

Mechanism of RNA Double Helix-Propagation at Atomic Resolution[†]

Srividya Mohan, Chiaolong Hsiao, Halena VanDeusen, Ryan Gallagher, Eric Krohn, Benson Kalahar, Roger M. Wartell, and Loren Dean Williams*

School of Chemistry and Biochemistry, Georgia Tech, Atlanta, Georgia 30332-0400

Received: May 6, 2008; Revised Manuscript Received: October 17, 2008

The conversion of a nucleic acid from single strands to double strands is thought to involve slow nucleation followed by fast double-strand propagation. Here, for RNA double-strand propagation, we propose an atomic resolution reaction mechanism. This mechanism, called the stack-ratchet, is based on data-mining of three-dimensional structures and on available thermodynamic information. The stack-ratchet mechanism extends and adds detail to the classic zipper model proposed by Porschke (Porschke, D. *Biophysical Chemistry* 1974, 2, pp. 97–101). Porschke's zipper model describes the addition of a base pair to a nucleated helix in terms of a single type of elementary reaction; a concerted process in which the two bases, one from each strand, participate in the transition state. In the stack-ratchet mechanism proposed here a net base-pairing step consists of two elementary reactions. Motions of only one strand are required to achieve a given transition state. One elementary reaction preorganizes and stacks the 3' single-strand, driven by base–base stacking interactions. A second elementary reaction stacks the 5' strand and pairs it with the preorganized 3' strand. In the stack-ratchet mechanism, a variable length 3' stack leads the single-strand/double-strand junction. The stack-ratchet mechanism is not a two-state process. A base can be (i) unstacked and unpaired, (ii) stacked and paired, or (iii) stacked and unpaired (only on the 3' strand). The data suggests that helices of DNA and of RNA do not propagate by similar mechanisms.

Introduction

The formation of double-stranded (ds) helices from single-stranded (ss) polynucleotides is a fundamental biological and technological process. Ss polynucleotides are converted to ds helices during RNA folding and DNA replication, detection, and sequencing. The ss to ds conversion is thought to involve a slow initial nucleation followed by fast propagation. Helix nucleation produces short helices of around three base pairs. Helix propagation is the formation of base pairs on a pre-nucleated helix. Aspects of base pairing kinetics have been recently reviewed.¹ Prentiss and co-workers studied the reverse process; the separation of DNA duplexes to single stranded-molecules at constant force. They observed rapid bursts of unzipping, punctuated by pauses.^{2,3}

Porschke^{4–6} approximated helix propagation as a zippering reaction (Figures 1 and 2). In zippering, each elementary step adds one base pair to the helix. Bases from opposing strands, in a concerted process, pair and stack on the ss-ds junction.

Here we propose an atomic resolution reaction mechanism based on available thermodynamic information and on data-mining of three-dimensional (3D) structures. The mechanism, called the stack-ratchet, may be considered to be an extension of Porschke's zipper mechanism. In the stack-ratchet, each net pairing step consists of two elementary reactions (Figures 3 and 4). One elementary reaction is the stacking of a base of the 3' single strand, resulting in a preorganized 3' single strand. A second elementary reaction is the pairing plus stacking of a base of the 5' strand. This reaction pairs a base of the 5' strand with

the preorganized (stacked) 3' single strand. A 3' stack of variable length leads the ss-ds junction.

The current study utilizes structural data-mining of large globular RNAs to dissect mechanisms for helix propagation. Many RNA fragments were observed that appear to be trapped intermediates or analogs of intermediates in propagation of helices. Among these, ss-ds junctions with preorganized 3' stacks were found at a much higher frequency than other putative intermediates. Thermodynamic data for RNA in solution^{7–12} is consistent with the data-mining frequencies, here and elsewhere.^{13,14}

Methods

Input Structures. Several large, structurally distinct RNAs, determined to high resolution, are contained within the structural database. The 23S rRNA from archaea *Haloarcula marismortui* (HM) large subunit (LSU)^{15,16} and the 16S rRNA from the bacterium *Thermus thermophilus* (TT) small subunit (SMU)¹⁷ are the highest resolution, and largest independent RNA structures in the database. The LSU of HM with 2914 observable 23S rRNA residues has a resolution of 2.4 Å resolution. The SMU of TT with 1581 observable 16S RNA residues has been determined to 2.8 Å resolution.

Secondary structural maps and data mining methods^{18–20} of these rRNAs were used to identify probable ss-ds junctions. Ss-ds junction candidates were inspected visually with Pymol²¹ and analytically with 3DNA.²² We employed a stringent definition for a junction, based on pairing and molecular interactions (below). The observed junctions are grouped and annotated (web.chemistry.gatech.edu/~williams/hel_prop).

A total of 31 ss-ds junctions were identified. Each junction when viewed in isolation appears to be partially duplex (A-form) and partially single-stranded. However, when viewed in

[†] Part of the "J. Michael Schurr Special Section".

* To whom correspondence should be addressed. E-mail: loren.williams@chemistry.gatech.edu. Address: 315 Ferst Drive, Georgia Institute of Technology, Atlanta GA 30332. Telephone: 404 894 9752.

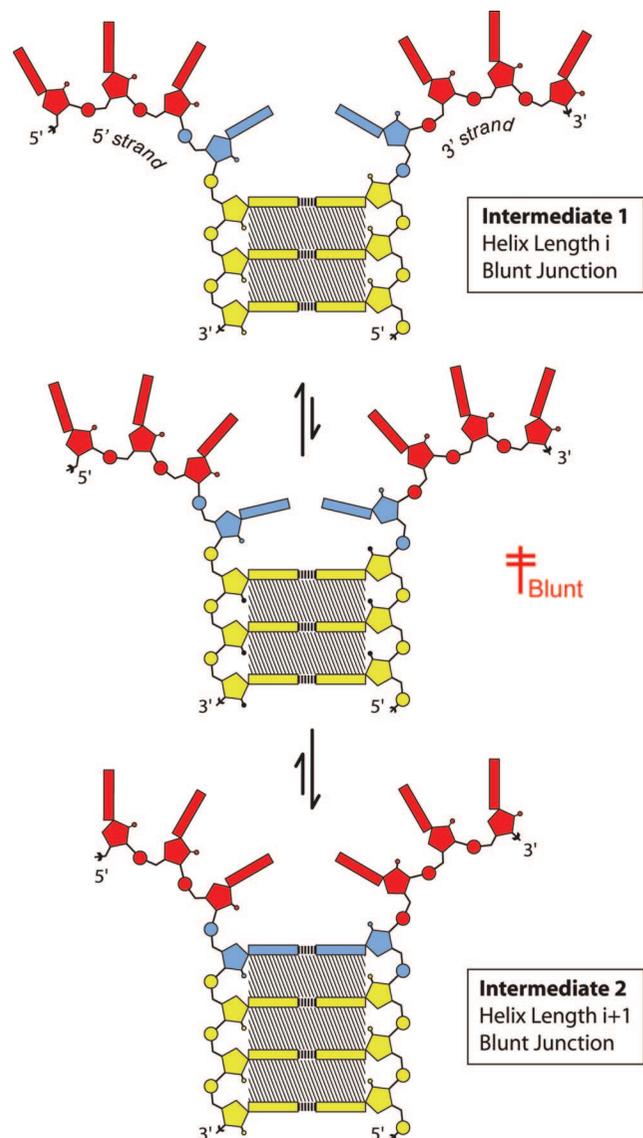


Figure 1. Helix propagation by zippering. The intermediates are blunt junctions (top and bottom panels). In the predicted transition state (middle panel), the incipient base pair is not paired or stacked. Those two residues are more restricted in conformation than single-stranded residues. Intermediate 2 (bottom) differs from Intermediate 1 (top) by an increase of one base pair. Stacking is indicated by shading. The helix is highlighted in yellow. The single-stranded region is red except for the two residues that are converted from ss to ds, which are blue.

the context of the full ribosomal assemblies, the single-stranded regions are seen to interact extensively with other RNA elements. Therefore our definition of single-stranded RNA does not imply that the RNA is not pairing with “remote” RNA elements (see below).

