Coarse-Grained Double-Stranded RNA Model from Quantum-Mechanical Calculations

Sergio Cruz-León,‡∥ Alvaro Vázquez-Mayagoitia,§ Simone Melchionna,‖ Nadine Schwierz,‡ and Maria Fyta‡∗

∥Institute for Computational Physics, Universität Stuttgart, Allmandring 3, 70569 Stuttgart, Germany
§Department of Theoretical Biophysics, Max Planck Institute of Biophysics, Max-von-Laue-Str. 3, 60438 Frankfurt, Germany
‖Argonne National Laboratory, 9700 S. Cass Avenue, Building 240, Argonne, Illinois, United States
‡Dipartimento di Fisica, ISC-CNR, Istituto Sistemi Complessi, Università Sapienza, P.le A. Moro 2, 00185 Rome, Italy

ABSTRACT: A coarse-grained model for simulating structural properties of double-stranded RNA is developed with parameters obtained from quantum-mechanical calculations. This model follows previous parametrization for double-stranded DNA, which is based on mapping the all-atom picture to a coarse-grained four-bead scheme. Chemical and structural differences between RNA and DNA have been taken into account for the model development. The parametrization is based on simulations using density functional theory (DFT) on separate units of the RNA molecule without implementing experimental data. The total energy is decomposed into four terms of physical significance: hydrogen bonding interaction, stacking interactions, backbone interactions, and electrostatic interactions. The first three interactions are treated within DFT, whereas the last one is included within a mean field approximation. Our double-stranded RNA coarse-grained model predicts stable helical structures for RNA. Other characteristics, such as structural or mechanical properties are reproduced with a very good accuracy. The development of the coarse-grained model for RNA allows extending the spatial and temporal length scales accessed by computer simulations and being able to model RNA-related biophysical processes, as well as novel RNA nanostructures.

INTRODUCTION

Despite the increase of available computer power in the last years, simulating complex biological systems at large scales often remains a computational challenge. One example of such complex systems involves nucleic acids forming biopolymers, which are the cornerstone of biology. A number of ongoing investigations try to understand physical and chemical properties of such biopolymers with the aim to manipulate and treat human diseases related to genetics and viral infections.1−4 From the theoretical perspective, understanding the chemistry and physics behind the nucleic acids at the atomic scale requires models that enable dynamic simulations at long temporal and spatial scales.

The ribonucleic acid (RNA) is responsible for performing vital processes within the cell, including transcription, mRNA, and protein synthesis, protein recognition, gene regulation, and others.5,6 RNA has been less investigated than the deoxyribonucleic acid (DNA), although in recent years, RNA has become the focus of many studies.6,7 RNA and DNA are long biopolymers that share many structural similarities.8 Both polymers form helical ladders composed of three building blocks: (1) a nitrogenous base, (2) a sugar, and (3) a phosphate group that builds up the helical structure. There are two main differences in the chemical composition of DNA and RNA that lead to variations in their geometrical, structural, and mechanical characteristics. The first difference is the sugar group; for DNA, an oxide-ribose (deoxyribose) is located in the backbone, whereas a ribose is present in RNA. The second difference is the set of nitrogenous base pairs (bps), which are as follows: adenine (A), thymine (T), guanine (G), and cytosine (C) for DNA, and they are canonically paired as A−T and G−C. In RNA, T is replaced by uracil (U), a molecule that regularly pairs with A. Both double-stranded DNA (dsDNA) and double-stranded RNA (dsRNA) can assume variations in their helical structures. For example, the incorporation of the 2′-hydroxyl group in the backbone of dsRNA generates an A-helix structure that is distinguishable from the B-DNA form. Both helices, dsRNA and dsDNA have a different rise per base pair, helix diameter, and helix pitch, and a different position of their helical axis.10 These differences result in variations in their properties, such as their mechanical stiffness. Many studies have indicated that the persistence length of DNA in buffers of moderate salt concentration is 45–
50 nm, whereas RNA has a persistence length around 60–65 nm under similar conditions. 

Both single-stranded RNA (ssRNA) and dsRNA are found in nature, ssRNA is more often involved in biological functions. Similarly to dsDNA, dsRNA is stabilized through the hydrogen bonding and stacking interactions of the nucleobases. RNA forms more complex tertiary structures. Some small separating fragments of RNA, which are not involved in coding for proteins, can be related to regulation of genes. These pieces can interfere with other RNA molecules, ending in silencing expression of genes. RNA-MARTINI, as well as its predecessor DNA-MARTINI, was derived following the Martini philosophy. On the other hand, RNA-MARTINI, as well as its predecessor DNA-MARTINI, was derived following the Martini philosophy.

Due to the huge potential of these RNA properties in biomedicine and biotechnology and the variety of potential relevant applications, a sustained effort to realize applications has been done. However, it is still a major computational challenge to simulate RNA structure and its properties efficiently. Biomolecules like DNA and RNA have been simulated with multiple approaches ranging from first-principles calculations to force-field schemes. Nevertheless, some force-field parametrizations are better suited to describe interactions with protein whereas others could incorporate qualitative solvent-free models or experimentally informed sequence-dependent coarse-grained models. Other models explicitly include electrostatics for DNA. Most of these models use empirical parameters, or intuitive assumptions are taken that strongly affect the transferability of the models.

Several coarse-grained models for RNA have been derived from different perspectives and using diverse optimization strategies. They can be categorized according to the parametrization philosophy into two groups: (i) top-down and (ii) bottom-up models. Within the first set of coarse-grained models, top-down models have been adjusted to reproduce various experimental evidence. Among such models one could find SPQR, OxnRNA, HiRE-RNA, and Xia’s model. For example, SPQR was designed to reproduce probability distributions of structural parameters sampled from structures obtained from the RNA 3D Motif Atlas. On the other hand, OxnRNA is an extension of oxDNA and was optimized to reproduce melting temperatures. By contrast, the Xia and HiRE-RNA were obtained from three-dimensional structures deposited in the Nucleic Acid Data Bank with different sizes and topologies. All of the top-down models rely completely on the availability of experimental data and suffer by the rather scarce amount of information at a given thermodynamic state. This limitation ultimately makes top-down models hardly transferable to generic thermodynamic conditions. The bottom-up approaches are mainly derived from atomistic simulations, some examples being the NATES-2P model or, more recently, the RNA-MARTINI. In NATES-2P, the interactions were obtained by the numerical integration of the energy landscape predicted by the AMBER force field. On the other hand, RNA-MARTINI, as well as its predecessor DNA-MARTINI, was derived following the Martini philosophy where the nonbonded interactions were tuned by potentials of mean force and free-energy calculations also using AMBER force field. The main issue inherent to this approach arises in the validity of underlying atomistic force fields where these coarse-grained approaches find their limits. As for top-down models, the derivation of coarse-grained potentials depends on the given thermodynamic state where the effective forces were parameterized. However, this limitation is somehow less restrictive when the force parametrization is derived from electronic structure calculations, that is, at virtually zero temperature. In fact, the modification of the forces by thermal effects can be incorporated a posteriori on the basis of physical considerations. At the same time, the form of the zero-temperature potential can be parameterized at high detail according to the local conformation of RNA, thus providing a stringent relation between structure and energetics.

