Constructing Optimal Backbone Segments for Joining Fixed DNA Base Pairs

J. Mazur,* R. L. Jernigan,# and A. Sarai§

*Frederick Biomedical Super Computing Laboratory, SAIC, NCI-FCRDC, Frederick, Maryland 21701 USA; #Laboratory of Mathematical Biology, DBS, National Cancer Institute, National Institutes of Health, Bethesda, Maryland 20892-5677 USA; and §RIKEN Life Science Center, Tsukuba, Ibaraki 305, Japan

ABSTRACT A method is presented to link a sequence of space-fixed base pairs by the sugar-phosphate segments of single nucleotides and to evaluate the effects in the backbone caused by this positioning of the bases. The entire computational unit comprises several nucleotides that are energy-minimized, subject to constraints imposed by the sugar-phosphate backbone segments being anchored to space-fixed base pairs. The minimization schemes are based on two stages, a conjugate gradient method followed by a Newton-Raphson algorithm. Because our purpose is to examine the response, or relaxation, of an artificially stressed backbone, it is essential to be able to obtain, as closely as possible, a lowest minimum energy conformation of the backbone segment in conformational space. For this purpose, an algorithm is developed that leads to the generation of an assembly of many local energy minima. From these sets of local minima, one conformation corresponding to the one with the lowest minimum is then selected and designated to represent the backbone segment at its minimum. The effective electrostatic potential of mean force is expressed in terms of adjustable parameters that incorporate solvent screening action in the Coulombic interactions between charged backbone atoms; these parameters are adjusted to obtain the best fit of the nearest-neighbor phosphorous atoms in an x-ray structure.

INTRODUCTION

There are many molecular computations in which some constraints are known from either experiment or calculation. The best known example is the use of nuclear magnetic resonance nuclear Overhauser effect distance constraints, which have been so successfully combined with distance geometry calculations (Havel et al., 1983a,b) to produce sets of viable molecular conformations (Moore et al., 1981). Another less constrained case is that of peptides, where titration experiments indicate the presence of a salt bridge; various calculations can be performed within the context of such constraints (Jacchieri and Jernigan, 1992). Another category of problems with constraints is entirely computational, where the problem is simply too large for a complete and rigorous evaluation, and the constraints are employed for expedient practical reasons, i.e., to reduce the size of calculations. The constraints themselves could have come from some limited calculations that indicated some conformational preferences or ranges of preferences. Two examples of large problems and appropriate simplifying limitations that come to mind are DNA double helices, where the range of conformational variables describing the excursions of neighboring base pairs with respect to one another is limited to various double-helical forms, and hydrophobic protein cores, where hydrophobic contacts are preferred (Yue and Dill, 1995). In the first case of the DNA double helix that we are considering in this paper, the base pairs are positioned explicitly to conform to double helices, but in various ways to sample the feasible range of positions, and the connecting backbone segments are then minimized. This is consistent with a description of DNA conformations in which the most important conformation-determining factor of the relatively small sequence dependence of the double-helical forms resides in the base-base interactions (Sarai et al., 1988). The short sugar-phosphate segments are relatively passive players, insofar as base-specific interactions are concerned. An interesting alternative to our present modeling of nucleic acids was presented by Srinivasan and Olson (1987). In that paper, attention was focused on the multiple structural solutions associated with the arrangements of base pairs. Optimization of the sugar-phosphate backbone, which is the main purpose of the current investigations, was not detailed there.

For the protein hydrophobic core, it is becoming clearer that there may be only a relatively limited number of ways to form hydrophobic cores for a given protein; hydrophobic interactions within these cores impart the dominant stability to a protein. The connecting, more polar loops are frequently observed to be flexible and the structures to be relatively indifferent overall to their sequences. In both of these cases, for DNA double helix and for protein folds, it is useful to have a computational method with which connecting segments of the chain can be minimized in the presence of rigid connections at the two ends. One approach to peptide loop conformations was presented by Zheng et al. (1993a,b, 1994). A different method of doing this for the DNA double helix is presented in this paper, together with some discussions about models of electric interactions in DNA. In the following paper, we apply this new method to

Received for publication 17 August 1995 and in final form 7 June 1996.

Address reprint requests to Dr. Robert Jernigan, Laboratory of Mathematical Biology, Bldg. 12B, Room B-116, DBS, National Cancer Institute, National Institutes of Health, Bethesda, MD 20892-5677. Tel.: 301-496-4783; 301-402-4724; E-mail: jernigan@lmmb.nci.nih.gov.

0006-3495/96/09/1493/14 $2.00

calculations on short DNA double helices of different sequences. By uniformly sampling over the acceptable ranges of variable positions between neighboring bases, and combining this with the minimized backbone segments, we are able to provide a relatively complete sampling of DNA double-helix conformations. This was also our intention in previous works on DNA conformations and their sequence dependence (Sarai et al., 1988; Mazur et al., 1989), but there base pairs were treated as independent unconnected bases.

In this paper, we describe a method that supplements those procedures to include a treatment of the backbone itself and its interactions with the space-fixed bases. It is intrinsically difficult to consider, in the same computational algorithm, translational and rotational degrees of freedom for individual bases, together with the internal rotational degrees of freedom for the backbone; however, other computational algorithms have been developed that encompass both base pairs and backbone in a single computational unit (e.g., Lavery, 1990, in his JUMNA algorithm). In the present work, first the stacking arrangement of the base pairs is chosen. Thus are the backbone segments connecting two base pairs constrained, and the backbone segments are energy minimized directly in a force field that includes valence and nonbonded interactions, using standard methods based on a Taylor expansion of coordinates about their minimum energy values.

The aim is to enumerate ranges of stacking arrangements, and the premise here is that small changes in stacking arrangements of the base pairs can be relatively easily accommodated by small changes among the six rotational degrees of freedom in each connecting segment of sugar-phosphate backbone atoms. Because the potential that hinders these rotations is much softer than potentials that control the structural rigidity of bases, the backbone on the scale of repeat units is, relatively speaking, conformationally flexible. The backbone flexibility, compared to the relative structural rigidity of the base pairs, provides the motivation for the development of the present computational algorithm, which optimizes the backbone segments attached to structurally rigid base pairs.

OUTLINE OF COMPUTATIONAL PROCEDURE

The molecular computational unit for the present calculations consists of four nucleotide units. Each unit includes two strands, so that a total of eight phosphate and eight deoxyribose or ribose groups are treated. These units are attached to the four space-fixed base pairs. The terminal atoms of the base pairs (the N_9 atoms of the purines and the N_1 atoms of the pyrimidines) serve as the attachment points for these dangling backbone segments. The number of atoms allowed to change their positions in the two strands connecting two base pairs is 36 atoms. The Hessian matrix for the computational unit in Cartesian coordinate space is then a manageable 432 × 432 in size. The rationale for this particular computational unit is also motivated by a need to avoid end effects, i.e., to have at least one base pair step that is not at a helix terminus.

The present paper and following paper present detailed studies of the energetics and geometry of short segments of backbone. Fig. 1 shows the details of one backbone segment of one strand. The backbone segment used for conformational analysis is internal and is not one connected to terminal nucleotides. Therefore, the backbone segment for calculations must contain at least four nucleotides. However, in many cases, longer backbone segments are required. How can these short computational units be extended to longer chains? Previous calculations indicated that, between base pairs, mostly short-range interactions between base pairs dominate in the total energy, so that longer segments can readily be constructed from shorter overlapping backbone segments. Two consecutive overlapping computational units could share their terminal nucleotides. That is, the first nucleotide of the second computational unit overlaps the last nucleotide of the first computational unit. Therefore, two computational units would span seven nucleotides. For this overlapping system, a backbone chain built of n computational units would have $3n + 1$ nucleotides. The terminal, or overlapping, nucleotide units are subjected to the energy minimization process twice. This procedure introduces a certain amount of stress at the junctions between two overlapping computational units. This stress is greatly reduced, but not entirely removed, by spatially constraining the location of the first atom in each computational unit.

The main aspect of the current investigation of combining backbone and base pairs into one computational procedure is to enable expression of the energy for the entire molecule in terms of elastic constants, which express the details of the deformability of DNA, for example, in the presence of a protein. The specificity of the base sequence and the anisotropy of degrees of freedom that determine the direction and magnitude of the deformation of the double helix can be

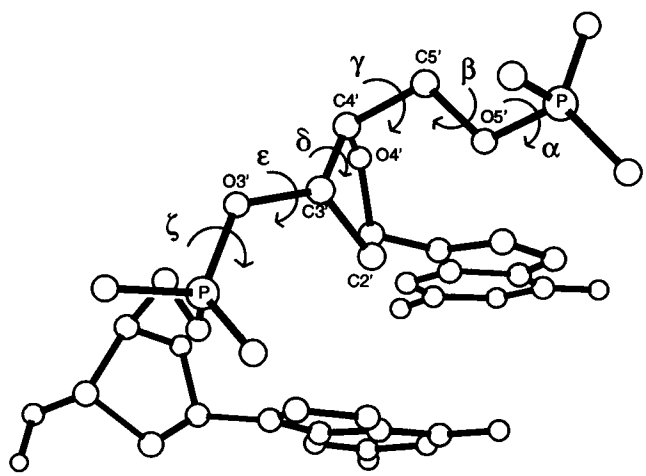

FIGURE 1 Definition of dihedral angles in the P-$O_{5'}$-$C_{5'}$-$C_{4'}$-$C_{3'}$-$O_{3'}$-P backbone segment.

