bioRxiv preprint doi: https://doi.org/10.1101/2021.08.31.458432; this version posted September 1, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

RNA secondary structure regulates fragments' adsorption onto flat substrates

Simón Poblete,[*,†,‡] Anže Božič,[¶] Matej Kanduč,[¶] Rudolf Podgornik,[§,‖,⊥] and Horacio V. Guzman[*,¶]

†*Instituto de Ciencias Físicas y Matemáticas, Universidad Austral de Chile, Valdivia 5091000, Chile*

‡*Computational Biology Lab, Fundación Ciencia & Vida, Santiago 7780272, Chile*

¶*Department of Theoretical Physics, Jožef Stefan Institute, SI-1000 Ljubljana, Slovenia*

§*School of Physical Sciences and Kavli Institute for Theoretical Sciences, University of Chinese Academy of Sciences, Beijing 100049, China and Institute of Physics, Chinese Academy of Sciences, Beijing 100190, China*

‖*Wenzhou Institute of the University of Chinese Academy of Sciences, Wenzhou, Zhejiang 325000, China*

⊥*Also affiliated with Department of Physics, Faculty of Mathematics and Physics, University of Ljubljana, SI-1000 Ljubljana, Slovenia*

E-mail: simon.poblete@uach.cl; horacio.guzman@ijs.si

Abstract

RNA is a functionally rich molecule with multilevel, hierarchical structures whose role in the adsorption to molecular substrates is only beginning to be elucidated. Here, we introduce a multiscale simulation approach that combines a tractable coarse-grained RNA structural model with an interaction potential of a structureless flat adsorbing

bioRxiv preprint doi: https://doi.org/10.1101/2021.08.31.458432; this version posted September 1, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

substrate. Within this approach, we study the specific role of stem-hairpin and multi-branch RNA secondary structure motifs on its adsorption phenomenology. Our findings identify a dual regime of adsorption for short RNA fragments with and without secondary structure, and underline the adsorption efficiency in both cases as a function of the surface interaction strength. The observed behavior results from an interplay between the number of contacts formed at the surface and the conformational entropy of the RNA molecule. The adsorption phenomenology of RNA seems to persist also for much longer RNAs as qualitatively observed by comparing the trends of our simulations with a theoretical approach based on an ideal semiflexible polymer chain.

Abbreviations

RNA, PMF, STMV, wiss, noss, 3D

Keywords

RNA secondary structure, RNA adsorption, polyelectrolytes, free energy, RNA tertiary structure

Introduction

In the past decades, the ability of ribonucleic acid (RNA) to influence biological processes occurring inside living cells[1] has generated a lot of interest and has provided a driving force for innovative strategies in ambitious nanomedical applications.[2–6] RNA is a flexible polyelectrolyte with highly adaptable conformations, and with self-associating base pairs creates a variety of complex structural motifs, which sets it apart from the commonly more rigid and structurally much less diverse DNA.[7–9] Unlike proteins, RNA acquires its structure in a hierarchical way, first assuming a secondary structure—a pattern of base pairs—followed

bioRxiv preprint doi: https://doi.org/10.1101/2021.08.31.458432; this version posted September 1, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

by the formation of a three dimensional tertiary structure.[10–15] While RNA differs fundamentally from DNA with its pervasive, stable double-stranded form, its self-association bears some similarity with protein folding, though the structural motifs present in RNA tend to be in general much softer and less globular.[16] This conformational softness furthermore implies that RNA might structurally respond to the presence of interactions with other macromolecular substrates sharing its biological environment. An important mode of these interactions is related to RNA adsorption to proteinaceous substrates, such as the capsid shells of viruses, where the RNA-capsid interactions are important for the efficiency of virion assembly and nanoparticle stability.[17–19] In this context, the fundamental question is whether the self-assembled RNA conformation is modified as a result of the adsorption or, equivalently, whether different RNA structural motifs modify its adsorption phenomenology.

From a fundamental polymer theory point of view,[20–23] molecular simulations have offered important insights into the adsorption of semiflexible macromolecules to a molecular substrate,[24] while the adsorption of macromolecules with either annealed or quenched internal structure onto a molecular substrate remains much less understood. The latter problem is particularly relevant in the context of RNA-virus and RNA-nanoparticle assembly phenomena,[25–32] where the soft, malleable RNA structure can respond to the adsorption process. The shortage of theoretical conceptualization and prediction makes it difficult to convert the observed RNA adsorption phenomenology into a robust parametrization of the underlying adsorption interaction potential and consequently modify and/or control the RNA-substrate interactions. Such insight into the adsorption characteristics would be especially valuable for the optimization and control of RNA assembly into carrier vesicles or virus-like nanoparticles for efficient RNA-cargo delivery,[33–36] which could potentially speed up high-throughput RNA nanocarrier fabrication for applications in nanomedicine.

The lack of a solid understanding of the connection between self-assembled structures of biopolymers such as RNA,[37,38] induced by specific internucleotide interactions, and their modification as a result of the adsorption process, induced by less specific interactions with

bioRxiv preprint doi: https://doi.org/10.1101/2021.08.31.458432; this version posted September 1, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

the adsorbing substrate, is a challenge whose general aspects we aim to address in this work. We have performed an extensive study of RNA-substrate interaction using a tractable multiscale[39–41] model to understand the adsorption mechanisms of RNA in proximity to a flat, featureless model of an adsorbing substrate. In order to elucidate the role of the secondary structure on the adsorption mechanisms of RNA, we performed coarse-grained molecular simulations of RNA fragments extracted from the Satellite Tobacco Mosaic Virus (STMV) genome.[42] One particular question we address is the role of different soft RNA secondary structure motifs in the adsorption phenomenology and the connection between the inherent RNA structure and the structure imposed by the adsorption itself. To this end, we compare the adsorption phenomenology of RNA fragments in two distinct configurations, with and without secondary structure. The model interaction potential between the adsorbing substrate and the proximal structured or unstructured RNA molecule is allowed to vary in a noninvasive manner so that the original secondary structures, if they exist, remain fixed.

We identify a transition between two interaction regimes for structured and unstructured RNA as the attractive substrate strength is varied. For structured RNA (with secondary structure), base-paired regions are preferentially adsorbed at lower surface interaction strengths when compared to the unstructured RNA. For unstructured RNA (without secondary structure), the opposite is true, and they are preferentially adsorbed at larger adsorption strengths compared to structured RNA. While it is difficult to scale up our simulations to significantly longer RNA sequences, we do simulate the adsorption behavior for different RNA sizes and compare their adsorption behavior to the ground state behavior of ideal polymer chains with the same adsorption potential. Finally, we discuss the possible role that the different adsorption regimes could play in the nanomedicine applications that can be complemented with our simulations.

bioRxiv preprint doi: https://doi.org/10.1101/2021.08.31.458432; this version posted September 1, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

System and Model Description

RNA fragments

Our study focuses on three short RNA fragments with different basic secondary structure motifs (shown in Figure 1): a short hairpin (S1), a large hairpin including several bulges (S2), and a two-hairpin multibranch structure of a similar length (S3). All three RNA fragments were extracted from previous experimental studies on the genome of STMV, whose secondary structures has been determined by SHAPE chemical probing method.[43,44] It is important to remark that the three fragments of this work represent archetypal structures of the entire STMV genome, as previously obtained by chemical probing methods.[42] Our coarse-grained model generates a 3D representation of the three RNA fragments and supports secondary structure restraints based on primary sequence and interactions deconvolved from an X-ray structure database (see Methods for details).

In order to explore the role of the secondary structure in RNA adsorption to a model substrate, we consider representative structured ("wiss"—with secondary structure) and unstructured ("noss"—no secondary structure) versions of the first two fragments, S1 and S2, while the third fragment complements the set by introducing a multibranch secondary structure (as illustrated in Figure 1). We did not consider the unstructured counterpart of the third fragment, as it is of a similar length to the second one (S2).

