JACS
JOURNAL OF THE AMERICAN CHEMICAL SOCIETY

Article

pubs.acs.org/JACS

RNA Unwinding from Reweighted Pulling Simulations

Francesco Colizzi* and Giovanni Bussi*

SISSA – Scuola Internazionale Superiore di Studi Avanzati, via Bonomea 265, 34136 Trieste, Italy

Supporting Information

ABSTRACT: The forming and melting of complementary base pairs in RNA duplexes are conformational transitions required to accomplish a plethora of biological functions. Yet the dynamic steps of these transitions have not been quantitatively characterized at the molecular level. In this work, the base opening process was first enforced by atomistic pulling simulations and then analyzed with a novel reweighting scheme, which allowed the free-energy profile along any suitable reaction coordinate, for example, solvation, to be reconstructed. The systematic application of such approach to different base-pair combinations provides a molecular motion picture of helix opening, which is validated by comparison with an extensive set of experimental observations and links them to the enzyme-dependent unwinding mechanism. The RNA intrinsic dynamics disclosed in this work could rationalize the directionality observed in RNA-processing molecular machineries.

INTRODUCTION

The ability of ribonucleic acid (RNA) to adopt peculiar three-dimensional structures that mediate a variety of biological functions makes it the most versatile regulatory factor in the cell.[1] Virtually involved in all cellular processing of the genetic information, RNA is able to achieve such a functional diversity by adaptively acquiring very distinct conformations in response to specific conditions of the cellular environment.[2] Among the structural rearrangements engaged by RNA, the opening of complementary base pairs is an ubiquitous process required to accomplish a wide range of metabolic activities such as transcription, pre-mRNA splicing, ribosome biogenesis, or translation initiation.[3] In the cell, this is usually catalyzed by enzymes called RNA helicases, which have been shaped by the evolution to unwind double-stranded (ds) RNA according to its intrinsic dynamic properties.[4,5]

From a molecular standpoint, the opening and forming of individual base pairs are fundamental, yet poorly understood, events, which provide the structural framework to large-scale RNA conformational transitions and folding.[2,6−12] In this respect and related to the work presented herein, insightful investigations have been reported only for short deoxyribonucleic acids (DNAs) in the B-form helical geometry.[13−15] Using transition-path sampling, Hagan et al.[14] have fully characterized the energetics of (un)pairing for a 5′-end cytosine. However, the mechanism underlying the complete opening of the duplex has not been systematically faced nor analyzed. Moreover, differences in topology and thermodynamic parameters between B-form DNA and A-form RNA suggest that the mechanism of duplex separation might obey different rules.

Recently, combining thermodynamic information with the relative population of unpaired terminal nucleotides (dangling ends) observed in large ribosomal RNA (rRNA) crystal structures, Mohan et al.[16] have proposed that stacking and pairing reactions are not simultaneous, and that 3′-single-strand stack leads the base pairing of the 5′-strand. Nevertheless, collecting an unbiased data set of dangling-end population is not trivial, and, when viewed in the context of the full ribosomal assemblies, the single-stranded (ss) regions are seen to interact extensively with other RNA elements.[16] On top of that, because the formation and opening of base pairs is a dynamic process, both the ensemble-averaged thermodynamic properties and the detailed but static X-ray picture have to be complemented with other methods able to directly and quantitatively capture the dynamics of the investigated event. Likely, this gap will be efficiently bridged by ad hoc designed spectroscopic approaches.[2,9,12,17] For instance, femtosecond time-resolved fluorescence spectroscopy is emerging as a powerful technique for the quantitative analysis of base-stacking pattern and base motion,[18] although its applications to probe RNA dynamics are still in their infancy, and the method presents several limitations.[11] As a matter of fact, the integration of spectroscopic approaches with other powerful techniques is presently needed to gain molecular details on the RNA intrinsic dynamics. Among the possible methodological choices, atomistic simulations[19] allow any base sequence to be characterized and all of the microscopic parameters to be controlled. Additionally, when combined with state-of-the-art free-energy methods, they can provide an unparalleled perspective on the mechanism and dynamics of the biomolecular process of interest. As far as the reconstruction of free-energy profiles is concerned, the capability of estimating those profiles along any suitable reaction coordinate, without any further computational cost, would offer researchers a powerful and versatile tool

Received: November 9, 2011
Published: February 7, 2012

 © 2012 American Chemical Society dx.doi.org/10.1021/ja210531q | J. Am. Chem. Soc. 2012, 134, 5173−5179

enabling both the disclosure of intermediate states and the multifaceted analysis of complex conformational transitions.

With this spirit, here we report an in silico study elucidating the mechanism for strand separation in the RNA double helix. In particular, we used atomistic steered molecular dynamics (MD) simulations[20] to enforce the unbinding of nucleobases into the surrounding explicit water. To allow a systematic analysis of different base sequences, we devised a novel Jarzynski-equation-based reweighting scheme, which allowed the free-energy landscape to be reconstructed as a function of different reaction coordinates and the unbinding energies to be straightforwardly estimated. The computed free-energy differences are consistent with experimental observations and suggest that the strand separation mechanism occurs by a stepwise process in which the probability of unbinding of the base at the 5′ terminus is significantly higher than that at the 3′ terminus. The biological implications of these findings are discussed and related to the unwinding mechanism catalyzed by RNA processing machineries. Given the general nature of our approach, the introduced methodology can be directly applied to analyze a broad range of molecular unbinding processes.