Molecular Interactions. The data-mining approach here requires application of explicit and consistent geometric definitions of ds and ss RNA and of stacked and unstacked bases. Each state is defined by a set of interatomic distances, which are interpreted in terms of molecular interactions. A duplex region is defined by base-pairing interactions. A single-stranded region is defined by the absence of base-pairing interactions. The hydrogen-bonding threshold is 3.4 Å. In a partial base pair, one or more, but not all, Watson–Crick or Wobble hydrogen bonds would be absent. Partial base pairs are not observed. The closing base-pair is the terminal base-pair of the helix at the

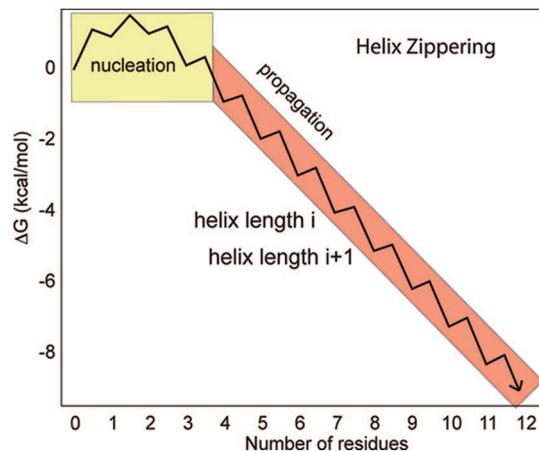


Figure 2. Energetic profile of helix zippering.⁶ There is one transition state for the addition of each base pair to the helix. Helix nucleation events are highlighted in yellow. Helix propagation is in pink.

ss-ds junction. The terminal base on the 3′ strand is called the 3′ closing base and that on the 5′ strand is called the 5′ closing base.

A ss-ds junction consists of a duplex linked to two single strands (Figure 5). One of the strands proceeds in the 5′ to 3′ direction from the closing base pair of the duplex to the terminus of the single strand and is called here the 3′ strand. The other strand proceeds in the 3′ to 5′ direction from the closing base pair of the duplex to the terminus of the single strand and is called the 5′ strand. A junction is either blunt (Figure 5A) or stacked (Figure 5B,C).

Definition of a Duplex. A double-stranded region requires at least three contiguous base-pairs with no bulges or inserts. Pairing interactions are restricted to Watson–Crick and G-U wobble pairs. Sheared and noncanonical base pairs in helical regions were disallowed. In subsequent work, this conservative definition will be expanded to determine the effects of helix length, purine-purine mismatches and other helical defects.

Definition of a Single-Strand. A single-stranded region consists of at least three contiguous residues whose bases do not engage in hydrogen bonding interactions with bases of the opposing strand. The opposing strand is defined in terms of the adjoining duplex. Allowed interactions in single-stranded regions are (i) base-backbone and backbone-backbone hydrogen bonding between opposing strands, (ii) base–base stacking interactions between opposing strands, and (iii) base–base hydrogen bonding interactions with bases not of the opposing strand.

Stacked and Blunt Junctions. In a blunt junction, one face of the closing base pair is unfettered; the ss bases do not stack on the closing base pair (Figure 5A). In a stacked junction, a ss base stacks on one or both bases of the closing base pair (Figure 5B,C).

We have found it informative to cluster stacked junctions by several criteria including the strand of the stacked ss base(s) (5′ or 3′), the stacking mode with respect to the closing base-pair (intrastrand, interstrand, or both-strand), the length of the stack, and the sequence of the stack.

Stacking: Intrastrand, Interstrand, and Both Strand. Stacking of RNA has been geometrically defined and quantified previously by Turner and co-workers¹⁴ and by Chattopadhyaya and co-workers.²³ In that work, two bases are considered to be stacked if the rise between them is not greater than 4 Å, the roll or tilt angles are not greater than 30°, and the bases overlap when projected onto the helical axis with at least one ring atom overlapping with the ring of the base upon which it is stacked.¹⁴

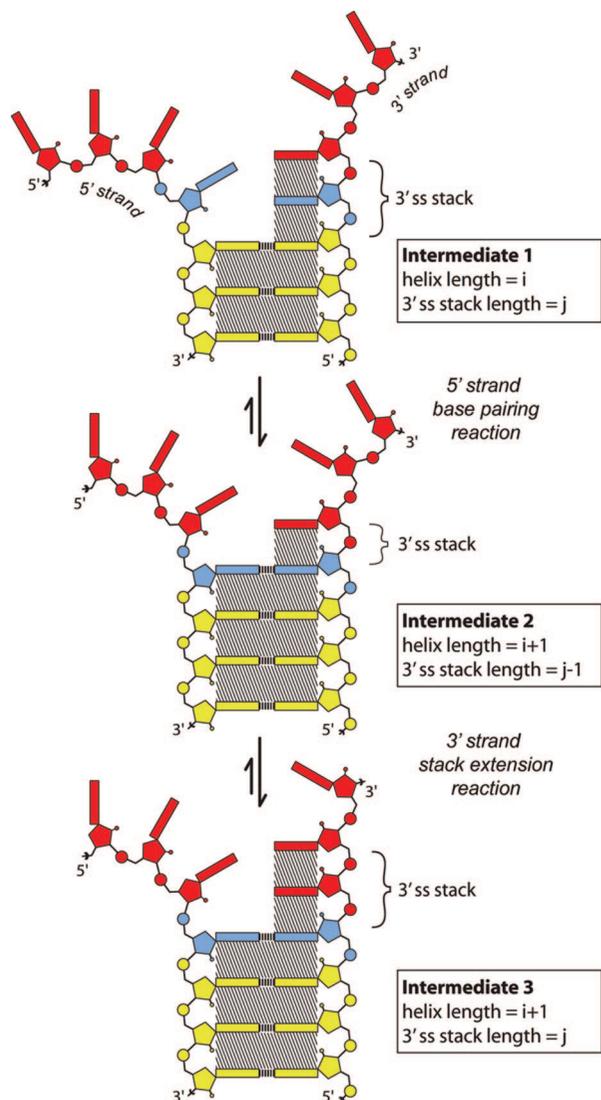


Figure 3. The stack-ratchet mechanism of helix propagation. In Intermediate 1 (top panel), some 3' bases are stacked but not paired. 5' ss bases are neither stacked nor paired. To reach Intermediate 2 (middle panel), the complex must pass through the 5'BP transition state. In Intermediate 2, the 5' base is stacked and paired. To reach Intermediate 3 (bottom panel) the complex must pass through the 3'SkEx transition state. In Intermediate 3, the 3' single strand stack has been extended. Stacking is indicated by shading. The length of the 3' ss stack is variable. The initial nucleated helix is highlighted in yellow. The single-stranded region is red except for the residues that are converted in this step from ss to ds, which are blue.

To characterize stacking here, Olson's program 3DNA²² was used to determine local helical parameters such as rise, shift, slide, roll, and tilt along with "area of overlap" (also see Chattopadhyaya). Junctions were analyzed in a stepwise process: (i) Each ss-ds junction was treated as two distinct single-strands to quantify intrastrand stacking. (ii) Each ss-ds junction was treated as a fully double-stranded duplex to quantify interstrand stacking in the junction region. Numerical output obtained through 3DNA was confirmed by visual inspection with Pymol. Pairwise van der Waals (vdw) contacts of atoms were used to further characterize stacking. The axial projection and pairwise vdw contacts are used to distinguish between intrastrand and interstrand stacking. Stacking of the first ss base at the closing base pair is characterized here as intrastrand, interstrand, or both-strand. Intrastrand stacking gives base-base overlap of greater than 0.1 \AA^2 with an adjacent paired base on the same strand,

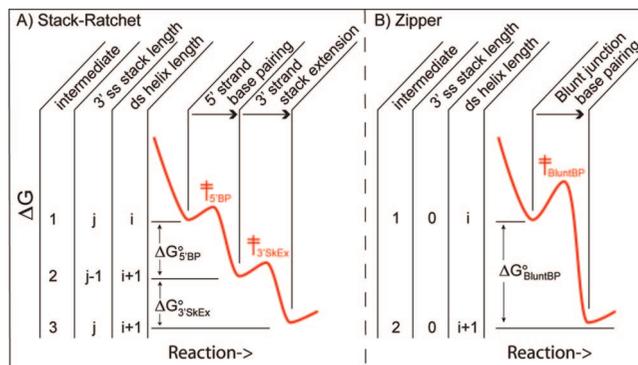


Figure 4. Energetic profile for helix propagation by stack-ratchet and zipper mechanisms. (A) The stack-ratchet, with two elementary reactions and two transition states. In one reaction, a base pairs and stacks on the 5' strand. In a second reaction, the 3' ss stack lengthens. (B) Zippering, with one elementary reaction and one transition state. The activation energy is higher for the zipper than the stack-ratchet. Conformational changes of both strands are required to reach the transition state for the zipper while changes of only one strand are required to reach each stack-ratchet transition state.