In this work, we focus on extending a four-bead coarse-grained model for dsDNA to dsRNA, an approach entirely based on quantum-mechanical calculations and excluding empirical parameters, while reproducing structural and mechanical properties. The aim of this work is to provide a parametrization that can enable large-scale simulations of RNA. We focus here on dsRNA and present a complete parametrization for the Watson–Crick pairing in the A-helix form of RNA. The parametrization of one mismatched bp is made to illustrate how the present model can potentially manage the vast possibilities for noncanonical pairing. The model for a dsRNA in a B-helix shape is also used for the parameterization with the purpose of allowing for a comparison with the parameters of B-DNA and serve as a starting point for future development of ssRNA where a B-helix conformation is possible. Note that here, we provide a pathway for obtaining interaction parameters for these models but extended additional work is needed to obtain a complete model for nucleic acids.

This article is organized as follows: First, we briefly present the methodology used for the generation of the parameters for the two coarse-grained dsRNA; in the next section, the parameterization procedure for A-helix and B-helix of dsRNA is introduced. Later, in the sections Implementation of the Coarse-Grained dsRNA Model and Model Performance, the implementation of the four-bead coarse-grained scheme is given and validated, respectively. In the last section, we present the conclusions.

### METHODOLOGY

We have performed density functional theory (DFT) simulations on a set of configurations of nucleic basis pairs to obtain analytic energy functions. With this information, we parameterize a model to reproduce the interactions present in A-helix of dsRNA, as well as in its B-helix model. DFT has been proven a good method to study several properties of nucleic acids. The DFT calculations were carried out using the code SIESTA. For the relaxation, a tolerance in the force of 0.04 eV/Å was taken. Efficient pseudopotentials and triple-ζ basis plus polarization orbitals were computed with an energy cutoff of 100 Ry. All molecules considered were simulated as neutral molecules in vacuum. In this respect, environmental factors such as interaction with the solvent or the temperature were not incorporated in the individual DFT energies. Nevertheless, these effects will be taken into account during the implementation of the coarse-grained model for dsRNA.

The methodology followed includes the use of atomic-like orbitals, and the DFT energies had to be corrected using the counterpoise correction and basis set superposition error.
Four-Beads Coarse-Grained Model

The coarse-grained model maps the all-atom structure onto a four-bead model. Two beads represent one bp. One bead is located at the position of the N1 atom in U and C, and the second is located at the position of the N9 atom in A and G. Two additional beads located at the C1’ atom positions map the sugar—phosphate backbone group attaching to each “base-bead”. The atom notations, as well as the all-atom and coarse-grained model, are sketched in Figure 1, for both the A-helix and B-helix conformations. For these, we have simulated separately the different energy contributions and fitted analytic functions, leading to the parameterization of the four-bead model. We follow for this our efficient scheme for a four-bead dsDNA model. An important difference between the A-helix and B-helix is their helical axis. The B-helix axis points out of the plane at one-third of the distance between the C4 atom of the pyrimidine (U or C) and the C6 atom of the purine (A or G). The A-helix axis is shifted 4.0 Å outward from the bp level and the location of the axis of the B-helix (see Figure 1). Both helices have different features regarding the raise per bp, helix diameter, and average twisting angle.

Our quantum-mechanical simulations provide information regarding the contributions from hydrogen bonding ($E_{hb}$), stacking between neighboring bps ($E_{st}$), and backbone ($E_{bb}$) interactions. These interactions are decomposed into contributions with a physical meaning. The model assumes independence and additivity of these energy components, and the total energy for the coarse-grained model reads

$$E_{tot} = E_{hb} + E_{st} + E_{bb} + E_{el}$$

where $E_{el}$’s are the electrostatic interactions, which will be added for validating the model and will be discussed in Electrostatics section. For each term in eq 2, explicit functions were obtained and will be discussed in detail in the following.

Hydrogen Bonding. Explicit functions for the hydrogen bonding interactions of two complementary bps in eq 2 are obtained. For the dsRNA A-helix, parameters for both A–U and G–C bp are collected. For the B-helix, we have developed parameters for the A–U bp and taken those for the G–C bp from our previous work. For the hydrogen bonding energy, $E_{hb}$, three variables are taken into account: (i) $r_{hb}$ the vector connecting the N1–N9 atoms and representing the base–base distance, (ii) $\theta_{hb}$ the dihedral angle measuring the planarity between bases, and (iii) $\phi_{hb}^i$ with $i \in \{1, 2\}$, an angle estimating the in-plane base rotation. The index 1 corresponds to the pyrimidine (U or C) and 2 corresponds to the purine (A or G) rotations (see Figure 1).

Base–Base Distance Dependence. We begin the discussion with the dependence of the hydrogen-bond energy on the base–base distance $r_{hb}$. The relaxed geometry of the bp is

Figure 1. Definition of the coarse-grained model for dsRNA. The all-atom model of dsRNA for an A-helix (left) and a B-helix (right) form. A top and side view at the U–A bp level is shown in the upper two panels. Red, blue, green, and white colors represent oxygen, nitrogen, carbon, and hydrogen atoms, respectively. The two lower panels show the four-bead model on the same plane. The angles and distances shown on the figure correspond to those defining the equations of the coarse-grained model (eq 2). Note that the helical axis for both forms is located at a different position.