Model RNA-substrate interaction

We aim to investigate the general role of RNA secondary structure in adsorption processes. To that purpose, we model the substrate as a flat, featureless surface. Excluding topographical and molecular features of the substrate, which could possibly intervene in the interpretation of the results, allows us to isolate the influence of the RNA secondary structure in its adsorption. We model the attraction of the RNA phosphate groups to the adsorbing substrate by a Debye-Hückel-like interaction potential, which can be rationalized as stemming

bioRxiv preprint doi: https://doi.org/10.1101/2021.08.31.458432; this version posted September 1, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

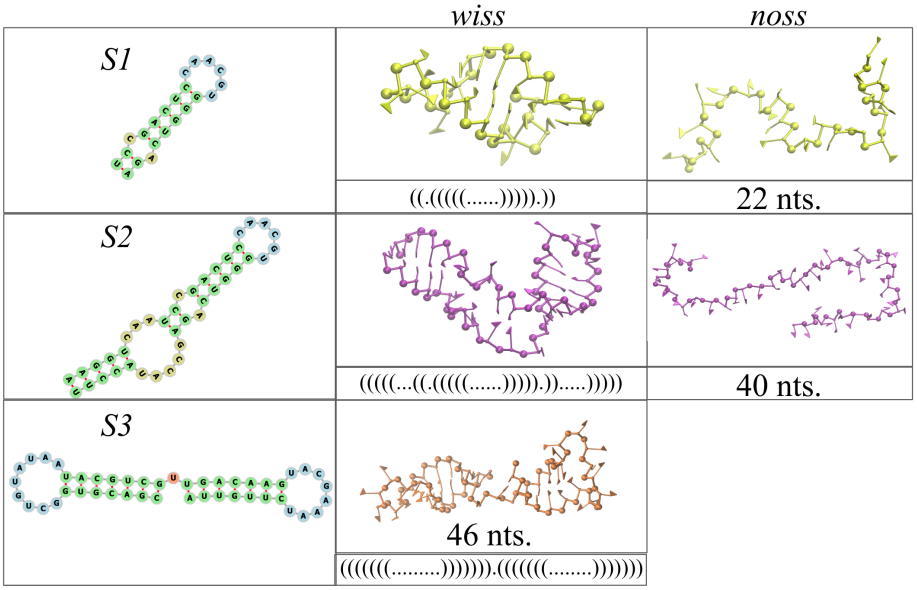

Figure 1: Secondary structure (left column), 3D model with secondary structure (middle column), and 3D model without secondary structure (right column) for the three RNA fragments used in this study (S1, S2, and S3).

from the electrostatic interactions between the dissociated RNA phosphates and the substrate charges.[45] Combined with a generic short-range repulsive term, the surface potential acting on the RNA then assumes the form

$$U_{\text{surf}}(z) = \varepsilon_{\text{r}} \left(\frac{\lambda_{\text{r}}}{z}\right)^9 - \varepsilon_{\text{surf}} \exp\left(-\frac{z}{\lambda_{\text{D}}}\right), \tag{1}$$

where $\varepsilon_{\text{r}} = 1\ k_{\text{B}}T$, $\lambda_{\text{r}} = 0.1$ nm is the distance of activation of the repulsive Lennard-Jones term (well below the size of the RNA phosphates ≈ 0.3 nm, as defined in our model) and $\lambda_{\text{D}} = 1$ nm, considering RNA under typical physiological conditions as described in Refs. 46,47. The strength of the attractive potential, $\varepsilon_{\text{surf}}$, is varied in the range between 0.44 $k_{\text{B}}T$ and 1.78 $k_{\text{B}}T$. For comparison, we also considered the scenario in which the RNA interacts with the substrate via the Mie 9-3 potential.[48] The latter Lennard-Jones-based potential can be rationalized as stemming only from the van der Waals interactions. The results are reported in the Supplementary Information.

Temperature and structural restraints of the system are chosen in such that they do

bioRxiv preprint doi: https://doi.org/10.1101/2021.08.31.458432; this version posted September 1, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

not destroy the secondary structure imposed, corresponding to a scenario below the melting temperature (for details, see the Supplementary Information). Figure 2 shows a schematic representation of the most prominent features of a simple RNA hairpin with (right) and without (left) the secondary structure, together with the action of the surface interaction potential and the center-of-mass coordinate d_{CM}.

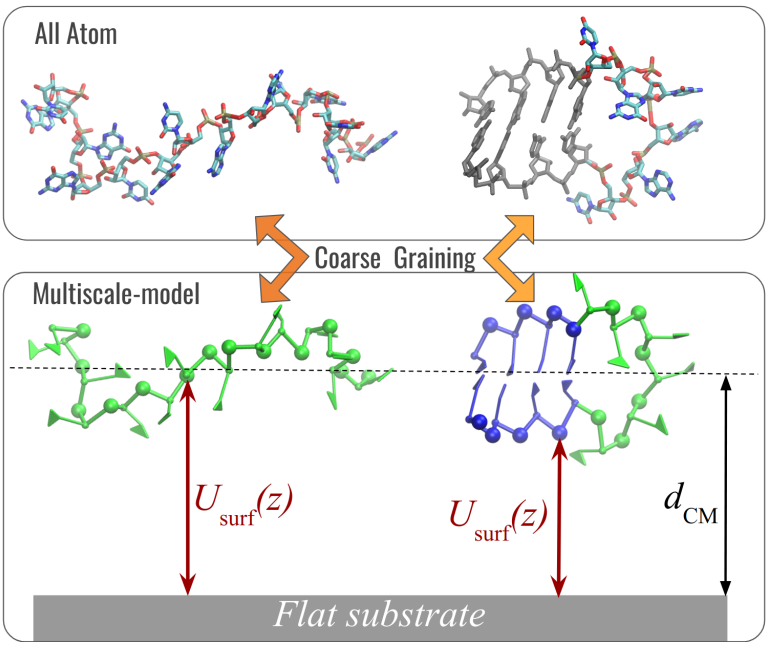

Figure 2: Two representative snapshots of a structured hairpin RNA fragment (right) and an unstructured (left) in All Atom (top) and multiscale representation (bottom). Coarse-grained nucleotides are comprised of five beads: a triangle in the bases, made of three sites, a bead representing the sugar, and a bead representing the phosphate. RNA phosphates are subjected to the RNA-substrate interaction potential U_{surf}, Equation 1. The schematic hairpin shows the base-paired nucleotides in blue and the unpaired ones in green. All-atom representation is sketched here as part of the illustration of the coarse-graining within the multiscale model.

bioRxiv preprint doi: https://doi.org/10.1101/2021.08.31.458432; this version posted September 1, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Results and Discussion

Small hairpin (RNA fragment S1)

The first RNA motif we address is a hairpin with a small bulge, comprising 22 nucleotides with a total of 7 base pairs, as shown in Figure 1. The behavior of the calculated potential of mean force (PMF) as a function of the distance of the RNA center-of-mass from the substrate is plotted in Figure 3. The PMF was calculated by averaging different conformations of the flexible regions, as well as orientations of the base-paired fragments; see also Supplementary Information. Figure 3 shows the results for both the structured (A) and the unstructured (B) S1 fragments. The PMF has the form of an attractive well, with a longer range for the unstructured fragment, and with a minimum whose location and depth depend on the strength of the RNA-substrate attraction (controlled by $\varepsilon_{\mathrm{surf}}$; see Figure 3(B)). Given the restraints of its secondary structure, the S1-wiss fragment cannot get closer than 0.9 nm from the substrate, while the single-stranded fragment can deform and adsorb more efficiently if the attraction to the substrate is strong enough.