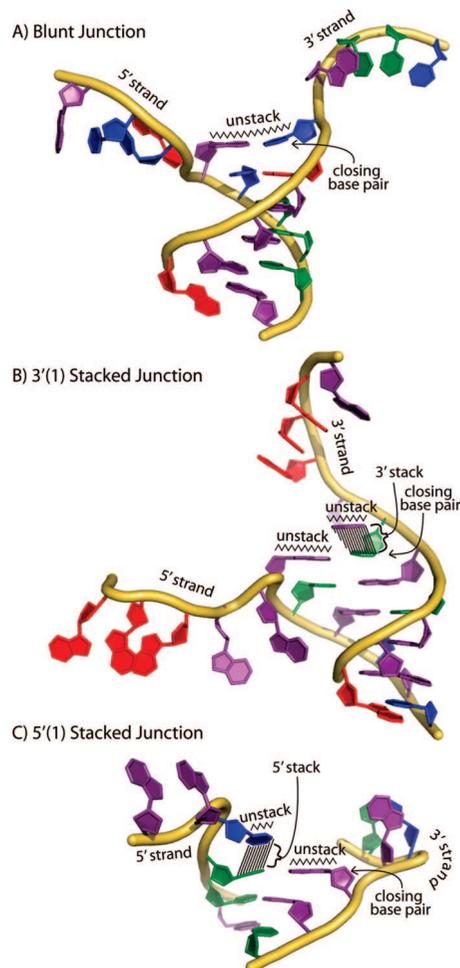


Figure 5. Representative ss-ds junctions observed in the three-dimensional database. (A) A blunt junction. (B) A 3'(1) stacked junction. (C) A 5'(1) stacked junction. Adenosine is red, guanosine is violet, uridine is blue, and cytosine is green.

along with poor interstrand stacking (Figure 5B,C). Interstrand stacking gives overlap with a base on the opposite strand of the helix of greater than 0.1 \AA^2 , along with poor intrastrand stacking (less than 0.1 \AA^2). Both-strand stacking gives comparable base-base overlap with each base of the closing base pair, of at least 0.1 \AA^2 for each.

In general, our criteria are in accordance with previous definitions.^{13,14} The “area of overlap” parameter in 3DNA is a useful quantitative measure of the extent of stacking. For single stranded regions, the area of overlap sometimes incorrectly identifies unstacked bases as stacked. Visual inspection reveals that highly noncoplanar and twisted bases are problematic. In these cases, additional parameters were used as criteria. In addition to requirements for the base–base rise, tilt and roll, helical twist not between 0 and 60° excludes bases from the stacked classification. If any single stacking criterion is not satisfied, the area of overlap value is set to “zero”.

To quantify the dispersion of area of overlap, an ideal A-form RNA 20-mer was built using the 3DNA program. The area of overlap values in the single strand treatment and double strand treatment were determined. With respect to the single-strand treatment, whenever overlap values were zero, visual inspection was used to assess overlap. If there was no overlap, that particular base-step was eliminated. Nine of 53 base steps were thus eliminated. These were Pyr–Pyr or Pyr–Pur steps. For the remaining steps, the average overlap is 3.5 Å² (SD=2.4). Therefore 0.1 Å² overlap is less than 2 SDs from the mean.

In general one can describe a base as being in one of two states, either stacked or unstacked. Although the frequency is low, one also observes partially stacked structures that cannot be assigned to either stacked or unstacked states. These ‘partially stacked states’ are indicated by low overlap values, and by only one or two pairs of atoms in vdw contact.

Calculation of Junction Stability. The stacked junctions represent possible intermediates in helix propagation reactions. Thermodynamic calculations were utilized to evaluate possible helix propagation intermediates. Within each helix junction, the three terminal Watson–Crick base pairs of the ds regions were identified along with the first three unpaired bases of the 3′ single strand. Nearest neighbor stacking free energies of each base pair were estimated using UNAFold.²⁴ Additionally, stacking free energies and probable secondary structures of each junction was estimated using the Vienna RNA package.²⁵ The free energy contributions for the first stacked, unpaired bases were estimated with Turner’s RNA energy rules for dangling ends.^{8,26,27} Free energy contributions for the second unpaired, stacked bases were estimated from results of Serra and co-workers.^{10,12}

Possible Kinetic Traps. Kinetically trapped ss-ds junctions (off-pathway) can be identified by interactions that would require disruption before helix propagation could proceed. Extruded bulges within the 3′ ss stack are considered to be candidates for kinetic traps. A junction is classified as a possible kinetic trap if such “incorrect” interactions are observed. These interactions might be within either single strand, or between the two ss regions, or between a single strand and the duplex region. A treatment of kinetically trapped ss-ds junctions is in progress.

Results

In the stack-ratchet mechanism of RNA helix propagation, each net pairing step consists of two elementary reactions (Figure 3). Significant motions of one strand only are required to achieve a particular transition state.

We call one elementary reaction the base-pairing step (5′BP), (figure 3, top). In this step an unpaired base of the 5′ strand stacks on closing base-pair at the junction and forms a base pair with the preorganized 3′ strand. We refer to the other elementary reaction as the stack extension step (3′SkEx), (Figure 3, bottom). In this reaction a base within the ss region joins the

stack of the 3′ strand. The 3′SkEx reaction preorganizes the 3′ single strand; stacking of the 3′ strand is distinct from base pairing.

The Pairing Reaction (5′BP). The 5′(BP) reaction appears to be a two state process. The stacking and base–base hydrogen bonding interactions of the incipient base-pair at the 5′BP step appear to form simultaneously. Structures with only one type of interaction, stacking or hydrogen bonding, are not observed. Similarly we do not observe partially formed base pairs.

The Stacking Reaction (3′SkEx). The stacking mode appears to be dominated by the tendency of the first 3′ ss base to stack upon the closing purine (Table 1). Stacking mode at a junction is classified here as intrastrand (same strand only), interstrand (opposing strand only) and both-strand stacking. For a 3′ stacked junction with a closing 5′Pu–Py3′ base pair one observes primarily interstrand or both-strand stacking. When the closing base pair is 5′Py–Pu3′ one observes primarily intrastrand stacking. It appears that the ss base finds and stacks upon the purine of the closing base pair.

Stacking within the ss region is most common between bases that are contiguous on the backbone (i.e., between adjacent residues). However in rare cases noncontiguous bases stack, resulting in bulges and more complex conformations (see Kinetic Traps in Methods).

2-State, 3-State..n-State. The stack ratchet mechanism is formally a three-state process. One step of the reaction converts a 5′ unpaired/unstacked base and a 3′ unpaired/stacked base to a stacked paired state in the 5′BP step, thereby adding a base pair to the duplex. To maintain the leading stack, another step of the reaction converts a 3′ unpaired/unstacked base to the unpaired/stacked state in the 3′SkEx step.

However stacking does not appear to be a two-state (“all or none”) process. Our results indicate a continuum between stacked and unstacked such that some states are best described as intermediate. These junctions have partially stacked bases in the ss region. The non-two-state behavior of base stacking in solution has been noted previously.²⁸ We have used nonintegral numbers for the stack length to indicate partial stacking, as in 3′(1.5) junction, 3′(2.5) junction, 5′(0.5) junction, and so forth. Of the 31 junctions, 6 show intermediate stacked/unstacked states. In partially stacked states, base–base overlap is low and two or fewer pairs of atoms are in vdw contact. Partial stacking is seen in 3′ stacked junctions *o*, *y*, *z* and *u* and in 5′ stacked junctions *bc* and *bd* (Table 1). Partial stacking is most commonly seen in ss stacks greater than one base in length. Therefore the three-state reaction mechanism of Figure 3 must be considered a simplification. The observation of partial stacking suggests that the 3′ SkEx reaction is not a simple two-state process. By contrast, partially base-paired structures (i.e., structures in which all possible hydrogen bonds are not formed between pairing bases) are not observed. Therefore the 3′SkEx step but not the 5′BP step most probably consists of a composite of several more subtle elementary reactions.

