Journal of
Computational and Theoretical Nanoscience
Vol. 3, 63–77, 2006

RESEARCH ARTICLE

Structural and Dynamical Classification of RNA Single-Base Bulges for Nanostructure Design

Whitney A. Hastings,[1,2] Yaroslava G. Yingling,[1] Gregory S. Chirikjian,[2] and Bruce A. Shapiro[1,*]

[1] Center for Cancer Research Nanobiology Program, National Cancer Institute, NCI-Frederick, Frederick, MD 21702, USA
[2] Department of Mechanical Engineering, The Johns Hopkins University, Baltimore, MD 21218, USA

A comprehensive examination of the bulge motif is presented for a set of twenty single-base bulge structures obtained from the Protein Data Bank. Examples of the bulge motif found in X-ray-crystal and NMR structures are analyzed using molecular dynamics simulations. Three classes of the single-base bulge motif are defined according to the bulge residue type, and its surrounding base-pairs. The first class contains bulges in a stacked conformation, while the other two classes have the bulge predominantly or exclusively in a looped-out conformation. In the first class the bulges participate in hydrogen bond interactions with one of the neighboring base-pairs. While this modifies the backbone of the structure, the overall backbone shape typically remains relatively constant due to the absence of distal bonding across the helix grooves. In contrast, most of the bulges in the looped-out conformations (second and third class) create a more flexible and a more distinctive kink or induced bending along the backbone. In the second class, the orientation of bulge bases depends on their type and surrounding sequences, whereas the third class contains only cytosine bulges that prefer to remain exclusively in the looped-out conformation. An ultimate goal of this study is to utilize the bulge classes and the structural characteristics of the bulges for the design of RNA-bulge-based nanotemplates.

Keywords: RNA Structural Motif, RNA Bulge, Molecular Dynamics, RNA Design.

1. INTRODUCTION

The most common structural element in RNA is the A-form double helix which accounts for nearly 50% of the residues in a standard RNA structure.[1] The remaining structural elements or motifs account for the wide variety of topological features observed in RNA structures. Examples of RNA motifs include base triplets, quadruplets, internal and hairpin loops, multibranch loops, and more complex motifs including pseudoknots, kissing hairpin loops, and ribose zippers. Through the detailed examination of X-ray crystallography and nuclear magnetic resonance (NMR) structures, insight can be gained into the resulting functionality of many RNA structures and motifs which could be used in drug design and nanotechnology applications.

Twenty five years ago Seeman introduced the branch of nanotechnology which uses the structural complexities found in nature's biological molecules to engineer functional devices and materials.[2] Until recently the focus of this technology has been mostly on DNA building blocks. DNA molecules have already been used to create complex well-defined nanostructure arrays in the form of crystals, ribbons, and octahedrons.[3–5] Other recent advances in DNA nanotechnology include nanoscale mechanical devices, such as nanotweezers and molecular lithography on substrate DNA molecules for the design of novel molecular-scale electronic devices.[6,7] Currently studies are finding RNA to be an equally, if not more desirable material for designing functional nanostructures. RNA's wider variety of three-dimensional structural motifs allows for greater diversity and a broader range of array patterns and designs.[8–10] Recently, secondary structure design paradigms of RNA multibranch loop structures have been studied using RNA thermodynamics.[11] Jaegar et al. have shown that manipulation of RNA tertiary interactions can yield three-dimensional self-assembling molecular units or tectoRNAs.[12] Chworos et al. have designed programmable RNA building blocks that consist of four tectoRNA blocks that assemble into two-dimensional RNA

*Author to whom correspondence should be addressed.

 1546-198X/2006/3/063/015 doi:10.1166/jctn.2006.005

tectosquares like a jigsaw puzzle.[13] This work emphasizes the importance of RNA hierarchy; small RNA motifs form the topology of large molecular structures. These structural scaffolds created by simple yet functional tectoshapes can be used to build much larger, more complex structures. Three-dimensional RNA superstructures or tectoshapes could be designed to have a biological function, such as binding to a target molecule or changing conformation as a function of the environment. However, in order to construct nanoscale functional molecules we should first understand the particular characteristics of different RNA motifs. It has been previously indicated that non-canonical base-pair motifs can be used to tune the overall helical twist.[12] Similarly, bulge motifs are ideal structures for nanotemplate design due to their ability to induce varying degrees of bending and conformational changes in RNA structures. Therefore, the optimization of bulge characteristics such as position, type, and surrounding context can be very important for the design of nanostructures requiring these qualities.

The RNA single-base bulge motif is an unpaired residue within a strand of several complementary base-pairs. Single and multiple base bulges frequently found in RNA structures are important for the tertiary folding process.[14] Moreover, the bulge regions are known to be specific sites for RNA-protein recognition[15,16] and are frequently involved in the binding of metal ions, especially magnesium ions.[17] The geometry of the bulge region typically induces a helical bend and a widening of the major or minor groove of the helix to allow for the possible binding of proteins, ligands, and ions. RNA structures determined with X-ray crystallography and NMR show single-base bulges functioning in both a stacked-in and looped out conformation. For example, in two crystal structures of 5SrRNA, the cytosine bulge is found in two different looped-out conformations, one oriented up in the major groove and the other oriented down in the minor groove.[18] Since the nearby helix motif is not disrupted, this paper concludes that the two different bulge orientations represent the bulge as a flexible hinge and most likely a protein recognition mechanism. In another structure, two looped-out adenine bulges found in the X-ray crystal structure of the initiation site of genomic HIV-1 RNA form a base grip structure available for intermolecular interactions, possibly as a recognition signal.[19] Interestingly, when the structure is solved in a solution and without the presence of magnesium ions the bulge goes from a looped-out conformation to a stacked conformation.[20] The solution structure of another molecule, the SL1 VBS RNA, suggests that an adenine bulge spends its time in both conformations and that the position of the bulge base relative to the rest of the helical structure is important to Cap-Pol protein binding.[21]

There have been many studies aimed at understanding the significance of mismatched and bulged nucleotides in nucleic acid structures. Experimental studies have used techniques such as fluorescence resonance energy transfer (FRET) and gel electrophoresis to measure the helical twist and kink of RNA and DNA helices by varying the number of bulged nucleotides.[22,23] A study using temperature gradient gel electrophoresis (TGGE) characterized the stability of single-base bulges due to identical and non-identical nearest neighbor context.[24] Computational studies have used search algorithms to find sterically possible structures and molecular dynamics simulations to examine the dynamic behavior of RNA structures. Recently, a hierarchical method to search for energetically favorable conformations of the single-base bulges in A-form DNA and RNA was employed.[25] The energetics were evaluated for 340 distinct bulge conformations and three primary low-energy conformations were identified: bulge base stacked between flanking nucleotides or looped-in bulge (I), the bulge base in the minor groove (II), and a continuous stacking of the flanking helices with a looped-out bulge base (III). In a molecular dynamics study of DNA four different start conformations of a double-stranded DNA fragment with a single adenine bulge was simulated.[26] In another study, an RNA uridine bulge was examined in two different start conformations.[27] Computational studies for other motifs including the ribose zipper, the hairpin loop and non-canonical base-pairs have also been performed. Nearly one hundred RNA ribose zippers were grouped into eleven different classes based on sequence and structure conservation.[28] The conformations and sequence conservation of the hairpin loop motifs have been extensively studied[29–31] and a continuum solvent analysis has been performed on non-canonical base-pairs such as G:A mismatches.[31] Molecular dynamics simulations have been described for a number of structures containing RNA motifs including the A-minor motif[33] and the kissing-loop motif.[34,35] There have also been studies that examine structural motifs found in specific functional RNA's, such as ribosomal RNA's.[36–40] Also, the SCOR database was developed to provide a survey of three-dimensional motifs found in X-ray and NMR structures.[41] However, existing computational studies of RNA bulge motifs are limited and they mostly focus on particular bulge residue types (mostly adenines), use generated substructures instead of naturally occurring RNA structures, and don't consider the dynamic aspects of the bulge. Furthermore, molecular dynamics is rarely used to examine the behavior of nucleic acid motifs and to our knowledge no studies that examine multiple RNA fragments with different types of single-base bulges exist.

In this paper, we analyze bulges in X-ray-crystal and NMR structures and use molecular dynamics simulations to capture the essential structural information necessary to understand the influence of the bulge and its surrounding helical structure. We establish degrees of similarity in the bulge structures according to bulge type and surrounding base-pairs. The inspection of individual X-ray-crystal or NMR structures (i.e., a static analysis of geometric characteristics), allows an examination of the bulge in an equilibrium context with the inclusion of external factors such as proteins and other nucleic acid interactions. A dynamical analysis allows us to classify the bulge motifs according to their temporal conformational behavior. Twenty single-base bulge structures are grouped into three different classes based upon the bulge behavior found during these analyses. The curvature for each is compared within and across the classification scheme to quantify the flexibility and other conservative aspects of the conformations and provide a model for the design of novel nanostructures. Our ultimate goal is to perform a similar analysis on other RNA motifs, such as hairpin loops, base mismatches, and multiple base bulge loops, and to create a toolbox of nanostructures where one could pick the pieces to design a structure that has the needed shape and desired behavior.

2. METHODS

2.1. Structure Selection

In this study we perform a static and dynamic analysis on twenty different RNA structures containing the single-base bulge motif. The RNA structures containing the bulge motif are obtained from the Protein Data Bank (PDB).[42] No two structures selected are from the same source (23SrRNA, Ribozyme, Viral RNA, etc.) and species (C. Elegan, E. Coli, Homo sapiens, etc.); therefore the dataset used in the analysis is comprised of a non-redundant and diverse set of structures. To focus on the bulge motif in its simplest form and minimize the effects of surrounding molecule interactions, only single-based bulges with three Watson-Crick or G:U wobble base-pairs above and below were selected as structures in this study. The bulge portion of this RNA motif was extracted to form the standard motif used in this analysis. By looking at the extracted segment only, we neglect the overall conformational variations found in the twenty different RNA structures, and instead consider a smaller number of residues pertinent to the bulge motif. A description of each bulge motif in its original context is found in Table I. If multiple NMR structures are given, then the all-atom RMSD between all pairs of the models given in the PDB file is computed. The structure with the lowest average RMSD is used as a representative structure.

