Conformations of an Adenine Bulge in a DNA Octamer and Its Influence on DNA Structure from Molecular Dynamics Simulations

Michael Feig,* Martin Zacharias,† and B. Montgomery Pettitt*
*Department of Chemistry and Institute for Molecular Design, University of Houston, Houston, Texas 77204-5641 USA; and †AG Theoretische Biophysik, Institut für Molekulare Biotechnologie, 07708 Jena, Germany

ABSTRACT Molecular dynamics simulations have been applied to the DNA octamer d(GCGCA-GAAC) · d(GTTCGCGC), which has an adenine bulge at the center to determine the pathway for interconversion between the stacked and extended forms. These forms are known to be important in the molecular recognition of bulges. From a total of ~35 ns of simulation time with the most recent CHARMM27 force field a variety of distinct conformations and subconformations are found. Stacked and fully looped-out forms are in excellent agreement with experimental data from NMR and x-ray crystallography. Furthermore, in a number of conformations the bulge base associates with the minor groove to varying degrees. Transitions between many of the conformations are observed in the simulations and used to propose a complete transition pathway between the stacked and fully extended conformations. The effect on the surrounding DNA sequence is investigated and biological implications of the accessible conformational space and the suggested transition pathway are discussed, in particular for the interaction of the MS2 replicase operator RNA with its coat protein.

INTRODUCTION

Errors during replication and recombination of double-stranded DNA may result in mismatched basepairs or the insertion/deletion of nucleotides. Although both types of mutation might compromise the biological fidelity of DNA, the most severe consequences can arise when nucleotide base bulges constitute the starting point for frameshift mutagenesis (Streisinger et al., 1966). Deficiencies in the natural repair mechanisms increase genetic instability through homologous recombination (Modrich and Lahue, 1996) and have been identified as the cause of both hereditary and sporadic tumors (Kolodner, 1996; Eshleman and Markowitz, 1995). Bulge-containing DNA is initially recognized by MutS protein during prokaryotic MutHLS-type mismatch repair (Kolodner, 1996) and MSH2 · MSH6 complexes with a high degree of sequence similarity to MutS in prokaryotic organisms (Marsischky et al., 1996; Genschel et al., 1998), but only very limited detailed structural information is available on the recognition mechanism. Specific binding to DNA with bulged nucleotides has also been reported for antitumor drugs (Kappen and Goldberg, 1993, 1997), intercalating agents (Nelson and Tinoco, 1985; Woodson and Crothers, 1988a; Lippard and Berg, 1993), and metal complexes (Cheng et al., 1999).

In RNA, nucleotide bulges play an important role as secondary structure elements (Turner, 1992) for specific recognition by proteins (Wyatt and Tinoco, 1993), RNA splicing (Query et al., 1994), and RNA folding (Tang and Draper, 1990). Prominent examples for the involvement of bulged bases in protein-RNA interactions include binding of ribosomal proteins to ribosomal RNA (Peattie et al., 1981; Baudin and Romaniuk, 1989; Zhang et al., 1989), binding of Tat protein to the transactivation response region (TAR) RNA of HIV-1 (Puglisi et al., 1992; Aboula-ela et al., 1995) and binding of phage R17, MS2, and GA coat proteins to an RNA stem-loop containing the translational repression operator and initiation site for the replicase gene (Wu and Uhlenbeck, 1987; Valegard et al., 1994, 1997).

Although the structure of regular double-stranded nucleic acids is generally well understood, relatively few studies have examined the effect of extra unpaired nucleotides on the helical structure of nucleic acids. Comparisons of electrophoretic mobilities in polyacrylamide gels have suggested the introduction of significant bending at the presence of bulge bases (Rice and Crothers, 1989; Bhattacharyya and Lilley, 1989; Wang and Griffith, 1991) with bend angles of 10–20° per extra base as measured by transient electric birefringence (Zacharias and Hagerman, 1995). Similar conclusions have also been found from electron microscopy (Hsieh and Griffith, 1989) and fluorescence energy transfer experiments (Gohlke et al., 1994). Higher-resolution structural data of bulge-containing double-stranded DNA and RNA molecules is available from NMR and x-ray diffraction studies, mostly for single bulges. The data on DNA support two main conformation types for bulge-containing double-stranded nucleic acids with bulge bases either stacked with the rest of the helix or looped-out in a more extended conformation that have been postulated earlier (Fresco and Alberts, 1960). From NMR experiments it appears that in solution purine bulge bases prefer the stacked conformation (Patel et al., 1982; Hare et al., 1986; Nikonowicz et al., 1989, 1990; Rosen et al., 1992; Woodson and Crothers, 1988b) independent of sequence context and

Received for publication 19 May 2000 and in final form 9 February 2001.

Address reprint requests to Dr. B. Montgomery Pettitt, Dept. of Chemistry, University of Houston, Houston, TX 77204-5641. Tel.: 713-743-3263/2701; Fax: 713-743-2709; E-mail: pettitt@uh.edu.

Michael Feig's present address is Dept. of Molecular Biology, Scripps Research Institute, TPC6, 10550 North Torrey Pines Rd., La Jolla, CA 92037.

0006-3495/01/07/352/19 $2.00

temperature (Kalnik et al., 1989a). Pyrimidine bulge bases, however, are reported in an equilibrium between both forms (Morden et al., 1983, 1990; van den Hoogen et al., 1988a) that is shifted toward the stacked or looped-out form depending on temperature and/or sequence context (Kalnik et al., 1989b, 1990). X-ray crystal diffraction experiments of DNA fragments containing single adenine bulges with a very similar sequence as those studied by NMR have found extra-helical conformations (Miller et al., 1988; Joshua-Tor et al., 1988, 1992). Although these findings are in contrast with the NMR results, they may indicate a shifted equilibrium toward a looped-out conformation as a consequence of the close packing in the crystal environment, especially when the looped-out base is intercalated with a symmetrically related neighboring helix (Joshua-Tor et al., 1992). A preference of purine bulges for stacked conformations compared to pyrimidine bulges in equilibrium between stacked and looped-out forms as seen in solution is also more consistent with the finding that helix bending is more pronounced with purine bulges than with pyrimidine bulges (Wang and Griffith, 1991). Helix bending can be easily explained by extra intrahelical bases on one strand while a looped-out base allows for stacking of the flanking bases so that a straighter helix structure can be assumed (Joshua-Tor et al., 1992).

In bulge-containing RNA stacked and extended bulge conformations depend on the environment in a similar way as for DNA. Under solution conditions, stacked (Borer et al., 1995; Kerwood and Borer, 1996; Smith and Nikonowicz, 1998; Thiviyanathan et al., 2000) and partially looped-out conformations (Greenbaum et al., 1996) are found for a single adenine bulge, while a single uridine base has been reported to be completely looped-out (van den Hoogen et al., 1988b). In the crystal environment single adenine bulges are extended in unbound RNA (Cate et al., 1996; Portmann et al., 1996; Golden et al., 1998; Ennifar et al., 1999; Sudarsanakumar et al., 2000) and in complexes with phage MS2 coat protein (Valegard et al., 1994, 1997; Rowsell et al., 1998; van den Worm et al., 1998), suggesting again that crystal and protein environments shift the equilibrium between stacked and extended conformations toward the extended form from a preference for the stacked form in solution.

Recently, continuum solvent modeling was used to examine possible conformations of the DNA fragment d(CG-C*A*GAA) · d(TTCGCG) with an adenine bulge (Zacharias and Sklenar, 1997) and compare them with experimental data for an extra adenine base in the same (Patel et al., 1982; Nikonowicz et al., 1989, et al., 1990; Joshua-Tor et al., 1992) or a very similar (Hare et al., 1986) sequence context. Both experimentally observed structures with the bulge base stacked between flanking bases or looped-out were found as local energy minima during the conformational search. The energy for the best stacked form was found to be 2.5 kcal/mol lower than for the lowest energy looped-out form. Although the difference of the total energy lies within the accuracy of the calculation method, significant differences in individual energetic contributions could be determined. The results suggest that the stacked form is stabilized mainly by favorable electrostatic interactions within the nucleic acid backbone, while the looped-out benefits most from favorable electrostatic contributions to the solvation energy.

Interestingly, another low-energy conformation was identified that can be classified as intermediate between the stacked and fully looped-out forms. It involves the formation of a base triple with the adjacent basepairs in the minor groove and is energetically very close to the stacked form (+0.5 kcal/mol). However, the modeling approach used in the above study did not propose a path for interconversion. Another limitation due to the lack of explicit modeling of solvent molecules is that this study did not take into account dynamic properties that may be relevant for coupling between DNA and solvent, and are expected to give rise to varying entropic contributions for different conformations (Zacharias and Sklenar, 1997). This appears to be an important factor, as recent studies have found interactions with explicit ions to be a key aspect in shifting the equilibrium of regular nucleic acid structures between A and B conformations (Feig and Pettitt, 1999b; Jayaram et al., 1998).

Molecular dynamics simulations with an explicit ionic solvent representation in atomic detail can be used to address some of the limitations in previous modeling studies. More extensive conformational sampling in such simulations provides a more detailed view of the complex energy landscape and the accessible conformational space. Different bulge orientations in the extended form that occur in the available x-ray structures raise the question whether there are corresponding distinct stable subconformations. Particularly interesting in the biological context is the transition between extended and stacked forms because experimental data appear to suggest that bulge bases may be mostly stacked in solution, but fully extended in contact with a protein (Valegard et al., 1994; Borer et al., 1995). Information about a possible transition pathway and its energetics as well as the role of the surrounding sequence would be a valuable addition to understanding RNA-protein interactions that involve bulge bases.

Many molecular dynamics simulations studies have been performed on regular DNA and RNA molecules (Auffinger and Westhof, 1998; Beveridge et al., 1994, 1997; Cheatham et al., 1997b). Following recent methodological and computational advances it is now possible to run stable simulations over many nanoseconds (Weerasinghe et al., 1995a; Feig and Pettitt, 1998b; Cheatham and Kollman, 1997b; Cheatham et al., 1997a; Young et al., 1997b; Young and Beveridge, 1998; Auffinger and Westhof, 1997; Duan et al., 1997) that are increasingly successful in matching and extending experimental data with respect to structure of the nucleic acid itself (Feig and Pettitt, 1998b; Cheatham and

Kollman, 1996, 1997b; Young et al., 1997b; Young and Beveridge, 1998; MacKerell, 1997; Lee et al., 1995) and the surrounding solvent (Feig and Pettitt, 1998a, 1999a; Cheatham et al., 1997a; Cheatham and Kollman, 1997a; Young et al., 1997a). Although molecular dynamics simulations of nucleic acid structures with base mismatches have been reported (Weerasinghe et al., 1995b), to our knowledge no such studies have been performed on DNA with bulged nucleotides. Here, we will present results from molecular dynamics simulations on the multi-nanosecond time scale of the same DNA fragment as in the previous modeling study, but with an extra C · G basepair on either side. We find that five main conformations can be distinguished where the base is stacked (S), fully extended (E), forms a base triple in the minor groove (T), associated with the DNA in the major groove (M), or along the minor groove (G). These conformers form our postulated transition coordinate.

This paper is organized as follows. Following a description of the methods used we will discuss the conformational variety of bulge base structures from the simulation data and evaluate the results by comparison with the available experimental data from NMR and x-ray crystal diffraction. Then, we will continue with an analysis of structure and dynamics of the remaining basepairs in the simulated nucleic acid fragment at different bulge conformations and examine the role of the explicit solvent. Finally, the results are discussed within the biological context, in particular with respect to implications for bulge-specific interactions between proteins and nucleic acids.

METHODS

Simulations

In this paper we report results from four molecular dynamics simulation runs of the double-stranded DNA fragment d(GCGC*A*GAAC) · (GTTC · GCGC) with an extra adenine bulge on the first strand. They are summarized in Table 1. The solvent is represented by explicit water molecules, sodium counterions to balance the DNA charge, and additional sodium chloride ion pairs to arrive at 0.5 M salt concentration. The bulk of the simulations were run with a recently improved version of the CHARMM force field (Foloppe and MacKerell, 2000; MacKerell and Banavali, 2000) that provides a well-balanced representation of nucleic acid conformations (MacKerell and Banavali, 2000). Preliminary simulations for comparison, not discussed in detail in this paper, were also run with the AMBER force field (Cornell et al., 1995). In the AMBER simulations stacked and extended bulge conformations were only stable for up to a few nanoseconds during the beginning of the simulations. From the stacked and extended forms the bulge base moved into the minor and major grooves, respectively, where they remained for the rest of the simulations. These structures, termed triple (T) and major groove (M), are shown in Fig. 1 for reference. Similar conformations have also been found by continuum solvent modeling (Zacharias and Sklenar, 1997), although the lack of support by experimental data suggests that at best these conformations represent only transient structures. In contrast, structural results from simulations with the CHARMM27 force field presented in this paper are in good agreement with available experimental evidence.

The simulation program was developed in this laboratory (Smith et al., 1996). It uses periodic boundary conditions, the velocity Verlet integration schemes (Allen and Tildesley, 1987), and the SHAKE algorithm (Ryckaert et al., 1977) to enforce holonomic constraints of all chemical bonds and allow an integrator time step of 2 fs. Electrostatic interactions were calculated using a twin-range implementation (Smith and Pettitt, 1995) of the exact Ewald summation technique (de Leeuw et al., 1980; Smith and Pettitt, 1994). The direct contribution to the Ewald sum was calculated every time step within a first cutoff of 1.2 nm and updated every 10 steps from 1.2 nm to the second cutoff set at 2.0 nm, corresponding to half of the box size. A convergence factor α of 1.5 and 15 reciprocal space vectors in each direction achieved optimal performance. All simulations used TIP3P water parameters (Jorgensen et al., 1983) and recently improved ion parameters (Roux et al., 1995) and were run at 300 K within an NVT ensemble.