Junctions. Clear trends in stacking are evident from the data-mining results of the ss-ds junctions of HM-23S and TT-16S rRNAs. These trends correlate with the results of Turner¹⁴ and Chattopadhyaya¹³ on the thermodynamics and data-mining of dangling ends on RNA duplexes. In addition to the 31 helical junctions used to support the stack-ratchet mechanism proposed here, eight potential kinetic traps were identified, which were not included in the analysis.

Blunt Junctions. Blunt junctions, with clean unstacked helical termini (example shown in Figure 5A) are rare in HM-23S and TT-16S rRNAs and are not considered probable

TABLE 1: Area of Overlap between Closing Base-Pair and the First ss Base in Stacked Junctions: Intrastrand, Interstrand and Cross-Strand Stacking

junction	PDB ID/Closing base pair residue numbers ^a	closing base pair ^b	first ss base ^c	intrastrand overlap (Å ²) ^d	interstrand overlap (Å ²) ^e	stacking type
(A) 3'(1) junctions						
a	HM G887-C774	G-C	G	0.05	2.25	interstrand
g	TT G9-C25	G-C	A	0.04	4.47	interstrand
i	HM G2293-C2315	G-C	G	0.08	3.68	interstrand
b	HM G539-C617	G-C	G	1.64	0.87	both-strand
c	TT G577-C764	G-C	G	1.82	0.38	both-strand
e	TT G1184-C1116	G-C	G	1.24	3.41	both-strand
j	HM G747-C658	G-C	A	0.40	0.19	both-strand
h	HM G1986-C2002	G-C	U	5.29	0.00	intrastrand
f	TT G567-C883	G-C	U	0.31	0.00	intrastrand
(B) 3'(2) junctions						
k	TT G39-C403	G-C	U	0.00	0.91	interstrand
o	TT G144-C178	G-C	A	0.00	5.00	interstrand
t	TT G548-C36	G-C	U	1.36	0.90	both-strand
s	HM G661-C685	G-C	A	0.55	2.93	both-strand
l	TT C240-G286	C-G	U	7.15	0.00	intrastrand
m	TT G406-C436	G-C	U	1.32	0.00	intrastrand
n	TT C1113-G1187	C-G	A	5.31	0.00	intrastrand
q	HM G1045-C1069	G-C	A	5.55	0.00	intrastrand
r	HM A2118-U2276	A-U	U	1.87	0.00	intrastrand
(C) 3'(3)+ junctions						
z	HM A1494-U1511	A-U	G	0.00	5.09	interstrand
ba	TT G316-C337	G-C	A	0.00	2.20	interstrand
w	TT G289-C311	G-C	C	0.00	1.25	intrastrand
v	HM G636-C1365	G-C	C	0.00	1.00	interstrand
d	TT G821-C879	G-C	C	0.03	0.38	interstrand
y	HM C915-G928	C-G	A	5.53	0.00	intrastrand
x	HM C2084-G2660	C-G	U	4.65	0.00	intrastrand
u	HM C905-G1300	C-G	C	2.18	0.00	intrastrand
(D) 5' junctions						
be	TT G541-C504	C-G	G	5.03	0.00	intrastrand
bd	HM C2409-G2418	G-C	A	5.35	0.00	intrastrand
bc	TT G881-C569	C-G	C	0.03	0.00	blunt/intrastrand

^a Allows identification of junctions in the PDB, TT and HM represent *Thermus thermophilus* and *Haloarcula Marismortui* respectively. The PDB IDs are HM, IJ2 and TT, 2J00. ^b Closing base pair (5'–3'). The base that stacks on the first ss base is bold. ^c The first ss stacked base from the 3' strand in the 3' stacked junctions (sections A–C) and by the 5' strand in the 5' stacked junctions (section D). ^d Area of overlap in single-strand treatment determined by 3DNA. ^e Area of overlap in double-strand treatment determined by 3DNA.

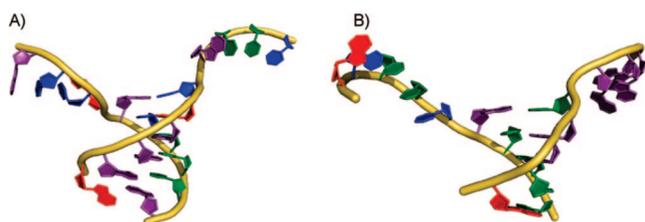


Figure 6. Both blunt junctions observed here. The 3' ss terminus is at the top right of each junction. The coloring scheme is the same as in Figure 5.

intermediates in the A-form helix propagation reaction. Only 2 of 31 observed ss-ds junctions are blunt. Both blunt junctions are shown in Figure 6. Blunt junctions are the lone reaction intermediate expected in the zipper mechanism (Figure 1), and they would be expected at high frequency if zippering were the primary mechanism of helix propagation.

Stacked Junctions. Stacked junctions (examples are shown in Figure 5B,C) are probable intermediates in helix propagation. Twenty-nine of 31 observed ss-ds junctions are stacked. In stacked junctions, the bases of the ss region stack upon the closing base-pair; the stacking within the helix extends into the ss region. Such stacking is observed for the 3' strand or the 5' strand but generally not for both strands simultaneously. This

is to say that simultaneously stacked yet non-hydrogen bonded “base pairs” are not observed.

The most general pattern observed for ss-ds junctions is 5'G-C3' closing base-pairs (22 of 31 junctions, Figure 9). This closing base-pair is preferred for all classes of junctions but most strongly for those with short ss stacks. A is not observed on the 3' side of the closing base pair. The frequency of C on the 3' side of the closing base pair decreases with increasing length of the 3' ss stack.

Stacked junctions have been grouped here by several criteria including the strand of the stack (the 3' strand, Figure 5B, or the 5' strand, Figure 5C), the stacking mode (intrastrand, interstrand, or both-strand), the stack length and the stack sequence. We have developed a nomenclature to describe parameters such as length and strand of ss stack. In a 3'(1) junction, the first ss base of the 3' strand stacks upon the closing base pair, followed by a break in the stack (Figures 5B and 7). In a 3'(2) junction, a stacked 3' ss base is followed by another stacked ss base, then by a break in the stack (Figure 8). A 5'(1) junction is the same as a 3'(1) junction except that the stacked base is contributed by the 5' strand (Figure 5C).

3' Stacked Junctions. Bases of the 3' ss strand stack upon the closing base pair in 26 of 31 junctions identified here. Nine 3'(1) junctions are observed (Figure 7). All observed 3'(1) junctions close with 5'G-C3' base pairs (9 of 9 junctions, Figure