(BSSE). Accordingly, the corrected interaction energy ($\Delta E_{bind}$) between two molecules can be calculated through

$$\Delta E_{bind} = E_C - E_{ghost} - E_{ghost}$$

where $E_C$ is the total energy of both weakly bonded molecules and structurally relaxed in the complex and $E_{ghost}$ is the energy of the monomer 1 calculated on the ghost basis of monomer 2, similar for $E_{ghost}$. To obtain the four-bead coarse-grained model, the interactions within nucleic acids were split into three terms: (i) hydrogen bonding, (ii) stacking interactions, and (iii) backbone contributions. For (i), the Perdew–Burke–Ernzerhof exchange-correlation functional within the general gradient approximation was used. Stacking case (ii), the vdW-DF2 exchange-correlation functional was used. This can handle dispersion interactions that are essential among stacked nucleobases. The contribution of the backbone (iii) was obtained using the exchange-correlation functional for the local density approximation.
used as the starting point. The bases are incrementally separated along the vector \( r_{ab} \) as defined in Figure 1. The process for obtaining this energy dependence is represented through the atomic models on the top of Figure 2. For each

\[
\Delta E_{hb}(r_{ab}) = E_\theta(1 + a^*) e^{-a^*} \quad \text{with} \quad a^* = (r_{ab} - r_0)/l
\]

This expression has a minimum in the energy around \( r_0 \) with a value of \( E_\theta \). Fitting this function to the DFT data leads to the parameters listed in Table 1 for each bp. As mentioned earlier, the parameters for C−G in the B-helix are taken from the B-DNA model.53

**Dihedral Angle Dependence.** The dihedral angle \( \theta_d \) represents the angle formed by the planes of the purine (U or C) and the pyrimidine (A or G) bases. We perform rotations of the pyrimidine with respect to the purine, with the rotation axes along the vector \( r_{ab} \) to obtain the dependence of the binding energy on such a rotation. Calculations are done

For \( 0 \geq (\theta_d - \theta_0) \leq 120 \). For the B-helix form, the resulting DFT energies are fitted through

\[
E_{hb}(\theta_d) = E_\theta e^{k_1(\cos\theta_d-1)}
\]

Due to the structural differences in the A-helix, the relevant data cannot be fitted properly. Accordingly, for the A-helix, the following expression can better fit the DFT data

\[
E_{hb}(\Delta \theta_d) = E_\theta[1 + \alpha(\Delta \theta_d)^2 + \beta(\Delta \theta_d)^3] e^{k_3(\cos(\Delta \theta_d)-1)}
\]

For A-helix, the interaction energy is no longer a function of \( \theta_d \) but a function of \( \Delta \theta_d = \theta_d - \theta_0 \) since for the A-helix, the bps are slightly off plane compared with the B-helix model (see Figure 1). The DFT energies and the fits are shown in Figure 3. For the A−U bp in B-helix, the parameters in eq 4 are \( \theta_0 = 0^\circ, \alpha = 0.0\text{ deg}^{-2}, \beta = 0.0\text{ deg}^{-3} \), and \( k_3 = 0.93 \). \( E_\theta \) is the same as in the previous section. For A-helix, the parameters in eq 5 are shown in Table 1.
In-Plane Angle Dependence. In the following, the dependence of the hydrogen-bond energy on the in-planar angles \(\phi_{hb}^{(i)}\) is evaluated. These angles explore rotations of the bps in plane and are formed between the atoms H9, H11, and N1 for the pyrimidine (mapped through angle \(\phi_{hb}^{(i)}\)) and between the atoms H1, H9, and N9 (mapped through angle \(\phi_{hb}^{(0)}\)) for the purine, as in Figure 1. In the DFT simulations, we have rotated one bp around the hydrogen atom attached to the N1 and N9 atoms, as the rotation should map the rotation with respect to the backbone. The DFT energies arising from these rotations are fitted through the function

\[
E_{hb}^{(i)}(\Delta\phi_{hb}^{(i)}) = E_{hb}^{(a)}(\Delta\phi_{hb}^{(i)}) + E_{hb}^{(b)}(\Delta\phi_{hb}^{(i)})
\]

where the interaction energy is written in terms of the shifted angle \(\Delta\phi_{hb}^{(i)} = 1[\phi_{hb}^{(i)} - \phi_{hb}^{(0)}]\) and \(\phi_{hb}^{(0)}\) corresponds to the minimum energy geometry. \(E_{hb}^{(a)}\) is an exponential function, and \(E_{hb}^{(b)}\) is described by a harmonic spring. Both of them are damped by a Heaviside function \(\Theta(\Delta\phi_{hb}^{(i)})\) and are given through

\[
E_{hb}^{(a)}(\Delta\phi_{hb}^{(i)}) = E_0\left[e^{-\left(\Delta\phi_{hb}^{(i)}\right)^2/2\sigma^2} + \Theta(\Delta\phi_{hb}^{(i)})\right]
\]

\[
E_{hb}^{(b)}(\Delta\phi_{hb}^{(i)}) = E_0\left[-\frac{(\Delta\phi_{hb}^{(0)})^2}{2\sigma^2} + \Theta(\Delta\phi_{hb}^{(i)})\right]
\]

For clarity, only data for the A-helix and the rotation of U around A and of G around A are shown in Figure 4. For A and both RNA helices. Note that the fitting function does not capture the energy barrier shown in Figure 4. In principle, it is possible to use a different function to improve the fitting of the data in Figure 4. However, we have chosen to use the same fitting functionals for both dsDNA and dsRNA (this work), to have a complete and consistent force field with the same functional form. For correctly describing the barrier region, we apply constraints in the respective rotations in the coarse-grained simulations (see section Model Performance). The complete set of the fitting parameters for both RNA helices and all bps for this rotation is shown in Table 2.

Stacking Interactions. As a next step, the contribution of the binding energy arising from neighboring stacked bps, as shown in Figure 5, is evaluated. Here, two stacked bps are included in the DFT simulations. The stacking interactions are evaluated through two variables: (i) \(r_{st}\), the distance between the stacking bps measured along the helical axis and (ii) \(\theta_{st}\), the twisting angle. For the B-helix, \(\theta_{st}\) corresponds to rotations around the vector connecting the C4 and C6 atoms of the upper bp with respect to an axis along the C6–C4 atoms of the lower bp (see the sketch in Figure 5). For the A-helix, the local axis is shifted by 4.0 Å (see Figure 1).