METHODS

Throughout this Article, the following nomenclature will be consistently used to define each elementary step involving one single base and occurring during the opening of a closed base pair: unpairing is used to define the process undergone by a single base for which both Watson–Crick hydrogen bonds and stacking interactions with adjacent bases are broken; unstacking is the process breaking the stacking interactions between a dangling terminal nucleotide and its adjacent bases. The opening of a base pair is thus composed by an initial unpairing followed by an unstacking (Figure 1). In this Article, we also use the term unbinding referring to both unpairing and unstacking processes by no means of specificity. Finally, the strand with the 5′ terminal (or 3′ terminal) nucleobase being pulled is referred to as the 5′-strand (or 3′-strand).

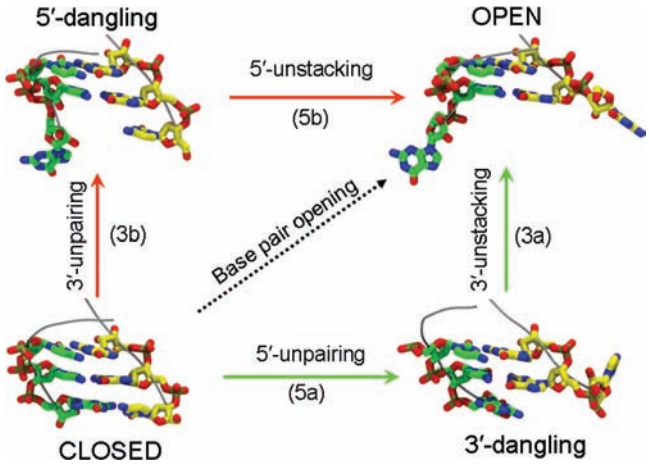

Figure 1. Elementary steps involved in the opening of a base pair. The thermodynamic cycle was used to characterize different base-pair combinations.

System Setup. We simulated the unpairing and unstacking of nucleobases at both 3′- and 5′-termini in dsRNAs of sequences 5′-CCGGGC-3′ / 3′-GGCCCG-5′ and 5′-GGCCCG-3′ / 3′-CCGGGC-5′ (Figure 2). Two sets of data can be obtained from each dsRNA, thus resulting in four systems with different combinations of Watson–Crick base pairing and stacking

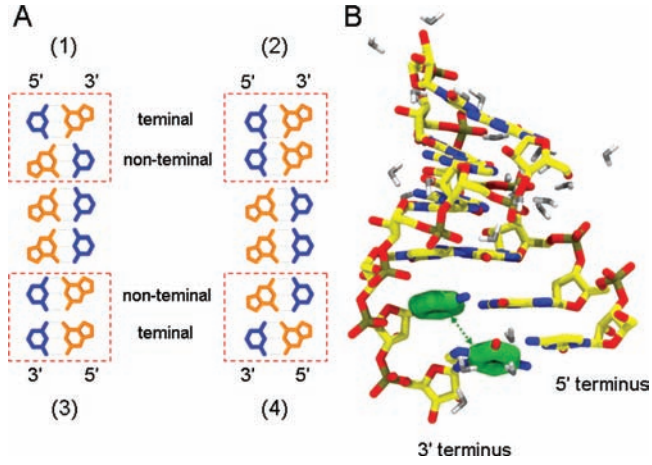

Figure 2. RNA double helix. (A) Schematic view of the combinations (red dotted boxes) of guanine (orange) and cytosine (blue) base-pairing and -stacking investigated. (B) Structural representation of the RNA duplex in water; the distance between the center of mass (green arrow) of the six-membered ring atoms (thick green sticks) of two stacked bases was used as a collective variable for the pulling simulations.

(Figure 2A). Both terminal and nonterminal base pairs (i.e., a base pair at the ss–ds RNA junction) were investigated. Nonterminal base pairs showed the same trend in relative stability observed for terminal ones and are reported in the Supporting Information. The A-form dsRNA was built using ASSEMBLE[21] and then solvated with ∼3600 water molecules, 20 Na^+, and 10 Cl^- ions, resulting in an excess salt concentration of about 0.15 M. The mobility of added ions was fairly diffusive during the simulations. After minimization and thermalization, each system (or intermediate) was then evolved for 30 ns in the isothermal–isobaric ensemble (300 K, 1 atm)[22,23] using the Amber99 force field[24] and TIP3P water.[25] Preliminary calculations carried out using the recent refinement of the Amber99 force field (parmbsc0)[26] have shown quantitatively similar results in the reconstructed free-energy profiles. This is probably due to the poor involvement of the refined α and γ dihedrals during the unbinding trajectories. Long-range electrostatic interactions were calculated with the particle mesh Ewald method.[27] Plain MD and biased steered MD trajectories were generated with GROMACS 4.0.7[28] combined with PLUMED 1.2.[29]

Pulling Simulations. The starting configurations for the pulling simulations were randomly sampled from the corresponding 30 ns-long runs. The distance between the center of mass of two stacked bases (Figure 2B) was used as pulled collective variable (CV), and thus harmonically restrained to a constant-velocity moving point, starting at a position equal to the equilibrium average of the CV and pulling it by 0.75 nm in 1.5 ns. This resulted in a biasing potential equal to $V_{bias}(q,t) = k/2[s(q(t)) - (s_0 + vt)]^2$, where k = 1200 kcal/mol/nm^2 is the spring constant of the restraint, q are the microscopic coordinates, $s(q)$ is the CV value for those coordinates, s_0 is the initial position of the restraint, v is the pulling velocity, and t is the time.

