Internal Loop−Helix Coupling in the Dynamics of the RNA Duplex $(GC^*C^*AGUUCGCUGGC)_2$

Edward C. Sherer and Christopher J. Cramer*

Department of Chemistry and Supercomputer Institute, University of Minnesota, 207 Pleasant St. SE, Minneapolis, Minnesota 55455-0431

Received: December 11, 2001; In Final Form: February 12, 2002

Molecular dynamics simulations are carried out for a double helical RNA tetradecamer containing a UUCG internal loop. Analysis of four separate trajectories, totaling 12 ns in length, identifies several persistent structural motifs adopted by the internal loop under various conditions. In addition to previously identified water-mediated C:U and G:U hydrogen bonding interactions, two key interpair interactions are identified, namely, cross-step hydrogen bonding between adjacent C:U pairs and cross-strand stacking interactions involving mispaired G, and cross-strand stacking interactions involving the C:U pairs. On the basis of principal component analysis of the full oligomer's motion, these local structures are associated with highly compressed helices, and underwound, elongated helices, respectively.

Introduction

Interactions between proteins and nucleic acid polymers depend critically upon the secondary and tertiary structural elements of the latter. With respect to biologically relevant forms, the structural freedom of RNA is more varied and complicated than DNA. Of the many secondary structural elements found in natural (and unnatural) RNA, tetraloops and internal loops are among the most common of noncanonical motifs. These internal unpaired loops, four bases in length, can either cap one end of an RNA helix when a single-stranded RNA forms duplex regions with itself upon folding, or they can be found intervening between two double-helical regions formed from the combination of two separate but self-complementary strands of RNA.

RNA strands that contain the sequence UUCG between otherwise self-complementary regions form tetraloops in solution.[1−4] When the concentration of the single strand is high, or when the sequence is crystallized, it is possible to obtain duplex formation with an internal loop of four mismatched pairs.[5−8] The loop is composed of two cytosine:uracil (C:U) and two guanine:uracil (G:U) noncanonical base pairs. The G:U base pair is known to preferentially adopt a wobble geometry, typically with an associated tightly bound water molecule.[1,2,9−17] The C:U pair also tends to adopt a water-mediated conformation, with somewhat greater variability in overall geometry, although the various motifs that have been observed tend to preserve a Watson−Crick-like spacing between the residue backbones.[5−8,17−23]

The intrinsic flexibility of such $(UUCG)_2$ loops is a question bearing on the degree to which they may influence molecular recognition. Schneider et al.[24] analyzed a 4 ns molecular dynamics (MD) simulation of (r-GGAC**UUCG**GUCC)$_2$ and noted four distinct C:U motifs, three of which included a water molecule tightly bound to the base pair; they concluded that the internal flexibility of the loop was high and subject to considerable conformational variability.[24] In addition, Anderson et al.[7] recently crystallized the tetradecamer duplex (r-GC^*C^*AG**UUCG**CUGGC)$_2$, where C* is 5-bromocytosine, and determined there to be three unique structures within the single-crystal unit cell, again suggesting that the loop contributes to structural variability, although Anderson et al. noted that the variation between structures was small.

To gain more insight into the nature of the flexibility intrinsic to the $(UUCG)_2$ loop, we have carried out MD simulations starting from each of the three structures found in the unit cell by Anderson et al.;[7] multiple molecular dynamics simulations have been shown to more efficiently explore phase space than a single long trajectory.[25−29] The system is modeled using the force field of Cornell et al.,[30] which has been demonstrated to provide good agreement with experiment for a variety of structural and dynamical properties of DNA and RNA, to include applications to species characterized by noncanonical nucleotide structures.[16,31−41] To better characterize the dynamics and to facilitate comparison between trajectories, we use principal components analysis (PCA) to determine the dominant macromolecular vibrational modes of the RNA.[36,42−49]

Computational Methodology

Starting structures for the MD simulations were taken from the single-crystal X-ray data of Anderson et al.[7] for (r-GC^*C^*AG**UUCG**CUGGC)$_2$ available as Protein Data Bank entry 1QBP—the unit cell in this study contains three unique RNA geometries. The sequence employed by Anderson et al. left one-base single-stranded overhangs after duplex formation, but these overhangs exhibited sufficient disorder that they were not located in the crystal structure; they have not been included in the simulations. Our nomenclature system will be to number this sequence with the 5′-G as 1, the 3′-C as 14, and the corresponding bases on the complementary strand as 15 and 28, respectively.

Because the sequence of Anderson et al. includes the nonstandard base 5-bromocytosine, we developed parameters for this base consistent with the force field of Cornell et al.[30] Partial atomic charges for C* were determined by fitting to the HF/6-31G* electrostatic potential and normalizing the resulting charges so that the net charge for C* was the same as for C. The necessary ab initio calculations were carried out using the Gaussian 98 software[50] package. All parameters for C* are provided as Supporting Information.

10.1021/jp014494d CCC: $22.00
Published on Web 04/20/2002

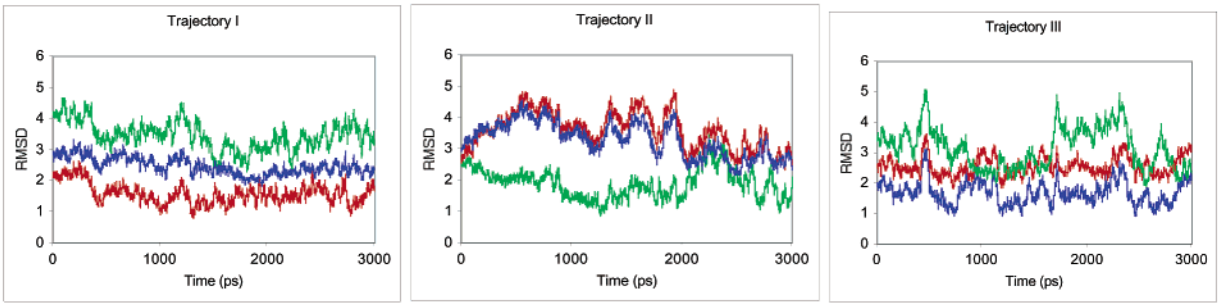

Figure 1. All-atom RMSDs (Å) for trajectories **I** (red), **II** (green), and **III** (blue) vs time-averaged structures for **I** (left), **II** (center), and **III** (right) over the final 3 ns of each trajectory.

MD trajectories were generated using the force field of Cornell et al.[7] as implemented in the AMBER5 software package[51] for each of the three unique helices. Each helix contains 14 residues (totaling 894 atoms), and 26 sodium ions were added to each so as to neutralize the charge of the polyelectrolyte. Periodic boundary conditions were established for simulation boxes containing 2641, 2621, and 2495 water molecules for the three different helices. The structures were equilibrated following a previously described protocol,[33,36,52] which is listed in detail in the Supporting Information. Following equilibration, each trajectory was propagated until the all-atom root-mean-square deviation (RMSD) was assessed to have been converged over at least 3 ns. During propagation, electrostatic interactions were evaluated without cutoffs using the Particle-Mesh-Ewald (PME) method.[53,54] Constant pressure (1 bar) and temperature (300 K) were maintained according to the Berendsen[55] coupling algorithm. SHAKE was used to constrain all covalent bond distances,[56] and a propagation step of 2 fs was used.

Time-averaged structures over the final 3.0 ns of each trajectory (and over a 9.0 ns trajectory created from catenation of the three individual ones) were calculated and minimized while all heavy atom positions were restrained. Helical geometrical parameters[57,58] for different structures were calculated using Curves.[59]

Principal component analysis of the helical dynamics was performed for the most part as described previously.[36,45] While PCA has been widely applied in the field of protein dynamics,[28,29,43,44,46,60−62] until recently[36,49,63,64] it has not seen as much application to polynucleotides. PCA, also called essential dynamics, ranks the dominant contributions to a given trajectory as a series of dynamical eigenvectors. The first few such vectors are often observed to account for a very large portion of the total structural variation, such that only 1−3 degrees of freedom may be needed to describe most of the large-scale macromolecular dynamics.

The final 3 ns of each of the three distinct trajectories were concatenated to form a single 9 ns trajectory. Next, the positional covariance matrix **C** was calculated after removal of rotational and translational motion. Given M snapshots of an N atom system, **C** is a $3N$ x $3N$ matrix with elements

$$C_{ij} = \frac{1}{M}\sum_{k=1}^{M}(q_{i,k} - \langle q_i \rangle)(q_{j,k} - \langle q_j \rangle) \qquad (1)$$

where $q_{i,k}$ is the value for snapshot k of the ith positional coordinate (x, y, or z coordinate for one of the N atoms), and $\langle q_i \rangle$ indicates the average of that coordinate over all snapshots. Diagonalization of **C** provides the eigenvectors that describe the dynamic motions of the structure, and the associated eigenvalues may be interpreted as weights describing the degree to which each mode contributes to the full dynamics.

As discussed in more detail below, terminal base pair 14:15 showed pronounced unpairing during the simulations. The chaotic motion associated with this unpairing decreased the efficacy of PCA analysis by convoluting the normal modes. To eliminate this effect, PCA was also carried out for a concatenated trajectory where base pair 14:15 was held fixed throughout in its time-averaged geometry as taken from one of the three separate MD runs (this being a postprocessing step, not a separate dynamics simulation). Overlaps, computed as vector dot products, between the first 10 eigenvectors of the frozen and unfrozen trajectories were calculated to evaluate correspondence. Qualitative characterization of PCA modes was accomplished by visual inspection of their animation.