Figure 4(A) shows adsorption free energy F_{m} (defined as the minimum of PMF) as a function of the surface attraction strength $\varepsilon_{\mathrm{surf}}$, for both structured and unstructured RNA fragments. Two regimes are immediately identifiable: the first for $\varepsilon_{\mathrm{surf}} < 0.89 k_{\mathrm{B}}T$, where the fragment with secondary structure adsorbs more strongly than the unstructured one (regime I), while for $\varepsilon_{\mathrm{surf}} > 0.89 k_{\mathrm{B}}T$, the unstructured fragment is the one exhibiting a stronger adsorption free energy (regime II).

An interesting question is how general is the existence of the two adsorption regimes. To provide a rough answer, we performed simulations using a different type of attractive potential, namely, a Lennard-Jones-based potential for a planar wall (Mie 9-3 potential). The outcomes of this much shorter-ranged potential are qualitatively the same as for the electrostatically-driven, longer-ranged Debye-Hückel potential (see Figures S1 and S2 in Supplementary Information). This suggests that the existence of the two adsorption regimes

bioRxiv preprint doi: https://doi.org/10.1101/2021.08.31.458432; this version posted September 1, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

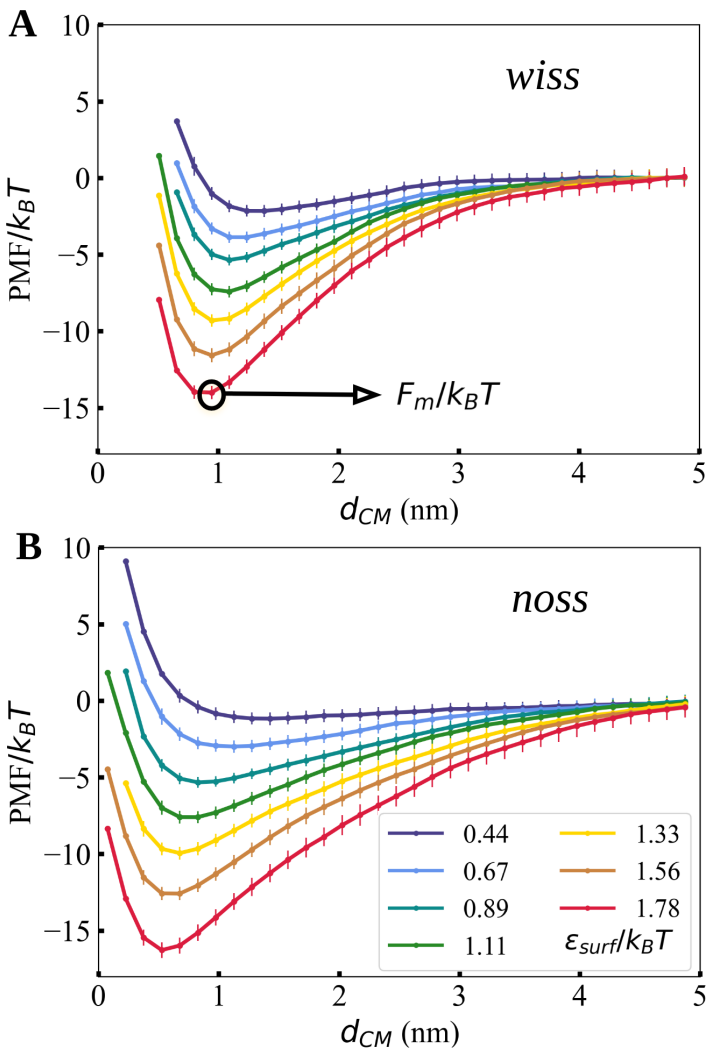

Figure 3: Potential of mean force as a function of the distance between the substrate and the center-of-mass of RNA fragment S1 **(A)** with and **(B)** without the secondary structure, shown for representative values of the attraction strength $\varepsilon_{\mathrm{surf}}$ in normalized units.

bioRxiv preprint doi: https://doi.org/10.1101/2021.08.31.458432; this version posted September 1, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

is not specific to the Debye-Hückel potential but may be considered a more general phenomenon, common to other substrate architectures as well.

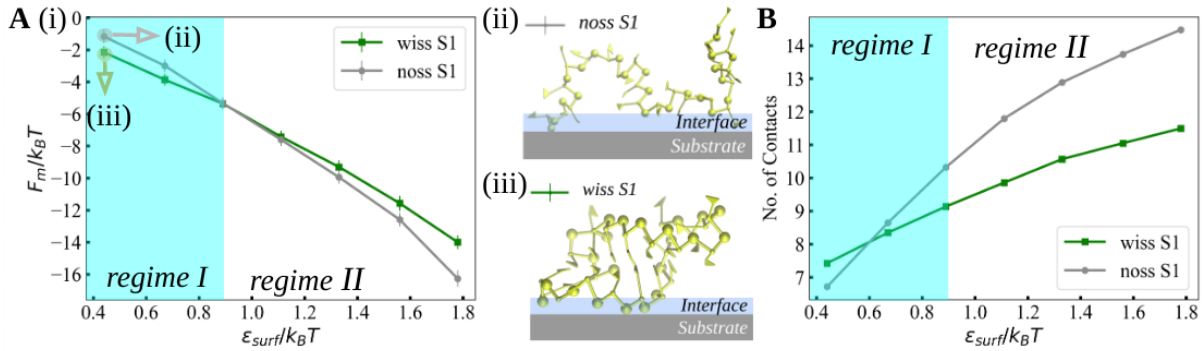

Figure 4: **(A)** Adsorption free energy as a function of the substrate attraction strength, $\varepsilon_{\text{surf}}/k_{\text{B}}T$, for RNA fragments S1-wiss and S1-noss. For a weak attraction of $\varepsilon_{\text{surf}}/k_{\text{B}}T = 0.44$, we show simulation snapshots of the (ii) unstructured and (iii) structured fragments. The illustrated interface layer is a schematic definition of a distance slightly thicker than the diameter of phosphates. **(B)** Number of contacts as a function of $\varepsilon_{\text{surf}}/k_{\text{B}}T$ for RNA fragments S1-wiss and S1-noss, calculated via Equation 2.

We quantify the number of contacts of RNA with the surface based on the normalized energy of the RNA-substrate interaction:

$$N_{\text{c}} = \frac{1}{\varepsilon_{\text{min}}} \int_0^\infty \rho(z)\, U_{\text{surf,a}}(z)\, \mathrm{d}z, \tag{2}$$

where $U_{\text{surf,a}}(z) = \min\{0, U_{\text{surf}}(z)\}$, ε_{min} is the minimum of the surface potential, and $\rho(z)$ is the monomer distribution for bound conformations normalized as $\int_0^\infty \rho(z)\, \mathrm{d}z = 1$ (for details and description, see Methods and Figure S3 in Supplementary Information). Equation 2 gives the exact contact fraction for a square-well surface potential, which is well-studied in the literature.[23]

As shown in Figure 4(B), on weakly attractive surfaces, the unstructured fragment forms fewer contacts than the structured one. This behavior changes at $\varepsilon_{\text{surf}} = 0.6\ k_{\text{B}}T$, which is below the point where the adsorption free energies F_{m} coincide (Figure 4(A)). Nevertheless, this behavior suggests that the adsorption free energy is influenced by the number of contacts formed with the surface and also by the conformation entropy of the RNA molecules. At

bioRxiv preprint doi: https://doi.org/10.1101/2021.08.31.458432; this version posted September 1, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

high interaction strengths, *i.e.*, $\varepsilon_{\rm surf} > 0.6\ k_{\rm B}T$, the unstructured fragment creates more contacts with the substrate than the structured one (Figure 4(B)). This result implies that the secondary structure of the fragment controls the number of contacts that can be made with the substrate. Furthermore, it suggests that the unstructured fragment can exhibit a saturated adsorption with all monomers being in contact with the substrate, like an RNA "landing pad." Note also that in our simulations the secondary structure, if it exists, remains fixed. Monomer distributions for RNA fragments S2 and and S3 are shown in Figure S4 in Supplementary Information, and the number of contacts for the two fragments are shown in Figure S5 in Supplementary Information. For fragment S2, the intersection between the wiss and noss occurs at slightly higher attraction strength than for S1, namely at $\varepsilon_{\rm surf}/k_{\rm B}T = 0.7$. The exact value of the intersection point depends on the fraction of structured domains. Moreover, hairpins, internal loops or junctions might contribute in a different manner to the phosphate distribution from the surface and the contact fraction.