The mechanical work done during the process was obtained by integrating the force exerted on the system along the biased reaction coordinate. After collecting about 400 realizations for each nucleobase-unbinding process, the Jarzynski nonequilibrium work theorem[30] was exploited to discount the dissipated work and to reconstruct the free-energy profile as a function of the restraint distance ($s_0 + vt$). Although employing Jarzynski's equality in principle allows unbiased free-energy differences to be estimated, its direct application is limited by the number of collectable realizations as well as by the complexity of the system.[31] A typical free-energy profile is shown in Figure 3A as a function of the restraint distance. The blue plot shows how, after a steep rise, a series of alternating shoulders and local plateaus gradually brought the system to higher free-energy states. Moreover, between distances ranging from 1 to 1.2 nm (for the exemplified system), the

dx.doi.org/10.1021/ja210531q | J. Am. Chem. Soc. 2012, 134, 5173–5179

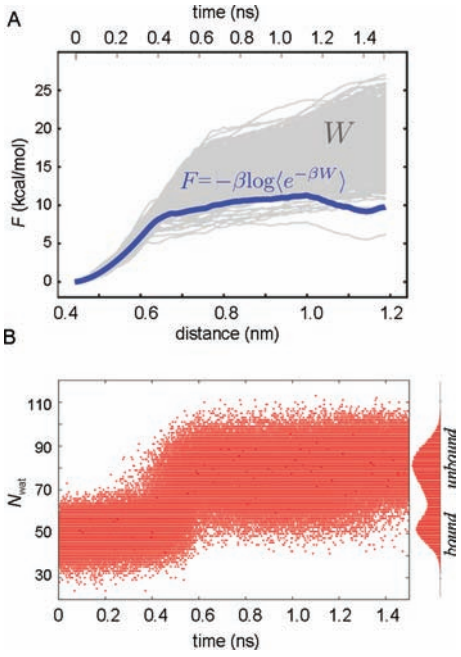

Figure 3. Typical nucleobase unbinding process, obtained by pulling along the distance between the center of mass of two stacked bases. (A) Mechanical work (W) performed (gray plot) and its exponential average (blue plot) as in Jarzynski equality, plotted as a function of the restraint position. The distance is practical for biasing the system but hardly allowed defining the bound and unbound states. (B) Number of water molecules coordinating the unbinding nucleobase (N_{wat}) as a function of time (main panel), and its probability distribution (right panel). The coordination with water is a useful metrics for identifying the bound and unbound states.

profile was strongly dominated by an outlier low-work realization, and the difference in reconstructing the free-energy profile with or without the outlier was more than 3 kcal/mol (see the Supporting Information).

Within this framework, there was no clean way to automatically detect when the nucleobase had reached the unbound configuration, and it was difficult to avoid systematic errors in the comparison of many profiles with small differences as we were interested in. To tackle this problem, we decided to analyze our simulations in terms of CVs different from the one used for the pulling. This a posteriori analysis could be done quickly, as a postprocessing, and allowed us to choose optimal CVs capable of describing in a user-independent manner all of the unbinding events.

Reweighting Scheme. To project the free-energy landscape on putative CVs, we devised a proper reweighting scheme. Whereas suitable schemes have been proposed to reweight other types of nonequilibrium simulations (e.g., ref 32), a reweighting algorithm for steered MD has not been reported. For a different purpose, Hummer and Szabo developed a method that enables the reconstruction of the free energy as a function of the pulled coordinate.[33,34] Here, we generalize this scheme so as to compute the free energy as a function of any a posteriori chosen variable.

Two different sorts of bias affect the steered MD trajectories and needed to be removed: (a) the nonequilibrium nature of the pulling and (b) the presence of artificial harmonic restraints on the pulled CV. The nonequilibrium bias is removed by noticing that the equilibrium probability $P_{eq}(q,t)$, for a restraint statically kept in its position at time t, can be obtained from the nonequilibrium one $P^{(i)}_{neq}(q,t)$ as observed in the ith trajectory exploiting a relation first reported by Crooks:[35,36]

$$P_{eq}(q,t) = \sum_i e^{-\beta[W_i(t)-F(t)]} P^{(i)}_{neq}(q,t) \tag{1}$$

where $W_i(t)$ is the work done on the ith trajectory up to time t and $\beta = 1/k_BT$ is the inverse thermal energy. Here, the free energy $F(t)$ represents the normalization factor corresponding to the instantaneous position of the moving restraint at time t. The bias of the harmonic restraint can then be removed by applying the weighted-histogram analysis method.[37] Whereas weighted histograms are traditionally used to combine independent simulations performed with different static biasing potentials, here we used it to combine snapshots obtained at different stages of the pulling, thus writing the unbiased equilibrium probability as

$$P_u(q) \propto \frac{\int_0^\tau dt P_{eq}(q,t)}{\int_0^\tau dt e^{-\beta[V(q,t)-F(t)]}} \tag{2}$$

where τ is the length of each pulling simulation. Finally, the free energy as a function of an arbitrary, a posteriori chosen CV $\bar{s}$ is defined as $F(\bar{s}) = -k_BT \log \int dq\, P_u(q)\delta(\bar{s} - \bar{s}(q))$. The scheme described so far closely resembles the one used by Hummer and Szabo.[33] However, it is conceptually different, as here the free energy can be reconstructed also with respect to a variable different from the pulled one. Thus, it potentially enables the disclosure and characterization of otherwise hidden features of the investigated process. To further simplify the data manipulation and to avoid building multidimensional histograms, with a further dependence on technical choices such as binning size, we recast our approach assigning a weight to each of the sampled configurations, in the same spirit as in ref 38. After simple manipulation, the weight can be shown to be equal to

$$w_i(t) \propto \frac{e^{-\beta[W_i(t)-F(t)]}}{\int_0^\tau dt' e^{-\beta[V(q_i(t),t')-F(t')]}} \tag{3}$$

The normalization factor for each time, $F(t)$, is then computed iteratively up to convergence as $e^{-\beta F(t)} = \sum_i \int_0^\tau dt'\ w_i(t') e^{-\beta V(q_i(t'),t)}$. Usually a few tens of iterations are enough to converge.