Additional Simulations. To probe the sensitivity of the MD results to both salt concentration and the presence of 5-bromocytosine in place of cytosine, we ran three additional simulations. The first, referred to as trajectory **Ia**, was prepared and propagated in a fashion identical to **I** except that 10 pairs of Na^+ and Cl^- ions were added to the simulation cell so as to double the monovalent ionic strength and place it in close correspondence to that used by Schneider et al.[24] in their analogous simulations of this system. As a separate point, the simulation of Schneider et al.[24] propagated a single trajectory for a double helix of different sequence and with "normal" C in place of C*. To further assess the impact of the unnatural base, we carried out implicit-solvent MD using a generalized Born formalism including nonzero ionic strength[63,65,66] for double helices with C and C* at the appropriate positions. In each case, 10 ns of trajectory were propagated, and good convergence was observed after a 4 ns equilibration period. More complete analysis of the generalized Born results compared to the explicit simulations will be reported elsewhere.

Results

Gross Helical Dynamics. Following the initial equilibration, all three trajectories, hereafter denoted **I**, **II**, and **III**, were propagated for 4.0 ns each. As noted in the methodology section, the propagation of each trajectory was halted when, on the basis of comparison to its own time-averaged structure, it was judged to have demonstrated converged behavior over at least 3 ns—this represents a fairly standard approach to assessing convergence. However, Figure 1 compares the all-atom root-mean-square deviations (RMSDs) for each trajectory against *each* of the three distinct time-averaged structures derived from these trajectories. It is evident that each trajectory, compared to its own time-averaged structure, behaves in a fashion usually associated with convergence. But if one examines the RMSD of trajectory **II** relative to the time-averaged structure from trajectory **I**, for example, it is clear that some conformational changes not manifest in the comparison of each structure with its *own* average are taking place. Having three trajectories in

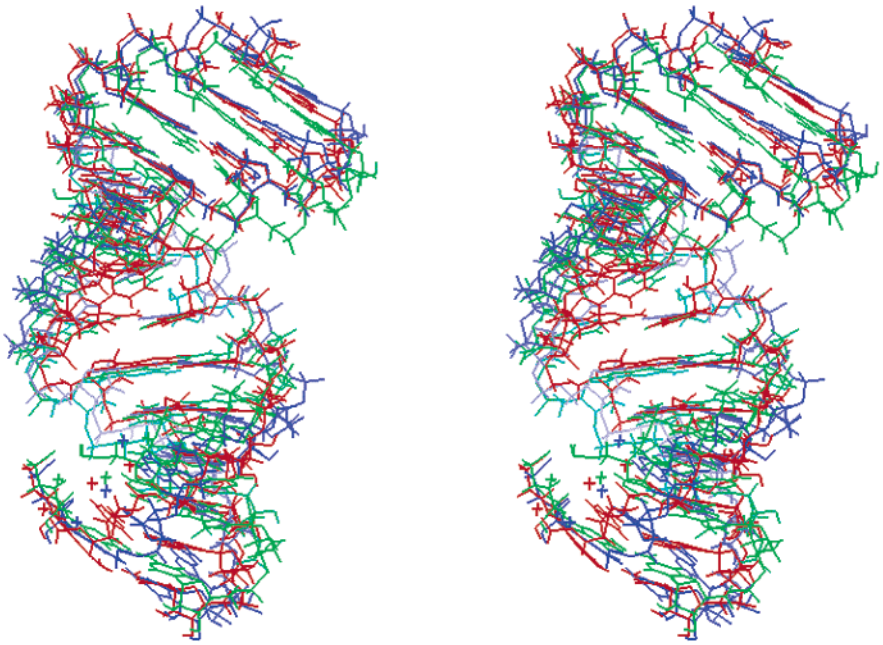

Figure 2. Stereo representation of the optimal overlap of the time-averaged structures of **I** (red), **II** (green), and **III** (blue). The orientation is such that the C14:G15 base pair is at the bottom of the helix, and its unpairing is evident. Bromine atoms appear as crosses.

hand, we can illustrate the somewhat arbitrary nature of assessing convergence in this usual fashion, at least within the context of structures exhibiting a high degree of flexibility.

Indeed, the large number of different structures sampled by the internal loop makes "convergence" in some sense a meaningless goal for the internal loop-containing system—our trajectories hopped back and forth between a number of metastable structures. While the key goal of this work is to explore the nature of those structures and their influence on the dynamics in more detail, there is still value in making the usual comparisons between trajectories to assess "similarity". Thus, the time-averaged structures from each trajectory are overlapped with one another in Figure 2. There is overall a high degree of correspondence, as might be expected given the sizable fraction of double-helical RNA, which *does* maintain a fairly constant structure, in the overall sequences. One feature of primarily technical concern is that the 14:15 terminus is observed to be unpaired to various degrees in all three trajectories.

A quantitative measure of the differences between the overall structures is provided by the RMSDs between the different time-averaged structures, both one with another and also with the three crystal structures from one of which each began. These RMSDs are provided in Table 1 for the entire double-helical structure as well as separately for the loop and helical regions (the time-course of the RMSDs for each trajectory relative to its starting structure as a function of time are provided in the Supporting Information). Also included in Table 1 are the equivalent RMSDs for the concatenated 9 ns trajectory, labeled $\langle \mathbf{I}-\mathbf{III} \rangle_{MD}$. The catenated trajectory more closely resembles every crystal structure than does any one of the individual trajectories, even when individual trajectories are matched with the crystal structure from which they were started. This sort of behavior has been seen previously,[28] and reflects the more balanced distribution of the dynamics in the multitrajectory catenated average structure compared to any one structure.

The RMSDs do not appear to be influenced to a greater extent by the loop regions in comparison to the helical ends in either the simulations or the crystal structures. As is pointed out in the crystal study, the RMS deviation of the UUCG loop region of 1QBP differed from 165D, 255D, and AR005 by 0.65, 1.13, and 1.2 Å, respectively.[7] The RMSDs for the overall trajectory average in reference to the three crystal loop regions of 1.7, 1.7, and 1.5 Å lie slightly outside of this range. Overall, the simulations provide a picture of a somewhat more dynamic helix than was inferred from the crystal study. Of course, the periodicity imposed by a crystal environment may reduce the distribution of structures energetically available to a given polynucleotide, so that a more complete sampling of a large accessible volume of phase space might be expected from an MD trajectory.[26,28,29,46,67−69]

TABLE 1: Heavy-Atom RMSDs (Å) between Time-Averaged MD and X-ray Crystal Structures

structure	$\langle \mathbf{I} \rangle_{MD}$	$\langle \mathbf{II} \rangle_{MD}$	$\langle \mathbf{III} \rangle_{MD}$	$\mathbf{I}_{X-ray}$	$\mathbf{II}_{X-ray}$	$\mathbf{III}_{X-ray}$
			Full Structure			
$\langle \mathbf{I-III} \rangle_{MD}$	1.4	1.8	1.2	2.0	1.9	2.0
$\langle \mathbf{I} \rangle_{MD}$		3.0	1.8	2.3	2.2	2.2
$\langle \mathbf{II} \rangle_{MD}$			2.7	2.8	2.6	2.8
$\langle \mathbf{III} \rangle_{MD}$				2.4	2.5	2.6
$\mathbf{I}_{X-ray}$					0.5	0.9
$\mathbf{II}_{X-ray}$						0.8
			Loop Region Only			
$\langle \mathbf{I-III} \rangle_{MD}$	1.4	1.1	1.1	1.7	1.7	1.5
$\langle \mathbf{I} \rangle_{MD}$		2.2	2.2	2.3	2.3	2.2
$\langle \mathbf{II} \rangle_{MD}$			1.7	2.1	2.1	1.9
$\langle \mathbf{III} \rangle_{MD}$				1.8	1.8	1.6
$\mathbf{I}_{X-ray}$					0.2	0.7
$\mathbf{II}_{X-ray}$						0.6
			Double-helical Region Only			
$\langle \mathbf{I-III} \rangle_{MD}$	1.2	1.8	1.1	2.1	2.0	2.1
$\langle \mathbf{I} \rangle_{MD}$		2.8	1.5	2.1	2.1	2.2
$\langle \mathbf{II} \rangle_{MD}$			2.7	2.7	2.5	2.6
$\langle \mathbf{III} \rangle_{MD}$				2.6	2.6	2.8
$\mathbf{I}_{X-ray}$					0.5	0.8
$\mathbf{II}_{X-ray}$						0.7

A somewhat more finely detailed comparison is provided in Table 2, where various helical parameters computed separately over double-helical base-pair steps and loop base-pair steps are listed. These data suggest that the double-helical structures for all 3 trajectories are very similar (all parameters are identical to within their computed standard deviations), as already suggested by Figure 2. While the simulations showed reduced

TABLE 2: Helical Parameters for Trajectories I-III[a]

		trajectory		
parameter	region	**I**	**II**	**III**
twist, deg	double helix[b]	29.2 (7.2)	30.1 (11.1)	29.8 (9.4)
	loop[c]	36.9 (21.2)	24.1 (42.6)	36.2 (47.7)
tilt, deg	double helix	0.1 (5.2)	−0.3 (7.8)	1.1 (6.3)
	loop	−1.2 (11.4)	−8.1 (14.1)	2.4 (13.7)
rise, Å	double helix	3.6 (0.5)	3.6 (0.6)	3.5 (0.6)
	loop	4.2 (0.9)	4.1 (1.0)	3.9 (0.8)
roll, deg	double helix	9.1 (9.4)	7.9 (10.3)	9.6 (8.8)
	loop	11.8 (14.3)	10.8 (17.5)	10.7 (13.6)
total length, Å	full dodecamer	38.7 (1.0)	33.3 (3.2)	37.3 (1.7)

[a] Expressed as average (standard deviation). Averages were taken over the helix remaining after removal of the terminal base pairs. [b] Includes base pair steps 2, 3, 4, 10, 11, and 12. [c] Includes base pair steps 6, 7, and 8.

helix twist around the 5-bromocytosine base pairs, similar to the crystal structure,[7] no other interesting behavior was seen for the nonstandard bases. The loop regions, on the other hand, effectively have rather ill defined structure, as judged by the very large computed standard deviations, which derive only in part from the limited number of base-pair steps in the sample. We now focus more carefully on this region.