Two simulations were started from the lowest energy structures from the continuum solvent modeling studies with the adenine bulge base in the stacked (S-C, stacked-CHARMM) and looped-out (E-C, extended-CHARMM) conformations. In the following we will abbreviate the simulations according to Table 1, with the first letter describing the starting conformation. The other two shorter simulations were started from the intermediate conformations found in the preliminary AMBER simulations after several nanoseconds' simulation time. The first (T-C, triple-CHARMM) was started with the bulge base in the minor groove forming a base triple with the previous C · G basepair. The second simulation (M-C, major groove-CHARMM) started with the bulge base associated with the major groove.

The protocol for initial equilibration for simulations S-C and E-C follows a previously described procedure (Weerasinghe et al., 1995a; Feig and Pettitt, 1998b). An initial 20-step steepest descent minimization is followed by alternating runs with either the solvent or the solute fixed and velocity reassignment every 50 steps from a canonical Maxwell distribution at 300 K for up to 500 ps to allow sufficient equilibration of the nucleic acid structure with the solvent. Configurations were collected every 100 fs over total simulation times of 10.1 ns (S-C), 20.0 ns (E-C), 2.6 ns (T-C), and 2.7 ns (M-C).

We note in particular the sodium ions, which were initially added to the simulation box by randomly replacing water molecules, require at least several hundred picoseconds to form an ion atmosphere around DNA (Feig and Pettitt, 1999b). Preliminary studies from this laboratory (M. Feig and

TABLE 1 Parameters of simulations discussed in this paper

Name	Starting Form	Waters	Na^+	Cl^-	Box Size	Length
S-C	Stacked	1617	25	10	$3.7 \times 3.7 \times 4.0$ nm^3	10.1 ns
E-C	Looped-out	1945	25	10	$4.0 \times 4.0 \times 4.0$ nm^3	20.0 ns
T-C	Triple	1617	25	10	$3.7 \times 3.7 \times 4.0$ nm^3	2.6 ns
M-C	Major groove	1945	25	10	$4.0 \times 4.0 \times 4.0$ nm^3	2.7 ns

The simulations will be referred to by the abbreviation in the first column. The given length of the simulations is the total time, including all initial equilibration. The "stacked" and "looped-out" starting conformations refer to low-energy structures from continuum solvent modeling (Zacharias and Sklenar, 1997), the "triple" and "major groove" forms are stable conformations with the AMBER force field.

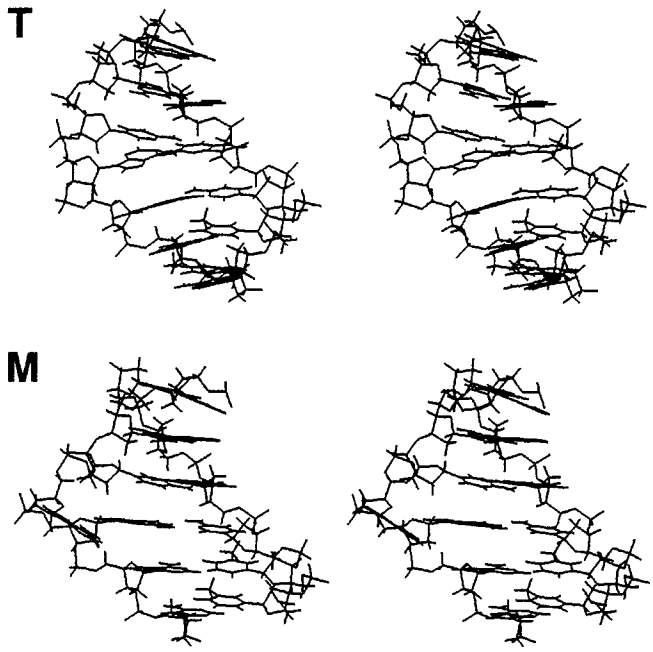

FIGURE 1 Stereo diagrams (wall-eyed) of average structures in T and M configurations from simulation with AMBER force field. Edge basepairs are omitted for clarity.

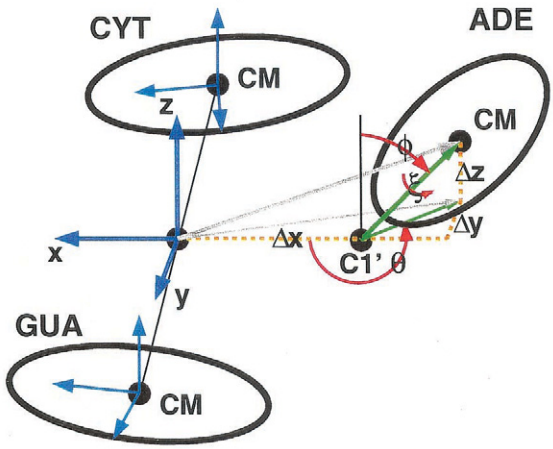

FIGURE 2 Parameter definition for specifying the bulge base orientation. A reference frame is constructed from the flanking basepairs (*blue*). The translational parameters Δx (slide), Δy (shift), and Δz (rise) describe the position of the center of mass (CM) of the bulge base. The rotational parameters θ and ϕ are defined for the orientation of the vector from the C1′ atom to the bulge center of mass. ξ measures the rotation of the base around this vector.

B. M. Pettitt, unpublished results) suggested an important role for ionic effects in stabilizing DNA with a bulge base in different conformations.

Previous experiences with DNA simulations under similar conditions (Feig and Pettitt, 1998b) suggest time scales of several nanoseconds before a dynamical and structural equilibrium between DNA and solvent is fully established in simulations started with completely unequilibrated water molecules and random ion positions. Full dynamical equilibria were found to be established after 3-4 ns. In this paper we will also only consider simulation data after 3 ns in the following analysis of the simulations S-C and E-C. For simulations T-C and M-C such long equilibration times are not necessary, because they were continued from structures that were already fully equilibrated with respect to the DNA-solvent interface. Considering only the first few hundred picoseconds as "equilibration time" appears to be sufficient for these trajectories.

All simulation runs were carried out in parallel on IBM SP-2 computers at San Diego Supercomputing Center and at the Texas Center for Computational and Information Sciences at the University of Houston.

Analysis

For the analysis of the backbone and geometry of the paired non-bulge bases, standard procedures were used (Saenger, 1984; Dickerson et al., 1989). As in earlier studies (Feig and Pettitt, 1998b) the NEWHEL93 algorithm (Dickerson, 1992) was chosen for the calculation of helical parameters because its use of a local helical axis as a reference frame seems appropriate for the analysis of short DNA fragments, as in this study.

The orientation of the adenine bulge base will be described by three translational and three rotational parameters, θ, ϕ, and ξ, as shown in Fig. 2. Their definition is explained in detail in the Appendix.

RESULTS

Bulge conformations

The two long simulations, S-C and E-C, generated a wide variety of structures covering stacked as well as partially and completely looped-out conformations, including some new minor conformations that have not been reported previously from experiment or modeling studies to our knowledge.

The most characteristic parameter that describes the conformational variation of the bulge base is the rotation angle θ introduced above that easily distinguishes between stacked and looped-out base conformations. For some conformations the bulge base tilts significantly toward the helical axis of the DNA. Therefore, the angle ϕ measuring this inclination angle represents another important quantity to describe the bulge conformation. Variations in the bulge displacements for fixed θ/ϕ values are relatively small because of the limited flexibility in the DNA backbone, and therefore less informative. If the pivoting of the bulge base around the glycosidic linkage (measured by ξ) is neglected, the angles θ and ϕ provide a good description of the bulge orientation and will be used in the following as the main quantities to distinguish different conformations.

Fig. 3 shows the sampling of θ and ϕ values during all simulations. Five main conformations can be distinguished where the base is either stacked (S), fully extended (E), forms a base triple in the minor groove (T), associated with the DNA in the major groove (M), or along the minor groove (G). T and M conformations are stable with the AMBER force field and represent the starting points for the T-C and M-C simulations. The S, E, and G conformations can be divided further into subconformations, denoted by S1, E1, etc., and are discussed in more detail below. The time series of θ and ϕ are given in Fig. 4 for S-C and E-C and in Fig. 5 for T-C and M-C. The stacked form is maintained throughout the whole simulation in S-C. By

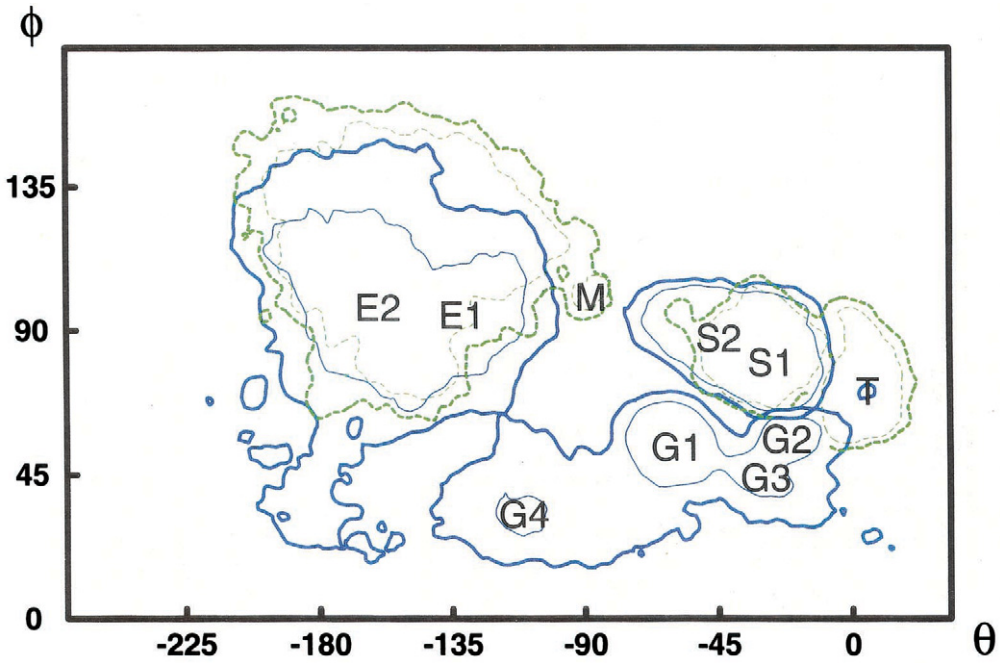

FIGURE 3 Extent of conformational sampling for different bulge orientations as measured by spherical coordinates θ and ϕ (see text) during simulations S-C, E-C (*blue*), and M-C, T-C (*green*). Contour lines are shown at population densities of $1.875 \cdot 10^{-5}$/deg^2 (*thick lines*) and $6.25 \cdot 10^{-5}$/deg^2 (*thin lines*). Labels for local minima are shown according to the text.

contrast, during the 20-ns E-C simulation the bulge first moves from the initial structure to the E1/E2 region remaining completely extended up to 10-ns simulation time. At that time a rotation around the backbone toward the minor groove is initiated that brings the bulge base near the triple form. The bulge base remains associated with the minor groove in different conformations G1–G4 until 16.5 ns, when it moves back into the extended E2 form. In the M-C simulation the bulge base leaves the major groove almost immediately and moves to the extended forms E1, and then E2. In simulation T-C the initial triple form is stable for a longer period of ~1 ns before a transition into the stacked conformation occurs. Simulations M-C and T-C indicate that the M and T starting conformations, preferred with the AMBER force field, are not particularly favorable conformations with the CHARMM27 force field, although the T conformation can be somewhat stable.

The population of θ and ϕ angles for the S and E conformations is shown at higher resolution in Fig. 6. The more detailed view shows separate minima in the E and S conformations, suggesting a distinction between subconformations S1/S2 and E1/E2. Because each of these contour diagrams is calculated from a single trajectory (S: S-C, E: E-C) they represent the actual free energy landscape within each conformation. Thus, it is possible to examine the free energy profile for transitions between subconformations

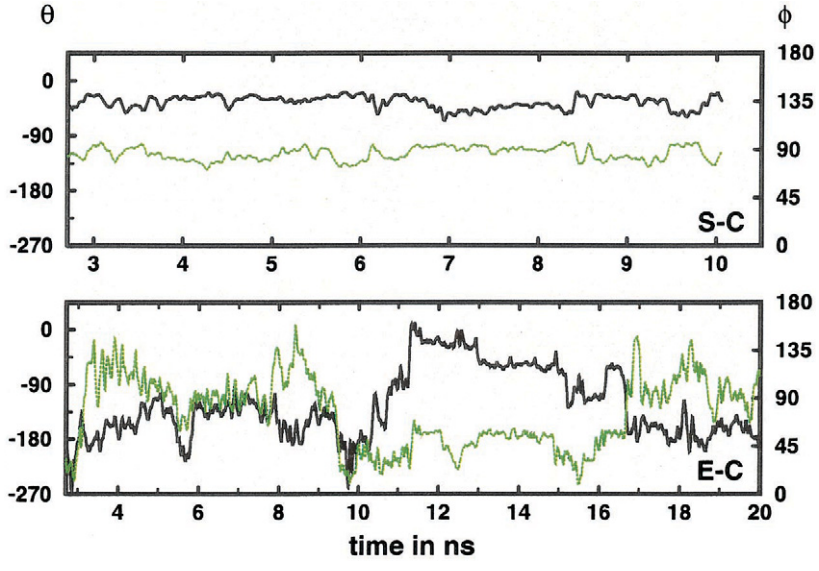

FIGURE 4 Time series of spherical coordinates θ (*black*) and ϕ (*green*) for simulations S-C and E-C.