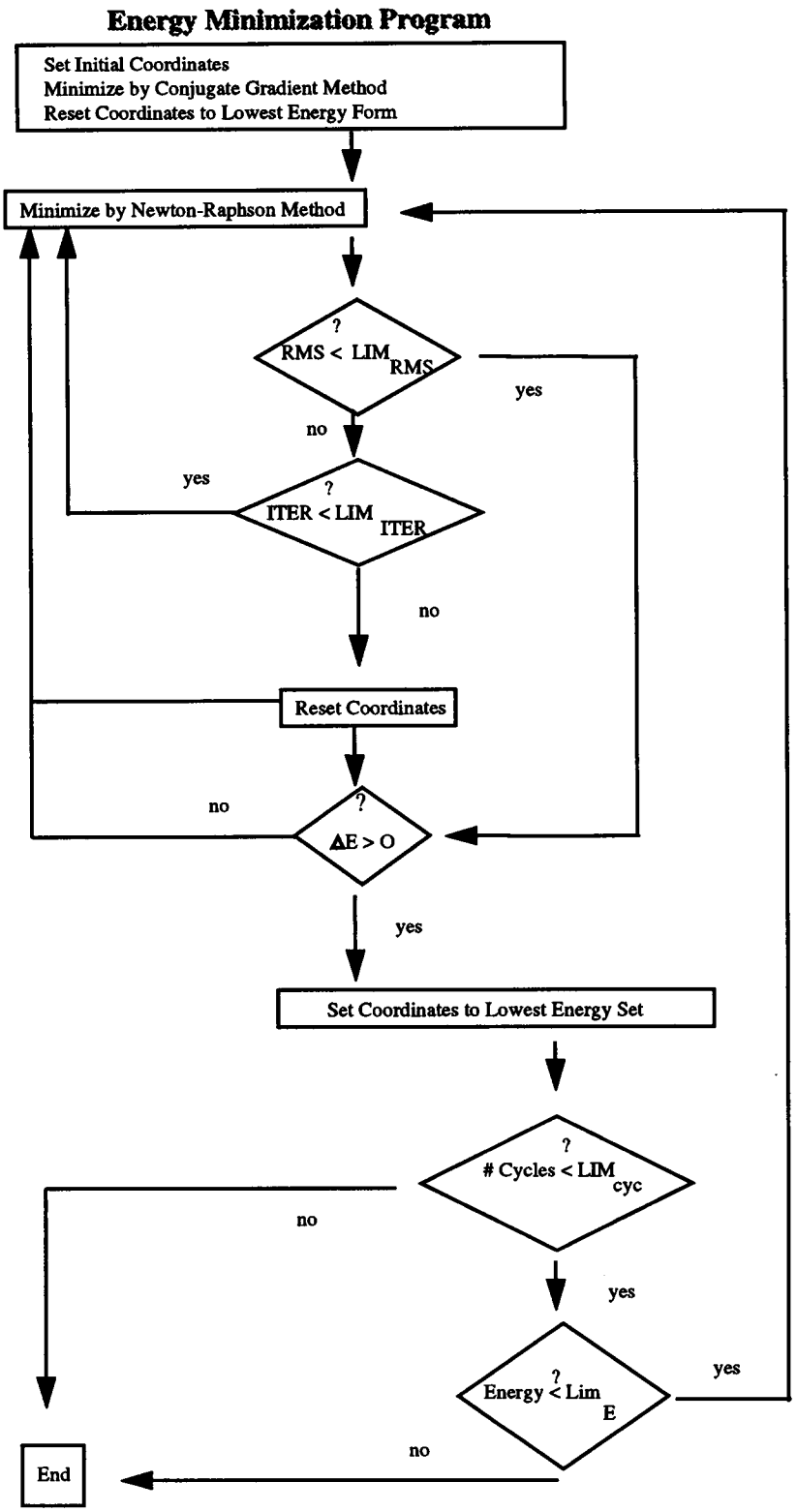

FIGURE 2 Flow diagram for the energy minimization program. ITER is the consecutive number of iterative steps performed to reach the limiting value of RMSD. ΔE is the difference in total energy between two consecutive conformers that satisfy the limiting value for RMSD.

useful in understanding the molecular mechanism for deformations.

The present work deals with constructing a minimum energy backbone that conforms structurally and energetically to a given sequence of base pairs in a given conformation. Figure 2 presents a simplified flow diagram for the energy minimization scheme employed in this and the following paper. The force field that represents the short-range, hard-core repulsion and the London dispersion forces is the same as that employed previously

(Sarai et al., 1988) in the adiabatic energy minimization of base pairs.

The form of the force for the Coulombic interactions between pairs of atoms requires special attention and examination. This is because the electrostatic interactions depend strongly on the environment of the DNA, that is, on solvent, ionic strength, temperature, etc. The backbone segments, and in particular the atoms located outside the grooves, are more exposed to solvent action than are the bases. Therefore, it is appropriate to develop a mean field form for the electrostatic interactions associated with backbone that differs from that employed earlier when dealing with the bases only.

These mean field solvent effects are incorporated through adjustable parameters in the potential function for the electrostatic interactions. These parameters are 1) the size of the low dielectric cavity, representing the part of the molecule impenetrable to direct solvent action; 2) ionic screening parameters; 3) the bulk solvent dielectric constant; and 4) the dielectric constant within the cavity. In this model, the screening parameter is assumed to be distance dependent, and it controls the rate at which the dielectric constant inside the cavity approaches its solvent value. Evaluation of these parameters and their interpretations in terms of solvent action will be given in detail below.

STRUCTURE OF THE COMPUTATIONAL UNITS: EXPLORATION OF CONFORMATIONAL SPACE IN A SEARCH FOR THE LOWEST ENERGY CONFIGURATION

The computational unit is constructed from single nucleotides. The data for the prototype nucleotides that serve as the building blocks for the assembly of the computational unit, comprising four nucleotides, were taken from x-ray data. The purpose here is to construct a larger molecule from single nucleotides. A similar approach was adopted in our previous work, where a sequence of base pairs was constructed from individual base pairs.

The coordinates of these four nucleotides were determined as follows. The base pair geometries were adapted from the work of Arnott and Hukins (1972), and the geometries of bases were fixed for the present analysis. However, the relative orientations within the paired bases were further energy optimized, using energy parameters previously employed by us (Sarai et al., 1988). This was necessary because we introduced propeller twist to base pairs as one of the conformational variables in our calculations. Thus, each base is rotated and translated in the same plane to obtain an optimized base pair. After the base pair geometry was fixed, backbone atoms were attached to these optimized base pairs. The starting backbone coordinates for B-DNA were taken from the Dickerson and Drew (1981) dodecamer (1BNA in PDB). The backbone atoms for A-DNA were taken from the crystal structure of the octamer d(GGGGCCCC) by McCall et al. (1985) (2ANA in PDB). The starting backbone structures were selected from the nucleotides located in the interior of these polynucleotides, to minimize, as much as possible, the end effects. The coordinates of the atoms in the unit nucleotides, employed in this and the following paper, are available upon request. Different crystal structures could have been used as well for extracting the starting backbone coordinates. However, because all starting structures are subjected to energy optimization in the force field, the particular selection from the various crystal structures is of minor importance. It is the force field that dominates in the process of building the DNA double helix from the optimized backbone segments.

The backbone is subjected to the same rotational and translational operations as base pairs, except for the three intra-base rotational degrees of freedom which describe the relative orientation of the two paired hydrogen bonded bases, i.e., propeller twist, buckle and opening. These internal rotations here remain an intrinsic property of the base pair. The motivation for not moving the backbone segment when intra-base rotations are performed is based on the above-described initial optimization of base pairs before their attachments to backbone atoms. The discontinuities in the calculated energies and conformations at the junctions between the backbone and the base pairs are, therefore, rather small. Those "gaps" are subsequently eliminated through the energy minimization procedure. In a way, the backbone easily "absorbs" any inhomogeneities and discontinuities in the energy and the spatial conformations created by internal rotations of the base pairs. Only propeller twist is included in the present calculations, but other degrees of freedom that spatially orient the two bases within a base pair, such as buckle, could easily be incorporated into this scheme.