In summary, in our reweighting scheme, we first compute the weight of each of the configurations saved along the MD simulations from eq 3, then estimate free energies as a function of any a posteriori chosen CV as

$$F(\bar{s}) = -k_BT \log \sum_i w_i(t)\delta(\bar{s} - \bar{s}(q_i(t))) \tag{4}$$

RESULTS

Using the reweighting scheme outlined in the previous section, we were able to investigate several order parameters. Because solvent interactions are known to affect the conformational state of nucleic acids,[39] we considered the solvation of an unbinding base as an effective metrics for the progression of the underlying process. This choice allowed us to define the unbinding in a manner that was totally independent of both the terminus and the specific base, and, in our explicit-solvent simulation, could be computed as the coordination among heavy atoms of the base and water oxygens (Figure 3B). In this metrics, the bound and unbound states could be clearly and unambiguously identified and corresponded to approximately harmonic basins. Sample free energies computed as a function of the number of water molecules coordinating the unpairing base are shown in Figure 4. The free-energy profile reconstructed along such a reaction coordinate is in no way biased by the absolute number of coordinated water molecules, which is merely used to distinguish one configuration from the other and to properly collect the corresponding weights along the simulation as in eq 4. Next, to compute accurately the

dx.doi.org/10.1021/ja210531q | J. Am. Chem. Soc. 2012, 134, 5173−5179

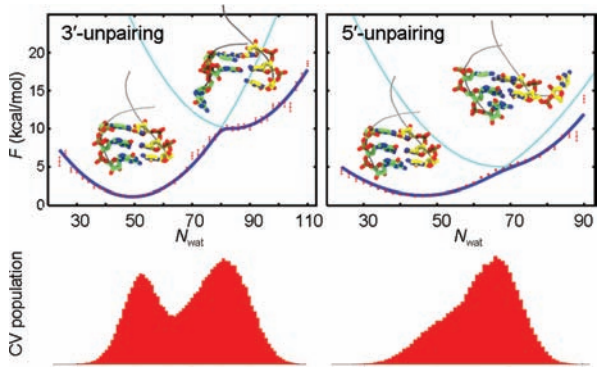

Figure 4. Reconstruction of the free-energy profile as a function of the number of water molecules (N_{wat}) surrounding the unbinding base. In the left and right panels are shown typical free-energy profiles (red dots with error bar) for the unpairing of a 3′-strand guanine and a 5′-strand cytosine, respectively. The quadratic potentials obtained from the double-well fitting are shown in light blue color, whereas their combination (eq 5) is in blue. Underneath each panel, the unnormalized population of the CV is also shown.

bound/unbound free-energy differences, we fit the free-energy profiles with the combination of two quadratic functions:[40,41]

$$e^{-\beta F(\bar{s})} = \sigma_1^{-1} e^{-((\bar{s}-\bar{s}_1)^2/2\sigma_1^2)-\beta F_1} + \sigma_2^{-1} e^{-((\bar{s}-\bar{s}_2)^2/2\sigma_2^2)-\beta F_2} \quad (5)$$

where F_1 and F_2 are the free energies of bound and unbound states. Both when the two states were clearly resolvable (e.g., Figure 4, left panels), and when the corresponding CV population was more overlapped (e.g., Figure 4, right panels), the fitting procedure that resulted was robust and poorly sensitive to outlier work realizations, thus enhancing convergence of the results (e.g., the difference in performing the fit with or without the outlier low-work realization in Figure 3A was less than 0.3 kcal/mol; see the Supporting Information). Furthermore, this approach showed very stable outcomes with respect to the choice of the details in the definition of the solvation order parameter and allowed us to compare systematically several similar situations without incurring of large statistical errors or, worst, human biases in the interpretation of the results.

Having an optimized statistical-mechanics tool able to provide free-energy differences in a flexible and automatic manner, we pursued a systematic step-by-step approach to investigate the feasibility of different opening paths for the four possible combinations of G−C base stacking and Watson−Crick pairing. The general procedure, as outlined in Figure 1, relied on two subsequent steps: first, the Watson−Crick base pair was partially opened by the unpairing of the base on either the 5′ or the 3′ terminus; second, the resulting dangling intermediate, on the 3′ or 5′ terminus respectively, was unstacked, and the base pair opening was completed.

The relative stability of putative intermediates involved in the opening of a base pair was estimated from the individual base-unbinding free energies (Table 1 and Figure 1). For all of the considered combinations, the difference in base-pair-opening free energy computed biasing the system along paths (a) and (b) in Figure 1 was lower than 1 kcal/mol. For the sake of clarity, it should be reminded that a finite number of unidirectional pulling simulations performed within a Jarzynski-like scheme are known to provide overestimates of absolute free-energy differences.[31] However, highly accurate estimates of unbinding constants were not needed to characterize the strand separation mechanism, and the free-energy differences we estimated were exploited as a quantitative tool to assess the relative stability of different configurations.