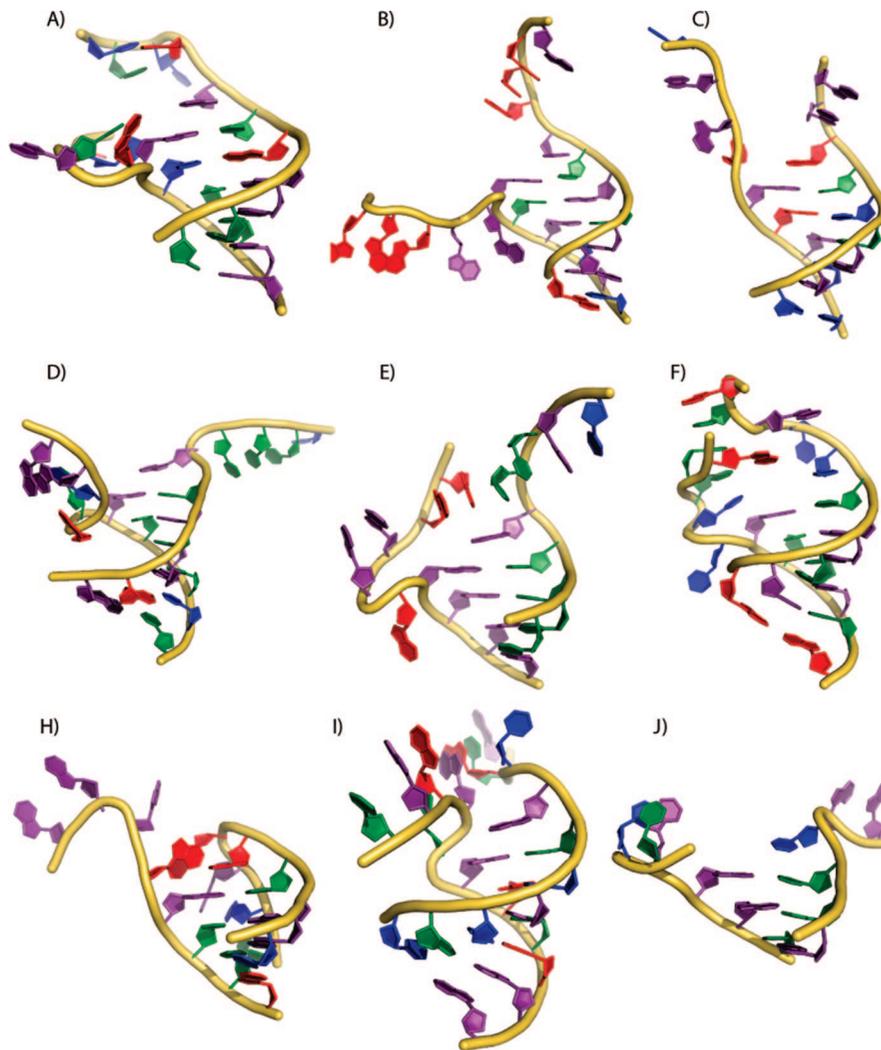


Figure 7. Nine 3'(1) junctions observed here. The 3' ss terminus is at the top right of each junction. The coloring scheme is the same as in Figure 5.

9A, Table 1). The 3'(2) group with nine members (Figure 8) is observed with the same frequency as 3'(1) junctions. 3'(2) Junctions most commonly close with 5'G-C3' base pairs (6 of 9 junctions, Figure 9b, Table 1). Two 3'(2) junctions close with 5'C-G3' base-pairs and one with a 5'A-U3' base-pair. The first 3' ss residue is commonly U (5 of 9) or A (4 of 9). There is no obvious sequence preference at the second 3' ss position. The 3'(3+) group contains stacked ss regions varying in length from three to eleven residues. Eight 3'(3+) junctions are identified among the 31 ss-ds junctions. These junctions show greater variation in closing base pair than the 3'(2) junctions. Four of the 3'(3+) junctions close with 5'G-C3' base pairs, three close with 5'C-G3' and one closes with 5'A-U3' (Table 1). The first 3' ss base is most commonly C (4 of 8 junctions), while A occurs twice and U and G once each (Figure 9C).

5' Stacked Junctions. Stacking of the 5' single strand on the closing base-pair is infrequent and is observed in only three of 31 junctions (Table 1). Two 5' stacked junctions close with 5'C-G3' base-pairs and one closes with 5'G-C3'. In one junction, a C on the 5' strand stacks on the closing base-pair. In the other two junctions a purine stacks upon the closing base pair. Observed 5' stacked junctions exhibit intrastrand stacking exclusively.

Local Free Energy Minima along the Helix Propagation Reaction Coordinate. Free energies were evaluated for the RNA duplex segments and dangling ends of the 3' stacked

junctions using Turner's parameters.^{8,26,27} Table 2 lists the average contributions of the duplex segment and 3' ss stack to the three classes of junctions. The results imply that 3'(1) junctions are most common when the duplex segment is most stable ($\Delta G^{\circ}_{\text{folding}} \sim -5.4$ kcal/mol) and that the first 3' ss base contributes significant stability ($\Delta\Delta G^{\circ}_{\text{folding}} \sim -1.6$ kcal/mol). The 3' ss stacks of 3'(1) junctions produce the most stable dangling ends (-1.7 kcal/mol for a G or A and -1.2 kcal/mol for a C; Figure 9). The combined stability of the helix and the first unpaired 3' ss residue (helix + ss1) decreases with the length of the 3' ss strand. The $\Delta G^{\circ}_{\text{folding}}$ (average helix + ss1) is -7.0 kcal/mol for 3'(1) junctions, -6.1 kcal/mol for 3'(2), and -5.7 kcal/mol for 3'(3+). When a 3' stacked junction has a pyrimidine at the first ss position, additional 3' stacked bases are common. This pattern is reflected by the thermodynamics of the 3'(2) and 3'(3+) stacked junctions (Table 2). Commonly, with a pyrimidine at the first position, the contribution of the first base to stability is relatively small ($\Delta\Delta G^{\circ} \sim -0.8$ kcal/mol) and appears to require augmentation from additional stacked bases.

Discussion

RNA Conformational Transitions. RNA conformational transitions help control processes in small systems such as riboswitches²⁹⁻³¹ and in large systems such as ribosomes.³²⁻³⁴ Riboswitches undergo conformational changes in response to

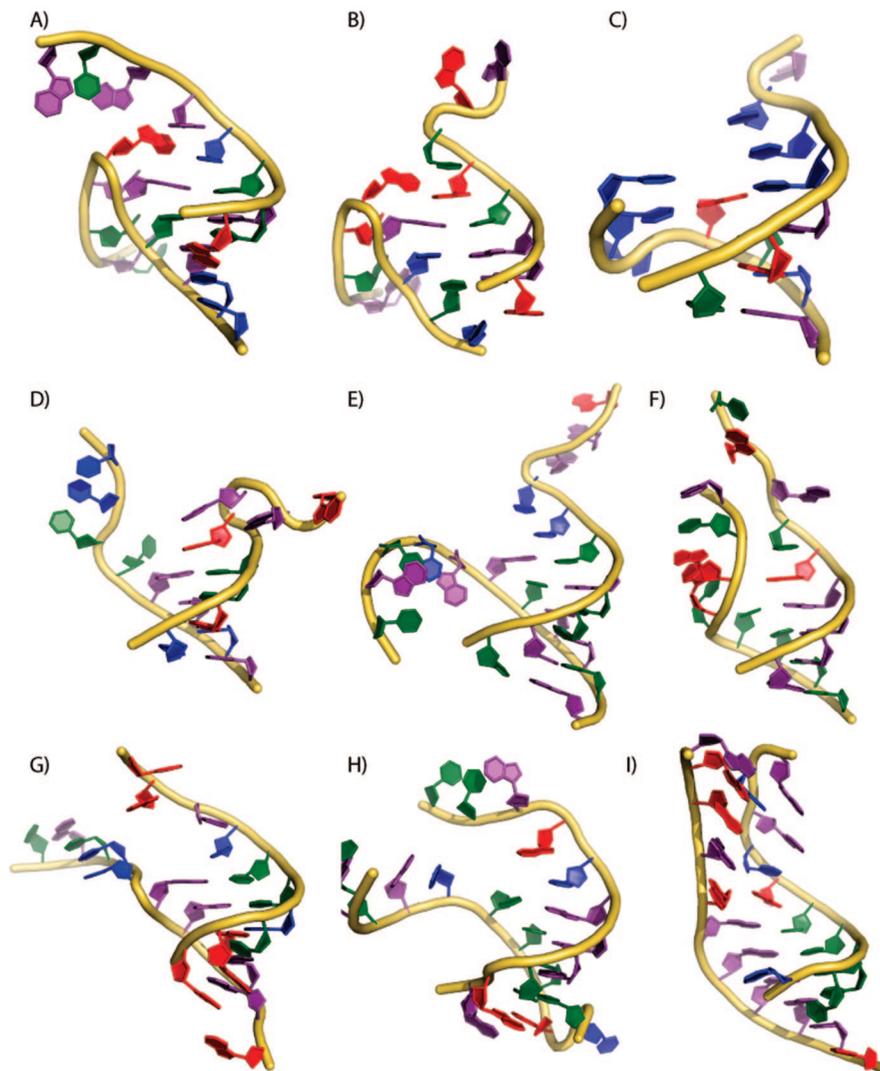


Figure 8. Nine 3'(2) junctions observed here. The 3' ss terminus is at the top right of each junction. The coloring scheme is the same as in Figure 5.

small-molecule binding. Ribosomes undergo conformational changes during translation.

Database Mining. RNA conformational transitions can be understood by analysis of static crystal structures. Relative populations over a large number of crystal structures reflect populations and relative energies in solution.^{35,36} Structural databases allow determination of averages and deviations of hydrogen bond and covalent bond lengths, bond angles and dihedrals.^{37,38} Structural databases also allow determination of coordination sphere geometry^{39–41} and reaction coordinates and transition pathways.^{42–47}

Ho and co-workers proposed a reaction coordinate for the transition of DNA between B-conformation and A-conformation, based on a series of DNA crystal structures.⁴⁸ There the transition is frozen at various points along the reaction coordinate by lattice forces and by intramolecular restraints (some of the DNA molecules are modified to alter their conformation). A series of structures was sorted, starting with the structure that most closely resembles canonical B-conformation, and ending with the structure that most closely resembles canonical A-conformation.

