) *J. Biol. Chem. 260*, 1427–1431.
Rozovskaya, T. A., Chenchik, A. A., & Beabealashvilli, R. S. (1982) *FEBS Lett. 137*, 100–104.
Ruetsch, N. R., & Dennis, D. (1987) *J. Biol. Chem. 262*, 1674–1679.
Schinkel, A. H., Koerkamp, M. J. A. G., & Tabak, H. F. (1988) *EMBO J. 7*, 3255–3262.
Schulz, W., & Zillig, W. (1981) *Nucleic Acids Res. 99*, 6889–6905.
Shapiro, R., & Klein, R. (1966) *Biochemistry 5*, 2358–2362.
Shimamoto, N., Kamigochi, T., & Utiyama, H. (1986) *J. Biol. Chem. 261*, 11859–11865.
Siebenlist, U. (1979) *Nature 279*, 651–652.
Siebenlist, U., Simpson, R., & Gilbert, W. (1980) *Cell 20*, 269–281.
Stackhouse, T. M., Telesnitsky, A. P., & Meares, C. F. (1989) *Biochemistry* (in press).
Straney, D. C., & Crothers, D. M. (1985) *Cell 43*, 449–459.
Straney, D. C., & Crothers, D. M. (1987) *J. Mol. Biol. 193*, 267–278.
Straney, S. B., & Crothers, D. M. (1987) *Cell 51*, 699–707.
Studier, F. W., & Rosenberg, A. (1981) *J. Mol. Biol. 153*, 503–527.
Telesnitsky, A. P. W., & Chamberlin, M. J. (1989) *J. Mol. Biol. 205*, 315–330.
von Hippel, P. H., Bear, G. G., Morgan, W. D., & McSwiggen, J. A. (1984) *Annu. Rev. Biochem. 53*, 389–446.

Sequence Dependence of DNA Conformational Flexibility[†]

A. Sarai,[‡,§] J. Mazur,[‖] R. Nussinov,[⊥] and R. L. Jernigan*[,‡]

Laboratory of Mathematical Biology, National Cancer Institute, National Institutes of Health, Bethesda, Maryland 20892, Advanced Scientific Computing Laboratory, Program Resource Inc., NCI-Frederick Cancer Research Facility, Frederick, Maryland 21701, and Sackler Institute of Molecular Medicine, Sackler Faculty of Medicine, Tel-Aviv University, Ramat Aviv, 69978 Israel

Received April 10, 1989; Revised Manuscript Received May 31, 1989

ABSTRACT: By using conformational free energy calculations, we have studied the sequence dependence of flexibility and its anisotropy along various conformational variables of DNA base pairs. The results show the AT base step to be very flexible along the twist coordinate. On the other hand, homonucleotide steps, GG(CC) and AA(TT), are among the most rigid sequences. For the roll motion that would correspond to a bend, the TA step is most flexible, while the GG(CC) step is least flexible. The flexibility of roll is quite anisotropic; the ratio of fluctuations toward the major and minor grooves is the largest for the GC step and the smallest for the AA(TT) and CG steps. Propeller twisting of base pairs is quite flexible, especially of A·T base pairs; propeller twist can reach 19° by thermal fluctuation. We discuss the effect of electrostatic parameters, comparison with available experimental results, and biological relevance of these results.

Various experimental studies indicate that DNA is a flexible molecule and that the flexibility is sequence dependent. The conformational flexibility can be manifested in various motions such as bending and twisting. The sequence dependence of the flexibility is observed in various experiments: DNAs with different base composition show different characteristic decay constants in optical measurements (Hogan et al., 1983). Poly(A)·poly(T) and poly(G)·poly(C) do not bind to nucleosome core proteins (Rhodes, 1979; Simpson & Kunzler, 1979; Kunkel & Martinson, 1981; Prunell, 1982). Certain types of sequences exhibit abnormal gel mobility, which has been attributed to the bending of DNA (Koo & Crothers, 1986; Hagerman, 1986). The flexibility of DNA appears to affect

---

[†] Research sponsored, at least in part, by the NCI, DHHS, under Contract NO1-CO-74102 with PRI.

[‡] National Cancer Institute.

[§] Present address: RIKEN Life Science Center, 3-1-1 Koyadai, Tsukuba-city, Ibaraki-ken, 305 Japan.

[‖] NCI-Frederick Cancer Research Facility.

[⊥] Tel-Aviv University.

the way proteins bind to it. The structure of the *Eco*RI–DNA complex showed that the conformation of DNA is significantly altered upon the binding of the enzyme (McClarin et al., 1986). The study of 434 repressor–operator binding has suggested that the binding affinity of 434 repressor appears to be affected by the flexibility of an AT-rich central region of the operator, where there is no direct contact between the protein and DNA bases (Koudelka et al., 1987). The *trp* repressor operator co-crystal shows no direct contacts between amino acid side chains and base pairs (Otwinowski et al., 1988). It was suggested that its sequence specificity may originate indirectly in the specific positioning of the atoms of the phosphate backbone to facilitate protein binding. The flexibility of DNA may also contribute to the cooperative binding of proteins at a distance [see Schleif (1988)]. Analyses of DNA wrapping around histones and CAP have suggested that AA/AT steps tend to be positioned such that their minor groove faces the histone, while the major groove of GG/GC steps tends to face the histone (Drew & Travers, 1985; Satchwell et al., 1986). This binding tendency indicates that the DNA flexibility may have anisotropic characteristics depending on the sequence. This anisotropic nature of flexibility was further supported by the experiment of CAP binding, in which DNA wraps around the protein in such a way that AA/AT steps tend to bend into the minor groove while GG/GC steps tend to bend into the major groove (Gartenberg & Crothers, 1988). These experiments indicate that the sequence-dependent flexibility of DNA may be important in achieving certain specificities in the sequence recognition by proteins.