Table 1. Context-Dependent Base-Unbinding Free Energy (kcal/mol) Corresponding to the Elementary Steps Shown in Figure 1

construct		opening steps			
n		3b	5b	5a	3a
1	5′-CG.. 3′-GC..	7.6	0.9	2.6	5.6
2	5′-CC.. 3′-GG..	8.9	0.3	3.9	4.5
3	5′-GG.. 3′-CC..	7.0	2.2	4.9	3.9
4	5′-GC.. 3′-CG..	7.4	2.0	5.7	2.8

COMPARISON WITH EXPERIMENTS

Below, we discuss the results of the first and second unbinding steps (Figure 1) and compare them with crystal structures conformer distributions,[16,42] relative population of stacked/unstacked bases detected by femtosecond time-resolved fluorescence spectroscopy,[43] and thermodynamic data based on dsRNA melting experiments.[1,44,45] The consistency with experimental observations and the capability of our simulations to complement those results are highlighted.

The more general outcome arising from the comparison of free-energy differences is that the paired base on the 5′ terminus always interacted more weakly than the complementary base on the 3′ terminus (steps 5a, 3b in Table 1 and Figure 1). This could be directly related to the A-form helical geometry of RNA in which the bases at the 5′ end of a ss−ds junction are less buried into the neighboring environment and expose a wider portion of their surface to water molecules, thus facilitating fraying events. The different stability of the nucleobase on the 5′ terminus can be reflected in the probability of observing a certain type of blunt closing base pair at ss−ds junctions. In this context, the stronger interaction was estimated for the 5′-guanine in $\begin{smallmatrix}5'\text{-GC..}\\3'\text{-CG..}\end{smallmatrix}$ (construct **4**), which was ∼1.7 kcal/mol weaker than the complementary cytosine on the 3′-terminus. Accordingly, the combination $\begin{smallmatrix}5'\text{-GC..}\\3'\text{-CG..}\end{smallmatrix}$ (construct **4**) is the most abundant closing base-pair pattern observed at ss−ds junctions in large rRNA crystal structures.[16] It can be further noticed that among the dangling ends (steps 5b and 3a in Table 1 and Figure 1), the most stable ones are those on the 3′ terminus, consistent with ultrafast spectroscopy experiments, which have detected a large subpopulation of stacked conformers for a 3′-dangling fluorescent purine probe, but only a relatively small one for a 5′-dangling purine probe.[43] In particular, we found the most stable 3′-dangling end in construct **1** ($\begin{smallmatrix}5'\text{-C}\\3'\text{-GC..}\end{smallmatrix}\!\!\text{G..}$), which has also been counted as the most common dangling end pattern in rRNA crystal structures.[16] Further agreement can be found considering dsRNA optical melting experiments, which have shown that single nucleotides overhanging at 3′-ends of an RNA helix increase the stability of the duplex in a sequence-

dx.doi.org/10.1021/ja210531q | J. Am. Chem. Soc. 2012, 134, 5173−5179

dependent manner. Notably, such a stabilization has been interpreted as the capability of the 3′ dangling ends to stack over the hydrogen bonds of the closing base pair protecting them from water exchange.[46] Those 3′-dangling bases that are more likely stacked would thus provide a larger contribution to duplex stabilization. In this light, the trend that we observed in the unstacking energies of the four dangling constructs ($^{5'\text{-}C}_{3'\text{-}GC..}$ > $^{5'\text{-}C}_{3'\text{-}GG..}$ > $^{5'\text{-}G}_{3'\text{-}CC..}$ > $^{5'\text{-}G}_{3'\text{-}CG..}$) is in agreement with duplex stabilization observed in dsRNA melting experiments.[1,44,45] It should be remarked that the duplex stabilization induced by 5′-dangling ends might not reflect the stacking energy of the dangling end itself because of its small overlap with the hydrogen bonds of the closing base pair. Further discussion on the comparison of computed and experimental dangling-end stabilities can be found in the Supporting Information.

Summarizing, the unbinding of the base on the 5′-strand was, in all of the considered cases, favored over the unbinding of the complementary 3′-strand base. Whereas the relative probability of 3′- and 5′-unbinding event can be modulated by the sequence, the general trend remains unchanged.

From a structural standpoint, the unpairing could proceed through two qualitatively different paths: one in which the twisting and breaking of Watson−Crick hydrogen bonds occurred before the rupture of stacking interactions, and the other in which the unbinding followed the concerted rupture of both hydrogen bonds and stacking interactions with, in some cases, the unbinding base stacking over the dangling end of the opposite strand (Figure 5). Similar unbinding geometries have also been described in other studies.[14,47] In our simulations, these intermediate states occurred with a context-dependent frequency. For instance, along the unbinding pathway, the 5′-terminal guanine in $^{5'\text{-}GC..}_{3'\text{-}CG..}$ (construct **4**) had ∼15% of probability to stack upon the 3′-dangling cytosine of the opposite strand. Interestingly, this probability dropped to ∼3% in $^{5'\text{-}GG..}_{3'\text{-}CC..}$ (construct **3**). Such an interstrand stacking pattern (Figure 5, panel t_2), exchanging with the conventional stacking of a paired base, could account for the delayed quenching of fluorescence detected by ultrafast fluorescence spectroscopy for a construct similar to $^{5'\text{-}GC..}_{3'\text{-}CG..}$ (construct **4**).[43]