Distance Dependence. The energy dependence with respect to \(r_{st}\) is obtained by varying the distance (in the range 2.5 and 4.8 Å) between the two stacked bps along the vector \(r_{st}\), as defined in Figure 5. For each case, we have calculated the interaction energies at the level of dispersion correction without further relaxation. The simulation data are summarized in Figure 6 together with the corresponding fits. The data are fitted using a Lennard-Jones-like potential

\[
E_{st}(r_{st}) = -\epsilon\left[6\left(\frac{r_m}{r_{st}}\right)^6 - 9\left(\frac{r_m}{r_{st}}\right)^4\right]
\]

where \(\epsilon\) and \(r_m\) are the fitting parameters. With this potential form, the minimum in the energy \(E_{st}(r_{st}) = \epsilon\) is found at a distance \(r_{st} = r_m\), which leads to the choice of the respective exponents (eq 8). The parameters obtained through fitting the data in Figure 6 with eq 8 are given in Table 3.

Table 2. Fitting Parameters for the In-Plane Rotation Described by Equation 6 and Parameters for the Base–Backbone Parameters<sup>4,6</sup>

<table>
<thead>
<tr>
<th>bp</th>
<th>(E_{0}) (eV)</th>
<th>(\phi_{hb}^{(0)}) (deg)</th>
<th>(\sigma^2) (deg)</th>
<th>(\theta_{st}^{(0)}) (deg)</th>
<th>(\theta_{st}^{(0)}) (deg)</th>
<th>(\kappa_{sb})</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>-0.734</td>
<td>51.72</td>
<td>6.203</td>
<td>83.72</td>
<td>68.16</td>
<td>1.92</td>
</tr>
<tr>
<td>U</td>
<td>-0.734</td>
<td>52.35</td>
<td>7.626</td>
<td>82.95</td>
<td>69.23</td>
<td>1.79</td>
</tr>
<tr>
<td>G</td>
<td>-1.390</td>
<td>50.17</td>
<td>18.45</td>
<td>83.41</td>
<td>67.51</td>
<td>2.16</td>
</tr>
<tr>
<td>C</td>
<td>-1.390</td>
<td>52.64</td>
<td>15.27</td>
<td>83.83</td>
<td>67.91</td>
<td>2.14</td>
</tr>
<tr>
<td>A’</td>
<td>-0.813</td>
<td>54.69</td>
<td>5.243</td>
<td>97.52</td>
<td>53.51</td>
<td>0.031</td>
</tr>
<tr>
<td>U’</td>
<td>-0.813</td>
<td>53.14</td>
<td>8.070</td>
<td>86.28</td>
<td>73.80</td>
<td>0.053</td>
</tr>
<tr>
<td>G’</td>
<td>-0.853</td>
<td>43.59</td>
<td>8.014</td>
<td>83.72</td>
<td>68.16</td>
<td>1.92</td>
</tr>
<tr>
<td>A</td>
<td>-0.843</td>
<td>44.03</td>
<td>10.67</td>
<td>83.67</td>
<td>67.51</td>
<td>2.16</td>
</tr>
</tbody>
</table>

<sup>4</sup>Section Base–Backbone Interactions, eq 12; \(k_{sb}\) is in units of \(10^{-2}\) eV/deg<sup>2</sup>. <sup>5</sup>The top two results are for the bps in the A-helix and the middle results are for A–U in the B-helix. The lower results: A’ and G’ are for the mismatched bp G–A (see section Noncanonical Base Pairing). The parameters for G and C in the B-helix are the same as those in our previous work.<sup>53</sup>
obtaining the bp distances, the vectors $\mathbf{d}_i$, $\mathbf{d}_j$, $\mathbf{d}_u$, and $\mathbf{d}_v$ are calculated (see Figure 5 for definitions). These vectors point to the distances between the N1, N9, C1, and C1′ atoms of adjacent base pairs. $\mathbf{d}_i$ is the vector pointing from the N1 atom of base $i$ to the N1$_{i+1}$ atom of the adjacent stacked base $i + 1$. The vectors $\mathbf{d}_j$, $\mathbf{d}_u$, and $\mathbf{d}_v$ are related to the distances between C1$′$ and C1$_{i+1}$, N9, and N9$_{i+1}$, and C1, and C1$_{i+1}$, respectively. The distance $r_{ij}$ is calculated as the average among all $z_p$ where $z_i = r_{ij}$.

$E(\theta_{tw}) = a_0 + \sum_{n=0}^{\infty} \left[ a_n \cos(n\theta_{tw}) + b_n \sin(n\theta_{tw}) \right]$  \(\theta_{tw}\) = twisting angle

The values for the Fourier coefficients and the different bps and conformations considered here are summarized in Table 4. For the calculation of the twisting interaction in the A-helix, a local rotation axis for a subset of two bps is defined. This local axis for the A-helix is different from the main rotation axis of the biomolecule. In the case of the B-RNA model (as in B-DNA), the local rotation axis coincides to the biomolecule axis. To define the local axis of rotation for the A-helix, the eigenvector of the inertia tensor perpendicular to the bp plane is used. This vector is located in the position shown in Figure 1 for the A-helix and points out of the bp plane. It is important to note that this axis definition is taken only for the

Table 3. Stacking Parameters for the Distance-Dependent Interactions of A- and B-Helix Model (Equation 8)$^a$

<table>
<thead>
<tr>
<th>bp</th>
<th>A-helix</th>
<th>B-helix</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>$r_{st}$ (Å)</td>
<td>$\epsilon$ (eV)</td>
</tr>
<tr>
<td>AU–AU</td>
<td>3.620 (2.702)</td>
<td>−0.450</td>
</tr>
<tr>
<td>GC–AU</td>
<td>3.880 (2.735)</td>
<td>−0.507</td>
</tr>
<tr>
<td>AU–UA</td>
<td>3.873 (2.813)</td>
<td>−0.487</td>
</tr>
<tr>
<td>GC–UA</td>
<td>3.889 (2.810)</td>
<td>−0.492</td>
</tr>
<tr>
<td>GC–CG</td>
<td>3.775 (2.808)</td>
<td>−0.717</td>
</tr>
<tr>
<td>AU–GC</td>
<td>3.945 (2.850)</td>
<td>−0.399</td>
</tr>
<tr>
<td>GA–GA</td>
<td>4.813 (3.159)</td>
<td>−0.809</td>
</tr>
<tr>
<td>GA–GC</td>
<td>4.763 (2.856)</td>
<td>−0.629</td>
</tr>
<tr>
<td>GA–GA</td>
<td>4.648 (2.894)</td>
<td>−0.664</td>
</tr>
<tr>
<td>GA–AU</td>
<td>4.367 (2.778)</td>
<td>−0.609</td>
</tr>
</tbody>
</table>

$^a$For the A-helix, $r_{st}$ correspond to the calculated distance. The values in parentheses are the average $z$ distance (along the axis perpendicular to a planar bp). For the B-helix, $r_{st}$ and the average $z$ distance are identical. The lower panel shows the results for the noncanonical bp G–A in A-helix form.
parametrization. The further implementation of our coarse-grained model does not require the definition of this axis.