Despite numerous experimental data, the physical basis of the sequence-dependent flexibility of DNA is not well understood. Several groups have been studying the flexibility of DNA from a theoretical viewpoint (Olson, 1979; Barkley & Zimm, 1979; Levitt, 1984; Kollman et al., 1982; Zhurkin et al., 1982; Ulyanov & Zhurkin, 1984; Calladine & Drew, 1986; Hogan & Austin, 1987). Most of these studies concern the overall or backbone flexibility of DNA, while a few studies have examined the sequence dependence. Calladine and Drew (1986) empirically determined the extent of sequence-dependent bending flexibility, along with the preferential bend angle of DNA. Zhurkin et al. (1982) and Ulyanov and Zhurkin (1984) studied the sequence-dependent flexibility of DNA in some detail by energy calculations. Previously, we have developed a method to calculate thermodynamic quantities such as conformational free energies and fluctuations, by generating a wide range of base pair conformations (Sarai et al., 1988a). The motivation for such analyses arose from the fact that DNA is a polymorphic molecule with many different conformations. Thus, conformational analyses of DNA require that large numbers of conformations be sampled. By using this method, we have previously been able to explain why DNA assumes particular helical structures, why DNA exhibits local structural variation depending on sequence (Sarai et al., 1988b), and how the transition between B- and A-form DNA depends on the sequence (Mazur et al., 1989). One of the important conclusions was that the conformation of DNA and its sequence dependence are mainly determined by the chemical structures of base pairs and their interactions, and that the backbone plays a rather passive role. Thus, we also would expect the sequence dependence of DNA flexibility to be explained by the interactions of base pairs. This assumption is examined and discussed later, with some calculations on the possible effects of backbone on flexibility and anisotropy. Because of the exhaustive sampling, this method is particularly suitable for studying the flexibility. In this paper, we examine the sequence dependence of conformational flexibility, and its anisotropy, in detail. We show that the flexibility is strongly sequence dependent and that the calculated results agree for the most part with available experimental observations.

METHOD OF CALCULATION

The approach here is to consider only base pairs and generate large numbers of conformations to estimate thermodynamic quantities such as conformational free energy and fluctuations. The details of the method of calculation were described previously (Sarai et al., 1988a,b; Mazur et al., 1989). Here, we only describe it in outline. The coordinate system describing the relative orientation of two neighboring base pairs and their transformations is the same as before. The rotational transformations are performed with three Euler angles, and the conventional three angles, "twist", "roll", and "tilt" are related to these through the nine direction cosines of a transformation matrix. Twist is the rotation of a base pair about the axis perpendicular to the base plane. Positive twist corresponds to a right-hand helix. Roll is the rotation of a base pair about its long axis, and the sign of roll is defined as positive if the major groove is compressed. Tilt is the rotation of a base pair about its short axis. Positive tilt means opening the bases on the strand of specified sequence. The energy of base pairs is calculated as a sum overall atom pair contributions in the form of dispersion and short-range repulsion, with 6–12 potentials, and 10–12 potentials for the hydrogen bonds. The parameters for the 6–12 potential are taken from Zhurkin et al. (1981), and for the 10–12 potential from Poltev and Shulyupina (1986). Electrostatic energy is given in the form

$$332\sum_{ij}(1/\epsilon r_{ij})C_iC_j \exp[-\alpha(r_{ij} - r_0)] \text{ (in kcal/mol)}$$

for $r_{ij} > r_0$, and

$$332\sum_{ij}(1/\epsilon r_{ij})C_iC_j$$

for $r_{ij} < r_0$. α is the ionic screening parameter in units of Å^{-1}, and ϵ is the limiting value of dielectric constant at distances less than r_0. C_i and C_j are partial atomic charges [taken from Zhurkin et al. (1981)]. r_0 is the contact distance between interacting charges, and it is qualitatively related to the sum of van der Waals radii. We take r_0 equal to 3.0 Å, which is usually the closest distance between atoms located in the adjacent base planes. The conformational states of base pairs and their energies can be described in the phase space spanned by the three translations, three rotations, and propeller twist.

In the present model, restricted to B-DNA, there are only four independent degrees of freedom, which describe base rotations. There are twist, roll, tilt, and propeller twist, all angular variables. We treat each of these variables separately. For each one of these angular variables, we define the partition function for a canonically distributed ensemble in which the particular angular variable is in adiabatic contact with all angular variables except with itself.

For this purpose, we perform summations over Boltzmann factors for all values of the other three angular variables. We assume 7 values for each one of those angular variables, for a total of 343. These three angular variables, over which the summations are performed, cover a span from −12° to +12° at 4° intervals. From the partition function, the free energy is calculated.

In the present calculation, the angular variable for which the fluctuations and the related anisotropy are to be determined is given a series of consecutive values with a constant increment. In the calculation of adiabatic potential, 81 values are selected around the value corresponding to an energy minimum taken at an 0.5° interval for the twist, and tilt angular variables. For the propeller twist, values are taken at a 1° interval, for a total of 45 values.

The ranges for twist, roll, and tilt are from −20° to +20°, and for propeller twist, from −22° to +22°. We need these ranges to be centered around the point of the minimum value of the free energy. However, we must first determine the approximate value at which the free energy is at its minimum. We therefore performed preliminary "short" calculations, with increments of 2°, over the ranges from −20° to +20°. This assumes that the minimum free energy is located within this range. Except for the twist angular variable for the AT step, this is true. After location of the value of the angular degree of freedom near its free energy minimum, the calculations are repeated, using the same total extent of the angular variable, but the ranges are taken equally on the two sides of the free energy minimum.

The calculations over the sequences of the base pairs are performed iteratively. One base pair at a time is averaged. That is, the average values of the conformational variables are calculated in succession. After the averages of these variables are computed, one base per each iteration step, the entire process is repeated, until a stationary steady-state solution is obtained. In order to take account of longer range interactions and cooperativity among base pairs, we never treat fewer than four base pairs. The calculated free energy, rather than the total energy, is used as the variable in the results, some of which are represented in Figures 1–3.

A typical fluctuation calculation for one degree of freedom for one sequence tetramer requires about 12 cpu minutes on Cray X-MP. The calculations are performed for all possible 16 base steps, regardless of their inherent symmetries e.g., for both AA and TT steps. The complementary sequences yield similar results, and results presented here are the averages over two complementary sequences. The total number of calculations, for one set of electrostatic parameters, is 64. Several different sets of electrostatic parameters have been considered.