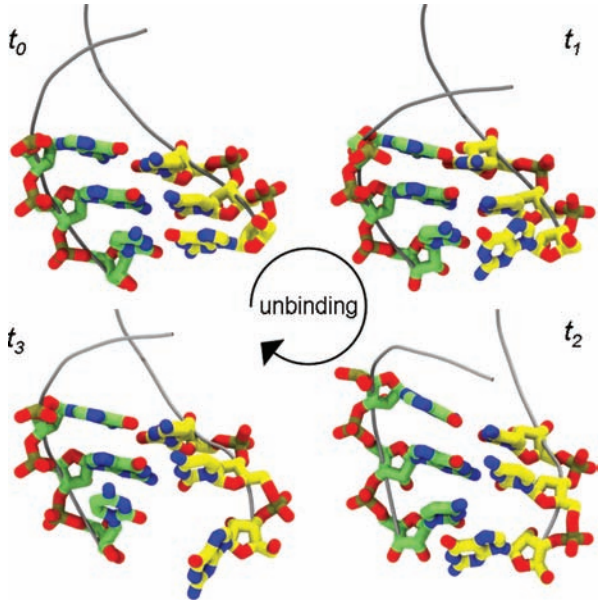

Figure 5. Snapshots sampled from the opening of a base pair. The unbinding base (here a 5′-terminal guanine in construct **4**) could transiently stack over the dangling end of the opposite strand (t_2).

As final experimental evidence to corroborate our results, Xia and co-workers[43] have reported that the dynamic behavior of a 3′-terminal purine is not affected by the presence of the opposite complementary base. Vice versa, the conformational dynamics of a 5′-terminal purine is drastically influenced by the presence of an opposite 3′-terminal pyrimidine, which would be likely stacked and potentially able to shift the population of the complementary 5′-terminal base toward a paired and stacked ensemble. Consistent with our systematic study, these data depict the formation of a stable base pair as generally driven by the stacking of the 3′-terminal base, and then by the energy gained by the system from both the stacking of the 5′-terminal base and the Watson−Crick hydrogen bonds.

BIOLOGICAL IMPLICATIONS

The motion picture of duplex separation emerging from the outcome of our simulations complements and augments with dynamic details and energetic considerations the helix propagation model based on the analysis of static 3D structures.[16] Our computations link experimental data from different fields, creating a common reading frame among them. Taken together, these results suggest that RNA unwinding occurs by a stepwise process in which the probability of unbinding of the base on the 5′ strand is significantly higher than that on the 3′ strand. What could be the biological implications of this finding?

When considering the RNA as the substrate of molecular motors such as helicases and other remodeling enzymes, the results could likely be interpreted from an evolutionary point of view, which could allow deciphering the basis of the evolutionary pressure responsible for the unwinding mechanism catalyzed by RNA-duplex processing enzymes.

The RNA unwinding catalyzed by helicases is coupled to adenosine triphosphate (ATP) binding and hydrolysis. The underlying mechanism would reasonably minimize the use of ATP, especially in a low-nutrient environment. Provided that the intrinsic RNA dynamics implies that at ss−ds junctions the unbinding of the 5′-strand base is favored over the unbinding of the complementary 3′-strand base, an enzymatic unwinding model would include a mechanism in which the separation of the two complementary strands is accomplished by acting on the weakest portion, that is, the 5′-strand base. Thus, an ancestral enzyme using the 3′-strand as running track rail (with 3′ to 5′-directionality) without perturbing its conformation and causing the displacement of the 5′-strand by mechanical exclusion could satisfy some energy-saving requirements.

The viral RNA helicase NS3 of hepatitis C virus, which is a prototypical DEx(H/D) RNA helicases essential for viral replication, could satisfy the above-mentioned requirements.[4,5] NS3 is a potentially relevant drug target and has been structurally and functionally characterized in various contexts.[48−51] It unwinds duplexes by first loading onto a single-stranded 3′-terminus region and then processively translocating with 3′ to 5′ directionality along this loading strand, thereby peeling off the complementary 5′-strand bases. In particular, the

dx.doi.org/10.1021/ja210531q | J. Am. Chem. Soc. 2012, 134, 5173−5179

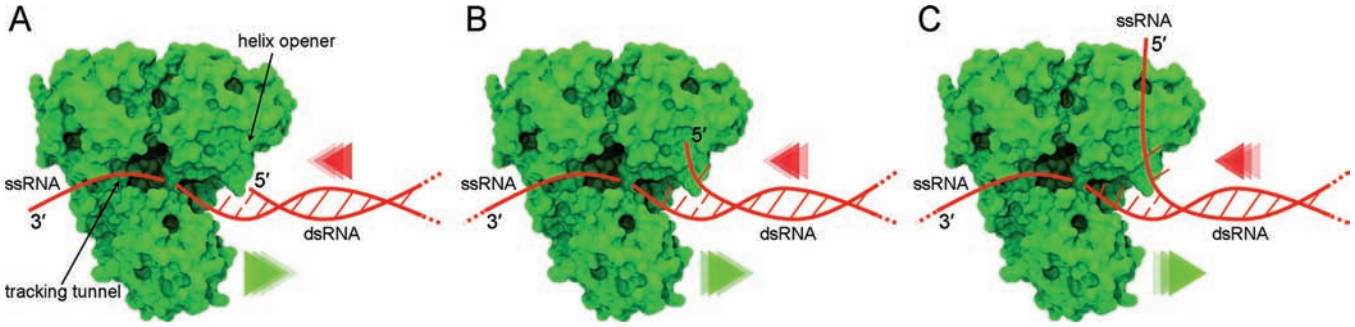

Figure 6. Model for RNA unwinding catalyzed by NS3 helicase. (A) The single-stranded 3′ terminus is loaded into the tracking tunnel. (B) The 5′ terminus is mechanically displaced by the helix opener. (C) The helicase proceeds with 3′ to 5′ directionality, displacing the 5′-strand.