**Influence of the Backbone.** An additional influence on the energetics within the dsRNA helices is related to the contribution due to the presence of a backbone. For this, two different variables need to be accounted for: (i) $r_{\text{CN}}$ the sugar–phosphate vector denoting stretching of the backbone along this direction and representing the contribution within the ribose–phosphate groups and (ii) $\theta_{\text{r}, \text{d}}$, $\theta_{\text{r}, \text{g}}$, the angles imposing changes to the $r_{\text{CN}}$ vector. The latter evaluate the binding energy as a function of the distance and angles between the base and the backbone. All variables are sketched in the insets of Figures 8—10.

**Interactions within the Sugar–Phosphate Backbone.** The interaction within the dsRNA backbone is considered as the interaction between the sugar and phosphates groups and is a measure of the stiffness of the backbone. To obtain an analytical expression for this interaction, a short backbone as in the inset of Figure 8 was taken. Simulations are performed by incrementally moving the atoms along the vector $r_{\text{NN}}$ compressing or stretching the backbone. The BSSE-corrected data, as well as the fitted function, are depicted in Figure 8. The analytical function used to fit these data is a harmonic potential centered at $r_{\text{NN}}$:

$$E_{\text{hh}}(r_{\text{NN}}) = c_2(r_{\text{NN}} - r_{\text{NN0}})^2 + c_3(r_{\text{NN}} - r_{\text{NN0}})^4$$  \hspace{1cm} (10)

For the A-helix, the fitting parameters are $c_2 = 5.10$ eV/Å², $c_3 = 17.82$ eV/Å⁴, and $r_{\text{NN0}} = 5.60$ Å, whereas $c_2 = 9.66$ eV/Å², $c_4 = 23.20$ eV/Å⁶ for the B-helix.

<table>
<thead>
<tr>
<th>bp</th>
<th>$a_1$</th>
<th>$a_2$</th>
<th>$a_3$</th>
<th>$a_4$</th>
<th>$b_1$</th>
<th>$b_2$</th>
<th>$b_3$</th>
<th>$b_4$</th>
</tr>
</thead>
<tbody>
<tr>
<td>AU-AU</td>
<td>35.72</td>
<td>1.258</td>
<td>3.920</td>
<td>3.506</td>
<td>0.117</td>
<td>0.152</td>
<td>0.236</td>
<td>8.989</td>
</tr>
<tr>
<td>GC-GC</td>
<td>18.59</td>
<td>0.880</td>
<td>1.656</td>
<td>0.125</td>
<td>5.283</td>
<td>5.890</td>
<td>7.672</td>
<td>17.06</td>
</tr>
<tr>
<td>GC-UA</td>
<td>22.46</td>
<td>0.860</td>
<td>1.910</td>
<td>0.000</td>
<td>4.999</td>
<td>4.265</td>
<td>6.726</td>
<td>1.940</td>
</tr>
<tr>
<td>GC-UA</td>
<td>25.39</td>
<td>0.856</td>
<td>1.910</td>
<td>0.000</td>
<td>4.999</td>
<td>4.265</td>
<td>6.726</td>
<td>1.940</td>
</tr>
</tbody>
</table>

Table 4. Fourier Coefficients for the Expression in Equation 9

All values are given in units of 0.01 eV.
Our results show an equilibrium distance \( r_{ss0} \) for the A-helix is in agreement to standard structural parameters obtained in previous studies. The position of a nucleobase is varied between 1.0 and 2.6 Å by translating it along the vector \( \vec{r}_{CN} \), and the variation of this distance is denoted by the blurred adenine replicas.

Backbone Interactions. The base–backbone interactions are further studied with respect to the variation of the base–backbone distance, \( r_{CN} \), and the flipping of the base with respect to the backbone is captured through the angles \( \theta_s \) and \( \theta_s' \). The position of a nucleobase is varied between 1.0 and 2.6 Å by translating it along the vector \( \vec{r}_{CN} \) as defined in the inset of Figure 9. In the same figure, the resulting DFT energies are fitted with a modified UBER expression:

\[
E_{bb-CN}(r_{CN}) = E_0(1 + a_1 + f_2 a_2 + f_3 a_3) e^{-a r_{CN}}
\]

For the dsRNA A-helix, fitting the respective DFT data resulted in the parameters \( E_0 = -4.689 \text{ eV}, r_0 = 1.464 \text{ Å}, f_2 = 0.044, f_3 = 0.084, \) and \( l = 0.391 \text{ Å} \). For the B-helix, these values are \( E_0 = -3.128 \text{ eV}, r_0 = 1.477 \text{ Å}, f_2 = -0.227, f_3 = 0.130, \) and \( l = 0.426 \text{ Å} \). The fits to the data are shown in Figure 9. These values are calculated by taking a section of the backbone with an adenine base. The results for other bases are almost identical.