Conformational fluctuations of all four angular variables are then calculated from an ensemble of conformations and defined in terms of root mean square deviations from averages:

$$\sigma \equiv \sqrt{\langle (X - \langle X \rangle)^2 \rangle} \qquad (1)$$

for a variable X, where broken brackets denote the Boltzmann average. The anisotropic fluctuations are defined by

$$\sigma_+ \equiv \sqrt{\langle (X - \langle X \rangle)^2 \rangle_+} \qquad (2)$$

$$\sigma_- \equiv \sqrt{\langle (X - \langle X \rangle)^2 \rangle_-} \qquad (3)$$

where $\langle\ \rangle_+$ and $\langle\ \rangle_-$ mean that the averaging is performed only for those values of X larger than or smaller than $\langle X \rangle$, respectively. In both averages, we use the total partition function covering the whole range. Thus, there is a simple relation among these quantities, $\sigma^2 = \sigma_+^2 + \sigma_-^2$. In order to measure the degree of anisotropy of the fluctuations, we introduce the anisotropy ratio

$$A = \sigma_+ / \sigma_- \qquad (4)$$

Although not explicitly given in this paper, the values of σ_+ and σ_- can be readily obtained from the value of σ and the anisotropy A.

Choice of Reference. The reference point, that is, the center of the range of fluctuations for the anisotropy calculations, is located at the average value for the angular variable under consideration. Another possibility would be to locate the reference point for the fluctuation calculations at the free energy minimum. Centering in this latter way, exaggerates the anisotropy. For cases where there is little anisotropy, that is, the anisotropy ratio is near 1, these two reference points will coincide, but for highly anisotropic systems, the anisotropy ratio will be even further away from 1. As a result, the calculated ranges for the anisotropy ratios will have a wider range of possible values if the reference point is the free energy minimum. For example, the range of anisotropy ratios for roll (Table IVA) goes from 0.92 to 2.53, instead of the range reported here of 0.82–1.35.

Table I: Sequence Dependence of Total Fluctuations of Twist, Roll, and Tilt[a]

σ (deg)					
twist		roll		tilt	
AT	>25.0	TA	10.7	AA(TT)	5.1
AC(GT)	14.6	TC(GA)	8.8	TC(GA)	4.8
TA	12.5	CA(TG)	8.6	GG(CC)	4.8
GC	11.1	AA(TT)	8.3	CG	5.0
CA(TG)	10.9	GC	8.0	CA(TG)	4.5
CG	8.9	AC(GT)	7.9	TA	4.3
TC(GA)	8.8	CT(AG)	7.5	CT(AG)	4.1
CT(AG)	8.2	AT	6.2	GC	3.7
AA(TT)	7.6	CG	6.0	AC(GT)	3.7
GG(CC)	7.2	GG(CC)	5.2	AT	3.3
mean	11.5		7.7		4.3

[a] Calculations are for sequence quartets NNMM. $\alpha = 1.20$ and $\epsilon = 2.5$. The range of angles varied is ±20° relative to averages with an increment of 0.5°.

Table II: Sequence Dependence of Total Fluctuations of Twist, Roll, and Tilt[a]

σ (deg)					
twist		roll		tilt	
AT	>25.0	TA	8.9	AA(TT)	4.8
AC(GT)	15.2	AA(TT)	8.2	TC(GA)	4.7
GC	14.3	GA(TC)	8.2	GG(CC)	4.2
CA(TG)	13.1	CG	7.9	CA(GT)	4.2
TA	12.8	CA(TG)	7.7	TA	4.2
CG	11.8	GC	7.4	CT(AG)	4.1
GG(CC)	11.6	AC(GT)	7.2	CG	4.1
TC(GA)	11.5	AG(CT)	6.6	AC(GT)	3.8
CT(AG)	11.4	AT	6.6	GC	3.7
AA(TT)	10.1	GG(CC)	6.2	AT	3.4
mean	13.3		7.5		4.1

[a] See footnote to Table I, except $\alpha = 1.0$ and $\epsilon = 5.0$.

RESULTS

Fluctuations of Twist. The total fluctuations of twist are shown in Tables I and II for two different sets of electrostatic parameters. The results show that the AT step exhibits exceptionally large fluctuations. The free energy potential along the twist coordinate is very flat, and we did not find any significant minima in the twist range between 0 and 80°. Also, the AC(GT) step is considerably more flexible than the mean. On the other hand, GG(CC) is least flexible at the lower dielectric constant, while the AA(TT) step is least flexible in a higher dielectric environment. Note that since this calculation does not include backbones, it is likely that the absolute values of the fluctuations should be smaller in the actual DNA. The fluctuations are reduced by introducing the potential of mean force due to backbones so as to match the observed variation of twist angles (Sarai et al., 1988a). This simply has the form of a harmonic term in the twist deviations. However,

Table III: Anisotropy of Twist Fluctuation[a]

base step	av twist (deg)	A
CG	24.1	0.90
AC(GT)	34.8	0.90
GC	43.7	0.91
CT(AG)	35.9	0.94
AA(TT)	39.4	0.98
TA	21.0	1.03
TC(GA)	34.9	1.05
GG(CC)	41.3	1.09
CA(TG)	26.7	1.15
AT		

[a] Conditions of calculations are the same as in Tables I and II. The free energy curve for the AT step is so flat that it is not possible to find any significant minima in the range considered. $\alpha = 1.2$ and $\epsilon = 2.5$.

Table IV: Anisotropy of Roll Fluctuation[a]

(A) $\alpha = 1.2$, $\epsilon = 2.5$			(B) $\alpha = 1.0$, $\epsilon = 5.0$		
base step	av roll (deg)	A	base step	av roll (deg)	A
CG	12.3	0.82	TA	5.3	1.02
AA(TT)	2.2	0.95	AA(TT)	0.8	1.05
AC(GT)	0.6	1.03	AT	−1.9	1.06
TA	5.1	1.03	AC(GT)	0.0	1.10
AT	−3.0	1.09	CG	7.1	1.13
CA(TG)	7.7	1.16	CA(TG)	5.7	1.13
CT(AG)	0.1	1.17	CT(AG)	1.3	1.15
TC(GA)	−1.5	1.20	TC(GA)	0.5	1.20
GG(CC)	−0.4	1.22	GG(CC)	1.1	1.21
GC	−5.0	1.35	GC	−0.9	1.34

[a] Conditions of calculations are the same as in Tables I and II.