The rigid base pair atoms are not subject to valence forces associated with bond stretching, valence angle deformations, and torsional rotation about bonds. However, some valence coordinates incorporate some of the connecting atoms from the base pairs, in addition to the atoms that belong to the backbone, and are, therefore, subject to displacements. These base pair atoms must be included in the backbone computations. These are the atoms N_9, C_8, and C_4 of the purines and the atoms N_1, C_6, and C_2 of the pyrimidines.

MODEL FUNCTIONS FOR THE POTENTIAL ENERGY

General description of the force field

The mean field that represents interactions between charges is the least well established, because the way the solvent enters and screens the electrostatic forces acting between partial atomic charges is complex. Because of this uncertainty, there is little point in presenting a detailed force field for the valence forces, as their exact values play a weak role in the energy minimization compared to the Coulombic and

van der Waals forces. The exception is the PO_4 unit. Because most of the initial stress originates within this group, and because it straddles the boundaries between backbone units, it is found expedient to cross-link the interactions of bond lengths and valence angles in atoms attached to phosphorous atoms. This is of particular importance for A-DNA and RNA; the presence of lateral translation between adjacent base pairs creates different initial strains at the $O_{5'}$-P-$O_{3'}$ bond angles located on the two backbone strands linked by a base pair. For example, a slide in the + direction along the axis linking the C_8 purine atom with the C_6 pyrimidine atom will initially expand this angle at one end of this axis and contract it at the other end. We approach this problem by adding a harmonic term to the $O_{5'}$-$O_{3'}$ distance with a large "penalty" force constant, to maintain the same $O_{5'}$-$O_{3'}$ distances on both strands. This additional force constant, which is a part of a Urey-Bradley force field, effectively cross-links the bond lengths and the bond angles within the PO_4 group. A general description of the force field used in the present calculations is provided in the Appendix. In this appendix only highlights rather than all details of the force field are given, with the exception of the torsional potentials. The force field associated with the torsional angles is also described in this appendix.

Parameters in the electrostatic potential of mean force

Charges on backbones are more exposed to solvent screening action than are the charges located on bases. In particular, the phosphate groups are located outside the grooves of the double helix and are more exposed to the high dielectric solvent than are more interior atoms. The backbone conformation is expected to depend strongly on the nature of electrostatic interactions and on the way they are screened by water and small ions. For this reason, part of this work examines, in detail, the electrostatic potential between the charged atoms of the backbone. We previously published similar investigations (Sarai et al., 1988; Mazur et al., 1989), but they concerned interactions between base pairs, excluding backbone. Below, we explore how solvent screening affects the geometry and energetics of the backbone and how the base sequence specificity might be involved in the structural stability of the various polymorphic forms of DNA.

To expand the concept of dielectric screening to include the sugar-phosphate backbone, it is advantageous to introduce a low dielectric cavity in the effective electrostatic potential. This cavity encompasses short-range interactions between charged atoms located within the solvent restricted space of the double helix. As we will show in the following description, the simple introduction of a solvent-restricted cavity of fixed size is especially important for dealing with the backbone in A-form DNA and in RNA. The dielectric model employed for the backbone resembles that used for certain polyampholytes (Mazur et al., 1959), where the ionization states of the nearest-neighbor COOH and NH_2 groups are controlled by the dielectric medium.

Several functional forms for the dielectric function were proposed by Mazur and Jernigan (1991) and others (Hingerty et al., 1985; Ramstein and Lavery, 1988; Friedman and Honig, 1992). The electrostatic mean field potential incorporates adjustable parameters to describe the effect of molecular environment on charge-charge interactions. A rather wide range of values has been given to some of these parameters. One of the most useful parameters for comparison is the average charge-charge distance at which the dielectric term reaches its half-solvent, or plateau, value. However, another parameter in the electrostatic energy also plays an important role in the determination of the backbone; this parameter is the radius of the dielectric cavity, and it determines the range of intercharge interactions that are unaffected by solvent screening action.

For stacking interactions between bases, the radius of the dielectric cavity coincides approximately with the sum of two atomic van der Waals radii (approximated by a cavity sphere with radius of 3.0 Å). The cavity radius employed for systems of base pair steps not connected by a backbone specifies only the starting distance from which the dielectric term becomes distance dependent. The space between base pairs is considered to be filled with a dielectric of low dielectric constant. This distinction is of some computational importance because the commonly used dielectric functions, such as the Langevin function, allow for smaller initial increases in the dielectric term near the application point of the dielectric function, rather than at larger distances. By taking the starting distance for the application of the distance-dependent dielectric term at the cavity surface, rather than at zero distance, a more gradual approach of this term toward its saturation value occurs at distances beyond the van der Waals contact distances.

In our past calculations on systems composed of unconnected base pairs (Mazur and Jernigan, 1991), a different functional form for the effective potential of mean electrostatic forces was suggested, based on the Langevin function for the distance-dependent term. The proposed form for V_{es} (in kcal/mol), the electrostatic energy, is expressed in terms of screening parameters and the radius of dielectric cavity:

$$V_{es} = 332 \sum_{ij} (1/\epsilon_{eff} r_{ij}) C_i C_j, \qquad (1)$$

in which ϵ_{eff}, the effective dielectric constant, is replaced by a Langevin function:

$$\epsilon_{eff} = (\epsilon_{bulk} - \epsilon_0)((\exp(\beta) + \exp(-\beta))/$$

$$(\exp(\beta) - \exp(-\beta)) - 1/\beta) + \epsilon_0 \quad r_{ij} > r_0,$$

$$\epsilon_{eff} = \epsilon_0 \qquad r_{ij} < r_0,$$

where $\beta = (r_{ij} - r_0)^2 \kappa_0^2$ and κ_0 is the screening parameter. r_0 represents the cavity size, ϵ_0 is the dielectric constant inside the cavity, and C_i and C_j are partial atomic charges. In this formulation, the effective dielectric constant con-

verges at large distances to ϵ_{bulk}, the water value, taken here as 78.

For calculations including backbone, we require different sets of electrostatic parameters and cavity sizes (r_0) than the ones previously employed for the base pairs alone. With base pairs, distances smaller than the sum of van der Waals radii for individual atoms seldom occur. With the backbones, the situation is quite different. For one thing, valence forces can affect the parameterization of the effective electrostatic potential. In addition, the balancing of the strong electrostatic interactions between charged dipoles at short distances precludes short-range screening between individual charged atoms.

Our calculations show that the typical geometry of the backbone of the B-form, characterized by P-P distances between adjoining P atoms of about 7 Å and by the sugar ring maintaining its $C_{2'}$-*endo* pucker mode, persists over a wide range of dielectric function parameters. The rather drastic energy changes when the dielectric function is altered are not usually accompanied by large changes in backbone geometry. For this reason, experimental data for the B-form of DNA are of limited use for the evaluation of the mean field electrostatic dielectric function in a backbone. The situation is different with "contracted" backbones, which are typical of RNA and of DNA in its A-form. The experimentally determined P-P distance of 5.9 Å, which is related to placing the sugar ring in its $C_{3'}$-*endo* pucker mode, tends to bring other atoms to distances that are equal to or less than the sums of their van der Waals radii. In particular, the H atom connected to the C_8 (in purines) or to the C_6 (in pyrimidines) atom forms a short distance contact with the $O_{5'}$ atom of the backbone, which is typical of a hydrogen bond distance. Although this may be a hydrogen bond (Taylor and Kennard, 1982), this H atom is not a conventional H donor because of the small polarity of the C-H dipole. The stability of the C-H $\cdots$ $O_{5'}$ contact is due to the electrostatic interactions, which involve not only the charges on this "hydrogen" bond but also those in its immediate environment.

We have examined different sets of electrostatic parameters to see how well the calculated structures agree with experimental stuctures. Because the charged phosphate group should be most affected by the electrostatic parameters, we took the distance between adjacent phosphate atoms as well as the backbone dihedral angles and sugar conformation to be important for choosing a set of electrostatic parameters. Calculations detailed below are based on the minimization procedure presented later in this paper, which leads to a set of dielectric parameters in the proposed expression for the electrostatic potential of mean force.

Fig. 3 shows the effect of the cavity size on the P-P distances. In this figure κ_0, the screening parameter, has a value of 0.55 $Å^{-1}$. This figure shows that small P-P distances that are observed for the A-form (around 6 Å) are more readily attained in the pyrimidine strand than in the purine strand. Both CC and TT strands display a minimum in the P-P distance when the size of the cavity is about 4.6 Å. The GG/CC step displays a steeper decrease in the P-P distances with increasing cavity size than is the case with the AA/TT step. Fig. 4 shows the effect of the screening parameter on the P-P distances. The data given are for a cavity size of 4.6 Å. As was the case with the variable cavity size, there are differences between purine and pyrimidine strands, with the pyrimidine strands always displaying lower P-P distances than the purine strand. There is some spread in the values of the screening parameter for which the P-P distance is at a minimum. For both the CC and the TT strands, P-P distance versus κ_0 plots indicates that P-P distances at or below their experimental values of 5.9 to 6.0 Å may be observed at a κ_0 of 0.5 $Å^{-1}$ or higher. The minimum P-P distance for the AA strand exceeds the observed experimental values.