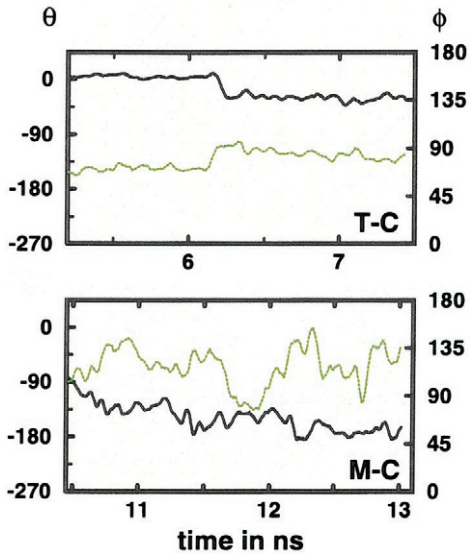

FIGURE 5 Time series of spherical coordinates θ (*black*) and ϕ (*green*) for simulations T-C and M-C.

S1/S2 and E1/E2. S2 has a relative free energy of ~0.2 kcal/mol compared to S1, with a separating barrier of ~0.4 kcal/mol. E1 and E2 differ by <0.1 kcal/mol with a barrier of ~0.2 kcal/mol. For comparison, the minimum energy structure for the S conformations from continuum solvent modeling (Zacharias and Sklenar, 1997) is marked in Fig. 6 and matches the energetically slightly higher S2 subconformation.

Using Fig. 4, time intervals have been determined during which the bulge base maintains a specific conformation. They are listed in Table 2 and will be used in the following for calculating average structural properties for each conformation.

The average three-dimensional structures in Fig. 7 give a first view of the structural features arising in different conformations. Not only the orientation of the bulge itself, but also variations in the surrounding DNA structure are interesting. In most structures the bulge introduces some degree of bending toward the strand without the bulge or toward the major groove. In the more extended conformations E, G, and M the backbone structure around the bulge base is altered significantly to accommodate different orientations of the bulge.

TABLE 2 Time intervals used for averaging over bulge conformations

Simulation	Conformation	Averaging Intervals (ns)
S-C	S1	3.9–4.4/4.6–5.0/8.7–9.3
	S2	7.1–8.2/9.5–9.7
E-C	E1	6.0–7.7
	E2	3.6–4.5/8.0–8.5/17.1–20.0
	G1	13.0–14.8/16.2–16.5
	G2	11.6–12.0
	G3	12.1–13.0
	G4	10.8–11.2/15.3–15.6

Looking at the structures in more detail, it is possible to induce some aspects of how they are stabilized to form distinct conformations. In the S conformation the bulge is sandwiched between the cytosine and guanine bases in 5′ and 3′ direction, respectively. However, rather than stacking equally with both bases, the bulge stacks preferentially either with the cytosine base in the S2 conformation or with the guanine base in the slightly lower energy S1 conformation.

A larger conformational space is accessible to the bulge base when it is fully extended. The average structures only reflect the midpoints of considerable tilting and swiveling around the glycosidic linkage. However, the backbone remains fairly rigid within each subconformation, effectively limiting and partitioning the conformational region that is accessible to the bulge base into distinct but energetically very similar subregions. In the E1 and E2 conformations different backbone structures cause the base to be oriented more along the major groove. In E2 the backbone follows an S-shape where the phosphate group between the (fourth) cytosine base and the bulge base is actually located further down along the helical axis than the next phosphate group between the bulge and the (sixth) guanine base. In the E1 conformation a similar, although less pronounced, effect is repeated one base further down toward the 3′ end with the phosphate group between the fifth and sixth base being almost on the same height as the phosphate group between the sixth and seventh base. The furanose ring is pulled down as a result and points toward the edge of the major groove. In the E-C simulation, transitions occur between the E2 and G1/G4 conformations. The G1 and G4 structures also have an S-shaped phosphate backbone very similar to the E2 structure, but the sugar ring at the bulge is rotated around the C3′-C4′ bond into the minor groove with the attached base oriented along the backbone in front of the minor groove. G1 and G4 are almost identical structures except for the orientation of the base itself. In the G1 conformation the amino group of the base points into the groove. Although the base is located >4.0 Å from the DNA, which is too far to form direct hydrogen bonding contacts, solvent molecules mediate the interaction with the second and third basepairs. The bulge base is positioned more outside of the groove in the G4 conformation oriented perpendicular to the G1 base orientation. The S-shaped backbone is unwound in the G2 and G3 conformations, pushing the furanose ring and the base further into the minor groove. As between G1 and G4, the difference between the G2 and G3 conformations is also the degree of association of the bulge base with the rest of the DNA. The interaction now occurs mostly with the opposite strand around the second and third basepairs, and in the more closely interacting G2 conformation a hydrogen bond is formed from the H61 atom to the sugar ring oxygen O4′ of the guanine in the second basepair (distance: 2.1 Å).

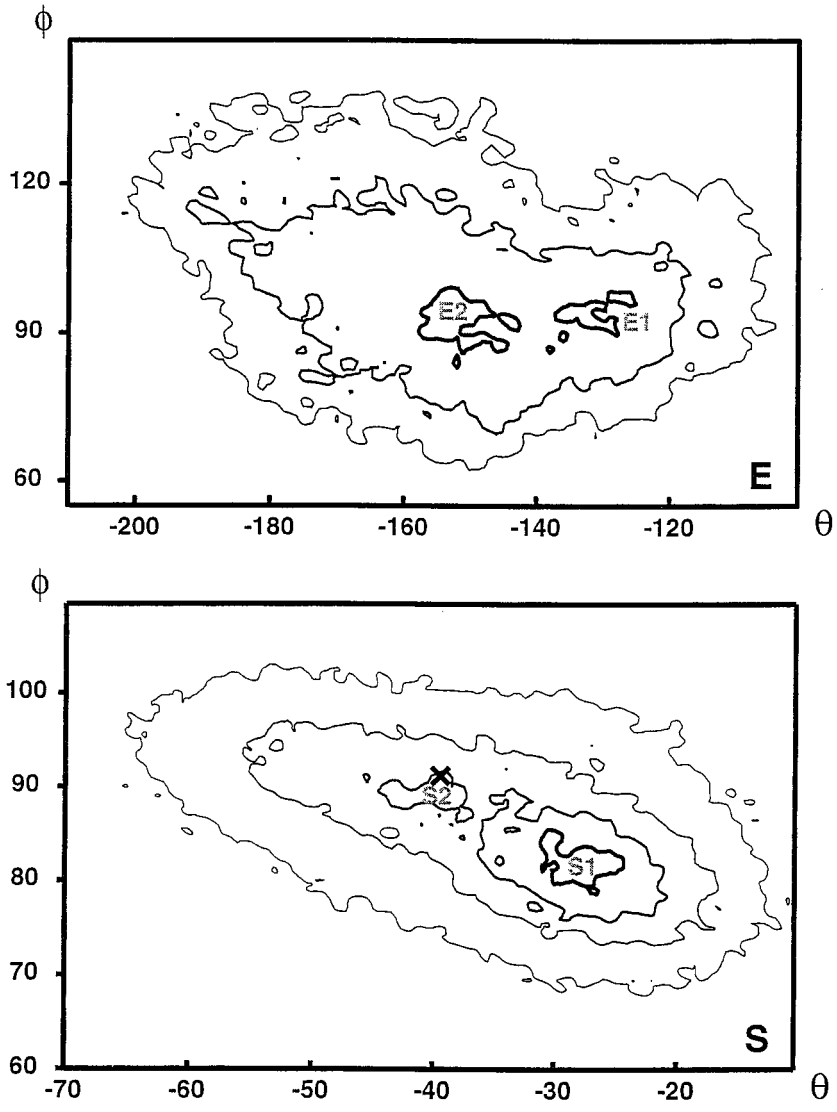

FIGURE 6 Population densities for stacked (S) and extended (E) bulge conformations measured by spherical coordinates θ and ϕ from simulations S-C and E-C, respectively. Contour levels are shown at $2.0 \cdot 10^{-4}/\text{deg}^2$, $8.0 \cdot 10^{-4}/\text{deg}^2$, $1.4 \cdot 10^{-3}/\text{deg}^2$, and $2.0 \cdot 10^{-3}/\text{deg}^2$ for S forms. The levels for E are $8.0 \cdot 10^{-5}/\text{deg}^2$, $1.6 \cdot 10^{-4}/\text{deg}^2$, and $3.0 \cdot 10^{-4}/\text{deg}^2$. Lowest energy stacked conformation from continuum solvent modeling (Zacharias and Sklenar, 1997) is indicated by $\times$.

Table 3 contains a compilation of average bulge parameters for each conformation to provide a more quantitative description. It is interesting to look at the fluctuations given in the table as well. As one would expect, they are generally much less for the conformations in which the bulge base is more closely associated with the rest of the DNA as in the S1, S2, G1, and G2 conformations. However, very large fluctuations in the ξ angle in G2 indicate that the base pivots quite extensively around its glycosidic linkage. The largest fluctuations are found for the extended conformations in the rotational and translational parameters.

More important than local dynamics within a conformation are transitions between the quite distinct main conformations that have occurred in our simulations giving rise to our proposed transition mechanism (see below). Most interesting is the almost completed transition from E1/E2 through G4/G1 and G3/G2 to the T form that provides a continuous pathway between the stacked and extended forms together with the transition from T to S1/S2 implied by simulation T-C. Fig. 8 shows snapshots along the pathway during the E-C simulation from an E2 structure to a G2 conformation closest to the T form. The transition occurs by a rotation around the backbone into the minor groove leading to a vertical position parallel to the helical axis and then a tilt back to a perpendicular orientation inside the groove.

Intuitively, one might except a movement where the base remains mostly perpendicular to the helical axis throughout the rotation around the backbone and then slides directly into the triple form, but the simulation indicates that such a pathway cannot be easily accommodated by the backbone. The backbone also imposes a barrier to the remaining transition from G2 to T that could not be observed and will be discussed below. Another important aspect is the role of sodium ions and coordinated water molecules that appear to stabilize the G conformations and may also prevent a tran-

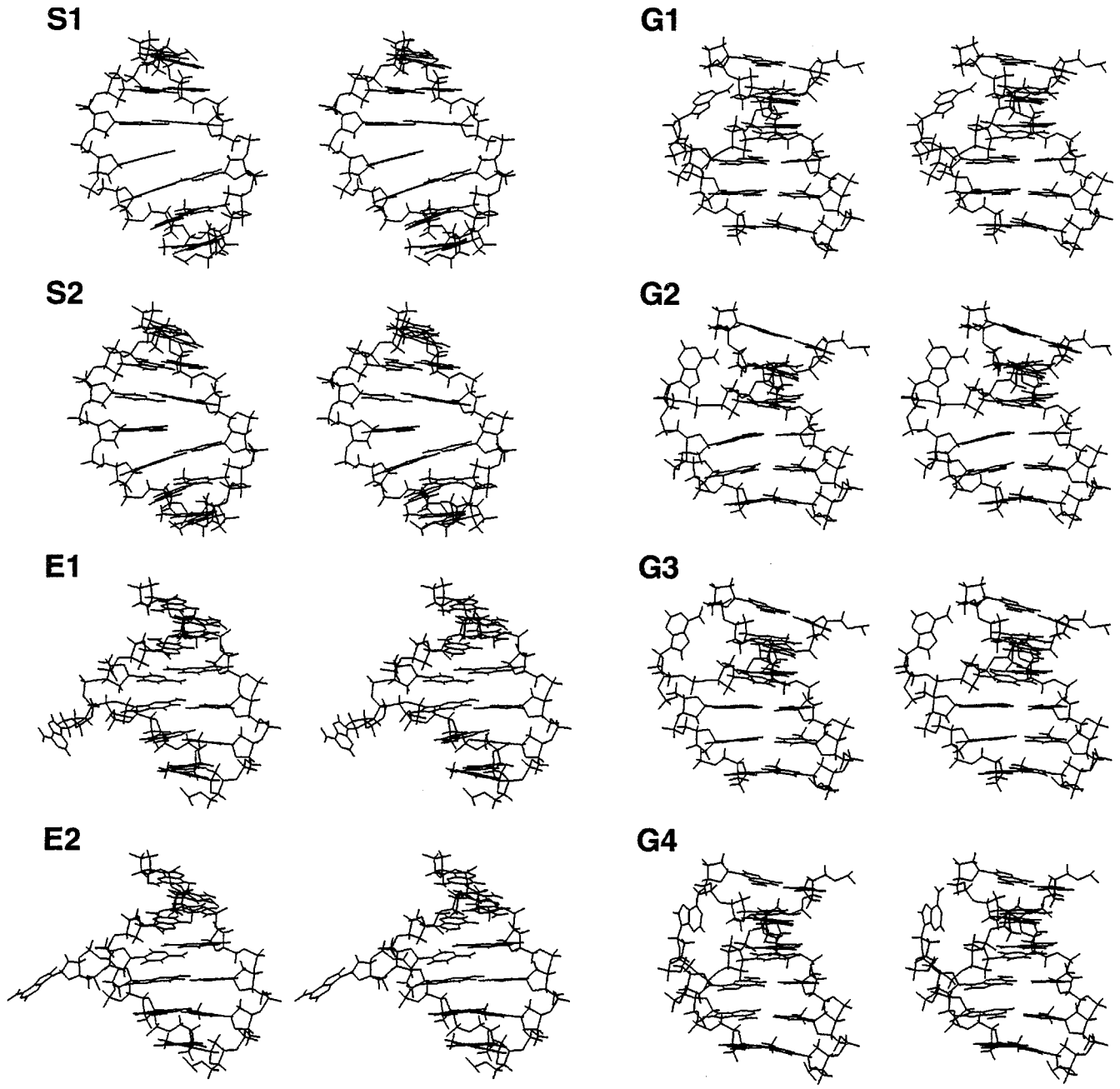

FIGURE 7 Stereo diagrams (wall-eyed) of average structures in S, E, and G configurations. Edge basepairs are omitted for clarity.

sition into the T form in the simulation. Fig. 9 shows the solvent arrangement at the last snapshot closest to the triple conformation in Fig. 8. The base is sandwiched between two sodium ions and their solvation shells, one of which occupies the space where the bulge base would have to move to in order to assume the triple conformation. Similarly strong solvent interactions are also found for the other G conformations. At some times even direct ion-base associations have been observed. A more detailed analysis of the role of solvent could provide additional insight into the stabilization of different bulge conformations and would go beyond the scope, but will be discussed in a future paper.