Sundaralingam examined some conformational transitions during protein folding.⁴⁵ He proposed that ground-state protein structures contain trapped intermediates. These intermediates are not at local minima in energy but are trapped in a global

energy minimization. He inferred that a water molecule can “pry” open an α -helix, converting it to a reverse turn. In that work, the intermediates are trapped by intramolecular forces within a globular protein and not by lattice forces.

In our work, structural data-mining of large globular rRNA has been utilized to define and characterize ss-ds junctions. Our data-mining results support the stack-ratchet mechanism of helix propagation. 3' Strands with preorganized unpaired stacks of one or more bases at helical junctions are very frequent. Unstacked (blunt) junctions are the least frequent. Known thermodynamic data (below) at helix junctions strongly correlates with the frequency of observation. The results suggest that unstacked intermediates are not favored during helix propagation. The low frequency of 5' stacked junctions combined with a high frequency of 3' stacked junctions, suggests that 5' stacked intermediates are improbable in RNA helix propagation. The frequencies suggest that during helix propagation, the 3' strand nearly always has at least one unpaired stacked base (Figure 3).

Helix Propagation in RNA. Our interests are in determining molecular-level mechanisms of RNA conformation transitions. The methods utilize analysis of 3D databases and published thermodynamic data and are generalizable to a variety of RNA conformational transitions. Here we focus on RNA helix propagation.

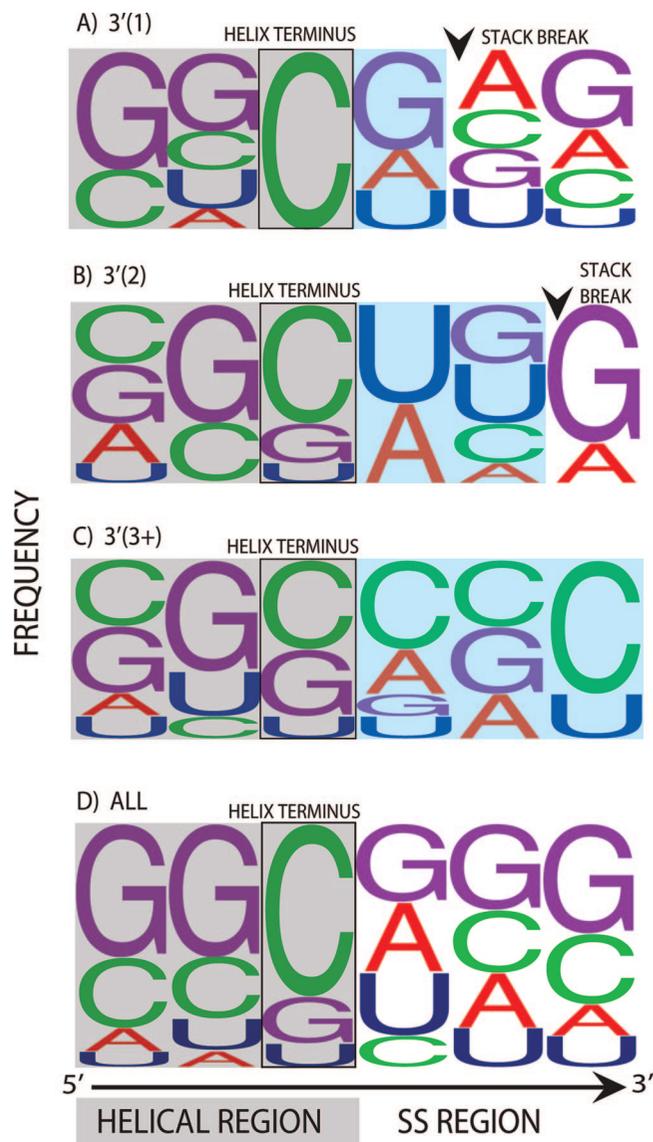


Figure 9. Base frequencies⁵⁹ of 3' stacked ss-ds junctions. (A) 3'(1) junctions. (B) 3'(2) junctions. (C) 3'(3+) junctions. (D) All 31 3' stacked junctions combined. The letter size corresponds to the relative frequency of that base at that position on the 3' strand. Six positions on the 3' strand are represented. The closing base pair is indicated by a box. The ds regions are shaded gray. The ss stacked regions are shaded blue. The first three residues are within the ds region and the last three are within the ss region. The stack break is indicated by a black arrowhead.

Ss-ds helix junctions extracted from the 3D database (1JJ2; 23S rRNA, and 2J00; 16S rRNA) appear to contain imbedded intermediates in helix propagation reactions. Our premise is that these ss-ds helix junctions on average reflect the same stabilizing and destabilizing influences as intermediates in helix propagation processes in solution. Ss-ds junctions in crystal structures are trapped by their surroundings and sometimes by their sequence. For a small number of examples the idiosyncrasies of a particular trapping environment would overwhelm information intrinsic to the junction in isolation. For a large number of junctions the specific effects average out, and one can infer information that is relevant to solution behavior. The results allow evaluation of possible mechanisms of helix propagation.

The Zipper. Porschke approximated ss to ds propagation of DNA and RNA as zippering reactions (Figures 1 and 2). Zippering is series of reversible elementary steps, with uniform

forward and uniform reverse rate constants.^{4,5} Each elementary step adds one base pair to the helix (Figure 1). Zippering is a two state process in that a base is in either a single-stranded state or a base-paired state. Paired bases are stacked. Unpaired bases are unstacked. Zippering is a concerted process in that two bases, one from each strand, participate in the transition state (Figure 1, center panel). The transition state requires restricted conformation, and the absence of hydrogen bonding and stacking interactions for both members of the incipient base pair. This limitation in the two-state approximation inherent in the zipper model is discussed by Porschke,^{4,5} who observed that in reality, "...base pairing... is not a simple conversion between two-states only."

The Stack-Ratchet. The atomic resolution mechanism for helix propagation proposed here for RNA, the stack-ratchet, is an extension of Porschke's zipper model. The combined thermodynamic and structural data support the stack-ratchet as a reasonable approximation of the mechanism of helix propagation for RNA.

In the stack-ratchet model, each net pairing step consists of two elementary reactions (Figures 3 and 4). Motions of one strand only are required during each elementary reaction to achieve a given transition state. One elementary reaction is the pairing and stacking of a base of the 5' strand of the junction to a preorganized 3' single strand. This reaction passes through the 5'BP (5' base pairing) transition state. A second elementary reaction preorganizes the 3' single strand, driven by stacking interactions. This step passes through the 3'SkEx (3' stack extension) transition state. This reaction stacks bases of the 3' single strand. Stacking of the 3' single-strand of the junction is not simultaneous with base pairing with the 5' single-strand. A 3' ss stack leads the ss-ds junction. The length of the leading stack is expected to be variable, depending on sequence, temperature, etc.

The stack-ratchet mechanism (Figures 3 and 4A) is not a two state process. Bases can be (i) unstacked and unpaired, (ii) stacked and paired, or (iii) stacked and unpaired (on the 3' strand only). Each elementary step seems facile and consistent with known behaviors of nucleic acids. The stack-ratchet does not require concerted motions or problematical transition states.

The Stacking Reaction: Competing Parallel Mechanisms.

The three-state stack-ratchet appears to be a simplification of true A-form helix propagation. The 3'SkEx step in particular, probably represents a composite of several more subtle elementary steps, with the possibility of competing parallel mechanisms. The stacking reaction of one base upon another does not appear to be a two-state process in our data mining or in solution.^{26,49} In addition it is likely that stacked bases (single-stranded) can join the 3' stack in groups of various sizes in a single step, allowing parallel mechanisms.