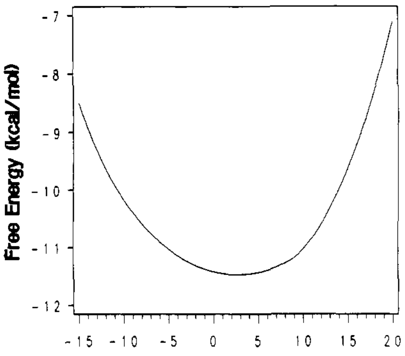

FIGURE 1: Dependence of free energy on roll for the base step AA. Free energy is calculated for the quartet sequence AAAA by changing roll at the central base step and by allowing all the other degrees of freedom to fluctuate for each fixed roll value. Conformations are averaged one base step at a time iteratively as described in text. $\alpha = 1.2$ and $\epsilon = 2.5$.

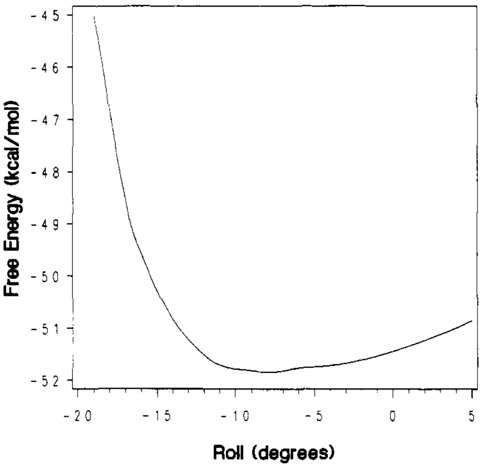

FIGURE 2: Dependence of free energy on roll for the base step GC. Other conditions are the same as in Figure 1.

the relative sequence dependence is only slightly affected by this additional term. Its affect is discussed further under Discussion.

Fluctuations in Roll. The average fluctuation in roll is about 8°, as shown in Tables I and II. The conformational free energy usually rises rapidly beyond 10° of roll because of the van der Waals repulsions between adjacent base pairs. The TA step is the most flexible of all along the roll coordinate. On the other hand, GG(CC) is the least flexible step for both dielectric constants.

Fluctuations in Tilt. These fluctuations are much smaller than those in roll. This is because tilt causes larger steric clashes between base pairs than roll. In fact, we did find that large values of tilt require increases in the axial rise to avoid steric clashes. The sequence-dependent variation of tilt fluctuation is also small compared to that of roll. There is a strong tendency for pyrimidine–purine steps to be more flexible than purine–pyrimidine steps.

Anisotropies of Twist Fluctuation. As shown in Table III, the ratio of fluctuations for positive to negative directions relative to the average twist is the smallest for CG and AC(GT). Thus, these steps are more susceptible to undertwist than overtwist of helix. The anisotropy ratio is the largest for the CA(TG) step. The AT step is exceptional in that it has a very flat energy potential. Thus, this step is strongly susceptible to either undertwist or overtwist. The AC(GT) step is also very susceptible to undertwist, with the largest value of σ_-.

Anisotropies of Roll Fluctuation. In general, there is more sequence-dependent variation in the anisotropy of roll fluctuations than for twist fluctuations (see Tables IVA and IVB). The sequences with A/T composition have small anisotropy ratios, while GC and GG have the largest ratios. In other words, A/T is more susceptible to roll into minor groove, with GC and GG into the major groove. The dependences of the conformational free energies on roll for the AA(TT) and the GC steps are shown in Figures 1 and 2, in which the asymmetry of the curves corresponds to the anisotropy of roll fluctuation. It should be noted that this tendency in the anisotropy does not correlate with the sequence dependence in the average roll. For example, CG has a smaller anisotropy ratio, but this step is already strongly rolled into the major groove (cf. the average roll in Table IVA).

The electrostatic parameters affect the fluctuations and anisotropies to some extent. This effect is noteworthy for the roll fluctuations. Tables IVA and IVB are for different values of the electrostatic parameters, which affect the strengths of electrostatic interactions between charges. In general, for weaker electrostatic interactions, variations in both the total fluctuation and its anisotropy ratio are smaller. The largest change in relative position in the table occurs in CG, as indicated by an increase in the anisotropic ratio at a low dielectric constant. In Table IVB, AA(TT), AT, and TA are clustered at small values of the anisotropy ratio, whereas GC, GG(CC), and CG are clustered at the other end, for the higher dielectric constant.

Anisotropies of Tilt Fluctuation. As shown in Table V, there is a very small variation in the anisotropy ratio. Because of this, it is difficult to derive consistent sequence dependence of the anisotropy over a range of electrostatic parameters. There are also very small variations of average tilt values and total tilt fluctuations.

Fluctuations and Anisotropies of the Propeller Twists. Table VI shows the fluctuation and anisotropy of propeller twist for AA(TT) and GG(CC) steps with two sets of elec-

Table V: Anisotropy of Tilt Fluctuation[a]

base step	av tilt (deg)	A
AA(TT)	2.7	0.90
TA	0.3	0.91
CA(TG)	1.5	0.92
GC	2.2	0.93
AT	0.6	0.96
TC(GA)	−0.4	0.97
GG(CC)	0.8	0.97
AC(GT)	1.4	0.97
CT(AG)	−1.8	1.03
CG	−3.3	1.14

[a] Conditions of calculations are the same as in Tables I and II. The average over base steps that are symmetry-related is based on an assumption that the average tilt in the complementary base is of opposite sign. Therefore, the average tilts for base pairs enclosed in parentheses have their signs reversed. The anisotropy ratios, A, for these bases are the inverse of the ones given in the last column.