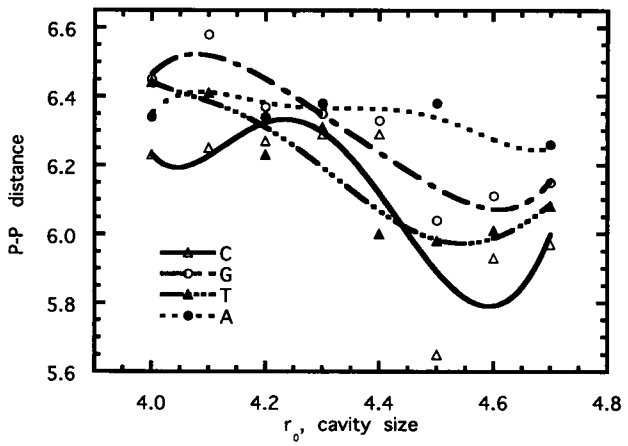

FIGURE 3 Effect of cavity size on the backbone morphology on nearest-neighbor P-P distances. Data are for GG, CC, AA, and TT strands in the A-form.

Closer examination of the effects of cavity size and screening parameter on dihedral angles and pseudorotation angles of the sugar provides an opportunity for investigating the effects of environment, such as hydrophobicity and

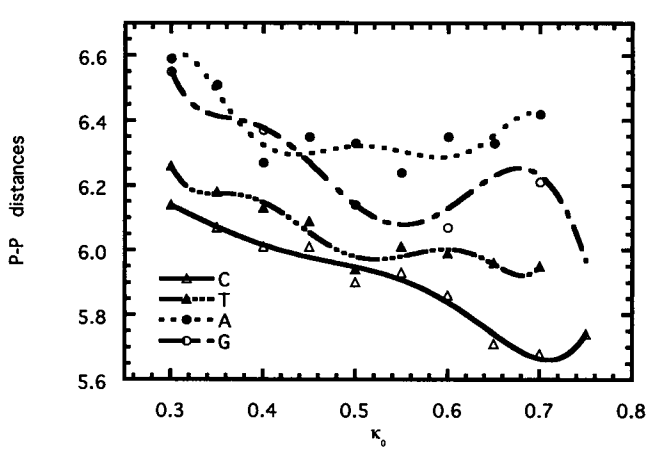

FIGURE 4 Effect of screening parameter on the backbone morphology for nearest-neighbor P-P distances. Data are for GG, CC, AA, and TT strands in the A-form.

water activity, on the DNA conformation in the A-form. To investigate the environmental effect, as represented by the dielectric screening and the size of the low dielectric cavity on dihedral angles, it is advantageous to introduce the common concept of crankshaft rotation. The crankshaft rotation of two dihedral angles occurs when the intervening bond remains in the *trans* configuration. Because the backbone is usually characterized by having two dihedral angles, β and ϵ, in trans configurations, the neighboring two pairs of dihedral angles, (α, γ) and (δ, ζ), display crankshaft-like rotations (i.e., they rotate in opposite directions) about the $O_{5'}$-C5′ bond and the $C_{3'}$-$O_{3'}$ bond, respectively. The two sets of three adjoining dihedral angles, α, β, γ and δ, ϵ, ζ, are designated as the β and ϵ crankshafts, respectively.

Fig. 5 shows the effect of r_0, the cavity size, on the β crankshaft rotation. This figure shows that the dihedrals α and γ tend toward their *gauche*$^-$ and *gauche*$^+$ values, respectively, at $r_0 \sim 4.6$ Å, the value at which the P-P distances are found to be in agreement with the experimentally observed ones. Differences between the purine and pyrimidine strands are found to be insignificant.

The total energy and, in particular, the electrostatic interactions strongly depend on the cavity size and the screening parameter. This energy is of limited physical interest, because it is highly affected by the short-range interactions between charges and, therefore, is subject to large numerical fluctuations. Nonetheless, it is of interest to note that the system has its lower energies for sets of electrostatic parameters for which the P-P distances are within their experimental range. Fig. 6 shows the dependence of the total energy on the cavity size for the GG/CC and AA/TT steps. Fig. 7 shows the total energy as a function of the screening parameter for the same two sequences. Both figures show that the energy drops off rapidly with increases in the cavity size and in the screening parameter, until they reach critical values at which the energy curve tends to either flatten out or reach low energy values of about −30 kcal/mol, which

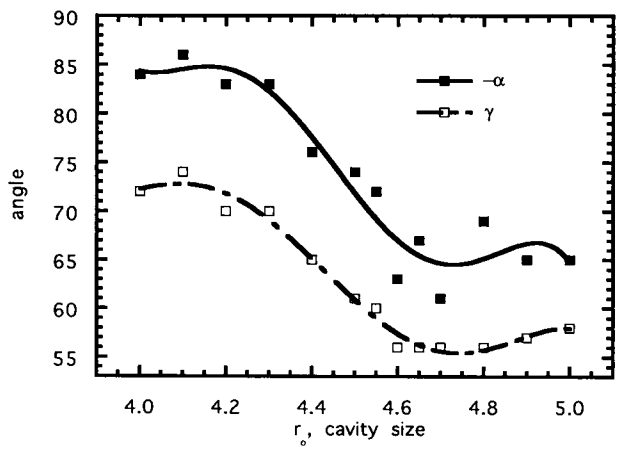

FIGURE 5 Effect of cavity size on the crankshaft rotation about the $O_{5'}$-$C_{5'}$ bond. Data are for GG sequences in the A-form. Angle values are averages of purine and pyrimidine strands.

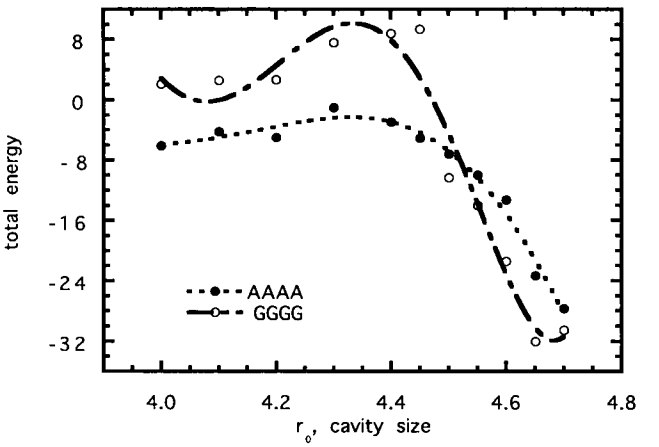

FIGURE 6 Effect of cavity size on the total energy of the GGGG and AAAA sequences in the A-form.

are typical calculated energies here for the A-form. At these critical values of the electrostatic parameters ($r_0 = 4.6$ Å, $\kappa_0 = 0.65$ Å^{-1}) the P-P distances are at or near their experimentally observed ranges.

The effects of the cavity size and the screening parameter on the P-P distances and the sugar ring conformations for RNA have been preliminarily investigated. Tables 1 and 2 show that both of these electrostatic parameters maintain values similar to those for A-form DNA, to preserve the observed P-P distances and the sugar pucker modes corresponding to RNA (Saenger, 1983). These tables also present the total energies. It is of interest to notice the narrow range of electrostatic parameters for which RNA is energetically stable. The differences between purine and pyrimidine strands in their responses to the changes in the electrostatic parameters are less noticeable for RNA than was the case for A-form DNA. The data presented in these tables are an example of the structural conservatism of the RNA and its inability to display the kind of structural polymorphism associated with the B-form of DNA.

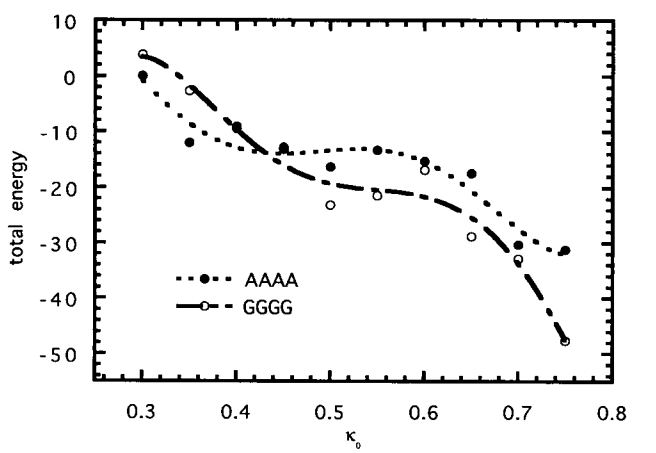

FIGURE 7 Effect of screening parameter on the total energy of the GGGG and AAAA sequences in the A-form.