For the base–backbone interactions, the flipping of the base with respect to the backbone has a significant contribution to the energetics. To account for this, we introduce the angles \( \theta_s \) and \( \theta_s' \), which are the angles between the vectors \( \vec{r}_{CN} \) and \( r_{CC3} \) and \( r_{CN} \) and \( r_{CC5} \), respectively. For this, we define \( r_{CN} \) as the backbone–base distance, which is the distance between the C1’ atom of the backbone and the N1 (or N9) atom of the nucleobase (see Figure 1). The vectors \( r_{CC3} \) and \( r_{CC5} \) point to the C1’ atom of the backbone sugar group and to the C1’ atom of the adjacent sugar group along the 5’–3’ and the 3’–5’ directions, respectively. The latter quantities are presented in the inset of Figure 10. These angles enter the base–backbone interaction as a function of angle \( \theta_s \). The lines are fits according to eq 12. The plots show the results for cytosine (upper left) and guanine (bottom left); both in the A-helix. The sketch on the right defines the angles \( \theta_s \) and \( \theta_s' \). The involved atoms of the nucleobase and the backbone are colored, whereas the blurred nucleobase denotes the rotation.

Noncanonical Base Pairing. Base pairing in RNA structures goes far beyond the Watson–Crick pairing. RNA
pairing geometry presents a rich variety,\textsuperscript{74} and many possible mismatches in RNA have been reported from crystallographic studies. An unambiguous and descriptive nomenclature with well-defined and nonoverlapping parameters has been proposed and describes concisely the structural information of pairs.\textsuperscript{75} Within this Leontis–Westhof classification, 12 basic geometric families have been found. This classification includes local orientation of the strand (parallel or antiparallel), the edge of interaction between nucleobases (Watson–Crick edge, Hoogsteen edge, and the sugar edge), and the bond glycosidic isomerism (cis and trans). According to Leontis–Westhof classification, the Nucleic Acid Database (NDB) presents an RNA base-pair catalog that reports around 150 bp possibilities.

From this large set of possible mismatches, one representative mismatch was chosen to assess the influence of mismatches in dsRNA. We have chosen the A–G cWW pair corresponding to a base pair with antiparallel local strand orientation, Watson–Crick/Watson–Crick interaction edges, and a cis glycosidic bond orientation, according to Leontis–Westhof classification. For simplicity, this mismatched bp is denoted in this paper just as A–G. This mismatch was chosen because A–G single mismatches are the most frequently occurring single mismatch type found in the secondary structure database.\textsuperscript{76} It is also the purine–purine base pairing with the largest occurrence in the NDB (191), and its structure was solved up to 1.2 Å resolution. To illustrate the inclusion in the model of non-WC bp, a complete set of parameters for A–G was derived and included in the dsRNA model. The parameterization for the mismatch was done according to the methodology presented in sections Methodology and Four-Beads Coarse-Grained Model. A complete set of parameters for the hydrogen bonding and stacking interactions was derived. In the following, the derivation of the parameters for the mismatch is presented.

For A–G, the initial structure was extracted from the Protein Data Bank, using the structure with identifier 1MIS. From this RNA structure, only the atoms of the nucleobases A–G were considered for the calculations. The A–G structure used for the parameterization is depicted in Figure 11. A geometry relaxation using DFT was performed on the noncanonical A–G bp. For the optimized geometry, the interaction energy was calculated and was BSSE-corrected according to eq 1. The value obtained was 18.97 kcal/mol (−0.823 eV). This interaction energy is in close agreement to the reported value (18.90 kcal/mol) in the benchmark database based on more accurate simulations such as MP2 and CCSD(T).\textsuperscript{77} In this mismatched pair, the O6(G)−N6(A) and N1(G)−N1(A) distances were calculated at 2.75 and 2.89 Å, respectively, with an error of less than 3% in both cases compared with more accurate simulations.\textsuperscript{77}

To include the presence of mismatches in our model, the assignment of the beads is naturally extended. For the noncanonical A–G bp, the beads that represent the nucleobases are located at atom N9 of each base and r_{hb} is defined as the vector connecting these beads (see Figure 11). All other definitions are analogous to the WC cases. The results on the hydrogen bonding dependence on the base–base distance, the in-planar angles, and the dihedral angle among bases are represented through the filled symbols and dashed lines in Figures 2, 3, and 4 respectively. The respective fitting parameters are summarized in Table 1 for base–base distance and dihedral angle dependence, whereas Table 2 shows the fitting values for the in-plane angle dependence. For the stacking interactions, the combinations GA–AG, GC–GA, and GA–AU were taken. For each of these, the stacking interactions were parameterized on the stacking distance and twisting angle. The same fitting functions as those for the WC bps were taken (see eqs 8 and 9), and the parameters are given in Tables 3 and 4. The respective data are depicted in Figures 6 and 7. Note that in all cases, the energy variations with respect to distance or angle for the mismatch follow the trends of the WC pairs. The main difference comes from the shifting of the minimum to larger distances in the case of the mismatch, which is expected to impose a local structural distortion in the helix diameter. In the hydrogen-bond energy, with respect to the in-plane angle (see Figure 4), a larger barrier was observed for the mismatch at an angle close to 40°, denoting that for certain angles, A and G come to close and need to impose a larger distortion to the molecule to overcome this rotation barrier. Finally, assuming the independence and additivity of the energy components in our model, no extra parameters are needed for the backbone interactions, as A and G are already parameterized.

■ IMPLEMENTATION OF THE COARSE-GRAINED DSrna MODEL

As a first step toward implementation of this model, the all-atom structures are mapped to the four-bead model and the interactions are described through eq 2. All parameters are calculated from the positions of the beads. In addition to the description in section Four-Beads Coarse-Grained Model, a normal vector n^{(i)} = r_{CC} × r_{hb} for each base (i = 1, 2) is calculated. This determines the dihedral angle (θ_{i}) using \cos \theta = \frac{n^{(i)} \cdot n^{(i)}}{|n^{(i)}||n^{(i)}|}. To avoid large in-planar angles, a constraint has been imposed on the model (see section In-Plane Angle Dependence). Our task was to test the performance and accuracy of the model. To realize that, we use simulate our coarse-grained model using the Langevin dynamics. Our code implementation makes use of the RESPA algorithm for time-step integrations in the NVE ensemble and the multiple time-step stochastic integrator for the Langevin dynamics.\textsuperscript{80} The performance of this algorithm was already tested for a XEON X5650 processor with a 100 ns simulation of 250 bps at 0.1 M salt concentration. The running time was about 200 min. This time is very short compared to an

![Figure 11. All-atom model of dsRNA for the A–G mismatched bp providing an extension of the definition of the coarse-grained model for dsRNA for noncanonical WC bps. The color coding is the same as in Figure 1. The definition of the bead position for the bases has to be modified from the definition given in section Four-Beads Coarse-Grained Model. In this case, we assign one bead for each of the N9 atoms in both bases, G and A.](image-url)
atomistic simulation of such a system.\(^{53}\) In this way, we can reach the goal of this work to realize large-scale simulations of RNA biomolecules. The good validation of our model, as described next, also confirms this.