Table VI: Fluctuations and Anisotropy of Propeller Twist[a]

base step	av propeller twist (deg)	σ (deg)	A
	$\alpha = 1.2, \epsilon = 2.5$		
AA(TT)	9.4	10.7	1.08
	9.9[b]	10.8	1.02
GG(CC)	12.7	7.9	0.73
	$\alpha = 1.0, \epsilon = 5.0$		
AA(TT)	9.7	10.0	1.11
	10.4[b]	10.4	1.03
GG(CC)	6.7	7.6	1.09

[a] Calculations are for sequence quartets NNNN. The range of angles is ±22° relative to averages with an increment of 1.0°. [b] This calculation permits the bifurcated hydrogen-bond interactions between N6 (A) and O4 (T) of adjacent base pairs.

trostatic parameters. The calculation shows that the free energy curves along propeller twist have minima between 7° and 16°. The average values of propeller twist are about 10° for AA(TT) and 7° for GG(CC) in the high dielectric environment. The lower dielectric constant increases the propeller twist of GG to nearly 13°. The total fluctuations are about 10° for the A·T base pair and 8° for the G·C pair, regardless of electrostatic environment. This difference is probably due to the smaller number of hydrogen bonds for A·T, that permits it to be more flexible. Figure 3 shows the free energy changes with propeller twist for the AA step. The larger propeller twist of G·C base pairs in the low dielectric environment is partly offset by larger fluctuation in the negative direction, which results in a smaller anisotropy ratio. Introduction of the possibility of bifurcated hydrogen bonding between N6 of A and O4 of T of adjacent base pairs slightly increases the propeller twist and lowers the anisotropy ratio.

At this point, we want to compare the present method of computing the fluctuations of a single angular variable with the one reported previously (Sarai et al., 1988a). There, we did not deal separately with each of the angular variables. Instead, the entire canonical ensemble consisted of seven values for each one of these four angular variables which define the four rotational degrees of freedom of a base pair presently under consideration. The partition function of the entire system was calculated. In that ensemble, the average values for each of these four angles of rotation, their average second moments, and their fluctuations were computed.

We did repeat some of the calculations for this ensemble, using the same ranges for roll, tilt, twist, and propeller twist employed in the current work. That is, the ranges for roll and tilt are −20°, +20°, while for the twist and propeller twist, the ranges are −12°, +12°. Since these calculations were primarily concerned with the fluctuations and anisotropies of

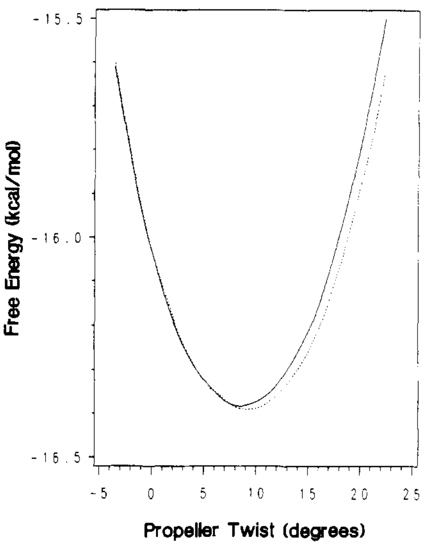

FIGURE 3: Dependence of free energy on propeller twist for the base step AA. The dotted line includes the effect of permitting the bifurcated hydrogen-bond interactions between N6 of A and O4 of T of adjacent base pairs. $\alpha = 1.2$ and $\epsilon = 2.5$.

roll and tilt, we took intervals of 1° for these variables for a total of 41 points for each of them. For twist and propeller twist, we took seven values, at 4° intervals. Anisotropies of both positive and negative roll and tilt were determined by calculating separately the partition function, second moments, and the fluctuations for positive and negative rolls and tilts. As mentioned above, fluctuations become smaller and in closer agreement with the fluctuations reported in Sarai et al. (1988a). For $\alpha = 1.2$ and $\epsilon = 2.5$, the mean fluctuation for roll is now 5.29, and for tilt, it is 2.95 as compared with the corresponding values of 7.72 and 4.30 from Table I. Similarly, for electrostatic parameters $\alpha = 1.0$, $\epsilon = 5.0$, the fluctuations become 5.18 and 2.60 for roll and tilt, respectively, compared with their values of 7.47 and 4.04 in Table II. The larger fluctuations reported in the present work depend strongly on the method of constructing the ensemble. In the present work, the individual angular variable being used for fluctuation calculations is in thermal contact with all other angular variables except with itself. In the previous calculations, averages were performed simultaneously over all angular variables in the ensemble. Generally, these two different canonical ensembles lead to consistent results in terms of relative flexibilities and anisotropies of the base steps and their dependence on the electrostatic parameters of the system.

DISCUSSION

A large body of experimental data indicates that the flexibility of DNA is sequence dependent and may affect the function of DNA. In this paper, we have examined the sequence dependence of DNA flexibility by conformational free energy analysis. There have been several conformational studies of DNA (Levitt, 1978; Kollman et al., 1982; Zhurkin et al., 1982; Ulyanov & Zhurkin, 1984; Tung & Harvey, 1986; Srinivasan et al., 1987), but no extensive calculations of sequence-dependent fluctuations of DNA molecule have been performed. It seems that DNA manifests many conformations with similar energies. The present method enables us to calculate thermal fluctuations as well as average conformations of DNA, and their sequence dependences. The results show that there is a strong sequence dependence of DNA flexibility, and that the flexibility is anisotropic.

The calculation of fluctuations along the twist coordinate shows that the AT step is extremely flexible. In our previous study of B-DNA, we showed that base pairs by themselves have a strong tendency to stack in a helical form (Sarai et al., 1988a,b). However, AT is an exception; its twist is extremely unrestricted by interactions among base pairs. The AT step can take any twist angle insofar as the interactions of base pairs are concerned. Although backbones restrict the range of fluctuations, the multiple rotational degrees of freedom of the backbone will allow base pairs to move within a wide range. For confirmation of this, crystallographic study of the binding complex between *Eco*RI and its recognition sequence containing the AATT segment has shown that the AT step is strongly undertwisted by about 25° (McClarin et al., 1986). Such a conformation change is likely to be important in facilitating specific contact between base pairs and amino acids. It is likely that the extreme flexibility of this step may be used by this enzyme as one of the recognition features.