2.2. Analysis of the X-ray-Crystal and NMR Structures

In this analysis, the bulge motif described above is extracted from its corresponding NMR or X-ray-crystal structure. Since each single-base bulge adjoins three standard base-pairs, it is assumed that during minimization the global features and characteristics of the bulge will not change significantly when removed from the original context of the entire X-ray crystal or NMR structure. To confirm this, the six backbone torsion angles (α, β, γ, δ, ε, and ζ) and the torsion angle of the bond between the ribose ring and the base (χ) are calculated. Each segment is then minimized and the angles are calculated again to check for congruence between the minimized structure and the X-ray or NMR structure. Motif selection and validation are important to ensure that the extracted bulge structures maintain their basic conformation. To examine the effect of the bulge on its surrounding context in the X-ray and NMR structures, the two pseudotorsion angles (η and θ) are calculated using *AMIGOS*.[43] The twenty bulges are compared by residue type, position, and effect on the nearby residues. In addition, the global base-pair

Table I. Structure descriptions.

No.	PDB ID	Bulge	Secondary Structure	Method	Interactions
1	17RA	6	19 14 3'-GUAUCC 5'-CGUAGG 3 A 9	NMR	None
2	1LMV	6	8 3 3'-UGAUGU-5' 5'-ACUACA-3' 2 A 8	NMR	None
4	1FJG	31	555 550 3'-CCAUUG 5'-GGUAAC 28 G 34	X-ray	None
3	1GID	130	195 190 3'-GGCAGU 5'-CCGUCA 127 U 133	X-ray	None
5	1JJ2	2437	2457 2452 3'-UAACGG 5'-AUUGCC 2434 A 2440	X-ray	Water
6	1JJ2	943	1026 1021 3'-CCGCAG 5'-GGUGUC 940 A 946	X-ray	Protein
7	1J5A	2581	2584 A 2578 3'-CUGCAA 5'-GACGUU 2568 2573	X-ray	Nucleic Acid
8	1JJ2	2637	2640 A 2634 3'-CUGCAG 5'-GACGUC 2624 2629	X-ray	Water
9	1J5A	601	681 676 3'-AUCGGG 5'-UAGCCU 598 A 604	X-ray	None
10	1JJ2	2896	2899 A 2893 3'-GGGGCU 5'-CCCCGA 2858 2863	X-ray	Water, Nucleic Acid
11	1JJ2	1137	1228 1223 3'-CCGCGG 5'-GGUGUC 1134 G 1140	X-ray	Water, Nucleic Acid
12	1J5A	2854	2857 G 2851 3'-CUCUGG 5'-GAGGCC 2811 2816	X-ray	None
13	1F7F	6	25 20 3'-UUCGGC 5'-AAGCCG 3 U 9	NMR	None
14	1P5M	6	53 48 3'-GACCUC 5'-CUGGAG 3 U 9	NMR	None
15	1S9S	319	348 343 3'-GGCCCA 5'-CCGGGU 316 U 322	NMR	None
16	1AQO	7	26 21 3'-CGCAGG 5'-GUGUUC 4 C 10	NMR	None
17	1NBR	7	26 21 3'-CGCAGG 5'-GUGUUC 4 C 10	NMR	None
18	1BVJ	6	21 16 3'-GCUCUA 5'-CGAGGU 3 C 9	NMR	None
19	1Z31	262	297 292 3'-CCACCG 5'-GGUGGC 259 C 265	NMR	None
20	1DK1	47	50 C 43 3'-UGCCAC 5'-ACGGUG 19 24	X-ray	Water

parameters and overall curvature of the helices were calculated using CURVES 5.3.[44]

2.3. Structure Preparations for Molecular Dynamics Simulations

In this analysis twenty RNA molecule segments are monitored for bulge-induced changes in the helix. Initially, the original (thirteen residue) bulge segment was used for molecular dynamics simulations. However, due to the small length of the segment some structures began experiencing non-standard behavior, base-pair separation at the helical ends, and in rare cases eventual rearrangement of the helix. Therefore, to ensure stability of the helix and reasonable behavior of the bulge region during the molecular dynamics simulations additional residues were appended to each bulge segment. A UUCG tetraloop and a C:G base-pair were added to cap one end of the bulge segment and two G:C base-pairs were added to the opposite end of the bulge segment (Fig. 1). Note that the 5′ and 3′ sides of the bulge segments are consistent with their respective original X-ray or NMR structure. The residues in the resulting structure are numbered from 1 to 23 for consistency and comparisons with the other structures. Both the stem extension and hairpin loop were selected for their low energy and stability. The UUCG tetraloop is one of the most common and well studied tetraloops found in RNA.[14,45] The three-dimensional structure of the UUCG tetraloop and C:G base-pair is taken from the NMR structure of the P1 Helix from the Group I Self-splicing Introns (PDB 1HLX). This structure was determined at extremely high precision with an all-atom RMSD of 1.22 Å and local tetraloop RMSD of 0.6 Å with respect to the twenty structural models given by NMR. Again, the best representative structure was used for our model as described above. After attaching the stem and hairpin loop to the bulge segment, the structure is cleaned up to remove steric clashes.

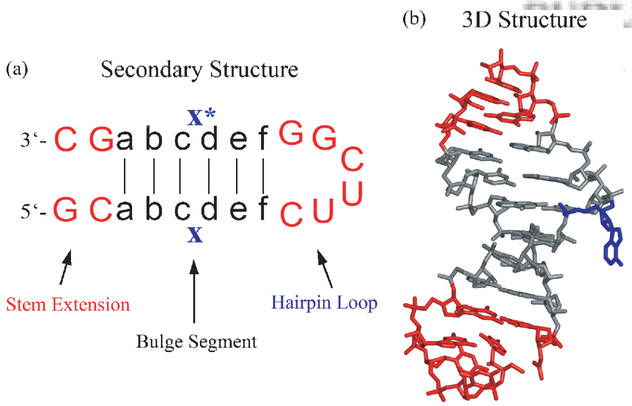

Fig. 1. RNA Model Structure of the stem extension, bulge segment, and hairpin loop. (a) Secondary structure of the RNA model structure. The upper case letters indicate actual residues, the lower case letters a–f indicate base-pairs surrounding the bulge, and the lower case letter *x* or *x** indicates the bulge residue. (b) Three-dimensional structure of the RNA model structure. The added segments including the stem and hairpin are indicated by red, the base-pairs surrounding the bulge by gray, and the bulge by blue.

2.4. Analysis of the Molecular Dynamics Simulations

All simulations in this study were performed using the molecular dynamics software AMBER 7.0[46] and the ff99 Cornell force field for RNA.[47] The simulations used the Generalized Born (GB) implicit solvent model which is implemented in the Sander module of AMBER. GB implicit solvent methods were chosen due to the large number of structures used in this study. Moreover, the GB method has been demonstrated to be an accurate and reliable method for various biomolecules.[48,49] Each structure was minimized, followed by slow 20 kcal/mol constrained heating up to 300 K, and four consecutive MD equilibrations with declining constraints from 2 kcal/mol to 0.1 kcal/mol over a time period of 250 ps. The heating and equilibration simulations were omitted from the analysis. The simulations used a salt concentration equal to 0.5 mol/L and the temperature was maintained at 300 K with a Berendsen thermostat.[50] The production simulations were performed for 4 ns using a 1 fs time step. The simulations were computed on a SGI-Altix computer using four processors. For each structure a thorough analysis of the dynamic simulations was performed, including RMSD comparisons, energy calculations, and curvature parameters. The RMSD for all molecular dynamics simulations is an all-atom RMSD calculated between the first structure and structures generated at 10 ps intervals during the simulation. Amber's Carnal and Ptraj modules were used for analysis of the RMSD and molecular energies.

The CURVES 5.3 was used for the curvature analysis of the RNA structures. The CURVES algorithm finds a helical axis that best fits the structure's conformation and provides a description of both the global and local geometric parameters. Every 10 ps a snapshot of the MD trajectory structure is taken and analyzed by CURVES. The global inter-base pair parameters, global curvature, and shortening were used in the comparison analysis of the structures. The global curvature (UU) was calculated as the angle between the local helical axes of the 2nd and n-1 base-pairs of the bulge segment (gray segment b–e, in Fig. 1). The percentage of helix shortening is derived from one minus the ratio of the distance along the line between the first and last base-pair (end-to-end distance) and the length of the global helical axis.