We obtain further insight into the structural configurations of the RNA molecules by looking at the parallel and perpendicular contributions to the radius of gyration (defined in Supplementary Information). In Figure 5(A), the ratio between the normal ($\langle R^2_{\rm g\perp}\rangle$) and parallel ($\langle R^2_{\rm g\parallel}\rangle$) contributions to the radius of gyration of fragment S1 in contact with the surface is shown to decrease monotonically as the substrate attraction gets stronger. A slower decrease in $\langle R^2_{\rm g\perp}\rangle/\langle R^2_{\rm g\parallel}\rangle$ is observed in the structurally constrained fragment given its flexible loop, while for unstructured RNA the values fall well below 0.32, in agreement with theoretical predictions for semiflexible polymer chains.[23] Representative snapshots of fragment S1 are shown as a side view in Figures 5(B) and (C) for the structured and unstructured cases, respectively. Whereby, we also observe more flattened conformations for the unstructured fragment and for stronger substrate attractions, which is a consequence of an increased formation of contacts in the unstructured RNA fragments and their consequent entropy loss. Figures 5(D) and (E) show the top view of the same fragments, with a discernible extension in the xy-plane for the unstructured fragment and a reorganization of

bioRxiv preprint doi: https://doi.org/10.1101/2021.08.31.458432; this version posted September 1, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

structural constrains for the structured fragment. The flattening of the unstructured RNA fragments reflects also in the decrease of the ratio between the normal and parallel radii of gyration with increasing surface attraction strength (shown in Figure 5(A)), and the fact that the parallel contribution is less sensitive to ϵ_{surf} than its normal counterpart (see Tables S1 and S2 in the Supplementary Information).

Long hairpin (RNA fragment S2)

We have performed the same analysis as above also for the longer RNA fragment, a hairpin with several additional bulges (S2-wiss), illustrated in Figure 1. This fragment includes more unpaired nucleotides than the first fragment, consequently providing a hinge-like structure of considerable flexibility, which is a factor that contributes to the structural arrangement and substrate adsorption efficiency. Interestingly, despite the differences, we observe that the two distinct regimes of adsorption behavior described before essentially persist also for the longer RNA fragment S2 (as can be seen in Figures 4 and 6).

For the longer fragment S2, the value of $\varepsilon_{\mathrm{surf}}$ for which the interaction of the unstructured fragment S2-noss becomes more favorable than the structured S2-wiss is shifted towards larger values ($\varepsilon_{\mathrm{surf}}/k_{\mathrm{B}}T \approx 0.97$) when compared to the smaller fragment S1. Such a shift is consistent with a point of balance between more rigid (base-paired) and more flexible (loops) regions, where, for longer unstructured molecules, a higher conformational barrier is present. The value of the transition point between the two regimes depends on the shape of the secondary structure of RNA, as seen for fragments S1 and S2. Moreover, the differences in adsorption free energies between structured and unstructured RNAs in regime I grow with the fragment size (see snapshots of the monomers trapped at the adsorption interface in Figure 6(i) and (ii)). In other words, for longer and more rigid fragments, a stronger total adsorption is expected in regime I. In regime II, however, a stronger adsorption occurs for the unstructured fragments. The resulting PMF curves are displayed in Figure S6 in Supplementary Information, while the components of the radii of gyration are listed in

bioRxiv preprint doi: https://doi.org/10.1101/2021.08.31.458432; this version posted September 1, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

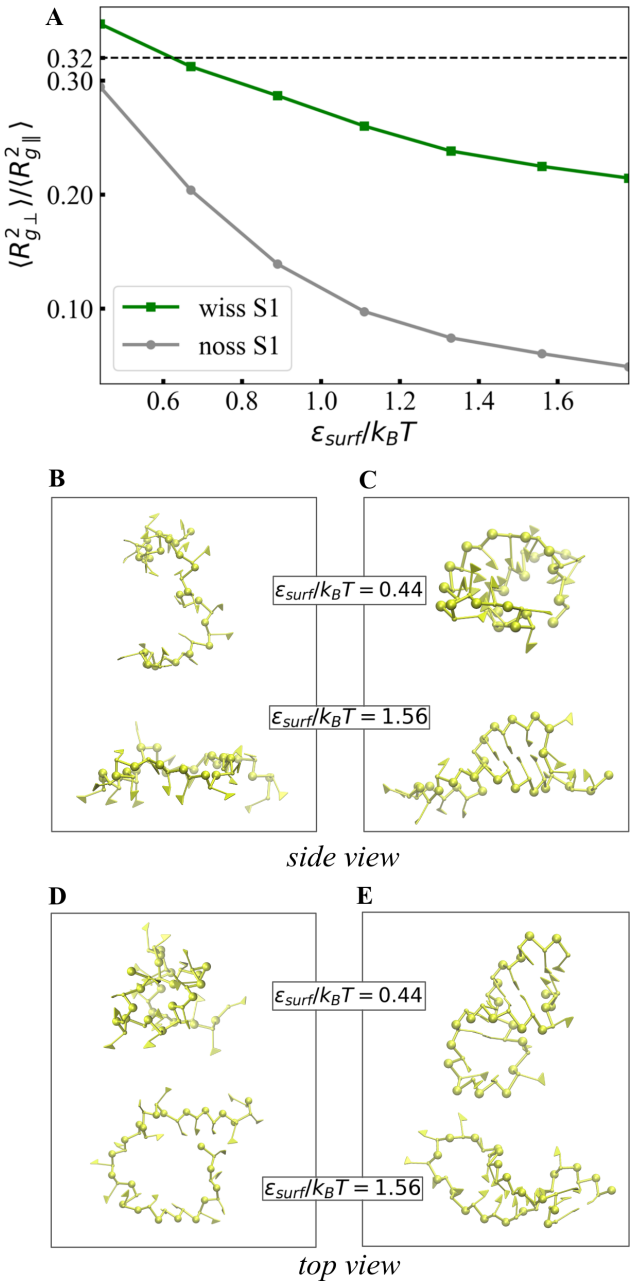

Figure 5: (A) Ratio of the normal and parallel radii of gyration $\langle R^2_{\mathrm{g}\perp}\rangle/\langle R^2_{\mathrm{g}\parallel}\rangle$ of RNA fragment S1 as a function of $\varepsilon_{\mathrm{surf}}/k_{\mathrm{B}}T$. Snapshots of the unstructured [(B) and (D)] and structured [(C) and (E)] fragments projected as side and top views with respect to the flat surface for given $\varepsilon_{\mathrm{surf}}/k_{\mathrm{B}}T$ values.

bioRxiv preprint doi: https://doi.org/10.1101/2021.08.31.458432; this version posted September 1, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