Helix structure

After discussing the simulated structures of the bulge base with respect to the DNA helix, we will now focus on how the structure of the rest of the DNA is influenced by the presence of a bulge base. The most obvious effect is bending of the DNA helix. Table 4 shows the magnitudes and directions of bending at the bulge site calculated from roll

TABLE 3 Average bulge parameters for bulge conformations as defined in Fig. 1 accumulated over time intervals according to Table 2

	θ	ϕ	ξ	Δx	Δy	Δz
S1	−27.4 (6.97)	80.6 (5.49)	−9.6 (8.34)	−0.05 (0.53)	0.05 (0.64)	−0.06 (0.20)
S2	−45.6 (8.55)	90.9 (5.34)	−3.2 (8.09)	−0.12 (0.54)	0.17 (0.59)	0.09 (0.18)
E1	−127.5 (15.28)	92.0 (13.40)	−46.1 (34.33)	−11.31 (1.43)	−1.95 (1.48)	−3.27 (1.62)
E2	−164.7 (18.50)	106.6 (21.02)	−183.3 (60.44)	−11.91 (0.87)	3.94 (2.02)	−0.79 (2.97)
G1	−59.0 (7.50)	55.9 (6.42)	−161.2 (8.85)	−6.93 (0.64)	1.17 (0.83)	5.41 (0.58)
G2	−18.8 (7.32)	57.8 (4.15)	−77.0 (91.78)	−3.09 (0.59)	5.47 (0.64)	5.64 (0.46)
G3	−23.9 (26.61)	40.2 (9.60)	−100.5 (36.81)	−5.51 (1.13)	4.97 (1.06)	6.28 (0.63)
G4	−91.7 (27.58)	28.8 (12.01)	−61.8 (81.59)	−8.03 (1.22)	2.25 (1.18)	6.81 (1.04)
T	2.5 (5.82)	76.1 (7.34)	8.3 (12.20)	−0.42 (0.42)	4.62 (0.42)	1.53 (0.48)
M	−81.6 (12.03)	105.7 (13.28)	−3.3 (37.09)	−6.54 (0.98)	−0.82 (0.86)	−2.44 (1.04)

Displacements are given in angstroms. Standard deviations are included in parentheses. For comparison, values for average structures (T and M) from simulations with the AMBER force field (M. Feig and B. M. Pettitt, unbpublished results) are included as well.

and tilt angles between the bulge-enclosing basepairs. For the S conformations the bulge acts as a wedge, causing significant bending mostly directed toward the opposite strand (90°) and slightly into the major groove (0°). For the conformations E1, G2, and G3 with quite different bulge orientations, the DNA is bent moderately toward the major groove. The remaining E2, G1, and G4 conformations result in mostly straight DNA helices.

Tables 5 and 6 show average rise and twist values along the helix depending on the bulge conformation. Most of the deviations from canonical geometries occur for the C · G/G · C base step flanking the bulge base and the previous base step at basepairs three and four. In some cases they extend to the base steps at basepairs two and three as well as five and six. In the S conformation the presence of the bulge base causes significantly larger twist values around the bulge. An increased distance between basepairs is necessary to accommodate a stacked basepair, while the other base steps exhibit fairly regular DNA structures. However, although twist values are similar, the helical rise is consistently lower for the S2 conformation than for S1 along the helix in a range that is more typical of A-DNA than B-DNA. This coincides with a larger fraction of A-type C3′-endo furanose ring conformations for S2 than for S1 around the bulge base. The extended conformations E1 and E2 have only a small effect on the DNA structure that involves an increase in rise and an increase in twist at the following base step starting at basepairs three (E2) and four (E1), respectively. All of the G conformations have increased twist angles around the bulge base. Rise values are fairly typical of regular DNA with the exception of G2, where the fourth basepair (C · G) is moved closer to the third basepair (G · C). The fluctuations provided in Tables 5 and 6 give an indication of the flexibility of DNA at different conformations. Higher degrees of flexibility are found for S1 and S2 and for E2. However, the G conformations are generally more rigid.

Backbone structure

As indicated above, the conformational freedom of the bulge base is determined to a large extent by the structure of

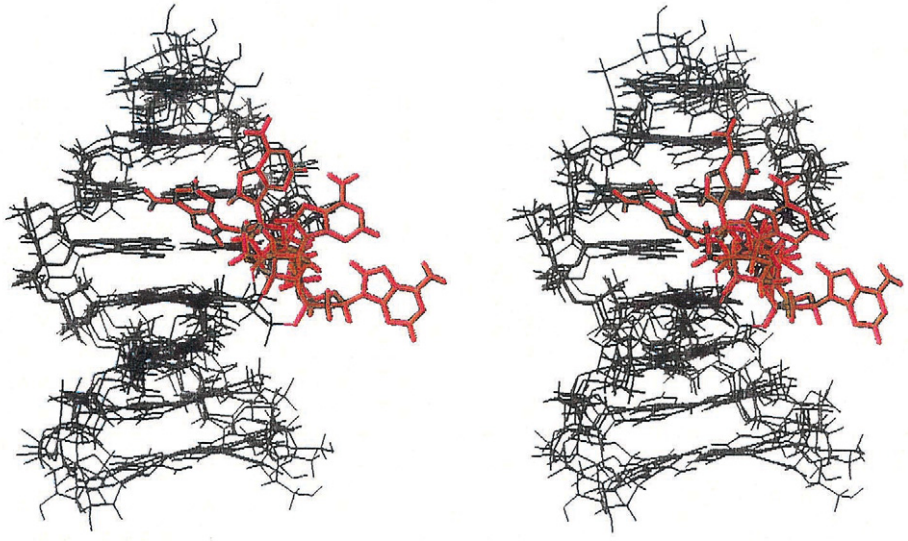

FIGURE 8 Stereo diagram (wall-eyed) of transition in simulation E-C from the looped-out conformation (E) into minor groove near the triple conformation (T). Snapshots are shown at 9.3, 9.5, 9.8, and 11.4 ns.

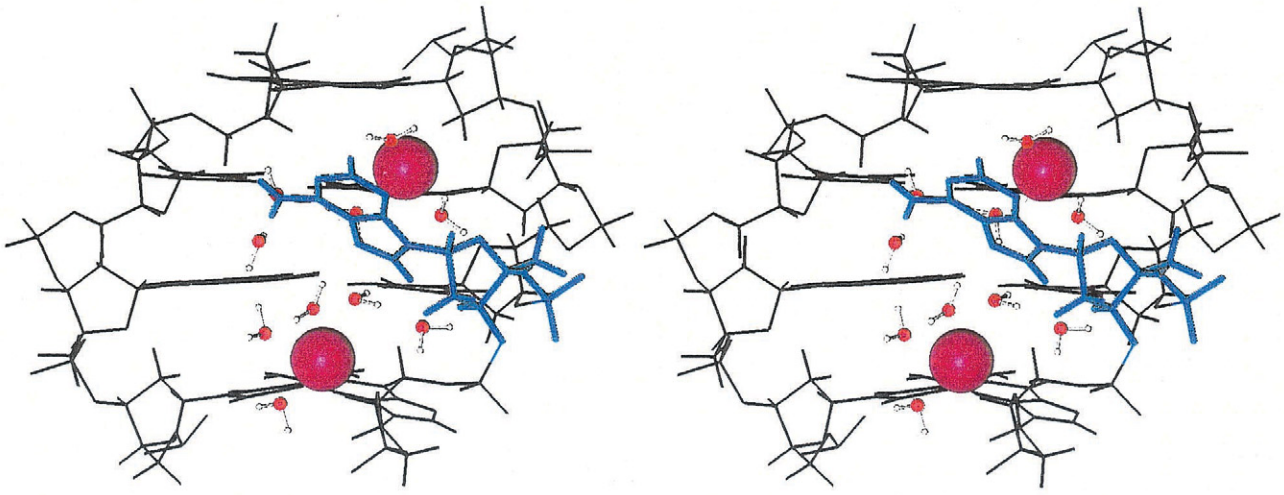

FIGURE 9 Stereo diagram (wall-eyed) of water (*red/white*) and sodium ions (*magenta*) in minor groove at 11.4 ns during simulation E-C. The bulge base near the triple conformation is shown in blue.

the backbone around the bulge, because the backbone fixes the orientation of the sugar ring where the bulge base is attached through the glycosidic bond. Most of the flexibility in the DNA backbone is realized by transitions involving the ϵ-ζ dihedral pair and sugar puckering between B-type C2′-endo/C1′-exo/O4′-endo and A-type C3′-endo conformations. Table 7 shows the main populations of furanose ring conformations at the bulge and surrounding bases. The sugar pucker is found mostly in B form between O4′-endo and C2′-endo, as would be expected for a B-DNA helix. Major populations of the A form C3′-endo sugar pucker occur during S and G conformations. In the G1, G2, and G3 structures the sugar at the bulge base adopts the C3′-endo or C2′-exo forms almost exclusively, but remains in C1′-exo/C2′-endo in the G4 structure. In the S2 structure, both the bulge base and the previous cytosine base predominantly populate C3′-endo conformations, while at the same time the rise is also reduced to values typical of A-DNA, as discussed above. However, the sugar at the bulge base is in B form in the S1 structure, while the flanking cytosine base still has 25% population in the C3′-endo range. Interestingly, a significant C3′-endo population is also found at the cytosine base of the fifth basepair on the opposite strand.

TABLE 4 Bend angle θ and orientation ϕ relative to major groove calculated from averge roll and tilt angles

	Bend Angle θ	Direction ϕ
S1	17.9	76
S2	24.8	64
E1	11.4	358
E2	8.6	320
G1	7.3	286
G2	14.3	10
G3	13.1	353
G4	5.7	295

Angles as follows (Zhurkin, 1985; Olson et al., 1988; Young et al., 1995): $\theta = \sqrt{\rho^2 + \tau^2}$; using $\alpha = \tan^{-1}(\tau/\rho)$, for $\rho > 0$ and $\tau > 0$: $\phi = \alpha$, for $\rho > 0$ and $\tau < 0$: $\phi = 360 + \alpha$, for $\rho < 0$: $\phi = 180 + \alpha$.

The dihedral angles ϵ and ζ show the most complex variation depending on the bulge conformation. Usually ϵ and ζ are highly correlated (Dickerson, 1983) with typical values of $\epsilon = 185°$, $\zeta = 265°$ (B_I), $\epsilon = 245°$, $\zeta = 175°$ (B_{II}), and $\epsilon = 205°$, $\zeta = 290°$ (A) (Schneider et al., 1997). Due to the presence of the bulge base significant deviations from canonical B_I conformations are apparent only at the phosphates immediately flanking the bulge base, and in some cases also at the next phosphate group in 3′ direction between the sixth and seventh base (GpA). Because ζ measures the torsion along the atoms C3′-O3′-P-O5′, compared to ϵ measuring the torsion for C4′-C3′-O3′-P, it is better suited to distinguish conformational variations of the phosphate backbone from the furanose pucker to the extent that they are not directly correlated. A clearer picture of the variation of ζ before and after the bulge base depending on the bulge conformation can be obtained by mapping ζ values onto θ and ϕ. The resulting pictures are shown in Fig. 10. It can easily be seen that transitions between bulge conformations coincide with distinctively different backbone conformations as measured by ζ in at least either above or below the bulge base. S1 and S2 conformations are distinguished by a region where ζ assumes values around 160° (B_{II}) at the ApG base step, occupying most of the conformational space covered by S1. In the T form the ApG base step has ζ values around 280°, but the CpA step is in B_{II} form. A transition from T to G2 would maintain the backbone at the CpA step, but require a transition in the ApG step to ζ values around 360° to accommodate the corresponding rotation of the bulge base. G1 and G4 structures are distinguished from G2 and G3 by a reduction to 90° in ζ at the CpA step, while an equilibrium between ζ =

TABLE 5 Average values for rise in Å between basepairs calculated with NEWHEL93 (Dickerson, 1992) and accumulated over time intervals according to Table 2

	C · GpG · C	G · CpC · G	C · G*G · C	G · CpA · T	A · TpA · T
S1	3.10 (0.49)	3.08 (0.51)	5.54 (0.50)	3.01 (0.49)	3.26 (0.40)
S2	2.94 (0.65)	2.83 (0.44)	5.08 (0.55)	2.92 (0.44)	3.07 (0.35)
E1	3.47 (0.38)	3.28 (0.30)	3.81 (0.39)	3.29 (0.36)	3.25 (0.32)
E2	3.52 (0.51)	3.71 (0.53)	3.24 (0.53)	3.31 (0.43)	3.17 (0.36)
G1	3.42 (0.33)	3.94 (0.29)	3.05 (0.31)	3.38 (0.30)	3.40 (0.33)
G2	3.69 (0.32)	2.79 (0.36)	4.32 (0.26)	3.22 (0.33)	3.09 (0.26)
G3	3.57 (0.31)	3.22 (0.36)	3.58 (0.46)	3.31 (0.36)	3.07 (0.29)
G4	3.90 (0.45)	3.63 (0.27)	3.44 (0.31)	3.32 (0.34)	3.29 (0.27)

Standard deviations are included in parentheses.