3. RESULTS

3.1. Static X-ray Crystal and NMR Structure Conformations

An analysis of each X-ray crystal and NMR structure was performed to examine the similarity of the bulge motif with respect to each other and to examine the extent to which

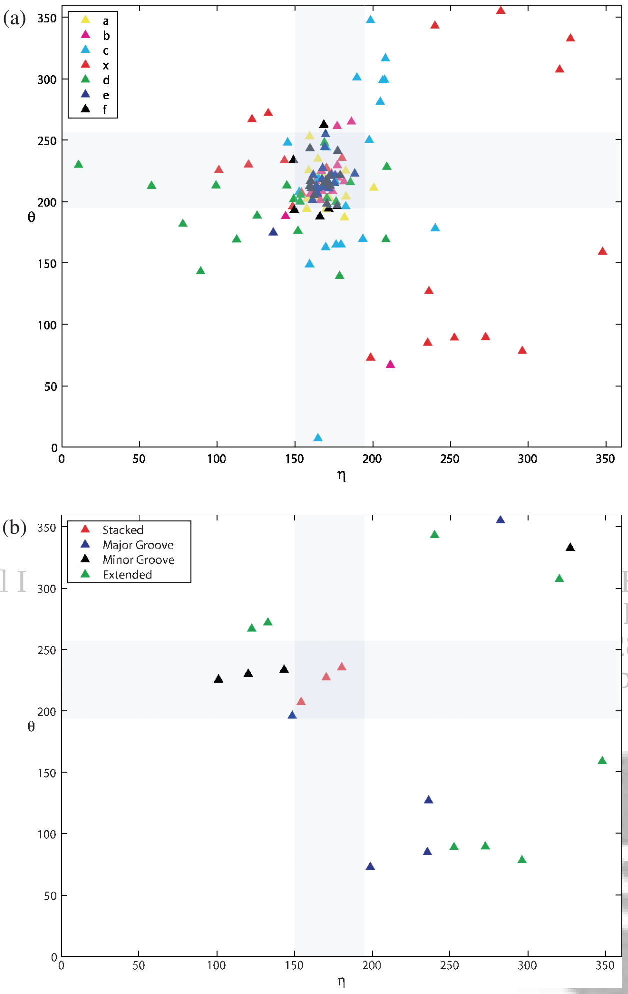

Fig. 2. Amigos angles for the twenty X-ray and NMR structures.[43] (a) The amigos angles for the bulge and the three flanking residues above and below the bulge. (b) The amigos angles by bulge position.

the surrounding context affects the bulge motif. We examine the impact of the structure's context on the bulge residue and the resulting helical properties. The pseudotorsion angles are shown in Figure 2. Only the bulge residue and the residues directly above and below the bulge deviate from a normal helical structure (indicated by the region defined by the crossing of the two gray bars in Fig. 2a). Additionally, it is evident that bulges that are looped-in or stacked maintain a standard helical form along the backbone while the looped-out conformations do not (Fig. 2b). Pseudotorsion angles of the bulges that loop towards the minor groove occur around a common region (125°, 230°), while pseudotorsion angles of the bulge that loop towards the major groove or loop out away from the helix vary. It should also be noted that the bulge type (A, C, G, or U) does not influence the value of the pseudotorsion angles.

The global curvature does not show a well defined pattern specific to the bulge position and orientation. This is possibly due to the differences in the surrounding contexts of the structures, including proteins and nucleic acid interactions.

3.2. Dynamic Conformations During the Molecular Dynamics Simulations

Now we examine the dynamical changes in the structure when it is free from environmental constraints. The molecular dynamics study shows that the dynamical behavior of the structures results in three different structural classes that depend not only on the bulge type but also on the surrounding base-pairs (Table II). To obtain a quantitative measure of the structural fluctuations of the base-pairs in the bulge region, the hydrogen bond distances are examined for the time course of the trajectory (Table III). The hydrogen bonding table lists Watson-Crick hydrogen bond occupancies attained for the bulge and the two flanking base-pairs on each side of the bulge. Class I contains bulges that compete for the standard hydrogen bonds of a neighboring base-pair and have at least one A:U base-pair on one side. This class contains adenine, guanine, and uracil bulges, with the bulge residue residing in a stacked conformation for at least 18% of the time (Table III). The stacked conformation of the bulge disrupts the standard hydrogen bonding of the nearby bases in all cases. In Table III, the base-pair that the stacked bulge disrupts is listed in the bulge column according to the notation in Figure 1. Note that for all class I bulges, the bulge residues are on the 5′ side and thus the bulge residues hydrogen bond to bases on the 3′ side of the structure. Class II is different from class I in that there is no competition between the bulge residue and the surrounding base-pairs. This is shown by the 0% Watson-Crick hydrogen bonding in the bulge column of Table III and the high percentages shown for the neighboring base-pairs. In addition, the bulge residues that are surrounded by G:C or wobble G:U base-pairs on both sides, have a predominately looped-out confirmation, and contain adenine, guanine, and uracil bulges. Depending on the structure, the class II bulge will associate with one of the helix grooves to form hydrogen bonds with the backbone or form base triples, protrude out away from the helix with no hydrogen bonding, or stack into

Table II. A description of the bulge classification scheme.

Class	Description	PDB Structures
I	A, G, U bulges with an A:U base-pair flanking the bulge. All stack into the helix during the simulation and compete with one of the base-pairs.	17RA, 1LMV, 1JJ2_2437, 1GID, 1FJG_31
II	A, G, U bulges with a G:C or G:U base-pairs flanking on the bulge. All predominately loop-out during the simulation and do not compete with flanking base-pairs.	1J5A_2854, 1JJ2_943, 1JJ2_2896, 1J5A_601, 1J5A_2581, 1JJ2_1137, 1JJ2_2637, 1F7F, 1S9S_319, 1P5M
II	C bulges regardless of surrounding sequence. All loop-out during the simulation.	1DK1, 1AQO, 1NBR, 1BVJ, 1Z31

Table III. Watson-Crick hydrogen bond occupancy table.

Structure	b–b	c–c	Bulge	d–d	e–e
Classification I					
17RA	G-U	U-A	A(x):U(d)	A-U	G-C
	79.33	86.65	88.78	7.12	98.88
1LMV	C-G	U-A	A(x):U(d)	A-U	C-G
	97.20	92.29	84.38	7.58	87.40
1FJG_31	G-C	U-A	G(x):A(c)	A-U	A-U
			G(x):U(d)		
	94.40	51.82	10.63/9.28	65.06	79.39
1GID	C-G	G-C	U(x):A(d)	U-A	C-G
	96.23	95.24	90.97	2.90	3.04
1JJ2_2437	U-A	U-A	A(x):A(c)	G-C	C-G
	51.73	66.27	18.78	96.34	97.11
Classification II					
1JJ2_943	G-C	U-G	A	G-C	U-A
	95.63	82.43	0.00	89.85	89.20
1J5A_2581	A-U	C-G	A	G-C	U-A
	89.50	88.52	0.00	80.19	81.91
1JJ2_2637	A-U	C-G	A	G-C	U-A
	98.03	97.44	0.00	85.92	76.63
1J5A_601	A-U	G-C	A	C-G	C-G
	83.04	97.54	0.00	97.74	94.65
1JJ2_2896	C-G	C-G	A	C-G	G-C
	97.80	86.60	0.00	96.90	90.51
1JJ2_1137	G-C	U-G	G	G-C	U-G
	97.12	90.23	0.00	97.86	79.13
1J5A_2854	A-U	G-C	G	G-U	C-G
	86.77	96.46	0.00	84.57	94.59
1F7F	A-U	G-C	U	C-G	C-G
	51.48	93.83	0.00	79.52	94.52
1P5M	U-A	G-C	U	G-C	A-U
	38.99	94.52	0.00	96.76	82.60
1S9S	G-C	G-C	U	G-C	C-G
	89.06	89.55	0.00	90.40	96.39
Classification III					
1AQO	U-G	G-C	C	U-A	U-G
	57.48	95.93	0.00	59.38	65.67
1NBR	U-G	G-C	C	U-A	U-G
	81.49	94.32	0.00	41.57	32.32
1BVJ	G-C	A-U	C	G-C	G-U
	95.63	94.42	0.00	95.61	74.04
1Z31	G-C	U-A	C	G-C	G-C
	91.18	80.08	0.00	97.28	96.13
1DK1	C-G	G-C	C	G-C	U-A
	92.31	97.29	0.00	89.82	82.01

This table represents the Watson-Crick hydrogen bond occupancies of the bulge segment for the entire 4 ns run. For class I, the bulge residue is indicated by an *x* and the residue that the bulge forms Watons-Crick bonds to is indicated by a (c) or (d) with the notation shown in Figure 1. For all class I bulges, the bulge is on the 5′ side (indicated by *x*) and thus the residue that the bulge hydrogen bonds to is on the 3′ side.

the helix without disrupting the existing base-pairs. Only adenine bulges attempt to stack into the helix. Class III contains only cytosine bulges that prefer the looped-out conformation regardless of the nearby sequence. Therefore cytosine bulges never compete against existing base-pairs for standard hydrogen bonds (Table III) and like most of class II, form hydrogen bonds with the backbone, form base triples, or protrude away from the helix with no hydrogen bonding. Examples of the different conformations found in the three classes are shown in Figure 3.

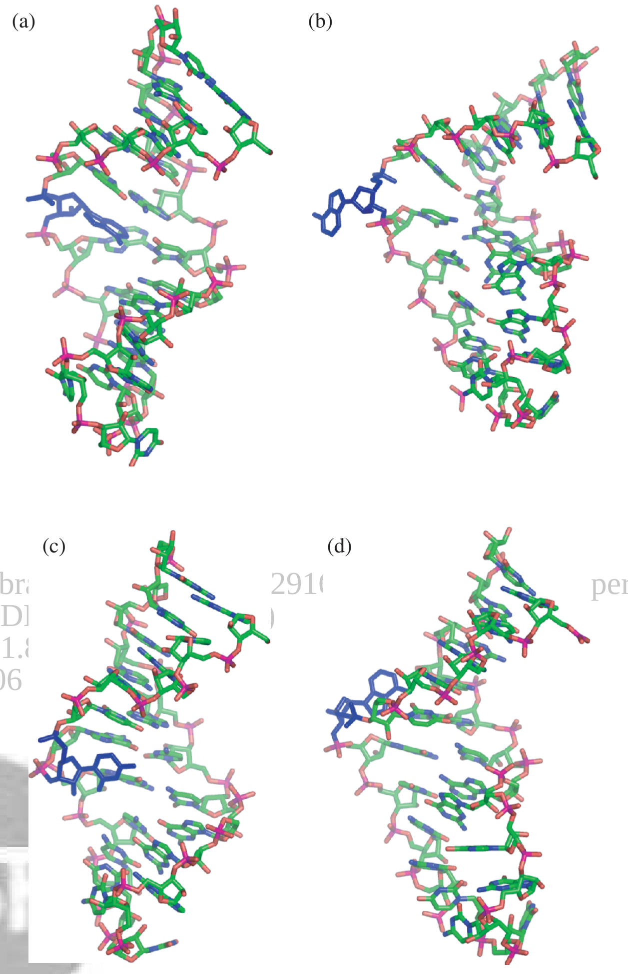

Fig. 3. Common conformations of the bulge structures with the bulge residue shown in blue. (a) The bulge stacks into the helix and disrupts one of the flanking base-pairs. These bulges are representative of class I. (b) The bulge is looped-out away from the helix. (c) The bulge is looped-out towards the major groove. (d) The bulge is looped-out towards the minor groove. Bulges found in (b–d) are representative of classes II and III. See Table IV for the conformations found in each class and each bulge structure.