**Electrostatics.** To perform further tests of the model and include the charges on the dsRNA, as well as the surrounding solution, we have used Coulomb potential to account for electrostatic interactions. The Coulomb potential is regulated through a distance-dependent dielectric function, which includes the screening effect produced by the ions in the medium. The dielectric function has the form\(^{53}\)

$$
\varepsilon(r) = \begin{cases} 
\varepsilon_{\text{int}} & \text{for } r < r_0 \\
\varepsilon_{\infty} e^{a(r-r_0)} & \text{for } r_0 < r < r_1 \\
\varepsilon_{\infty} e^{-\kappa r} & \text{for } r > r_1 
\end{cases}
$$

(13)

where \(\varepsilon_{\infty} = 78\) is the dielectric constant of water and \(\varepsilon_{\text{int}} = 3\) is the dielectric constant inside the helix.\(^{79}\) The values \(r_0\) and \(r_1\) define the boundary between unscreened and screened electrostatic interactions. In the Debye–Hückel theory of screening, the dielectric constant of water in bulk is recovered approximately at 5 times the mean water–oxygen–water distance so \(r_1\) is chosen as 13.0 Å. On the other hand, \(r_0\) is set equal to 4.0 Å, as the averaged size of the chemical groups represented by the beads of the model. The quantity \(\kappa\) corresponds to the inverse of the Debye length \(\kappa^{-1} = \sqrt{\frac{q^2 e_0 k_B T}{2 N_A e^2 l}}\). In this expression, \(e_0\) is the permittivity of empty space, \(k_B T\) is the thermal energy, \(N_A\) is Avogadro’s number, and \(l\) is the ionic strength.\(^{73}\) The Coulomb potential can then be described as

$$
E_d = \frac{e^2}{4\pi\varepsilon_0 \varepsilon(r) r}
$$

(14)

Note that in the coarse-grained model, a negative charge was placed in the position of the sugar site to simulate the polar group located on the phosphate group of the nucleic acids. Additional information on the electrostatics can be found elsewhere.\(^{53}\)

### MODEL PERFORMANCE

To validate our model, we have performed a number of simulations. A 350 bps long A-RNA molecule, with random bps, was simulated in a salt concentration of 0.1 M up to 1 μs, with a temperature of 300 K. The system is highly stable. Simulations were run on an Intel Xeon E5-2670 processor, spending around 76 CPU hours for the complete calculation. After relaxation, we could observe that helical structure is well preserved and the total energy fluctuations and its components are on the order of \(10^{-5}\) per site. During the simulations, we have calculated averages of structural parameters that can be directly compared with experimental reference data. The results were found in close agreement with relevant crystallographic values, as compared in Table 5. Overall, the structure of the A-helix RNA is well reproduced by our coarse-grained model.

**Persistence Length.** As an additional validation and accuracy test, we investigate the mechanical properties of the dsRNA molecules. As a representative quantity, we calculate the persistence length, which is a measure of the stiffness of polymers\(^{85}\) and is typically dependent on the salt concentration. It is well known that RNA is stiffer compared with DNA.\(^{12,13}\) For the latter, the persistence length is around 50 nm for a salt concentration around 0.1 M.\(^{86,87}\) For dsRNA, recent measurements with atomic force microscopy reported a persistence length of \(l_p = 63.8 \pm 0.7\) nm, whereas experiments with magnetic tweezers led to \(l_p = 62 \pm 2\) nm, both at moderate salt concentration.\(^{88}\) Our coarse-grained simulations of 350 bps long coarse-grained poly-A–U, poly-G–C, and random sequences in the RNA A-helix form in a salt concentration of 0.1 M resulted in persistence lengths of \(l_p = 69.7, 51.7,\) and \(58.3\) nm, respectively. These values reproduce the experimental data mentioned above within a very good accuracy. For a poly-GC with 30% of mismatches, the model predicts a persistence length of 72.8 nm. Unfortunately, we could not find relevant experimental data. Nevertheless, our model predicts that a certain amount of mismatches, although imposing local distortions in the dsRNA, does not effectively alter the mechanical properties of the biomolecule. In addition, the coarse-grained model prognosticates a sequence-dependent persistence length, which agrees with experimental results.\(^{85}\)

As mentioned before, the persistence length is strongly influenced by environmental factors,\(^{85}\) such as the ion concentration in the aqueous solution or equivalently the Debye length. Recently, persistence lengths under different ionic conditions have been reported for dsRNA\(^{89}\) and were found to decrease with increasing salt concentration. Similarly, the nonlinear Poisson–Boltzmann theory for uniformly charged cylinders predicts the persistence length to be proportional to the square of the Debye length according to

$$
l_p = l_b + \frac{1}{4 \kappa^2 l_b^3}
$$

(15)

where \(l_b\) is the Bjerrum length, which is the distance at which thermal fluctuations and electrostatic interactions are comparable.\(^{90}\) To investigate the behavior of the coarse-grained model for different salt concentrations, we have performed 1 μs long simulations for each molecule and calculated the persistence length through the decay of the orientation correlation function

$$
\langle \mathbf{n}(i) \cdot \mathbf{n}(j) \rangle = e^{-l_p / l_b}
$$

(16)

where \(\mathbf{n}(i)\) is the normal vector of the base \(i\), \(l\) is the arc length between the bps \(i\) and \(j\), and \(l_b\) is the calculated persistence length. Our results together with the estimated errors in our calculations are given in Figure 12 and fitted using eq 15. The solid lines in the figure correspond to the fitting, whereas the

<table>
<thead>
<tr>
<th>Table 5. Comparison of Average Simulation Results (sim.) on the Basis of Our Coarse-Grained Model with Reference Values (ref.) for Structural Parameters Used in the Model</th>
<th>RNA A-helix</th>
</tr>
</thead>
<tbody>
<tr>
<td>parameter</td>
<td>unit</td>
</tr>
<tr>
<td>(r_{\text{ho}})</td>
<td>Å</td>
</tr>
<tr>
<td>(\theta_{\text{w}})</td>
<td>deg</td>
</tr>
<tr>
<td>(r_{\text{e}})</td>
<td>Å</td>
</tr>
<tr>
<td>(r_{\text{w}})</td>
<td>Å</td>
</tr>
<tr>
<td>(r_{\text{CN}})</td>
<td>Å</td>
</tr>
<tr>
<td>(\phi_{\text{lb}})</td>
<td>deg</td>
</tr>
</tbody>
</table>