Internal Loop Dynamics. Other than the unpairing of base pair 14:15, no anomalous behavior was seen in the canonically paired regions of the helices (including the C*:G base pairs). The loop regions by contrast, as suggested by their large standard deviations in helical parameters, showed highly dynamical behavior.

This behavior was associated primarily with the central two base pairs of the loop. The G:U pairs showed only a low degree of fluctionality (G:U hydrogen bond distances are available in the Supporting Information). In trajectory **III**, both G:U pairs maintained classical wobble geometries throughout the simulation. In the other two trajectories as well, the G9:U20 pair was always in a wobble configuration, but in trajectory **I** the U6:G23 pair showed some propeller twisting and bending of both bases into the minor groove, while in trajectory **II** the U6:G23 pair sampled not only the classical wobble configuration but also geometries where both bases bent outward into the major groove or where G23 engaged in a cross-strand purine stack (vide infra).

As for the C:U pairs, pairing between the two bases in a quasi-Watson−Crick fashion took place over a range of possible hydrogen bonding motifs, including H42−O4, H41−O4, and H41−O2 interactions. Snapshots spanning this continuum are provided in Figure 3, with tightly bound water molecules indicated in those cases where they were prevalent. Structure **a** involves a fair degree of dislocation of both base pairs into the minor groove while maintaining a hydrogen bond between H42 and O4. Structure **b** has the same hydrogen bonding interaction, but the bases are better incorporated into the overall helix. This structure corresponds to one previously identified by Schneider et al. (called by them "conformation I")[24] as being reasonably persistent in their prior simulation of a UUCG-loop containing dodecamer and it is close to the geometry found in a crystal structure of that same dodecamer.[5] Structure **c** swaps H42 for H41 as the hydrogen-bond donor to O4. Structure **d** is transitional—the cytosine amino group essentially passes over the amide proton of U while the individual bases for the most part maintain their "usual" positions. Finally, structure **e** is characterized by H41 to O2 hydrogen bonding. This arrangement is qualitatively similar to another motif observed by Schneider et al. (their "conformation II")[24] that has also been found in the crystal structure of the *Escherichia coli* 23 S rRNA sarcin/ricin domain.[70]

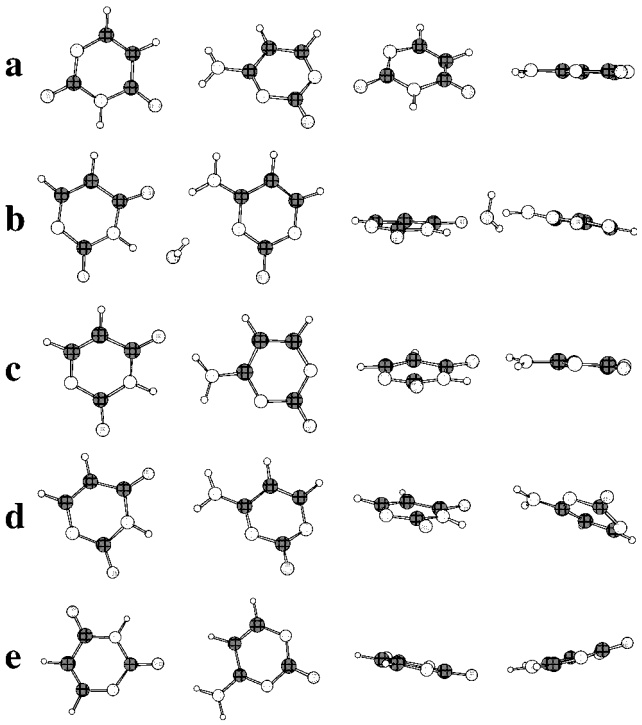

Figure 3. Top (left) and side (right) views of selected geometries across the continuum of C:U base pairing involving hydrogen bonding interactions. Structure **b** includes a water molecule; one is consistently found bridging the two bases in this geometry. The particular structures **a**−**e** derive from (trajectory, time, C:U): (**I**, 800, 7:22), (**II**, 1100, 8:21), (**II**, 500, 8:21), (**II**, 1700, 8:21), and (**II**, 3500, 8:21).

Structures analogous to **b** were observed to incorporate a bridging water molecule over long periods of time (in agreement with similar observations by Schneider et al.[24]), and the relevant water is included in the representation of **b**. In structures **a**, **c**, **d**, and **e**, chains of two or more water molecules were typically observed to bridge hydrogen bonding groups not involved in direct base-base interactions, but no particular geometric arrangement had sufficient lifetime to warrant it being singled out as integral to the base pair structure.

In addition to long-term sampling of distinct regions in this pairing continuum, C:U pairs were also observed to be formally unpaired for significant amounts of time. Transitions between all of these regimes tended to be reasonably sharply defined, as indicated by the time plots of C H41 and H42 distances from U O2 and O4 presented in Figure 4. At the beginning of each trajectory, all C:U base pairs showed hydrogen bonding to O4, although the mode varied between **a**, **b**, and **c**. Switching between structures **b** and **c** occurred at several points (see, for instance, U7:C22 in trajectory **I** over the span from 800 to 1450 ps). In both trajectories **I** and **II**, the C8:U21 base pair switches to a structure **e** motif after 1500 to 2000 ps; in trajectory **I** the U7:C22 pair also adopts this structure at the same time as C8:U21 but shows more dynamical behavior.

In trajectory **III**, there are significant time periods during which the 8:21 pair propeller twists so that H42 of C8 hydrogen bonds to either N3 or O2 of C21, i.e., across the base-pair step (hydrogen bond distances can be found in the Supporting Information). This structure is illustrated in Figure 5 and persists during the periods from 400 to 500, 1350−1550, and after 2050 ps in trajectory **III**. It is also found very briefly in trajectory **II** between 2100 and 2500 ps. In trajectory **III**, the cross-step

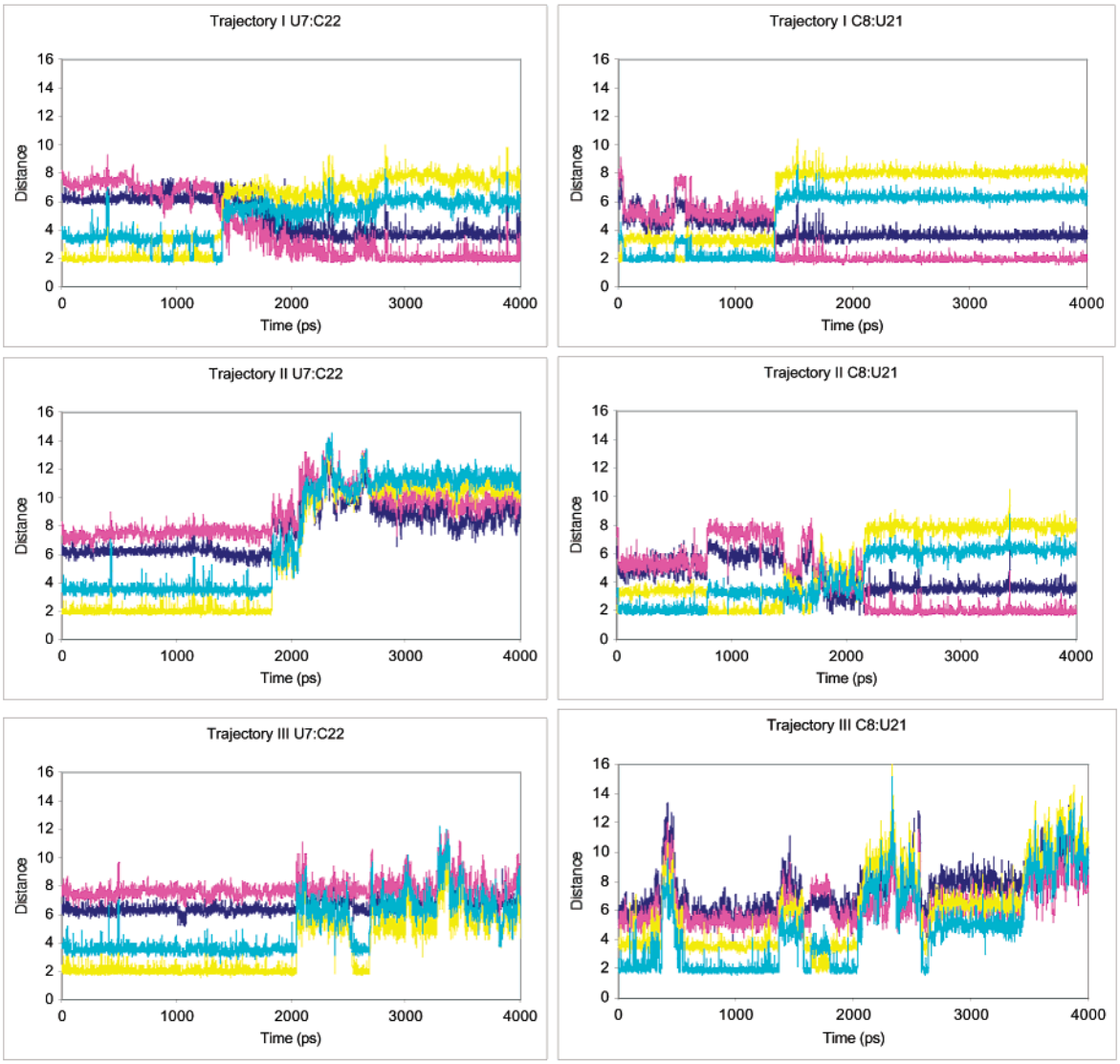

Figure 4. Time plots (ps) of H42 to O4 (cyan), H41 to O4 (yellow), H42 to O2 (magenta), and H41 to O2 (blue) distances (Å) for U7:C22 (left) and C8:U21 (right) base pairs for trajectories **I** (top), **II** (center), and **III** (bottom).