3.2.1. *Class I—Competing Bulges*

Class I contains A, G, and U bulges that compete for the Watson-Crick hydrogen bonds of a neighboring base-pair. The bulge residue exists predominately in a stacked conformation with at least one flanking A:U base-pair. There are five structures in class I, three structures with A:U base-pairs on both sides of the bulge and two structures with one A:U base-pair and one G:C base-pair surrounding the bulge.

Structures 17RA and 1LMV have the same base-pair context directly surrounding the bulge and behave almost identically during the molecular dynamics run. The bulge in structure 17RA is initially looped-in without forming hydrogen bonds with surrounding bases. After approximately 50 ps the A6 bulge breaks A7:U18 and forms a standard Watson-crick A6:U18 base-pair, thereby creating an A7 bulge that forms a Hoogsteen/sugar base triple with U18. The A6 bulge has a hydrogen bond occupancy of 88.78% throughout the simulation. During the short transitional phase the structure's overall RMSD quickly changes by 3 Å and then the structure fluctuates around this conformation by ±1 Å. The total energy decreases by 15 kcal/mol due to the bulge switching. The global curvature shows a slight drop from 64 degrees to 50 degrees with consistent fluctuations. The change in conformation is easily seen in the shortening value that changes from 40% to 50% during the transition period and then remains relatively constant at 50% shortening. The mean and standard deviations for the structures are given in Table IV. Structure 1LMV has a dynamic behavior pattern that is similar to 17RA with the A6 bulge competing with residue A7 for a standard Watson-Crick hydrogen bonding to U18. The only difference between these two structures is the transition of A7 from a stacked conformation with intermittent U18 and U17 Watson-Crick hydrogen bonds to a U18 Hoogsteen/sugar base-triple just before 1100 ps. During the simulation the hydrogen bond occupancy for the A6 bulge is 84.38%. The structures overall RMSD fluctuates between ±1.5 Å until the bulge switch at which point the RMSD fluctuations decrease to ±0.75 Å and the energy begins to

Table IV. Mean and standard deviations for the global curvature and helix shortening.

Structure	Type (ACGU:SMNE)	Global curvature mean	Global curvature stdev	% Shortening mean	% Shortening stdev
Class I					
17RA	(A:S)	52.8	13.4	50.1	3.5
1LMV	(A:S)	52.4	20.0	34.8	8.8
1FJG_31	(G:E, S)	45.4	16.4	36.3	8.3
1GID	(U:S)	79.1	21.0	28.2	5.3
1JJ2_2437	(A:N, S, M)	46.9	14.9	28.9	7.5
Class II					
1JJ2_943	(A:N, S)	50.6	16.9	32.0	5.6
1J5A_2581	(A*:N, S)	37.3	16.4	23.4	6.0
1JJ2_2637	(A*:M)	47.0	15.3	24.6	4.7
1J5A_601	(A:E)	58.9	14.4	32.0	4.9
1JJ2_2896	(A*:N)	34.6	13.8	23.4	4.3
1JJ2_1137	(G:M)	36.7	14.0	34.6	3.8
1J5A_2854	(G*:M)	41.5	13.1	26.1	5.1
1F7F	(U:N)	50.6	24.1	19.5	9.2
1P5M	(U:N, E, M)	49.1	14.7	31.1	5.6
1S9S	(U:N)	29.2	16.1	30.6	6.6
Class III					
1AQO	(C:N)	43.8	17.0	33.4	5.5
1NBR	(C:M)	43.1	21.4	37.7	8.4
1BVJ	(C:E)	54.7	13.6	36.7	4.6
1Z31	(C:M)	43.3	15.3	31.7	4.8
1DK1	(C:M)	40.6	13.9	20.9	4.1

Each structure has a structure type notation consisting of three parts. The first part is the bulge residue type (A, C, G, U) and the second part is the favored conformation(s) of the bulge where S indicates stacked into the helix, M indicates looped-out towards the major groove, N indicates looped-out towards the minor groove, and E indicates looped-out, extended away from the helix. The third notation is an asterisk and indicates a bulge is on the 3′ side of the helix. No asterisk indicates a bulge on the 5′ side of the helix.

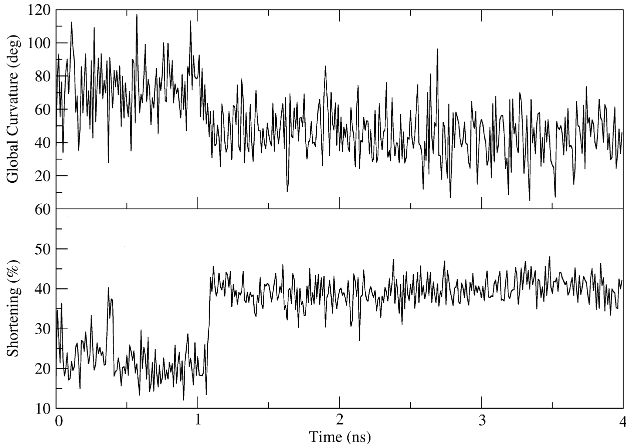

Fig. 4. Global curvature and shortening for structure 1LMV.

decrease. Again the A7 transition is clearly indicated by the change in helix shortening which goes from 20% to 40% during this period and can be seen in Figure 4. After this period, the global curvature stabilizes at 46 degrees with consistent fluctuations similar to that of 17RA. The relatively small fluctuations in global curvature and the helix shortening for these A bulges can also be seen in Figure 4. Hence this sequence has a predictable competing dynamic bulge characterization.

Structure 1FJG31 has an initially looped-out bulge with A:U base-pairs surrounding the bulge, but because it is a G bulge, the possible competing base-pair interactions are the G:U or G:A. The energetics for these competing base-pairs are not as favorable as the A:U base-pairs. Like 1LMV the bulge base switch to the stacked conformation occurs after approximately 1 ns, with RMSD fluctuations of 2 Å until the switch. After the formation of the G6:A19 hydrogen bonds, the RMSD fluctuates less than 1 Å and the energy drops by 5 kcal/mol until the G6 bulge is slightly pushed out and the standard base-pairing scheme resumes intermittently. At 3400 ps the bulge again resumes Watson-Crick hydrogen bonding, but with U18 to form a G:U wobble base-pair. This base-pairing scheme continues intermittently between the bulge and the standard base-pair until the end of the simulation. While the total hydrogen bond occupancy is lower for this bulge, it forms many base triples and temporary hydrogen bonds to A19 and U18 to reduce their hydrogen bonding occupancy to 51.82% and 65.06% respectively. Thus, the initially looped-out bulge does stack into the helix and replaces a standard Watson-crick base-pair like the other stacked bulge structures, further proving that it's the sequence that matters for bulge competition not whether the bulge is initially stacked or looped-out. The hydrogen bond occupancy for this G bulge is only 19.91% and thus does not consistently maintain the stacked equilibrium conformation throughout the simulation like the A bulges.

The remaining class I structures have one G:C base-pair next to the bulge, but also have competing bulges, albeit in

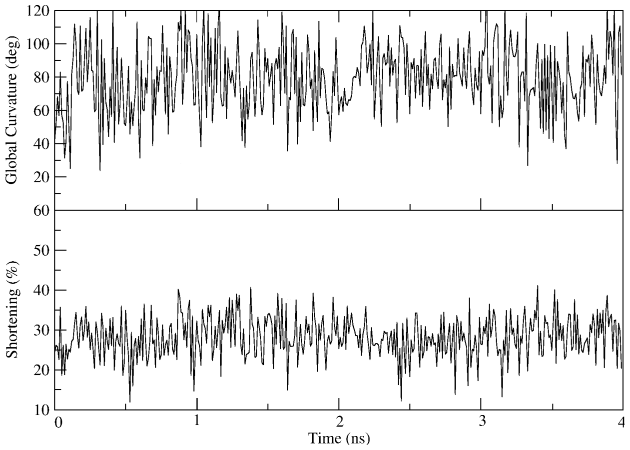

Fig. 5. Global curvature and shortening for structure 1GID.

a less pronounced manner. In structure 1GID, the initially looped-out bulge residue U6 quickly loops into the helix and begins base-pairing with A18. This causes a shift in base-pairing for the two standard base-pairs below the bulge such that U7:G17 and C8:U16 form hydrogen bonds on their Watson-Crick edges and A9 is forced out into the minor groove of the helix. While the A9 residue no longer forms a standard hydrogen bond with U16, it does form a base triple on the Watson-Crick/sugar edge (O2 and H62). The shift in base-pair hydrogen bonds for U7 and U8 can be seen by the very low hydrogen bond occupancies of 2.90% and 3.04% in Table III. The Watson-Crick hydrogen bond occupancy for the A6 bulge is 90.97%. The RMSD of this structure after the switch varies only slightly more than the previous structures, but the fluctuations in global curvature and shortening vary considerably as seen in Figure 5. It is also important to note that while this is a U bulge, it competes for a standard U:A hydrogen bond similar to the stable structures 17RA and 1LMV.