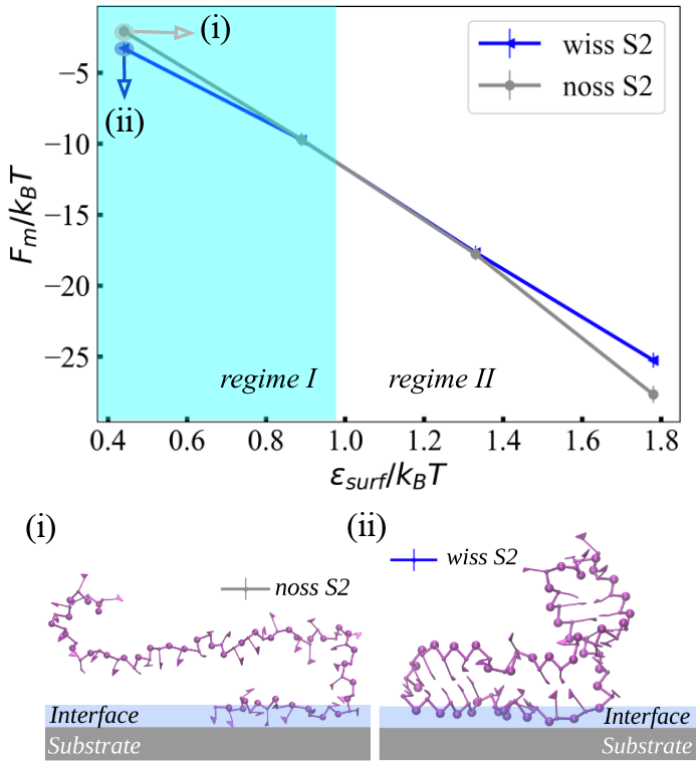

Figure 6: Adsorption free energy F_m as a function of the substrate attraction strength $\varepsilon_{\text{surf}}$ for RNA fragments S2-wiss and S2-noss. Snapshots of (i) unstructured and (ii) structured fragments at $\varepsilon_{\text{surf}} = 0.44\ k_{\text{B}}T$.

Tables S3 and S4.

Multibranch fragment S3 and qualitative scaling

The last fragment with secondary structure addressed in this work is a multibranch fragment (S3-wiss), which is comprised of two hairpins joined together by a single unpaired nucleotide hinge (see Figure 1). The normalized adsorption free energy of fragment S3 as a function of ε_{min} is shown in Figure 7(A). Full PMF as a function of distance from the surface is shown in Figure S6, while the components of the radii of gyration are shown in Table S5. Specifically, we show that the adsorption behavior of the multibranch fragment S3, once the results are normalized for the different number of nucleotides, almost overlap with fragments S1 and S2. Here, we note that base-pair fractions are comparable for all three fragments, namely 63% (S1), 60% (S2), and 61% (S3), and we can thus conclude such an overlap only for this

bioRxiv preprint doi: https://doi.org/10.1101/2021.08.31.458432; this version posted September 1, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

particular base-pair fraction.

Figure 7(B) shows the comparison between two simulated unstructured RNA fragments with different lengths and the theoretical prediction based on the continuum Edwards model of an infinite ideal chain with the same adsorption potential. As seen previously in Figure 4(A) and Figure 6, the adsorption becomes more favorable as the structures grow in size. Although the difference between structured and unstructured fragments is of the order of $k_{\mathrm{B}}T$ for low values of $\varepsilon_{\mathrm{min}}$, the adsorption properties, shown here for smaller and larger fragments, indicate that such a difference grows with the fragment size. This difference is therefore expected to become large for very long RNA fragments, for instance, in the case of the whole genome of a virus.

Moreover, for large $\varepsilon_{\mathrm{min}}$, the behavior of fragments with secondary structure resembles a straight line with a slope approximately given by $\varepsilon_{\mathrm{min}}/2$ (see Figure 7(A)). This can be rationalized by the fact that only half of the phosphates (monomers) are in direct contact with the surface because of the shape of the base-paired regions. On the contrary, fragments without secondary structure vary at a faster rate without reaching a linear regime, suggesting that more contacts may form, but the change in the shape of the molecule still plays a role in its adsorption. The result of the Edwards model for the infinitely-long ideal chain in Figure 7(B) displays essentially the same features as the ones observed in the simulations, although with a considerably larger entropy contribution, resulting in a slower variation of the free energy per monomer with the adsorption strength. Nevertheless, the decay is nonlinear, suggesting that a crossover between regimes I and II might persist, regardless of the system size.

Our results, based on the comparison between bulged-hairpin and multibranch RNA fragments show that the total fraction of base-pairs in the secondary structure determines the adsorption behavior in regime I rather than the exact topology of the RNA motifs and the additional freedom given by relative positions and orientations of the stems, at least when it comes to adsorption onto flat substrates. As for the difference of adsorption energies

bioRxiv preprint doi: https://doi.org/10.1101/2021.08.31.458432; this version posted September 1, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

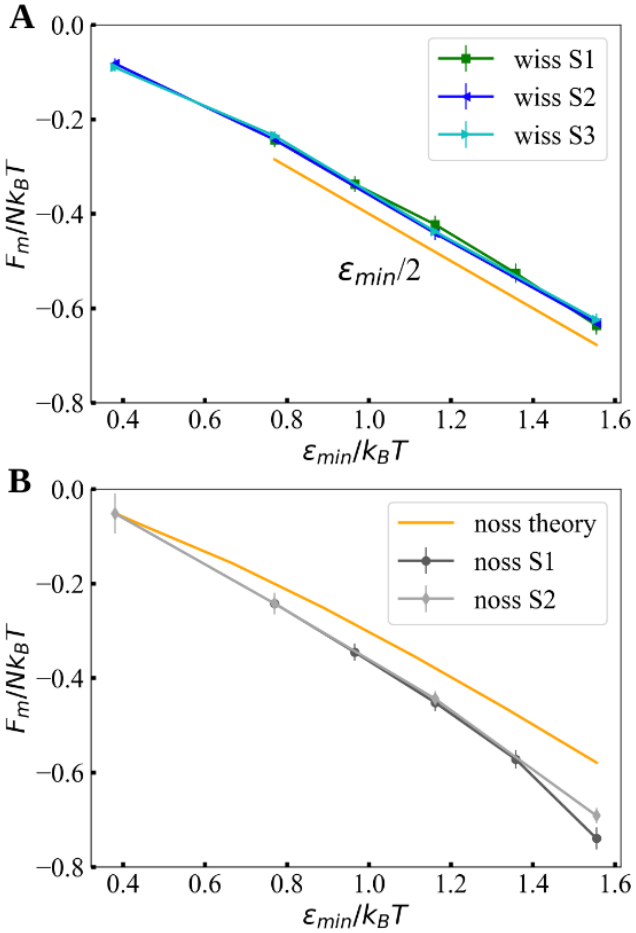

Figure 7: Adsorption free energy per nucleotide as a function of $\varepsilon_{\min}$ for the three fragments S1, S2, and S3 (A) with a secondary structure and (B) without a secondary structure. The fragments with a secondary structure show a linear tendency, shown by the slope $\varepsilon_{\min}/2$ (in orange). The gray curves correspond to the unstructured fragments and indicate similar trends. The orange curve in panel (B) corresponds to a one-dimensional Edwards model for a long ideal chain adsorbed onto a surface (for details, see Supplementary Information and Figure S7).

between structured and unstructured RNA fragments with the same sequence under regime I, it increases with the number of contacts. For instance, for a weak surface interaction with adsorption strength $\varepsilon_{\text{surf}}/k_{\text{B}}T$=0.44, the difference between the noss and wiss structure for the small hairpin (fragment S1) is $\Delta F_{\text{m}} \approx 1.0\ k_{\text{B}}T$, and for the longer bulged-hairpin (fragment S2) $\Delta F_{\text{m}} \approx 1.5\ k_{\text{B}}T$.

For stronger attractions in regime II, ΔF_{m} increases monotonically with the number of contacts (after the conformation barrier is overcome) and depends instead on the number of

bioRxiv preprint doi: https://doi.org/10.1101/2021.08.31.458432; this version posted September 1, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

contacts formed with the substrate (see Figures 4 and 6 as well as Figure S5 in Supplementary Information).