320° and $\zeta = \sim 30°$ at the ApG step is sampled for both structures. In the extended conformations ζ values are predominantly found around 60° at the ApG step, whereas the backbone at the CpA correlates with the bulge orientation. For E2 it populates two states at 121° and 156° (B_{II}), and for E1 it is at 290° in the A-DNA range.

Comparison with experimental data

After characterizing the observed structures in detail, we will now compare the results from the simulation to available experimental data. The most important benchmarks are NMR measurements that have been performed on the same sequence in a similar solution environment (Hare et al., 1986; Nikonowicz et al., 1989). Using their published proton distances or distance constraints from NOE for the bulge base and the two flanking basepairs on each side, root-mean-square (RMS) and maximum violations have been calculated for the average structure of each conformation (Table 8). Results for some of the lowest energy conformations from continuum solvent modeling (Zacharias and Sklenar, 1997) as well as the M and T conformations from the AMBER force field simulations are also included for comparison. It should be noted that Hare et al. have published interval distance constraints, while Nikonowicz et al. only report single values. The RMS deviations are calculated from the violations outside the specified intervals for the data by Hare et al., but with respect to the distance from the single values provided by Nikonowicz et al. As a consequence, the RMS values for the data by Hare et al. in the first two columns are much smaller than for the data by Nikonowicz et al. Not surprisingly, the extended E, M, and G conformations violate the experimental distance constraints significantly. The T conformations agree reasonably well, only if the distances between exchangeable protons are disregarded, but show large maximum deviations of several angstroms otherwise. Confirming the previous conclusions from the NMR data (Hare et al., 1986; Nikonowicz et al., 1989) that the bulge base is stacked within the helix, the agreement for the S conformations is good. The best match with experimental data is found for the S1 conformation. Average violations of ~0.04 Å with respect to the constraint intervals given by Hare et al. and average deviations of 0.2 Å from the distances by Nikonowicz et al. with maximum deviations of <1 Å from both experimental data sets are considered very good results from free simulations without restraints. For comparison, typical violations of NOE distance restraints in *restrained* simulations of nucleic acid structures are 0.03–0.05 Å on average, with maximum violations of 0.3–0.5 Å (Dornberger et al., 1998). Very encouragingly, S1 also coincides with the lowest energy conformation in the S-C and T-C simulations.

The comparison with crystallographic data on bulge-containing DNA is less straightforward because interactions of the bulge bases with neighboring DNA molecules may influence the reported structures. This appears to be the case particularly for the structure by Joshua-Tor et al. (1992) where the looped-out base from one helix stacks into the

TABLE 6 Average values for twist angles between basepairs calculated with NEWHEL93 (Dickerson, 1992) and accumulated over time intervals according to Table 2

	C · GpG · C	G · CpC · G	C · G*G · C	G · CpA · T	A · TpA · T
S1	38.6 (4.9)	32.5 (4.9)	51.8 (5.9)	34.2 (4.7)	34.9 (4.5)
S2	39.5 (6.4)	33.5 (3.8)	52.4 (5.4)	31.8 (4.1)	35.2 (3.7)
E1	38.3 (4.0)	35.2 (4.2)	38.9 (5.4)	40.1 (4.3)	36.6 (3.6)
E2	38.0 (4.4)	27.8 (6.6)	43.8 (4.7)	37.3 (4.0)	37.4 (3.6)
G1	40.1 (3.5)	25.9 (4.1)	49.7 (3.6)	38.7 (3.5)	35.8 (3.9)
G2	35.6 (3.5)	36.7 (3.0)	41.6 (3.6)	39.1 (3.9)	38.9 (2.5)
G3	37.3 (3.5)	34.4 (3.5)	44.1 (4.2)	38.5 (3.3)	37.2 (3.6)
G4	36.0 (5.1)	29.5 (3.3)	45.6 (4.2)	38.8 (4.4)	37.9 (3.1)

Standard deviations are included in parentheses.

TABLE 7 Predominant furanose pseudo rotation angle populations during time intervals according to Table 2

	C	A	G	C	G	
S1	C2′-endo C3′-endo	C2′-endo C1′-exo	C2′-endo	C2′-endo C3′-endo	C2′-endo	
S2	C3′-endo C2′-endo	C3′-endo	C2′-endo C3′-exo	C2′-endo	C2′-endo	
E1	C2′-endo	C1′-exo C2′-endo	O4′-endo C2′-endo C1′-exo	C2′-endo	C2′-endo	
E2	C1′-exo	C2′-endo	C2′-endo C1′-exo	C1′-exo	C2′-endo	C2′-endo
G1	C2′-endo C1′-exo	C3′-endo C2′-exo	C2′-endo C1′-exo	C2′-endo	C2′-endo	
G2	C2′-endo	C3′-endo	C1′-exo	C2′-endo	C2′-endo	
G3	C2′-endo	C3′-endo	O4′-endo C1′-exo C2′-endo	C2′-endo	C2′-endo	
G4	C2′-endo C1′-exo	C1′-exo C2′-endo	C1′-exo C2′-endo	C2′-endo	C2′-endo	

C2′-exo: −36–0, C3′-endo: 0–36, O4′-endo: 72–108, C1′-exo: 108–144, C2′-endo: 144–180, C3′-exo: 180–216.

neighboring helix and it is less obvious to what extent such results are applicable to non-interacting DNA helices. Nevertheless, quite good agreement is found between the structure by Joshua-Tor et al. with an adenine bulge in a similar sequence context as in our simulations and the E2 conformation. The comparison finds that the backbone conformation is essentially the same, although the orientation of the bulge base deviates, most likely due to particular stacking interaction with a neighboring helix that is not present in the simulation.

Similar conformations are also found for looped-out adenine bulges in RNA structures. The bulge in the phage MS2 replicase gene operator RNA bound to its coat protein is close to the E2 conformation (Valegard et al., 1997) and an adenine bulge in the P4-P6 domain of group I self-splicing intron RNA from *Tetrahymena thermophila* (Cate et al., 1996). Another example of an extended adenine bulge conformation in the crystal environment is found in a recent investigation of a DNA-RNA hybrid (Sudarsanakumar et al., 2000), but most interesting is experimental evidence for G-type bulge conformations as well. An adenine bulge in the first stem loop of the SL1 RNA of *Caenorhabditis elegans* has a conformation similar to G2 (Greenbaum et al., 1996), while in a structure of the HIV-1 genomic RNA dimerization initiation site a bulge conformation close to G1 is reported (Ennifar et al., 1999).

DISCUSSION

Although many computer simulations have been reported for canonical DNA, only a few simulations are available for non-canonical nucleic acid structures important in a variety of biological processes. To our knowledge, no simulations have been reported for bulged DNA or RNA structures despite their biological relevance, e.g., in DNA mismatch repair and RNA-protein and RNA-RNA interactions. Experimental data on DNA and double-stranded RNA molecules with single bulge bases distinguish between structures where single bulge bases are either stacked in with the rest of the helix or completely looped-out. Which conformation is found depends on the base type and the environment and/or the experimental technique used for structure determination. No clear picture is available from experiment how bulge-containing nucleic acid structures interact with proteins. The conformational flexibility of bulge bases, and in particular the transition from stacked conformation apparently favored in solution by purine bases to an extended conformation in nucleic acid-protein complexes, is only poorly understood.

The simulations presented here offer a more detailed view of the complex range of conformations that may be accessible to bulge-containing nucleic acid helices at room temperature in an aqueous solution environment and suggest a transition pathway between the stacked and extended form through the minor groove. In addition to the stacked (S) and fully extended (E) conformations that agree well with NMR (stacked form) and crystallographic data (extended), other conformations were identified in which the bulge base associates with the nucleic acid structure in the minor groove to varying extents. Some of these conformations were already suggested from earlier continuum solvent modeling (Zacharias and Sklenar, 1997), but our simulations provide a more detailed picture and have also revealed completely new possible conformations.

Both the S-C and T-C simulations converge toward the S1 conformation, which is in excellent agreement with available NMR distance restraints, in particular when considering that no restraints were used during our simulations.

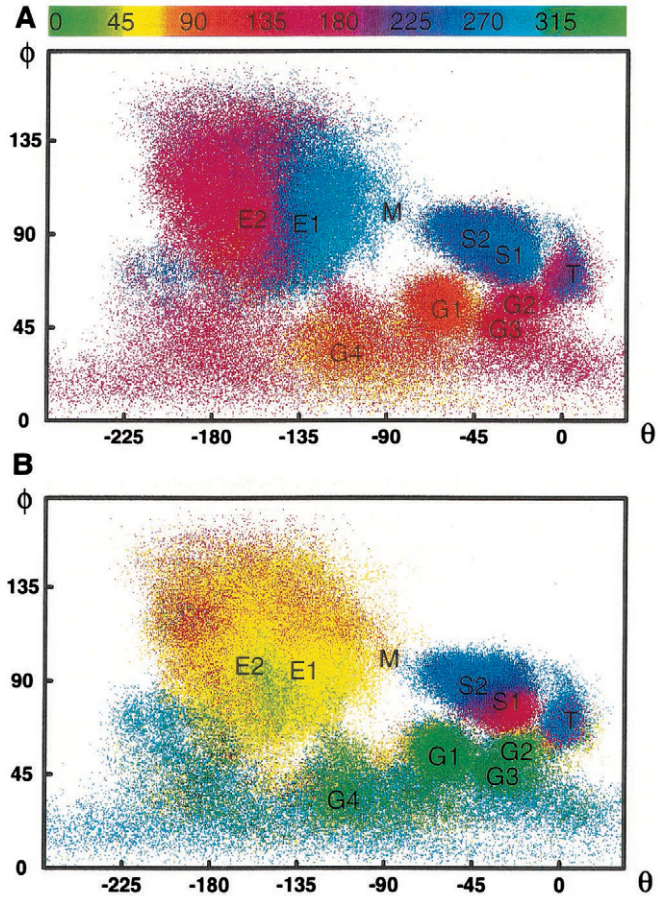

FIGURE 10 Distribution of backbone dihedral angle ζ at the bulge base depending on bulge orientation measured by spherical coordinates θ and ϕ, as in Fig. 3. ζ values are color-coded for CpA (*A*) and ApG base steps (*B*) surrounding the bulge base according to the color bar shown, and accumulated from simulations S-C, E-C, T-C, and M-C.

At the same time, the S1 conformation also exhibits the degree of bending expected from gel retardation experiments. Because the extended E and G conformations observed in the simulations E-C and M-C violate NMR distance constraints significantly, they are not expected to be populated >10% of the time to be consistent with NMR data. They correspond, however, to similar structures found in the crystal environment where the extended bulge conformations appear to be stabilized by interactions with other molecules and/or by the crystal environment. Although these conformations are stable on the relatively short nanosecond time scale accessible in simulations, we assume that these conformations are energetically higher than the stacked conformation, as concluded from previous modeling studies. We therefore expect a complete transition toward the stacked form if sufficient time is given to cross the kinetic barrier between the extended and stacked form. Due to the apparent involvement of extended bulges in protein interactions, it may be more important, though, in the biological context to gain a good understanding of the energetic landscape of extended bulge conformations and possible transition pathways than the favored stacked solution conformation.

In this respect the simulations presented here offer new insight not available from previous experimental or theoretical studies. In particular they provide the first direct evidence of transitions between different bulge conformations on the nanosecond timescale. Transitions are observed from E1/E2, through G4, G1, G3, to G2, and from T to S. Although a complete transition between the stacked and extended forms has not been observed in a single simulation, we deduce from our simulation data that the most likely pathway includes a transition between the G2 and T conformations that are structurally quite close. Although there is no steric hindrance for a G2 ↔ T transition, the simulation data indicate that such a transition is apparently limited mostly by a barrier in the backbone around the ζ dihedral that hinders the required conformational rearrangement, but ionic solvent effects that stabilize G conformations in particular seem to play an important role as well. An alternative transition pathway through the major groove would be conceivable in theory, but is not supported by the simulation data. The pronounced bending of the DNA helix toward the major groove with the stacked bulge base imposes steric hindrance for opening of the bulge base into the major groove, while the widened minor groove is easily accessible from the S1 conformation.