TABLE 1 Effect of cavity size on the sugar-backbone morphology in RNA

		P-P distance		Pseudorotation	
r_0	Energy	Y	R	Y	R
3.00	5.7	6.52	6.54	63	72
3.25	2.9	6.26	6.30	22	23
3.50	−9.8	6.23	6.37	29	7
3.75	1.4	6.35	6.16	30	11
4.00	−14.3	6.29	6.57	16	53
4.25	−13.7	6.28	6.46	10	36
4.50	−59.0	5.74	6.05	−5	17
4.75	−60.7	5.94	6.11	−7	6
5.00	−49.2	6.19	6.28	2	4

Calculations are for the GG step in the GGGG sequence. $\kappa_0 = 0.550$ Å^{-1}, twist = 31°, slide = −1.75 Å, shift = −0.5 Å. Notice the sharp increase in the energy for r_0 of 4.25 Å and up, accompanied by large increases in the P-P distances.

Data presented in Figs. 3–7 and in Tables 1 and 2 clearly demonstrate the significant effects of cavity size and screening parameters on the phosphate-phosphate distances, the dihedral angles, and the pseudorotation angles of sugar rings. On the basis of the material presented in these figures and tables, it is suggested that acceptable values for the dielectric parameters that are compatible with the experimental data for the A-form of DNA and for RNA are r_0 (the cavity size) = 4.6 ± 0.1 Å, and κ_0 (the screening parameter) = 0.5 ± 0.15 Å^{-1}.

These are the values adopted for the dielectric parameters employed in most of the following stress-strain analysis carried out on the A-form of the DNA.

In summary, to achieve the observed P-P distances (5.9 ~ 6.0 Å) in A-form DNA and RNA (Saenger, 1983), the cavity size must be increased for the calculations with backbone structures. Simultaneously, long-range solvent effects must be taken into consideration by adopting a range of screening parameters that will reproduce the observed P-P distances.

TABLE 2 Effect of screening parameter on the sugar-backbone morphology in RNA

		P-P distance		Pseudorotation	
κ_0	Energy	Y	R	Y	R
0.30	13.4	6.11	6.41	0	15
0.35	−5.7	6.04	6.23	−3	12
0.40	−23.2	5.97	6.15	−5	11
0.45	−39.2	5.92	6.08	−5	10
0.50	−49.7	5.90	6.07	−6	10
0.55	−60.7	5.88	6.08	−6	14
0.60	−70.1	5.74	6.07	−6	13
0.65	−78.9	5.68	6.03	−9	14
0.70	−86.1	5.66	6.02	−6	13

Calculations are for the middle GG step in the GGGG sequence. $r_0 = 4.6$ Å, twist = 31°, slide = −1.75 Å, shift = −0.5 Å. Notice a sharp decrease in energy at $\alpha_0 = 0.40$ Å^{-1} and higher, accompanied by large decrements in the P-P distances.

Because the present calculations may be sensitive to the partial atomic charges, we also repeated these calculations with a different set of charges (Weiner et al., 1984). The dependence of P-P distance on the cavity size was more pronounced, as expected, because these charges are larger. However, the general trends are quite similar to the results presented above for the charge set from Zhurkin et al. (1981). The experimentally observed P-P distances were also reproduced with a cavity size in the range of 4.5–4.8 Å.

It is of interest to compare our parameterized mean field electrostatic potential with the electrostatic potential obtained by Jayaram et al. (1989) and by Friedman and Honig (1992), by solving the nonlinear Poisson-Boltzmann (PB) equation. This analytical approach treats the DNA as a medium of low dielectric constant, impenetrable to solvent and immersed in a solvent with a uniform dielectric constant. A significant physical difference between the method based on solving the PB equation and the parameterization method as employed in the current work is the way screening effects from solvent and from added electrolyte are treated. The parameterization method combines these two screening effects into a single screening parameter, whereas the method based on solving the PB equation treats them separately. In that work it was postulated that the solvent screening due to phosphate-phosphate interactions differs significantly from the one resulting from base-stacking interactions. The low dielectric medium in these works is currently being replaced by a low dielectric cavity. However, in both cases, the electrostatic potential obtained by solving the Poisson-Boltzmann equations or from the parameterized potential of mean force in the current work leads to potential discontinuities at boundaries between a low dielectric interior for the DNA and its high-dielectric surroundings.

The paper by Friedman and Honig (1992) presents some calculations of phosphate-phosphate effective dielectric constants, based on a finite-difference Poisson-Boltzmann equation, as functions of distance between interacting charges. To compare the effective dielectric constants for backbone interactions in the Friedman and Honig paper with ours, it is convenient to use the distance at which the dielectric constant reaches the midpoint of its limiting solvent saturation value. The above-quoted values for dielectric parameters (based on data for A-DNA and RNA) lead to a value of $r_{1/2} = 6.9$ Å, which compares reasonably well with the data presented in the Friedman-Honig paper (for B-DNA). It should be emphasized that the calculated phosphate-phosphate effective dielectric constants, based on the PB equation, do not cover short intercharge distances that determine the size of the low dielectric cavity.

ENERGY MINIMIZATION

It is not always possible to find the molecular conformation with the lowest energy minimum in a molecule as complex as DNA. However, it is feasible to explore the conforma-

tional space located near the initial configuration. It is possible, nevertheless, with the help of the following computational algorithm, to explore a larger part of conformational space, not necessarily so restricted to the immediate vicinity of the initial polynucleotide.

This section comprises three parts. In the first part, the constraints imposed on the energy-optimized backbone are described. These constraints are unique to the methodology of constructing a longer backbone from computational nucleotide segments. In the second part, a brief description of conjugate gradient method is presented, as it applies to the energy constrained systems. The last part details the modified Newton-Raphson algorithm.

Treatment of constraints

In the first section of this paper, two computational routes are developed. First, if the goal is to create long backbones by overlapping shorter segments that are already energy minimized, the terminal atoms in the overlapping repeat units must be spatially constrained. Otherwise, a gap will form between the two overlapping repeat units upon a subsequent energy minimization of the attached backbone segment. However, for the purpose of performing a thorough study of the interdependence of conformational variables in the computational units, backbone atoms ought not to be positionally constrained. Otherwise, a backbone subsegment that connects a space-fixed base atom (either N_1 or N_9) to a space-fixed backbone atom would exert an additional strain on the energy-minimized backbone segment. The resulting stress would then affect the sought-for interdependences between the conformational variables in the double helix. Therefore, the two computational routes described previously differ in their treatments of molecular constraints.

Our present investigations are limited to backbones that link four base pairs. The purpose of these investigations is to examine how the energy and the geometry of this computational unit respond to changes in the external constraints imposed on the backbone unit. These constraints include rotational and translational degrees of freedom of the base pairs to which the backbone is anchored. Constraints should not be imposed on the terminal atoms of the backbone units for the following reason: the parts of the backbone that connect either the terminal P atom at the 5′ end or the terminal $O_{3'}$ atom at the 3′ end of the backbone strand to the space-fixed N_9 atom of a purine or to the N_1 atom of a pyrimidine base have different geometries. In the first case, the "dangling" part of the backbone is a six-atom P-$O_{5'}$-$C_{5'}$-$C_{4'}$-$O_{4'}$-$C_{1'}$ segment. In the second case, the "dangling" part is a four-atom $O_{3'}$-$C_{3'}$-$C_{2'}$-$C_{1'}$ segment. These two dangling segments differ considerably in their flexibilities, mostly because of their differing numbers of torsional rotations.

Keeping the constraints at the two end points of each backbone strand in each computational unit, one at the 5′ end and the other at the 3′ end, a procedure adopted for construction of longer backbone segments destroys the symmetries between the complementary base pair sequences, such as d(AAAA) and d(TTTT). There is another aspect of the present program that further necessitates removal of constraints on the terminal atoms. The bases are allowed to alter their initial spatial arrangements, through changes in their internal degrees of freedom that are associated with rotations and translations about rotation axes in each base pair step, such as twist, slide, etc. These changes introduce additional stresses at the junction points between the nucleotides. These stresses must be allowed to dissipate along the portions of the backbone segments bounded by the atoms directly linked to the bases. If another atom located in the same nucleotide were also constrained (e.g., the first P atom or the $O_{3'}$ atom, depending on whether the terminal point is at the $O_{5'}$ or $O_{3'}$ atom), the backbone unit that links the space-fixed terminal atom and the atom linked to a base could be subject to different stresses than is the case for the backbone segments anchored at both ends to spatially fixed base pairs. The stresses associated with the terminal atoms being held in fixed positions will depend not only on the direction of the backbone unit (that is, toward the $O_{5'}$ or $O_{3'}$ atom), but also on the type of terminal base, i.e., whether it is a purine or a pyrimidine.