Hogan et al. (1983) measured triplet anisotropy decay of DNA for some sequences and found that poly(dG)·poly(dC) has the largest torsional stiffness, whereas poly(dA-dC)·poly(dT-dG) and poly(dA)·poly(dT) are 40 and 20 times less stiff than poly(dG)·poly(dC), respectively. In our calculation, the GG(CC) step is the stiffest for low dielectric constant, whereas the AA(TT) step is the stiffest for high dielectric environment. The experimental results agree with the calculated results at low dielectric constant. Although poly(dA)·poly(dT) is much more flexible than poly(dG)·poly(dC) in the above experiment, it is stiffer than random-sequence DNA. Our results show that AC(GT) is the second most flexible step, which is consistent with the above experimental observation. The fact that poly(G)·poly(C) and poly(A)·poly(T) do not bind to nucleosome core proteins (Rhodes, 1979; Simpson & Kunzler, 1979; Kunkel & Martinson, 1981; Prunell, 1982) may have an explanation in the stiffness of these homopolymers.

The experiments of Hogan et al. (1983) also showed that the bending stiffness of poly(dG)·poly(dC) is at least twice that of poly(dA·dC)·poly(dG-dT) or random-sequence DNA and is at least 3-fold greater than that of poly(dA)·poly(dT). This observation is in agreement with our results, in which GG(CC) is among the least flexible steps, the flexibility of the AC(GT) or CA(TG) step is about the mean value, and the AA(TT) step is somewhat more flexible than the mean. Chen et al. (1985) observed that poly(dA-dT)·poly(dA-dT) is about twice as flexible as random-sequence DNA. In our calculation, TA is the most flexible of all the steps in its roll, whereas AT is less flexible than average. The alternating AT/TA steps would exhibit greater flexible than the mean value.

Our conformational free energy calculations show that the fluctuations are anisotropic (Table IV). The extent of the anisotropy is especially large for roll. In general, GC and GG(CC) steps have a larger anisotropy ratio, which makes these steps tend to roll (bend) into the major groove of DNA. On the other hand, AA(TT), AT, and TA have smaller anisotropy ratios. Zhurkin et al. (1988) noted similar tendencies in their energy calculations. This ranking of anisotropy is consistent with the experimental observations of CAP binding to DNA (Gartenberg & Crothers, 1988), in which DNA wraps around the protein in such a way that AA/AT steps tend to bend into minor groove while GG/GC steps tend to bend into major groove. The experimental dinucleotide ranking of anisotropy (Gartenberg & Crothers, 1988) is very similar to the results of calculations presented in Table IV. Also, analysis of DNA wrapping around histones has suggested that the AA/AT steps tend to be positioned such that their minor groove faces the histone, while the major groove of GG/GC steps tends to face the protein (Drew & Travers, 1985; Satchwell et al., 1986).

The origin of the sequence-dependent anisotropic fluctuation of roll appears to have a simple explanation in the chemical structure of base pairs and their interactions. The absence of an NH_2 group at C2 in the minor groove of A·T base pairs makes dinucleotide combinations of A and T more flexible into the minor groove than combinations of G and C. In order to test this, we performed similar calculations by substituting inosine (I) for guanine. Sequences with I·C base pairs appear to have flexibilities similar to A·T base pairs, indicating that the NH_2 group is probably the main cause of the anisotropy difference.

It has been known that pyrimidine (Y)–purine (R) steps tend to have positive roll (bend into the major groove), whereas R–Y steps have negative roll (bend into minor groove) (Calladine, 1982; Dickerson, 1983). This tendency has also been found in conformational calculations (Ulyanov & Zhurkin, 1984; Tung & Harvey, 1986; Sarai et al., 1988a,b; see also the calculated average roll in Table IV). The present results show that there is no apparent correlation between this equilibrium directional bias and the anisotropy of roll fluctuation. The binding of proteins to DNA can induce conformation changes of DNA. The amount of the induced conformational changes, which depends on the flexibility of DNA, is likely to be comparable or even larger than the equilibrium conformational variation reported here. Since the fluctuations obtained in the present calculations agree with the experimentally observed anisotropy, then the inducible conformation change or the flexibility could be more important than the equilibrium variation in providing sequence dependence to protein–DNA binding.

The anisotropic fluctuations in twist and tilt are also sequence dependent, though to a lesser extent than roll. The anisotropy in the twist fluctuation might affect the extent of underwinding or overwinding of DNA when it is bound to proteins or when DNA is under torsional stress. The anisotropy of the tilt fluctuation would affect, like that of the roll, the direction of DNA bends.

The present calculations show that the free energy curves along propeller twist have minima at positive values between 7° and 16°. The stabilization of positive propeller twist is due to the improved stacking interactions between adjacent bases in the same strand (Levitt, 1978). The flexibility of propeller twist is quite large; in particular, A·T base pairs have up to 11° of thermal fluctuation, mainly because of the smaller number of hydrogen bonds than G·C base pairs. This enables an A·T base pair to fluctuate from 3° to 19° of propeller twist at room temperature. Crystallographic studies of A tract segments of DNA have suggested that this segment of sequence exhibits a very large propeller twist, larger than 20°, and N6 of A and O4 of T can form a bifurcated hydrogen bond between adjacent base pairs (Nelson et al., 1987; Coll et al., 1987; Yoon et al., 1988). The average propeller twist value in the present calculation is around 10°. The introduction of the inter-base pair hydrogen bond between adjacent base pairs, that has been neglected in most previous calculations, however, appears to increase propeller twist only slightly (Table VI and Figure 3). Of course, the extent of this is manifested through the energetics of propeller twist that are sensitive to the hydrogen-bond parameters in the calculation. Furthermore, there may be other factors, such as hydration free energy, which can affect the propeller twist. In fact, we have been investigating the effects of hydration on conformations of DNA and find that hydration tends to stabilize larger propeller twist for the AA(TT) step (Raghunathan et al., 1989). This, together with the large flexibility in propeller twist, would indicate the large propeller twist of A tracts, as reported in experiments, to be energetically feasible.