3′-strand would migrate through a tracking ssRNA tunnel running within the protein, whereas the complementary 5′-strand is forced toward the back of the protein by the "helix opener" hairpin (Figure 6).[52,53] In light of the free-energy calculations discussed above, it could be suggested that this mechanism has been optimized according to the intrinsic RNA unwinding dynamics disclosed in this work.

Arguably, the processing machineries are being constantly shaped by the evolutionary pressure of a plethora of (often unknown) factors contributing to the optimization of metabolism in the whole living system, rather than to the local biochemical process. As a consequence, the preference for a well-defined RNA processing directionality cannot be ubiquitously observed.[4,5]

We speculate that all of the biochemical processes involving RNA in which directionality plays a role (e.g., transcription) could be related to the energetics of RNA double helix forming and fraying discussed in this Article.

CONCLUSIONS

This study lays down the basis for the molecular-level understanding of intrinsic RNA dynamics and its role in function. The asymmetric behavior of the 3′- and 5′-strand could be responsible for the directionality observed in RNA processing. From a computational perspective, the approach we introduced can be generalized to analyze any kind of (un)binding event. Indeed, it allowed the free-energy landscape to be reconstructed along different reaction coordinates and the unbinding energies to be easily computed in an automatic and user-independent manner, therefore removing statistical and human biases. We foresee the application of our approach to a wider range of molecular systems, including the typical ligand-target complex faced in drug discovery.[54]

ASSOCIATED CONTENT

Supporting Information

Details of methodology and computations. Unbinding free-energy profiles for both terminal and nonterminal base pairs. Further discussion on the comparison of computed and experimental dangling-end stabilities. This material is available free of charge via the Internet at http://pubs.acs.org.

AUTHOR INFORMATION

Corresponding Author

colizzi@sissa.it; bussi@sissa.it

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

We thank Francesco Di Palma, Rolando Hong, and Vittorio Limongelli for critically reading this manuscript and an anonymous referee for several useful suggestions. We acknowledge the CINECA Award no. HP10BLIT9Z, 2011 for the availability of high performance computing resources and MIUR grant "FIRB - Futuro in Ricerca" no. RBFR102PY5 for funding. Developed software is available on request.

REFERENCES

(1) Bloomfield, V. A.; Crothers, D. M.; Tinoco, I. J. *Nucleic Acids: Structures, Properties, and Functions*; University Science Books: Sausalito, CA, 2000.
(2) Haller, A.; Soulière, M. F.; Micura, R. *Acc. Chem. Res.* **2011**, *44*, 1339−1348.
(3) Rocak, S.; Linder, P. *Nat. Rev. Mol. Cell Biol.* **2004**, *5*, 232−241.
(4) Jankowsky, E.; Fairman, M. *Curr. Opin. Struct. Biol.* **2007**, *17*, 316−324.
(5) Pyle, A. M. *Annu. Rev. Biophys.* **2008**, *37*, 317−336, PMID: 18573084.
(6) Norberg, J.; Nilsson, L. *Acc. Chem. Res.* **2002**, *35*, 465−472.
(7) Zhuang, Z.; Jaeger, L.; Shea, J. *Nucleic Acids Res.* **2007**, *35*, 6995.
(8) Li, P.; Vieregg, J.; Tinoco, I. Jr. *Annu. Rev. Biochem.* **2008**, *77*, 77−100.
(9) Al-Hashimi, H.; Walter, N. *Curr. Opin. Struct. Biol.* **2008**, *18*, 321−329.
(10) Orozco, M.; Noy, A.; Pérez, A. *Curr. Opin. Struct. Biol.* **2008**, *18*, 185−193.
(11) Xia, T. *Curr. Opin. Chem. Biol.* **2008**, *12*, 604−611.
(12) Rinnenthal, J.; Buck, J.; Ferner, J.; Wacker, A.; Fürtig, B.; Schwalbe, H. *Acc. Chem. Res.* **2011**, *44*, 1292−1301.
(13) Várnai, P.; Lavery, R. *J. Am. Chem. Soc.* **2002**, *124*, 7272−7273.
(14) Hagan, M. F.; Dinner, A. R.; Chandler, D.; Chakraborty, A. K. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 13922−13927.
(15) Giudice, E.; Lavery, R. *J. Am. Chem. Soc.* **2003**, *125*, 4998−4999.
(16) Mohan, S.; Hsiao, C.; VanDeusen, H.; Gallagher, R.; Krohn, E.; Kalahar, B.; Wartell, R. M.; Williams, L. D. *J. Phys. Chem. B* **2009**, *113*, 2614−2623.
(17) Lee, M.-K.; Gal, M.; Frydman, L.; Varani, G. *Proc. Natl. Acad. Sci. U.S.A.* **2010**, *107*, 9192−9197.
(18) Zhao, L.; Xia, T. *Methods* **2009**, *49*, 128−135.
(19) Karplus, M.; McCammon, J. A. *Nat. Struct. Biol.* **2002**, *9*, 646−652.
(20) Sotomayor, M.; Schulten, K. *Science* **2007**, *316*, 1144−1148.
(21) Jossinet, F.; Ludwig, T. E.; Westhof, E. *Bioinformatics* **2010**, *26*, 2057−2059.
(22) Bussi, G.; Donadio, D.; Parrinello, M. *J. Chem. Phys.* **2007**, *126*, 014101.
(23) Parrinello, M.; Rahman, A. *J. Appl. Phys.* **1981**, *52*, 7182−7190.
(24) Wang, J.; Cieplak, P.; Kollman, P. A. *J. Comput. Chem.* **2000**, *21*, 1049−1074.