Removal of constraints from the first atom in the energy-minimized backbone segment creates certain computational difficulties, because the minimization procedure starts at a coordinate location that is allowed to change its position as the backbone is being energy optimized. It is therefore desirable that the "dangling" segment, from its terminal atom to the first space-fixed atom (which is N_9 of a purine base or N_1 of a pyrimidine base), be as short and as conformationally stiff as possible. For this reason, it is operationally preferable to begin the energy minimization from the 3′ terminus. The preference for the 3′ end over the 5′ end as the starting point for the matrix computations is motivated by the different geometries and lengths of the dangling segments that link the terminal atoms (P atom for the 5′ end and $O_{3'}$ atom for the 3′ end) to the anchor point of the base. Specifically, only one torsional degree of freedom, δ, separates the 3′ terminus from the space-fixed base. The 5′ terminus is characterized by a dangling crankshaft, in which the two terminal dihedrals (α, γ) can freely rotate in opposite directions about the middle $O_{5'}$-$C_{5'}$ bond, which is usually in the *trans* conformation. This is the reason for the greater initial rigidity of the dangling segment attached to the 3′ end compared to the structural flexibility of the dangling segment located at the 5′ end.

Description of the energy minimization algorithm

The energy minimization algorithm is based on the conjugate gradient method, followed by a modified Newton-Raphson method combined with a Gaussian elimination method. Conjugate gradient and the Newton-Raphson meth-

ods treat the entire molecule. A conjugate gradient procedure removes most of the initial strains due to gaps at the phosphate groups, where the boundaries between repeat units are. In particular, the conjugate gradient method restores the original regular tetrahedral arrangement of atoms in the phosphate groups at junction points between individual nucleotides, where they are linked together to form a polynucleotide. The conjugate gradient method, like all other first-order methods, is not so reliable when a more complete energy minimization is needed. For the purpose of the present studies, we must obtain a configuration that will simulate, as closely as possible, the local lowest minimum energy conformer. With first-order methods, such as conjugate gradient, large fluctuations in the energies and geometries would render our comparative studies of the response of the double helix to varying amounts of stress quite unreliable for the following reasons. The main difficulty with the conjugate gradient method is that it terminates the calculations at some local energy minimum or at a saddle point on the energy surface that might be far removed from the global minimum. In addition, our electrostatic force field is discontinuous at boundaries between cavities with a constant dielectric term and their environment, where the dielectric term is taken as being distance dependent.

The presence of discontinuities in the force field can lead to a premature termination of calculations when based on first-order methods. Occasionally, when starting with very strained conformers, the conjugate gradient method can lead to choosing high-energy rotational states when either the β or the ϵ dihedral angles are significantly distorted from their original *trans* forms. If one of these two dihedral angles falls into a *gauche* potential well, it can lead to the formation of a highly unfavorable *gauche*$^-$-*gauche*$^+$ pair at two adjacent dihedral bonds. As an example, if the β dihedral angle falls into a *gauche*$^-$ potential well, while the adjacent γ dihedral angle maintains its original *gauche*$^+$ orientation, a conformer with two adjoining *gauche* dihedral angles of opposite signs ensues that is characterized by a high configurational energy but is locally stable conformationally because of low energies associated with the two individual *gauche* potential wells. The high energy of this conformer, called the "pentane effect" (Flory, 1969), is relatively stable and cannot be easily removed by subsequent second-order minimization methods. This conformation is avoided altogether and a lower energy conformer is obtained if the initial conjugate gradient algorithm is removed from computations.

Comments on the conjugate gradient method

The conjugate gradient program employed in this work, which was run with Unicos version 6.1.5a and cft77 version 5.0.2.2, is taken from the IMSL Math/Library (1991). It is based on the Fletcher-Reeves algorithm (1964), and it requires a user-supplied gradient. The energy gradient is calculated analytically and is used subsequently to calculate the second derivative matrix for the Newton-Raphson method. Customarily, the conjugate gradient calculations are terminated either when the energy cannot be further decreased (that is, when changes in the potential energy between two consecutive iterative steps are smaller than a predetermined value) or when the norm of the gradient is less than a specified default value.

The presence of space-fixed atoms that interact with the energy-minimized backbone produces a certain amount of stress on the backbone. Therefore, the norm to the gradient cannot reach a zero value. As a matter of fact, the final value of the calculated gradient provides a qualitative estimate of the applied stress. The nonvanishing gradient is typical of the situation encountered in systems in which a deformable macromolecular chain is subject to the constraint of surrounding chains when it must fit into a crystal lattice (Reneker and Mazur, 1987). Therefore, it would be impossible to specify a satisfactory default value for the norm of the gradient. We therefore let the norm of the gradient reach zero asymptotically, even though the limiting value of this norm is finite and is determined by the interactions between optimized backbone atoms and the atoms of the space-fixed base pairs.

The large size of the optimized system subjects the calculations to the possibility of large roundoff errors, which are further exacerbated by the presence of energy discontinuities at dielectric cavity boundaries. However, because the conjugate gradient method is always followed by a Newton-Raphson procedure, the exact value of the energy at the termination point of this preliminary computational step is of little significance. In a typical calculation, the number of atoms whose positions are altered is 144. Therefore, there are 426 independent degrees of freedom. The norm of the gradient at the termination point is found to range between 0.025 and 0.04 kcal/mol-Å per degree of freedom.

Description of the second-order energy minimization algorithm

The above description of the conjugate gradient method clearly indicates that the shortcomings inherent in this algorithm render the method unsatisfactory for our purpose, i.e., a detailed and systematic investigation of the response of the double-helix backbone to applied external forces. In spite of the nature of the potential energy surface, with many local minima often separated by low energy barriers, it is desirable to approach a locally lowest energy as closely as possible. Therefore, our aim is to develop a large set of local energy minima, from which the best energy minimum is to be taken. The conjugate gradient method provides just one such minimum, which is often considerably higher than the locally lowest minimum.

The energy computational algorithm described below is directed toward developing a sufficient number of these local energy minima. The first energy minimum is found via the conjugate gradient method. All other minima are calcu-

lated from Hessian matrices of second derivatives, using the Newton-Raphson method. The text by McCammon and Harvey (1987) provides a detailed description of the conjugate gradient and the Newton-Raphson methods. The problem of the existence of multiple local minima had been tackled by Srinivasan and Olson (1987). They found the presence of multiple minima on the potential energy surface to be useful for comprehending the conformational complexities of the backbone and for understanding the transitions between different backbone morphologies. Tung (1992) employed Monte Carlo simulations to overcome this multiple minimum problem. However, for our problem of deriving stress-strain relations among conformational variables, it is essential to set apart the locally lowest energy minimum.

The adopted computational route is described below. The unit of eight nucleotides, joined by two backbones, is energy minimized using the standard Newton-Raphson method. The initial conformer is usually the one obtained at the end of the conjugate gradient method. The following parameters are initially assigned to the computation program: 1) a limiting value for the root mean square deviation (RMSD) between two consecutive iterative solutions of the Hessian matrix, 2) a specific maximum number of iterative steps for reaching the limiting value of the RMSD, and 3) the number of times that the total energy is allowed to increase between two consecutive iterative steps. Once the required RMSD is obtained, a check is made to determine whether the energy is also at a local minimum. This is done by performing another set of iterative steps, starting with the conformer found with the limiting value of the RMSD. This process is known as "overshooting" the target. If the energy shows a decrease from its previous value, the whole process of reaching another value for the RMSD is repeated. The cycle of calculations is considered to be satisfactorily concluded only when both conditions needed for the molecule to be at its local energy minimum are satisfied: the conformer is near the bottom of its energy potential well and the RMSD-limiting criterion is satisfied. The convergence of the iterative steps implies that the energy gradient vanishes. At this point, the energy need not be at its local minimum. This is a consequence of the highly nonquadratic character of the energy surface and of the existence of many local energy minima, which are often separated by low-potential barriers.

The above cycle of computations establishes one particular conformer at a local energy minimum. This conformer is stored, together with the parameters that are related to its geometry and energetics. The "overshot" conformer, at the last iterative step that establishes the end of a cycle of calculations leading to the local energy minimum, serves as the starting conformer for another cycle of calculations. The whole process of calculations is then repeated, leading to another local energy minimum. The calculation cycle is repeated a predetermined number of times. At the end, we have a set of local energy minima. The lowest minimum in this set is then selected and "declared" to be the lowest local minimum in the given set of energy minima. Because the "overshot" conformer has a higher energy than the preceding one, we are, in a way, surmounting an energy barrier.

In most computations, the number of iteration cycles, that is, the number of generated conformers at local energy minima, was set to 80. A lower number of cycles could have been employed for low-energy initial conformers. In reality, it is incorrect to assume that the number of potential energy wells equals the number of local energies found in calculations. Many of the energy minima can fall into the same potential energy well but have different values for the RMSD. That is, the number of distinct potential energy wells that are visited at least once during the search for a locally lowest minimum is considerably smaller than the number of local energy minima investigated in the computations. Fig. 2 (above) presents a simplified flow diagram for the energy minimization procedure.