The final structure in this class, 1JJ2_2437, is the least stable and the most flexible structure in class I due to the possibility of bulge-base interactions via non-canonical A:C or A:A interactions. These somewhat unfavorable bulge interactions result in a continual temporary competing bulge A:A. The bulge moves from an initially stacked conformation, to looped-out (minor groove), to a stacked conformation, to looped-out (major groove), to a stacked conformation, to looped-out (minor groove) and has a Watson-Crick hydrogen bond occupancy of only 18.78%. While this bulge can move freely through the helix, the competing base-pairs found in the sequence combination do not appear strong enough to completely sever the standard U5:A19 bonding to become a stable replacement base-pair. The RMSD, energy, shortening, and global curvature values show small changes in the bulges orientation, which are not much different than that of standard helix fluctuations. The small changes that leave the helical structure relatively undisturbed are due to the A6 bulge participating in a base triple interaction with A19 via Hoogsteen/sugar hydrogen bonds when positioned in the minor and major grooves of the helix.

In summary, the mean global curvature values for class I are very similar. After the bulge competition occurs, the purine bulges all have a mean global curvature of approximately 50 degrees and a standard deviation of approximately 16 degrees. The one pyrimidine structure, 1GID, has a global curvature mean of 79 degrees with a standard deviation of 21. The mean shortening values can vary considerably from structure to structure, with the means ranging from 50.1% to 28.2%.

3.2.2. Class II—G:C and G:U Surrounded Bulges

In class II, there is no competition between the bulge residue and the surrounding base-pairs. All ten structures have A, G, or U bulges that are surrounded by G:C or G:U base-pairs on both sides and primarily stay in a looped-out confirmation. These bulges rarely participate in standard hydrogen bonding with the surrounding base-pairs (Table III). For most of the structures in this class it is difficult for the bulge to interrupt the standard hydrogen binding for any significant period of time because the base-pairs surrounding the bulge are GC or GU wobble base-pairs. Therefore the bulge base is almost always shifting positions to find potential hydrogen bonds and frequently changes the helix's conformation. The bulge typically forms hydrogen bonds to the backbone or forms sugar/Hoogsteen base triples with residues in the major or minor grooves of the helix. In some cases, however, the bulge protrudes away from the helix without any hydrogen bonding or stacks into the helix without disruption (A bulges only). The bulges that are oriented towards the major or minor groove of the helix in the original X-ray or NMR structure typically remain oriented that way. There are several factors that keep the bulge residue near its starting state during the molecular dynamics simulation: hydrogen bonding to nearby residues, the inability for the bulge residue to break the G:C or G:U base-pairs or rotate from one helix groove to the other, and the unfavorable alternative in becoming an unbound bulge extending away from the helix. However, this is not always the case; some structures start oriented towards a groove and change conformations frequently. In almost half of the structures in class II, the bulge bases deviate from the initial start configuration by going from looped-out to a position across the major or minor groove and visa versa. The orientation towards one of the helix grooves is preferred due to hydrogen bond formation and a more stable, energetically favorable state. There are two cases in this class where the bulge remains unbound, protruding away from the helix. This could be due to a unique sequence composition that consists of any combination of CAG's for the bulge and flanking bases. This sequence combination exists for only these two structures. The use of this sequence with the bulge could potentially be used as a bulge signaling mechanism for other nucleic acids or proteins. However, we do not have enough data to confirm this.

Adenine Bulges. Adenine bulges have been frequently studied due to the discrepancies between the looped-out and stacked conformations given by NMR and X-ray structures (see discussion section). For the class II structures, we indeed find that this bulge is the only one where both the stacked and looped-out conformations are possible. However, the simulations show that the most frequent conformation for adenine bulges is across the minor groove. We also note that adenine and uracil bulges are the only type in class II to orient towards the minor groove. The bulge in structure 1JJ2_943 crosses the minor groove in the initial X-ray structure to hydrogen bond with the side chain of the amino acid methionine, approximately 3.3 Å away. This bulge structure is in a complex environment with several nucleic acid chains, amino acids, and water molecules which presumably hold the bulge in place. During the molecular dynamics simulation, the bulge residue reaches diagonally across the minor groove to form two hydrogen bonds on the Hoogsteen edge of G19 and U5 (H21(G19)...N7(A6), O2(U5)...H62(A6)). At 1100 ps the bulge changes conformation and aligns itself with the base-pairs in the helix and forms two Hoogsteen/sugar hydrogen bonds with G19 (N7(A6)...H22(G19) and H62(A6)...N3(G19). Finally, at approximately 1860 ps the bulge stacks into the helix and remains there for the rest of the simulation. During the stacked conformation the bulge does not disrupt any base-pair nor does it participate in any significant hydrogen bonding. The two conformational transitions can be seen by the slight change in helix shortening (Fig. 6). The total RMSD fluctuates by ±1.5 Å about the mean throughout the entire trajectory. The fluctuations in global curvature and helix shortening shown in Figure 6 are representative of those in class II. A similar adenine bulge structure, 1J5A_2581, initially appears to interlock with another part of the RNA helix at U2785 (a residue distal in sequence but proximal in distance to the bulge) and is therefore looped-out protruding away from the helix. This is one of the bulges found on the 3′ side of the bulge segment and is indicated by an asterisks in the tables and figures. After 60 ps the A18 bulge stacks into the helix without disrupting the helix. However, unlike the previous structure, at 810 ps the bulge moves into the minor groove of the helix where it forms two hydrogen bonds with the sugar edge from the base diagonally below it (H22(G6)...N7(A18) and N3(G6)...H62(A18)). While the only significant change in RMSD of 1 Å occurs at 60 ps, the helix shortening decreases by 10% at 810 ps when the bulge residue relocates to the minor groove. In addition, the energy stabilizes when the bulge is in the minor groove in contrast to the constantly changing energy during the stacked conformation. Another adenine bulge structure, 1JJ2_2637, comes close to lining up with the helix for a stacked conformation, but does not. Initially this structure is in contact with a water molecule and is looped-out away from the helix with a slight angle towards the major groove. The structure moves into the major groove with a 1 Å change in RMSD during the first 500 ps and then remains in this conformation for the entire simulation with limited hydrogen bonding (less than 3% occupancy). The total energy of this structure fluctuates by ±5 kcal/mol around the mean, which can be associated with the limited attempts at hydrogen bonding. Structure 1J5A_601 is not in contact with anything in the initial X-ray structure, and does not form any hydrogen bonds. The structure is initially completely looped-out perpendicular to the backbone of the helix and stays looped-out throughout the entire simulation. The mean RMSD changes by less than 0.5 Å and the energy drops by approximately 7 kcal/mol as the molecule adjusts. The last adenine bulge structure, 1JJ2_2896, is initially looped-out perpendicular to the backbone and facing the minor groove where it base pairs with residue U2756 a few base-pairs away and in the opposite chain. During our simulation the A18 bulge is moving along the backbone and across the minor groove of the helix for approximately 700 ps until it finally forms hydrogen bonds with G7 (26% occupancy) on its sugar edge (N1(A18)...H22(G7) and H61(A18)...N3(G7)). As the bulge moves toward the minor groove the RMSD changes by 0.7 Å, the helix shortening decreases by 5%, and the energy decreases by 7 kcal/mol. This is the only adenine bulge in this class that is in the minor groove and does not line up with the base-pairs for sugar edge hydrogen bonding or stacking into the helix.

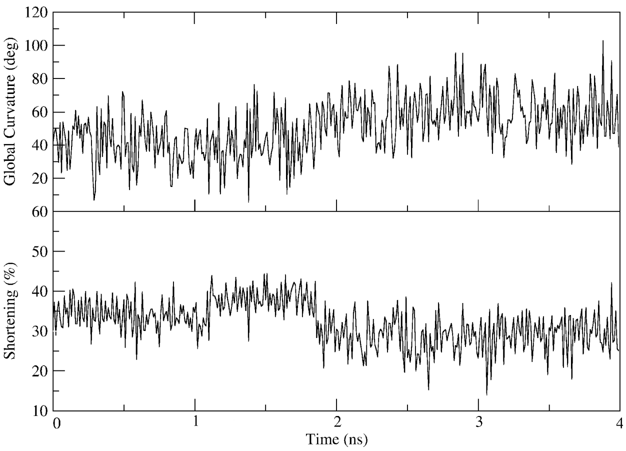

Fig. 6. Global curvature and shortening for structure 1JJ2_943.

Guanine Bulges. There are only two guanine bulges in this class and both are oriented towards the major groove of the helix to form unusual hydrogen bonds. The guanine bulge in structure 1JJ2_1137 is similar to the adenine bulge in structure 1JJ2_2637. It has the G6 bulge initially looped-out, however, the bulge quickly moves into the major groove of the helix and perpendicular to the base-pairs, but does not form hydrogen bonds. Within a few picoseconds the bulge changes slightly in orientation and remains in this conformation for the entire simulation with limited non-standard hydrogen bonding to N7(G7), O6(G7), H5(C9, U8, G10), and H42(C9, G10) with less than 10% hydrogen bonding occupancy. Here, the changes in bulge orientation during the early stages of the simulation could be attributed to the contacts between the bulge residue and distal residues in the original X-ray structure that are not included in the extracted bulge structure. The energy decreases by only 5 kcal/mol and the RMSD fluctuates ±3.5 Å about the mean during what appears to be a low energy 250 ps transition phase from one non-bonded form to another. The other G bulge found in structure 1J5A_2854 also orients itself in the major groove and initially weakly bonds to other residues. During the molecular dynamics simulation inconsistent hydrogen bonding occurs between the G18 bulge and the nonstandard, unoccupied hydrogen bonds on the Hoogsteen or C–H edge of C2, A4, and G5. Due to the strong G:C hydrogen bonds flanking the bulge, the bulge base is forced to reach over these base-pairs for alternative bonding. The bulge base encounters these somewhat distal bases at an angle that prevents long term bonding. After 1170 ps, the stem region stabilizes as the bulge residue tries to find other atoms for hydrogen bonding and the bulge residue moves to the helix backbone. While this transition is clearly seen by a 1.5 Å change in RMSD and a change in helix shortening by 7%, there is a minimal change in total energy. Interestingly, in its original context the bulge base stacks in-between two base-pairs three residues below the bulge on the opposite side to create a kinked shape on the exterior of the structure. In summary, the large purine bases have difficulty finding stable long term bonding with the helix in the looped-out conformation.