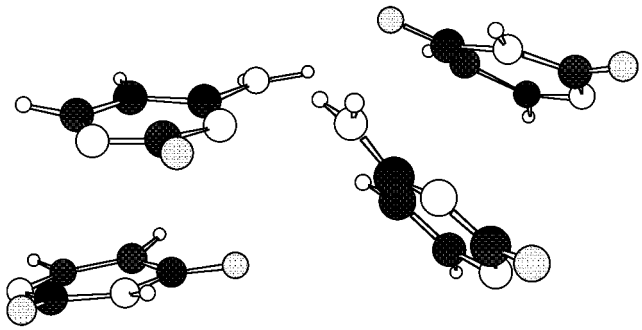

Figure 5. Cross-step hydrogen bonding between the two cytosines viewed from the minor groove (7:22 above, 8:21 below).

interaction causes the 7:21 base pair to become unpaired, beginning at about 2100 ps, although it briefly repairs at about 2350 ps, so the disruption does not appear to be irreversible. Reversible unpairing events have been observed previously in noncanonical base pairs, e.g., adenine−difluorotoluene.[35]

Finally, in trajectory **II**, another unusual bonding scheme persists over a long time period. In this case, a cross-strand stack[21,30,71,72] within the mispaired loop, as displayed in Figure 6. The intercalation of G23 between bases U6 and C22 leads to a highly regular stacked single-stranded helix within the loop that is composed of residues 5, 6, 23, 22, 21, and 20. In the process of forming the stack, U7 becomes translated into the minor groove.

As depicted on the right in Figure 6, three distinct cross-strand stacking interactions were observed. The first, present in principal component (PC) 2, is the cross-strand purine stack mentioned above. The other two cross-strand stacking interactions are observed in animations of PC 3 and involve pyrimidines. At the minimum coefficient of PC 3, the loop region adopts a double cross-strand pyrimidine stack. The C:U pair geometry depicted in Figure 3e has a hydrogen bond to the O2 of uracil. This pairing requires a significant amount of base pair slide within the plane of the base pair. Normally, this slide would disrupt the canonical stacking of each helical strand. However, in the case of the double cross-strand stack, bases 7 and 22 stack on opposite strands, effectively recovering favorable stacking energy. At the opposite extreme of PC 3, the double cross-strand stack is lost to only a single cross strand pyrimidine stack of base 21.

Principal Components Analysis. To assess the degree of coupling, if any, between loop dynamics and overall helical dynamics, we used PCA to analyze the final 3 ns of each

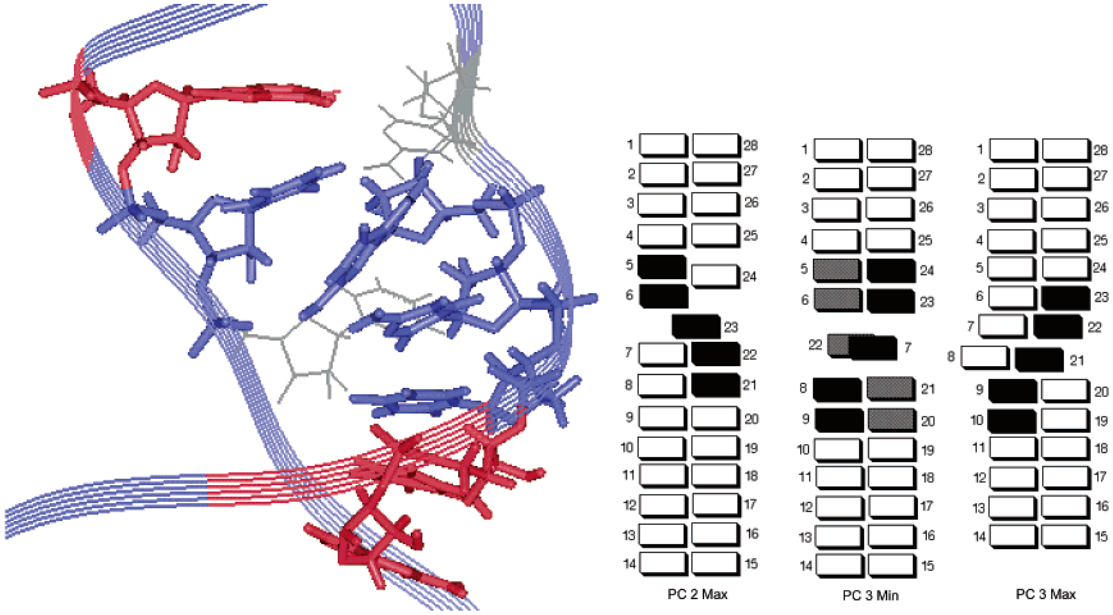

Figure 6. Cross-strand stack persisting in trajectory **II** after 2500 ps. The helix backbone is represented as a ribbon. From top to bottom, the stacked bases are 5 (red), 6, 23, 22, 21, and 20 (red). The three unique cross-strand stacking interactions found in components 2 and 3 are shown to the right. Normal stacking would follow each strand in white, while cross-strand stacking is shown in either black or gray. The helix shown corresponds to the left-most cross-strand stack.

trajectory. Initially, the analysis was carried out over the 9 ns concatenated trajectory (after removal of water) without additional processing. However, the unpairing of the 14:15 terminal base pair led to some ambiguity in assessing the character of the principal components. Modifying the concatenated trajectory so as to hold this pair fixed at a single geometry led to considerably clearer separation between distinct helical motions. The dot products between the first three components of the original and modified trajectories (each in order of eigenvalues) were 0.83, 0.76, and 0.63, indicating that the trajectory processing did not fundamentally change the nature of the key eigenvectors. All further discussion refers to the eigenvectors from the processed trajectory.

The first three eigenvectors capture 42, 16, and 10% (total of 68%) of the total motion, and we restrict our analysis to these modes, effectively reducing the degrees of freedom from 2676 ($3N-6$) to 3. The distribution of structures defined within this basis of eigenvectors is roughly Gaussian for each mode in each trajectory, as illustrated in Figure 7, consistent with their interpretation as fundamental modes. If each simulation were to *completely* sample phase space in an ergodic fashion, then all three distributions for each eigenvector in Figure 7 would perfectly overlap. Since complete sampling of phase space is not a realistic goal in a system of biologically relevant size, we employ multiple simulations to more efficiently sample phase space using standard protocols. Insofar as there is clearly a significant amount of overlap in the individual portions of phase space explored by each simulation, it is clear that the three simulations are not somehow isolated in distinct energy wells.

The first mode, characterized by the largest eigenvalue, is well described as helix twisting. With alterations in helix twist come changes in the length (high twist corresponds to compressed helices) and bending of the helix; we find a linear relationship between the eigenvector coefficient and the computed average helical twist over the full concatenated trajectory (data not shown). The second mode is best described, from visual inspection, as a helix bending about the axis of the major groove in conjunction with a sliding motion of the loop base pairs, and the third mode corresponds to a breathing of the minor groove that effectively bends the 1:28 end of the helix down into the major groove. In Figure 8, the structures characterized by the minimum and maximum coefficients for each eigenvector are provided, and these extrema are consistent with the assignment of the modes.

A simplified measure of the extent to which the individual trajectories overlap in phase space is provided by projection of the trajectories onto a plane defined by two principal modes. The plots in Figure 9 indicate that, while each trajectory is centered about a unique area, regions of overlap exist. In Figure 10 the individual trajectories are projected onto the 3-dimensional phase space defined by these three eigenvectors. This projection best illustrates the degree to which the three trajectories are related, with maximum overlap being at the edges of the phase space regions sampled by trajectories **I** and **III**. Finally, in Figure 11, a 2-dimensional RMSD plot shows that trajectories **I** and **III** are more similar then trajectory **II**. Substates in each of the individual simulations are evident from the subtle shading patterns of the plot, and the latter third part of trajectory **II** has a notable overlap with the whole of both trajectories **II** and **III**.