Parameters that control the Newton-Raphson method

To reduce the number of iterations required to reach the RMSD limiting value and to ensure convergence of the iterative steps, two modifications in the Newton-Raphson procedure were made. One modification requires a large attenuation of the atomic displacements after each inversion of the Hessian matrix. The way we attenuate the atomic displacements follows the modification introduced by R. Boyd and co-workers at the University of Utah in their "Molecular Builder" program. In this modification, the attenuation of the calculated new atomic positions is introduced through

$$\Delta x(\text{attenuated}) = \Delta x(\text{computed})/(1 + A \times \text{RMSD}), \qquad (2)$$

where RMSD is the mean root square deviation between unattenuated displacements in two consecutive iterations, and A is the attenuation factor. A large attenuation factor is particularly critical in the initial stages of calculations, when the eigenvectors of the Hessian matrix can be significantly large. After the first few iterative cycles, the RMSD becomes sufficiently small, so that the attenuation factor does not seriously affect the calculated atomic displacements. Values between 10 and 15 have most frequently been employed for the attenuation factor. Lower values for the attenuation parameter can lead to a "blow-up" of calculations, as both the RMSD and the total energy can become excessively large and the calculations will fail. Should this happen, computations are abandoned at this stage. The cycle with the lowest energy is then selected as representing the locally lowest energy minimum. In the present calculations, the computations stop arbitrarily when the total energy exceeds 3000 kcal/mol.

The second modification calls for employing a penalty function for the bond-stretching coordinate. This penalty function technique adds potential energy with a large positive constant to constrain a given bond to its specified

length. A harmonic term has been added to the bond stretching coordinates. The penalty function is a simplified version of the SHAKE (Ryckaert et al., 1977) algorithm, commonly employed in molecular dynamics simulations of large biological molecules. It serves a purpose similar to that of the SHAKE algorithm, which is to constrain bond lengths to their specified values. The penalty function is employed to prevent serious computational divergences in the Newton-Raphson iterative processes, rather than to reduce CPU time, which is the raison d'être for the employment of the SHAKE algorithm in molecular dynamics. The penalty function cannot be employed with the conjugate gradient method, nor is it needed there, as it does not work well with conformers that have large initial energies.

With the Newton-Raphson method, the total energy is employed only as an external parameter, which serves as an indicator for when the bottom of the potential energy well has been reached, to terminate a given cycle of energy minimization steps. Therefore, its inclusion in the calculated energy is not required. Whether this penalty function is included or removed from the computed energy has little bearing on the final results. The penalty function essentially eliminates bond lengths as degrees of freedom in the minimization procedure. This exerts some extra strain on bond angle coordinates, leading to somewhat higher energies. Our computations performed with and without penalty functions show that the geometries of the backbone are not much affected. The computational efficiency is greatly improved through use of the penalty function, because the number of matrix inversions required in a given cycle of calculations is reduced by almost one-half.

The minimization procedure presented above can be extended to cases where the purpose is to investigate transitions between different polymorphic forms of DNA. Usually different polymorphs of DNA are separated by large potential barriers. Therefore, direct minimization methods cannot be used, because they are not designed for finding reaction coordinates across potential barriers. Monte Carlo and molecular dynamics calculations have been applied to sample the conformational space of sugar ring puckering (Tung, 1992; Gabb and Harvey, 1993). Here we circumvented this difficulty by driving several bonds artificially across potential barriers (Reneker et al., 1977). In the bond-drive method, a given coordinate (usually a dihedral bond) is initially frozen at its starting value while all other coordinates are subjected to the energy minimization algorithm. The selected bond that is "driven" across the potential barrier is successively incremented by predesignated amounts, and the energy minimization algorithm is then repeated. The bond-drive method, as applied to the problem of a pseudorotating sugar across the potential barrier that separates the two sugar pucker modes associated with the B- and A-forms of DNA, has not been published. Here we mention that, starting with the B-form, the A-form can be reasonably simulated by driving the δ dihedral angle, or any one of the five endocyclic torsional bonds of the sugar ring, toward the value it maintains in the A-form of DNA. The principal advantage of the "bond-drive" method is that it leads to a reasonable estimate of the height of the potential barrier that separates the two sugar pucker modes associated with these two polymorphs of DNA, and of the dependence on solvent action. The problem of estimation of the energy barrier that separates the sugar pucker modes of A-DNA and B-DNA forms of DNA was investigated by Tung (1992) using Metropolis Monte Carlo simulations. In agreement with our unpublished data, by crossing over the potential barrier separating the $C_{2'}$-*endo* and the $C_{3'}$-*endo* pucker modes, the B-A transition can be induced.

SUMMARY

A novel model for DNA is presented, based on having energy-optimized sugar-phosphate backbone segments linking a sequence of space-fixed base pairs. This molecular model effectively combines backbone and base pairs into one computational procedure. The investigations presented in this and in the following paper require a computational method that reaches a conformer with the best energy minimum in the multiminimum potential energy surface. The present paper develops an energy minimization scheme, based on combining a conjugate gradient method with a suitably modified Newton-Raphson algorithm.

The backbone segments are energy minimized in a force field that takes into account solvent screening action on different polymorphic forms of the double helix. This screening action is incorporated in the electrostatic potential of mean force through a set of adjustable parameters. These parameters are adjusted to obtain the best fit of the nearest-neighbor phosphorous atoms as determined in an X-ray structure. The proposed molecular model of DNA serves several purposes. One of them is directed toward building long backbones linked to a sequence of base pairs by overlapping the previously energy-minimized shorter backbone segments. The main purpose of the computational algorithm developed in this paper, however, is to provide a molecular model for a detailed study of the conformational polymorphism of DNA and of the interrelations among the conformational variables in the double helix. These latter investigations are described in the following paper.

APPENDIX: GENERAL DESCRIPTION OF THE FORCE FIELD

Bond stretching

Because of the use of a penalty function in the energy minimization procedure, the physically relevant quantities are the equilibrium bond lengths. The force constants only enter the total energy calculations, which are used as an indicating parameter for the local energy minima. The penalty function enters the calculation of the Hessian matrix of the second derivatives. For conjugate gradient calculations, the penalty function is not employed. The data for the bond lengths were compiled from the pertinent literature.

Bond bending

For these constants, the force constants of CHARMM (Brooks et al., 1983) are employed. For the PO_4 group, all distances between oxygen atoms in this tetrahedron are assumed to be equal to 2.52 Å. Because the P-O_1' and P-O_2' distances differ considerably from the P-O_5' and P-O_3' distances (1.48 Å and 1.6 Å, respectively), the bond angles in this group must reflect these bond length differences, while maintaining the assumed equalities in all O-O distances in the PO_4 group. The actual bond angles closely follow this model. A regular tetrahedral model is assumed for the C atoms, with a C-C distance of 1.53 Å and a C-H distance of 1.09 Å. For the sugar rings, the same bond angles of 108° for all C-C-C and O-C-C angles are employed. This value is close to the equilibrium, or ideal, value of 109.5° based on a tetrahedral structure (See Olson, 1982).

Torsional angles

A single threefold periodic cosine rotational potential is usually employed. An exception is for the rotation around the N-C bond (glycosyl), for which a much lower barrier of 0.14 kcal/mol with a sixfold periodicity is used. For the CCOH rotor in RNA, a threefold rotational potential is assumed, with a barrier of 1.0 kcal/mol, based on the data for aliphatic alcohols (Herzberg, 1945). Most of the dihedral torsional angles listed below (Table 3) were adapted from Erie et al. (1993). We add to the standard threefold potential energy for the γ dihedral angle, defined as the O_5'-C_5'-C_4'-C_3' rotor, a twofold rotational symmetry term. The twofold torsional potential is shifted by a phase angle of 240°, so that the rotational *trans/gauche* energy difference remains zero, while the location of the *gauche*$^+$ rotational energy minimum is slightly shifted to angles smaller than the true *gauche* angle of 60°. The inclusion of this *gauche* effect maintains the calculated γ dihedral angles at values smaller than 60°, which follows the crystallographic data of Arnott and Hukins (1972).

van der Waals terms

In the present work, the 1–4 electrostatic and van der Waals interactions are not scaled, because the cavity effects employed here (see the following section) effectively scale these interactions. The force field for the van der Waals nonbonded interactions is the same as used previously by us for the system composed of base pairs not connected by backbone.

Point charges

The partial charges are based on the set given by Poltev and Shulyupina (1986). For RNA, partial charges for the O-H group were taken from Renugopalakrisnan et al. (1971). The charges of the C_2' and H_2' atoms were modified slightly to maintain the same total charge for the C_2' and its attached atoms. We follow the Zhurkin et al. (1981) data for the charges assigned to the backbone and to the bases. These charges have been modified recently (Zhurkin, private communication), so that the total charge of the base connected to the sugar through the C_1'-N_9 (or the C_1'-N_1) bond is identical regardless of the composition of the base.