Local Free Energy Minima. The most frequently observed junctions in 3D structures appear to reasonably represent local minima in the free energy surface in solution. Dangling ends confer significant stability to ds RNA when attached to the 3' but not to the 5' end.^{50,51} RNA helices with no dangling ends are generally less stable than helices with dangling ends. Turner and co-workers previously examined the structural database and, concluded that sequence-modulated probabilities of stacking at ss-ds junctions correlate with solution free energies of stacking.¹⁴

Relationships among 3' dangling end sequence,⁷⁻¹⁰ length,^{10,11} stacking geometry,^{13,14} and phylogeny¹² have been investigated. Chattopadhyaya and co-workers analyzed specific stacking geometries of dangling ends¹³ and concluded that stabilization is proportional to the extent to which a ss base stacks on the

TABLE 2: Thermodynamic Evaluation of 3' Stacked junctions

stacked junction	average free energy of formation of duplex (ΔG_{37}°) (kcal/mol)	average free energy of formation of unpaired, stacked 3' ss base ($\Delta\Delta G_{37}^{\circ}$) (kcal/mol)			average free energy of formation of helix + first ss base (ΔG_{37}°) (kcal/mol)	average total free energy of formation at junction (ΔG_{37}°) (kcal/mol)
		first ss base	second ss base	third ss base		
3'(1)	-5.4 ± 0.8	-1.6 ± 0.2	NA	NA ^a	-7.0 ± 0.7	-7.0 ± 0.7
3'(2)	-5.3 ± 0.6	-0.8 ± 1.1	-0.2 ± 0.3	NA ^a	-6.1 ± 1.4	-6.3 ± 1.6
3'(3+)	-4.8 ± 1.3	-0.9 ± 0.4	-0.2 ± 0.3	-0.1^b	-5.7 ± 1.4	-5.9 ± 1.4

^a Not applicable. ^b Estimated.

hydrogen bonds of the closing base pair. In sum, the stabilities of stack-ratchet intermediates compared to alternatives are consistent with known thermodynamic effects of 3' dangling ends.

One anticipates that the stabilities of various intermediates in RNA helix propagation would be modulated by sequence. Indeed closing 5'G-C3' base pairs with stacked 3' purines are observed much more frequently than other sequences (Figure 9). Thus the regular sawtooth pattern of Figure 4A would in reality be irregular with valleys of various depths. It is likely that RNA ss-ds junctions with closing 5'G-C3' base pairs with stacked 3' purines might be quasi-pause sites in the helix propagation reaction. The most frequent junctions would represent the deepest local wells in free energy. Thermodynamic calculations (Table 2) do indeed suggest that the relative stabilities of the frequently observed ss-ds junctions are due to the sequence-dependent stability of the helix and the 3' stacked region. Less stable helices appear to be compensated by longer ss stacks.

Reaction Rates and Transition States. One of the steps proposed here in helix propagation is the stacking of a 3' single-stranded residue onto the adjacent unpaired base (3' strand stack extension reaction, Figure 3). The transition state for this process is anticipated to be conformationally restrained but not stabilized by hydrogen bonding or stacking interactions. It seems reasonable to approximate the transition state free energy using the entropy of the stacking reaction, which is similarly accompanied by a loss of conformational freedom.⁵² The appropriate equilibrium entropic parameters are available from reported thermodynamic measurements on 3' double-nucleotide overhangs.¹⁰

For a helix with one stacked/unpaired residue on the 3' side, the ΔS° of stacking of an additional residue ranges from -7 to -20 eu.¹⁰ These parameters give a $\Delta G^{\circ\ddagger}$ of 2.1 to 6.0 kcal/mol at 298 K for the 3' strand stack extension reaction. The range in $\Delta G^{\circ\ddagger}$ values reflects sequence variation as well as experimental error. Utilizing this range of $\Delta G^{\circ\ddagger}$ in the Eyring equation gives $k_{3'SkEx} = kT/h \exp(-\Delta G^{\circ\ddagger}/RT)$, where $k_{3'SkEx}$ is the first order rate constant of 3' ss base stacking, k is the Boltzmann constant, and h is Planck's constant, one obtains $3 \times 10^8 \text{ s}^{-1} < k_{3'SkEx} < 2 \times 10^{11} \text{ s}^{-1}$. This range for the $k_{3'SkEx}$ is roughly equivalent to the rate constant observed for flavin ethenoadenine dinucleotide ($k_{Stack} = 1.3 \times 10^8 \text{ s}^{-1}$).⁵³ The latter analog has five more bonds linking the chromophores than a dinucleoside monophosphate and hence is expected to have a slower rate of stacking. Our estimated $k_{3'SkEx}$ is more than the experimental rate obtained by Porschke for stacking of poly(A) ($2-5 \times 10^7 \text{ s}^{-1}$).⁵⁴ Torsional restraints imposed by the junction are expected to increase the rate of stacking in the 3' stack extension reaction (SkEx) relative to that in purely single-stranded polynucleotides. The sequence-dependence of the entropy of stacking would modulate the activation energy for the SkEx reaction.

In the second step in helix propagation, a 5' single-stranded residue stacks on the helical junction and pairs with the opposing base (5' strand base pairing reaction, Figure 3). In this case,

the entropy for stacking of 5' single-nucleotide overhangs may not provide a good estimate for $-\Delta G^{\circ\ddagger}/T$, because their stacking on helical junctions is so poor,⁸ and the residual entropy of the 5' dangling residue is so high.⁵⁵

However one may estimate rate constants by another method. Here we use the rate constant for the bimolecular nucleation of two complimentary strands ($\sim 5 \times 10^5 \text{ M}^{-1} \text{ s}^{-1}$)⁵⁶ and correct for the estimated effective concentration of the 5' base in the vicinity of the 3' stacked base for the junction. We estimate that the spherical volume available to the 5' base in the vicinity of the complimentary 3' stacked base is around 10^{-23} to 10^{-24} L, giving a concentration from 0.4 to 3.0 M (assuming that the radius of the sphere is between 5 and 10 Å). This estimate of the concentration gives a pseudo first order rate constant of 5' base pairing in the range of 10^5 to 10^6 s^{-1} .

If these simple models are reasonably accurate representations of reality, then the long 3' stacks are seen to arise naturally; the rate of 3' stack extension is greater than the rate of 5' base pairing, which is generally rate limiting.

Off-Pathway Species. The data-mining results are relatively clean in that nearly all observed ss-ds junctions appear to fall reasonably along the stack-ratchet reaction coordinate. However at low frequency we observe bulged-stacks in which a residue is excluded from the ss stack and several other species with hydrogen bonding interactions that do not fall on the reaction coordinate. These may represent off-pathway species (possible kinetic traps) that must be disrupted for helix propagation to proceed. In addition to the 31 junctions, we observe 8 putative kinetically trapped junctions (Table 1). A complete reaction coordinate, including kinetic traps, is work in progress.

DNA versus RNA. Differences in the thermodynamic effects of dangling ends on DNA versus RNA suggest that the mechanisms of helix propagation for DNA and RNA might differ. 3' Dangling ends confer less stability to DNA than to RNA duplexes.^{11,57,58} Further 5' dangling ends confer equivalent or greater stability than 3' dangling ends to DNA duplexes. Therefore one cannot propose models of DNA helix propagation from the data presented here, except to note that the mechanisms of RNA and DNA helix propagation probably differ.

Conclusions

We propose that, during RNA folding, double helices propagate via the stack-ratchet mechanism. In the stack-ratchet mechanism, stacking and pairing reactions are not simultaneous; a 3' single-strand stack leads the base pair forming reaction. One elementary reaction of the stack-ratchet mechanism is the stacking plus pairing of the 5' strand base to the stacked, unpaired 3' strand. The second elementary reaction is the stacking of this unpaired 3' strand. The presence of two elementary reactions gives rise to two relatively stable transition states. Our data-mining results and previously published thermodynamic information on the relative stabilities of 3' dangling ends on RNA double helices support the stack-ratchet mechanism of RNA helix propagation.

Acknowledgment. We acknowledge funding from the NASA Astrobiology Institute and thank Dr. Michael Schurr for helpful discussions.

Supporting Information Available: This material is available free of charge via the Internet at <http://pubs.acs.org>.