Proteins have been observed to populate multiple discrete minima, and to jump over free-energy barriers between such wells.[46,62] Multiple wells are characterized by distinct and highly disjoint regions observed on a plane spanned by the first two principal components.[46] We do not see such distinct wells separated by fast transitions in any of the individual trajectories. When the energy of each structure is plotted as a function of the first three modes (available in Supporting Information), the results suggest either a single well, or a flat surface with low energy barriers for this helix, since the barriers are not nearly as pronounced as those seen by de Groot et al.[62]

PCA also proves useful in assigning a quantitative measure of the similarity of the different trajectories with respect to their dynamics. In particular, we have taken the root-

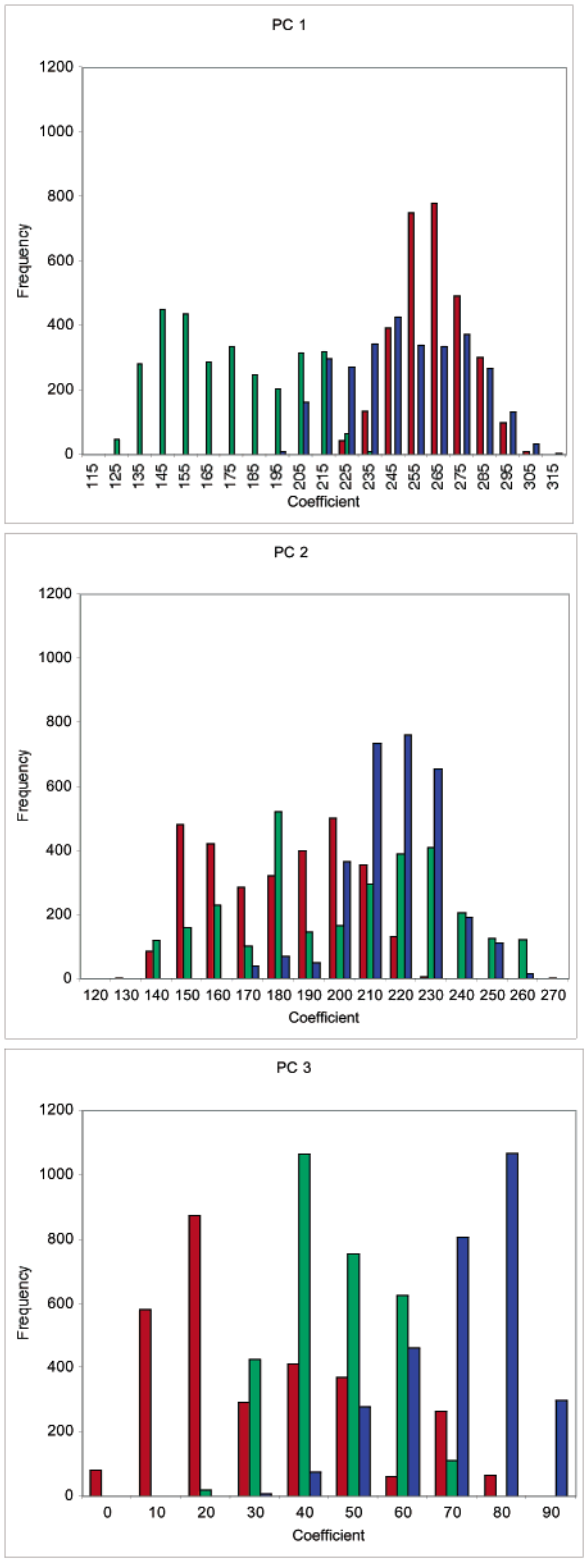

Figure 7. Distributions of eigenvector coefficients describing the structures in trajectories **I** (red), **II** (green), and **III** (blue).

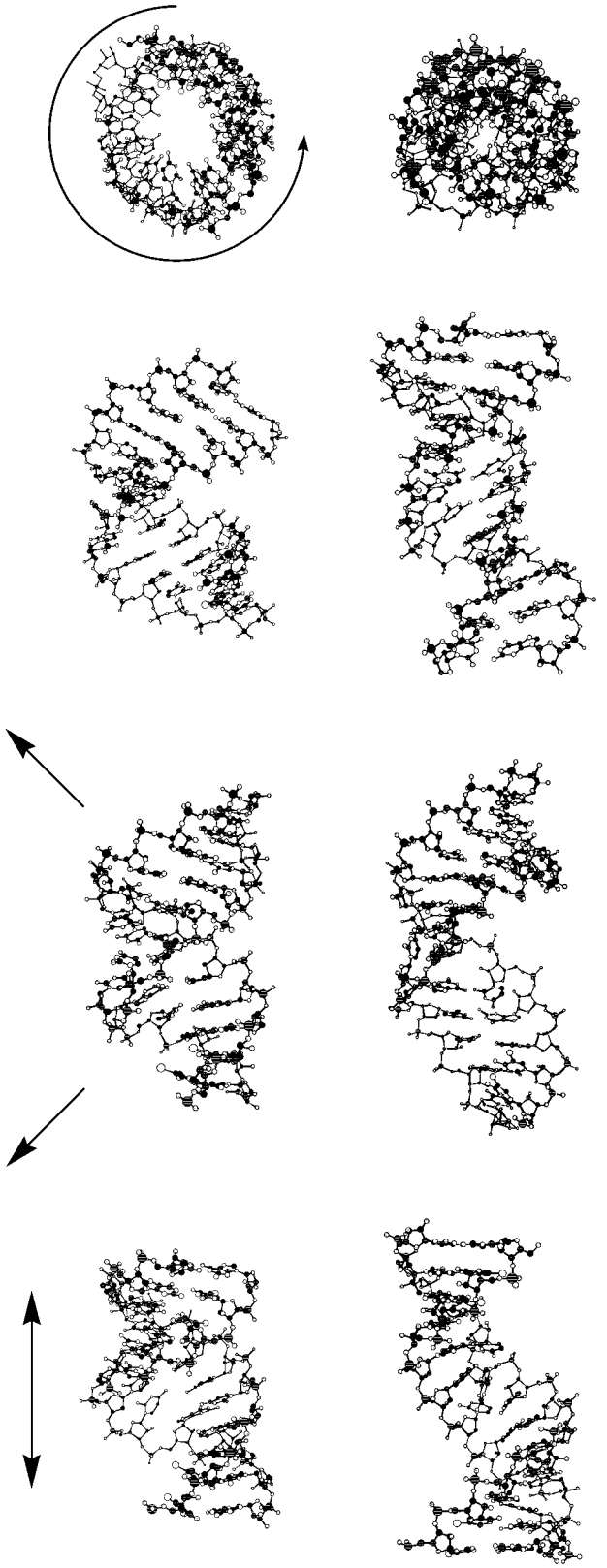

Figure 8. Structures corresponding to extreme values (minimum, left; maximum, right) in the three principal components that have highest weight in the PCA analysis. Arrows indicate the sense of each motion. At top, twisting is shown for both side-on and end-on views of the helix (mode 1). Below this is helix bending (mode 2) and at the bottom is minor groove breathing leading to an asymmetric bend of the 1:28 helix end (mode 3).

mean-square inner product (RMSIP)

$$\mathrm{RMSIP} = \left[\frac{\sum_{i=1}^{N}\sum_{j=1}^{N}(\mathbf{r}_i \cdot \mathbf{r}_j)^2}{N}\right]^{1/2} \qquad (2)$$

where i and j run over the first N normalized PCA components **r** of two different trajectories. If the two sets of eigenvectors span exactly the same dynamical space, RMSIP = 1, while if

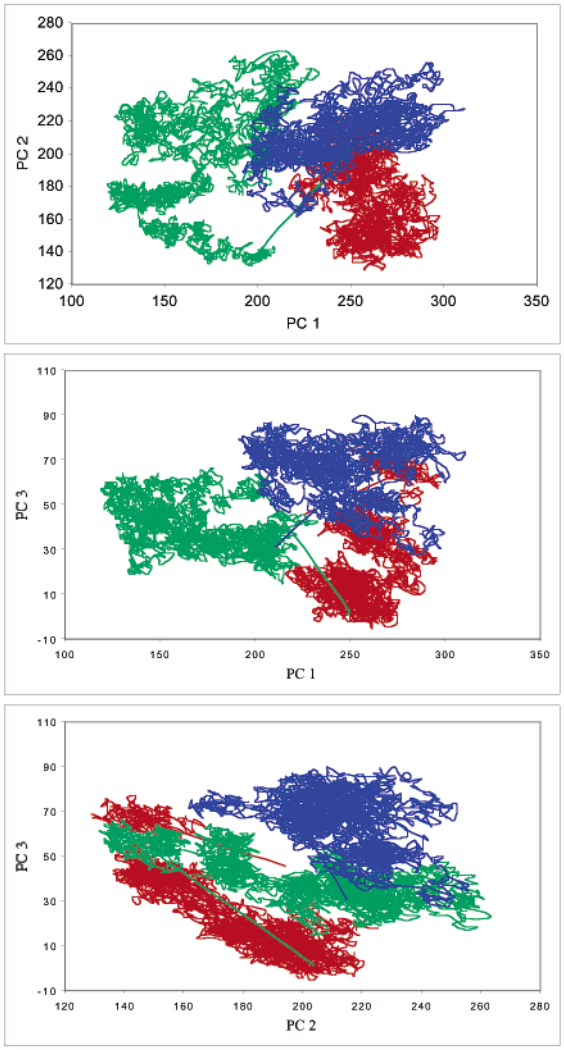

Figure 9. Trajectories **I** (red), **II** (green), and **III** (blue) projected onto the 2-dimensional spaces formed by pairings of the first three largest principal components.