The present calculations show that electrostatic parameters affect the amount of fluctuation and its anisotropy. This can be seen clearly in Tables I, II, and IV, in which larger electrostatic screening decreases the extent of sequence-dependent variation of fluctuation and anisotropy. It also affects some of the sequence dependence of total fluctuations and anisotropy ranking [e.g., AA(TT) and GG(CC) in twist fluctuation and CG in roll anisotropy]. This result suggests that the balance between electrostatic interactions and van der Waals interactions may be important for the sequence-dependent fluctuations and anisotropy. It would appear to be particularly important to assess the electrostatic effects by measuring the variation of DNA flexibility and anisotropy in various experimental electrostatic environments.

In this study, we have not considered backbone atoms explicitly. Although the several rotational degrees of freedom permit the backbone to be quite flexible, they can still affect the base pair fluctuations. The backbone is most likely to affect the limits to the twist fluctuations since the tilt fluctuations are already relatively restricted. Such an effect could be taken into account partially by introducing a potential of mean force empirically determined from observed conformational variations in crystal structures (Sarai et al., 1988a). Backbone segments can also introduce directional bias in bending. Ulyanov and Zhurkin (1984) found from energy calculations that DNA bends in groove directions (rolls) more readily than in a perpendicular direction (tilt). Thus, backbone may further enhance the stiffness of tilt over roll.

From some of our unpublished calculations in which backbone atoms are included and the energy is minimized, we observed that the twist coordinate is considerably more restricted than the actual numbers reported here. This observation would indirectly support the inclusion of the potential of mean force in twist that was mentioned before which is based on experimentally measured variations in the twist coordinate in the crystal of the *Eco*RI dodecamer (Sarai et al., 1988a). To explore the effects of backbone on the twist flexibility, we have performed additional calculations with this harmonic term and the electrostatic parameters used for Tables I and III. The average twist was found to be 5.4° instead of the value of 11.5° in Table I. Its range for different base steps was only from 4.7° to 6.6°. However, the dependence of the order of the fluctuations on the base step is almost the same as in the present tables. That is, the most flexible steps are AT and AC(GT). The least flexible steps are the GG(CC) and the AA steps. Also, this introduction of backbone does not seem to affect very much the anisotropies of the twist fluctuations. The anisotropy ratios range from the lowest values of 0.89 to the highest value of 1.12. The anisotropy ratio is the largest for the CA(TG) and TC(GA) steps. It is the smallest for the AC(GT), CG, and TA steps.

These additional calculations demonstrate that including backbone atoms can reduce the extent of fluctuations, especially for twist. However, the sequence dependence of fluctuations and anisotropy studied here seems unlikely to be affected in major ways by the neglect of backbone.

In a series of conformational analyses of DNA, we have studied the sequence dependence of DNA conformations and the transitions between different conformations. We have successfully explained, in terms of the chemical structure of base pairs and their interactions, why DNA assumes a particular helical structure, why DNA exhibits local sequence-dependent conformational variations (Sarai et al., 1988b). We have also shown how the sequence affects the transition between B- and A-form DNA (Mazur et al., 1989). In this paper, we have applied the same approach to examine elastic properties of DNA. The calculations show that the flexibility of DNA is strongly sequence dependent and anisotropic. There is increasing evidence that suggests the importance of DNA flexibility in providing specificity for protein–DNA binding. The agreement between all of these calculations and various experimental data is quite encouraging. We believe that such conformational treatments can provide insights into the mechanisms of sequence dependence of various conformational properties of DNA.

ACKNOWLEDGMENTS

We thank Dr. Jacob V. Maizel, Jr., for his constant interest in this work. Computations were performed on the Cray X-MP at the ASCL.

REFERENCES

Barkley, M. D., & Zimm, B. H. (1979) *J. Chem. Phys. 70*, 2991–3007.

Calladine, C. R. (1982) *J. Mol. Biol. 161*, 343–352.

Calladine, C. R., & Drew, H. R. (1986) *J. Mol. Biol. 192*, 907–918.

Chen, H. H., Rau, D. C., & Charney, E. (1985) *J. Biomol. Struct. Dyn. 2*, 709–719.

Coll, M., Frederick, C. A., Wang, A. H.-J., & Rich, A. (1987) *Proc. Natl. Acad. Sci. U.S.A. 84*, 8385–8389.

Dickerson, R. E. (1983) *J. Mol. Biol. 166*, 419–441.

Drew, H. R., & Travers, A. A. (1985) *J. Mol. Biol. 186*, 773–790.

Gartenberg, M. R., & Crothers, D. M. (1988) *Nature 333*, 824–829.

Hagerman, P. J. (1986) *Nature 321*, 449–450.

Hogan, M. E., & Austin, R. H. (1987) *Nature 329*, 263–266.

Hogan, M., LeGrange, J., & Austin, B. (1983) *Nature 304*, 752–754.

Kollman, P., Keepers, J. W., & Weiner, P. (1982) *Biopolymers 21*, 2345–2376.

Koo, H. S., Wu, H.-M., & Crothers, D. M. (1986) *Nature 320*, 502–506.

Koudelka, G. B., Harrison, S. C., & Ptashne, M. (1987) *Nature 326*, 886–888.

Kunkel, G. R., & Martinson, H. G. (1981) *Nucleic Acids Res. 9*, 6869–6888.

Levitt, M. (1978) *Proc. Natl. Acad. Sci. U.S.A. 75*, 640–644.

Levitt, M. (1984) in *Specificity in Biological Interactions* (Chagas, C., & Pullman, B., Eds.) pp 43–57, D. Reidel, Boston.

Mazur, J., Sarai, A., & Jernigan, R. L. (1989) *Biopolymers* (in press).

McClarin, J. A., Frederick, C. A., Wang, B.-C., Green, P., Boyer, H. W., Grable, J., & Rosenberg, J. M. (1986) *Science 234*, 1526–1541.

Nelson, H. C. M., Finch, J. T., Luisi, B. F., & Klug, A. (1987) *Nature 330*, 221–226.

Olson, W. K. (1979) *Biopolymers 18*, 1235–1260.

Otwinowski, Z., Schevitz, R. W., Zhang, R.-G., Lawson, C. L., Joachimiak, A., Marmorstein, R. Q., Luisi, B. F., & Sigler, P. B. (1988) *Nature 335*, 321–329.