Uracil Bulges. Like the adenine bulges, uracil bulges prefer a minor groove orientation. All three of the uracils in this class prefer a looped-out conformation, angled toward the minor groove. In structure 1F7F, the U6 bulge reaches across the minor groove on the 5′ side to form intermittent hydrogen bonds with U20. The bulge forms a hydrogen bond on the sugar edge and the Watson-Crick edge (H3(U6)…O2(U20) and O4(U6)…H3(U20)). As shown in Table III, the Watson-Crick bond does cause some disruption to the A4:U20 base-pair. The structure shows large fluctuations with the RMSD deviating by as much as 4 Å. The energy initially increases slightly by 5 kcal/mol, but decreases as the bulge forms stable hydrogen bonds. The bulge in structure 1P5M initially bridges the minor groove like 1F7F. However, after approximately 200 ps the bulge loops out away from the helix where it stays until approximately 3275 ps without any hydrogen bonding. At 3275 ps, the bulge residue moves in along the major groove where it remains until the end of the simulation, again without any hydrogen bonding. As the bulge loops away from the helix and as the bulge loops back toward the helix the RMSD changes by 1 Å. During the period when the U6 bulge is looped away from the helix the total energy begins fluctuating by as much as 10 kcal/mol. The final uracil bulge, 1S9S is initially looped-out away from the helix, but quickly comes into the minor groove. Like the other two U bulge structures, it also bridges the minor groove to the 5′ side of the helix where it maintains temporary hydrogen bonds to the unoccupied H22 atom of the residues G21, G22, and C4.

3.2.3. *Class III—Cytosine Bulges*

Cytosine bulges prefer to remain in the looped-out conformation regardless of surrounding sequence. However, the mean global curvature does change as a result of surrounding sequence. Cytosine bulges also do not appear to favor a particular looped-out orientation with the 1AQO bulge pointing towards the minor groove, 1NBR, 1Z31, and 1DK1 bulges pointing towards the major groove, and 1BVJ in the fully extended looped-out conformation. These bulges have the highest structural flexibility of all three classes.

In structure 1AQO the bulged base is perpendicular to the base-pairs in the minor groove of the helix with a slight tilt towards the stem. During the molecular dynamics simulation the bulge participates in non-standard hydrogen bonds with bases G20 and C21. The strongest binding occurs between N3(C6) and H22(G20) with 27.9% occupancy. This formation of a temporary hydrogen bond with G20 occurs due to a somewhat weaker Watson-crick binding of the original U4:G20 base-pair, the strong affinity of GC bonding, and the absence of any other strong hydrogen bonding partner in the stem region of the helix. The hydrogen bonding of the A6 bulge with C21 is weaker (less than 19.8% occupancy) and is caused by the inability to achieve consistent bonding with G20. The low hydrogen bond occupancies are reflected by high RMSD values and fluctuating energy values. The global curvature and shortening values are shown in Figure 7. The somewhat high fluctuations in global curvature and helix shortening shown in Figure 7 are representative of

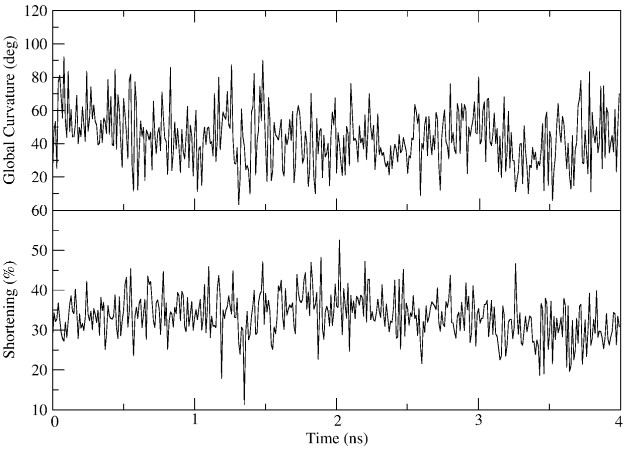

Fig. 7. Global curvature and shortening for structure 1AQO.

class III structures. Initially the bulged base of 1NBR protrudes out into the major groove and perpendicular to the sugars of the surrounding bases with no apparent hydrogen bonding. Within a few picoseconds the bulge forms hydrogen bonds with the helix backbone for approximately 950 ps. The bulge then goes back into the major groove of the helix to form standard Watson-crick hydrogen bonds with G17 (O2(C6)…H21(G17), N3(C6)…H1(G17), H41(C6)…O6(G17)) for the rest of the molecular dynamics simulation. The residue G17 shares these bonds with the bulge residue and its original base-pair U8 preventing the C6 bulge from remaining fully stacked into the helix. Stacking into the helix is also prohibitive for the bulge due to the U7:A18 base-pair in-between the bulge and its hydrogen bonding partner G17. Additional binding occurs between H42(C6) and O4(U4) due to the curvature of the helix. This conformation is just slightly more favorable energetically and deviates by about a 0.5 kcal/mol from the initial structure. The helix shortening decreases by an average of 15% and the global curvature decreases by an average of 15 degrees for this new conformation. The structure 1BVJ has an initial conformation that is very similar to that of 1NBR. It protrudes out perpendicular to the sugars of the surrounding bases with a slight tilt towards the stem without hydrogen bonding. However, within 30 ps the structure moves away from the helix and becomes a fully looped-out bulge. For the remaining simulation the structure fluctuates away from the helix without any hydrogen bonding. The completely looped-out conformation has a minimal effect on the overall RMSD, helix shortening, and global curvature, but decreases the energy by 20 kcal/mol. The bulge in structure 1Z31 is initially positioned in the major groove. However during the first 400 ps the bulge moves away from the helix in an extended conformation, then across the minor groove, and then back in the major groove where it remains throughout the rest of the simulation. The conformational changes during the first 400 ps decrease the energy by approximately 30 kcal/mol and are easily seen by large RMSD changes. Once the bulge stabilizes in the final major groove orientation the RMSD fluctuates by ± 0.75 Å. Overall the global curvature and shortening values show less fluctuation around the mean as compared to other structures in this class. The structure 1DK1 is the only structure in this class with two G:C base-pairs surrounding the bulge. The C18 bulge is initially looped-out, angled slightly towards the major groove of the helix and is unbound. During the first 200 ps the bulge fluctuates about the major groove and causes large deviations in the overall motion of the helix. Relatively quickly the O2 of the bulge forms alternating bonds to the unoccupied C—H edge of residues C19 and C17 flanking the bulge (H42(C19) and H5(C17)). Once in this position, the bulge continues this pattern with the majority of the fluctuations occurring only in the stem and hairpin loop region. The total energy drops by 25 kcal/mol with a minimal change in RMSD within the first 500 ps. For this particular structure, it is important to note that this type of hydrogen bonding can only occur due to the size and arrangement of the two C′s (pyrimidine) and one bulge C (pyrimidine) perpendicular to them.

Class III bulges with one G:C and one A:U flanking base-pair have very similar mean global curvature values in comparison to the other classes, with values between 43.1 and 54.7 degrees. They are also quite flexible given the high associated standard deviations averaging a little more than 16 degrees. The same is true for the shortening values which have mean values between 31.7 and 37.7 with standard deviations averaging 5.5 degrees.

3.2.4. Curvature Analysis Based on Bulge Position

In addition to the classification based on sequence, it is also insightful to examine the parameter statistics such as the global curvature mean and variance from the mean for the different bulge positions. Table IV shows the global curvature and shortening means and standard deviations for each class broken down into looped-out direction (stacked, minor groove, major groove, extended away from the helix). The values are based on the entire trajectory of 4 ns. When analyzing the mean global curvature, each grouping stands out. The bulges looped-out towards the minor groove have means that vary by approximately 21 degrees, while the bulges looped-out towards the major groove have means that vary by only 10 degrees. The least flexible structures are the bulges that are looped-out away from the helix, with means that vary by 4 degrees. Note that due to the limited set of looped-out structures, especially those that are looped-out away from the helix, it is hard to verify these patterns with any statistical certainty. However, the results presented directly represent conformational patterns that are clearly seen during the molecular dynamics run.

For example, the bulges that are looped-out away from the helix maintain a very constant overall helix curvature with the bulge residue having little impact on the overall helical structure, despite the wide range of motion the bulge residue exhibits. A more surprising feature was the large difference in mean global curvature values between the bulges looped-out towards the major groove as opposed to the bulges looped-out towards the minor groove. Again this is also clearly seen during the molecular dynamics run.

In addition to the looped-out type, we also examined the differences between the bulges located on the 5′ verses the 3′ side. Here, there is a definite difference in the shortening values with the 3′ bulges ranging from 23.4% to 26.1% and the 5′ bulges ranging from 19.5 to 50.1%. The global curvature value ranges are not as large with the 3′ bulges having values between 34.6 and 47.0 degrees and the 5′ bulges having values between 29.2 and 79.1 degrees. It is evident that the 5′ bulges have a slightly higher variance of 49.9 degrees compared to 12.4 degrees. When the global curvature and shortening values were examined by bulge type A, C, G, and U no clear pattern was recognized.