Conclusions

We have studied the role of RNA secondary structure in its adsorption to planar substrates. We use a simple and general coarse-grained RNA model interacting with the surface via a Debye-Hückel potential, which captures the basic electrostatic part of the dominant interaction in the adsorption processes on charged surfaces. Our investigation highlights the existence of two adsorption regimes concerning the RNA structure and the substrate attraction. In the first regime, operative at a weak surface attraction, base-paired segments of structured RNA behave as rigid objects and attach more easily to the substrate than unstructured, undulating RNA fragments (i.e., lacking any secondary structure). Increasing surface attraction leads to a second regime, which favors the adsorption of unstructured RNA fragments over structured ones. The existence of this turning point in adsorption does not depend on the exact nature of the interaction potential but appears to be more general. Namely, we have demonstrated that the Mie potential, based on van der Waals interaction, yields a similar behavior in adsorption. We have rationalized the origin of the two adsorption regimes as an interplay between the conformational entropy of the RNA fragments and the surface attraction. We complemented our simulations with an analytical scaling theory of the ideal chain, which provides deeper insight into the qualitative trends of adsorption of long polyelectrolytes as a function of substrate attraction strength. Based on the theoretical and simulation trends, we expect that the two observed adsorption regimes should persist even for very long RNA molecules (i.e., $\gtrsim 1000$ nucleotides in length).

Our results, which indicate a selectivity in adsorption between single- and double-stranded regions of RNA, underline the importance of RNA structure in regulating its adsorption to various substrates. We expect that the selective adsorption of one RNA structure over the

bioRxiv preprint doi: https://doi.org/10.1101/2021.08.31.458432; this version posted September 1, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

other could be experimentally controlled by tuning the interaction strength, for instance, by changing the salt concentration or pH. Moreover, the lower interaction free energies of unstructured RNAs compared to structured, double-stranded ones at high attractions suggest that possibly highly attractive surfaces may promote the unfolding of a double-stranded RNA structure. This would add an additional layer of complexity to the RNA-substrate interactions, differentiating the behavior of RNA and DNA information carriers when interacting with biological substrates. These observations can have crucial and far-reaching biological implications since RNA adsorption is linked to numerous biological functions, including phosphate positioning in virus shells, stability of RNA nanocarriers, and transfection of short RNAs.[25,44,49–51] A particularly noticeable example is RNA packaging in viral capsids, where it has been observed that unstructured RNA molecules can compete with the native RNA genome when simultaneously present in the solvent.[19,52]

In the future, it would be insightful to extend the studies to other (longer) RNA structures and different specific molecular substrate interactions and geometries. Further optimization of our code will allow us to extend and test those predictions. As for short RNA molecules, we already demonstrated that our computational model is versatile enough and generalizable to tackle more complex interaction potentials between RNA bio-macromolecules and molecular substrates.

We want to emphasize this point because several experimental techniques for secondary RNA structure characterization can be much improved by connecting them with reliable computational modeling when combined with high-throughput experimental data.[53] RNA structure modeling combined with computational methodology is thus quickly developing and reaching higher precision/resolution, with our computational efforts fitting nicely into this newly emerging paradigm, providing additional insights into the interpretation of phenomenology as well as all the way to improving the assessment of secondary structure candidates.

bioRxiv preprint doi: https://doi.org/10.1101/2021.08.31.458432; this version posted September 1, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Methods

RNA 3D structure modeling

To simulate the RNA fragments, we adopt a coarse-grained representation[54] that offers a good balance between computational efficiency and structural detail, as shown for the 3D domain reconstruction of the STMV genome.[55] In our model, we have implemented the potential of the flat substrate (Equations 1 and S1) and further analysis for the conformational sampling (see Supplementary Information for details). The resulting multiscale method is available in this work and hence could be further implemented in models/codes[56–60] of preference. Figure 2 illustrates the coarse-grained representation, where each RNA nucleotide is composed of an oriented particle with a virtual site which represents its nucleoside (sugar ring plus nitrogenous base) and of a point particle which represents the phosphate group. The interactions present in the simulations are built based on an X-ray structure database, which are the backbone connectivity (through bonds and angles) and excluded volume, while the structure of the stems is enforced by an energy restraint. The restraints take on a form of harmonic potential and are applied to the nucleotides belonging to a stem in such a way as to minimize the $\mathcal{E}$RMSD[14] metric with respect to a structure template, which is referenced to the A-form. The definition of the metric and the parameters used can be found in Supplementary Information.

RNA simulations of adsorption and sampling analysis

The potential of mean force (PMF) was obtained systematically by means of Umbrella Sampling[61] Monte Carlo simulations together with the Weighted Histogram Analysis Method (WHAM).[62] The employed Collective Variable (CV) was the minimum distance between the center of mass of the RNA molecule and the surface. In this specialized procedure, each distance is consistently sampled by an individual simulation using a harmonic restraint, which improves sampling and convergence of the PMF. Histograms of the distance d_{CM} were

bioRxiv preprint doi: https://doi.org/10.1101/2021.08.31.458432; this version posted September 1, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

used to define the PMF,

$$F(d_{\mathrm{CM}}) = -k_{\mathrm{B}}T \log P(d_{\mathrm{CM}}) \tag{3}$$

using a well-known WHAM implementation.[63] The error bars were estimated by the bootstrapping method after carefully determining the auto-correlation time of each Monte Carlo simulation by means of blocking analysis. Additional parameters of the simulations can be found in Supplementary Information. Tables S6, S7, and S8 in Supplementary Information contain details on the CV constraint parameters of each run. Histograms of the CV are given in Figures S8 and S9 in Supplementary Information, which highlight the quality of the systematic sampling used in our calculations. Further simulation scripts and analysis files can be found at https://zenodo.org/record/4646934.

Acknowledgement

H.V.G., A.B., and M.K. acknowledge financial support from the Slovenian Research Agency ARRS (Funding No. P1-0055 and J1-1701). S.P. acknowledges the project FONDECYT Iniciación en Investigación 11181334 for financial support.

Supporting Information Available

Detailed statistical and computational details can be found in Supplementary Information. Animated simulation snapshots for each of the RNA fragments are also provided in the Supplementary Information.

References

1. Strobel, E. J.; Watters, K. E.; Loughrey, D.; Lucks, J. B. RNA Systems Biology: Uniting Functional Discoveries and Structural Tools to Understand Global Roles of RNAs. *Curr.*

bioRxiv preprint doi: https://doi.org/10.1101/2021.08.31.458432; this version posted September 1, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Op. Biotechnology **2016**, *39*, 182–191.

2. Delebecque, C. J.; Lindner, A. B.; Silver, P. A.; Aldaye, F. A. Organization of Intracellular Reactions with Rationally Designed RNA Assemblies. *Science* **2011**, *333*, 470–474.

3. Dai, X.; Li, Z.; Lai, M.; Shu, S.; Du, Y.; Zhou, Z. H.; Sun, R. In situ structures of the genome and genome-delivery apparatus in a single-stranded RNA virus. *Nature* **2017**, *541*, 112–116.

4. Jasinski, D.; Haque, F.; Binzel, D. W.; Guo, P. Advancement of the Emerging Field of RNA Nanotechnology. *ACS Nano* **2017**, *11*, 1142–1164.

5. Manfredonia, I.; Nithin, C.; Ponce-Salvatierra, A.; Ghosh, P.; Wirecki, T. K.; Marinus, T.; Ogando, N. S.; Snijder, E.; vanHemert, M. J.; Bujnicki, J. M.; Incarnato, D. Genome-Wide Mapping of SARS-CoV-2 RNA Structures Identifies Therapeutically-Relevant Elements. *Nucleic Acids Research* **2020**, *48*, 1243612452.

6. Sahin, U.; Oehm, P.; Derhovanessian, E.; Jabulowsky, R. A.; Vormehr, M.; Gold, M.; Maurus, D.; Schwarck-Kokarakis, D.; Kuhn, A. N.; Omokoko, T.; Kranz, L. M.; Diken, M.; Kreiter, S.; Haas, H.; Attig, S.; Rae, R.; Cuk, K.; Kemmer-Brück, A.; Breitkreuz, A.; Tolliver, C. *et al.* An RNA Vaccine Drives Immunity in Checkpoint-Inhibitor-Treated Melanoma. *Nature* **2020**, *585*, 107–112.