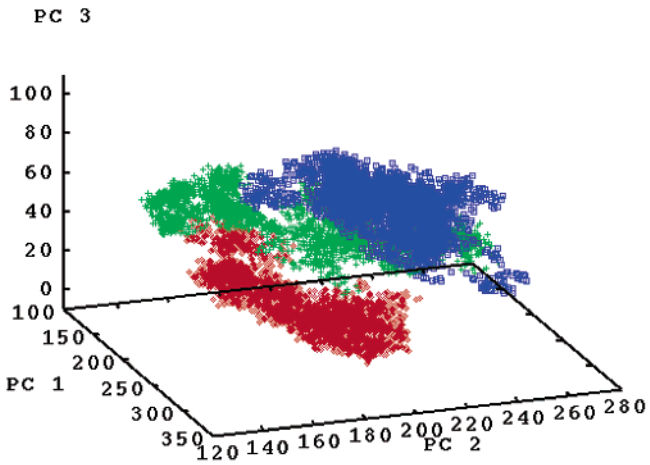

Figure 10. Trajectories **I** (red), **II** (green), and **III** (blue) projected onto the 3-dimensional space formed by the three largest principal components.

there is no correspondence of any kind RMSIP = 0. We report the RMSIP over the first 10 PCA eigenvectors between various trajectories in Table 3. To evaluate both the influence of salt on the dynamics and the impact of including 5-bromocytosine in place of cytosine (the former is present in the crystal structure as well as our simulations, but not in the prior simulation of Schneider et al.[24]), trajectory **Ia** and the generalized Born trajectories with and without 5-bromocytosine are included in this analysis.

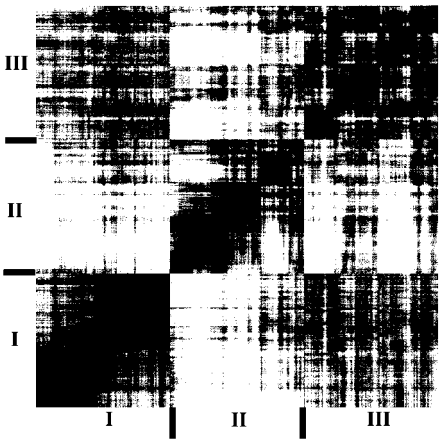

Figure 11. Two-dimensional RMSD plot of the 9 ns concatenated trajectory. The plot ranges from 0.0 Å (black) to 6.0 Å (white). Each main block lying along the diagonal represents an individual trajectory, and substates within each trajectory are evident as smaller shaded blocks.

TABLE 3: Root-Mean-Square Inner Product between First 10 PCA Eigenvectors for Various Pairs of Trajectories

	RMSIP, %					
structure	**I**[a]	**II**[a]	**III**[a]	**Ia**[a]	$\mathbf{GB}_{C}$[b]	$\mathbf{GB}_{C*}$[b]
I−**III**[c]	74.6	48.9	71.7	70.7	48.4	45.3
I[a]		51.9	74.5	81.1	54.6	51.0
II[a]			53.6		60.2	48.8
III[a]					50.6	49.6
$\mathbf{GB}_{C}$[b]						60.5

[a] Final 3 ns of 4 ns trajectory. [b] Final 6 ns of 10 ns trajectory. [c] 9 ns trajectory formed from catenation of final 3 ns of trajectories **I**, **II**, and **III**.

Discussion

Assessing Convergence in Flexible Models. Two articles of faith typically endorsed in biomolecular simulations are that (i) most biopolymers have definite structures about which they fluctuate to only a small extent and (ii) the best mark of convergence in the simulation of such a biopolymer is a roughly flat RMSD plot of the trajectory against some standard, e.g., the average structure. It is clear that the tetradecamer studied here, which contains a UUCG loop, does not adhere particularly well to the first stricture, as we see rather significant variations in overall structure, and as discussed in more detail below, variations in internal loop structure couple with variations in double-helical structure in such a way as to make a very large volume of phase space accessible to this particular RNA duplex.

Given that situation, it is perhaps unsurprising that the second tenet noted above also fails to be of much utility. The RMSD plots in Figure 1 convey remarkably little information. Indeed, the standard practice of comparing the trajectory to its own average structure would tend to argue for convergence, but it is clear from examining other comparisons (e.g., one trajectory against another trajectory's average structure) or from focusing more finely on particular regions (e.g., via hydrogen bonding analysis or computation of helical parameters) that lack of persistent structure in some portion of the polymer can easily be missed when only global data analysis is undertaken. Perhaps more critically, in structures having a high degree of flexibility,

it is *inherently* unrealistic to expect an RMSD plot to tend toward "flatness". Oscillations about some median, associated with the large-scale motions of the molecule, will instead be characteristic, and such oscillations may be difficult to differentiate from unconverged behavior.

An alternative measure of convergence, which enjoys some virtue of simplicity, is principal components analysis. Because the principal components can be ordered in terms of importance (measured as percentage of variance accounted for), the full 3*N*-dimensional trajectory in Cartesian coordinate space can be projected very efficiently into a space of considerably reduced dimensionality with minimal loss of information. A converged trajectory should in principle sample phase space in an approximately Gaussian fashion about some center of mass, and an inspection of the distribution of structures projected into the reduced phase space expressed in the PCA coordinates provides a rapid approach to such an evaluation. Using Figure 7 and judging by that criterion, trajectories **I** and **II** appear stable and converged. In PC 2, trajectory **II** shows some bimodality, but the distribution is not so asymmetric as to make further analysis dubious.

An additional point meriting discussion is that on the time scale of the simulation, all 3 trajectories appear to be reasonably stable, but presumably over longer time scales than can easily be addressed with current simulation technology, an individual trajectory might transition from its own region of phase space to another's. We would argue, echoing others,[25−29] that for biomolecules where reduced structural rigidity is an issue, running several simulations of moderate length is likely to offer more comprehensive information than running a single simulation of longer length. That being said, it is obviously true that any *finite* number of trajectories may miss an important region of phase space, and there is no way to completely avert that possibility.

C:U Pair Flexibility. The vast majority of structural data for C:U (mis)pairs, whether experimental or computational, comes from single mismatches in otherwise canonical double helices,[22] mismatches marking the head of a hairpin loop,[19,20] or from internal loops having two consecutive C:U pairs.[5,7,8] In the case of an isolated C:U base pair, the ability of a strongly bound water to improve the otherwise rather weak base pairing available within an overall structure that preserves a canonical double helical form[5,17] has led Brandl et al.[17] to suggest that the C:U:water complex should be considered as a composite structural unit. Structure **b** of Figure 3 is representative of this water-mediated motif, which involves direct hydrogen bonding between O4 of uracil and H42 of cytosine.

While such a structure is indeed found in many situations,[5−8,20,22] it is not universal. Other structures analogous to **e** of Figure 3 involving H41−O2 hydrogen bonding have also been identified.[18,19,23,70] It is clear, then, that there is a fair degree of flexibility associated with the C:U mispair, and one expects this flexibility to increase if the local region is characterized by additional mispairs, either of the C:U variety or otherwise, as is the case for the simulations carried out here (and, of course, is evidenced by the three distinct geometries found in the unit cell of the X-ray crystal structure[7]). This contrasts somewhat with the well characterized G:U mispair, where a water-mediated wobble geometry appears to be quite stable in almost every observed instance. Indeed, in the current simulations such a geometry was also observed to be highly persistent, albeit with a few noteworthy deviations—the presence of 4 mispaired bases in an internal loop allows other interactions such as stacking and bifurcated hydrogen bonding also to be explored.

Within the context of the simulations reported here, many reversible changes between structures **a**, **b**, and **c** of Figure 3 are observed to occur (see Figure 4), and structure **e** is also observed to be capable of significant persistence. One item not previously observed that is dependent on the adjacency of two consecutive C:U pairs is the possibility of cross-step hydrogen bonding, as illustrated in Figure 5. This motif was observed in five separate instances in trajectories **II** and **III** and persisted anywhere from 50 to about 450 ps. A second motif dependent on multiple mismatches is the cross-strand stacking of G23 noted in Figure 6, which, once adopted, also proved to be highly persistent, lasting the better part of 2 ns in trajectory II. The tendency for the loop region to adopt any particular one of these several different, metastable geometries is coupled with the overall helix dynamics, as assessed by PCA.

Loop−Helix Coupling. In past studies where PCA has been applied to DNA, the modes of deformation that characterize the most important motions of DNA are twisting and bending motions.[36] These modes, when animated, show a general motion uniformly affecting the entire helix. In the RNA modes identified here, by contrast, the motions are evenly distributed over the double-helical regions, but they are also coupled with specific loop geometry changes in various instances. This coupling is particularly interesting to the extent it illustrates how local changes in loop structure may potentially drive more global changes in overall helix structure.

By plotting the PCA coefficients as a function of time (data in the Supporting Information), it is straightforward to identify correlations between specific loop motifs and particular principal components values. For instance, maximal values of the first principal component, a situation that corresponds to a highly twisted helix, are observed for those time periods where both C:U pairs adopt geometries similar to that of structure **b** in Figure 3 (e.g., from 1000 to 1300 ps in trajectory **I**).