The simulations demonstrate that the observed transitions between the distinct bulge conformations may occur readily without significantly disrupting the surrounding structure of the double-stranded helix, and the same appears to be the case for the suggested G2 ↔ T transition. Thus, different bulge conformations are accessible after formation of a double helix without major structural rearrangements or temporary dissociation of both strands. Although the simulations here do not provide sufficient sampling to calculate reliable relative free energies or populations between the distinct conformational families, the direction of the observed transitions combined with the fact that transitions could be observed on nanosecond simulation timescales allows a tentative sketch of a complete free energy profile as shown in Fig. 11. Although a more accurate study of the transition path is in progress and will be reported at a later time, the available data from this simulation combined with experimental evidence already allows an estimate of relative free energies. The free energy of the S conformations appears to be lower than the T conformation by at least 2-3 kcal/mol, since a T → S transition was observed within ~1 ns in simulation T-C, but the reverse S → T did not occur during 10 ns in simulation S-C. S1 was found to be slightly lower than S2. Within the extended conformations E and G smaller transitional barriers seem to be present between E2 and G4/G1, G4/G1, and G2/G3, and between E1 and E2. Because it takes ~10 ns until the E2 → G4 transition

TABLE 8 Root-mean-square and maximum deviations in average simulated bulge conformations from NOE-derived distances (Nikonowicz et al., 1989) or distance intervals (Hare et al., 1986)

Conformation	NOE Violation (Å)					
	Hare et al.†		Hare et al.‡		Nikonowicz et al.§	
S1	0.035	0.90	0.037	0.90	0.195	0.80
S2	0.047	1.49	0.050	1.49	0.287	1.48
I-1*	0.066	1.45	0.070	1.45	0.241	1.19
E1	0.245	9.75	0.115	4.61	0.896	4.91
E2	0.294	12.55	0.098	3.60	0.823	3.74
G1	0.226	10.49	0.091	3.57	0.811	4.19
G2	0.204	8.98	0.104	4.46	0.919	4.92
G3	0.237	10.99	0.099	4.19	0.884	4.59
G4	0.236	9.32	0.152	6.32	1.128	5.94
T¶	0.105	4.34	0.051	1.21	0.184	0.90
IIa-1*	0.063	2.18	0.053	1.17	0.230	1.23
IIa-2*	0.104	5.12	0.057	1.44	0.225	1.07
M¶	0.113	3.40	0.088	2.79	0.571	3.09

*Lowest energy conformations from continuum solvent modeling (Zacharias and Sklenar, 1997).
†Sixty-two nonexchangeable and exchangeable proton distance constraints for bulge base and two flanking basepairs on each side (Hare et al., 1986).
‡As above but excluding exchangeable proton distance constraints.
§Eight NOE-derived distances involving bulge base (Nikonowicz et al., 1989).
¶Converged structures from simulations with AMBER force field.

occurs, but only ~3 ns for the reverse G4/G1 → E2 transition, E2 is expected to be lower in free energy than G4/G1. G4/G1 and G2/G3 should have similar free energies because they readily interconvert. The relative free energy difference between G2 and T or extended and stacked conformations cannot be estimated from the simulations. However, experimental evidence from NMR measurements that only observe stacked conformations imposes a population limit of <10% for other conformations. This translates into an energetic difference of at least 2-3 kcal/mol between the lowest-energy S1 and E2 conformations. A minimum barrier height of several kcal/mol is estimated in either direction of the G2 ↔ T transition because neither transition was observed in the simulations. Finally, the M conformation has a higher free energy relative to the E conformations because the M → E transition occurs rapidly in simulation M-C. Following this rough estimate, the rate-limiting barrier would be >4 kcal/mol relative to S1 between T and G2. It should be emphasized again that this suggested free energy profile is by no means quantitative, but should give a useful first qualitative estimate that will be improved in future studies by other computational techniques that would go beyond the range of this paper.

Another interesting aspect that can be addressed in more detail with explicit solvent simulations is the influence of solvent on different bulge conformation and possible implications for a change in equilibrium under different solvent conditions. From a first analysis of water and ion distributions it appears that ionic effects play a role in the stabilization of some G conformations possibly also inhibiting the completion of the G2 → T transition. If this is the case, increased salt concentrations might cause higher populations of extended E and G conformations. Indirectly, increased salt concentrations are also expected to favor S2 over S1, because S2 exhibits A-DNA features to a larger extent than S1, and A-DNA is preferentially stabilized under high salt conditions. However, another difference between S1 and S2 is given by stacking with either the base above or below. Because stacking interactions depend on the involved base types (Norberg and Nilsson, 1995a, b), the surrounding sequence should affect the equilibrium between S1 and S2 as well.

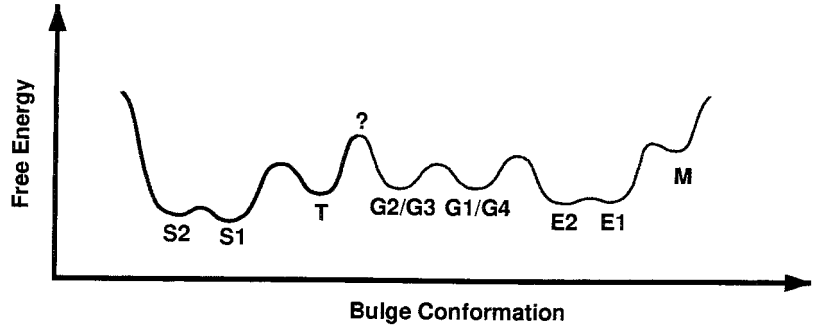

FIGURE 11 Suggested free energy profile between bulge conformations.

Biological implications

The conformational landscape depicted above consists of several well-defined main conformations, some of which having distinct subconformations, mainly as a consequence of discrete backbone states. The information available from the data presented here gives a comprehensive view of the conformational freedom of a free nucleic acid with a single bulge base. The related energetic aspects are important in understanding formation of bulge-specific protein-DNA and protein-RNA complexes.

The mechanism of natural mismatch repair for unpaired nucleotides is provided by an interplay among bacterial MutH, MutL, and MutS proteins (Modrich, 1991; Kolodner, 1995) or eukaryotic homologs (Kolodner, 1996; Modrich and Lahue, 1996). The repair mechanism is initiated through MutL/MutH after specific structural recognition of bulge-containing DNA fragments by MutS independent of the bulge base type or surrounding sequence, although NMR experiments indicate that stacked and looped-out conformations are populated differently, depending on the bulge base and sequence context (Kalnik et al., 1989a, 1990; Morden et al., 1990). Unless the repair mechanism can recognize extra bases in both conformations, it would require transitions to a common conformation before it is recognized by the repair enzyme. A transition from the stacked to the extended form is also implied for a phage MS2 stem-loop RNA fragment at the replicase gene translational repression operator and initiation site. An extra adenine bulge within this structure is found to be stacked within the stem helix (Borer et al., 1995; Kerwood and Borer, 1996; Smith and Nikonowicz, 1998) in solution but looped-out when bound to the MS2 coat protein for translational repression (Valegard et al., 1994, 1997; Rowsell et al., 1998). Crystal structures of a pyrimidine dimer specific excision repair enzyme (Vassylev et al., 1995) and a bacterial methylase (Reinisch et al., 1995) complexed with DNA indicate that the transition from a regular stacked form to an extended conformation is also interesting in different contexts.

It is also interesting to compare the results found here to base flipping in regular double-stranded DNA upon methyltransferase binding that requires a completely flipped-out base for the methylation reaction to take place. Crystallographic data are available for complexes of DNA with cytosine-5 *Hha*I methyltransferase (O'Gara et al., 1996; Kumar et al., 1994; Klimasauskas et al., 1994). The data indicate that the methyltransferase makes contact with DNA in the major groove while the flipped-out base is oriented somewhat toward the minor groove, suggesting a transition pathway through the minor groove. However, modeling studies of base opening in enzyme-free canonical B-DNA have identified only major groove pathways to be energetically feasible (Ramstein and Lavery, 1988; Chen et al., 1998). Enzyme participation has been suggested to allow a minor groove base opening pathway, but no detailed model is available. Within the limitations that a complementary base poses additional constraints on basepair opening pathways, our results for an unpaired base may suggest that a minor groove pathway similar to the one described here could become favorable if the DNA helix is transiently bent toward the major groove.

The results obtained here for an extra unpaired base in double-stranded DNA are expected to be transferable to unpaired bases in RNA structures, at least to some degree. As reviewed in more detail in the Introduction, experimental evidence finds for the most part very similar structures of DNA and RNA with single bulge bases under the same environmental conditions. For adenine bulges, stacked conformations are preferred in solution, while extended conformations occur in the crystal environment and/or complexed with proteins. In general, this suggests similar energetics with respect to bulge conformations in DNA and RNA. Projection of the proposed transition pathway between the stacked and extended form from the DNA simulations into an RNA context introduces no apparent steric hindrance, suggesting that a very similar pathway is likely in RNA as well. Despite some degree of uncertainty in applying the results from DNA structures to RNA, it is interesting to evaluate the proposed transition pathway for a well-documented RNA-protein complex involving specific interactions with an adenine bulge.

The interaction of the coat protein in phage MS2 with its replicase gene operator RNA is the most extensively studied example of nucleic acid-protein interactions involving an extra nucleotide, and it is interesting to discuss the simulation results with respect to this complex in more detail. Experimental data have established that apart from the nucleotides in the loop region, only a bulged purine at position −10 is essential for tight binding of the RNA operator to the closely related R17 coat protein (Wu and Uhlenbeck, 1987; Romaniuk et al., 1987). This becomes obvious in the crystal structure of the complex where highly specific interactions are visible between the extended bulge nucleotide and residues Thr-45 and Ser-47 of the protein (Valegard et al., 1997). From the available experimental data it has been suggested that the coat protein first recognizes and binds the RNA loop region before the adenine bulge loops out into the extended form and also binds to the protein. This would imply limitations for the transition path and the conformational flexibility of nucleotide bases next to the bulge during the transition due to the presence of the already associated protein.

As shown in Fig. 12 the suggested transition path from our simulations presents a feasible pathway that would allow the base to move from the stacked form, which is apparently more favorable in free RNA in solution, to the extended form, more favorable in the complex, without clashing with the protein bound to the loop region. An interesting additional effect of the transition from the

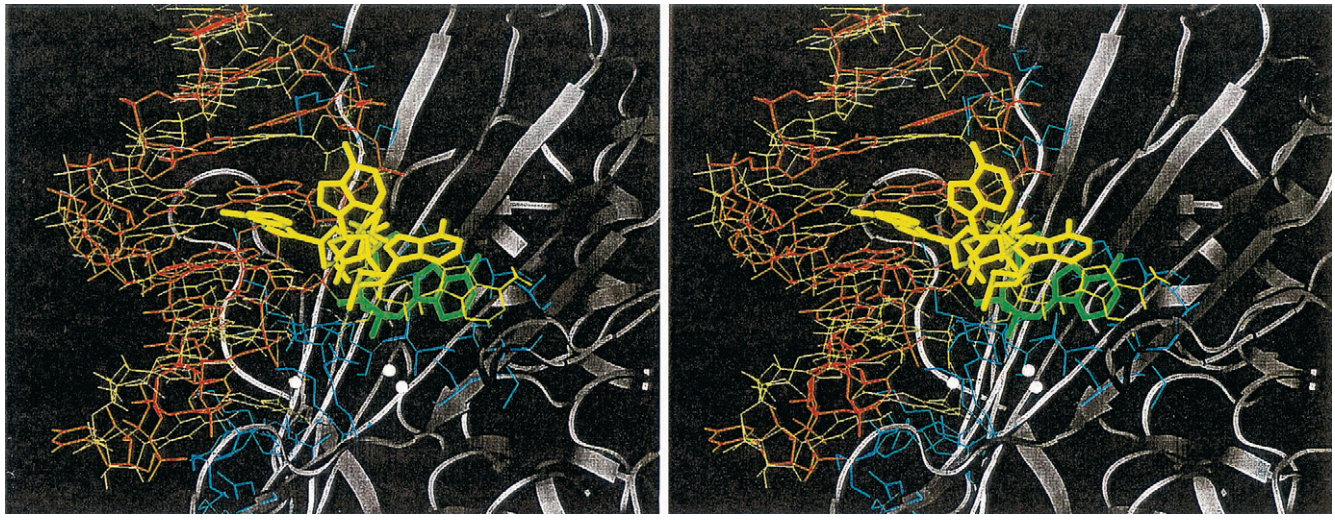

FIGURE 12 Complex of RNA fragment representing bacteriophage MS2 repression operator and initiation site for replicase gene with MS2 coat protein (Valegard et al., 1997) in comparison with simulated bulge conformations along the suggested transition path. Experimental RNA structure is shown in red and green, protein residues within 5 Å are shown in blue. The E2 conformation from the simulation matched with the RNA structure at the backbone around the bulge base is given in yellow. Bulge conformations of structures at 9.5, 9.8, and 11.4 ns from simulation E-C matched to the RNA in the same way are highlighted and shown in yellow as well. White balls represent water molecules in the crystal structure.

stacked to the extended form is a straightening of the helix at the bulge site that enables an additional RNA-protein contact point between the RNA A-13 nucleotide and the Arg-49 protein residue further enhancing binding specificity for the looped-out conformation. The biological function of the repressor operator also requires some form of modulation that allows dissociation for eventual gene transcription. It appears that this function is provided by the conformational flexibility of the adenine bulge because the binding affinity of the stem-loop is expected to vary greatly between a stacked and extended bulge conformation. Our simulations indicate that one possibility to affect the equilibrium between stacked and extended conformations may be given by changes in the ionic solvent environment. Because RNA stem-loops are a common motif in RNA-protein interactions, it is likely that similar mechanisms are relevant for other RNA-protein complexes. However, further studies will be required for a more complete understanding, including the fact that relative free energies for both conformations and barriers are expected to differ between DNA and RNA.

CONCLUSIONS

In this study molecular dynamics simulations over 35 ns total simulation time have provided new insight into the range of conformations that are accessible to a DNA fragment with a single bulge base with implications for bulge-specific protein-DNA and protein-RNA interactions. In excellent agreement with experimental data, the simulations produce stable structures where the bulge base remains either stacked within the flanking bases or fully extended for many nanoseconds. Furthermore, the extensive sampling done here allows a division of each family into distinct subconformations that occur as a result of discrete backbone states and alternative stacking interactions with either one of the flanking basepairs in the stacked form. In addition, for significant periods of simulation time partially extended conformations are observed in which the bulge base associates with the minor groove to varying extents. Although a complete transition between the stacked and extended forms has not been observed in the simulations, the combination of all observed structures suggests a likely transition path that would leave the surrounding DNA structure mostly unaffected. For the interaction of the MS2 replicase gene operator RNA with its coat protein such a pathway allows initial recognition and binding of the RNA fragment with a stacked bulge base found in solution before the bulge loops out to an extended form that is more favorable in the complex.