7. Lipfert, J.; Skinner, G. M.; Keegstra, J. M.; Hensgens, T.; Jager, T.; Dulin, D.; Köber, M.; Yu, Z.; Donkers, S. P.; Chou, F.-C.; Das, R.; Dekker, N. H. Double-Stranded RNA Under Force and Torque: Similarities to and Striking Differences From Double-Stranded DNA. *Proceedings of the National Academy of Sciences* **2014**, *111*, 15408–15413.

8. Bao, L.; Zhang, X.; Shi, Y.-Z.; Wu, Y.-Y.; Tan, Z.-J. Understanding the Relative Flexibility of RNA and DNA Duplexes: Stretching and Twist-Stretch Coupling. *Biophysical Journal* **2017**, *112*, 10941104.

bioRxiv preprint doi: https://doi.org/10.1101/2021.08.31.458432; this version posted September 1, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

9. Marin-Gonzalez, A.; Vilhena, J. G.; Moreno-Herrero, F.; Perez, R. Sequence-Dependent Mechanical Properties of Double-Stranded RNA. *Nanoscale* **2019**, *11*, 2147121478.

10. Brion, P.; Westhof, E. Hierarchy and Dynamics of RNA Folding. *Annu. Rev. Biophys. Biomol. Struct.* **1997**, *26*, 113–137.

11. Tinoco, I.; Borer, P. N.; Dengler, B.; Levine, M. D.; Uhlenbeck, O. C.; Crothers, D. M.; Gralla, J. Improved Estimation of Secondary Structure in Ribonucleic Acids. *Nature New Biology* **1973**, *246*, 40–41.

12. Tinoco, I.; Bustamante, C. How RNA Folds. *Journal of Molecular Biology* **1999**, *293*, 271 – 281.

13. Mathews, D. H.; Sabina, J.; Zuker, M.; Turner, D. H. Expanded Sequence Dependence of Thermodynamic Parameters Improves Prediction of RNA Secondary Structure. *Journal of Molecular Biology* **1999**, *288*, 911 – 940.

14. Bottaro, S.; DiPalma, F.; Bussi, G. The Role of Nucleobase Interactions in RNA Structure and Dynamics. *Nucleic Acids Research* **2014**, *42*, 13306–13314.

15. Miao, Z.; Westhof, E. RNA Structure: Advances and Assessment of 3D Structure Prediction. *Annual Review of Biophysics* **2017**, *46*, 483–503.

16. poner, J.; Bussi, G.; Krepl, M.; Ban, P.; Bottaro, S.; Cunha, R. A.; Gil-Ley, A.; Pinamonti, G.; Poblete, S.; Jureka, P.; Walter, N. G.; Otyepka, M. RNA Structural Dynamics as Captured by Molecular Simulations: A Comprehensive Overview. *Chemical Reviews* **2018**, *118*, 4177–4338.

17. Hu, Y.; Zandi, R.; Anavitarte, A.; Knobler, C. M.; Gelbart, W. M. Packaging of a Polymer by a Viral Capsid: The Interplay Between Polymer Length and Capsid Size. *Biophys. J.* **2008**, *94*, 1428–1436.

bioRxiv preprint doi: https://doi.org/10.1101/2021.08.31.458432; this version posted September 1, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

18. Castellanos, M.; Pérez, R.; Carrasco, C.; Hernando-Pérez, M.; Gómez-Herrero, J.; de Pablo, P. J.; Mateu, M. G. Mechanical Elasticity as a Physical Signature of Conformational Dynamics in a Virus Particle. *Proceedings of the National Academy of Sciences* **2012**, *109*, 12028–12033.

19. Beren, C.; Dreesens, L. L.; Liu, K. N.; Knobler, C. M.; Gelbart, W. M. The effect of RNA secondary structure on the self-assembly of viral capsids. *Biophysical journal* **2017**, *113*, 339–347.

20. Eisenriegler, E.; Kremer, K.; Binder, K. Adsorption of Polymer Chains at Surfaces: Scaling and Monte Carlo Analyses. *The Journal of Chemical Physics* **1982**, *77*, 6296–6320.

21. Egorov, S. A.; Milchev, A.; Virnau, P.; Binder, K. Semiflexible Polymers Under Good Solvent Conditions Interacting With Repulsive Walls. *The Journal of Chemical Physics* **2016**, *144*, 174902.

22. Nikoubashman, A.; Vega, D. A.; Binder, K.; Milchev, A. Semiflexible Polymers in Spherical Confinement: Bipolar Orientational Order Versus Tennis Ball States. *Phys. Rev. Lett.* **2017**, *118*, 217803.

23. Milchev, A.; Binder, K. How Does Stiffness of Polymer Chains Affect Their Adsorption Transition? *The Journal of Chemical Physics* **2020**, *152*, 064901.

24. Netz, R. R.; Andelman, D. Neutral and Charged Polymers at Interfaces. *Physics Reports* **2003**, *380*, 1–95.

25. Schneemann, A. The Structural and Functional Role of RNA in Icosahedral Virus Assembly. *Annu. Rev. Microbiol.* **2006**, *60*, 51–67.

26. Stockley, P. G.; Ranson, N. A.; Twarock, R. A New Paradigm for the Roles of the Genome in ssRNA Viruses. *Future Virology* **2013**, *8*, 531–543.

bioRxiv preprint doi: https://doi.org/10.1101/2021.08.31.458432; this version posted September 1, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

27. Garmann, R. F.; Comas-Garcia, M.; Koay, M. S.; Cornelissen, J. J.; Knobler, C. M.; Gelbart, W. M. Role of Electrostatics in the Assembly Pathway of a Single-Stranded RNA Virus. *J. Virol.* **2014**, *88*, 10472–10479.

28. Hernandez-Garcia, A.; Kraft, D. J.; Janssen, A. F. J.; Bomans, P. H. H.; Sommerdijk, N. A. J. M.; Thies-Weesie, D. M. E.; Favretto, M. E.; Brock, R.; de Wolf, F. A.; Werten, M. W. T.; van der Schoot, P.; Stuart, M. C.; de Vries, R. Design and Self-Assembly of Simple Coat Proteins for Artificial Viruses. *Nature Nanotechnology* **2014**, *9*, 698–702.

29. Perlmutter, J. D.; Hagan, M. F. The Role of Packaging Sites in Efficient and Specific Virus Assembly. *J. Mol. Biol.* **2015**, *427*, 2451–2467.

30. Božič, A. L.; Micheletti, C.; Podgornik, R.; Tubiana, L. Compactness of Viral Genomes: Effect of Disperse and Localized Random Mutations. *Journal of Physics: Condensed Matter* **2018**, *30*, 084006.

31. Martínez, M.; Cooper, C. D.; Poma, A. B.; V Guzman, H. Free Energies of the Disassembly of Viral Capsids From a Multiscale Molecular Simulation Approach. *J. Chem. Inf. Model.* **2019**, *60*, 974981.

32. Zandi, R.; Dragnea, B.; Travesset, A.; Podgornik, R. On Virus Growth and Form. *Physics Reports* **2020**, *847*, 1 – 102.

33. Pardi, N.; Tuyishime, S.; Muramatsu, H.; Kariko, K.; Mui, B. L.; Tam, Y. K.; Madden, T. D.; Hope, M. J.; Weissman, D. Expression Kinetics of Nucleoside-Modified mRNA Delivered in Lipid Nanoparticles to Mice by Various Routes. *Journal of Controlled Release* **2015**, *217*, 345–351.