4. DISCUSSION

This is the first study of RNA single-base bulges that utilizes molecular dynamics to examine the behavior of bulges from real structures and includes a comprehensive sampling of all bulge types. Given our structure selection method, we have confidence that the structures studied computationally will have a similar sequence, fold, and dynamic behavior to the structures studied experimentally. We examine our results and compare our data with other studies of RNA and DNA bulges found in the literature. Often experimental studies show RNA and DNA bulges to be similar within the same environmental context. In both DNA and RNA, adenine bulges prefer a looped-out or stacked conformation depending on whether the structure is solved in crystal form or in a solution.[51, 52] However, there is some level of uncertainty and therefore not all of the characteristics and properties of DNA can be directly transferred to RNA. There have been several studies of DNA and RNA bulges that include (1) analyses of X-ray or NMR structures, (2) algorithms that search the conformational space for energetically favorable structures, and (3) molecular dynamics simulations of the structures in (1) and (2). First let us compare the analysis of several bulges obtained from X-ray and NMR derived structures. The question of why a bulge base stacks into the helix, loops out, or orients towards the minor groove has been a long contested debate. In the case of adenine bulges, there appears to be some discrepancies in NMR structures verses X-ray structures over whether the adenine base is looped-out or stacked into the helix. NMR solution structures and matrix refinement methods consistently find adenine bases stacked into the helix[53–56] while X-ray crystallographic structures always find the adenine base looped-out.[57–59] Thiviyanathan attributes this to the differences in solution and solid state conditions necessary for each method.[53] The stacked-in conformation causes the structure to adopt a bent geometry that is unfavorable for stacking and crystallization, while the looped-out conformation is typically stabilized by inter-molecular contacts and is energetically favorable. In the dilute solutions used by NMR, intermolecular contacts are not as relevant and the bulge can more easily stack inside the helix to obtain a low energy conformation with little helical disruption. While this reasoning may explain the discrepancies based on experimental methods, this does not explain what the preferred conformation is for a dynamic molecule found in nature. Through the use of molecular dynamics, we have shown that not only can both conformations exist for adenine, but bulge competition can occur depending on the surrounding sequence and that the energy difference between the two states is only around 7–10 kcal/mol for structures that have total energies of approximately 4200 kcal/mol. It is shown that for all other single-base bulge types this is not the case; cytosine prefers not to stack into the helix and the stacking of uracil and guanine is based on the neighboring sequence context. Hence, our findings for RNA are somewhat different than some of the existing experimental studies on DNA bulges; we can't classify the structures based on base type (purine or pyrimidine) alone.[26] We classify based on the base type and the surrounding context. It should be noted that in the previous studies, cytosine stacks into the helix only at elevated temperatures. In addition to bulge orientation, experimental studies provide some initial insight into the bending of the helix induced by bulge residues in equilibrium conditions, bound and unbound. Using gel electrophoresis and transient electric birefringence, it has been shown that bulges introduce pronounced kinks or bends into RNA and DNA helices, with the bulge type and number of bases in the bulge determining the magnitude of the kink.[60, 61]

Another method used to examine the characteristics of bulge residues in RNA or DNA employs a conformational search to find all possible energetically favorable conformations. The hierarchical search by Zacharias and Sklenar was performed on the bulge and the immediate neighboring nucleotides in a continuum (implicit) solvent model with the constraint that the motif fits into a continuous dsRNA.[25] While our structures were not of identical sequence context, all three low energy classes found by the hierarchical method were observed in our twenty structures. Additionally, our structures of similar sequence context favored one of the lowest five energy conformations given by the hierarchical method. The differences in overall sequential context and the difference in constraints prevent us from directly comparing energies for all structures and explain minor differences in results between our structures and those of the hierarchical method. However, our structures that have both the looped-out and stacked conformations during the simulations energetically prefer the stacked conformation, which is the lowest energy conformation found by the hierarchical method.

There are relatively few studies that use computational approaches to examine the dynamics of non-canonical motifs such as bulge motifs. One study simulates four different start conformations of a double-stranded DNA fragment with a single adenine bulge and produces similar results to our RNA structures.[26] Two long simulations were performed when the adenine was stacked and looped-out, while two shorter simulations used start states with the bulge in the major groove and in the minor groove forming a base triple. Like the results from the present study, if the bulge starts in the stacked conformation, the bulge remains in the conformation throughout the whole simulation. However, if the DNA bulge begins in the extended looped-out conformation it rotates around the backbone to associate with the minor groove in different conformations, and then moves back into the extended looped-out form.[26] This is in agreement with our results in that nearly all of the adenine bulges in the looped-out conformation eventually associate with the minor groove. The bulge in the two shorter simulations initially associate with the major groove and minor groove to form base triple conformations. Another study ran molecular dynamics on a single-base uridine bulge in both the stacked and looped-out conformations.[27] The sequence of the structure was of Class I and was easily amendable to both the stacked and looped-out conformations. Higher base mobility and local backbone flexibility was found with the looped-out conformation as expected.

The goal of our study is to provide information on bulge motifs that will lead to the design of bulge nanotemplates. The bulge structures presented in this paper were taken from their original context in order to isolate the bulge motif for the purposes of nano-design. Several initial parameters of these structures, including helix curvature and bulge orientation, change during the molecular dynamics simulations. In many cases the bulge in its original context was initially either interacting with distal loops in the RNA far from the bulge motif or bonding with proteins, ions, or water. As the constraints are removed the bulge is free to find a new equilibrium, most probably a stable energetically favorable conformation that will be recognizable to proteins and nucleic acids. Our results indicate that the equilibrium conformation of the bulge base interacting with a protein or distal part of the helix is looped-out away from the helix for all residues. The equilibrium state of the bulge not undergoing these interactions is predominantly in the stacked position for adenine, extended position for cytosine, and sequence dependent for the other two residues. This statement concurs with experimental studies.

For the purposes of nanostructure design, the structures of class I have a predictable flexibility and stability determined by base-pair composition. A bulges surrounded by A:U base-pairs (structures 17RA and 1LMV) would be a good choice for the design of a stacked bulge nanostructure that has internal bulge switching, but maintains a constant shape. If a more flexible stacked structure is wanted, a U bulge with one G:C base-pair would be suggested. This is seen by comparing the global curvature in Figures 4 and 5. If the bulge also needs to be able to move in and out of the helix structures then a G bulge is more ideal. A design using 1JJ2_2437 would be ideal if a flexible moving bulge is needed that imposes little conformational change on the helix. It is possible that structures 1GID and 1FJG31 could be used as stacked-in looped-out switches that are initially stacked and change to looped-out when distal or remote tertiary interactions are possible. This suggestion comes from the observation that these bulges are looped-out in their original X-ray structure but stack into the helix when in isolation.

The curvatures and helix shortening for class II bulges have no clear pattern. This is possibly due to the variety of bulge conformations and bulge types that exist within the class. However, if the individual bulge type and bulge position are examined, then much stronger patterns develop. The bulge type identifies whether the bulge prefers the looped-out or stacked conformation while the bulge position is an excellent way to get a desired global helix curvature and flexibility. Extended bulges are ideal for a more rigid, highly curved helical structure because of the conserved high global curvature mean values around 57 degrees and lower associated standard deviations. Bulges contacting the minor groove are more suited for a very flexible helix where exact curvature is not as important due to the wider range of mean global curvature values that range from 29.2 to 50.6 degrees with standard deviations from 13.8 to 24.1. If Class III structures were incorporated into a nanostructure design it would be for their slightly higher flexibility and looped-out conformation.

In summary, adenine bulges surrounded by A:U base-pairs will give a very stable helix with the bulge in a stacked conformation and a global curvature mean around 53 degrees. If a more dynamic stacked bulge is needed, guanine or uracil bulges should be used. The guanine bulge maintains a 45 degree mean global curvature while the uracil bulge has a much higher global curvature mean of 79 degrees. If a looped-out conformation is desired it is important to know what type of curvature flexibility is desired, what mean curvature variance is acceptable, and if there is a preferential location for the bulge (minor groove, major groove, extended). Class III cytosine bulges consistently have a looped-out conformation, consistent global curvature means based on surrounding sequence, and high standard deviations indicating that the structure overall is quite flexible. The class II global curvature means vary extensively based on bulge type and bulge orientation, but overall have slightly lower means and standard deviations indicating a less flexible, more standard helical structure.

Future studies include fine tuning environmental factors such as ion concentration; this could rapidly induce structural changes in RNA and thus could be used for targeting specific genetic sites and environmental sensing. For example, a recent study showed a strong correlation between the magnitudes of bends for adenine and uracil bulges based on sequence and counterion valence.[61] Applying the knowledge obtained through a combination of experimental and computational simulations has therefore laid the foundation necessary for one to choose the appropriate single-base bulge sequence to deliver the desired features for the bulge-based nano-building block.

5. CONCLUSION

X-ray crystal and NMR structures show single-base bulges in a variety of conformations. We employ molecular dynamics and curvature analysis techniques to create a protocol that singles out the essential structural information for these motifs. This protocol establishes the degree of conservation in the single-base bulge due to sequence composition, bulge type and helix curvature with the goal of attaining predictable models necessary for a bulge nano-template. Through the examination of static single-base bulge structures it is clear that each structure's environment plays a critical role in its conformation. However, through the use of molecular dynamics simulations and curvature analysis, a sequence dependent bulge classification system was identified to classify the dynamic behavior of the bulges. Essential structural features, including curvature and hydrogen bonding stability are critical for a single-base bulge nano-template motif. Class I frequently has competing bulge structures and a curvature variability based on bulge type. Class II has a slightly lower curvature variability and, with the exception of a few adenine bulges, all structures remain in a looped-out conformation. Class III includes only cytosine bulges in a looped-out conformation with highly variable conformations and curvature. Additional motifs can now be examined using an analysis protocol that is similar to the one we applied to single-base bulge motifs.