APPENDIX

Here we describe the details of a convenient set of coordinates for studying bulges. First, a local coordinate system is calculated for both flanking basepairs by applying a previously introduced definition (Feig and Pettitt, 1999a). The line through the centers of mass of each base to the C1 carbon defines the x axis. The average between the vectors from the C4 (pyrimidines) or C6 (purines) atom to the C2 atom is used to approximate the y axis and span the x/y plane. The z axis can then be derived by taking the cross product between x and the approximate y axes. The proper orthogonal y axis follows from the cross product between z and x. The reference frame for calculating the bulge base orientation is then defined as the average between the coordinate systems of both basepairs. To arrive at an orthogonal coordinate system, only the x and approximate y axes are averaged from both basepairs. From these averages the z and proper y axes are then

computed in the same way as described above. The origin of the new reference frame is set at the midpoint on the connection line between the centers of mass of the flanking bases on the strand that contains the bulge base so that the center of mass of the bulge base will be close to the origin in the stacked conformation. Displacements of the bulge base center of mass from this origin along the x, y, and z directions in the new reference frame are termed slide, shift, and rise, in correspondence with the regular helical parameters with the same name (Dickerson et al., 1989). The vector from the furanose C1′ atom at the bulge base to the bulge base center of mass is used to describe the angular orientation of the bulge base. Representing this vector in spherical coordinates gives the angles θ and ϕ. θ measures the counterclockwise rotation in the x/y plane so that bulge base is in the stacked form at $\theta \approx -30°$, in the minor groove for $\theta \approx 0°$, fully extended at $\theta \approx -180°$, and in the major groove for $\theta < -90°$. The angle ϕ specifies the inclination of the bulge base out of the x/y plane. Calculated as the angle toward the z axis it is 90° when the bulge lies in the x/y plane, and 0° or 180° when the bulge base is completely perpendicular to the plane along the z axis in positive or negative direction, respectively. The third rotational degree of freedom consists of the rotation angle around the vector from C1′ to the bulge base center of mass that will be called ξ. It is calculated as the rotation of the vector from C6 to C2 around the x axis after rotating the bulge base back so that the vector from C1′ to the bulge base center of mass is aligned with the x axis. This is accomplished by a rotation around the z axis by $-\theta$ and then around the y axis by $-\phi + \pi/2$.

We thank the Texas Coordinating Board, the Robert A. Welch Foundation, and the NIH for partial support of this research. We thank Alex MacKerell for making the new CHARMM27 parameter set available to us before publication. Gillian C. Lynch is thanked for valuable suggestions and discussions. NPACI is acknowledged for a generous allocation of computer time at San Diego Supercomputing Center. MSI is acknowledged for providing graphics software support.

REFERENCES

Aboula-ela, F., J. Karn, and G. Varani. 1995. The structure of the human immunodeficiency virus type-1 TAR RNA reveals principles of RNA recognition by Tat protein. *J. Mol. Biol.* 253:313–332.

Allen, M. P., and D. J. Tildesley. 1987. Computer Simulation of Liquids, 1st Ed. Oxford University Press, New York.

Auffinger, P., and E. Westhof. 1997. RNA hydration: three nanoseconds of multiple molecular dynamics simulations of the solvated $tRNA^{Asp}$ anticodon hairpin. *J. Mol. Biol.* 269:326–341.

Auffinger, P., and E. Westhof. 1998. Simulations of the molecular dynamics of nucleic acids. *Curr. Opin. Struct. Biol.* 2:227–236.

Baudin, F., and P. J. Romaniuk. 1989. A difference in the importance of bulged nucleotides and parent base pairs in the binding of transcription factor IIIA to Xenopus 5S RNA and 5S RNA genes. *Nucleic Acid Res.* 17:2043–2056.

Beveridge, D. L., K. J. McConell, R. Nirmala, M. A. Young, S. Vijayakumar, and G. Ravishanker. 1994. Recent progress in molecular dynamics simulations of DNA and protein-DNA complexes including solvent. *ACS Symposium Series.* 568:381–394.

Beveridge, D. L., M. A. Young, and D. Sprous. 1997. Modeling of DNA via molecular dynamics simulation: structure, bending, and conformational transitions. *In* Molecular Modeling of Nucleic Acids. N. B. Leontis, editor. ACS, Washington, DC, 260–284.

Bhattacharyya, A., and D. M. J. Lilley. 1989. The contrasting structures of mismatched DNA sequences containing looped-out bases (bulges) and multiple mismatches (bubbles). *Nucleic Acids Res.* 17:6821–6840.

Borer, P. N., Y. Lin, S. Wang, M. W. Roggenbuck, J. M. Gott, O. C. Uhlenbeck, and I. Pelczer. 1995. Proton NMR and structural features of a 24-nucleotide RNA hairpin. *Biochemistry.* 34:6488–6503.

Cate, J. H., A. R. Gooding, E. Podell, K. Zhou, B. L. Golden, C. E. Kundrot, T. R. Cech, and J. A. Doudna. 1996. Crystal structure of a group I ribozyme domain: principles of RNA packing. *Science.* 273: 1678–1685.

Cheatham, T. E., M. F. Crowley, T. Fox, and P. A. Kollman. 1997a. A molecular level picture of the stabilization of A-DNA in mixed ethanol-water solutions. *Proc. Natl. Acad. Sci. USA.* 94:9626–9630.

Cheatham, T. E., and P. A. Kollman. 1996. Observation of the A-DNA to B-DNA transition during unrestrained molecular dynamics in aqueous solution. *J. Mol. Biol.* 259:434–444.

Cheatham, T. E., and P. A. Kollman. 1997a. Insight into the stabilization of A-DNA by specific ion association: spontaneous B-DNA to A-DNA transitions observed in molecular dynamics simulations of $d(ACCCGCGGGT)_2$ in the presence of hexaamminecobalt(III). *Structure.* 15:1297–1311.

Cheatham, T. E., and P. A. Kollman. 1997b. Molecular dynamics simulations highlights the structural differences among DNA:DNA, RNA:RNA, and DNA:RNA hybrid duplexes. *J. Am. Chem. Soc.* 119: 4805–4825.

Cheatham, T. E., J. L. Miller, T. I. Spector, P. Cieplak, and P. A. Kollman. 1997b. Molecular dynamics simulations on nucleic acid systems using the Cornell et al. force field and particle mesh Ewald electrostatics. *In* Molecular Modeling of Nucleic Acids. N. B. Leontis, editor, ACS, Washington, DC, 285–303.

Chen, Y. Z., V. Mohan, and R. H. Griffey. 1998. Effect of backbone ζ torsion angle on low energy single base opening in B-DNA crystal structures. *Chem. Phys. Lett.* 287:570.

Cheng, C.-C., Y.-N. Kuo, K.-S. Chuang, C.-F. Luo, and W. J. Wang. 1999. A new Co^{II} complex as a bulge-specific probe for DNA. *Angewandte Chemie.* 38:1255–1257.

Cornell, W. D., P. Cieplak, C. I. Bayly, I. R. Gould, K. M. Merz, D. M. Ferguson, D. C. Spellmeyer, T. Fox, J. W. Caldwell, and P. A. Kollman. 1995. A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *J. Am. Chem. Soc.* 117: 5179–5197.

de Leeuw, S. W., J. W. Perram, and E. R. Smith. 1980. Simulation of electrostatic systems in periodic boundary conditions. I. Lattice sums and dielectric constants. *Proc. R. Soc. Lond. A.* 373:27–56.

Dickerson, R. E. 1983. Base sequence and helix structure variation in B- and A-DNA. *J. Mol. Biol.* 166:419–441.

Dickerson, R. E. 1992. NEWHELIX Program Suite. University of California at Los Angeles.

Dickerson, R. E., et al. 1989. Definitions and nomenclature of nucleic acid structure parameters. *J. Mol. Biol.* 205:787–791.

Dornberger, U., J. Flemming, and H. Fritzsche. 1998. Structure determination and analysis of helix parameters in the DNA decamer $d(CATGGCCATG)_2$. Comparison of results from NMR and crystallography. *J. Mol. Biol.* 284:1453–1463.

Duan, Y., P. Wilkosz, M. Crowley, and J. M. Rosenberg. 1997. Molecular dynamics simulation study of DNA dodecamer d(CGCGAATTCGCG) in solution: conformation and hydration. *J. Mol. Biol.* 272:553–572.

Ennifar, E., M. Yusupov, P. Walter, R. Marquet, C. Ehresmann, B. Ehresmann, and P. Dumas. 1999. The crystal structure of the dimerization initiation site of genomic HIV-1 RNA reveals an extended duplex with two adenine bulges. *Structure.* 7:1439–1349.

Eshleman, J. R., and S. D. Markowitz. 1995. Microsatellite instability in inherited and sporadic neoplasms. *Curr. Opin. Oncol.* 7:83–89.

Feig, M., and B. M. Pettitt. 1998a. Crystallographic water sites from a theoretical perspective. *Structure.* 6:1351–1354.

Feig, M., and B. M. Pettitt. 1998b. Structural equilibrium of DNA represented with different force fields. *Biophys. J.* 75:134–149.

Feig, M., and B. M. Pettitt. 1999a. Modeling high-resolution hydration patterns in correlation with DNA sequence and conformation. *J. Mol. Biol.* 286:1075–1095.

Feig, M., and B. M. Pettitt. 1999b. Sodium and chlorine ions as part of the DNA solvation shell. *Biophys. J.* 77:1769–1781.

Foloppe, N., and A. D. MacKerell. 2000. All-atom empirical force field for nucleic acids. I. Parameter optimization based on small molecule and condensed phase macromolecular target data. *J. Comp. Chem.* 21: 86–104.

Fresco, J. R., and B. M. Alberts. 1960. The accommodation of noncomplementary bases in helical polyribonucleotides and deoxyribonucleic acids. *Proc. Natl. Acad. Sci. USA.* 46:311–321.

Genschel, J., S. J. Littman, and P. Modrich. 1998. Isolation of MutS beta from human cells and comparison of the mismatch repair specificities of the MutS beta and MutS alpha. *J. Biol. Chem.* 273:19895–19901.

Gohlke, C., A. I. H. Murchie, D. M. J. Lilley, and R. M. Clegg. 1994. Kinking of DNA and RNA helices by bulged nucleotides observed by fluorescence energy transfer. *Proc. Natl. Acad. Sci. USA.* 91: 11660–11664.

Golden, B. L., A. R. Gooding, E. R. Podell, and T. R. Cech. 1998. A preorganized active site in the crystal structure of the tetrahymena ribozyme. *Science.* 282:259–264.

Greenbaum, N. L., I. Radhakrishnan, D. J. Patel, and D. Hirsh. 1996. Solution structure of the donor site of a trans-splicing RNA. *Structure.* 4:725–733.

Hare, D., L. Shapiro, and D. J. Patel. 1986. Extra-helical adenosine stacks into right-handed DNA: solution conformation of the d(CGCA-GAGCTCGCG) duplex deduced from distance geometry analysis of nuclear Overhauser effect spectra. *Biochemistry.* 25:7456–7464.

Hsieh, C.-H., and J. D. Griffith. 1989. Deletions of bases in one strand of duplex DNA, in contrast to single-base mismatches, produce highly kinked molecules: possible relevance to the folding of single-stranded nucleic acids. *Proc. Natl. Acad. Sci. USA.* 86:4833–4837.

Jayaram, B., D. Sprous, M. A. Young, and D. L. Beveridge. 1998. Free energy analysis of the conformational preferences of A and B forms of DNA in solution. *J. Am. Chem. Soc.* 120:10629–10633.

Jorgensen, W., J. Chandrasekhar, J. Madura, R. Impey, and M. Klein. 1983. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* 79:926–935.

Joshua-Tor, L., F. Frolow, E. Appela, H. Hope, D. Rabinovich, and J. L. Sussman. 1992. Three-dimensional structures of bulge-containing DNA fragments. *J. Mol. Biol.* 225:397–431.

Joshua-Tor, L., D. Rabinovich, H. Hope, F. Frolow, E. Appella, and J. L. Sussman. 1988. The three-dimensional structure of a DNA duplex containing looped-out bases. *Nature.* 334:83–84.

Kalnik, M. W., D. G. Norman, B. F. Li, P. F. Swann, and D. J. Patel. 1990. Conformational transitions in thymidine bulge-containing deoxytridecanucleotide duplexes. *J. Biol. Chem.* 265:636–647.

Kalnik, M. W., D. G. Norman, P. F. Swann, and D. J. Patel. 1989a. Conformation of adenosine bulge-containing deoxytridecanucleotide duplexes in solution. *J. Biol. Chem.* 264:3702–3712.

Kalnik, M. W., D. G. Norman, M. G. Zagorski, P. F. Swann, and D. J. Patel. 1989b. Conformational transitions in cytidine bulge-containing deoxytridecanucleotide duplexes: extra cytidine equilibrates between looped out (low temperature) and stacked (elevated temperature) conformations in solution. *Biochemistry.* 28:294–303.

Kappen, L. S., and I. H. Goldberg. 1993. Site-specific cleavage at a DNA bulge by neocarzinostatin chromophore via a novel mechanism. *Biochemistry.* 32:13138–13145.

Kappen, L. S., and I. H. Goldberg. 1997. Characterization of a covalent monoadduct of neocarzinostatin chromophore at a DNA bulge. *Biochemistry.* 36:14861–14867.

Kerwood, D. J., and P. N. Borer. 1996. Structure refinement for a 24-nucleotide RNA hairpin. *Magn. Reson. Chem.* 34:S136–S146.

Klimasauskas, S., S. Kumar, R. J. Roberts, and X. Cheng. 1994. HhaI methyltransferase flips its target base out of the DNA helix. *Cell.* 76:357–369.

Kolodner, R. 1996. Biochemistry and genetics of eukaryotic mismatch repair. *Genes Devel.* 10:1433–1442.