TABLE 3 Torsion angles

Dihedral angle	Potential barrier (kcal/mol)	Periodicity	Phase angle
α(C_5'-O_5'-P-O_3')	1.0	3	0
β(P-O_5'-C_5'-C_4')	2.5	3	0
γ(O_5'-C_5'-C_4'-C_3')	2.8	3	0
γ(O_5'-C_5'-C_4'-C_3')	1.3	2	240
δ(C_5'-C_4'-C_3'-O_3')	2.5	3	0
ϵ(P-O_3'-C_3'-C_4')	2.5	3	0
ζ(C_3'-O_3'-P-O_5')	1.0	3	0
χ(O_1'-C_1'-N-C_2')	0.3	6	30
C_1'-C_2'-O_4'-H (RNA)	0.5	3	0

Part of the calculations of the displacement vectors of the atomic coordinates were based on the "Molecular Builder" program provided many years ago by Professor R. H. Boyd of the University of Utah and subsequently modified by us. In particular, the algorithm for the molecular topology was particularly useful for structuring the double helix in terms of the valence coordinates. Computations were performed on the Cray Y-MP at the Frederick Biomedical Super Computing Laboratory (Frederick, MD).

This research was partly supported by the National Cancer Institute, Department of Health and Human Services, under contract NO1-CO-74102 with PRI. The contents of this article do not necessarily reflect the views of the DHHS, nor does mention of trade names, commercial products, or organizations imply endorsement by the U.S. Government.

REFERENCES

Arnott, S., and D. W. L. Hukins. 1972. Optimised parameters for A-DNA and B-DNA. *Biochem. Biophys. Res. Commun.* 47:1504–1510.

Brooks, B. R., R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan, and M. Karplus. 1983. CHARMM: A program for energy, minimization, and dynamics calculations. *J. Comp. Chem.* 4:187.

Dickerson, R. E, and H. R. Drew. 1981. Structure of a B-DNA dodecamer. II. Influence of base sequence on helix structure. *J. Mol. Biol.* 149: 761–786.

Erie, D. A., K. J. Breslauer, and W. K. Olson. 1993. A Monte-Carlo method for generating structures of short-stranded DNA sequences. *Biopolymers.* 33:75–105 (and supplemental material available from journal).

Fletcher, R., and C. M. Reeves. 1964. Function minimization by conjugate gradients. *Comput. J.* 7:149.

Flory, P. J. 1969. Statistical Mechanics of Chain Molecules. Wiley-Interscience, New York.

Friedman, R. A., and B. Honig. 1992. The electrostatic contribution to DNA base-stacking interactions. *Biopolymers.* 32:144–159.

Gabb, H. A., and S. Harvey. 1993. Conformational transitions in potential and free energy space for furanoses and 2'-deoxynucleosides. *J. Am. Chem. Soc.* 115:4218–4227.

Havel, T. F., G. M. Crippen, I. D. Kuntz, and J. M. Blaney. 1983a. The combinatorial distance geometry method for the calculation of molecular conformation. II. Sample problems and computational statistics. *J. Theor. Biol.* 104:383–400.

Havel, T. F., I. D. Kuntz, and G. M. Crippen. 1983b. The combinatorial distance geometry method for the calculation of molecular conformation. I. A new approach to an old problem. *J. Theor. Biol.* 104:359–381.

Herzberg, Z. 1945. Infrared and Raman Spectra of Polyatomic Molecules. D. van Nostrand Company, New York.

Hingerty, B. E., R. H. Ritchie, T. L. Ferrell, and J. E. Turner. 1985. Dielectric effects in biopolymers: the theory of ionic saturation revisited. *Biopolymers.* 24:427–439.

IMSL. 1991. International Mathematical and Statistical Libraries. IMSL, Houston, TX.

Jacchieri, S., and R. L. Jernigan. 1992. Variable ranges of interactions in polypeptide conformations with a method to complement molecular modeling. *Biopolymers.* 32:1327–1388.

Jayaram, B., K. A. Sharp, and B. Honig. 1989. The electrostatic potential of B-DNA. *Biopolymers.* 28:975–993.

Lavery, R. 1990. JUMNA IIIe. Laboratorie de Biochimie Theorique CNRS, Institut de Biologie Physico-Chimique, Paris.

Mazur, J., and R. L. Jernigan. 1991. Distance-dependent dielectric constants and their application to double-helical DNA. *Biopolymers.* 31: 1615–1629.

Mazur, J., and R. L. Jernigan. 1995. Comparison of rotation models for describing DNA conformation: Application to static and polymorphic forms. *Biophys. J.* 68:1472–1489.

Mazur, J., A. Sarai, and R. L. Jernigan. 1989. Sequence dependence of the B-A conformational transition of DNA. *Biopolymers.* 28:1223–1233.

Mazur, J., A. Silberberg, and A. Katchalsky. 1959. Potentiometric behavior of polyampholytes. *J. Polym. Sci.* 35:43–47.

McCall, M., T. Brown, and O. Kennard. 1985. The crystal structure of d(G-G-G-G-C-C-C-C). A model for poly(dG)-poly(dC). *J. Mol. Biol.* 183:385–398.

McCammon, J. A., and S. C. Harvey. 1987. Dynamics of Proteins and Nucleic Acids. Cambridge University Press, Cambridge, England.

Moore, J. M., D. A. Case, W. J. Chazin, G. P. Gippert, T. F. Havel, R. Powls, and P. E. Wright. 1988. Three-dimensional structure of plastocyanin from the green alga *Scenedesmus obliquus. Science.* 240: 314–317.

Olson, W. K. 1982. How flexible is the furanose ring? 2. An updated potential energy estimate. *J. Am. Chem. Soc.* 104:278–286.

Poltev, V. I., and N. V. Shulyupina. 1986. Simulation of interactions between nucleic acid bases by refined atom-atom potential functions. *J. Biomol. Struct. Dyn.* 3:739–765.

Ramstein, J., and R. Lavery. 1988. Energetics coupling between DNA bending and base pair opening. *Proc. Natl. Acad. Sci. USA.* 85: 7231–7235.

Reneker, D. H., B. M. Fanconi, and J. Mazur. 1977. Energetics of defect motion which transports polyethylene molecules along their axis. *J. Appl. Phys.* 48:4032–4042.

Reneker, D. H., and J. Mazur. 1987. Small defects in crystalline polyethylene. *Polymer.* 29:3–14.

Renugopalakrisnan, V. A., V. Lakshminarayana, and V. Sasisekharan. 1971. Stereochemistry of nucleic acids and polynucleotides. III. Electronic charge distribution. *Biopolymers.* 10:1159–1167.

Ryckaert, J. P., G. Ciccoti, and H. J. C. Berendsen. 1977. Numerical integration of the Cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *J. Comp. Phys.* 23:327.

Saenger, W. 1983. Principles of Nuclear Acid Structure. Springer-Verlag, New York. Figs. 2 and 3.

Sarai, A., J. Mazur, R. Nussinov, and R. L. Jernigan. 1988. Sequence dependence of DNA conformations: means and fluctuations. *In* Structure and Expressions, Vol. 3. W. K. Olson, M. H. Sarma, and M. Sundaralingam, editors. Adenine Press, Schenectady, New York. 213–223.

Srinivasan, A. R., and W. K. Olson. 1987. Nucleic acid model building: the multiple backbone solutions associated with a given base morphology. *J. Biomol. Struct. Dyn.* 4:895–938.

Taylor, R., and O. Kennard. 1982. Crystallographic evidence for the existence of C-H···O, C-H···N, and C-H···Cl hydrogen bonds. *J. Am. Chem. Soc.* 104:5063–5070.

Tung, C.-S. 1992. A reduced set of coordinates for modeling DNA structures: (I) B-to-A transition pathway driven by pseudorotation angle. *J. Biomol. Struct. Dyn.* 9:1185–1194.

Weiner, S. J., P. A. Kollman, D. A. Case, J. C. Singh, C. Ghio, G. Alagona, S. Profeta, and P. J. Weiner. 1984. A new force field for molecular mechanical simulation of nucleic acids and proteins. *J. Am. Chem. Soc.* 106:765–784.

Yue, K., and K. A. Dill. 1995. Forces of tertiary structural organization of globular proteins. *Proc. Natl. Acad. Sci. USA.* 92:146–150.

Zheng, Q., R. Rosenfeld, C. DeLisi, and D. J. Kyle. 1994. Multiple copy sampling in protein loop modeling: computational efficiency and sensitivity to dihedral angle perturbations. *Protein Sci.* 3:493–506.

Zheng, Q., R. Rosenfeld, S. Vaida, and C. Delisi. 1993a. Loop closure via bond scaling and relaxation. *J. Comp. Chem.* 14:556–565.

Zheng, Q., R. Rosenfeld, S. Vaida, and C. Delisi. 1993b. Determining protein loop conformation using scaling-relaxation techniques. *Protein Sci.* 2:1242–1248.

Zhurkin, V. B., V. I. Poltev, and V. I. Florent'ev. 1981. Atom-atom potential functions for conformational calculations of nucleic acids. *Mol. Biol.* 14:887–895.