\(a\)The relative error between our results and those from the literature is also given.
To further interrogate our model, we have performed multiscale simulations that involve a continuum description of the salt solution solved by the lattice Boltzmann (LB) approach and a particle description of the RNA beads by the molecular dynamics scheme. Our multiscale methodology is able to reproduce the electrokinetics of the charged species in solution by providing the distinct response of the three-component system, neutral, positive, and negative charges, representing water, cations, and anions, respectively. As a result, one has access to the full set of hydrodynamic fields, that is, density, velocity, and stress tensor, together with the complete information related to RNA. In the past, the computational methodology has been developed and tested against known analytical solutions and compared to experimental data regarding the electrophoretic flow of polyelectrolytes and polymer translocation. Even in no-flow conditions, utilizing the multiscale methodology is important to verify the correctness of the conformational response and related fluctuations. We then simulated a 350 bps long random sequence RNA in its A-helix conformation at 300 K in a 0.1 M salt concentration and for 100 ns. Overall, the calculations produced similar conformations to the Langevin simulations obtained with the uniform dielectric background and static solvent. The RNA molecules remained stable and no significant deviation from the equilibrium bond lengths and bond angles in Table 5 were observed. The persistence length for 0.1 M was found to be 63.52 nm, in very good agreement with the Langevin model, as given in Figure 12. In the end, although our model should be tested in RNA-related electrokinetic processes, the agreement between the Langevin and multiscale computational methods is a very positive step toward its use in more complex situations.

Figure 12. Salt dependence of the persistence length at 300 K for A-helix conformations: poly-A−U, poly-G−C, and a random sequence. Solid lines are fits to the theoretical model in eq 15. The error bars are obtained from the statistics of the simulations. The results are based on the Langevin Dynamics simulations. The data point from the multiscale lattice Boltzmann (LB) model is also shown for comparison.

SUMMARY

In this work, we have developed a coarse-grained four-bead model for dsRNA. We have considered its A-helix and also parameterized a more unstable B-helix configuration. All energetic contributions within dsRNA were split into several contributions with a specific physical meaning. These include the hydrogen bonding, stacking, and backbone interactions. Each contribution is calculated using DFT simulations and fitted with a certain potential function. Overall, similar functional forms of the fit functions for four-bead coarse-grained dsDNA model and the dsRNA model could be used despite the different chemical and structural specificities of the two molecules. It was also shown, in which way noncanonical bps can be efficiently included in the coarse-grained model. However, a small modification that takes the nonplanarity of the bases into account and a different definition of the helical axis of A-RNA is introduced in the parameterization process. To account for the effect of electrostatics due to the presence of charges on the backbone of dsRNA and the solvent, a position-dependent dielectric function is included in the model. Our four-bead coarse-grained dsRNA model promotes large-scale simulations of dsRNA, while reproducing experimental structural values. In addition, mechanical properties such as the persistence length are in good agreement with recent measurements. The behavior of the dsRNA for different salt concentrations is in accordance with nonlinear Poisson–Boltzmann predictions for uniformly charged cylinders, as in previous studies. Multiscale simulations based on a mesoscopic solvent and an electrokinetic approach of our model could also reproduce our findings. It is very encouraging that a similar parametrization based on DFT can be used for DNA and RNA without relying on experimental data. The four-bead model for dsRNA is found to be computationally very efficient and easy to implement to reach large temporal and spatial scales and to model long biomolecules, processes these are involved in, as well as novel RNA nanostructures.

At this point, our dsRNA model seems to be very promising but would be inefficient in modeling single-stranded RNA, as single strands were not included in our parameterization. We have first presented a parametrization for double-stranded RNA, as this was more straightforward than that for single-stranded RNA (ssRNA). As ssRNA, as well as ssDNA, is also very essential in biophysical processes and in designing novel nanostructures, we aim to generalize our four-bead coarse-grained model also for these single strands. Such an extension could potentially deal with folding back processes, tertiary structure formation, or more complex architectures that are currently out of the scope of our model. Nonetheless, as with dsDNA, the parametrization for B-RNA can become useful as a starting point since ssRNA can form B-form-like structures. Further work should address solvent and ion-specific effects that are found to be important for the structure of RNA. We have delineated the pathway to developing a generic coarse-grained potential for nucleic acids. However, the amount of work for developing a comprehensive model for both DNA and RNA in different forms, including single and double strands and including base-pair mismatches, is substantial. Regarding the limitations of the current model, it should be made clear that the potentials developed in this work cannot adequately represent the backbone of a single-stranded molecule and they cannot capture transitions between conformations. Also, at the moment, it is not possible to
model transitions from one form of RNA to another, as the parameters for the backbone are targeted toward A-form helices and B-form helices separately since the structure was fixed in one or the other conformation. In further developments, it should always be checked how well the in vacuo quantum-mechanical calculations can describe the energetics in the presence of a solvent. In this study, we could show that our DFT calculations performed for two canonical base pairs in vacuo can be exported to a situation where the system is in an effective (Langevin) solvent. The difference of one hydrogen bond between AU and GC gives a significant energy difference that will remain qualitatively true even when solvent is added. This could probably not be the case for all noncanonical base pairs involving two hydrogen bonds and for which the energy difference depends on the conformational details. In these cases, the presence or absence of the solvent could make a significant difference.

Finally, we were able to develop an efficient (in terms of accuracy and computational performance) coarse-grained four-bead model for dsRNA. This builds upon dsDNA parameterization. Accordingly, we have unified a coarse-grained model for two of the most important nucleic acids, both of which are essential in biophysical processes. In addition, both RNA and DNA are the core blocks in producing novel and promising essential in biophysical processes. In addition, both RNA and DNA are the core blocks in producing novel and promising nanostructures. In these cases, the presence or absence of the solvent could make a significant difference.

REFERENCES

The Journal of Physical Chemistry B