Poltev, V. I., & Shulyupina, N. V. (1986) *J. Biomol. Struct. Dyn. 3*, 739–1102.

Prunell, A. (1982) *EMBO J. 2*, 173–179.

Raghunathan, R., Sarai, A., Ting, K.-L., & Jernigan, R. L. (1989) *Biochemistry* (submitted for publication).

Rhodes, D. (1979) *Nucleic Acids Res. 6*, 1805–1816.

Sarai, A., Mazur, J., Nussinov, R., & Jernigan, R. L. (1988a) in *Structure and Expression* (Olson, W. K., Sarma, M. H., Sarma, R., & Sundaralingam, M., Eds.) Vol. 3, pp 213–223, Alan Liss, New York.

Sarai, A., Mazur, J., Nussinov, R., & Jernigan, R. L. (1988b) *Biochemistry 27*, 8498–8502.

Satchwell, S., Drew, H. R., & Travers, A. A. (1986) *J. Mol. Biol. 191*, 659–675.

Schleif, R. (1988) *Science 240*, 127–128.

Simpson, R. T., & Kunzler, P. (1979) *Nucleic Acids Res. 6*, 1387–1415.

Srinivasan, A. R., Torres, R., Clark, W., & Olson, W. K. (1987) *J. Biomol. Struct. Dyn. 5*, 459–496.

Tung, C.-S., & Harvey, S. C. (1986) *J. Biol. Chem. 261*, 3700–3709.

Ulyanov, N. B., & Zhurkin, V. B. (1984) *J. Biomol. Struct. Dyn. 2*, 361–385.

Yoon, C., Privê, G. G., Goodsell, D. S., & Dickerson, R. E. (1988) *Proc. Natl. Acad. Sci. U.S.A. 85*, 6332–6336.

Zhurkin, V. B., Poltev, V. I., & Florentiev, V. I. (1981) *Mol. Biol. (Moscow) 14*, 887–895.

Zhurkin, V. B., Lysov, Y. P., Florentiev, V. I., & Ivanov, V. I. (1982) *Nucleic Acids Res. 10*, 1181–1830.

Zhurkin, V. B., Ulyanov, N. B., & Ivanov, V. I. (1988) in *Structure and Expression* (Olson, W. K., Sarma, M. H., Sarma, R., & Sundaralingam, M., Eds.) Vol. 3, pp 169–190, Alan Liss, New York.

Binding of a Hoechst Dye to d(CGCGATATCGCG) and Its Influence on the Conformation of the DNA Fragment[†]

Maria A. A. F. de C. T. Carrondo,[‡,§] Miquel Coll,[‡] Joan Aymami,[‡] Andrew H.-J. Wang,[‡,∥] Gijs A. van der Marel,[⊥] Jacques H. van Boom,[⊥] and Alexander Rich*[,‡]

Department of Biology, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, and Gorlaeus Laboratory, Leiden State University, 2300 RA Leiden, The Netherlands

Received January 25, 1989

ABSTRACT: Hoechst dye 33258 is a planar drug molecule that binds to the minor groove of DNA, especially where there are a number of A·T base pairs. We have solved the structure of the Hoechst dye bound to the DNA dodecamer d(CGCGATATCGCG) at 2.3 Å. This structure is compared to that of the same dodecamer with the minor-groove-binding drug netropsin bound to it, as well as to structures that have been solved for this Hoechst dye bound to a DNA dodecamer containing the central four base pairs with the sequence AATT. We find that the position of the Hoechst drug in this dodecamer is quite different from that found in the other dodecamer since it has an opposite orientation compared to the other two structures. The drug covers three of the four A·T base pairs and extends its piperazine ring to the first G·C base pair adjacent to the alternating AT segment. Furthermore, the drug binding has modified the structure of the DNA dodecamer. Other DNA dodecamers with alternating AT sequences show an alternation in the size of the helical twist between the ApT step (small twist) and the TpA step (large twist). In this structure the alternation is reversed with larger twists in the ApT steps than in the TpA step. In addition, there is a rotation of one of the thymine bases in the DNA dodecamer that is associated with hydrogen bonding to the Hoechst drug. This structure illustrates the considerable plasticity found in the DNA molecule when it binds to different planar molecules inserted into the minor groove.

There are a number of planar drug molecules that bind to DNA in the minor groove (Zimmer & Wahnert, 1986). These molecules, which include netropsin, distamycin, and the Hoechst dye 33258, all have the property of being formed of repetitions of a structural motif that produce an arc conformation of the planar molecule. This arc matches somewhat the turn of the DNA double helix, and these drugs fit snugly into the double helix in regions in which there are predominantly no G·C base pairs. The N2 amino group of guanine effectively blocks the binding of these drugs both by its presence and by its effect in widening the minor groove of DNA. In the absence of an N2 amino group on guanine, the minor groove has a narrow form that permits these planar molecules to bind using both hydrogen-bonding and van der Waals interactions.

Hoechst dye 33258 (Figure 1) is widely used as a generalized DNA stain, but it is also active an an antihelminthic agent (Lammler et al., 1971). This drug is known to have both a low-level tight-binding mode and a high-level low-affinity binding that are differentiated by their fluorescent characteristics. The interaction of this drug with DNA has been studied extensively. Early work has shown that the Hoechst dye binds preferentially to A·T-rich DNA (Latt & Wohlleb,

---

[†] This work was supported by grants from NIH, NSF, NASA, and ONR to A.R. M.A.A.F.C.T.C. acknowledges support from the Junta Nacional de Investigação Científica e Tecnológica. J.v.B. was supported by the Netherlands Organization for the Advancement of Pure Research (ZWO). M.C. was a Fulbright Fellow. J.A. was a NATO Fellow.

[*] To whom correspondence should be addressed.

[‡] Massachusetts Institute of Technology.

[§] Present address: Centro de Química Estructural, Instituto Superior Técnico, 1096 Lisboa Codex, Portugal.

[∥] Present address: Department of Physiology and Biophysics, University of Illinois at Urbana–Champaign, Urbana, IL 61801.

[⊥] Leiden State University.

0006-2960/89/0428-7849$01.50/0