In addition, the lowest values of the first principal component are observed to correlate with the existence of the cross-strand stack in trajectory **II**, provided the third principal component simultaneously takes on large values. This situation corresponds to an underwound, elongated helix. Within the loop region, movement of the helix in such a fashion has an effect rather like closing a zipper: the base "teeth" slip between one another at the step indicated in Figure 6. Cross-strand purine stacks have previously been observed in other situations.[21,71,72]

Helix compression, which is present in structures characterized by high coefficients of principal components 1 and 3, is associated with the cross-step hydrogen bonding observed in trajectories **II** and **III** and illustrated in Figure 5. This correlation is certainly reasonable, insofar as the different base pairs are brought nearer to one another by movement along principle component 3 in the compression direction. Careful inspection of individual trajectories suggests that when principal component 2 reaches its maximal values, there is some translation of C8 and U7 into the minor groove.

Roughly speaking, low coefficients for modes 1, 2, and midrange values of 3 correspond to the most canonical A-form helices. High coefficients for mode 2 are associated with the purine cross-strand stack, while the extrema of mode 3 are associated with single and double cross-strand pyrimidine stacking interactions. Variations in structural energy over the course of the three trajectories indicate that the area of phase space explored by trajectory **I** has lowest energy, followed by **III**, and then **II**, but it must be recalled that this is a measure of potential energy, not free energy. On the basis of potential energy analysis, the formation of a cross-strand purine stack

seen predominantly in trajectory **II** is destabilizing compared to a canonical conformation, and a double cross-strand pyrimidine stack is more favorable than just a single cross-strand pyrimidine stack.

Possible Sources of Error. All of our different trajectories indicate the loop region of the (r-GC*C*AG**UUCG**CUGGC)$_2$ tetradecamer to be highly dynamic, although certain discrete structural motifs are repeatedly sampled. Schneider et al.[24] also observed dynamic behavior in a single trajectory of a related dodecamer, but the number of different loop structures sampled was substantially smaller. By contrast, Anderson et al., on the basis of their comparison of three different (r-GC*C*AG**UU-CG**CUGGC)$_2$ tetradecamer structures in a single-crystal X-ray unit cell, inferred the UUCG loop structure to be robust and well defined as a unique structural unit. These variations in observation/interpretation motivate some discussion on possible sources of difference between the X-ray single-crystal experiment and the two simulations.

First and foremost, the simulations take place in an environment that compares more closely to a solution structure than to a single-crystal structure, and we consider it most likely that the constraints imposed by crystal periodicity significantly reduce the conformational options available to the UUCG loop (it should be noted that it is certainly possible that different crystal polymorphs exist for different UUCG conformations, but only the one X-ray crystal study has been reported). Alternatively, of course, the force fields used in the simulations may be inadequate to the task. Obviously, no classical force field can be perfect; however, the force field of Cornell et al.[30] has proven remarkably successful in predicting structural and dynamical properties of DNA and RNA,[38] to include properties of UUCG loops.[4,16,73,74]

Other possible sources of differences between the three analyses are variations in the environmental conditions. To assess the effect of monovalent ionic strength, we generated one trajectory (**Ia**) at approximately twice the ionic strength of **I**, **II**, and **III**; the ionic strength for **Ia** (0.1 N) matched that used by Schneider et al.[24] for their dodecamer simulation. Table 3 indicates that the variation in ionic strength appears to have negligible impact on the helix dynamics: the RMSIP between **I** and **Ia** is very high, at 81.1%, and the RMSIP between **Ia** and the other trajectories is essentially the same as that between **I** and the other trajectories.

A separate source of varation is the presence or absence of C* in place of C. In this work, we have parametrized C* using standard methods, but limited data are available against which to validate this parametrization. To assess this issue more closely, we carried out less costly implicit-solvent dynamics using a generalized Born formalism including ionic strength effects[63,65,66,75] for the tetradecamer either with C* or C at positions 2, 3, 16, and 17. The RMSIP between the final 6 ns of these two trajectories is 60.5% (Table 3). This value is just slightly higher than the RMSIP between either one of the two implicit-solvent trajectories and any one of the 3 explicit-solvent trajectories and is very similar to values observed previously in comparing GB trajectories to explicitly solvated trajectories.[63] We infer from this analysis that irrespective of the absolute quality of our C* parameters, the substitution of C* for C does not have any significant impact on the helix dynamics. The higher degree of flexibility observed in our simulations compared to the single trajectory of Schneider et al.[24] may thus be associated with the limited sampling of phase space possible for any single trajectory (or small collection of trajectories) that is always a limitation in molecular dynamics.

Conclusions

Molecular dynamics simulations suggest that UUCG loops within RNA double helices are even more flexible than has been inferred from analysis of X-ray crystal structures, NMR, and prior dynamics studies. In the examination of four separate trajectories, new geometric motifs within the loops involving interpair interactions have been identified. In particular, cross-step hydrogen bonding between the adjacent C:U pairs and cross-strand stacking interactions involving mispaired G, and cross-strand stacking of pyrimidines have been identified as persistent structural elements under appropriate conditions. Principal component analysis of global motion in an internal loop-containing tetradecamer suggests that specific loop structures are correlated with particular overall helical structures, and the potential to drive one by modification of the other offers interesting possibilities for nucleic acid recognition and processing.

Acknowledgment. We thank the Alfred P. Sloan Foundation, the Spanish Ministry of Education and Culture, and Fundación BBV for financial support, and the Minnesota Supercomputer Institute for computational resources. Profs. Javier Luque and Modesto Orozco are thanked for discussions and hospitality. Prof. Charles Laughton and Dr. Sarah Harris are thanked for discussions and technical assistance.

Supporting Information Available: 5-Bromocytosine parameters, equilibration protocol, RMSD plots of trajectories against starting structures, time propagation of PCA coefficients, energies as a function of coefficient, C8−C22 hydrogen bond distances, G:U wobble pair hydrogen-bond distances, and hydrogen-bond distances for loop pairs in trajectory **Ia**. This material is available free of charge via the Internet at http://pubs.acs.org.

References and Notes

(1) Tuerk, C.; Gauss, P.; Thermes, C.; Groebe, D. R.; Gayle, M.; Guild, N.; Stormo, G.; D'Aubenton-Carafa, Y.; Uhlenbeck, O. C.; Tinoco, I.; Brody, E. N.; Gold, L. *Proc. Natl. Acad. Sci. U.S.A.* **1988**, *85*, 1364.

(2) Cheong, C.; Varani, G.; Tonoco, I. *Nature* **1990**, *346*, 680.

(3) Sakata, T.; Hiroaki, H.; Oda, Y.; Tanaka, T.; Ikehara, M.; Uesugi, S. *Nucl. Acids Res.* **1990**, *18*, 3831.

(4) Williams, D. J.; Hall, K. B. *J. Mol. Biol.* **2000**, *297*, 251.

(5) Holbrook, S. R.; Cheong, C.; Tinoco, I.; Kim, S. H. *Nature* **1991**, *353*, 579.

(6) Cruse, W. B. T.; Saludjian, P.; Biala, E.; Strazewski, P.; Prangé, T.; Kennard, O. *Biochemistry* **1994**, *33*, 4160.

(7) Anderson, A. C.; O'Neil, R. H.; Filman, D. J.; Frederick, C. A. *Biochemistry* **1999**, *38*, 12577.

(8) Tanaka, Y.; Fujii, S.; Hiroaki, H.; Sakata, T.; Tanaka, T.; Uesugi, S.; Tomita, K.-i.; Kyogoku, Y. *Nucl. Acids Res.* **1999**, *27*, 949.

(9) Crick, F. H. C. *J. Mol. Biol.* **1966**, *19*, 548.

(10) Betzl, C.; Lorenz, S.; Furste, J. P.; Bald, R.; Zhang, M.; Schneider, T. R.; Wilson, K. S.; Erdmann, V. A. *FEBS Lett.* **1994**, *351*, 159.

(11) Limmer, S.; Reif, B.; Ott, G.; Lubos, A.; Sprinzl, M. *FEBS Lett.* **1996**, *385*, 15.

(12) Biswas, R.; Wahl, M. C.; Ban, C.; Sundaralingam, J. *J. Mol. Biol.* **1997**, *267*, 1149.

(13) Ramos, A.; Varani, G. *Nucl. Acids Res.* **1997**, *25*, 2083.

(14) Vogtherr, M.; Schübel, H.; Limmer, S. *FEBS Lett.* **1998**, *429*, 21.

(15) Mueller, U.; Schübel, H.; Sprinzl, M.; Heinemann, U. *RNA* **1999**, *5*, 670.

(16) Nagan, M. C.; Kerimo, S. S.; Musier-Forsyth, K.; Cramer, C. J. *J. Am. Chem. Soc.* **1999**, *121*, 7310.

(17) Brandl, M.; Meyer, M.; Sühnel, J. *J. Phys. Chem. A* **2000**, *104*, 11177.

(18) Mirmira, S. R.; Tinoco, I. *Biochemistry* **1996**, *35*, 7664.

(19) Kolk, M. H.; Heus, H. A.; Hilbers, C. W. *EMBO J.* **1997**, *16*, 3685.

(20) Lynch, S. R.; Tinoco, I. *Nucl. Acids Res.* **1998**, *26*, 980.

(21) Correll, C. C.; Munishkin, A.; Chan, Y. L.; Ren, Z.; Wool, I. G.; Steitz, T. A. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 13436.