Kolodner, R. D. 1995. Mismatch repair: mechanisms and relationship to cancer susceptibility. *Trends Biochem.* 20:397–402.

Kumar, S., X. Cheng, S. Klimasauskas, S. Mi, J. Posfai, R. J. Roberts, and G. G. Wilson. 1994. The DNA (cytosine-5) methyltransferases. *Nucleic Acids Res.* 22:1–10.

Lee, H., T. A. Darden, and L. Pedersen. 1995. Molecular dynamics simulation studies of a high resolution Z-DNA crystal. *J. Chem. Phys.* 102:3830–3834.

Lippard, S. J., and J. M. Berg. 1993. Principles of Bioinorganic Chemistry. University Science Books, Mill Valley, CA.

MacKerell, A. D. 1997. Observations on the A versus B equilibrium in molecular dynamics simulations of duplex DNA and RNA. *In* Molecular Modeling of Nucleic Acids. N. B. Leontis, editor. ACS, Washington, DC. 304–311.

MacKerell, A. D., and N. K. Banavali. 2000. All-atom empirical force field for nucleic acids. II. Application to molecular dynamics simulations of DNA and RNA in solution. *J. Comp. Chem.* 21:105–120.

Marsischky, G. T., N. Filosi, M. F. Kane, and R. Kolodner. 1996. Redundancy of *Saccharomyces cerevisiae* MSH3 and MSH6 in MSH2-dependent mismatch repair. *Genes Devel.* 10:407–420.

Miller, M., R. W. Harrison, A. Wlodawer, E. Appella, and J. L. Sussman. 1988. Crystal structure of 15-mer DNA duplex containing unpaired bases. *Nature.* 334:85–86.

Modrich, P. 1991. Mechanisms and biological effects of mismatch repair. *Annu. Rev. Genet.* 25:229–253.

Modrich, P., and R. S. Lahue. 1996. Mismatch repair in replication fidelity, genetic recombination, and cancer biology. *Annu. Rev. Biochem.* 65: 101–133.

Morden, K. M., Y. G. Chu, F. H. Martin, and I. Tinoco. 1983. Unpaired cytosine in the deoxyoligonucleotide duplex $d(CA_3CA_3G) \cdot d(CT_6G)$ is outside of the helix. *Biochemistry.* 22:5557–5563.

Morden, K. M., B. M. Gunn, and K. Maskos. 1990. NMR studies of a deoxyribodecanucleotide containing an extrahelical thymidine surrounded by an oligo(dA) · oligo(dT) tract. *Biochemistry.* 29:8835–8845.

Nelson, J. W., and I. Tinoco. 1985. Ethidium ion binds more strongly to a DNA double helix with a bulged cytosine base than to a regular double helix. *Biochemistry.* 24:6416–6421.

Nikonowicz, E., R. P. Meadows, and D. G. Gorenstein. 1990. NMR structural refinement of an extrahelical adenosine tridecamer d(CGCA-GAATTCGCG)$_2$ via a hybrid relaxation matrix procedure. *Biochemistry.* 29:4193–4204.

Nikonowicz, E., V. Roongta, C. R. Jones, and D. G. Gorenstein. 1989. Two-dimensional ^{1}H and ^{31}P NMR spectra and restrained molecular dynamics structure of an extrahelical adenosine tridecamer oligodeoxynucleotide duplex. *Biochemistry.* 28:8714–8725.

Norberg, J., and L. Nilsson. 1995a. Potential of mean force calculations of the stacking-unstacking process in single-stranded deoxyribodinucleoside monophosphates. *Biophys. J.* 69:2277–2285.

Norberg, J., and L. Nilsson. 1995b. Stacking free energy profiles for all 16 natural ribodinucleoside monophosphates in aqueous solution. *J. Am. Chem. Soc.* 117:10832–10840.

O'Gara, M., S. Klimasauskas, R. J. Roberts, and X. Cheng. 1996. Enzymatic C5-cytosine methylation of DNA: mechanistic implications of new crystal structures for HhaI methyltransferase-DNA-AdoHcy complexes. *J. Mol. Biol.* 261:634–645.

Olson, W. K., A. R. Srinivasan, M. Hao, and J. L. Nauss. 1988. Base Sequence-dependent Models of DNA Curvature. The Geometric Interdependence of Complementary Residues. Adenine Press, Schenectady, NY.

Patel, D. J., S. A. Kozlowski, L. A. Marky, J. Rice, C. Broka, K. Itakura, and K. J. Breslauer. 1982. Extra adenosine stacks into the self-complementary d(CGCAGAATTCGCG) duplex in solution. *Biochemistry.* 21:445–451.

Peattie, D. A., S. Douthwaite, R. A. Garrett, and H. F. Noller. 1981. A "bulged" double helix in an RNA-protein contact site. *Proc. Natl. Acad. Sci. USA.* 78:7331–7335.

Portmann, S., S. Grimm, C. Workman, N. Usman, and M. Egli. 1996. Crystal structures of an A-form duplex with single-adenosine bulges and conformational basis for site-specific RNA self-cleavage. *Chem. Biol.* 3:173–184.

Puglisi, J. D., R. Tan, B. J. Calnan, A. D. Frankel, and J. R. Williamson. 1992. Conformation of the TAR RNA-arginine complex by NMR spectroscopy. *Science.* 257:76–80.

Query, C. C., M. J. Moore, and P. Sharp. 1994. Branch nucleophile selection in pre-mRNA splicing: evidence for the bulged duplex model. *Genes Devel.* 8:587–597.

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

Reinisch, K. M., L. Chen, G. L. Verdine, and W. N. Lipscomb. 1995. The crystal structure of HaeIII methylase covalently complexed to DNA: an extrahelical cytosine and rearranged base pairing. *Cell.* 82:143–153.

Rice, J. A., and D. M. Crothers. 1989. DNA bending by the bulge defect. *Biochemistry.* 28:4512–4516.

Romaniuk, P. J., P. T. Lowary, H. N. Wu, G. Stormo, and O. C. Uhlenbeck. 1987. RNA binding site of R17 coat protein. *Biochemistry*. 26: 1563–1568.

Rosen, M. A., D. Live, and D. J. Patel. 1992. Comparative NMR study of A_n bulge loops in DNA duplexes: intrahelical stacking of A, A-A, and A-A-A bulge loops. *Biochemistry*. 31:4004–4014.

Roux, B., B. Prodhom, and M. Karplus. 1995. Ion transport in the gramicidin channel: molecular dynamics study of single and double occupancy. *Biophys. J.* 68:876–892.

Rowsell, S., N. J. Stonehouse, M. A. Convery, C. J. Adams, A. D. Ellington, I. Hirao, D. S. Peabody, P. G. Stockley, and S. E. V. Phillips. 1998. Crystal structures of a series of RNA aptamers complexed to the same protein target. *Nat. Struct. Biol.* 5:970–975.

Ryckaert, J. P., G. Ciccotti, 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: 327–341.

Saenger, W. 1984. Principles of Nucleic Acid Structure. Springer Verlag, New York.

Schneider, B., S. Neidle, and H. M. Berman. 1997. Conformations of the sugar-phosphate backbone in helical DNA crystal structures. *Biopolymers.* 42:113–124.

Smith, P. E., M. E. Holder, L. X. Dang, M. Feig, and B. M. Pettitt. 1996. ESP. University of Houston.

Smith, J. S., and E. P. Nikonowicz. 1998. NMR structure and dynamics of an RNA motif common to the spliceosome branch-point helix and the RNA-binding site for phage GA coat protein. *Biochemistry*. 37: 13486–13498.

Smith, P. E., and B. M. Pettitt. 1994. Modeling solvent in biomolecular systems. *J. Phys. Chem.* 98:9700–9711.

Smith, P. E., and B. M. Pettitt. 1995. Efficient Ewald electrostatic calculations for large systems. *Comp. Phys. Commun.* 91:339–344.

Streisinger, G., Y. Okada, J. Emrich, J. Newton, A. Tsugita, E. Terzaghi, and M. Inouye. 1966. Frameshift mutations and the genetic code. *Cold Spring Harbor Symp. Quant. Biol.* 31:77–84.

Sudarsanakumar, C., Y. Xiong, and M. Sundaralingam. 2000. Crystal structure of an adenine bulge in the RNA chain of a DNA · RNA hybrid, d(CTCCTCTTC) · r(gaagagagag). *J. Mol. Biol.* 299:103–112.

Tang, C. K., and D. E. Draper. 1990. Evidence for allosteric coupling between the ribosome and repressor binding sites of a translationally regulated mRNA. *Biochemistry.* 29:5232–5237.

Thiviyanathan, V., A. B. Guliaev, N. B. Leontis, and D. G. Gorenstein. 2000. Solution conformation of a bulged adenosine base in an RNA duplex by relaxation matrix refinement. *J. Mol. Biol.* 300:1143–1154.

Turner, D. H. 1992. Bulges in nucleic acids. *Curr. Opin. Struct. Biol.* 2:334–337.

Valegard, K., J. B. Murray, P. G. Stockley, N. J. Stonehouse, and L. Liljas. 1994. Crystal structure of an RNA bacteriophage coat protein-operator complex. *Nature.* 371:623–626.

Valegard, K., J. B. Murray, N. J. Stonehouse, S. van den Worm, P. G. Stockley, and L. Liljas. 1997. The three dimensional structures of two complexes between recombinant MS2 capsids and RNA operator fragments reveal sequence-specific protein-rNa interactions. *J. Mol. Biol.* 270:724–738.

van den Hoogen, Y. T., A. A. van Beuzekom, E. de Vroom, G. A. van der Marel, J. H. van Boom, and C. Altona. 1988b. Bulge-out structures in the single-stranded trimer AUA and in the duplex (CUGGUGCGG) · (CCGCCCAG). A model-building and NMR study. *Nucleic Acids Res.* 16:5013–5030.

van den Hoogen, Y. T., A. A. van Beuzekom, H. van den Elst, G. A. van der Marel, J. H. van Boom, and C. Altona. 1988a. Extra thymidine stacks into the d(CTGGTGCGG) · d(CCGCCCAG) duplex. An NMR and model-building study. *Nucleic Acids Res.* 16:2971–2986.

van den Worm, S. H. E., N. J. Stonehouse, K. Valegard, J. B. Murray, C. Walton, K. Fridborg, P. G. Stockley, and L. Liljas. 1998. Crystal structures of MS2 coat protein mutants in complex with wild-type RNA operator fragments. *Nucleic Acids Res.* 26:1345–1351.

Vassylev, D. G., T. Kashiwagi, Y. Mikami, M. Ariyoshi, S. Iwai, E. Ohtsuka, and K. Morikawa. 1995. Atomic model of a pyrimidine dimer-specific excision repair enzyme complexed with a DNA substrate: structural basis for damaged DNA recognition. *Cell.* 83:773–782.

Wang, Y.-H., and J. Griffith. 1991. Effects of bulge composition and flanking sequence on the kinking of DNA by bulged bases. *Biochemistry.* 30:1358–1363.

Weerasinghe, S., P. E. Smith, V. Mohan, Y.-K. Cheng, and B. M. Pettitt. 1995a. Nanosecond dynamics and structure of a model DNA triple helix in saltwater solution. *J. Am. Chem. Soc.* 117:2147–2158.

Weerasinghe, S., P. E. Smith, and B. M. Pettitt. 1995b. Structure and stability of a model pyrimidine-purine-purine DNA triple helix with a GC-T mismatch by simulation. *Biochemistry.* 34:16269–16278.

Woodson, S. A., and D. M. Crothers. 1988a. Binding of 9-aminoacridine to bulged-base DNA oligomers from a frameshift hot spot. *Biochemistry.* 27:8904–8914.

Woodson, S. A., and D. M. Crothers. 1988b. Structural model for an oligonucleotide containing a bulged guanosine by NMR and energy minimization. *Biochemistry.* 27:3130–3141.

Wu, H.-N., and O. C. Uhlenbeck. 1987. Role of a bulged a residue in a specific RNA protein interaction. *Biochemistry.* 26:8221–8227.

Wyatt, J. R., and J. Tinoco. 1993. Cold Spring Harbor Laboratory Press, New York. 465–496.

Young, M. A., and D. L. Beveridge. 1998. Molecular dynamics simulations of an oligonucleotide duplex with adenine tracts phased by a full helix turn. *J. Mol. Biol.* 281:675–687.

Young, M. A., B. Jayaram, and D. L. Beveridge. 1997a. Intrusion of counterions into the spine of hydration in the minor groove of B-DNA: fractional occupancy of electronegative pockets. *J. Am. Chem. Soc.* 119:59–69.

Young, M. A., G. Ravishanker, and D. L. Beveridge. 1997b. A 5-nanosecond molecular dynamics trajectory for B-DNA: analysis of structure, motions, and solvation. *Biophys. J.* 73:2313–2336.

Young, M. A., G. Ravishanker, D. L. Beveridge, and H. M. Berman. 1995. Analysis of local helix bending in crystal structures of DNA oligonucleotides and DNA-protein complexes. *Biophys. J.* 68:2454–2468.

Zacharias, M., and P. J. Hagerman. 1995. Bulge-induced bends in RNA: quantification by transient electric birefringence. *J. Mol. Biol.* 247: 486–500.

Zacharias, M., and H. Sklenar. 1997. Analysis of the stability of looped-out and stacked-in conformations of an adenine bulge in DNA using a continuum model for solvent and ions. *Biophys. J.* 73:2990–3003.

Zhang, P., P. Popienick, and P. B. Moore. 1989. Physical studies of 5S RNA variants at position 66. *Nucleic Acid. Res.* 17:8645–8656.

Zhurkin, V. B. 1985. Sequence-dependent bending of DNA and phasing of nucleosomes. *J. Biomol. Struct. Dyn.* 2:785–804.