References and Notes

- (1) Bundschuh, R.; Gerland, U. *Eur. Phys. J. E* **2006**, *19*, 319.
- (2) Lee, C.; Danilowicz, C.; Coljee, V.; Prentiss, M. *Eur. Phys. J. E* **2006**, *19*, 339.
- (3) Cocco, S. *Eur. Phys. J. E* **2006**, *19*, 345.
- (4) Porschke, D. *Biophys. Chem.* **1974**, *2*, 97.
- (5) Porschke, D. *Biophys. Chem.* **1974**, *2*, 83.
- (6) Porschke, D. *Mol. Biol. Biochem. Biophys.* **1977**, *24*, 191.
- (7) Romaniuk, P. J.; Hughes, D. W.; Gregoire, R. J.; Neilson, T.; Bell, R. A. *J. Am. Chem. Soc.* **1978**, *100*, 3971.
- (8) Freier, S. M.; Alkema, D.; Sinclair, A.; Neilson, T.; Turner, D. H. *Biochemistry* **1985**, *24*, 4533.
- (9) Sugimoto, N.; Kierzek, R.; Turner, D. H. *Biochemistry* **1987**, *26*, 4554.
- (10) O'Toole, A. S.; Miller, S.; Serra, M. J. *RNA* **2005**, *11*, 512.
- (11) Ohmichi, T.; Nakano, S.; Miyoshi, D.; Sugimoto, N. *J. Am. Chem. Soc.* **2002**, *124*, 10367.
- (12) O'Toole, A. S.; Miller, S.; Haines, N.; Zink, M. C.; Serra, M. J. *Nucleic Acids Res.* **2006**, *34*, 3338.
- (13) Isaksson, J.; Chattopadhyaya, J. *Biochemistry* **2005**, *44*, 5390.
- (14) Burkard, M. E.; Kierzek, R.; Turner, D. H. *J. Mol. Biol.* **1999**, *290*, 967.
- (15) Ban, N.; Nissen, P.; Hansen, J.; Moore, P. B.; Steitz, T. A. *Science* **2000**, *289*, 905.
- (16) Klein, D. J.; Schmeing, T. M.; Moore, P. B.; Steitz, T. A. *EMBO J.* **2001**, *20*, 4214.
- (17) Selmer, M.; Dunham, C. M.; Murphy, F. V.; Weixlbaumer, A.; Petty, S.; Kelley, A. C.; Weir, J. R.; Ramakrishnan, V. *Science* **2006**, *313*, 1935.
- (18) Richardson, J. S.; Schneider, B.; Murray, L. W.; Kapral, G. J.; Immormino, R. M.; Headd, J. J.; Richardson, D. C.; Ham, D.; Hershkovits, E.; Williams, L. D.; Keating, K. S.; Pyle, A. M.; Micallef, D.; Westbrook, J.; Berman, H. M. *RNA* **2008**, *14*, 1.
- (19) Hsiao, C.; Mohan, S.; Hershkovitz, E.; Tannenbaum, A.; Williams, L. D. *Nucleic Acids Res.* **2006**, *34*, 1481.
- (20) Hershkowitz, E.; Sapiro, G.; Tannenbaum, A.; Williams, L. D. *IEEE/ACM Trans. Comput. Biol. Bioinf.* **2006**, *3*, 33.
- (21) DeLano, W. L. *The PyMOL Molecular Graphics System*; DeLano Scientific LLC: San Carlos, CA; <http://www.pymol.org>.
- (22) Lu, X. J.; Olson, W. K. *Nucleic Acids Res.* **2003**, *31*, 5108.
- (23) Acharya, P.; Acharya, S.; Cheruku, P.; Amirkhanov, N. V.; Foldesi, A.; Chattopadhyaya, J. *J. Am. Chem. Soc.* **2003**, *125*, 9948.
- (24) Markham, N. R.; Zuker, M. *Nucleic Acids Res.* **2005**, *33*, W577.
- (25) Hofacker, I. L. *Nucleic Acids Res.* **2003**, *31*, 3429.
- (26) Petersheim, M.; Turner, D. H. *Biochemistry* **1983**, *22*, 256.
- (27) Xia, T.; SantaLucia, J., Jr.; Burkard, M. E.; Kierzek, R.; Schroeder, S. J.; Jiao, X.; Cox, C.; Turner, D. H. *Biochemistry* **1998**, *37*, 14719.
- (28) Kroon, P. A.; Kreishman, G. P.; Nelson, J. H.; Chan, S. I. *Biopolymers.* **1974**, *13*, 2571.
- (29) Winkler, W.; Nahvi, A.; Breaker, R. R. *Nature* **2002**, *419*, 952.
- (30) Winkler, W. C.; Cohen-Chalamish, S.; Breaker, R. R. *Proc. Natl. Acad. Sci. U.S.A.* **2002**, *99*, 15908.
- (31) Winkler, W. C.; Nahvi, A.; Sudarsan, N.; Barrick, J. E.; Breaker, R. R. *Nat. Struct. Biol.* **2003**, *10*, 701.
- (32) Mitra, K.; Frank, J. *Annu. Rev. Biophys. Biomol. Struct.* **2006**, *35*, 299.
- (33) Spahn, C. M.; Gomez-Lorenzo, M. G.; Grassucci, R. A.; Jorgensen, R.; Andersen, G. R.; Beckmann, R.; Penczek, P. A.; Ballesta, J. P.; Frank, J. *EMBO J.* **2004**, *23*, 1008.
- (34) Frank, J.; Agrawal, R. K. *Nature.* **2000**, *406*, 318.
- (35) Allen, F. H.; Harris, S. E.; Taylor, R. *J. Comput. Aided Mol. Des.* **1996**, *10*, 247.
- (36) Taylor, R. *Acta Crystallogr., Sect D.* **2002**, *58*, 879.
- (37) Taylor, R.; Kennard, O.; Versichel, W. *J. Am. Chem. Soc.* **1983**, *105*, 5761.
- (38) Taylor, R.; Kennard, O.; Versichel, W. *J. Am. Chem. Soc.* **1984**, *106*, 244.
- (39) Bock, C. W.; Kaufman, A.; Glusker, J. P. *Inorg. Chem.* **1994**, *33*, 419.
- (40) Bock, C. W.; Markham, G. D.; Katz, A. K.; Glusker, J. P. *Theor. Chem. Acc.* **2006**, *115*, 100.
- (41) Markham, G. D.; Glusker, J. P.; Bock, C. W. *J. Phys. Chem. B* **2002**, *106*, 5118.
- (42) Burgi, H. B. *Inorg. Chem.* **1973**, *12*, 2321.
- (43) Burgi, H. B.; Dunitz, J. D.; Shefter, E. *J. Am. Chem. Soc.* **1973**, *95*, 5065.
- (44) Allen, F. H.; Mondal, R.; Pitchford, N. A.; Howard, J. A. K. *Helv. Chim. Acta* **2003**, *86*, 1129.
- (45) Sundaralingam, M.; Sekharudu, Y. C. *Science* **1989**, *244*, 1333.
- (46) Bandyopadhyay, D.; Bhattacharyya, D. *J. Biomol. Struct. Dyn.* **2003**, *21*, 447.
- (47) Hays, F. A.; Teegarden, A.; Jones, Z. J.; Harms, M.; Raup, D.; Watson, J.; Cavaliere, E.; Ho, P. S. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 7157.
- (48) Vargason, J. M.; Henderson, K.; Ho, P. S. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 7265.
- (49) Porschke, D.; Eggers, F. *Eur. J. Biochem.* **1972**, *26*, 490.
- (50) Martin, F. H.; Uhlenbeck, O. C.; Doty, P. *J. Mol. Biol.* **1971**, *57*, 201.
- (51) Uhlenbeck, O. C.; Martin, F. H.; Doty, P. *J. Mol. Biol.* **1971**, *57*, 217.
- (52) Zhang, W.; Chen, S. *J. Biophys. J.* **2006**, *90*, 765.
- (53) Barrio, J. R.; Tolman, G. L.; Leonard, N. J.; Spencer, R. D.; Weber, G. *Proc. Natl. Acad. Sci. U.S.A.* **1973**, *70*, 941.
- (54) Porschke, D. *Eur. J. Biochem.* **1973**, *39*, 117.
- (55) Liu, J. D.; Zhao, L.; Xia, T. *Biochemistry.* **2008**, *47*, 5962.
- (56) Cantor, C.; Schimmel, P. *Biophysical Chemistry*; Academic Press: New York, 1984; Vols. I–III.
- (57) Bommarito, S.; Peyret, N.; SantaLucia, J., Jr. *Nucleic Acids Res.* **2000**, *28*, 1929.
- (58) Riccielli, P. V.; Mandell, K. E.; Benight, A. S. *Nucleic Acids Res.* **2002**, *30*, 4088.
- (59) Schneider, T. D.; Stephens, R. M. *Nucleic Acids Res.* **1990**, *18*, 6097.