34. Lam, P.; Steinmetz, N. F. Delivery of siRNA Therapeutics Using Cowpea Chlorotic Mottle Virus-Like Particles. *Biomater. Sci.* **2019**, *7*, 3138–3142.

bioRxiv preprint doi: https://doi.org/10.1101/2021.08.31.458432; this version posted September 1, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

35. Chen, P.; Xu, Z.; Zhu, G.; Dai, X.; Yan, L.-T. Cellular Uptake of Active Particles. *Phys. Rev. Lett.* **2020**, *124*, 198102.

36. Xu, Z.; Gao, L.; Chen, P.; Yan, L.-T. Diffusive transport of nanoscale objects through cell membranes: a computational perspective. *Soft Matter* **2020**, *16*, 3869–3881.

37. Schlick, T.; Pyle, A. M. Opportunities and Challenges in RNA Structural Modeling and Design. *Biophysical Journal* **2017**, *113*, 225234.

38. Schroeder, S. J. Perspectives on Viral RNA Genomes and the RNA Folding Problem. *Viruses* **2020**, *12*, 1126.

39. Praprotnik, M.; Site, L. D.; Kremer, K. Multiscale Simulation of Soft Matter: From Scale Bridging to Adaptive Resolution. *Annual Review of Physical Chemistry* **2008**, *59*, 545–571.

40. Potestio, R.; Peter, C.; Kremer, K. Computer Simulations of Soft Matter: Linking the Scales. *Entropy* **2014**, *16*, 4199–4245.

41. Guzman, H. V.; Tretyakov, N.; Kobayashi, H.; Fogarty, A. C.; Kreis, K.; Krajniak, J.; Junghans, C.; Kremer, K.; Stuehn, T. ESPResSo++ 2.0: Advanced Methods for Multiscale Molecular Simulation. *Comput. Phys. Commun.* **2019**, *238*, 66 – 76.

42. Larman, B. C.; Dethoff, E. A.; Weeks, K. M. Packaged and Free Satellite Tobacco Mosaic Virus (STMV) RNA Genomes Adopt Distinct Conformational States. *Biochemistry* **2017**, *56*, 2175–2183.

43. Weidmann, C. A.; Mustoe, A. M.; Weeks, K. M. Direct Duplex Detection: An Emerging Tool in the RNA Structure Analysis Toolbox. *Trends in Biochemical Sciences* **2016**, *41*, 734–736.

44. Weidmann, C. A.; Mustoe, A. M.; Jariwala, P. B.; Calabrese, J. M.; Weeks, K. M.

bioRxiv preprint doi: https://doi.org/10.1101/2021.08.31.458432; this version posted September 1, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Analysis of RNA–protein networks with RNP-MaP defines functional hubs on RNA. *Nature Biotechnology* **2021**, *39*, 347–356.

45. Podgornik, R.; Harries, D.; DeRouchey, J.; Strey, H. H.; Parsegian, V. A. *Interactions in Macromolecular Complexes Used as Nonviral Vectors for Gene Delivery*; 2008; pp 443–484.

46. Caliskan, G.; Hyeon, C.; Perez-Salas, U.; Briber, R. M.; Woodson, S. A.; Thirumalai, D. Persistence Length Changes Dramatically as RNA Folds. *Phys. Rev. Lett.* **2005**, *95*, 268303.

47. Singh, N.; Willson, R. C. Boronate Affinity Adsorption of RNA: Possible Role of Conformational Changes1Presented in part at the 1998 International Symposium on Preparative Chromatography, Ion-Exchange and Adsorption/Desorption Processes and Related Techniques, Washington, DC, 31 May3 June 1998.1. *Journal of Chromatography A* **1999**, *840*, 205–213.

48. Theodorakis, P. E.; Mller, E. A.; Craster, R. V.; Matar, O. K. Modelling the superspreading of surfactant-laden droplets with computer simulation. *Soft Matter* **2015**, *11*, 9254–9261.

49. Dai, X.; Li, Z.; Lai, M.; Shu, S.; Du, Y.; Zhou, Z. H.; Sun, R. In situ structures of the genome and genome-delivery apparatus in a single-stranded RNA virus. *Nature* **2017**, *541*, 112–116.

50. Beren, C.; Cui, Y.; Chakravarty, A.; Yang, X.; Rao, A. L. N.; Knobler, C. M.; Zhou, Z. H.; Gelbart, W. M. Genome organization and interaction with capsid protein in a multipartite RNA virus. *Proceedings of the National Academy of Sciences* **2020**, *117*, 10673–10680.

51. Tomezsko, P. J.; Corbin, V. D. A.; Gupta, P.; Swaminathan, H.; Glasgow, M.; Persad, S.; Edwards, M. D.; Mcintosh, L.; Papenfuss, A. T.; Emery, A.; Swanstrom, R.; Zang, T.;

bioRxiv preprint doi: https://doi.org/10.1101/2021.08.31.458432; this version posted September 1, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

Lan, T. C. T.; Bieniasz, P.; Kuritzkes, D. R.; Tsibris, A.; Rouskin, S. Determination of RNA structural diversity and its role in HIV-1 RNA splicing. *Nature* **2020**, *582*, 438–442.

52. Thurm, A. R.; Beren, C.; Duran-Meza, A. L.; Knobler, C. M.; Gelbart, W. M. RNA Homopolymers Form Higher-Curvature Virus-like Particles Than Do Normal-Composition RNAs. *Biophysical journal* **2019**, *117*, 1331–1341.

53. Li, B.; Cao, Y.; Westhof, E.; Miao, Z. Advances in RNA 3D Structure Modeling Using Experimental Data. *Frontiers in Genetics* **2020**, *11*, 1147.

54. Poblete, S.; Bottaro, S.; Bussi, G. A Nucleobase-Centered Coarse-Grained Representation for Structure Prediction of RNA Motifs. *Nucleic Acids Research* **2017**, *46*, 16741683.

55. Poblete, S.; Guzman, H. V. Structural 3D Domain Reconstruction of the RNA Genome from Viruses with Secondary Structure Models. *Viruses* **2021**, *13*.

56. Boniecki, M. J.; Lach, G.; Dawson, W. K.; Tomala, K.; Lukasz, P.; Soltysinski, T.; Rother, K. M.; Bujnicki, J. M. SimRNA: a Coarse-Grained Method for RNA Folding Simulations and 3D Structure Prediction. *Nucleic Acids Research* **2015**, *44*, e63.

57. ulc, P.; Romano, F.; Ouldridge, T. E.; Doye, J. P. K.; Louis, A. A. A Nucleotide-Level Coarse-Grained Model of RNA. *The Journal of Chemical Physics* **2014**, *140*, 235102.

58. Bell, D. R.; Cheng, S. Y.; Salazar, H.; Ren, P. Capturing RNA Folding Free Energy With Coarse-Grained Molecular Dynamics Simulations. *Scientific Reports* **2017**, *7*, 45812.

59. Zhang, D.; Chen, S.-J. IsRNA: An Iterative Simulated Reference State Approach to Modeling Correlated Interactions in RNA Folding. *Journal of Chemical Theory and Computation* **2018**, *14*, 22302239.

60. Pasquali, S.; Derreumaux, P. HiRE-RNA: A High Resolution Coarse-Grained Energy Model for RNA. *The Journal of Physical Chemistry B* **2010**, *114*, 1195711966.

bioRxiv preprint doi: https://doi.org/10.1101/2021.08.31.458432; this version posted September 1, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted bioRxiv a license to display the preprint in perpetuity. It is made available under aCC-BY-ND 4.0 International license.

61. Frenkel, D.; Smit, B. *Understanding Molecular Simulation*, 2nd ed.; Academic Press, Inc.: USA, 2001.

62. Kumar, S.; Rosenberg, J. M.; Bouzida, D.; Swendsen, R. H.; Kollman, P. A. Multidimensional Free-Energy Calculations Using the Weighted Histogram Analysis Method. *Journal of Computational Chemistry* **1995**, *16*, 13391350.

63. An Implementation of WHAM: The Weighted Histogram Analysis Method. `http://membrane.urmc.rochester.edu/content/wham`.