(22) Tanaka, Y.; Kojima, C.; Yamazaki, T.; Kodama, T. S.; Yasuno, K.; Miyashita, S.; Ono, A. M.; Ono, A. S.; Kainosho, M.; Kyogoku, Y. *Biochemistry* **2000**, *39*, 7074.
(23) Ban, N.; Nissen, P.; Hansen, J.; Moore, P. B.; Steitz, T. A. *Science* **2000**, *289*, 905.
(24) Schneider, C.; Brandl, M.; Sühnel, J. *J. Mol. Biol.* **2001**, *305*, 659.
(25) Gibson, Q. H.; Regan, R.; Elber, R.; Olson, J. S.; Carver, T. E. *J. Biol. Chem.* **1992**, *267*, 22022.
(26) Carlson, M. L.; Regan, R. M.; Gibson, Q. H. *Biochemistry* **1996**, *35*, 1125.
(27) Auffinger, P.; Westhof, E. *J. Mol. Biol.* **1997**, *269*, 326.
(28) Caves, L. S. D.; Evanseck, J. D.; Karplus, M. *Protein Sci.* **1998**, *7*, 649.
(29) Schulze, B. G.; Evanseck, J. D. *J. Am. Chem. Soc.* **1999**, *121*, 6444.
(30) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. *J. Am. Chem. Soc.* **1995**, *117*, 5179.
(31) Cheatham, T. E., III.; Miller, J. L.; Fox, T.; Darden, T. A.; Kollman, P. A. *J. Am. Chem. Soc.* **1995**, *117*, 4193.
(32) Cheatham, T. E., III.; Crowley, M. F.; Fox, T.; Kollman, P. A. *Proc. Natl. Acad. Sci. U.S.A.* **1997**, *94*, 9626.
(33) Shields, G. C.; Laughton, C. A.; Orozco, M. *J. Am. Chem. Soc.* **1997**, *119*, 7463.
(34) Cheatham, T. E., III.; Kollman, P. A. In *Structure, Motion, Interaction, and Expression of Biological Macromolecules*; Sarma, R. H., Sarma, M. H., Eds.; Adenine Press: New York, 1998; p 99.
(35) Cubero, E.; Sherer, E. C.; Luque, F. J.; Orozco, M.; Laughton, C. A. *J. Am. Chem. Soc.* **1999**, *121*, 8653.
(36) Sherer, E. C.; Harris, S. A.; Soliva, R.; Orozco, M.; Laughton, C. A. *J. Am. Chem. Soc.* **1999**, *121*, 5981.
(37) Nagan, M. C.; Beuning, P.; Musier-Forsyth, K.; Cramer, C. J. *Nucl. Acids Res.* **2000**, *28*, 2527.
(38) Cheatham, T. E., III.; Kollman, P. A. *Annu. Rev. Phys. Chem.* **2000**, *51*, 435.
(39) Soliva, R.; Sherer, E.; Luque, F. J.; Laughton, C. A.; Orozco, M. *J. Am. Chem. Soc.* **2000**, *122*, 5997.
(40) Beuning, P. J.; Nagan, M. C.; Cramer, C. J.; Musier-Forsyth, K.; Gelpí, J.-Ll.; Bashford, D. *RNA*, in press.
(41) Kallick, D. A.; Nagan, M. C.; Beuning, P. J.; Kerimo, S.; Tessmer, M. R.; Cramer, C. J.; Musier-Forsyth, K. *J. Phys. Chem. B*, submitted for publication.
(42) de Groot, B. L.; van Aalten, D. M. F.; Amadei, A.; Berendsen, H. J. C. *Biophys. J.* **1996**, *71*, 1707.
(43) Balsera, M. A.; Wriggers, W.; Oono, Y.; Shulten, K. *J. Phys. Chem.* **1996**, *100*, 2567.
(44) van Aalten, D. M. F.; de Groot, B. L.; Findlay, J. B. C.; Berendsen, H. J. C.; Amadei, A. *J. Comput. Chem.* **1997**, *18*, 169.
(45) Wlodek, S. T.; Clard, T. W.; Scott, L. R.; McCammon, J. A. *J. Am. Chem. Soc.* **1997**, *119*, 9512.
(46) Kitao, A.; Hayward, S.; Go, N. *Proteins Struct. Funct. Genet.* **1998**, *33*, 496.
(47) Matsumoto, A.; Go, N. *J. Chem. Phys.* **1999**, *110*, 11070.
(48) Amadei, A.; Ceruso, M. A.; Di Nola, A. *Proteins Struct. Funct. Genet.* **1999**, *36*, 419.
(49) Zacharias, M. *Biopolymers* **2000**, *54*, 547.
(50) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Zakrzewski, V. G.; Montgomery, J. A.; Stratmann, R. E.; Burant, J. C.; Dapprich, S.; Millam, J. M.; Daniels, A. D.; Kudin, K. N.; Strain, M. C.; Farkas, O.; Tomasi, J.; Barone, V.; Cossi, M.; Cammi, R.; Mennucci, B.; Pomelli, C.; Adamo, C.; Clifford, S.; Ochterski, J.; Petersson, G. A.; Ayala, P. Y.; Cui, Q.; Morokuma, K.; Salvador, P.; Dannenberg, J. J.; Malick, D. K.; Rabuck, A. D.; Raghavachari, K.; Foresman, J. B.; Cioslowski, J.; Ortiz, J. V.; Baboul, A. G.; Stefanov, B. B.; Liu, G.; Liashenko, A.; Piskorz, P.; Komaromi, I.; Gomperts, R.; Martin, R. L.; Fox, D. J.; Keith, T.; Al-Laham, M. A.; Peng, C. Y.; Nanayakkara, A.; Challacombe, M.; Gill, P. M. W.; Johnson, B.; Chen, W.; Wong, M. W.; Andres, J. L.; Gonzalez, C.; Head-Gordon, M.; Replogle, E. S.; Pople, J. A. *Gaussian 98*, Revision A.10; Gaussian, Inc.: Pittsburgh, PA, 2001.
(51) Case, D. A.; Pearlman, D. A.; Cladwell, J. W.; Cheatham, T. E.; Ross, W. S.; Simmerling, C. L.; Darden, T. A.; Merz, K. M.; Stanton, R. V.; Cheng, A. L.; Vincent, J. J.; Crowley, M.; Ferguson, D. M.; Radmer, R. J.; Seibel, G. L.; Singh, U. C.; Weiner, P. K.; Kollman, P. A. *AMBER*, version 5.0; University of California: San Francisco, 1997.
(52) Shields, G. C.; Laughton, C. A.; Orozco, M. *J. Am. Chem. Soc.* **1998**, *120*, 5895.
(53) Darden, T.; York, D.; Pedersen, L. *J. Chem. Phys.* **1993**, *98*, 10089.
(54) Essmann, U.; Perera, L.; Berkowitz, M. L.; Darden, T.; Lee, H.; Pedersen, L. G. *J. Chem. Phys.* **1995**, *103*, 8577.
(55) Berendsen, H. J. C.; Postma, J. P. M.; van Gunsteren, W. F.; DiNola, A.; Haak, J. R. *J. Comput. Phys.* **1984**, *81*, 3684.
(56) Ryckaert, J.; Ciccotti, G.; Berendsen, H. J. C. *J. Comput. Phys.* **1977**, *23*, 327.
(57) Lavery, R.; Skelnar, H. *J. Biomol. Struct. Dynam.* **1988**, *1*, 63.
(58) Lavery, R.; Skelnar, H. *J. Biomol. Struct. Dynam.* **1989**, *4*, 655.
(59) Lavery, R.; Skelnar, H. *Curves*, version 5.3; Institut de Biologie Physico-Chimique: Paris, 1998.
(60) García, A. E.; Blumenfeld, R.; Hummer, G.; Krumhansl, J. A. *Physica D* **1997**, *107*, 225.
(61) Kitao, A.; Go, N. *Curr. Opin. Struct. Biol.* **1999**, *9*, 164.
(62) de Groot, B. L.; Daura, X.; Mark, A. E.; Grubmüller, H. *J. Mol. Biol.* **2001**, *309*, 299.
(63) Tsui, V.; Case, D. A. *J. Am. Chem. Soc.* **2000**, *122*, 2489.
(64) Bostock-Smith, C.; Harris, S. A.; Laughton, C. A.; Searle, M. S. *Nucl. Acids Res.* **2001**, *29*, 693.
(65) Srinivasan, J.; Trevathan, M. W.; Beroza, P.; Case, D. A. *Theor. Chem. Acc.* **1999**, *101*, 426.
(66) Bashford, D.; Case, D. A. *Annu. Rev. Phys. Chem.* **2000**, *51*, 129.
(67) Weber, G. *Adv. Protein Chem.* **1975**, *29*, 1.
(68) Cooper, A. *Proc. Natl. Acad. Sci. U.S.A.* **1976**, *73*, 2740.
(69) Elber, R.; Karplus, M. *Science* **1987**, *235*, 318.
(70) Correll, C. C.; Wool, I. G.; Munishkin, A. *J. Mol. Biol.* **1999**, *292*, 275.
(71) Butcher, S. E.; Dieckmann, T.; Feigon, J. *EMBO J.* **1997**, *16*, 7490.
(72) Mueller, U.; Maier, G.; Onori, A. M.; Cellai, L.; Heumann, H.; Heinemann, U. *Biochemistry* **1998**, *37*, 12005.
(73) Zacharias, M. *Biophys. J.* **2001**, *80*, 2350.
(74) Burkhardt, C.; Zacharias, M. *Nucl. Acids Res.* **2001**, *29*, 3910.
(75) Onufriev, A.; Bashford, D.; Case, D. A. *J. Phys. Chem. B* **2000**, *104*, 3712.