(25) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. F.; Impey, R. W.; Klein, M. L. *J. Chem. Phys.* **1983**, *79*, 926.
(26) Pérez, A.; Marchán, I.; Svozil, D.; Sponer, J.; Cheatham, T. III; Laughton, C.; Orozco, M. *Biophys. J.* **2007**, *92*, 3817−3829.
(27) Darden, T.; York, D.; Pedersen, L. *J. Chem. Phys.* **1993**, *98*, 10089−10092.
(28) Hess, B.; Kutzner, C.; Van Der Spoel, D.; Lindahl, E. *J. Chem. Theory Comput.* **2008**, *4*, 435−447.
(29) Bonomi, M.; Branduardi, D.; Bussi, G.; Camilloni, C.; Provasi, D.; Raiteri, P.; Donadio, D.; Marinelli, F.; Pietrucci, F.; Broglia, R.; Parrinello, M. *Comput. Phys. Commun.* **2009**, *180*, 1961−1972.
(30) Jarzynski, C. *Phys. Rev. Lett.* **1997**, *78*, 2690.
(31) Gore, J.; Ritort, F.; Bustamante, C. *Proc. Natl. Acad. Sci. U.S.A.* **2003**, *100*, 12564−12569.
(32) Bonomi, M.; Barducci, A.; Parrinello, M. *J. Comput. Chem.* **2009**, *30*, 1615−1621.
(33) Hummer, G.; Szabo, A. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 3658−3661.
(34) Gupta, A.; Vincent, A.; Neupane, K.; Yu, H.; Wang, F.; Woodside, M. *Nat. Phys.* **2011**, *7*, 631−634.
(35) Crooks, G. E. *Phys. Rev. E* **2000**, *61*, 2361−2366.
(36) Williams, S. R.; Evans, D. J. *Phys. Rev. Lett.* **2010**, *105*, 110601.
(37) Kumar, S.; Rosenberg, J. M.; Bouzida, D.; Swendsen, R. H.; Kollman, P. A. *J. Comput. Chem.* **1992**, *13*, 1011−1021.
(38) Souaille, M.; Roux, B. *Comput. Phys. Commun.* **2001**, *135*, 40−57.
(39) Cantor, R.; Schimmel, P. *Biophysical Chemistry*; W. H. Freeman: San Francisco, CA, 1980.
(40) Hummer, G.; Szabo, A. *Acc. Chem. Res.* **2005**, *38*, 504−513.
(41) Hummer, G.; Szabo, A. *Proc. Natl. Acad. Sci. U.S.A.* **2010**, *107*, 21441−21446.
(42) Mohan, S.; Hsiao, C.; Bowman, J. C.; Wartell, R.; Williams, L. D. *J. Am. Chem. Soc.* **2010**, *132*, 12679−12689.
(43) Liu, J. D.; Zhao, L.; Xia, T. *Biochemistry* **2008**, *47*, 5962−5975.
(44) Sugimoto, N.; Kierzek, R.; Turner, D. H. *Biochemistry* **1987**, *26*, 4554−4558.
(45) Turner, D. H.; Sugimoto, N.; Freier, S. M. *Annu. Rev. Biophys. Biophys. Chem.* **1988**, *17*, 167−192.
(46) Isaksson, J.; Chattopadhyaya, J. *Biochemistry* **2005**, *44*, 5390−5401.
(47) Pohorille, A.; Ross, W.; Tinoco, I. *Int. J. Supercomputer Appl. High Performance Computing* **1990**, *4*, 81−96.
(48) Dumont, S.; Cheng, W.; Serebrov, V.; Beran, R. K.; Tinoco, I. Jr.; Pyle, A. M.; Bustamante, C. *Nature* **2006**, *439*, 105.
(49) Büttner, K.; Nehring, S.; Hopfner, K.-P. *Nat. Struct. Mol. Biol.* **2007**, *14*, 647−652.
(50) Gu, M.; Rice, C. *Proc. Natl. Acad. Sci. U.S.A.* **2010**, *107*, 521.
(51) Raney, K. D.; Sharma, S. D.; Moustafa, I. M.; Cameron, C. E. *J. Biol. Chem.* **2010**, *285*, 22725−22731.
(52) Luo, D.; Xu, T.; Watson, R. P.; Scherer-Becker, D.; Sampath, A.; Jahnke, W.; Yeong, S. S.; Wang, C. H.; Lim, S. P.; Strongin, A.; Vasudevan, S. G.; Lescar, J. *EMBO J.* **2008**, *27*, 3209−3219.
(53) Serebrov, V.; Beran, R. K. F.; Pyle, A. M. *J. Biol. Chem.* **2009**, *284*, 2512−2521.
(54) Jorgensen, W. L. *Nature* **2010**, *466*, 42−43.