Acknowledgments: This research was supported by the Intramural Research Program of the National Institutes of Health, Center for Cancer Research, National Cancer Institute.

References

1. P. B. Moore, *Annu. Rev. Biochem.* 68, 287 (**1999**).
2. N. C. Seeman, *J. Theor. Biol.* 99, 237 (**1982**).
3. E. Winfree, F. Liu, L. A. Wenzler, and N. C. Seeman, *Nature* 394, 539 (**1998**).
4. H. Yan, S. H. Park, G. Finkelstein, J. H. Reif, abd T. H. Labean, *Science* 301, 1882 (**2003**).
5. W. M. Shih, J. D. Quispe, and G. F. Joyce, *Nature* 427, 618 (**2004**).
6. B. Yurke, A. J. Turberfield, A. P. Mills, Jr., F. C. Simmel, and J. E. Neumann, *Nature* 406, 605 (**2000**).
7. K. Keren, M. Krueger, R. Gilad, G. Ben-Yoseph, U. Sivan, and E. Braun, *Science* 297, 72 (**2002**).
8. S. Horiya, X. Li, G. Kawai, R. Saito, A. Katoh, K. Kobayashi, and K. Harada, *Nucleic Acids Res. Suppl.* 2, 41 (**2002**).
9. Y. Ikawa, K. Fukada, S. Watanabe, H. Shiraishi, and T. Inoue, *Structure (Cambridge)* 10, 527 (**2002**).
10. E. Westhof, B. Masquida, and L. Jaeger, *Fold Des.* 1, R78 (**1996**).
11. R. M. Dirks, M. Lin, E. Winfree, and N. A. Pierce, *Nucleic Acids Res.* 32, 1392 (**2004**).
12. L. Jaeger, E. Westhof, and N. B. Leontis, *Nucleic Acids Res.* 29, 455 (**2001**).
13. A. Chworos, I. Severcan, A. Y. Koyfman, P. Weinkam, E. Oroudjev, H. G. Hansma, and L. Jaeger, *Science* 306, 2068 (**2004**).
14. C. R. Woes, S. Winker, and R. R. Gutell, *Proc. Natl. Acad. Sci. USA* 87, 8467 (**1990**).
15. D. A. Peattie, S. Douthwaite, R. A. Garrett, and H. F. Noller, *Proc. Natl. Acad. Sci. USA* 78, 7331 (**1981**).
16. H. Moine, C. Cachia, E. Westhof, B. Ehresmann, and C. Ehresmann, *RNA* 3, 255 (**1997**).
17. E. Ennifar, P. Walter, and P. Dumas, *Nucleic Acids Res.* 31, 2671 (**2003**).
18. Y. Xiong and M. Sundaralingam, *RNA* 6, 1316 (**2002**).
19. E. Ennifar, M. Yusupov, P. Walter, R. Marquet, B. Ehresmann, C. Ehresmann, and P. Dumas, *Structure* 7, 1439 (**1999**).
20. F. Girard, F. Barbault, C. Gouyette, T. Huynh-Dinuh, J. Paoletti, and G. Lancelot, *J. Biomol. Struct. Dyn.* 16, 1145 (**1999**).
21. J. Yoo, H. Cheong, B. J. Lee, Y. Kim, and C. Cheong, *Biophys. J.* 80, 1957 (**2001**).
22. C. Gohlke, A. I. H. Murchie, D. M. J. Lilley, and R. M. Clegg, *Proc. Natl. Acad. Sci. USA* 91, 11660 (**1994**).
23. R. S. Tang and D. E. Draper, *Biochemistry* 29, 5232 (**1990**).
24. J. Zhu and R. M. Wartell, *Biochemistry* 38, 15986 (**1999**).
25. M. Zacharias and S. Heinz, *J. Mol. Biol.* 289, 261 (**1999**).
26. M. Feig, M. Zacharias, and B. M. Pettitt, *Biophys. J.* 81, 352 (**2001**).
27. J. Sarzynska, T. Kulinski, and L. Nilsson, *Biophys. J.* 79, 1213 (**2000**).
28. M. Tamura and S. R. Holbrook, *J. Mol. Biol.* 320, 455 (**2002**).
29. W. Li, B. Ma, and B. Shapiro, *J. Biomol. Struct. Dyn.* 19, 381 (**2001**).
30. A. Maier, H. Sklenar, H. F. Kratky, A. Renner, and P. Schuster, *Eur. Biophys. J.* 28, 564 (**1999**).
31. D. J. Williams and K. B. Hall, *Biophys. J.* 76, 3192 (**1999**).
32. G. Villescas-Diaz and M. Zacharias, *Biophys. J.* 85, 416 (**2003**).
33. F. Razga, J. Koca, J. Sponer, and N. B. Leontis, *Biophys. J.* 88, 3466 (**2005**).
34. N. Pattabiraman, H. M. Martinez, and B. A Shapiro, *J. Biomol. Struct. Dyn.* 20, 397 (**2002**).
35. K. Reblova, N. Spackova, J. E. Sponer, J. Koca, and J. Sponer, *Nucleic Acids Res.* 31, 6942 (**2003**).
36. N. B. Leontis and E. Westhof, *J. Mol. Biol.* 283, 571 (**1998**).
37. N. B. Leontis and E. Westhof, *RNA* 4, 1134 (**1998**).
38. D. J. Klein, T. M. Schmeing, P. B Moore, and T. A. Steitz *EMBO J.* 20, 4214 (**2001**).
39. P. Nissen, J. A. Ippolito, N. Ban, P. B. Moore, and T. A. Steitz, *Proc. Natl. Acad. Sci. USA* 98, 4899 (**2001**).
40. C. M. Duarte, L. M. Wadley, and A. M. Pyle, *Nucleic Acids Res.* 31, 4755 (**2003**).
41. P. S. Klosterman, M. Tamura, S. R. Holbrook, and S. E. Brenner, *Nucleic Acids Res.* 30, 392 (**2002**).
42. H. M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov, and P. E. Bourne, *Nucleic Acids Res.* 28, 235 (**2000**).
43. C. M. Duarte and A. M. Pyle, *J. Mol. Biol.* 284, 1465 (**1998**).
44. R. Lavery and H. Sklenar, *J. Biomol. Struct. Dyn.* 6, 63 (**1988**).
45. D. J. Williams and K. B. Hall, *J. Mol. Biol.* 297, 1045 (**2000**).
46. D. A. Case, D. A. Pearlman, J. W. Caldwell, T. E. Cheatham III, J. Wang, W. S. Ross, C. L. Simmerling, T. A. Darden, K. M. Merz, R. V. Stanton, A. L. Cheng, J. J. Vincent, M. Crowley, V. Tsui, H. Gohlke, R. J. Radmer, Y. Duan, J. Pitera, I. Massova, G. L. Seibel, U. C. Singh, P. K. Weiner and P. A. Kollman, University of California, San Francisco (**2002**).
47. J. M. Wang, P. Cieplak and P. A. Kollman, *J. Comput. Chem.* 21, 1049 (**2000**).
48. M. Feig, A. Onufriev, M. S. Lee, W. Im, D. A Case, and C. L. Brooks III, *J. Comput. Chem.* 25, 265 (**2004**).
49. L. Y. Zhang, E. Gallicchio, R. A. Friesner, and R. M. Levy, *J. Comput. Chem.* 22, 591 (**2001**).
50. H. J. C. Berendsen, J. P .M. Postma, W. F. van Gunsteren, A. DiNola, and J. R. Haak, *J. Chem. Phys.* 81, 3684 (**1984**).
51. L. Joshua-Tor, F. Frolow, E. Appella, H. Hope, D. Rabinovich, and J. L. Sussman, *J. Mol. Biol.* 225, 397 (**1992**).
52. K. Valegard, J. B. Murray, N. J. Stonehouse, S. van den Worm, P. G. Stockley, and L. Liljas, *J. Mol. Biol.* 270, 724 (**1997**).
53. V. Thiviyanathan, A. B. Guliaev, N. B. Leontis, and D. G. Gorenstein, *J. Mol. Biol.* 300, 1143 (**2000**).
54. D. J. Patel, S. A. Kozlowski, L. A. Marky, J. Rice, C. Broka, K. Itakura, and K. J. Breslauer, *Biochemistry* 21, 445 (**1982**).

55. M. A. Rosen, D. Live, and D. J. Patel, *Biochemistry* 31, 4004 (**1992**).
56. P. N. Borer, Y. Lin, S. Wang, M. W. Roggenbuck, J. M. Gott, O. C. Uhlenbeck, and I. Pelczer, *Biochemistry* 34, 6488 (**1995**).
57. L. Joshua-Tor, D. Rabinovich, H. Hope, F. Frolow, E. Appela, and J. L. Sussman, *Nature* 334, 82 (**1988**).
58. K. Valegard, J. B. Murray, P. G. Stockley, N. J. Stonehouse, and L. Liljas, *Nature* 371, 623 (**1994**).
59. J. R. Cate, A. R. Gooding, E. Podell, K. Thou, B. L. Golden, C. E. Kundrot, T. R. Cech, and J. A. Doudna, *Science* 273, 1678 (**1996**).
60. A. Bhattacharyya, A. I. Murchie, and D. M. Lilley, *Nature* 343, 484 (**1990**).
61. M. Zacharias, *J. Mol. Biol.* 247, 486 (**1995**).

Received: 3 September 2005. Accepted: 19 September 2005.